

NUCLEIC ACID RESEARCH AND MOLECULAR BIOLOGY

volume 75

Kivie Moldave

PROGRESS IN

Nucleic Acid Research
and Molecular Biology

Volume 75

PROGRESS IN

Nucleic Acid Research and Molecular Biology

edited by

KIVIE MOLDAVE

*Department of Molecular Biological Chemistry
University of California, Irvine
Irvine California*

Volume 75



ACADEMIC PRESS

An imprint of Elsevier, Inc.

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

This book is printed on acid-free paper. ♾

Copyright © 2003, Elsevier, Inc.

All Rights Reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

The appearance of the code at the bottom of the first page of a chapter in this book indicates the Publisher's consent that copies of the chapter may be made for personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per copy fee through the Copyright Clearance Center, Inc. (222 Rosewood Drive, Danvers, Massachusetts 01923), for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. Copy fees for pre-2003 chapters are as shown on the title pages. If no fee code appears on the title page, the copy fee is the same as for current chapters.

0079-6603/2003 \$35.00

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.com.uk. You may also complete your request on-line via the Elsevier Science homepage (<http://elsevier.com>), by selection "Customer Support" and then "Obtaining Permissions."

Academic Press

An Elsevier Imprint.

525 B Street, Suite 1900, San Diego, California 92101-4495, USA

<http://www.academicpress.com>

Academic Press

84 Theobald's Road, London WC1X 8RR, UK

<http://www.academicpress.com>

International Standard Book Number: 0-12-540075-6

PRINTED IN THE UNITED STATES OF AMERICA

03 04 05 06 07 08 9 8 7 6 5 4 3 2 1

This Page Intentionally Left Blank

Contents

SOME ARTICLES PLANNED FOR FUTURE VOLUMES	ix
--	----

Molecular Regulation, Evolutionary, and Functional Adaptations Associated with C to U Editing of Mammalian ApolipoproteinB mRNA	1
---	---

Shrikant Anant, Valerie Blanc, and Nicholas O. Davidson

I. RNA Editing: General Overview of Underlying Biochemical Mechanisms and Evolutionary Development	2
II. C to U RNA Editing	5
III. C to U ApoB RNA Editing: <i>cis</i> -Acting Elements	14
IV. C to U ApoB RNA Editing: <i>trans</i> -Acting Factors	17
V. Summary	33
Acknowledgments	33
References	33

Polymorphisms in the Genes Encoding Members of the Tristetraprolin Family of Human Tandem CCCH Zinc Finger Proteins	43
---	----

Perry J. Blackshear, Ruth S. Phillips, Johana Vazquez-Matias, and Harvey Mohrenweiser

I. Introduction	44
II. Materials and Methods	45
III. Results	52
IV. Discussion	63
Acknowledgments	66
References	66

Molecular and Cell Biology of Phosphatidylserine and Phosphatidylethanolamine Metabolism	69
--	----

Jean E. Vance

I. Membrane Phospholipids	70
II. Pathways for Biosynthesis of Phosphatidylserine and Phosphatidylethanolamine	71
III. Enzymes of Phosphatidylserine Biosynthesis	74
IV. Enzymes of Phosphatidylethanolamine Biosynthesis	82

V. Regulation of PS and PE Biosynthesis	85
VI. Interorganelle Transport of PS and PE	87
VII. Intramembrane Transport of PS and PE	92
VIII. Biological Functions of PS and PE	95
IX. Future Directions	97
References	98

DNA-Protein Interactions during the Initiation and Termination of Plasmid pT181 Rolling-Circle Replication 113

Saleem A. Khan

I. Introduction	114
II. Origins of Replication of the Plasmids of the pT181 Family	116
III. Initiator Proteins of the Plasmid pT181 Family	122
IV. Mechanism of Inactivation of the Plasmid pT181 Family Rep Proteins	126
V. Role of Host Proteins in pT181 RC Replication	129
VI. Replication of the Lagging Strand of Plasmid pT181	132
VII. Conclusions and Future Directions	133
Acknowledgments	133
References	133

Herpes Simplex Virus Type-1: A Model for Genome Transactions 139

Paul E. Boehmer and Giuseppe Villani

I. Introduction to <i>Herpesviridae</i>	140
II. HSV-1 Life Cycle, Genome Structure, and Replication Genes	140
III. Initiation of Replication	144
IV. Elongation	151
V. Interaction of the HSV-1 DNA Replication Machinery with DNA Damage	155
VI. Homologous Recombination	159
VII. Concluding Remarks	163
Acknowledgments	165
References	165

Enzymology and Molecular Biology of Glucocorticoid Metabolism in Humans 173

Andreas Blum and Edmund Maser

I. Introduction	174
II. Glucocorticoid Action	175

III. The Free Hormone Hypothesis	177
IV. Glucocorticoid Receptors	178
V. Nongenomic Action of Glucocorticoids	179
VI. Glucocorticoid Resistance	180
VII. 11β -Hydroxysteroid Dehydrogenase (11β -HSD) as Prereceptor Control	181
VIII. Hydroxysteroid Dehydrogenases: Physiological Role	182
IX. The History of Two Isoforms: 11β -HSD Type 1 and 11β -HSD Type 2	182
X. A (Patho)physiological Role for 11β -HSD 1 and 11β -HSD 2	183
XI. 11β -HSD 2 and the Syndrome of "Apparent Mineralocorticoid Excess" (AME)	184
XII. 11β -Hydroxysteroid Dehydrogenase Type 1 (11β -HSD 1)	187
XIII. Conclusions	203
References	203

Dancing with Complement C4 and the *RP-C4-CYP21-TNX* (*RCCX*) Modules of the Major Histocompatibility

Complex 217

C. Yung Yu, Erwin K. Chung, Yan Yang, Carol A. Blanchong,
Natalie Jacobsen, Kapil Saxena, Zhenyu Yang, Webb Miller,
Lilian Varga, and George Fust

I. Introduction	219
II. The Sophisticated Genetic Diversities of Human <i>C4</i> and <i>RCCX</i> Modules	221
III. Structural Diversities of the Human and Mouse <i>C4</i> Proteins	232
IV. Nucleotide Polymorphisms in Human <i>C4A</i> and <i>C4B</i> Genes	245
V. Expression of <i>C4A</i> and <i>C4B</i> Transcripts and Proteins	255
VI. The Endogenous Retrovirus that Mediates the Dichotomous Size Variation of <i>C4</i> Genes	260
VII. Evolution of Complement C3/C4/C5 and the <i>RCCX</i> Modules in the MHC	267
VIII. Afterthoughts: Eleven Outstanding Issues to be Addressed	274
Acknowledgments	276
References	276

The Roles and Regulation of Potassium in Bacteria . . . 293

Wolfgang Epstein

I. Paths of K^+ Movement	296
II. Regulation of K^+ Transport Systems	305

III. K As An Intracellular Activator	309
IV. Regulation of Internal pH by K^+	311
V. Future Directions	312
References	313

Conformational Polymorphism of $d(A-G)_n$ and Related Oligonucleotide Sequences 321

Nina G. Dolinnaya and Jacques R. Fresco

I. Introduction	321
II. α -DNA Helix \rightleftharpoons Acid-Dependent Duplex Equilibrium	323
III. Acid-Dependent Parallel Duplex \rightleftharpoons pH-Independent Parallel Duplex Equilibrium	332
IV. pH-Independent Parallel Duplex \rightleftharpoons Single-Hairpin Duplex \rightleftharpoons Two-Hairpin Tetraplex Equilibria	334
V. Concluding Remarks	342
Acknowledgments	344
References	344
INDEX	349

Some Articles Planned for Future Volumes

Clamp and Clamp Loaders in Eukaryotic DNA Metabolism

PETER M. J. BURGERS

Ccr4-Not Complex: A Regulatory Platform for Several Cellular Machineries

MARTINE A. COLLART AND MARC TIMMERS

DNA Methylation and the Regulation of Vertebrate Gene Expression

GORDON D. GINDER

The Yeast 2 Micron Plasmid: A Model for Optimized Molecular Selfishness

MAKKUNI JAYARAM, YURI VAZIYANOV AND SOUNDARAPANDIAN VELMURUGAN

Translational Control of Gene Expression by Hormones and Nutrients

LEONARD S. JEFFERSON AND SCOT R. KIMBALL

The Possible Origin of Features that Set Apart the Eukaryal and Prokaryal Replication Forks

GABRIEL KAUFMANN AND TAMAR NETHANEL

FGF3: a Gene with a Finely Tuned Spatiotemporal Pattern of Expression During Development

CHRISTIAN LAVIALLE

Specificity and Diversity in DNA Recognition by *E. coli* Cyclic AMP Receptor Protein

JAMES C. LEE

Oxygen Sensing and Oxygen-Regulated Gene Expression in Yeast

ROBERT O. POYTON

Ribonucleases in Cancer Chemotherapy

ROBERT T. RAINES, P. A. LELAND, M. C. HERBERT AND K. E. STANISZEWSKI

Broad Specificity of Serine/Arginine (SR)-Rich Proteins Involved in the Regulation of Alternative Splicing of Premessenger RNA

JAMES STEVENIN, CYRIL BOURGEOIS AND FABRICE LEJUNE

This Page Intentionally Left Blank

Molecular Regulation, Evolutionary, and Functional Adaptations Associated with C to U Editing of Mammalian ApolipoproteinB mRNA

SHRIKANT ANANT,*
VALERIE BLANC,* AND NICHOLAS
O. DAVIDSON*[†]

**Department of Internal Medicine; and*

*[†]Molecular Biology and Pharmacology,
Washington University School of
Medicine, St. Louis, MO 63110, USA*

I. RNA Editing: General Overview of Underlying Biochemical Mechanisms and Evolutionary Development	2
II. C to U RNA Editing	5
III. C to U ApoB RNA Editing: <i>cis</i> -Acting Elements	14
IV. C to U ApoB RNA Editing: <i>trans</i> -Acting Factors	17
V. Summary	33
Acknowledgments	33
References	33

RNA editing encompasses an important class of co- or posttranscriptional nucleic acid modification that has expanded our understanding of the range of mechanisms that facilitate genetic plasticity. Since the initial description of RNA editing in trypanosome mitochondria, a model of gene regulation has emerged that now encompasses a diverse range of biochemical and genetic mechanisms by which nuclear, mitochondrial, and t-RNA sequences are modified from templated versions encoded in the genome. RNA editing is genetically and biochemically distinct from other RNA modifications such as splicing, capping, and polyadenylation although, as discussed in Section I, these modifications may have relevance to the regulation of certain types of mammalian RNA editing. This review will focus on C to U RNA editing, in particular, the biochemical and genetic mechanisms that regulate this process in mammals. These mechanisms will be examined in the context of the prototype model of C to U RNA editing, namely the posttranscriptional cytidine deamination targeting a single nucleotide in mammalian apolipoproteinB (apoB). Other examples of C to U RNA editing will be discussed and the molecular mechanisms—where known—contrasted with those regulating apoB RNA editing. © 2003 Elsevier Science

I. RNA Editing: General Overview of Underlying Biochemical Mechanisms and Evolutionary Development

RNA editing is defined as a process through which the nucleotide sequence specified in the transcript is different from that in the genomically templated version. From a mechanistic standpoint, it may be viewed as the biochemical equivalent of a somatic mutation that targets nucleotides in RNA instead of DNA. RNA editing in different organisms employs a variety of genetic mechanisms, whose biochemical basis has been elucidated following the development of *in vitro* editing assays in which synthetic transcripts are incubated in the presence of cell or tissue extracts and changes in the targeted base examined. These technical advances have facilitated a mechanistic understanding of the precise modification in each case as well as the identification, in specific cases, of at least some of the factors involved. For purposes of understanding the underlying mechanisms, RNA editing is divided into two distinct classes, namely insertion–deletion and substitutional (reviewed in Refs. (1,2)). An understanding of these distinct reaction mechanisms illustrates the evolution and complexity of mammalian C to U RNA editing.

A. Insertion–Deletion RNA Editing

Insertion–deletion-type RNA editing was the original example of this form of genetic regulation. This model of transcript editing occurs in mitochondrial RNAs of kinetoplastid species and predominantly involves insertion (or occasional deletion) of varying numbers of uridine (U) residues into the RNA (3). These U insertions or deletions occur mostly within coding regions of mRNA and function to correct transcriptional frameshifts, thereby yielding translatable RNA templates (3). The mechanism involves a series of cleavage–ligation reactions catalyzed by a multicomponent protein complex that is appropriately targeted through base pairing with a guide RNA that permits editing to proceed in a 3' to 5' direction (3). Mitochondrial RNA editing in the slime mold *Physarum polycephalum* involves cotranscriptional nucleotide insertion, individually or as dinucleotides (CU, CG, GU), in a 5' to 3' direction close to the end of the nascent RNA (1). Insertional RNA-editing activity in mitochondrial extracts copurifies with large transcription elongation complexes (TECs) (4). The efficiency and specificity of insertional RNA editing is modulated by concentration of both the inserted nucleotide and its 5' neighbor as well as temporary pausing of the RNA polymerase component of the TECs, the latter serving to augment recognition of the editing site in the nascent transcript (5).

B. Substitutional RNA Editing

This second model of RNA editing occurs in both plants and mammals. Substitutional RNA editing in plants predominantly involves C to U transitions targeting mitochondrial and chloroplast mRNAs and certain tRNAs (1). In chloroplasts of higher plants approximately 30 editing sites have been described, all C to U and all confined to mRNAs (1). *In vitro* RNA-editing assays have established that synthetic transcripts containing ~30 nucleotides are accurately edited, suggesting that the *cis*-acting elements are located close to the edited site and are sufficient to support this biochemical modification (1). Although no conserved sequence elements have yet emerged and no consensus secondary structure has been found in these edited transcripts, studies in chloroplasts indicate that the identity of the closest neighboring nucleotide may be critical for editing (6,7). The development of an *in vitro* assay for chloroplast C to U editing has allowed refinement of the *cis*-acting requirements and has yielded two candidate genes, a site-specific RNA-binding protein (p25) that binds to a *cis*-acting element upstream of the edited base and a chloroplast ribonucleoprotein (RNP) (cp31) that may function as an adaptor protein (8). Extensive C to U changes have been proposed to exist in the mitochondrial RNAs of *Arabidopsis*, where over 6% of all cytosines in the coding region are edited (9), suggesting that RNA editing may be a more widespread mechanism for modifying gene expression than originally suspected.

1. SUBSTITUTIONAL RNA EDITING IN MAMMALS

RNA editing in mammals is exclusively of the substitutional type. Two types of substitutional RNA have been reported in mammals, namely adenosine to inosine (A to I) and cytosine to uracil (C to U) RNA editing (1,2). All forms of substitutional RNA editing result in an alteration of the amino acid sequence encoded by the edited mRNA, emphasizing the relevance of this model of genetic regulation in creating diversity at the level of the transcriptome. A to I RNA editing modifies pre-mRNA transcripts of the neuronal calcium-gated glutamate receptor (GluR) and serotonin (5HT) receptor-type 2c mRNA, as well as viral RNA such as hepatitis delta virus (2). A to I RNA editing is mediated through the enzymatic action of distinct members of a family of adenosine deaminases acting on double-stranded RNA (ADARs). At least three members of this gene family have been characterized, each with partially overlapping target specificity, suggesting that a range of potential targets exist for each member (1,2). A to I RNA editing proceeds via enzymatic deamination at the C6 position of the purine ring. The reading frame of the edited transcript is altered because the inosine residue is recognized as guanosine by the translational apparatus, making the net change

an A to G. The selectivity of A to I RNA editing depends upon an intronic sequence which can fold and undergo base pairing with an adjacent exonic complementary sequence, resulting in the formation of a double-stranded structure encompassing the editing site (2,10). This requirement for a double-stranded RNA template is an important distinction between A to I and C to U RNA editing since the former requires a pre-mRNA template containing intronic regions which form a double-stranded conformation with a complementary region of the exon containing the edited base. A to I RNA editing is thus confined to unspliced RNA transcripts. There is a second biochemical distinction between A to I and C to U RNA editing. Specifically, A to I RNA editing is mediated by ADAR gene products that are functionally competent as single recombinant proteins. ADARs do not exhibit a requirement for additional cofactors in order to mediate enzymatic activity since each protein contains both double-stranded RNA-binding domains as well as a deaminase domain which together permits them to act as modular-editing enzymes *in vitro* (2,10).

Three examples of C to U substitutional RNA editing have been described in mammals and each will be reviewed to provide a framework for a detailed discussion of the underlying mechanisms. These include the RNAs encoding apolipoproteinB (apoB), neurofibromatosis type 1 (NF1), and the novel target of apobec-1 (NAT1). ApoB RNA editing is the original and most extensively described form of C to U substitutional editing and will be the major focus of this review. NF1 RNA editing is less well understood, but the available information suggests that certain features may be shared with apoB RNA editing. NAT1 is a novel gene with homology to the translational repressor eIF4G and was discovered in a genetic screen for targets of apobec-1 (described in Section II.C).

In addition to these well-described examples, there are other reports of substitutional RNA editing in mammals, each involving U to C changes. These include RNA editing of the Wilms' tumor susceptibility gene (WT1) in rodent and human tissues, and U to C RNA editing of the mouse mitochondrial 16S transcript in certain tissues (11). In regard to the possible U to C editing of the WT1 RNA, an intriguing report almost a decade ago (12) indicated that a single nucleotide change was observed in both rat and human tissues at a conserved position (nucleotide 839) which resulted in a change in the amino acid sequence from leucine (CTC) to proline (CCC). U to C RNA editing modified ~30% of the endogenous WT1 transcripts and the change was further predicted to result in a loss of repressor activity with functional implications in both renal growth and malignant transformation (12). However, two more recent reports failed to confirm these earlier findings. In one report, a series of 15 Wilms' tumor samples was analyzed and revealed only the wild-type (CTC) sequence (13). A second report demonstrated less

than 1% RNA editing in rat and mouse kidney and no editing in a human sample (14). Our summary conclusions from these studies suggest that methodological or strain-specific differences may account for the divergent results. From a practical perspective, this uncertainty precludes any formal conclusions concerning the incidence, extent, or potential biological significance of WT1 RNA editing at this time.

Another example of U to C RNA editing was recently reported in which this modification occurred in a chimeric transcript containing 16S mitochondrial RNA joined by a linker sequence of 121 nt to the 5' end of ribosomal RNA (15). The workers found a single U to C change in the chimeric 16S RNA isolated from sperm and testis RNA, and further demonstrated both edited and unedited forms of the RNA in somatic tissues such as spleen and liver (11). The authors formally excluded such possibilities as cloning artefact, genetic polymorphisms, and amplification of a pseudogene, as well as RNA polymerase misincorporation. Although the mechanism of this apparent U-to-C-editing reaction was not addressed, potential explanations include transamination and transglycosylation of the targeted nucleoside base. Similarly, the function of the chimeric RNA remains unknown and the impact of this putative RNA-editing event awaits further study.

II. C to U RNA Editing

A. C to U RNA Editing of ApoB—An Overview of Mechanisms, Functional Biological Significance, and Evolutionary Physiology

The original descriptions of C to U editing in mammals centered on the finding that two isoforms of a protein involved in lipid metabolism, apoB, were encoded by a single gene (reviewed in Refs. (16,17)). ApoB is an abundant gene expressed in mammalian liver and small intestine where it plays an indispensable role in the mobilization, transport, and delivery of cholesterol and triglyceride from these tissues to cells throughout the body. Two forms of apoB are naturally found in the circulation. The larger form is referred to as apoB100 and represents the protein product of the unedited mRNA. In humans and higher primates, the liver produces exclusively apoB100. A smaller form is produced in the mammalian intestine as a result of RNA editing in which a CAA codon, specifying glutamine in apoB100 mRNA, is changed to UAA in the edited transcript (Fig. 1). Remarkably, this process targets a single nucleotide at position 6666 in apoB cDNA, which is located within a transcript of over 14,000 bases. C to U editing of apoB RNA exhibits

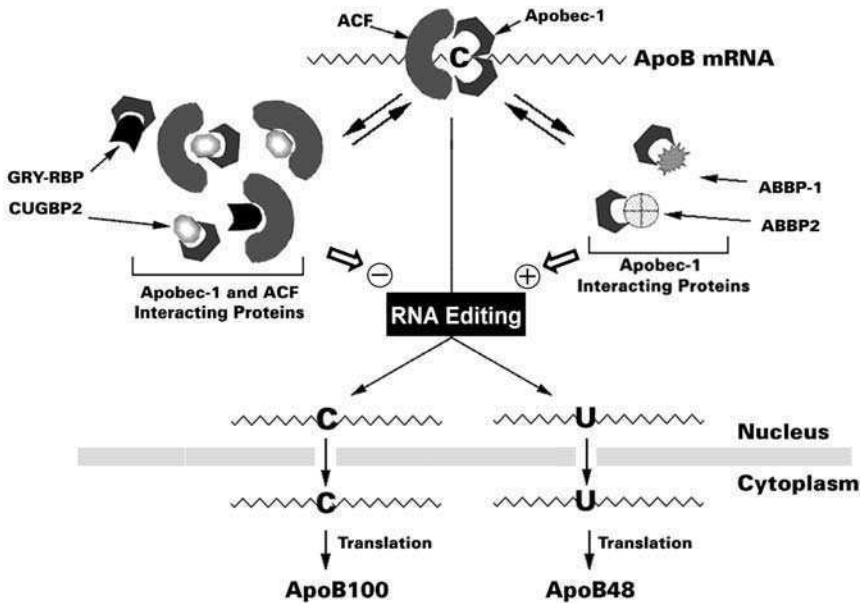


FIG. 1. C to U editing of apolipoproteinB mRNA. ApoB circulates in human plasma as two isoforms, apoB100 and apoB48, both encoded from a single *apoB* gene. In human liver, apoB pre-mRNA is spliced and transported to the cytoplasm and translated to form apoB100. In the human intestine, as well as in liver of some mammals such as rats and mice, the pre-RNA undergoes intranuclear splicing and RNA editing, resulting in the conversion of a CAA (glutamine) codon to a UAA, stop codon. Translation of the edited apoB mRNA generates apoB48. RNA editing is mediated by a holoenzyme containing apobec-1 and ACF, each of which binds to sequences flanking the targeted cytidine. Apobec-1 and ACF represent the core components of the editing holoenzyme. This core is in constant equilibrium with cellular factors that interact with either apobec-1 alone (ABBP-1, ABBP-2) or with apobec-1 and ACF (CUGBP2, GRY-RBP). These interactions modulate editing efficiency of the holoenzyme.

stringent requirements in both the *cis*-acting sequence elements and also in the protein components, which mediate this enzymatic modification (17). The core components of the apoB RNA-editing holoenzyme, each of which is *required and (in combination) sufficient* for *in vitro* editing, include apobec-1, the catalytic deaminase, and a competence factor, apobec-1 complementation factor (ACF).

In humans and all other mammals, apoB RNA editing occurs in the small intestinal enterocyte (17). The tissue-specific pattern of apoB RNA editing is accounted for by the expression of apobec-1, which in humans is usually confined to cells lining the gastrointestinal tract and is notably absent from the liver (18,19). Mice, rats, and certain other species, however, express

apobec-1 in the liver and these animals are capable of editing apoB RNA in hepatocytes, a functional adaptation described in more detail in the following paragraph (20). The edited apoB RNA specifies a UAA translational stop codon at position 2153 and encodes a truncated product referred to as apoB48—it is a peptide 48% the length of apoB100—and which is colinear with its amino terminus. As the major transport protein of cholesterol in plasma, apoB100 is responsible not only for the solubilization of lipid, but also for its targeted delivery to cells throughout the body. This latter function is accomplished through the interaction of apoB100 with a cell-surface receptor, the low-density lipoprotein receptor (LDLR). The ligand-binding domain in apoB100 responsible for interaction with the LDLR is located in the carboxyl terminus of the protein and is eliminated from apoB48 (16,21). Accordingly, apoB mRNA editing generates a protein that has no intrinsic capacity to interact with the LDLR. Lipid packaged into lipoprotein particles in association with apoB48 are secreted by the intestine as chylomicrons and are cleared by a distinct receptor pathway that recognizes another ligand on the surface of the particle, namely apoE (16). The emergence of apoB RNA editing is a central feature in mammalian lipid metabolism since it provides a dual ligand–receptor pathway to accommodate intestinal production of lipoproteins (i.e., apoB48) in response to dietary fat on the one hand, and the production of hepatic lipid and lipoproteins (i.e., apoB100) for distribution of cholesterol and other lipids throughout the body.

From an evolutionary perspective, the emergence of apoB RNA editing is a recent adaptation (20,22,23). Mammals absorb dietary lipid from the intestine and secrete chylomicron particles into lymphatic vessels that enter the peripheral circulation, and ultimately result in delivery of lipolytic products to the liver. Birds do not possess a lymphatic circulation and secrete intestinal lipid directly into the portal venous circulation for delivery to the liver (22,23). The evolutionary emergence of lymphatic transport of intestinal lipid thus roughly coincides with the appearance of apoB RNA editing. Reptiles and birds express but do not edit apoB RNA in any tissue, and as a consequence, assemble and secrete lipoproteins with apoB100 as the sole isoform (22,23). Among the factors that contribute to the absence of apoB RNA editing in birds is the sequence divergence in the RNA template and the absence of a catalytically active apoB RNA deaminase (22,23). More specifically, the chicken apoB RNA demonstrates sequence divergence from the conserved mammalian template in the region surrounding the edited base, and is thus not editable (Fig. 2) (23). In addition, the chicken homolog of apobec-1 does not mediate C to U editing of a mammalian apoB RNA (22). One suggestion arising from these observations is that apoB48 evolved as a more efficient mechanism to accommodate the absorption of dietary

		apobec-1 binding site (apoB)				Editing		
		Mooring Sequence				Canonical	Hyperediting	
apoB	Human	CUGCAGACAUUAUGAUA	*	AAUU	UGAUCAGUAUA	UUA	+	+
	Mouse	CUUGAGACAUACGCGAUA	C	AAUU	UGAUCAGUAUA	UUA	+	+
	Rat	CUUGAGACAUACGCGAUA	C	AAUU	UGAUCAGUAUA	UUA	+	+
	Chicken	CUUCAAGUGUAUCUGGUG	C	AAAU	AGAGCAGUACA	UCA	-	-
NAT-1/DAP5		AAACAUGGUUUACUUGCA	C	AAAAC	UGAUCAGUUUG	AGA	-	+
NF1	Human	CUCCUGGAUCCUUUAUUA	C	GAAUUG	UGAUCAUCCUC	UGA	-	+
	Mouse	CUCUUGGACCCUUUAUUA	C	GAGUCA	UCAUCUUCGGA	UUG	-	-
		apobec-1 binding site (NF1)						

FIG. 2. Sequence alignment of the apoB RNA-editing site with NAT1 and NF1. RNA sequences flanking the edited cytidine (asterisk) are depicted for apoB, NAT1/DAP5, and NF1. Their ability to serve as substrates for canonical C to U RNA editing or hyperediting is shown to the right. Sequences that differ from the murine apoB RNA are depicted by a grey background. The mooring sequence and apobec-1-binding sites in apoB RNA are bracketed above. The human, mouse, and rat apoB RNAs are grouped together because they undergo canonical C to U RNA editing. Chicken apoB RNA, containing multiple differences in the sequence from the mammalian apoB RNAs, is shown below. For NAT1/DAP5, the site that undergoes the highest levels of hyperediting is depicted. The human and mouse NF1 RNA sequence flanking the edited cytidine at nucleotide position 3916 is depicted along with the apobec-1-binding site.

lipid (predominantly triglyceride) from the intestine, and to direct receptor-mediated uptake to specific pathways in the liver that are adapted for high-capacity internalization of substrate. This suggestion is now experimentally testable. As will be detailed below, strains of mice have been generated in which apobec-1 has been disrupted by homologous recombination. These mice express apoB100 only (like birds), and consequently secrete apoB100-containing intestinal lipoproteins into their lymphatic compartment (like mammals). Detailed analysis of intestinal lipid transport in these animals and the consequences to chylomicron production and catabolism should be extremely illuminating.

There is yet another consequence of apoB RNA editing that has particular significance for human disease pathogenesis, namely the development of atherosclerotic heart and peripheral vascular disease. Humans and other species that are susceptible to the development of atherosclerosis are characterized by their predominant transport of plasma cholesterol in lipoproteins of the low-density class (LDL), in which the predominant protein is apoB100. LDL is cleared in humans through receptor-mediated uptake through the LDL receptor (LDLR). Mutations in this receptor result in familial hypercholesterolemia (FH), a disease characterized by high circulating levels of LDL and premature atherosclerosis. Mice and rats are naturally resistant to the development of atherosclerosis and, by virtue of the fact that hepatic apoB RNA is edited in these species, produce only small amounts of apoB100. These and other species in which hepatic apoB mRNA is edited have very low plasma levels of LDL (20). Accumulated epidemiological and direct experimental evidence has implicated plasma LDL cholesterol concentrations as a prime determinant of atherosclerosis in mammals, emphasizing the rationale for a multitude of therapeutic strategies to limit the accumulation of LDL particles in humans. By virtue of their low levels of apoB100, induction of experimental atherosclerosis in mice usually requires that the animals be placed on a diet rich in fat and cholesterol in order to develop the full manifestations of this disease. Mice in which apobec-1 was deleted through homologous recombination and which were then crossed into LDLR^{-/-} mice were found to have extremely high levels of LDL and of apoB100, and developed a pattern of spontaneous atherosclerosis similar to that observed in human FH while consuming a low fat, chow diet (24). These findings emphasize the importance of apoB RNA editing in cholesterol metabolism and lipoprotein transport in mammals, and suggest that its adaptive advantage to the organism may include the assembly and secretion of lipoprotein particles that, by virtue of their enrichment with apoB48, are intrinsically less atherogenic than particles containing apoB100. These suggestions are now experimentally testable using newer genetic crosses and should provide insight into these possibilities.

B. C to U RNA Editing of NF1

ApoB RNA editing demonstrates stringent requirements in the nucleotide sequences that flank the targeted base (Fig. 2). These sequence requirements have been mapped by several groups, the strategy for which was based on the observation that C to U RNA editing of apoB could be recapitulated using synthetic templates containing 30 nucleotides flanking the edited base (25–29). Emerging from the detailed characterization of the *cis*-acting elements required for C to U RNA editing of apoB (findings described in greater detail below), was the question of whether other transcripts might also undergo similar modification. Extensive database searches of transcripts containing similar *cis*-acting elements yielded a number of candidates demonstrating homology to the apoB-editing site. Based upon such a homology search, Skuse and coworkers identified the NF1 transcript as containing a potentially edited site, a cytidine base at nucleotide 3916 in the cDNA, positioned in a sequence context with ~50% identity to human apoB (Fig. 2) (30,31). C to U RNA editing of NF1 results in modification of a **C**GA (arginine) codon to a **U**GA, translational stop codon. Tumors from patients with NF1 manifested C to U RNA editing of the NF1 transcript, and follow-up studies suggested that the extent of RNA editing appeared to correlate with the malignancy grade of the tumor, more malignant tumors demonstrating higher levels of RNA editing (32). What might these findings imply for the biology of NF1 and how does this example fit with what is known about C to U RNA editing?

NF1 is a common genetic disease, affecting ~1 in 3000 individuals worldwide. It is also a cancer predisposition syndrome, which results from a combination of germline and somatic mutations in the neurofibromatosis tumor-suppressor gene product, neurofibromin (33). The tumor-suppressor function of neurofibromin is believed to reside in a region that contains intrinsic Ras-GTPase activating protein (GAP) activity and which is homologous to other Ras-GAP proteins conserved from yeast through mammals (reviewed in (33)). Loss of GAP activity leads to increased Ras activity, which in turn promotes cell proliferation and imparts a growth advantage (34). C to U RNA editing of NF1 is predicted to result in translational termination of the edited transcript, with elimination of the GAP domain in neurofibromin, thereby potentially contributing to the growth advantage demonstrated in these tumors. Thus, NF1 RNA editing could potentially represent a novel mechanism whereby the clinical manifestations of this disease may be conditionally modified. This becomes a relevant consideration in the context of understanding the clinical heterogeneity of NF1, which often yields marked differences in disease manifestations even within family members carrying identical germline mutations (33). Based upon these considerations, therefore,

a testable hypothesis is that C to U RNA editing of NF1 may function as a conditional genetic modifier of the neurofibromatosis tumor-suppressor gene by inactivating the remaining wild-type allele in subjects, who have inherited a germline mutation in the first allele. The mechanism of such conditional inactivation would reflect translational termination of neurofibromin in tumors that demonstrate C to U RNA editing of NF1. This said, the crucial piece of information yet to be provided is formal proof that the edited NF1 transcript encodes a truncated protein. An alternative possibility is that the edited mRNA, by virtue of the in-frame translational termination codon, becomes unstable and subject to nonsense-mediated decay (35). In either instance, the net effect would be to reduce production of the full-length neurofibromin tumor-suppressor gene product.

What is the relationship of this form of C to U RNA editing to that described for apoB? We have recently revisited the question of C to U RNA editing of NF1 in tumors from patients with neurofibromatosis. These latest findings suggest that C to U RNA editing occurs in only a subset of peripheral nerve sheath tumors and was undetectable in tissues from unaffected subjects. In exploring the features that distinguish tumors, in which C to U RNA editing was demonstrated, two defining characteristics emerged. First, tumors that demonstrated C to U RNA editing were all found to express apobec-1 mRNA, the catalytic deaminase of the apoB RNA-editing holoenzyme (31). This feature is considered particularly relevant since the underlying mechanism of NF1 RNA editing and its relationship, if any, to apoB RNA editing was previously unresolved. It bears emphasis that only those tumors, in which C to U RNA editing of NF1 was found, demonstrated the presence of apobec-1 mRNA. Secondly, alignment of the RNA sequence of human apoB and NF1 over a 40 nucleotide stretch encompassing their respective edited sites reveals ~50% identity, but more significantly revealed only 6 of 11 matches in a critical region (described in detail below), referred to as the “mooring sequence” (Fig. 2) (26,28). This degree of mismatch in the context of apoB mRNA would eliminate C to U RNA editing (26–28). Accordingly, we reasoned that there might be compensatory alterations in either the primary sequence of the NF1 transcript or in the secondary structure of the relevant region that permits presentation of the targeted cytidine in an appropriate configuration relative to the active site of the deaminase, apobec-1. These studies revealed the presence of a consensus binding site for apobec-1, UUUN[A/U]U, located immediately adjacent and 5′ to the edited base in NF1 RNA (Fig. 2) (31,36). This consensus binding site in apoB RNA is located 3′ of the edited base and spans the 5′ end of the mooring sequence (36). In addition to the consensus binding site for apobec-1, we demonstrated that C to U editing of NF1 RNA was confined to transcripts in which an alternatively spliced downstream exon was included in the template (31). Specifically,

a 63-nucleotide exon, 23A, located 3' of the edited base (23-1), was required for editing of cytidine 3916 in NF1 cDNA (31). Analysis of multiple tumors derived from different patients demonstrated a significant correlation between the extent of C to U RNA editing and the relative abundance of the alternatively spliced exon 23A (31). Modeling predictions further supported the hypothesis that the presence of the alternatively spliced downstream exon 23A in NF1 RNA allows the transcript to fold in a manner that resembles the stem-loop/bulge appearance of apoB RNA. This folding is predicted, in turn, to allow presentation of the targeted cytidine in a favorable configuration relative to apobec-1. Thus, the implications arising from these latest studies point toward NF1 RNA editing, representing an opportunistic outcome of several discreet events, including aberrant expression of apobec-1 and differential alternative splicing of the NF1 transcript (31). Even under these conditions, however, the extent of C to U editing of this transcript, generally less than 20%, is considerably less efficient than observed with the canonical substrate, apoB, which in the small intestine is typically $\sim 90\%$ (30,31,37). This difference in editing efficiency may be accounted for by sequence divergence in the immediate vicinity of the targeted base in NF1 compared to the conserved sequence in apoB RNA.

C. C to U RNA Editing of NAT1

Yet another example of C to U RNA editing was discovered some years ago during the course of studies aimed at understanding the biological functions of apobec-1. Innerarity and coworkers generated several lines of transgenic mice expressing high copy numbers of rabbit apobec-1 driven by a liver-specific promoter (38). These animals demonstrated extensive C to U RNA editing of apoB, not only at the canonical site (nucleotide 6666), but also at multiple cytidines predominantly downstream of the canonical site (39). Although editing at nucleotide 6802 in apoB had been previously reported to occur at low efficiency ($\sim 10\%$) (40), the findings from transgenic overexpression of apobec-1 represented the first example of C to U editing of apoB RNA at multiple sites and at high efficiency. Remarkably, virtually all of the animals expressing high levels of apobec-1 developed hepatic dysplasia and carcinomas, findings replicated in transgenic rabbits with similarly high-level forced overexpression (38). Although the mechanism underlying the cellular dysplasia and malignant transformation is yet to be fully explained, these workers proposed that promiscuous editing of other transcripts may play a significant role in the process.

Using a novel differential display strategy, Innerarity and coworkers went on to demonstrate that a novel transcript, identified as novel apobec-1 target (NAT1), contained multiple sites throughout the transcript in which C to U RNA editing had occurred (41). NAT1 is a gene whose biological significance

is poorly understood. The primary sequence is homologous to the translational repressor protein eIF4G, and it is presumed on this basis that NAT1 has important functions in the regulation of protein translation (41). More recent studies have demonstrated that targeted deletion of NAT1, using homologous recombination, results in embryonic lethality, and suggest that the gene product functions in an early stage of development that is, at least, partially retinoic acid sensitive (42). Specifically, embryonic stem cells derived from NAT1^{-/-} donor mice were resistant to the differentiation-inducing effects of retinoic acid (42). Other lines of evidence, however, indicate that NAT1 may operate in a retinoic acid-independent manner with effects mediated through alterations in growth and apoptosis (42). Taken together, the data suggest that C to U RNA editing of NAT1 may account for a loss of function phenotype in situations, where apobec-1 is either overexpressed or unconstrained. As proof-of-principle of the hypothesis of conditional gene inactivation, C to U RNA editing of NAT1 was demonstrated to introduce multiple translational termination codons into the reading frame, alterations shown in turn to result in the production of a truncated and presumably nonfunctional protein (41).

Certain features of C to U RNA editing of NAT1 are intriguing and additionally raise interesting questions germane to our understanding of the regulation of apoB RNA editing. The original description of NAT1 RNA editing in the I20 line of apobec-1 transgenic overexpressing mice demonstrated that cytidine deamination occurred throughout the NAT1 transcript, without the stringent requirements for an appropriate *cis*-acting context which characterizes apoB RNA editing (41). These findings were interpreted to suggest that under certain defined circumstances, where apobec-1 was greatly overexpressed in relation to either the substrate (i.e., apoB RNA) or other components of the editing enzyme (i.e., ACF or potentially others), the catalytic site of the deaminase was less discriminating in selecting target cytidine nucleotides. Innerarity and coworkers generated other lines of mice, in which the level of apobec-1 expression was more modest than in the first set of experiments, in order to address the possibility that the level of apobec-1 expression per se was an important consideration in the editing of NAT1. These latter animals (RE4) demonstrated increased levels of apoB RNA editing (~90%), which was confined to the canonical site and did not manifest editing of NAT1, suggesting that promiscuous editing is largely, if not exclusively, dependent upon high levels of apobec-1 expression (43).

More recent findings, however, have challenged the exclusivity of this conclusion. Specifically, recent studies by Hersberger and Innerarity have demonstrated an interesting paradox revealed through genetic experiments performed in the background of apobec-1^{-/-} mice. As alluded to above, apobec-1^{-/-} mice do not edit apoB RNA (44,45). Hersberger and Innerarity

introduced a tetracycline-inducible apobec-1 transgene into the background of apobec-1^{-/-} mice in order to characterize the effects of conditional reintroduction of apobec-1 on C to U RNA editing of apoB and other targets (46). These studies revealed that reintroduction of apobec-1 resulted in increased C to U editing of apoB RNA, at both the canonical and also other downstream sites in the apoB template. In addition, these downstream editing events took place under conditions of apparently modest apobec-1 expression as evidenced by the observation that there was less than 30% editing of apoB RNA at the canonical site (46). These latter findings are significant in view of the fact that wild-type mice generally edit at least 50–60% of the endogenous hepatic apoB RNA, yet do not edit NAT-1. These observations imply that the level of apobec-1 expression, as deduced from the extent of editing at the canonical site, is alone insufficient to account for the presence of promiscuous editing sites in the template. Recall that the modest transgenic overexpressing line (RE4) of Qian *et al.* demonstrated ~90% editing of apoB RNA at the canonical site, yet no C to U editing elsewhere (43). Hersberger and Innerarity have now demonstrated that conditional apobec-1 expression produces less than 30% editing of the endogenous apoB RNA, yet extensive editing of NAT1, at multiple sites throughout the transcript (46). These findings imply that considerations, other than simply the level of expression of apobec-1, must factor into whether other targets undergo C to U editing.

At present, these findings reveal an unresolved paradox in terms of the regulation of apoB RNA editing and the induction of C to U RNA editing in NAT1. On the one hand, unconstrained expression of apobec-1 is likely to result in promiscuous editing of targets beyond the canonical base in apoB. On the other hand, apobec-1 expression is normally regulated during embryonic development and in response to certain hormonal and nutritional cues, with large changes noted in the extent of C to U editing of apoB in the tissues involved (37,47–56). These various examples of physiological regulation of apoB RNA editing have not been associated with the appearance of editing of other sites in apoB RNA, nor with the appearance of NAT1 editing (39,41). The resolution of this paradox will likely require greater understanding of the compartmentalization of apobec-1 and the other *trans*-acting factors that regulate its biological activity.

III. C to U ApoB RNA Editing: *cis*-Acting Elements

ApoB mRNA editing is an enzymatic modification of a single cytidine in a large transcript spanning greater than 14,000 residues. In attempting to understand the sequence requirements for this process, the minimal effective

transcript size that supports *in vitro* C to U editing was demonstrated to be contained within ~ 30 nucleotides flanking the edited base (Fig. 2) (26–28), although greater efficiency is encountered when more distant elements, located ~ 150 nucleotides both 5' and 3' of the edited base, are included (29,57). The region surrounding the edited base is enriched in A + U residues, a feature that is again highly conserved in all mammalian species (26–28,57,58). A conserved motif comprising 11 nucleotides and located 4–6 nucleotides downstream of the edited base is particularly critical for *in vitro* RNA editing, virtually all mutations in this region eliminating C to U deamination of a synthetic template (26,28). In addition, optimal selection of the editing site, likely depends upon the template adopting an appropriate secondary structure. RNase mapping of short synthetic templates predicted that the apoB template presents the targeted cytidine located within an exposed loop of RNA in the context of a stem-loop structure (57,59). These predictions were independently confirmed using computer algorithms to fold short transcripts in the range of 100–200 nucleotides flanking the edited base (36). More recently, a consensus binding site for apobec-1 was obtained using circular permutation analysis of a 105-nucleotide rat apoB RNA. This consensus binding site [UUUN(A/U)U] is located three nucleotides downstream of the editing site, where it partially overlaps the 5' terminus of the mooring sequence element (36). Folding algorithms, using the 105-nucleotide rat apoB transcript, suggest that the consensus binding site is located at the apex of a stem-loop structure that positions the targeted cytidine in a configuration that allows optimal binding of the *trans*-acting factors to the AU-rich flanking sequences, and will simultaneously accommodate cytidine 6666 within the active site of the deaminase. These predictions are consistent with those obtained by RNase mapping, minor differences likely being accounted for by the use of human versus rat apoB to model the region immediately flanking the edited base (36,59). The consensus binding site for apobec-1 is similar to a functional motif known to be involved in regulating mRNA degradation, referred to as an AU-rich element (ARE) (35,60). AREs often contain multiple iterations of a sequence AUUUA, which is generally presented in the context of an A + U-rich sequence (35,60). Considered together, the findings that apoB RNA contains an A + U-rich sequence in which a canonical apobec-1 consensus site is embedded in proximity to a targeted cytidine, along with requisite flanking efficiency elements, suggests that this region of the transcript is highly adapted for efficient C to U editing. In addition, the functional identification of a consensus site for apobec-1 binding provides a potential screen by which to identify other RNAs that may be substrates for this deaminase. These candidates might reasonably include transcripts containing the apobec-1 consensus binding site in the 3'-untranslated region, where it might participate in the regulation

of mRNA stability in concert with other known destabilization elements (35,60).

The *cis*-acting elements that regulate apoB C to U RNA editing require discussion in regard to the subcellular compartment in which this modification occurs. Studies using nuclei isolated from rat hepatocytes demonstrated that apoB RNA editing was virtually complete in fully processed, i.e., spliced and polyadenylated, transcripts. Nuclear RNA, enriched for pre-mRNA or polyA-RNA, demonstrated much lower levels of C to U editing. These findings suggested that C to U RNA editing does not occur on unspliced mRNA, yet occurs within the nucleus (61). The mechanisms involved in the selection of RNA substrates for C to U editing were recently explored in studies in which an apoB cDNA-editing cassette containing the minimal *cis*-acting element, cloned within an intron, was transfected into hepatoma cells capable of editing endogenous apoB RNA (62). These studies revealed that intronic sequences reduced editing efficiency at the canonical site (62). However, mutations of the splice donor and acceptor sites rescued editing efficiency, despite the apoB RNA-editing cassette being presented within an intron (62). The apoB-editing cassette was then cloned into an intron, in proximity to a Rev response element (RRE) and the resultant plasmid cotransfected into cells along with a Rev expression construct. This experiment produced conditional export of unspliced apoB RNA (using Rev to mediate export of the RRE), yet rescued editing of the chimeric construct. These findings imply that unspliced apoB RNA is intrinsically capable of undergoing editing (62), and suggest that the splicing of apoB RNA per se is not a requisite for efficient editing. The demonstration that apoB RNA editing occurs in the nucleus and is selective for spliced transcripts, rather than pre-mRNA, is an important distinction from A to I RNA editing, which occurs *before* splicing and indeed requires the presence of an intron for efficient utilization of the exonic editing site (1,2). The observations concerning splicing and editing are of biological relevance in the context of mammalian apoB since C to U editing occurs in the middle of a large exon (exon 26, which spans > 7 kb). Accordingly, the splice donor and acceptor sites for exon 26 would presumably be located some distance from the site at which the apoB C to U editing enzyme assembles. However, it remains possible that folding of the apoB RNA encoded by exon 26 could bring components of the splicing complex into proximity with the C to U holoenzyme complex, and thereby interfere with editing efficiency. It bears emphasis that formation of a double-stranded template for A to I RNA editing involves base pairing between the exonic RNA containing the targeted base and an adjacent intron which may be immediately adjacent, or in some instances up to 2 kb downstream (10). Thus, higher-order folding of the RNA, along with the proteins with which the transcripts are associated, could result in steric hindrance to important steps in

C to U RNA editing. These considerations are relevant to the analysis of NF1 RNA editing that, as alluded to above, requires alternative splicing of a downstream exon. The availability of sophisticated structural models capable of incorporating information from large transcripts should allow experimental testing of these hypotheses.

IV. C to U ApoB RNA Editing: *trans*-Acting Factors

ApoB RNA editing requires the assembly of a sequence-specific higher-order “editosome” complex that is formed on the RNA. The size of the editosome was identified to be 27S in sucrose gradients using complexes assembled *in vitro* (63). This equates to a protein complex of about 450 kDa. The minimal enzyme complex that can mediate C to U editing of apoB RNA *in vitro* comprises two proteins: apobec-1, the catalytic subunit, and ACF, a novel RNA-binding protein that serves as the RNA recognition component of the editing enzyme (Fig. 1) (64–66). Other proteins have been identified using detection assays that reflect their ability to interact with apobec-1, ACF, or apoB RNA. These interactions may play a role in modulating the efficiency and specificity of the C to U RNA-editing reaction. Examples of proteins through these approaches include GRY-RBP, CUGBP2, ABBP-1, ABBP-2, AUX240, and hnRNP-C (67–72). Our current understanding of the role of these factors in apoB RNA editing is discussed in the next section.

A. ApoB RNA-Editing Core Components

1. APOBEC-1: AN RNA-SPECIFIC CYTIDINE DEAMINASE

Apobec-1, the catalytic subunit of the mammalian apoB mRNA-editing holoenzyme, is a developmentally regulated 27-kDa homodimeric protein that is expressed exclusively in the luminal gastrointestinal tract of humans, a distribution that coincides with the expression of its canonical target, apoB mRNA (66,73,74). However, in mice and rats, apobec-1 is expressed ubiquitously, including tissues where no apoB mRNA is expressed. The functional significance of apobec-1 in these other tissues is not completely understood, although one tantalizing speculation concerns its role as an AU-rich RNA-binding protein (see the next paragraph). Apobec-1 is the sole catalytic subunit of the editing enzyme; apobec-1^{-/-} mice established the lack of genetic redundancy as evidenced by the complete elimination of C to U editing of apoB RNA (44,45).

As alluded to above, apobec-1 is a cytidine deaminase with homology to other members of the deaminase superfamily (75). Members of this

gene family contain a signature motif of 32–38 amino acids [(H/C)-(A/V)-E-(X)_{24–30}-P-X-X-X-C] which coordinates a zinc residue (76–78). Interestingly, none of the other members of the cytidine deaminase gene family to which apobec-1 belongs can substitute for apobec-1 in regard to C to U editing of apoB RNA. Apobec-1 functions as a cytidine deaminase for both monomeric cytidines and deoxycytidines in solution, as well as acting upon a single specific cytidine in the context of apoB RNA (76,79). The difference in these functional properties of apobec-1, however, lies in the observation that while apobec-1 can deaminate monomeric substrates without the need for additional cofactors, it can only deaminate the cytidine at position 6666 in apoB RNA when additional factor(s) including ACF are present. In addition to its enzymatic activity, apobec-1 encodes an RNA-binding function (36,80,81). Mutagenesis studies have determined that RNA binding is essential for the RNA-editing activity (79–81). The region in apobec-1 required for RNA binding was mapped to the catalytic domain, and includes the active site (H₆₁, E₆₃, C₉₃, C₉₆) as well as two phenylalanine residues acting as the contact sites with apoB RNA (79,80). Other residues have been demonstrated to be involved in RNA substrate binding based on homology to the composite active site (77). These particular studies established that there is a significant amino acid sequence and structural homology between apobec-1 and *Escherichia coli* cytidine deaminase within the catalytic domain (77). The predicted structure of apobec-1, based on the known structure of the *E. coli* CDD, suggests that a cleft is formed using the active sites from two homodimeric molecules of apobec-1, which is predicted to span ~ 21 Å (77). The apoB RNA is postulated to adopt a stem-loop structure with a hairpin structure of ~ 30 nt containing the edited C within the terminal loop (77). This stem loop structure has been proposed to fit into the cleft of apobec-1, positioning the targeted C in close proximity to the glutamate residue, which functions as the proton donor in the RNA-editing reaction (77). Indirect support for this hypothesis is provided by RNase mapping and modeling of RNA substrates, and also with the demonstration that a six-nucleotide sequence UUUGAU, located immediately downstream of the edited C, functions as a consensus binding site for apobec-1 (36,57,59). Furthermore, immunoprecipitation and yeast two-hybrid assays demonstrated that apobec-1 forms homodimers in solution. These observations notwithstanding, it remains to be determined whether apobec-1 binds to the target apoB RNA as a monomer or as a dimer, and whether one or two apoB RNA molecules interact with the same enzyme complex simultaneously. In addition, the molecular form of apobec-1 in the higher-order enzyme complex is currently unknown. These issues will be resolved, once the crystal structure of the apoB RNA-editing holoenzyme is determined.

In regard to the RNA-binding function of apobec-1, recent studies have demonstrated that apobec-1 binds with high affinity (~ 50 nM) to an AU-rich element (ARE) with a consensus sequence UUUN(A/U)U (Fig. 2) (36). This sequence is part of a consensus ARE, present in the (3'UTR) region of many rapidly degraded mRNAs that encode cytokines and protooncogenes. AREs have been demonstrated to be essential for rapid degradation of certain mRNAs as evidenced by targeting studies in mice, where the AREs were deleted from both alleles of tumor necrosis factor (TNF)- α gene resulting in augmented expression of TNF- α , which in turn led to a clinical phenotype resembling inflammatory bowel disease (82). Recombinant apobec-1 was demonstrated to bind to AREs in c-myc, interleukin-2, TNF- α , and granulocyte-macrophage colony-stimulating factor with high affinity (50–100 nM) (36). Furthermore, transient apobec-1 overexpression led to increased stability of endogenous c-myc mRNA in F442A cells, a function accounted for by its RNA-binding activity. Apobec-1 mutants that retain cytidine deaminase activity, but which have abrogated RNA-binding properties, failed to modulate the half-life of c-myc mRNA (36). These data suggest a potential role for apobec-1 in the regulation of mRNA stability, an area likely to yield interesting insights in the future.

ApoB mRNA editing is a nuclear event, implying that apobec-1 should either be a resident nuclear protein or a cytoplasmic protein that shuttles between the nucleus and cytoplasm in order to mediate its RNA-editing function. Apobec-1 is a low-abundance protein and detection of its endogenous expression in native tissues or cells has proven challenging. Immunofluorescence microscopy has revealed that ectopically expressed apobec-1 is present in both the nucleus and cytoplasm (83–85). The nuclear localization signal resides within the N-terminal 172 amino acids (85). However, no canonical nuclear import signal has been identified for the protein. In contrast, residues 173–187 encode a sequence with significant homology to the known leucine-rich nuclear export signal (NES) sequence (L-X₃-L-X₂-L-X-L), which is recognized by the NES receptor CRM1. When this sequence was cloned in frame with a heterologous nuclear protein, it was sufficient to drive the protein to the cytoplasm, suggesting that the region encoded an authentic NES (85). However, unlike other proteins encoding the leucine-rich NES, apobec-1 export was not mediated by CRM1 as evidenced by the lack of inhibition of apobec-1 export by a CRM1 specific inhibitor, leptomycin B (85). These data suggest the possibility of additional CRM1-independent, leucine-rich export signal-dependent mechanism of apobec-1 export from the nucleus. Further studies on the mechanism of import and export of apobec-1 from the nucleus will be required to understand the transport kinetics and functional implications of its directed subcellular targeting.

The human *APOBEC1* gene spans ~18 kb and contains five coding exons. Two different groups have mapped transcription start sites to three different locations. Transcription from a site close to the initiator ATG results in a very small (34 nt) 5'UTR (86). A second transcription start site is located within an *Alu* sequence located further upstream of the translation start codon (87). This results in a 5'UTR of ~0.5 kb encoding ten additional AUG codons. The third transcription start site is located ~1 kb upstream of the AUG codon and is also located within an *Alu* sequence (87). The transcript generated from the third start site undergoes alternative splicing within the 5'UTR region, resulting in the inclusion of an additional ~0.2 kb in the 5'UTR and which contains two upstream AUG codons. This alternatively spliced mRNA demonstrates reduced translational activity because of the inhibitory effects of these upstream AUGs. One testable hypothesis emerging from these findings is that differential splicing of apobec-1 mRNA serves as a mechanism to restrict expression of the protein product. This possibility is discussed, in detail, in relation to rat *apobec-1* gene expression (see below). In addition to alternative promoter usage, human *APOBEC1* undergoes alternative splicing within the coding region (86–88). In normal human intestine, ~55% of the apobec-1 transcript undergoes alternative splicing with skipping of the coding exon 2 and a frame-shift insertion (88). Translation of the alternatively spliced apobec-1 mRNA generates a novel 36-amino acid polypeptide, apobec-T that retains none of the functional motifs required for cytidine deaminase, RNA binding, or RNA editing (88). The protein product from the alternatively spliced transcript (apobec-T) mRNA is detectable in primary human colon cancers and in hepatic metastasis (88). Of greater surprise was the observation that there is a significant shift in the pattern of alternative splicing of apobec-1 mRNA in colon cancer that generates increased quantities of the alternative spliced apobec-T form, coupled with a 3.5-fold increase in apobec-1 transcription (88). In view of the association of transgenic overexpression of apobec-1 with hepatic dysplasia and hepatocellular (38), these observations suggest that elaborate control mechanisms have been put in place to restrict apobec-1 overexpression as a way of controlling unregulated cellular proliferation.

As alluded to above, differential promoter usage may represent a model for regulating expression of the *APOBEC1* gene in the setting of colorectal cancer. Recent studies have demonstrated decreased translation of rat apobec-1 mRNA as a result of a regulated shift in promoter usage in the setting of experimental colon carcinogenesis (89). The rat *apobec-1* gene is transcribed from three promoters resulting in three distinct 5'UTRs accompanied by variable number of upstream AUGs, which in turn are predicted to generate upstream open-reading frames of different length (89,90). In the setting of azoxymethane-induced experimental colon cancer,

the promoter usage shifts to generate apobec-1 mRNA that is burdened with 5'AUGs (89). Western-blot analysis demonstrated a 90% decrease in apobec-1 protein levels compared to normal tissue, suggesting an inhibitory role of the uAUG-loaded 5'UTRs in the translation of apobec-1 (89). Confirmation of this suggestion emerged from mutagenesis studies, wherein point mutations were systematically introduced into all the upstream AUGs. These mutations preserved the overall length and flanking sequence of the reporter construct. The results of these experiments revealed that introducing AUG → UUG mutations throughout the 5'UTR, restored apobec-1 translation, both *in vitro* as well as in transfected cells (89). A final series of studies directly established that the AUG-burdened transcripts manifested defective polysomal loading, which is most plausibly the mechanism of restricting apobec-1 translation in the alternatively spliced transcripts (89). These studies point to the existence of elaborate mechanisms for restricting mammalian apobec-1 expression.

Homologs of apobec-1 have been recently identified through interrogation of both human and murine databases. These homologs include AID, apobec-2/ARCD-1, and the apobec-3 cluster in chromosome 22 (91–94). AID was identified in a differential RNA-display analysis of lymphocytes undergoing immunoglobulin (Ig) class switching. AID is a cytidine deaminase with activity on monomeric substrates, yet demonstrates no C to U editing activity directed toward any known RNA (94). The structural gene for AID is located on human chromosome 12p13.2, adjacent to the *APOBEC1* gene. This proximity lends indirect support to the hypothesis that AID and *APOBEC1* represent gene duplications. Although the target of AID is unknown, its biological function has been convincingly demonstrated. Targeted deletion of the murine AID gene resulted in failure of class-switch recombination and accumulation of IgM (95,96). In addition, inactivating germline mutations of the human AID gene have been demonstrated in association with the hyper-IgM syndrome (96). These features suggest that AID has a critical function in immunoglobulin gene rearrangement and class-switch recombination, despite the fact that the nucleotide target(s) of AID are yet to be determined. Apobec-2/ARCD-1 was identified in a Blast search of the GenBank database with the catalytic subunit of apobec-1 (92,93). It is also a cytidine deaminase with activity on monomeric substrates. While apobec-1/ARCD-1 has no apoB RNA-editing activity, it does competitively inhibit apobec-1-mediated editing activity (93). ARCD-1 inhibits C to U RNA editing of apoB by interacting with both apobec-1 and ACF (92,93). Apobec-3 gene cluster, also identified from a Blast search of the GenBank database, is a group of seven genes, which is thought to be present only in humans (91). The genomic organization of the apobec-3 genes is similar to that of apobec-1, suggesting that the apobec-1 genes may be functional homologs of apobec-1 (91). However, despite the

demonstration that several of these genes exhibit RNA-binding and dimerization activity, none encode functional apoB RNA-editing activity, suggesting that these may be orphan cytidine deaminases (91). Three of the apobec-1-related genes, apobec-3B, -3C, and -3G, were observed to be overexpressed in a variety of cancers, suggesting a potential role for members of this gene cluster in the pathogenesis of certain cancer (91). This possibility will require formal evaluation.

2. APOBEC-1 COMPLEMENTATION FACTOR ACF, THE SPECIFICITY DETERMINANT

Apobec-1 alone is necessary, but not sufficient, for catalyzing C to U editing of apoB RNA. Since the identification that other factors are necessary for RNA editing, many laboratories sought to identify protein(s) that may function as a complementation factor. An important advance in the field recently occurred with the identification of ACF (also known as ASP) by two independent groups using complementary methods (64,65). Driscoll and coworkers initially described the purification of a 65-kDa complementation activity from baboon kidney extracts, using affinity columns coupled with either apobec-1 or the minimal apoB RNA flanking the edited base (97,98). Their strategy, ultimately successful, was based on the likelihood that factors which bound either to the RNA or to apobec-1 would likely complement enzyme activity (65). In a simultaneous report of the cloning of ACF, Greeve and coworkers used a biochemical purification technique to enrich for editing complementation activity (64). They subjected rat nuclear extracts to sucrose gradient centrifugation and, subsequently, a single-stranded DNA cellulose column chromatography for purifying apobec-1 complementation activity. The enriched fractions contained three major protein bands, which upon further size fractionation by fast-phase liquid chromatography resolved the activity to a single ~65-kDa protein (64). Both groups performed microsequencing analysis and isolated peptides with no sequence homology to any known protein (64,65). However, they identified several matches to the mouse and human EST databases. Independently of these two groups, other laboratories, including our own, undertook ACF identification using a genetic two-hybrid screening with a chicken intestinal library and apobec-1 as bait (69). Subsequent cloning of the human ACF was performed by PCR amplification of human intestinal RNA, which demonstrated a high level of conservation in the amino acid sequence between human, rat, mouse, chicken, and baboon ACFs (65,69).

Recombinant ACF and apobec-1 together are sufficient to edit apoB RNA *in vitro*, suggesting that these two proteins make up the minimal functional core of the apoB RNA-editing holoenzyme. However, it is currently not clear whether other proteins exist that can substitute for ACF. Furthermore, the

demonstration of a large apoB RNA-editing enzyme complex in extracts prepared from cells that edit endogenous apoB RNA (63) suggests that additional proteins may play a modulatory role in C to U RNA editing. One piece of data that may suggest the lack of redundancy in ACF comes from the antisense-mediated knockdown studies in McArdle7777 cells, a rat hepatoma cell line. Antisense-mediated knockdown of ACF demonstrated almost complete elimination of apoB RNA editing, similar to that observed when apobec-1 expression was inhibited by antisense oligonucleotides (69). Further proof of this concept will likely require the development and characterization of mice with a targeted deletion of the *acf* gene.

Hepatic apoB synthesis in rats has been previously shown to undergo a thyroid hormone-mediated switch, resulting in the virtual elimination of apoB100 and preferential production of apoB48 (99). This was subsequently determined to be due to the induction of apoB mRNA editing in the livers of thyroid hormone-treated animals, although no effect on apobec-1 mRNA expression was observed (53). This suggested that the thyroid hormone effect might come from modulation of the complementary factor(s). We have recently determined that expression of the murine ACF gene is modulated by thyroid hormone using a strain of congenitally hypothyroid mice *Pax8*^{-/-} that lack thyroid follicles (100). Hepatic apoB RNA editing was significantly decreased in neonatal *Pax8*^{-/-} mice resulting in increased accumulation of triglycerides in the liver. ACF mRNA and protein abundance were significantly lower in the liver of *Pax8*^{-/-} mice (100). Administration of thyroid hormone to *Pax8*^{-/-} mice at levels sufficient to induce hyperthyroidism restored hepatic RNA editing to wild-type levels, and significantly increased ACF mRNA and protein abundance (100). Furthermore, thyroid hormone administration to *Pax8*^{-/-} mice dramatically increased ACF protein abundance in the nucleus (100). These data clearly demonstrate the role of thyroid hormone in the modulation of ACF and, thereby, apoB mRNA editing. Further studies will focus on the proximate-signaling pathways involved in this hormonal regulation of ACF gene expression.

Similar findings concerning the regulation of nuclear localization of ACF in hepatocytes were observed by Smith and coworkers, who employed either insulin or ethanol treatment (49). Previous studies have demonstrated that both insulin and ethanol stimulated hepatic apoB RNA editing (47,48,50,101–103). Furthermore, while apobec-1 mRNA abundance was induced by insulin, no effect on apobec-1 gene expression was observed with ethanol treatment (51,103). More recently, treatment of rat primary hepatocytes in culture, with either insulin or ethanol, demonstrated that the induction of apoB mRNA editing is the result of preferential localization of ACF to the nucleus (49). This is a significant finding since this demonstrates that enhanced C to U RNA editing can occur not only by transcriptional induction of expression of the

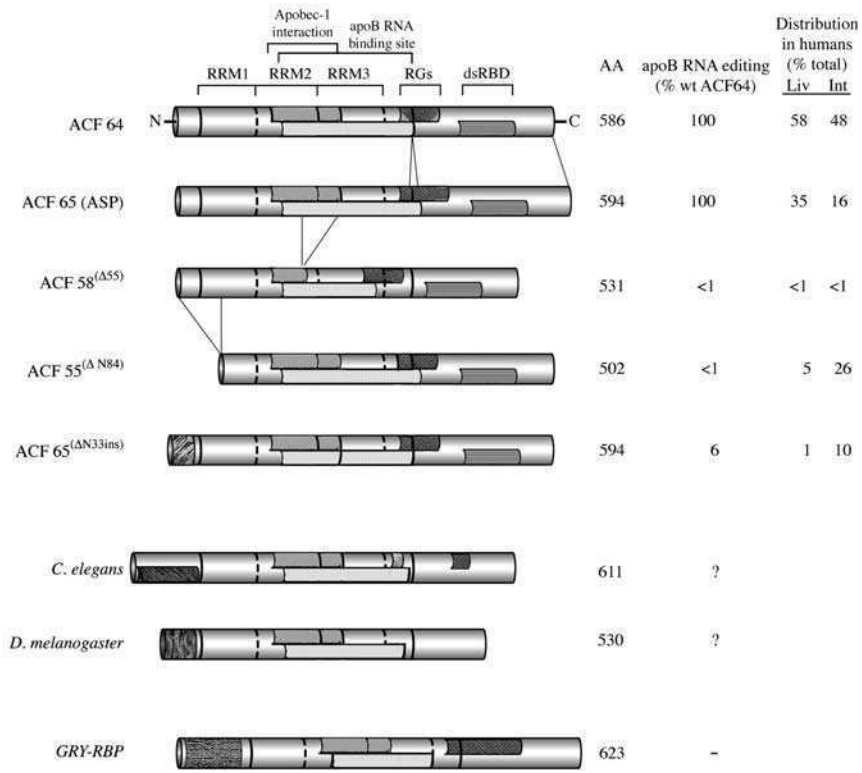


FIG. 3. Splice variants and homologs of human ACF. Five major splice variants of human ACF are illustrated along with their respective amino acids (AA), relative apoB RNA-editing activity, and distribution in liver (Liv) and small intestine (Int). The alignments have been extended to include the predicted conserved domains in *Caenorabditis elegans* ACF (GenBank accession # CAB 70238.1), *Drosophila melanogaster* (GenBank accession # AAL13706) and the related human homolog, GRY-RBP. These predicted domains include conserved RNA recognition motifs (RRMs), an apobec-1 interacting domain (light-shaded bar), an apoB RNA-binding domain (dark-shaded bar), an RG-rich domain (hashed bar), and a double-stranded RNA-binding domain. Unique domains at the amino terminus of ACF homologs from *C. elegans*, *D. melanogaster*, and in GRY-RBP are also indicated.

editing proteins apobec-1 and ACF, but also through their regulated distribution in the nucleus where they would be positioned for optimal activity.

Mutagenesis studies have delineated critical regions of ACF that mediate RNA binding and editing functions. ACF is a highly conserved 65-kDa protein with three nonidentical RNA-binding motifs of the RRM superfamily, located in tandem near the N-terminus of the protein (Fig. 3) (64,65). This is followed by an RG motif and a putative double-stranded RNA-binding motif. Based on

deletion analyses, the minimal functional ACF protein was encoded by 378 residues, located between amino acids 14 and 391 (83,104). However, residues 14–520 were needed for robust complementation activity (83,104). While RRM2 was sufficient for apobec-1 interaction, RRM2 and RRM3 as well as a leucine-rich region immediately downstream of RRM3 were essential for apoB RNA binding (83,104). Point mutations have demonstrated that L359 and L368 may be crucial for binding to apoB RNA (83,104).

With the identification of ACF, it has also been possible to determine the subcellular localization of apobec-1 in the context of the minimal enzyme complex. As alluded to above, apobec-1 is distributed in both the nucleus and cytoplasm of transfected cells (69). However, following cotransfection with ACF, the distribution of apobec-1 becomes predominantly nuclear, suggesting that ACF expression may be a primary determinant in localizing the minimal-editing enzyme complex to the nuclear compartment. With respect to a domain required for nuclear and cytoplasmic transport of the ACF, the primary sequence predicts the presence of a putative SV-40-type NLS between residues 144 and 150, located within RRM2 (83). However, deletion of the entire RRM2 domain or point mutations within this putative NLS did not affect the nuclear localization of ACF (83). Surprisingly, when apobec-1 was cotransfected with these ACF mutants, apobec-1 continued to localize to the cytoplasm (83). This suggests that the RRM2 domain in ACF may play an essential role in mediating protein–protein interactions with apobec-1. These findings imply that ACF may be the determining factor in the localization of the minimal-editing complex to the nucleus. Further studies on the role of chaperones involved in localization of apobec-1 and ACF may shed further light into the mechanisms that regulate this activity.

The human *ACF* gene was determined to span ~80 kb on chromosome 10q and to encode 15 exons (105). There are three noncoding exons and 12 coding exons. Primer extension mapped the site of transcription of the human *ACF* gene to one dominant start site (105). A minimal promoter showing activity in HepG2 cells, a human hepatoma cell line, was mapped to a 330-bp fragment (105). However, sequences further upstream are required for ACF promoter activity in CaCo-2 cells, a human intestinal cell line. Multiple alternative splice acceptor sites have been observed, generating at least nine different transcripts (105). Of these alternative spliced variants, two occur within the coding region, resulting in the presence of either an 8-amino acid insertion or a 55-amino acid deletion (105,106). The 8-amino acid insertion, resulting from the use of an alternative splice acceptor in exon 12, accounts for 14% of intestinal and 29% of hepatic ACF mRNA (105,106). The protein product of the 8-amino acid insertion demonstrated wild-type editing activity (105,106). The other splice variant, resulting from exclusion of exon 9, results in a 55-amino acid deletion in ACF (105,106). This transcript accounts for only

about 2–3% of the intestinal and hepatic transcript, and encodes a protein with no editing activity (105,106). However, the possibility that the protein encodes other functions in the cell cannot be excluded. Similarly, the other splice variants, which include alternative splicing within the coding region, make up a minor fraction of the total ACF mRNA and their cellular function is not known (105,106). In addition, it is not known whether alternative splicing within the 5'UTR adds another level of complexity to regulation of the ACF gene.

Essential splicing factors, such as serine/arginine (SR) proteins, modulate splice-site usage of the cellular splicing machinery in a concentration-dependent manner (107,108). Several SR protein-binding sites were observed in exon 12, in the region flanking the alternative splice acceptor site (106). Cotransfection of a minigene encompassing exons 11 and 12 and the intervening intron, together with selected RS expression plasmids, identified SRp40 as a viable candidate for determining the ratio of the two alternative spliced variants (106). However, the effect of SRp40 on exon-12 splicing of the endogenous ACF transcript, as well as the other alternatively spliced exons, is currently unknown (106). Final proof of the role of SR proteins will likely require gene-knockout studies, where individual SR proteins are targeted.

B. Other Factors—Physiological Modulators of C to U RNA Editing

1. GRY-RBP

GRY-RBP (also known as hnRNP-Q or NSAP1) was identified in a two-hybrid screen of a chicken intestinal RNA library using apobec-1 as bait, as well as by peptide sequencing of a purified fraction of chicken intestinal extracts that were found to demonstrate enrichment for apoB RNA binding (69). GRY-RBP closely resembles ACF in its primary structure (see Fig. 3). Similar to ACF, GRY-RBP contains three tandem, nonidentical RNA-binding motifs analogous to others in the RRM superfamily. However, unlike ACF, GRY-RBP contains an extended N-terminus. In addition, GRY-RBP contains a C-terminal putative bipartite nuclear localization signal sequence, 20 RG motifs and 8 RGG motifs, resulting in an overall 36% of arginine or glycine residues in the region (69). Transient transfection experiments demonstrated that GRY-RBP is localized to the nucleus (69). In addition, cotransfection of apobec-1 with GRY-RBP demonstrated nuclear localization of both proteins (69). Since apobec-1 is generally resident in the cytoplasm when transfected alone, these studies suggest that GRY-RBP may also act as a docking protein for transporting apobec-1 to the nucleus (69). GRY-RBP is an RNA-binding protein that interacts with U- and AU-rich targets based on its demonstrated

binding to homopolymeric sequences. This would be of significant relevance to apoB RNA since the region flanking the edited base is 70% AU-rich (81). Indeed recombinant GRY-RBP bound to apoB RNA, but mutations within the 30 nucleotides flanking the edited base did not identify any specific sequences for interaction with GRY-RBP (69). Thus, further studies will be needed to determine the binding affinity of GRY-RBP to U- and AU-rich sequences as a first step in identifying binding targets for GRY-RBP. GRY-RBP competitively inhibits C to U editing of apoB mRNA by interacting with apobec-1 and ACF (69). Addition of even small quantities of GRY-RBP to the *in vitro* editing system resulted in complete loss of RNA editing. By contrast, antisense oligonucleotide-mediated inhibition of GRY-RBP expression in cells revealed a two-fold increase in C to U editing of endogenous apoB RNA (69). Taken together, these data suggest that GRY-RBP is a potent inhibitor of the editing enzyme.

a. GRY-RBP: mRNA Splicing and Stability Functions Other studies have suggested a wider role for GRY-RBP in cellular metabolism. GRY-RBP was initially identified as an NS1-binding protein in a yeast two-hybrid screen of HeLa-cell mRNA library (109). The domain in GRY-RBP that was identified in this screen included the RGG motifs found close to the C-terminus of the protein (109). NS1 is a major nonstructural protein of the minute virus of mice, and encodes ATPase, DNA binding, helicase, site-specific endonuclease, and transcriptional activation domains (110). NS1 is a nuclear protein and can support the replication of a minigene containing the *cis*-acting internal replication sequences, but lacking the entire coding region (110). One prediction for these interactions is that GRY-RBP may act as a cofactor of NS1 protein in double-stranded RNA binding (109). Further studies are required to determine the precise role of the NS1–GRY-RBP interaction. A second function for GRY-RBP is related to the major protein-coding region determinant of instability (mCRD) in *c-fos* mRNA (111). The mCRD is encoded within the *c-fos* protein-coding region, and facilitates mRNA deadenylation and decay by a mechanism coupled to translation (112). The current hypothesis is that a protein complex interacts with mCRD and the polyA tail to stabilize the mRNA (113,114). However, when the transcript undergoes translation, binding of the complex to mCRD is disrupted, resulting in rapid deadenylation and hence transcript degradation (115,116). Indeed, a complex containing GRY-RBP was identified to bind to mCRD along with four other proteins: Unr, a purine-rich RNA-binding protein; PABP, a poly A-binding protein; PAIB, a PABP-interacting protein; and hnRNP-D, an ARE-binding protein. These observations lend support to the hypothesis that a physical bridge forms between mCRD and the poly(A) sequence (116).

GRY-RBP has also been recently identified as one of three alternatively spliced variants of hnRNP-Q (117). The function of hnRNP-Q was inferred through studies in which the protein was observed to bind to the survival of motor neurons (SMN) protein, mutations of which cause spinal muscular atrophy, a common neurodegenerative disease (117). SMN proteins are present in a complex that includes core spliceosomal proteins that bind to U-rich small nuclear RNAs, which in turn function in pre-mRNA splicing (118). Further *in vitro* studies demonstrated GRY-RBP binding to SMN, a function determined to be essential for efficient pre-mRNA splicing of adenovirus minigenes, suggesting that GRY-RBP plays an important function in pre-mRNA splicing (117). GRY-RBP was also identified to bind to the C-terminal repeat domain (CTD) of RNA polymerase II (119). CTD contains 52 repeats of the consensus heptamer sequence YSPTSPS, a region that becomes hyperphosphorylated during transcriptional elongation (119). Deletion of this consensus region results in lack of elongation of the transcript suggesting its importance in gene transcription. The interaction of GRY-RBP with the spliceosomal complex containing SMN, as well as with CTD, suggests that transcription-coupled splicing may involve a physical interaction between the two complexes perhaps involving a common protein, GRY-RBP. In this regard, it should be noted that the SMN complex was recently demonstrated to interact with RNA polymerase II (120).

2. CUGBP2

CUGBP2 (also known as ETR-3, Napor-2, or Brunol-3) was identified in a yeast two-hybrid screening of chicken intestinal RNA using apobec-1 as bait (68). Chicken intestinal RNA was used as a source of interacting proteins since previous studies revealed that intestinal extracts are an abundant source of complementation activity, yet lack apobec-1 (23). CUGBP2 is a 54-kDa protein containing three RRM, two of which are located in the N-terminus, while the third RRM is located in the C-terminus separated from the first two RRM by an alanine- and glutamine-rich bridge sequence (68). Size fractionation of bovine liver extracts revealed several fractions that were enriched in apoB RNA-editing complementation activity, in which CUGBP2 was initially observed to be present along with ACF and a number of other proteins that eluted in a size range of ~40–120 kDa (68). Addition of recombinant apobec-1 to these fractions shifted the size of the active complex to a size range of ~250–669 kDa, presumably more representative of the native-editing holoenzyme (68). This is a size range close to that reported by other investigators using glycerol gradient centrifugation of native-editing enzyme complexes isolated from rat liver (11S–27S) (63). This larger complex contained CUGBP2, in addition to apobec-1 and ACF, suggesting that CUGBP2 may be an integral member of the C to U RNA-editing holoenzyme.

CUGBP2 interacts with both apobec-1 and ACF as evidenced by confocal immunocytochemistry, far-western analyses, as well as immunoprecipitation assays (68). When cotransfected with ACF, both CUGBP2 and ACF localized in the nucleus. On the other hand, when cotransfected with apobec-1, CUGBP2 was observed to colocalize with apobec-1 in the cytoplasm (68). This is an apparent paradox since other studies demonstrate that transfection of CUGBP2 alone produces a predominantly nuclear distribution. Cytoplasmic localization of CUGBP2, when expressed in the presence of apobec-1, suggests several possible interpretations. Among these, it is possible that CUGBP2 does not contain a strong nuclear signal; alternatively, it is possible that apobec-1 is the determinant of localization of the apobec-1-CUGBP2 complex, or conceivably that the interaction between apobec-1 and CUGBP2 masks the nuclear location signal in CUGBP2. Further, mutational analysis of CUGBP2 will be required to better understand this process.

CUGBP2 is an RNA-binding protein that was originally identified to bind CUG triplet-expansion repeats (121). This was the second such protein that has been identified, hence its designation *CUG-binding protein 2* (CUGBP2). Although apoB RNA does not contain CUG triplet repeats, CUGBP2 was observed to bind to the RNA, with the major binding site mapped to a six-nucleotide sequence AUGAUA located immediately upstream of the edited cytidine (68). In addition, immunoprecipitation of CUGBP2 from McArdle cell extracts, coprecipitated apoB mRNA, implying that CUGBP2 is bound to apoB RNA in these cells. CUGBP2 also demonstrated binding to other targets of apobec-1-mediated mRNA editing, including NAT1 and NF1 RNAs (68).

The physical interactions of CUGBP2 with apobec-1, ACF, and apoB RNA suggest that CUGBP2 might play a role in regulating apoB RNA editing. Indeed, addition of the superdex 200 fractions containing apobec-1, ACF, and CUGBP2 to a synthetic apoB RNA produced efficient RNA editing, suggesting that these fractions contained a functional holoenzyme (68). Furthermore, immunodepletion of CUGBP2 from the extracts demonstrated a complete loss of editing activity (68). Disappointingly, addition of either the bead eluant or recombinant CUGBP2 failed to reconstitute editing activity. However, addition of recombinant ACF was sufficient to restore complementation activity, suggesting that ACF was also removed in the immunodepletion (68). This was confirmed by western-blot analysis, which demonstrated that ACF was present in the bead eluant and not in the supernatant of extracts immunodepleted with anti-CUGBP2 IgG. Failure of the bead eluant to restore complementation activity suggests among other possibilities that the ACF-CUGBP2 complex within the immunoprecipitate is present in a confirmation that does not allow complementation activity.

Further studies demonstrated that CUGBP2 is a competitive inhibitor of the RNA-editing reaction, which can be rescued by addition of either increasing amounts of apobec-1 or ACF, but not apoB RNA (68). This suggests that CUGBP2 exerts its effects by disrupting the apobec-1-ACF interaction. Further, protein-protein interaction and enzyme kinetic studies are required to fully deduce the mechanism of CUGBP2 action in the RNA-editing process.

a. CUGBP2: mRNA Splicing, Stability, and Translational Control Functions? In addressing its potential biological functions, it is worth emphasizing that CUGBP2 is ubiquitously expressed at low levels, but at higher levels in skeletal and cardiac muscle, sites where neither apobec-1 or apoB RNA is expressed and no C to U RNA editing occurs (68). An important feature in myotonic dystrophy is muscle wasting and weakness, features that emerge in the setting of expanded CUG repeats within the myotonic dystrophy-related protein kinase, DMPK (122). These expanded CUG repeats were initially postulated to bind to proteins and sequester them in the nucleus, resulting in loss of their normal RNA processing function. Indeed, CUGBP2 was initially identified to bind to the expanded CUG triplet repeats present in the 3'UTR of DMPK (121).

What cellular function of CUGBP2 is disrupted in myotonic dystrophy? CUGBP2 was identified to bind and regulate alternative splicing of the human cardiac troponin T pre-mRNA (123–125). The observation that splicing of cardiac troponin T is also disrupted in striated muscle of myotonic dystrophy patients, implies indirectly that CUGBP2 may exert a role in alternative splicing in this setting (126–128). In this regard, it should be noted that a muscle-specific chloride channel, CIC-1, and insulin receptor pre-mRNAs are aberrantly spliced in striated muscle of myotonic dystrophy patients (126,129,130). The mechanistic relationship of these observations to disease pathogenesis will require further study. Adding to the complexity in this disease, there is increased expression of CUGBP1, a homolog of CUGBP2, in the muscle of myotonic dystrophy patients (126). CUGBP1 is also implicated in the aberrant splicing function (126,131), and its role in the setting of myotonic dystrophy will need further direct evaluation. More recently, CUGBP2 has been demonstrated to modulate mRNA stability and translation functions (132). CUGBP2 was observed to bind AU-rich sequences in the 3'UTR of cyclooxygenase-2 mRNA and stabilize the mRNA, yet paradoxically inhibits its translation (132). Furthermore, CUGBP2 overexpression in a colon cancer cell line, HT-29, resulted in apoptosis, suggesting an important role for this protein in cell turnover and proliferation (132). Taken together, these studies suggest that CUGBP2 and its homologs may play a role in various cellular functions, including regulating apoB mRNA editing in epithelial cells

as well as possibly in disorders associated with altered RNA splicing and translation.

C. Additional Factors—Regulators of RNA Editing?

Four cellular proteins have been identified either as a apobec-1 RNA-binding protein (ABBP-1, ABBP-2, hnRNP-C) or a member of the editosome complex (AUX240).

1. ABBP-1/hnRNP-A/B

ABBP-1 is an alternatively spliced variant of hnRNP-A/B, which was originally identified in a two-hybrid screen of a human placenta cDNA library (71). Confirmation of apobec-1 interaction was performed by immunoprecipitation assays using *in vitro* translated proteins (71). In addition, the apobec-1-binding domain in ABBP-1 was mapped to a glycine-rich region, located in the C-terminus of the protein (71). ABBP-1 was also observed to bind apoB RNA *in vitro* by electrophoretic mobility shift assays (71). Immunodepletion of ABBP-1 in extracts from apobec-1-transfected HepG2 cells reduced apoB RNA editing *in vitro* (71). Furthermore, antisense cDNA-mediated knock-down of ABBP-1 in HepG2 cells resulted in loss of endogenous apoB mRNA editing (71). Taken together these data suggest that ABBP-1 or another spliced variant of hnRNP-A/B is required for apoB mRNA editing. These findings are intriguing in light of the information that HepG2 cells express ACF, the required complementation factor for C to U editing. The immediate question is how ABBP1 knockdown abrogates apoB RNA editing since ACF itself should continue to be expressed and has been demonstrated alone to complement apobec-1. Further analysis of these ABBP-1 antisense experiments may reveal whether the inhibition was specific for ABBP-1 downregulation or whether ACF expression was inadvertently altered.

2. ABBP-2

ABBP-2 (also known as HEDJ) also identified in a yeast two-hybrid screen of a human placenta library (70). This protein was originally identified as a member of the hsp40 cochaperone family, which is characterized by the presence of a highly conserved J domain involved in protein synthesis, folding, and secretion (70,133). The J domain is a signature of members of the DNAJ family. These chaperones are also involved in reverse translocation and degradation of misfolded proteins in the lumen of the endoplasmic reticulum (133). The J domain interacts with the endoplasmic reticulum-associated Hsp70 in an ATP-dependent manner and is capable of stimulating its ATPase activity (70,133). It is intriguing that apobec-1 interacts with this protein. Similar to ABBP-1, inhibition of ABBP-2/HEDJ expression by transfection of an antisense ABBP-2 cDNA resulted in downregulation of apoB RNA editing

(70). The mechanism of this response is unclear and additional studies will need to address the effects of ABBP-2 expression on the expression and function of ACF.

3. hnRNP-C1

hnRNP-C1 was identified in a yeast two-hybrid screen of a rat liver cDNA library using apobec-1 as bait (67). hnRNP-C1 is an AU-rich RNA-binding protein that binds to nascent transcripts and is involved in spliceosomal assembly and mRNA splicing (134,135). Recombinant hnRNP-C1 bound to the sense strand of apoB RNA, but not the antisense strand (67). Given that apoB RNA is ~70% AU-rich in content in the region flanking the edited cytidine, this RNA-binding activity profile suggests that hnRNP-C1 interacts with a specific AU-rich sequence in apoB RNA (67). Further studies narrowed the RNA-binding domain to the N-terminal 104 amino acids in hnRNP-C1, which contains a single RRM (67). Recombinant hnRNP-C1 was observed to inhibit C to U RNA editing, catalyzed by a partially purified fraction from rat liver extracts (67). The mechanism of this competitive inhibition was not due to binding of hnRNP-C1 to apoB RNA, since a deletion mutant of hnRNP-C1 that retained RNA-binding activity, but lacked the leucine-rich repeats, failed to inhibit the RNA-editing reaction (67). Rather, the findings suggest that hnRNP-C1 competitively inhibits apoB RNA editing through protein-protein interactions with apobec-1 and/or other factors in the holoenzyme.

4. AUX240

AUX240 was identified in rat liver extracts using monoclonal antibodies generated against the 27S editosome that was assembled *in vitro* on a synthetic apoB RNA (72). Immunodepletion of AUX240 from rat liver extracts resulted in significant inhibition of C to U RNA-editing activity (72). Furthermore, addition of the bead-eluted material to rat liver extracts enhanced RNA-editing efficiency (72). However, addition of the immunoadsorbed material alone to the editing reaction in the absence of the additional proteins did not demonstrate any RNA-editing activity (72). These data suggest that the immunoadsorbed material does not contain all the factors necessary to support C to U RNA editing. Taken together, these data offer no clear proof of the involvement of AUX240 in C to U RNA-editing activity. It should be noted that multiple proteins from rat liver extracts coeluted with AUX240 from rat liver extracts, including a protein of ~66 kDa which may conceivably be ACF (72). The nature of these other coeluting proteins is currently unknown, but further study will likely be revealing.

V. Summary

C to U editing of a single cytidine at nt 6666 in apoB RNA is mediated by a multicomponent-editing complex, and is regulated by developmental, nutritional, and hormonal factors. Apobec-1 and ACF are components of the minimal-editing enzyme, although other proteins, such as GRY-RBP and CUGBP2, represent likely auxiliary factors of the editing holoenzyme that regulate its activity. Much work remains in this area, including identification of all the factors that comprise the holoenzyme and a characterization of their role in C to U RNA editing. In addition, as implied throughout this report, the proteins that have been currently identified may exhibit other functions in RNA metabolism. It would be interesting to determine whether these other functions, such as pre-mRNA splicing, mRNA stability, and mRNA translation, are related to their intrinsic RNA-editing function. Future studies will likely provide an opportunity to consider these issues in the context of the evolutionary significance of RNA editing and the RNA world.

ACKNOWLEDGMENTS

Work cited from the authors' laboratories is supported by the National Institutes of Health grants DK-56260 and HL-38180 to NOD, DK-62265 to S.A., and the Digestive Disease Research Core Center Grant DK-52574. S.A. is a Research Scholar of the American Gastroenterology Association. The authors acknowledge valuable discussions with current and former laboratory members, particularly Jeff Henderson, Susan Kennedy, Debnath Mukhopadhyay, and Libby Newberry.

REFERENCES

1. J. M. Gott and R. B. Emeson, Functions and mechanisms of RNA editing, *Annu. Rev. Genet.* **34**, 499–531 (2000).
2. A. P. Gerber and W. Keller, RNA editing by base deamination: more enzymes, more targets, new mysteries, *Trends Biochem. Sci.* **26**, 376–384 (2001).
3. L. Simpson, O. H. Thiemann, N. J. Savill, J. D. Alfonzo and D. A. Maslov, Evolution of RNA editing in trypanosome mitochondria, *Proc. Natl. Acad. Sci. USA* **97**, 6986–6993 (2000).
4. Y. W. Cheng and J. M. Gott, Transcription and RNA editing in a soluble in vitro system from Physarum mitochondria, *Nucleic Acids Res.* **28**, 3695–3701 (2000).
5. Y. W. Cheng, L. M. Visomirski-Robic and J. M. Gott, Non-templated addition of nucleotides to the 3' end of nascent RNA during RNA editing in Physarum, *EMBO J.* **20**, 1405–1414 (2001).
6. M. L. Reed, N. M. Peeters and M. R. Hanson, A single alteration 20 nt 5' to an editing target inhibits chloroplast RNA editing in vivo, *Nucleic Acids Res.* **29**, 1507–1513 (2001).
7. R. Bock, Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing, *Biochimie* **82**, 549–557 (2000).

8. T. Hirose and M. Sugiura, Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system, *EMBO J.* **20**, 1144–1152 (2001).
9. P. Giege and A. Brennicke, RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs, *Proc. Natl. Acad. Sci. USA* **96**, 15324–15329 (1999).
10. P. H. Seeburg, A-to-I editing: new and old sites, functions and speculations, *Neuron* **35**, 17–20 (2002).
11. J. Villegas, I. Muller, J. Arredondo, R. Pinto and L. O. Burzio, A putative RNA editing from U to C in a mouse mitochondrial transcript, *Nucleic Acids Res.* **30**, 1895–1901 (2002).
12. P. M. Sharma, M. Bowman, S. L. Madden, F. J. Rauscher, 3rd and S. Sukumar, RNA editing in the Wilms' tumor susceptibility gene, WT1, *Genes Dev.* **8**, 720–731 (1994).
13. K. B. Gunning, S. L. Cohn, G. E. Tomlinson, L. C. Strong and V. Huff, Analysis of possible WT1 RNA processing in primary Wilms tumors, *Oncogene* **13**, 1179–1185 (1996).
14. C. Mrowka and A. Schedl, Wilms' tumor suppressor gene WT1: from structure to renal pathophysiologic features, *J. Am. Soc. Nephrol.* **11 Suppl. 16**, S106–S115 (2000).
15. J. Villegas, A. M. Zarraga, I. Muller, L. Montecinos, E. Werner, M. Brito, A. M. Meneses and L. O. Burzio, A novel chimeric mitochondrial RNA localized in the nucleus of mouse sperm, *DNA Cell Biol.* **19**, 579–588 (2000).
16. S. G. Young, Recent progress in understanding apolipoprotein B, *Circulation* **82**, 1574–1594 (1990).
17. N. O. Davidson and G. S. Shelness, Apolipoprotein B: mRNA editing, lipoprotein assembly, and presecretory degradation, *Annu. Rev. Nutr.* **20**, 169–193 (2000).
18. C. Hadjiagapiou, F. Giannoni, T. Funahashi, S. F. Skarosi and N. O. Davidson, Molecular cloning of a human small intestinal apolipoprotein B mRNA editing protein, *Nucleic Acids Res.* **22**, 1874–1879 (1994).
19. P. P. Lau, H. J. Zhu, A. Baldini, C. Charnsangavej and L. Chan, Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene, *Proc. Natl. Acad. Sci. USA* **91**, 8522–8526 (1994).
20. J. Greeve, I. Altkemper, J. H. Dieterich, H. Greten and E. Windler, Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins, *J. Lipid Res.* **34**, 1367–1383 (1993).
21. S. Anant and N. O. Davidson, Molecular mechanisms of apolipoprotein B mRNA editing, *Curr. Opin. Lipidol.* **12**, 159–165 (2001).
22. S. Anant, H. Yu and N. O. Davidson, Evolutionary origins of the mammalian apolipoprotein B RNA editing enzyme, apobec-1: structural homology inferred from analysis of a cloned chicken small intestinal cytidine deaminase, *Biol. Chem.* **379**, 1075–1081 (1998).
23. B. Teng and N. O. Davidson, Evolution of intestinal apolipoprotein B mRNA editing. Chicken apolipoprotein B mRNA is not edited, but chicken enterocytes contain in vitro editing enhancement factor(s), *J. Biol. Chem.* **267**, 21265–21272 (1992).
24. L. Powell-Braxton, M. Veniant, R. D. Latvala, K. I. Hirano, W. B. Won, J. Ross, N. Dybdal, C. H. Zlot, S. G. Young, N. O. Davidson, A mouse model of human familial hypercholesterolemia: markedly elevated low density lipoprotein cholesterol levels and severe atherosclerosis on a low-fat chow diet, *Nat. Med.* **4**, 934–938 (1998).
25. M. S. Davies, S. C. Wallis, D. M. Driscoll, J. K. Wynne, G. W. Williams, L. M. Powell and J. Scott, Sequence requirements for apolipoprotein B RNA editing in transfected rat hepatoma cells, *J. Biol. Chem.* **264**, 13395–13398 (1989).
26. R. R. Shah, T. J. Knott, J. E. Legros, N. Navaratnam, J. C. Greeve and J. Scott, Sequence requirements for the editing of apolipoprotein B mRNA, *J. Biol. Chem.* **266**, 16301–16304 (1991).

27. D. M. Driscoll, S. Lakhe-Reddy, L. M. Oleksa and D. Martinez, Induction of RNA editing at heterologous sites by sequences in apolipoprotein B mRNA, *Mol. Cell. Biol.* **13**, 7288–7294 (1993).
28. J. W. Backus and H. C. Smith, Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro, *Nucleic Acids Res.* **20**, 6007–6014 (1992).
29. M. Hersberger and T. L. Innerarity, Two efficiency elements flanking the editing site of cytidine 6666 in the apolipoprotein B mRNA support mooring-dependent editing, *J. Biol. Chem.* **273**, 9435–9442 (1998).
30. G. R. Skuse, A. J. Cappione, M. Sowden, L. J. Metheny and H. C. Smith, The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing, *Nucleic Acids Res.* **24**, 478–485 (1996).
31. D. Mukhopadhyay, S. Anant, R. M. Lee, S. Kennedy, D. Viskochil and N. O. Davidson, C → U editing of neurofibromatosis 1 mRNA occurs in tumors that express both the type II transcript and apobec-1, the catalytic subunit of the apolipoprotein B mRNA-editing enzyme, *Am. J. Hum. Genet.* **70**, 38–50 (2002).
32. A. J. Cappione, B. L. French and G. R. Skuse, A potential role for NF1 mRNA editing in the pathogenesis of NF1 tumors, *Am. J. Hum. Genet.* **60**, 305–312 (1997).
33. K. Cichowski and T. Jacks, NF1 tumor suppressor gene function: narrowing the GAP, *Cell* **104**, 593–604 (2001).
34. G. Bollag, D. W. Clapp, S. Shih, F. Adler, Y. Y. Zhang, P. Thompson, B. J. Lange, M. H. Freedman, F. McCormick, T. Jacks, K. Shannon, Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells, *Nat. Genet.* **12**, 144–148 (1996).
35. J. Ross, mRNA stability in mammalian cells, *Microbiol. Rev.* **59**, 423–450 (1995).
36. S. Anant and N. O. Davidson, An AU-rich sequence element (UUUN[A/U]U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3' untranslated region of c-myc increases mRNA stability, *Mol. Cell. Biol.* **20**, 1982–1992 (2000).
37. F. Giannoni, S. C. Chou, S. F. Skarosi, M. S. Verp, F. J. Field, R. A. Coleman and N. O. Davidson, Developmental regulation of the catalytic subunit of the apolipoprotein B mRNA editing enzyme (APOBEC-1) in human small intestine, *J. Lipid Res.* **36**, 1664–1675 (1995).
38. S. Yamanaka, M. E. Balestra, L. D. Ferrell, J. Fan, K. S. Arnold, S. Taylor, J. M. Taylor and T. L. Innerarity, Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals, *Proc. Natl. Acad. Sci. USA* **92**, 8483–8487 (1995).
39. S. Yamanaka, K. S. Poksay, D. M. Driscoll and T. L. Innerarity, Hyperediting of multiple cytidines of apolipoprotein B mRNA by APOBEC-1 requires auxiliary protein(s) but not a mooring sequence motif, *J. Biol. Chem.* **271**, 11506–11510 (1996).
40. N. Navaratnam, D. Patel, R. R. Shah, J. C. Greeve, L. M. Powell, T. J. Knott and J. Scott, An additional editing site is present in apolipoprotein B mRNA, *Nucleic Acids Res.* **19**, 1741–1744 (1991).
41. S. Yamanaka, K. S. Poksay, K. S. Arnold and T. L. Innerarity, A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme, *Genes Dev.* **11**, 321–333 (1997).
42. S. Yamanaka, X. Y. Zhang, M. Maeda, K. Miura, S. Wang, R. V. Farese, Jr., H. Iwao and T. L. Innerarity, Essential role of NAT1/p97/DAP5 in embryonic differentiation and the retinoic acid pathway, *EMBO J.* **19**, 5533–5541 (2000).
43. X. Qian, M. E. Balestra, S. Yamanaka, J. Boren, I. Lee and T. L. Innerarity, Low expression of the apolipoprotein B mRNA-editing transgene in mice reduces LDL levels but does not cause liver dysplasia or tumors, *Arterioscler. Thromb. Vasc. Biol.* **18**, 1013–1020 (1998).

44. K. Hirano, S. G. Young, R. V. Farese, Jr., J. Ng, E. Sande, C. Warburton, L. M. Powell-Braxton and N. O. Davidson, Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48, *J. Biol. Chem.* **271**, 9887–9890 (1996).
45. J. R. Morrison, C. Paszty, M. E. Stevens, S. D. Hughes, T. Forte, J. Scott and E. M. Rubin, Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism, *Proc. Natl. Acad. Sci. USA* **93**, 7154–7159 (1996).
46. M. Hersberger, S. Patarroyo-White, X. Qian, K. S. Arnold, L. Rohrer, M. E. Balestra and T. L. Innerarity, Regulatable liver expression of the rabbit ApoB mRNA-editing enzyme catalytic polypeptide 1 (APOBEC-1) in mice lacking endogenous APOBEC-1 leads to aberrant hyperediting, *Biochem. J.* **369**, 255–262 (2002).
47. P. P. Lau, D. J. Cahill, H. J. Zhu and L. Chan, Ethanol modulates apolipoprotein B mRNA editing in the rat, *J. Lipid Res.* **36**, 2069–2078 (1995).
48. A. Giangreco, M. P. Sowden, I. Mikityansky and H. C. Smith, Ethanol stimulates apolipoprotein B mRNA editing in the absence of de novo RNA or protein synthesis, *Biochem. Biophys. Res. Commun.* **289**, 1162–1167 (2001).
49. M. P. Sowden, N. Ballatori, K. L. Jensen, L. H. Reed and H. C. Smith, The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors, *J. Cell Sci.* **115**, 1027–1039 (2002).
50. D. Van Mater, M. P. Sowden, J. Cianci, J. D. Sparks, C. E. Sparks, N. Ballatori and H. C. Smith, Ethanol increases apolipoprotein B mRNA editing in rat primary hepatocytes and McArdle cells, *Biochem. Biophys. Res. Commun.* **252**, 334–339 (1998).
51. Y. Yang, M. P. Sowden and H. C. Smith, Induction of cytidine to uridine editing on cytoplasmic apolipoprotein B mRNA by overexpressing APOBEC-1, *J. Biol. Chem.* **275**, 22663–22669 (2000).
52. T. Funahashi, F. Giannoni, A. M. DePaoli, S. F. Skarosi and N. O. Davidson, Tissue-specific, developmental and nutritional regulation of the gene encoding the catalytic subunit of the rat apolipoprotein B mRNA editing enzyme: functional role in the modulation of apoB mRNA editing, *J. Lipid Res.* **36**, 414–428 (1995).
53. Y. Inui, F. Giannoni, T. Funahashi and N. O. Davidson, REPR and complementation factor(s) interact to modulate rat apolipoprotein B mRNA editing in response to alterations in cellular cholesterol flux, *J. Lipid Res.* **35**, 1477–1489 (1994).
54. B. Teng, M. Verp, J. Salomon and N. O. Davidson, Apolipoprotein B messenger RNA editing is developmentally regulated and widely expressed in human tissues, *J. Biol. Chem.* **265**, 20616–20620 (1990).
55. C. L. Baum, B. B. Teng and N. O. Davidson, Apolipoprotein B messenger RNA editing in the rat liver. Modulation by fasting and refeeding a high carbohydrate diet, *J. Biol. Chem.* **265**, 19263–19270 (1990).
56. K. Higuchi, K. Kitagawa, K. Kogishi and T. Takeda, Developmental and age-related changes in apolipoprotein B mRNA editing in mice, *J. Lipid Res.* **33**, 1753–1764 (1992).
57. M. Hersberger, S. Patarroyo-White, K. S. Arnold and T. L. Innerarity, Phylogenetic analysis of the apolipoprotein B mRNA-editing region. Evidence for a secondary structure between the mooring sequence and the 3' efficiency element, *J. Biol. Chem.* **274**, 34590–34597 (1999).
58. K. Bostrom, S. J. Lauer, K. S. Poksay, Z. Garcia, J. M. Taylor and T. L. Innerarity, Apolipoprotein B48 RNA editing in chimeric apolipoprotein EB mRNA, *J. Biol. Chem.* **264**, 15701–15708 (1989).
59. N. Richardson, N. Navaratnam and J. Scott, Secondary structure for the apolipoprotein B mRNA editing site. Au-binding proteins interact with a stem loop, *J. Biol. Chem.* **273**, 31707–31717 (1998).

60. C. Y. Chen and A. B. Shyu, AU-rich elements: characterization and importance in mRNA degradation, *Trends Biochem. Sci.* **20**, 465–470 (1995).
61. P. P. Lau, W. J. Xiong, H. J. Zhu, S. H. Chen and L. Chan, Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation, *J. Biol. Chem.* **266**, 20550–20554 (1991).
62. M. P. Sowden and H. C. Smith, Commitment of apolipoprotein B RNA to the splicing pathway regulates cytidine-to-uridine editing-site utilization, *Biochem. J.* **359**, 697–705 (2001).
63. H. C. Smith, S. R. Kuo, J. W. Backus, S. G. Harris, C. E. Sparks and J. D. Sparks, In vitro apolipoprotein B mRNA editing: identification of a 27S editing complex, *Proc. Natl. Acad. Sci. USA* **88**, 1489–1493 (1991).
64. H. Lellek, R. Kirsten, I. Diehl, F. Apostel, F. Buck and J. Greeve, Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex, *J. Biol. Chem.* **275**, 19848–19856 (2000).
65. A. Mehta, M. T. Kinter, N. E. Sherman and D. M. Driscoll, Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA, *Mol. Cell. Biol.* **20**, 1846–1854 (2000).
66. B. Teng, C. F. Burant and N. O. Davidson, Molecular cloning of an apolipoprotein B messenger RNA editing protein, *Science* **260**, 1816–1819 (1993).
67. J. Greeve, H. Lellek, P. Rautenberg and H. Greten, Inhibition of the apolipoprotein B mRNA editing enzyme-complex by hnRNP C1 protein and 40S hnRNP complexes, *Biol. Chem.* **379**, 1063–1073 (1998).
68. S. Anant, J. O. Henderson, D. Mukhopadhyay, N. Navaratnam, S. Kennedy, J. Min and N. O. Davidson, Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor, *J. Biol. Chem.* **276**, 47338–47351 (2001).
69. V. Blanc, N. Navaratnam, J. O. Henderson, S. Anant, S. Kennedy, A. Jarmuz, J. Scott and N. O. Davidson, Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing, *J. Biol. Chem.* **276**, 10272–10283 (2001).
70. P. P. Lau, S. H. Chen, J. C. Wang and L. Chan, A 40 kilodalton rat liver nuclear protein binds specifically to apolipoprotein B mRNA around the RNA editing site, *Nucleic Acids Res.* **18**, 5817–5821 (1990).
71. P. P. Lau, H. J. Zhu, M. Nakamuta and L. Chan, Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing, *J. Biol. Chem.* **272**, 1452–1455 (1997).
72. D. Schock, S. R. Kuo, M. F. Steinburg, M. Bolognino, J. D. Sparks, C. E. Sparks and H. C. Smith, An auxiliary factor containing a 240-kDa protein complex is involved in apolipoprotein B RNA editing, *Proc. Natl. Acad. Sci. USA* **93**, 1097–1102 (1996).
73. S. Yamanaka, K. S. Poksay, M. E. Balestra, G. Q. Zeng and T. L. Innerarity, Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc motif is essential for catalytic activity, and noncatalytic auxiliary factor(s) of the editing complex are widely distributed, *J. Biol. Chem.* **269**, 21725–21734 (1994).
74. M. Nakamuta, K. Oka, J. Krushkal, K. Kobayashi, M. Yamamoto, W. H. Li and L. Chan, Alternative mRNA splicing and differential promoter utilization determine tissue-specific expression of the apolipoprotein B mRNA-editing protein (Apobec1) gene in mice. Structure and evolution of Apobec1 and related nucleoside/nucleotide deaminases, *J. Biol. Chem.* **270**, 13042–13056 (1995).
75. P. Hodges and J. Scott, Apolipoprotein B mRNA editing: a new tier for the control of gene expression, *Trends Biochem. Sci.* **17**, 77–81 (1992).

76. N. Navaratnam, J. R. Morrison, S. Bhattacharya, D. Patel, T. Funahashi, F. Giannoni, B. B. Teng, N. O. Davidson and J. Scott, The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase, *J. Biol. Chem.* **268**, 20709–20712 (1993).
77. N. Navaratnam, T. Fujino, J. Bayliss, A. Jarmuz, A. How, N. Richardson, A. Somasekaram, S. Bhattacharya, C. Carter, J. Scott, Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition, *J. Mol. Biol.* **275**, 695–714 (1998).
78. C. Barnes and H. C. Smith, Apolipoprotein B mRNA editing in vitro is a zinc-dependent process, *Biochem. Biophys. Res. Commun.* **197**, 1410–1414 (1993).
79. A. J. MacGinnitie, S. Anant and N. O. Davidson, Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity, *J. Biol. Chem.* **270**, 14768–14775 (1995).
80. N. Navaratnam, S. Bhattacharya, T. Fujino, D. Patel, A. L. Jarmuz and J. Scott, Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site, *Cell* **81**, 187–195 (1995).
81. S. Anant, A. J. MacGinnitie and N. O. Davidson, apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, is a novel RNA-binding protein, *J. Biol. Chem.* **270**, 14762–14767 (1995).
82. D. Kontoyiannis, M. Pasparakis, T. T. Pizarro, F. Cominelli and G. Kollias, Impaired on/off regulation of TNF biosynthesis in mice lacking TNF AU-rich elements: implications for joint and gut-associated immunopathologies, *Immunity* **10**, 387–398 (1999).
83. V. Blanc, J. O. Henderson, S. Kennedy and N. O. Davidson, Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity, *J. Biol. Chem.* **276**, 46386–46393 (2001).
84. Y. Yang and H. C. Smith, Multiple protein domains determine the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing, *Proc. Natl. Acad. Sci. USA* **94**, 13075–13080 (1997).
85. Y. Yang, M. P. Sowden and H. C. Smith, Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA, *Exp. Cell Res.* **267**, 153–164 (2001).
86. K. Hirano, J. Min, T. Funahashi, D. A. Baunoch and N. O. Davidson, Characterization of the human apobec-1 gene: expression in gastrointestinal tissues determined by alternative splicing with production of a novel truncated peptide, *J. Lipid Res.* **38**, 847–859 (1997).
87. T. Fujino, N. Navaratnam and J. Scott, Human apolipoprotein B RNA editing deaminase gene (APOBEC1), *Genomics* **47**, 266–275 (1998).
88. R. M. Lee, K. Hirano, S. Anant, D. Baunoch and N. O. Davidson, An alternatively spliced form of apobec-1 messenger RNA is overexpressed in human colon cancer, *Gastroenterology* **115**, 1096–1103 (1998).
89. S. Anant, D. Mukhopadhyay, K. Hirano, T. A. Brasitus and N. O. Davidson, Apobec-1 transcription in rat colon cancer: decreased apobec-1 protein production through alterations in polysome distribution and mRNA translation associated with upstream AUGs, *Biochim. Biophys. Acta* **1575**, 54–62 (2002).
90. K. Hirano, J. Min, T. Funahashi and N. O. Davidson, Cloning and characterization of the rat apobec-1 gene: a comparative analysis of gene structure and promoter usage in rat and mouse, *J. Lipid Res.* **38**, 1103–1119 (1997).
91. A. Jarmuz, A. Chester, J. Bayliss, J. Gisbourne, I. Dunham, J. Scott and N. Navaratnam, An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22, *Genomics* **79**, 285–296 (2002).

92. W. Liao, S. H. Hong, B. H. Chan, F. B. Rudolph, S. C. Clark and L. Chan, APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family, *Biochem. Biophys. Res. Commun.* **260**, 398–404 (1999).
93. S. Anant, D. Mukhopadhyay, V. Sankaranand, S. Kennedy, J. O. Henderson and N. O. Davidson, ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant negative effect on C to U RNA editing, *Am. J. Physiol. Cell Physiol.* **281**, C1904–C1916 (2001).
94. M. Muramatsu, V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson and T. Honjo, Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells, *J. Biol. Chem.* **274**, 18470–18476 (1999).
95. R. S. Harris, J. E. Sale, S. K. Petersen-Mahrt and M. S. Neuberger, AID is essential for immunoglobulin V gene conversion in a cultured B cell line, *Curr. Biol.* **12**, 435–438 (2002).
96. H. Arakawa, J. Hauschild and J. M. Buerstedde, Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion, *Science* **295**, 1301–1306 (2002).
97. A. Mehta, S. Banerjee and D. M. Driscoll, Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA in vitro, *J. Biol. Chem.* **271**, 28294–28299 (1996).
98. A. Mehta and D. M. Driscoll, A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA, *Mol. Cell. Biol.* **18**, 4426–4432 (1998).
99. N. O. Davidson, R. C. Carlos and A. M. Lukaszewicz, Apolipoprotein B mRNA editing is modulated by thyroid hormone analogs but not growth hormone administration in the rat, *Mol. Endocrinol.* **4**, 779–785 (1990).
100. D. Mukhopadhyay, M. Plateroti, S. Anant, F. Nassir, J. Samarut and N. O. Davidson, Thyroid hormone regulates hepatic triglyceride mobilization and apolipoproteinB mRNA editing in a murine model of congenital hypothyroidism. *Endocrinology* **144**, 711–719 (2002).
101. A. Lorentz, D. Plonne, H. P. Schulze and R. Dargel, Dexamethasone enhanced by insulin, but not by thyroid hormones stimulates apolipoprotein B mRNA editing in cultured rat hepatocytes depending on the developmental stage, *FEBS Lett.* **391**, 57–60 (1996).
102. E. Levy, D. Sinnett, L. Thibault, T. D. Nguyen, E. Delvin and D. Menard, Insulin modulation of newly synthesized apolipoproteins B-100 and B-48 in human fetal intestine: gene expression and mRNA editing are not involved, *FEBS Lett.* **393**, 253–258 (1996).
103. M. A. von Wronski, K. I. Hirano, L. M. Cagen, H. G. Wilcox, R. Raghov, F. E. Thorngate, M. Heimberg, N. O. Davidson and M. B. Elam, Insulin increases expression of apobec-1, the catalytic subunit of the apolipoprotein B mRNA editing complex in rat hepatocytes, *Metabolism* **47**, 869–873 (1998).
104. A. Mehta and D. M. Driscoll, Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing, *Rna* **8**, 69–82 (2002).
105. J. O. Henderson, V. Blanc and N. O. Davidson, Isolation, characterization and developmental regulation of the human apobec-1 complementation factor (ACF) gene, *Biochim. Biophys. Acta* **1522**, 22–30 (2001).
106. G. S. Dance, M. P. Sowden, L. Cartegni, E. Cooper, A. R. Krainer and H. C. Smith, Two proteins essential for apolipoprotein B mRNA editing are expressed from a single gene through alternative splicing, *J. Biol. Chem.* **277**, 12703–12709 (2002).
107. M. L. Hastings and A. R. Krainer, Pre-mRNA splicing in the new millennium, *Curr. Opin. Cell Biol.* **13**, 302–309 (2001).
108. J. F. Cáceres and A. R. Kornblihtt, Alternative splicing: multiple control mechanisms and involvement in human disease, *Trends Genet.* **18**, 186–193 (2002).
109. C. E. Harris, R. A. Boden and C. R. Astell, A novel heterogeneous nuclear ribonucleoprotein-like protein interacts with NS1 of the minute virus of mice, *J. Virol.* **73**, 72–80 (1999).

110. C. R. Astell, Q. Liu, C. E. Harris, J. Brunstein, H. K. Jindal and P. Tam, Minute virus of mice cis-acting sequences required for genome replication and the role of the trans-acting viral protein, NS-1, *Prog. Nucleic Acid Res. Mol. Biol.* **55**, 245–285 (1996).
111. A. B. Shyu, M. E. Greenberg and J. G. Belasco, The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways, *Genes Dev.* **3**, 60–72 (1989).
112. M. E. Greenberg, A. B. Shyu and J. G. Belasco, Deadenylation: a mechanism controlling c-fos mRNA decay, *Enzyme* **44**, 181–192 (1990).
113. P. L. Bernstein, D. J. Herrick, R. D. Prokipcak and J. Ross, Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant, *Genes Dev.* **6**, 642–654 (1992).
114. I. Lemm and J. Ross, Regulation of c-myc mRNA decay by translational pausing in a coding region instability determinant, *Mol. Cell. Biol.* **22**, 3959–3969 (2002).
115. C. Y. Chen, Y. You and A. B. Shyu, Two cellular proteins bind specifically to a purine-rich sequence necessary for the destabilization function of a c-fos protein-coding region determinant of mRNA instability, *Mol. Cell. Biol.* **12**, 5748–5757 (1992).
116. C. Grosset, C. Y. Chen, N. Xu, N. Sonenberg, H. Jacquemin-Sablon and A. B. Shyu, A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex, *Cell* **103**, 29–40 (2000).
117. Z. Mourelatos, L. Abel, J. Yong, N. Kataoka and G. Dreyfuss, SMN interacts with a novel family of hnRNP and spliceosomal proteins, *EMBO J.* **20**, 5443–5452 (2001).
118. M. Sendtner, Molecular mechanisms in spinal muscular atrophy: models and perspectives, *Curr. Opin. Neurol.* **14**, 629–634 (2001).
119. S. M. Carty and A. L. Greenleaf, Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing, *Mol. Cell. Proteomics* **1**, 598–610 (2002).
120. L. Pellizzoni, B. Charroux, J. Rappsilber, M. Mann and G. Dreyfuss, A functional interaction between the survival motor neuron complex and RNA polymerase II, *J. Cell Biol.* **152**, 75–85 (2001).
121. X. Lu, N. A. Timchenko and L. T. Timchenko, Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy, *Hum. Mol. Genet.* **8**, 53–60 (1999).
122. A. Mankodi, E. Logigian, L. Callahan, C. McClain, R. White, D. Henderson, M. Krym and C. A. Thornton, Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat, *Science* **289**, 1769–1773 (2000).
123. A. V. Philips, L. T. Timchenko and T. A. Cooper, Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy, *Science* **280**, 737–741 (1998).
124. A. N. Ladd, N. Charlet and T. A. Cooper, The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing, *Mol. Cell. Biol.* **21**, 1285–1296 (2001).
125. B. N. Charlet, P. Logan, G. Singh and T. A. Cooper, Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing, *Mol. Cell* **9**, 649–658 (2002).
126. R. S. Savkur, A. V. Philips and T. A. Cooper, Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy, *Nat. Genet.* **29**, 40–47 (2001).
127. Z. Saba, R. Nassar, R. M. Ungerleider, A. E. Oakeley and P. A. Anderson, Cardiac troponin T isoform expression correlates with pathophysiological descriptors in patients who underwent corrective surgery for congenital heart disease, *Circulation* **94**, 472–476 (1996).
128. P. A. Anderson, N. N. Malouf, A. E. Oakeley, E. D. Pagni and P. D. Allen, Troponin T isoform expression in the normal and failing human left ventricle: a correlation with myofibrillar ATPase activity, *Basic Res. Cardiol.* **87 Suppl. 1**, 117–127 (1992).

- 129. B. N. Charlet, R. S. Savkur, G. Singh, A. V. Philips, E. A. Grice and T. A. Cooper, Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol. Cell* **10**, 45–53 (2002).
- 130. A. Mankodi, M. P. Takahashi, H. Jiang, C. L. Beck, W. J. Bowers, R. T. Moxley, S. C. Cannon and C. A. Thornton, Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy, *Mol. Cell* **10**, 35–44 (2002).
- 131. H. Suzuki, Y. Jin, H. Otani, K. Yasuda and K. Inoue, Regulation of alternative splicing of alpha-actinin transcript by Bruno-like proteins, *Genes Cells* **7**, 133–141 (2002).
- 132. D. Mukhopadhyay, C. Houchen, S. Kennedy, B. K. Dieckgraefe and S. Anant, Coupled mRNA stabilization and translational silencing of cyclooxygenase-2 by a novel RNA binding protein, CUGBP2, *Mol. Cell* **11**, 113–126 (2003).
- 133. M. Yu, R. H. Haslam and D. B. Haslam, EDJ, an Hsp40 co-chaperone localized to the endoplasmic reticulum of human cells, *J. Biol. Chem.* **275**, 24984–24992 (2000).
- 134. G. Dreyfuss, Structure and function of nuclear and cytoplasmic ribonucleoprotein particles, *Annu. Rev. Cell Biol.* **2**, 459–498 (1986).
- 135. T. Y. Nakagawa, M. S. Swanson, B. J. Wold and G. Dreyfuss, Molecular cloning of cDNA for the nuclear ribonucleoprotein particle C proteins: a conserved gene family, *Proc. Natl. Acad. Sci. USA* **83**, 2007–2011 (1986).

This Page Intentionally Left Blank

Polymorphisms in the Genes Encoding Members of the Tristetraprolin Family of Human Tandem CCCH Zinc Finger Proteins

PERRY J. BLACKSHEAR,^{*,†}
RUTH S. PHILLIPS,^{*,†}
JOHANA VAZQUEZ-MATIAS,[§] AND
HARVEY MOHRENWEISER[§]

^{*}*Office of Clinical Research and
Laboratory of Signal Transduction, A2-
05 National Institute of Environmental
Health Sciences, 111 Alexander Drive,
Research Triangle Park, NC 27709, USA;*

[†]*Department of Medicine and
Biochemistry, Duke University Medical
Center, Durham, NC 27710, USA; and*

[§]*Biology and Biotechnology Research
Program, Lawrence Livermore National
Laboratory, Livermore, CA 94551-9900,
USA*

I. Introduction	44
II. Materials and Methods	45
III. Results	52
IV. Discussion	63
Acknowledgments	66
References	66

The three known mammalian CCCH tandem zinc finger proteins of the tristetraprolin (TTP) class have recently been demonstrated to be mRNA-binding proteins. The prototype, TTP, functions in normal physiology to promote the instability of the tumor necrosis factor α (TNF α) and granulocyte-macrophage colony-stimulating factor mRNAs. Conversely, these mRNAs are stabilized in TTP-deficient mice, leading to an inflammatory phenotype characterized by overproduction of these cytokines. To explore sequence variations in TTP and its two related proteins, we sequenced genomic DNA encoding the TTP protein (*ZFP36*) and those of its two known mammalian relatives, *ZFP36L1* and *ZFP36L2*, from 72 to 92 anonymous human subjects from various geographical and ethnic backgrounds. We also sequenced *ZFP36* in genomic DNA from 92 subjects exhibiting evidence of excessive TNF α action. The resequencing strategy identified 13 polymorphisms in the protein-coding regions of these

three genes, of which six would result in amino acid changes; other putative polymorphisms were identified by EST searches. One mutation in *ZFP36L1* was a dinucleotide substitution that would prevent splicing of the single intron. This mutation was identified in only one allele of the original 144 sequenced from an adult female Aka Pygmy from the Central African Republic; a second individual with the same variant allele was found by genotyping 58 additional Aka DNA samples. Analysis of mRNA from one of these subject's lymphoblasts confirmed that *ZFP36L1* mRNA levels were approximately 50% of those in a comparable sample without the mutation. The functional significance of this and the other polymorphisms identified remains to be determined by both biochemical and population linkage studies. © 2003 Elsevier Science

I. Introduction

The tristetraprolin (TTP) family of CCCH tandem zinc finger proteins contains three known members in mammals and a fourth in *Xenopus laevis* and fish (1,2). TTP, also known as Tis11, Nup475, and G0S24, is encoded by the human gene *ZFP36* (OMIM number 190900) (3), and was discovered through its rapid transcriptional induction in cells stimulated with insulin, serum, or phorbol esters (4–7). Studies in TTP-deficient mice have shown that TTP is involved in the physiological regulation of the secretion of tumor necrosis factor α (TNF α) (8) and granulocyte-macrophage colony-stimulating factor (GM-CSF) (9). TTP somehow destabilizes the mRNAs encoding these proteins after binding to AU-rich elements (ARE) located in their 3'-untranslated regions (3'UTR), leading to the accelerated destruction of these mRNAs (1,8–11).

TTP deficiency in mice leads to a chronic inflammatory state, characterized by weight loss, severe erosive arthritis, conjunctivitis, dermatitis, splenomegaly, adenopathy, autoimmunity, and severe medullary and extra-medullary myeloid hyperplasia (12). Most aspects of this syndrome appear to be due to chronically elevated concentrations of TNF α (8,12,13), and the chronically elevated GM-CSF levels may be partly responsible for the myeloid hyperplasia (9,14). Thus, complete TTP deficiency in man would be expected to result in a severe inflammatory disease with myeloid hyperplasia, probably not compatible with life. The heterozygous state in the mouse is also associated with a similar syndrome, although its onset is later in life, and the syndrome is less severe (E. Carballo and P. J. Blackshear, unpublished data). Thus, it is possible that the heterozygous state in humans would be compatible with life, and that less severe mutations affecting TTP's biosynthesis, turnover or function might result in human disease.

Although the phenotypes of mouse knockouts of the two known mammalian *ZFP36* relatives, now officially designated as *ZFP36L1* (for

ZFP36-like 1) (15) and *ZFP36L2* (16), have not been published to date, we have recently shown that their encoded proteins can exert similar effects to TTP to destabilize TNF α and GM-CSF mRNAs in a cell cotransfection system (1). The genes encoding these proteins are regulated differently from the *ZFP36*, but given the similar mRNA binding and destabilizing properties to TTP of the encoded proteins, it seems possible that abnormalities in their expression or primary sequence might also be associated with clinical disease.

To begin to explore the natural variations in the protein-coding regions of these three genes in man, we initially sequenced genomic DNA encompassing their protein-coding regions from the 72 individuals from various ethnic groups that make up the initial study population for the Environmental Genome Project being carried out by the National Institute of Environmental Health Sciences (17–19). We also resequenced several sets of DNA from other pertinent groups of subjects, and searched for polymorphisms in the protein-coding regions of these genes by EST analysis. Altogether, we identified 11 polymorphisms in the protein-coding regions of these genes that would result in amino acid changes, including a dinucleotide mutation at an intron splice site that appears to lead to hemizygous expression of the *ZFP36L1* mRNA.

II. Materials and Methods

A. Nomenclature

The approved nomenclature for the human gene encoding TTP (also known as TIS11, GOS24, and Nup475) is *ZFP36* (3) (OMIM number 190700). This stands for zinc finger protein 36. The approved nomenclature for the second human member of the family is ZFP36-like 1, *ZFP36L1* (OMIM number 601064). Other names for the encoded protein are BRF1, TIS11B, cMG1, Berg-36, and ERF1. The approved nomenclature for the third family member is ZFP36-like 2, *ZFP36L2* (no OMIM number); the encoded protein is also known as TIS11d, ERF2, and BRF2. Approved gene symbols were obtained from the HUGO Gene Nomenclature Committee (HGNC). Further details are available at URL <http://www.gene.ucl.ac.uk/nomenclature/> or e-mail nome@galton.ucl.ac.uk.

B. Subjects and DNA

Lymphoblasts representing the 72 anonymous subjects in the initial arm of the Environmental Genome Project were obtained from the Coriell Collection. Details of this collection, as well as culture conditions and genomic DNA

extraction methods, are described in detail in Ref. (17). Briefly, the group labeled Asian (A) contained representatives from the Taiwanese Ami (5 subjects), Indo-Pakistan (4), Cambodian (3), Beijing Chinese (5), Japanese (3), and Melanesian (4) groups. The group labeled Black (B) contained African American (15) and Pygmy subjects from the Central African Republic (4) and Zaire (5). The group labeled Caucasian (C) contained Druze (5), Adygei (5), Moscow Russian (5), and Utah (9) subjects. In the case of *ZFP36*, we also sequenced the promoter, portions of the 5'- and 3'UTRs, and the single intron, from the original 72 subjects as well as from 20 additional North American individuals from the Coriell Polymorphism Discovery collection, 42 individuals that were either hypo- or hyperresponsive to inhaled endotoxin in terms of their bronchial resistance (kindly provided by Dr. David A. Schwartz, Duke University Medical Center, Durham, NC; see Ref. (20)), and 50 individuals with a systemic inflammatory syndrome resembling the Tumor necrosis factor Receptor-Associated Periodic Syndrome (TRAPS) (21) that were negative for mutations in the TNF α receptor (kindly provided by Dr. Ivona Aksentijevich and Daniel L. Kastner, National Institutes of Health, Bethesda, MD). For specific mutations, we used other genotyping strategies, as described in the next section, in DNA from 58 additional Aka Pygmy subjects (kindly provided by Dr. L. Luca Cavalli-Sforza and Alice A. Lin, Stanford University, Palo Alto, CA), 430 anonymous residents of Durham, NC of various ethnicities (kindly provided by Dr. Douglas Bell, NIEHS, Research Triangle Park, NC), and 20 DNA samples from germ cell testicular cancers (kindly provided by Dr. David Hogg, University of Toronto, Canada) (22).

C. Sequencing Strategy

The strategy of resequencing of PCR products containing the exons plus splice sites and the 5' and 3' regions of genes as an efficient strategy for identification of genetic variation, especially sequence variation resulting in amino acid substitutions, in heterozygote individuals in a population has been discussed (23–25). PCR primers were located so that amplification of the genomic sequence was initiated approximately 50 bp from the intron–exon boundary, and the PCR products were 350–450 bp.

D. PCR-Amplification Conditions

PCR primers were designed using the Oligo Primer Analysis Software (National Biosciences, Inc., Plymouth, MN). Appended to the 5' end of each of the PCR primers were sequences containing the primer-binding sites for the forward or reverse energy transfer DNA-sequencing primers (Amersham Life Science, Cleveland, OH). PCR primers were matched so that the sense

and the antisense PCR primers contained different sequencing primer-binding sites. Other procedures are as described (25).

E. DNA Sequencing

The PCR products were diluted 10-fold with TE [10 mM Tris-HCl (pH 8.0 at 20°C), 0.1 mM EDTA] and directly sequenced in both directions as described (25). Sequencing reactions were performed according to the manufacturer's instructions using the DYEnamic Direct cycle sequencing kit with the DYEnamic ET primers (Amersham Life Science, Inc., Cleveland, OH).

F. Sequence Analysis

The initial data analysis (lane tracking and base calling) was performed with the ABI Prism DNA Sequence Analysis Software (version 2.1.2). The chromatograms were reanalyzed with Phred (bases called and quality of sequence values assigned, version 0.961028), assembled with Phrap (version 0.960213), and the resulting data viewed with Consed (version 4.1). Description and documentation for Phred, Phrap, and Consed may be obtained at <http://www.genome.washington.edu>. "PolyPhred" (version 2.1), a software package that utilizes the output from Phred, Phrap, and Consed to identify nucleotide substitutions in heterozygote individuals (24,26), was utilized to identify presumptive variants.

G. Other Methods

A dinucleotide polymorphism identified in one chromosome of a single subject was predicted to destroy an intron splice site within the protein-coding region of *ZFP36L1*. In this case, the DNA sequence traces in both directions were retrieved to confirm the presence of this mutation, and a novel restriction fragment length polymorphism (RFLP) was identified at the site in the form of a new site for cleavage by Fnu4HI and TseI. The subject's DNA was used to confirm the presence of this novel restriction site by amplifying genomic DNA with PCR primers based on neighboring sequences, followed by cleavage with TseI under standard conditions. The forward primer used was 5'-GATGACCACCACCCTCGTGTC-3' (complementary to bp 650-670 of GenBank accession number X79066) and the reverse primer was 5'-CCCATTTGCCTTCCAAGACCC-3' (complementary to bp 781-800). To confirm that the splice site mutation resulted in failure of mRNA splicing, lymphoblasts from this subject were grown under standard conditions, and the cells were harvested in log phase growth and used for the preparation of total cellular RNA as described (1). This RNA, and a

parallel sample from another member of the same ethnic group, was used for northern blotting of ZFP36L1 mRNA, using a probe based on the protein-coding region of the orthologous rat cDNA sequence (GenBank accession number X52590, bases 110–1120) (10). ZFP36L1 mRNA expression was quantitated by PhosphorImager analysis, and corrected for gel loading by cohybridization with a GAPDH cDNA probe as described (8).

This RFLP assay was also used to screen the 58 additional Aka DNA samples described earlier.

Other genotyping was performed on various sample sets for selected polymorphisms as described (27) (QIAGEN Genomics, Inc., Bothell, WA). For the trinucleotide repeat polymorphism detected in ZFP36L2, resequencing of this region using selected sample sets was performed by methods similar to those described earlier for the primary resequencing project.

H. EST Searches

One class of EST searches focused on identifying ESTs that contained polymorphisms already documented by the current sequencing effort. In this case, approximately 30 bp on either side of the polymorphism was included in a blastn search of the current human EST database. No attempt was made to examine the original EST sequence traces for accuracy, with the following exception: in the case of one EST, a 6-bp CAG repeat polymorphism was identified in ZFP36L2 that would result in an increase from seven to nine consecutive encoded glutamine residues; in this case, the original GenBank traces were examined, and one of the EST clones was obtained and resequenced in both directions using dRhodamine Terminator Cycle Sequencing (Perkin-Elmer, Foster City, CA) to confirm the presence of the extra CAGCAG sequence.

A second EST searching effort was directed at finding new polymorphisms that had not been discovered by the resequencing effort. To this end, the edited reference sequences (see below) spanning only the protein-coding regions of the three genes under consideration were used to search the GenBank human EST database, using the program blastn and the “flat query-anchored with identities” alignment view. Putative polymorphisms that were seen three times or more were then checked by reblasting the database with the putative polymorphism flanked by 30 bp of genomic sequence on either side. If the polymorphism occurred three or more times, in what appeared to be an otherwise reliable sequence, it was counted as a putative polymorphism and included in Tables I, II and III, along with the frequency of occurrence in the GenBank human EST collection.

TABLE I
ZFP36 POLYMORPHISMS

Polymorphism	Location	B	Base change	Sequence	Amino acid change	Variant allele frequency in 92 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
ZFP36*1	Promoter	316	C → A	CCCCC(C/A)ATCCG		1.8		0	0	0
ZFP36*2	Promoter	359	A → G	CGGTC(A/G)CGGCT		47		29	30	32
ZFP36*3	Promoter	490	C → A	CCGGC(C/A)CCGGC		3.1		0	100	0
ZFP36*4	Promoter	492	C → T	GGCCC(C/T)GGCCC		0.6		0	100	0
ZFP36*5	Intron	1226	G → A	GGGAA(G/A)CCGGG		0.5		0	100	0
ZFP36*6	Intron	1256	C → G	TAAGG(C/G)CTCGG		0.5		0	100	0
ZFP36*7	PCD (ex. 2)	1525	C → T	CGGGA(C/T)CCTGG	P37 → S	0.6	0/127	0	0	0
ZFP36*8	PCD (ex. 2)	1725	C → T	TCGCG(C/T)TACAA	R103 → R	6.2	2/127 (1.6%)	22	44	33
ZFP36*9	PCD (ex. 2)	2235	T → C	CCCTC(T/C)GTACA	S273 → S	4.2	1/69 (1.4%)	0	50	50
ZFP36*10	3'UTR	2980	Del TT	TTTTT(delTT)GTAAT		7.6	64/249 (26%)	18	27	55
ZFP36*11 ^a	PCD (ex. 2)	1807	G → T	GCCTG(G/T)GCCGAG	G131 → C		5/118 (4.2%)			
ZFP36*12 ^a	PCD (ex. 2)	2112	C → T	GCCTT(C/T)TCTGC	F232 → F		4/97 (4.1%)			
ZFP36*13 ^a	PCD (ex. 2)	2184	C → A	AGGGC(C/A)ACTCC	A256 → A		10/94 (11%)			
ZFP36*14 ^b	Intron	783	C → T	TGCCT(C/T)CCGCT				0	0	0

These data are based on the following accession numbers for human ZFP36: gene, M92844; cDNA, M63625; protein, AAA61240. B refers to the base number in the genomic sequence M92844; AA refers to the amino acid number in AAA61240. The polymorphic changes are indicated as follows: C (original base or amino acid) → T (polymorphic base or amino acid). See Fig. 1 for a depiction of the gene, sequencing primers, and location of polymorphisms.

PCD, protein-coding domain; ex., exon; 3'UTR, 3'-untranslated region; 5'UTR, 5'-untranslated region; EGP, Environmental Genome Project. A, Asian; B, Black; C, Caucasian.

Variants shown as having frequencies of 0 for all three ethnic groups were found in the 20 added EGP samples of unknown ethnicity.

^aPolymorphisms identified by EST searches only.

^bFound in 1/162 (0.6%) of alleles from bronchial sensitivity and TRAPS-like syndrome subjects, not found in 92 EGP subjects.

TABLE II
ZFP36L1 POLYMORPHISMS

Polymorphism	Location	B	Change	Sequence	Amino acid change	Variant allele frequency in 72 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
ZFP36L1*1	5'UTR	644	G → T	CGAAC(G/T)CACAG		12.5	6/60 (10%)	0	56	44
ZFP36L1*2	PCD1	706	AG → GC	ATGCA(AG/GC)GTAAA	K19 → S	0.7	0/65 ^a	0	100	0
ZFP36L1*3	Intron 1	729	G → A	CATTT(G/A)CTTTT		0.7		0	100	0
ZFP36L1*4	Intron 1	772	C → CC	ACCCC(C/CC)AAAAG		18.8		7.4	56	37
ZFP36L1*5	Intron 1	804	A → G	GGGAA(A/G)GTGGT		6.2		0	0	100
ZFP36L1*6	Intron 1	845	G → C	TTTCT(G/C)CCAAG		4.2		100	0	0
ZFP36L1*7	PCD2	3685	G → A	ACCGG(G/A)CTGCT	G218 → G	0.7	0/40	100	0	0
ZFP36L1*8	PCD2	3915	C → A	CAGCC(C/A)TCAGG	P295 → H	0.7	0/46	100	0	0

These data are based on accession numbers X79066 and X79067 for the 5' and 3' regions of the gene, X99404 for the cDNA, and CAA67781 for the protein. The intron sequence between X79066 and X79067 was supplied by the chromosome 14 BAC clone AL 133313.2.2. The numbering system used for polymorphisms ZFP36L1*1–6 was from X79066. The numbering system for polymorphisms ZFP36L1*7 and P8 was based on interposing 1347 bp between X79066 and X79067, and numbering the bases consecutively starting at the 3' end of X79066. The corresponding base numbers for polymorphisms ZFP36L1*7 and ZFP36L1*8 in X79067 are 1365 and 1595, respectively. See the legend to Table I for abbreviations and other details.

^aOne EST continued into the intron (BF873589).

TABLE III
ZFP36L2 POLYMORPHISMS

Poly-morphism	Location	B	Change	Sequence	Amino acid change	Variant allele frequency in 72 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
ZFP36L2*1	5'UTR	1591	Del	CTGCC(GCTCCGGCCACTGCG/del) GGATC		0.7	0/11	0	100	0
ZFP36L2*2	Intron 1	1753	C → A	ACGCC(C/A)CTCCC		0.7		100	0	0
ZFP36L2*3	PCD2	2389	C → T	GCCGG(C/T)GGTCC	G93 → G	0.7	0/20	100	0	0
ZFP36L2*4	PCD2	2435	G → A	GCGGC(G/A)GCACA	G109 → S	2.8	1/27 (3.7%)	25	0	75
ZFP36L2*5	PCD2	2445	T → A	AGCCC(T/A)GCTCA	L112 → Q	1.4	0/28	0	100	0
ZFP36L2*6	PCD2	2719	C → T	ACCAT(C/T)GGCTT	I203 → I	6.25	20/78 (26%)	22	44	33
ZFP36L2*7	PCD2	2870	C → T	CCAAG(C/T)TGCAC	L254 → L	48	52/97 (54%)	20	57	20
ZFP36L2*8	PCD2	3055	C → T	ACGCC(C/T)TCGGG	P315 → P	2.1	0/70	33	0	67
ZFP36L2*9	PCD2	3323	G → T	CCCCC(G/T)CGCAG	A407 → S	1.4	0/204	0	100	0
ZFP36L2*10	3'UTR	3601	G → A	AGGGC(G/A)CCAGT		0.7	0/106	100	0	0
ZFP36L2*11 ^a	PCD2	3055	C → A	ACGCC(C/A)TCGGG	P315 → P		3/69 (4.3%)			
ZFP36L2*12 ^a	PCD2	3317	Insert (CAG)2	AGCAG(insertCAGCAG) GGCCT	Insert QQ after Q401		6/>200			
ZFP36L2*13 ^a	PCD2	3458	TC → CT	CGGAC(TC/CT)GCTGT	S452 → L		4/>200			
ZFP36L2*14 ^a	PCD2	3536	G → A	GCCTC(G/A)ACCCT	D478 → N		3/>150			
ZFP36L2*15 ^a	PCD2	3579	C → G	CATCT(C/G)CGACG	S492 → C		5/>100			

These data are based on accession numbers U07802 (gene), NM_006887.3 (cDNA) and NP_008818 (protein). The numbering of DNA and protein polymorphisms is based on the numbering systems used in U07802 and NP_008818, respectively. See the legend to Table I for abbreviations and other details.

^aPolymorphisms identified from EST searches only.

III. Results

A schematic representation of the results of this study is shown in Fig. 1. It depicts human *ZFP36*, *ZFP36L1*, and *ZFP36L2*, as well as the location of exons, introns, protein-coding regions, and the 64-amino acid tandem zinc finger domain within the genomic sequences. It also shows the orientation of sequencing primer pairs, the location of the PCR products generated, and the locations of the polymorphisms detected by the present resequencing effort. These polymorphisms are named, for example, *ZFP36*4*, to indicate polymorphism 4 within *ZFP36*; this nomenclature convention was approved by the Human Genome Variation Society <http://ariel.ucs.unimelb.edu.au/~cotton/mdi.htm> (S. E. Antonarakis, personal communication).

The polymorphisms determined by the resequencing effort are indicated below the line representing each gene by the arrowheads labeled, e.g., *4. The putative polymorphisms generated solely by the EST searches are indicated with an arrowhead and numbered asterisk above the gene. In the cases of *ZFP36L1* and *ZFP36L2*, our focus was on protein-coding regions, but some intronic sequence was determined as a byproduct of this effort. In the case of *ZFP36*, sequence was also determined for a regulatory region of the promoter (28); the complete intron, which also functions to regulate expression (29); and an AU-rich region in the 3'UTR, which is at least partly responsible for the instability of this message (W. S. Lai and P. J. Blackshear, unpublished data). Detailed information on the numbered polymorphisms identified in each of the three genes can be found in Tables I, II and III.

A. *ZFP36* Polymorphisms

The present resequencing effort confirmed the cDNA and protein GenBank reference sequences (RefSeq) NM_003407 and NP_003398 at every position. There were many differences between our results and the genomic sequence M92844; however, every one of the differences found in the resequencing effort agreed with the chromosome 19 sequence, AC0011500. These differences will not be described here. However, on the basis of these comparisons, we have concluded that the "wild-type" human genomic *ZFP36* sequence corresponding to part of the promoter, most of the two exons, and the single intron, is accurately represented by bp 11,119–14,350 of the chromosome 19 sequence AC0011500.

In the genomic DNA samples from the core 72 individuals, there were no polymorphisms in the *ZFP36* protein-coding region that resulted in amino acid changes. We therefore expanded the sequencing effort to include DNA from a separate collection of 20 anonymous North American subjects (Coriell Discovery Set) (Fig. 1 and Table I). The following discussion therefore applies to all 92 subjects or 184 alleles. In the key regulatory region of the promoter,

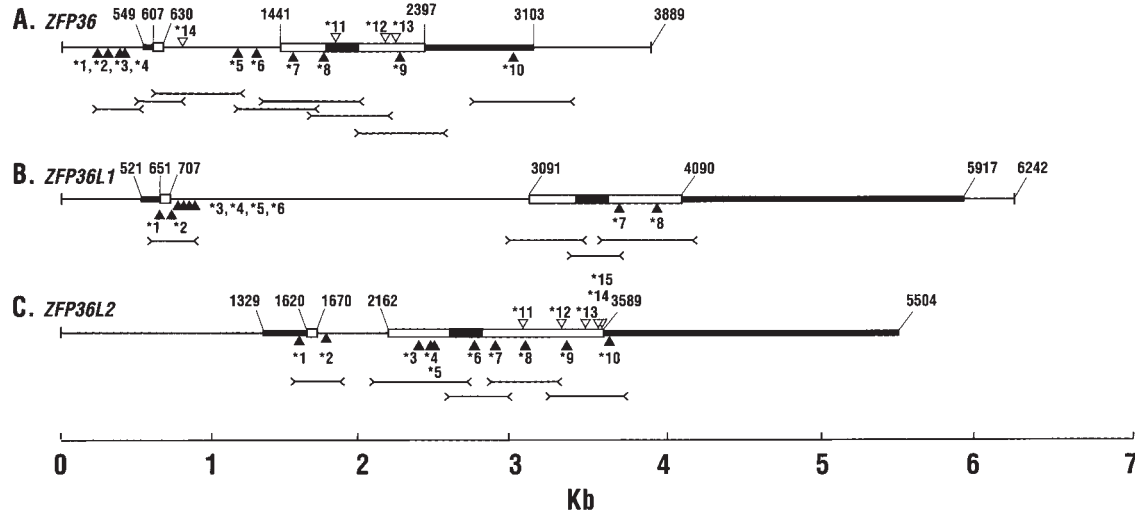


FIG. 1. Schematic representation of genes encoding human CCH proteins. The three genes encoding the known human CCH proteins with tandem zinc fingers are illustrated schematically. For all genes, the length in kb is indicated at the bottom of the figure. Genomic noncoding sequences, including introns, are indicated by a thin solid line; the 5'- and 3'UTRs by thick solid lines; and the protein-coding domains by thick open lines, in all three cases separated by the single intron. The gray regions within each of the second protein-coding regions represent the locations of the tandem zinc finger domains. The numbers at the top of each of gene show the location of specific bases within the primary sequence; the arrowheads labeled using numbers with asterisks below the indicated genes represent the sites of the numbered polymorphisms identified by sequence analysis, described in the text and in Tables I, II and III; the numbered arrowheads with asterisks above the genes represent the locations of putative polymorphisms based solely on EST analysis, but include those verified by direct sequencing of ESTs or by other means. The locations of the sequenced fragments are indicated below each gene by a solid line connecting two primer pairs, with > and < representing the binding sites of forward and reverse primers, respectively. The numbering systems used are: for *ZFP36* (A), the numbering system is from the genomic sequence in GenBank accession number M92844; for *ZFP36L1* (B), the numbering system begins with the genomic clone X79066, continues with 1347 bp of intronic sequence from AL132986.2, and then finishes with X79067, as described in detail in Table II; for *ZFP36L2* (C), the numbering system is from the genomic sequence U07802. See the text for further details.

four SNPs were identified (Fig. 1 and Table I). The first, polymorphism *ZFP36*1*, involved a single nucleotide polymorphism (SNP) at position 316 in GenBank accession number M92844. This was seen only once, in the additional 20 DNA samples from subjects of unknown ethnicity, and occurred at a 1.8% overall frequency. A second promoter SNP was seen at base 359 (*ZFP36*2*); this was very common, occurring in 47% of tested alleles, and was roughly equally distributed among the three major ethnic groups. It was found at a lower frequency in 15.4% of the alleles from subjects with bronchial hyperresponsiveness and in 34.1% of the alleles from patients with TRAPS-like syndromes. Two other promoter SNPs were identified, *ZFP36*3* and *ZFP36*4*, occurring at frequencies of 3.1% and 0.6%, respectively, in the 92 core samples. *ZFP36*3* was present in six alleles from Black individuals, and two alleles were from Americans of unknown ethnicity. Neither of these two uncommon SNPs was found in the bronchial sensitivity subjects or patients with TRAPS-like syndromes. None of the four promoter SNPs was in the three key enhancer elements previously identified as regulating serum-induced transcription of *ZFP36* (28).

Although the entire intron was sequenced in all of the 92 DNA samples, only two intron SNPs were identified (*ZFP36*5* and *P6*), and neither one was in the key regulatory regions that have previously been identified in the intron (29). Both were rare, occurring in only one allele of the 184 studied, in both cases in Black individuals. They were not found in the bronchial sensitivity subjects or the TRAPS-like patients.

Three SNPs occurred in the protein-coding region, in all cases within the second exon. The first, *ZFP36*7*, occurred at position 1525 in accession number M92844. This resulted in a change from a proline to a serine at position 37 in AAA61240. This was a rare polymorphism in this study sample, occurring at a frequency of 0.6% in individuals of unknown ethnicity, and was not seen in 127 ESTs spanning that region when searched on 11/15/02. Using a separate assay, we also evaluated genomic DNA from anonymous residents of Durham, NC, of varying ethnicities; of 844 alleles successfully evaluated, 834 contained C at this position, encoding a proline, and 10 contained T, encoding a serine, for an SNP frequency of 1.2% in this population. Only one instance of this SNP was found in 92 alleles successfully examined from the TRAPS-like patients (1.1%), and it was not seen in the bronchial sensitivity subjects. Examination of the same amino acid position in the protein sequences of other organisms showed that the wild-type proline was present in the bovine sequence (30) (accession number P53781), but that this was replaced by a leucine in both the mouse (10) (accession number P26651) and rat (P47973) proteins. Although TTP-like proteins have been sequenced from both *X. laevis* (2) (accession number AAD24207) and *D. rerio* (W. S. Lai and P. J. Blackshear, unpublished data), the protein sequences at this

position in these organisms were too unrelated to make similar phylogenetic comparisons.

Two other SNPs were identified in the protein-coding region, but neither resulted in an amino acid change. *ZFP36*8* resulted in no change to arginine 103; this SNP occurred roughly equally in the three major ethnic groups tested at 6.2% of alleles tested; it was not present in the samples from the bronchial sensitivity subjects or the TRAPS-like syndromes. It was seen in 2/127 (1.6%) relevant ESTs searched on 11/15/02. *ZFP36*9* occurred in 4.2% of alleles tested, and was seen in Black and Caucasian but not in Asian samples. This failed to change serine 273, and was seen in 1 of 69 ESTs (1.4%) that spanned this region when searched on 11/15/02. It was found in 3/76 (4%) of alleles successfully examined in the subjects with bronchial sensitivity, and in 1/92 (1.1%) of alleles from the patients with the TRAPS-like syndromes.

A final polymorphism was identified within the 3'UTR. Polymorphism *ZFP36*10* comprised a deletion of two T residues beginning at position 2980 in accession number M92844. This polymorphism was found in 7.6% of the alleles, in 64 of 249 (26%) ESTs in GenBank that spanned this region when searched on 11/15/02, and was roughly equally distributed among the three ethnic groups. It was also found at a frequency of 14.5% in the subjects with bronchial sensitivity, and 17.4% in the patients with TRAPS-like syndromes.

Three additional putative SNPs were identified within the protein-coding region of the second exon of *ZFP36* by the EST search. In the case of *ZFP36*11*, there was a G to T substitution at position 1807 in the genomic sequence, which resulted in a change of glycine at position 137 to a cysteine. This was found in 5/118 (4.2%) of ESTs examined on 11/15/02. This non-conservative change is between the two zinc fingers in the mRNA tandem zinc finger-binding domain of the TTP protein; however, it did not significantly affect TTP binding to an RNA probe in a cell-free gel shift assay (W. S. Lai and P. J. Blackshear, unpublished data). Two of the polymorphic ESTs (GenBank accession numbers AV661270 and AV661568) were from the same library prepared from noncancerous liver tissue from a Chinese subject, and one was from unknown tissue from a subject of unknown ethnicity (AW952247.1). Comparison of this amino acid position with TTP orthologs from other animal species revealed that glycine in this position was conserved in TTP from mouse, rat, and cow, but not in *X. laevis*.

Two other putative SNPs in the protein-coding region were identified by the EST search, but neither affected the encoded amino acid. *ZFP36*12* was a C to T substitution at position 2112, and occurred in 4/97 ESTs (4.1%), two of which were from the same human placenta library. *ZFP36*13* was a C to A substitution at position 2184, and occurred in 10/94 (11%) ESTs at this position.

A final SNP was identified in the resequencing of samples from the bronchial sensitivity subjects or those with TRAPS-like syndromes. This SNP,

*ZFP36*14*, was found in the intron at position 783 in the genomic sequence; it was found in only 1/162 alleles (0.6%) successfully examined. It was not in one of the known transcriptional regulatory elements in this intron (29).

We also analyzed the putative SNPs identified by GenBank and listed as “variants” in accession number NM_003407. Four variants have been identified and placed into dbSNP. In three of the four cases, these putative SNPs did not meet our criteria for inclusion when EST searches were conducted on 11/15/02: dbSNP # 2229272 was not seen in 121 ESTs, # 1803662 was seen in 2/250 ESTs, and #14869 was seen in 1/229 ESTs. However, # 1042905 was seen in 17/232 (7.3%) of ESTs searched on 11/15/02, and it therefore seems reasonable to include it as *ZFP36*15*. This SNP was an A to C change within the extreme 3' end of the mRNA (b 1692 of NM_003407.1, b 3049 of M92844), and was within an AU-rich region that could be involved in the well-known instability of TTP mRNA (5). All but four of the variant transcripts at this site were from tumor specimens.

B. *ZFP36L1* Polymorphisms

We compared the current consensus sequence derived from the resequencing effort with data available in GenBank. The current consensus sequence that included the first exon corresponded exactly to bp 558–899 of X79066. The current consensus sequence differed by one nucleotide within the intron portion of *ZFP36L1*, in that G845 in X79066 was changed to C in the consensus. G845 was present in X79066.1 and XM_007458.1, but C was present in AL132986.4, Z64730, and Z61811.1. That this is a true intronic SNP is suggested by the facts that 95.8% of the 144 alleles resequenced here contained the G residue, and 4.2% contained the C residue (all Asian subjects). This is represented as SNP *ZFP36L1*6* in Table I and Fig. 1.

The consensus sequence that included the second exon spanned from bp 648 to 1844 of X79067. There were two differences between the current consensus sequence and X79067, but these have now been corrected in the RefSeq NM_004926.2 and are reflected in the current RefSeq protein sequence NP_004917.2.

A total of eight polymorphisms was found in *ZFP36L1* (Fig. 1 and Table II). A single SNP was identified in the 5'UTR (*ZFP36L1*1*); it was present in 12.5% of the alleles. Fifty-six percent of the polymorphic alleles were in Black subjects, and 44% were in Caucasians. Six of 60 (10%) of ESTs in GenBank on 8/15/02 also contained this SNP; this was also noted in dbSNP as # 1051533.

Within the protein-coding sequence, a potentially important double nucleotide mutation (*ZFP36L1*2*) was found in a single allele at the junction of the first exon with the single intron. Reexamination of the original sequence traces from this sample confirmed high-quality reads spanning this mutation in both directions (Fig. 2A). This AG to GC mutation would result in

a potentially significant change of a lysine to a serine at position 19 in the protein (accession number NP_004917.2). More significantly, this change would result in the elimination of the 5'-donor portion of the splice site for the single intron in this gene. This mutation was found in a single allele, from an adult female Aka Pygmy subject from the Central African Republic (Coriell cell line AM10473A; subject X19). It was not found in 65 ESTs from GenBank that included this region on 11/15/02, although one EST (BF873589) contained unspliced 5'-intronic sequence. This mutation was not encountered in DNA from seven other Pygmies from the Central African Republic and Zaire that were in the original 72 samples.

This mutated splice site occurs at position 19 in the protein sequence, well before the tandem zinc finger domain that is the hallmark of this family of proteins. Theoretical translation of the mutated allele (Fig. 2B) indicates that after the lysine to serine conversion another 37 amino acids would be translated from intron sequence before a stop codon is reached. We suspect that the resulting hypothetical peptide, which would be 56 amino acids in length with an M_r of 6014 and a pI of 8.07, would have little biological significance, but this remains to be determined.

Mutation of both nucleotides in one allele would introduce two novel restriction sites into the genomic DNA, for the enzymes Fnu4HI and TseI, at position 704 of X79066. To confirm that both mutant bases were in the same allele and to devise an alternative assay for genotyping this dinucleotide mutation, we amplified a short region of DNA surrounding the mutation with appropriate PCR primers, and then cut the resulting 145 bp product with TseI. As shown in Fig. 2C, one allele from this DNA sample (X19) was cut into two fragments of approximately 50 and 95 bp, whereas the other allele was unaffected. Simultaneous digestion of a DNA sample from a subject of the same ethnic group did not reveal the presence of the mutation (Fig. 2C; X18). Thus, the presence of both mutations in the same allele was confirmed.

To determine whether this mutation would, as predicted, result in a failure to splice the mutant pre-mRNA, we subjected total lymphoblast RNA from subject X19 to northern-blotting analysis with a rat ZFP36L1 cDNA probe. As shown in Fig. 2D, lymphoblast RNA from a control subject from the same ethnic group (X15) produced the expected 3.9-kb band of the mature human ZFP36L1 mRNA. The amount of ZFP36L1 mRNA from the affected subject (X19) was decreased to 56% of that from the control cells, after PhosphorImager analysis of the northern blot and correction for GAPDH mRNA levels (Fig. 2D).

Because the two genomic sequences for *ZFP36L1* in GenBank did not meet in the middle of the intron, we searched GenBank for complete human genomic sequences in this region. This gene has been mapped previously to chromosome 14q22-24 (31). There was essentially a perfect match with

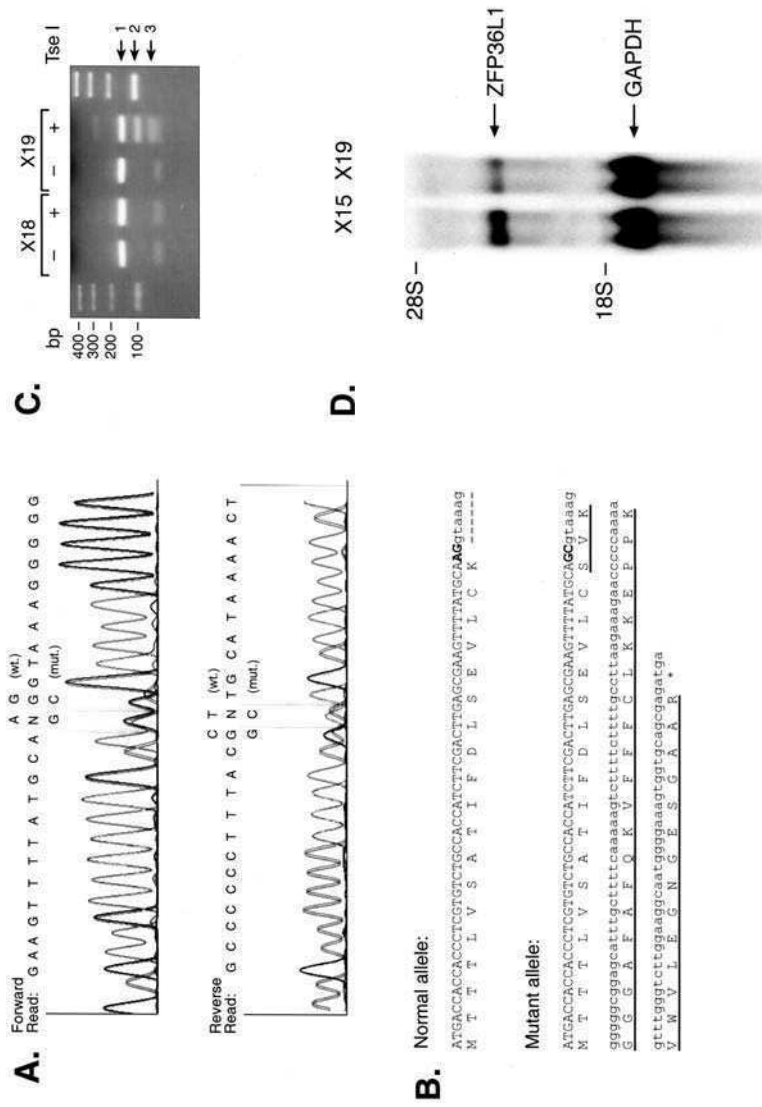


FIG. 2. Analysis of a *ZFP36L1* nonsplicing allele in subject X19. In (A) are shown the sequence traces from the sequencing of genomic DNA from subject X19; the wild-type and mutant base changes are shown in both strands. In (B) is shown the conceptual translation of the normal and X19 mutant *ZFP36L1* alleles at the amino terminal end of the *ZFP36L1* protein, along with the relevant DNA sequence extending into the first part of the intron. The normal DNA sequence begins with the ATG at position 651 of X79066. The corresponding DNA and amino acid sequences for the mutant allele from subject X19 are also shown; the DNA sequence was translated until a stop codon was encountered within the intron sequence. The normal and mutant dinucleotides are indicated in bold type; the intron sequence is indicated by lower case letters, and by a series of dashes at the end of the amino acid sequence; the hypothetical translated mutant “protein” sequence is indicated by the underlined amino acids; and the stop codon in the intron that terminates this hypothetical open-reading frame is indicated by an asterisk. In (C), DNA (approximately 1 μ g) comprising a 145-bp PCR product from subject X19 was digested using the restriction endonuclease TseI, and was compared with the same amount of DNA from a control subject (X18) from the same ethnic group, as indicated. The arrow labeled 1 indicates the undigested DNA from both the mutant and nonmutant alleles; the arrows labeled 2 and 3 indicate the new restriction fragments of 95 and 50 bp, respectively, derived from the TseI digestion of the mutant but not the wild-type allele. In (D) is shown a northern analysis of spliced and unspliced *ZFP36L1* mRNA from subject X19. Total cellular RNA (15 μ g) from lymphoblasts derived from subject X19 and a control subject from the same ethnic group (X15) was subjected to northern analysis using a *ZFP36L1* cDNA probe. The arrow labeled *ZFP36L1* indicates the correctly spliced *ZFP36L1* mRNA with a size of approximately 3.9 kb. The arrow labeled “GAPDH” indicates the GAPDH mRNA, which served as an internal control. After correction for the GAPDH mRNA values, the amount of correctly spliced *ZFP36L1* mRNA from subject X19 was 56% of the amount from control subject X15, as determined by PhosphorImager analysis. See the text for further details.

a BAC clone from chromosome 14 (GenBank accession number AL133313.2). Position 1 in X79066 is the equivalent to position 78926 in the BAC sequence; position 3922 in the 3'-genomic clone X79067 is the equivalent of 85172 in the BAC sequence. This indicates that the single intron in *ZFP36L1* is 2385 bp in size, as indicated in Fig. 1. It also indicates that a putative unspliced pre-mRNA should be about 6–6.5 kb. However, this putative unspliced pre-mRNA was not detected on the northern blot of the affected subject's RNA (X19), even after a much longer exposure of the autoradiograph depicted in Fig. 2D.

The RFLP assay described earlier was used to analyze DNA samples from 58 other Aka individuals. One other mutant allele was identified from a 54-year-old male Aka subject, also from the Central African Republic. The relationship between the two anonymous subjects with this mutation is not known. This dinucleotide variant was not found in the genotyping of DNA samples from 430 anonymous residents of Durham, NC.

Although we focused on the protein-coding regions of these genes, the sequencing strategy for *ZFP36L1* yielded several polymorphisms in the 5' region of the single intron, and this small region was quite polymorphic. Intron SNPs were identified at positions 729, 772, 804, and 845 (polymorphisms *ZFP36L1**3–6, respectively). *ZFP36L1**3 was present in only a single allele (Table II). *ZFP36L1**4 was found in 27 alleles, roughly distributed among the three racial groups. *ZFP36L1**5 was found in nine alleles, in every case in Caucasian subjects. *ZFP36L1**6 was found in six alleles, in every case in Asian subjects; of these, two were from the Taiwan Ami group, two were Cambodian, and one each was Beijing Chinese and Japanese.

Two other SNPs were identified in the protein-coding region of the second exon. In the first (*ZFP36L1**7), no change in the glycine residue at position 218 in the protein would occur; this was found in a single Asian allele, and in 0/40 ESTs spanning that region on 8/15/02. *ZFP36L1**8 was also found in a single Asian allele, but not in 46 ESTs spanning that region on 11/15/02. In this case, the resulting predicted amino acid change (a proline at position 295 changed to a histidine) is nonconservative and possibly significant. This portion of the *ZFP36L1* protein sequence was compared with those of the orthologous sequences from mouse (accession number P23950), rat (P17431), and *Xenopus* (AF061081). In all cases, the original proline in the human protein was present at the same position in the orthologs; it was in the middle of a highly conserved region of the proteins, with a core sequence of PSP*QD in human, mouse, and rat, and PSP*RD in *Xenopus*, with the asterisk indicating the proline changed to histidine in *ZFP36L1**8. This amino acid change removes a potential proline-directed protein kinase phosphorylation site on the serine immediately in front of the changed proline; in the case of TTP, members of this family of protein kinases are thought to phosphorylate

the protein in intact cells (30,32). Whether this potential site is phosphorylated by one or more of these kinases *in vivo*, and whether this polymorphism causes any other changes in the property of the protein, are subjects for future study.

No other polymorphisms for *ZFP36L1* were identified by the EST analysis of the protein-coding regions. However, it was apparent that there were at least two highly polymorphic regions within the 3'UTR of the mRNA: the region immediately 5' of b 2750 of NM_004926.2 contained at least five common variants of a series of A residues that was highly variable in length, and the region between 1730 and 1755 of the NM_004926.2 contained at least four common variants. These regions will require future resequencing to determine allele frequency and haplotypes of these common variants.

C. *ZFP36L2* Polymorphisms

Based on the consensus sequence for *ZFP36L2* generated by the current resequencing effort, the GenBank RefSeq NM_006887.3 now corresponds to what we believe to be the correct cDNA sequence. The corresponding protein sequence is NP_008818.3.

Polymorphisms in the *ZFP36L2* protein-coding region that were identified by the sequencing effort are indicated schematically in Fig. 1 and in more detail in Table III. A single polymorphism (*ZFP36L2*1*) was identified in the 5'UTR that comprised a 15 b deletion. This was found in a single allele, from an African-American subject, and in 0/11 EST from GenBank on 8/15/02. A single intron SNP was also identified (*ZFP36L2*2*) in one allele from an Asian subject. Seven SNPs were identified in the protein-coding domain of the second exon. The first (*ZFP36L2*3*) was found in a single allele and did not change glycine 93 of accession number AAA91778.1; it was not found in 20 relevant ESTs. A second (*ZFP36L2*4*) was found in 2.8% of alleles and in 1/27 ESTs (3.7%), and resulted in a semiconservative glycine to serine change at position 109. Examination of *ZFP36L2* sequences from other species in GenBank revealed that the glycine at position 93 was conserved in mouse (accession number C39590), and in the conceptual translations of ESTs for cow (AW465810), rat (BF522474), and pig (AW465810); in contrast, the *Xenopus* ortholog XC3H-3 contained a serine at that position (AAD24209). A third coding region SNP (*ZFP36L2*5*) occurred in two alleles and 0/28 ESTs, and resulted in a change from leucine to glutamine at position 112; the wild-type leucine residue was perfectly conserved in the five animal species listed earlier, using data from the same DNA sequences. The fourth (*ZFP36L2*6*) occurred in 6.25% of alleles and in 20/78 (26%) of ESTs; this resulted in no change to isoleucine 203. This has been identified by GenBank as dbSNP # 8098. The fifth (*ZFP36L2*7*) occurred in 48% of alleles and in 52/97 ESTs (54%), and resulted in no change to leucine 254. This SNP was found in all

three racial groups; it has been identified as dbSNP # 7933. The sixth (*ZFP36L2*8*) occurred in 2.1% of alleles and 0/70 ESTs, and resulted in no change in proline 315. The seventh (*ZFP36L2*9*) occurred in 1.4% of alleles and in 0/204 ESTs, and resulted in a conservative alanine to serine change at position 405. This SNP was in a region of the protein sequence that is relatively poorly conserved among animal species, so it was not possible to do a phylogenetic comparison at this site. Finally, a 3'UTR SNP (*ZFP36L2*10*) was identified in one allele but not in 106 ESTs; this was not in a canonical instability motif.

Five additional putative polymorphisms in the protein-coding region of *ZFP36L2* were uncovered by the EST analysis. The first, *ZFP36L2*11*, was a C to A substitution at position 3055 in the genomic sequence; this did not change proline 315, and was found in 3/69 (4.3%) of ESTs spanning this site. A C to T SNP at the same site (*ZFP36L2*8*) had previously been identified in the resequencing effort, also with no amino acid change.

A trinucleotide repeat polymorphism was identified in 6/>200 ESTs on 7/4/01; this will be called *ZFP36L2*12*. This change increased a series of seven CAG repeats to nine, resulting in two additional glutamine residues at position 393 in the *ZFP36L2* protein (NP_008818), for a total of nine consecutive glutamines. One of the ESTs with nine CAG repeats was resequenced in both directions, and the CAGCAG insert was confirmed. The five polymorphic ESTs came from two libraries. Three were from a mixed male germ cell tumor library (accession numbers AW589855.1, AI633753.1, and AI202919.1), and two were from a tonsillar germinal center B cell library (AA489889.1 and AA761736.1). A second complete cDNA and conceptually translated protein sequence for *ZFP36L2* were deposited in GenBank (AAH05010), in which there were 10 CAG repeats at this site, corresponding to 10 glutamine residues. This clone was derived from a pancreatic adenocarcinoma. These expanded CAG repeat polymorphisms were not present in the 144 alleles that form the basis of this paper, all of which contained the wild-type seven consecutive CAGs. In addition, resequencing of this portion of *ZFP36L2* in DNA from 20 germ cell testicular tumors did not reveal any differences from wild type in this potential microsatellite.

Three other potential polymorphisms that would result in amino acid changes were identified by the EST analysis. *ZFP36L2*13* was a TC to CT change that would result in a change of leucine 450 to serine. Although this could be a simple sequencing error, it was found in four of the more than 200 ESTs spanning this region, two of which were from the same prostate library. *ZFP36L2*14* was a G to A substitution that would lead to a change from asparagine 476 to aspartate, and was seen in three of the more than 150 relevant ESTs, two of which were from the same germ cell tumor library. *ZFP36L2*15* was a C to G substitution that would change a serine at

position 492 to a cysteine, and occurred in six of more than 150 ESTs spanning this region.

Recently, the genomic *ZFP36L2* sequence was identified on a BAC contig from chromosome 2, accession number AC010883.5. Base 1 in U07802, the *ZFP36L2* genomic clone, aligned with base 152080 in AC010883.5; base 5503 of U07802 aligned with base 146557 in AC010883.5. Although the sequence identity between these two clones was 97%, with 79 gaps, it seems likely that this position in chromosome 2 is the locus for *ZFP36L2*. The BAC sequence AC010883.5 contained the normal seven CAG repeats at the polyglutamine polymorphic region (*ZFP36L2*12*).

IV. Discussion

These studies provide initial information on the occurrence and type of polymorphisms among the three known human genes encoding proteins of the TTP CCCH tandem zinc finger type. These analyses focused on the protein-coding regions of all three genes. Out of 13 polymorphisms identified in these protein-coding regions by the resequencing effort, seven (54%) resulted in no change in amino acid, and 6/13 (46%) resulted in amino acid changes ranging from conservative to nonconservative. Notably, there were no amino acid changes within the 64-amino acid tandem zinc finger domains of these proteins, perhaps reflecting the critical importance of many residues in this domain to mRNA binding (1). An additional triplet repeat polymorphism was identified in *ZFP36L2* by EST searches (*ZFP36L2*11*); this resulted in an increase from seven to nine consecutive glutamine residues. This polymorphism was found in ESTs from two different libraries, from mixed male germ cell tumors, and from nonmalignant tonsillar B cell, and an even longer repeat of 10 glutamines was found in a clone derived from a pancreatic adenocarcinoma (AAH05010). Although this type of triplet repeat polymorphism was not found in the 144 alleles resequenced for this paper, it seems possible that it represents an example of microsatellite instability in tumor tissue. Matched samples of tumor and nontumor DNA from the same individuals are currently being evaluated to determine if this is the case.

For most of these polymorphisms, further study will be required to determine if they cause any functional change in the resulting mRNAs and proteins. An exception is in the case of the dinucleotide splice site mutation found in one *ZFP36L1* allele from the Aka Pygmy subject X19, and subsequently in another individual from the same ethnic group, but not in 430 anonymous subjects from various ethnic groups in Durham, NC, or in ESTs currently in GenBank. This mutation resulted in a decrease in steady-state *ZFP36L1* mRNA levels to about 50% of its normal level of expression in cells

derived from subject X19, suggesting normal expression from the unaffected allele; the absence of a detectable unspliced form of the mRNA suggests that the mutant allele is not expressed. The function of the normal *ZFP36L1* protein in physiology is unknown. However, when it is expressed in human embryonic kidney 293 cells, the *ZFP36L1* protein can, like TTP and *ZFP36L2*, promote the instability of ARE-containing mRNAs following direct binding to the class II ARE (1). In addition, its expression is increased in a form of acute myelogenous leukemia, where its overexpression is associated with enhanced myeloid cell proliferation in response to granulocyte colony-stimulating factor (33). Recent experiments with the mouse gene knockout indicate that total deficiency of *Zfp36L1* is lethal in early gestation, suggesting that even the hemizygous state might be associated with a disease or trait (D. J. Stumpo, R. S. Phillips, Y. Mishina and P. J. Blackshear, unpublished data). We are currently attempting to identify human kindreds with this splicing mutation, with the goal of examining them for coinheritance of a disease or trait with the mutant allele.

Another *ZFP36L1* polymorphism with potential functional consequences is *ZFP36L1**8, which occurred in a single allele from an Indo-Pakistani subject (sample W08, Coriell cell line GM11213) and in none of the 46 ESTs from GenBank on 11/15/02 that spanned this region. This SNP would result in the removal of a potential phosphorylation site for proline-directed protein kinases, such as the mitogen-activated protein (MAP) kinases. The normal sequence in this region is *sppsPqds*ls, with the proline indicated by the upper case P mutated to a histidine in the polymorphic sequence. It is not known at present if this is a physiologically relevant phosphorylation site; in the case of TTP, MAP kinase phosphorylation of the protein is known to occur in intact cells (30), and recent evidence from our laboratory suggests that phosphorylation of TTP by the MAP kinase p38 has functional effects on mRNA binding and destabilization (9). The serine-proline duo at positions 294, 295 in the human sequence is conserved in all orthologues of *ZFP36L1* examined, including mouse (accession number B39590), rat (P17431), *X. laevis* (AAD24208.1), and a presumed ortholog encoded by a chicken EST (AJ399312.1, reading frame +1). Ongoing experiments in our laboratory will test the hypothesis that this site is a phosphorylation site for one of these kinases under normal circumstances, and that this mutation affects not only phosphorylation of the protein but also one or more functional properties.

Another potential site for interesting polymorphisms is the ARE within the 3'UTR of the TTP mRNA, which is responsible in part for the lability of this mRNA. Direct resequencing of this region yielded only two potential polymorphisms, one TT deletion at position 2980 in the genomic sequence, and one putative T to A change at position 3072. The former dinucleotide deletion polymorphism (*ZFP36**10) appears to be a *bona fide*, common human

variant: it appeared in 7.6% of the alleles directly sequenced in this study, was represented in samples from all three major racial groups, and was found in 64/249 (26%) of ESTs in GenBank on 11/15/02. Although it is not within a canonical AUUUA instability motif, it is possible that it contributes in some way to TTP mRNA stability. The latter potential polymorphism identified in this region of the TTP mRNA is problematical. Specifically, it occurred in a region that could not be reliably sequenced using the present methods, and was not found in 144 ESTs in GenBank that spanned this region. At this point, its existence must be considered doubtful. A third polymorphism within the 3'UTR that was identified only by EST analysis is likely to be real; it (ZFP36*15) was found in 17/232 (7.3%) of EST examined on 11/15/02, and was within an AU-rich region of the mRNA that is likely to be involved in mRNA stability.

Although these polymorphisms in the TTP mRNA currently are of unclear significance to the function or steady-state levels of the mRNA or protein, they may be of interest in linkage studies. TTP has been localized to chromosome 19q13.1 (3), and its genomic sequence is contained within a BAC contig localized to chromosome 19 (accession number AC011500.6), where genomic sequence 1–3880 from accession number M92844 is equivalent to bp 45,571–41,684 of accession number AC011500.6. Based on the results to date in with the TTP knockout mice, complete TTP deficiency in man would be expected to result in a severe autoimmune and inflammatory condition with myeloid hyperplasia that would probably not be compatible with life when untreated; the hemizygous state might result in a less severe syndrome of later onset. However, as we demonstrated previously, the phenotype of the TTP knockout mice can be virtually normalized using an anti-TNF α monoclonal antibody (12) or interbreeding with mice deficient in both TNF α receptors (14). Similar treatments, using a “humanized” anti-TNF α antibody or a chimeric soluble TNF α receptor, have proven to be effective treatments for human rheumatoid arthritis and Crohn’s disease, and might well be used to delay or prevent the consequences of complete or partial TTP deficiency in man. These considerations suggest that linkage studies with TTP polymorphisms might be profitable in human autoimmune, autoinflammatory, or myeloproliferative disorders, but our limited knowledge of normal TTP physiology precludes identification of more likely patient groups.

As an initial exploratory step, we resequenced the protein coding and regulatory regions of the gene encoding TTP, *ZFP36*, from two groups of subjects: one group of otherwise normal subjects that exhibited bronchial hypo- or hypersensitivity in response to inhaled endotoxin (20), and a second group of subjects with a chronic inflammatory syndrome resembling that caused by activating mutations in one of the TNF α receptors (21). Only a single new intronic SNP was identified by this expanded resequencing effort;

to date, none of the other polymorphisms in *ZFP36* exhibited major differences in frequencies when the TRAPS-like subjects were compared with the other subjects, or data from the bronchial hypersensitivity group were compared with the hyposensitivity group.

These studies have begun to describe the normal sequence variability within members of this small human gene family. Future studies will focus on identifying which of the polymorphisms described here are associated with changes in the biosynthesis, turnover, or function of the encoded CCCH proteins. In addition, we hope to link each of these polymorphisms to specific haplotypes, identify kindreds with each haplotype, and then determine the possible association of these haplotypes with heritable traits.

ACKNOWLEDGMENTS

Supported in part by Interagency Agreement Yi-ES-8054-05, and by a Cooperative Research and Development Agreement with AstraZeneca Ltd. We are grateful to Dr. James Selkirk and Tina Xi for advice, to Mary Watson for the cultured lymphoblasts, and to Drs. Jack Taylor and Stephanie London for helpful comments on the manuscript.

REFERENCES

1. W. S. Lai, E. Carballo, J. M. Thorn, E. A. Kennington and P. J. Blackshear, Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA, *J. Biol. Chem.* **275**, 17827–17837 (2000).
2. J. De, W. S. Lai, J. M. Thorn, S. M. Goldsworthy, X. Liu, T. K. Blackwell and P. J. Blackshear, Identification of four CCCH zinc finger proteins in *Xenopus*, including a novel vertebrate protein with four zinc fingers and severely restricted expression, *Gene* **228**, 3454 (1999).
3. G. A. Taylor, W. S. Lai, R. J. Oakley, M. F. Seldin, T. B. Shows, R. L. Eddy, Jr. and P. J. Blackshear, The human TTP protein: sequence, alignment with related proteins and chromosomal localization of the mouse and human genes, *Nucleic Acids Res.* **19**, 3454 (1991).
4. R. N. DuBois, M. W. McLane, K. Ryder, L. F. Lau and D. Nathans, A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence, *J. Biol. Chem.* **265**, 19185–19191 (1990).
5. W. S. Lai, D. J. Stumpo and P. J. Blackshear, Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein, *J. Biol. Chem.* **265**, 16556–16563 (1990).
6. Q. Ma and H. R. Herschman, A corrected sequence for the predicted protein from the mitogen-inducible TIS11 primary response gene, *Oncogene* **6**, 1277–1278 (1991).
7. B. C. Varnum, R. W. Lim, V. P. Sukhatme and H. R. Herschman, Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 cells by the tumor promoter tetradecanoyl phorbol acetate, *Oncogene* **4**, 119–120 (1989).
8. E. Carballo, W. S. Lai and P. J. Blackshear, Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin, *Science* **281**, 1001–1005 (1998).

9. E. Carballo, W. S. Lai and P. J. Blackshear, Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability, *Blood* **95**, 1891–1899 (2000).
10. W. S. Lai, E. Carballo, J. R. Strum, E. A. Kennington, R. S. Phillips and P. J. Blackshear, Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA, *Mol. Cell. Biol.* **19**, 4311–4323 (1999).
11. W. S. Lai and P. J. Blackshear, Interactions of CCCH zinc finger proteins with mRNA: tristetraprolin-mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail, *J. Biol. Chem.* **276**, 23144–23154 (2001).
12. G. A. Taylor, E. Carballo, D. M. Lee, W. S. Lai, M. S. Thompson, D. D. Patel, D. I. Schenkman, G. S. Gilkeson, H. E. Broxmeyer, B. F. Haynes and P. J. Blackshear, A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency, *Immunity* **4**, 445–454 (1996).
13. E. Carballo, G. S. Gilkeson and P. J. Blackshear, Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (–/–) mice. Evidence that monocyte/macrophage progenitors may be responsible for TNFalpha overproduction, *J. Clin. Invest.* **100**, 986–995 (1997).
14. E. Carballo and P. J. Blackshear, Roles of tumor necrosis factor alpha receptor subtypes in the pathogenesis of the tristetraprolin-deficiency syndrome, *Blood* **98**, 2389–2395 (2001).
15. Z. Q. Ning, J. D. Norton, J. Li and J. J. Murphy, Distinct mechanisms for rescue from apoptosis in Ramos human B cells by signaling through CD40 and interleukin-4 receptor: role for inhibition of an early response gene, Berg36, *Eur. J. Immunol.* **26**, 2356–2363 (1996).
16. X. F. Nie, K. N. Maclean, V. Kumar, I. A. McKay and S. A. Bustin, ERF-2, the human homologue of the murine Tis11d early response gene, *Gene* **152**, 285–286 (1995).
17. E. Fritsche, G. S. Pittman and D. A. Bell, Localization, sequence analysis, and ethnic distribution of a 96-bp insertion in the promoter of the human CYP2E1 gene [In Process Citation], *Mutat. Res.* **432**, 1–5 (2000).
18. F. P. Guengerich, The environmental genome project: functional analysis of polymorphisms, *EHP* **106**, 365–368 (1998).
19. R. R. Sharp and J. C. Barrett, The environmental genome project and bioethics, *Kennedy Inst. Ethics J.* **9**, 175–188 (1999).
20. J. G. Moreland, R. M. Fuhrman, C. L. Wohlford-Lenane, T. J. Quinn, E. Benda, J. A. Pruessner and D. A. Schwartz, TNF-alpha and IL-1 beta are not essential to the inflammatory response in LPS-induced airway disease, *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280**, L173–L180 (2001).
21. I. Aksentjevich, J. Galon, M. Soares, E. Mansfield, K. Hull, H. H. Oh, R. Goldbach-Mansky, J. Dean, B. Athreya, A. J. Reginato, M. Henrickson, B. Pons-Estel, J. J. O'shea and D. L. Kastner, The tumor-necrosis-factor receptor-associated periodic syndrome: new mutations in TNFRSF1A, ancestral origins, genotype-phenotype studies, and evidence for further genetic heterogeneity of periodic fevers, *Am. J. Hum. Genet.* **69**, 301–314 (2001).
22. B. L. King, H. Q. Peng, P. Goss, S. Huan, D. Bronson, B. M. Kacinski and D. Hogg, Repeat expansion detection analysis of (CAG)_n tracts in tumor cell lines, testicular tumors, and testicular cancer families, *Cancer Res.* **57**, 209–214 (1997).
23. D. A. Nickerson, S. L. Taylor, K. M. Weiss, A. G. Clark, R. G. Hutchinson, J. Stengard, V. Salomaa, E. Vartiainen, E. Boerwinkle and C. F. Sing, DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene [see comments], *Nat. Genet.* **19**, 233–240 (1998).
24. M. J. Rieder, S. L. Taylor, V. O. Tobe and D. A. Nickerson, Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome, *Nucleic Acids Res.* **26**, 967–973 (1998).

25. M. R. Shen, I. M. Jones and H. Mohrenweiser, Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans, *Cancer Res.* **58**, 604–608 (1998).
26. D. A. Nickerson, V. O. Tobe and S. L. Taylor, PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing, *Nucleic Acids Res.* **25**, 2745–2751 (1997).
27. M. Kokoris, K. Dix, K. Moynihan, J. Mathis, B. Erwin, P. Grass, B. Hines and A. Duesterhoeft, High-throughput SNP genotyping with the Masscode system, *Mol. Diagn.* **5**, 329–340 (2000).
28. W. S. Lai, M. J. Thompson, G. A. Taylor, Y. Liu and P. J. Blackshear, Promoter analysis of Zfp-36, the mitogen-inducible gene encoding the zinc finger protein tristetraprolin, *J. Biol. Chem.* **270**, 25266–25272 (1995).
29. W. S. Lai, M. J. Thompson and P. J. Blackshear, Characteristics of the intron involvement in the mitogen-induced expression of Zfp-36, *J. Biol. Chem.* **273**, 506–517 (1998).
30. G. A. Taylor, M. J. Thompson, W. S. Lai and P. J. Blackshear, Phosphorylation of tristetraprolin (TTP), a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase in vitro. *J. Biol. Chem.* **270**, 13341–13347 (1995).
31. K. N. Mclean, C. G. See, I. A. McKay and S. A. Bustin, The human immediate early gene BRF1 maps to chromosome 14q22-q24. *Genomics* **30**, 89–90 (1995).
32. E. Carballo, H. Cao, W. S. Lai, E. A. Kennington, D. Campbell and P. J. Blackshear, Decreased sensitivity of tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of tristetraprolin in the p38 signaling pathway, *J. Biol. Chem.* **276**, 42580–42587 (2001).
33. H. Shimada, H. Ichikawa, S. Nakamura, R. Katsu, M. Iwasa, I. Kitabayashi and M. Ohki, Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF, *Blood* **96**, 655–663 (2000).

Molecular and Cell Biology of Phosphatidylserine and Phosphatidylethanolamine Metabolism

JEAN E. VANCE

*Canadian Institutes for Health Research
Group on Molecular and Cell Biology
of Lipids and Department of Medicine,
University of Alberta, 332 HMRC,
Edmonton, AB, Canada T6G 2S2*

I. Membrane Phospholipids	70
II. Pathways for Biosynthesis of Phosphatidylserine and Phosphatidylethanolamine	71
III. Enzymes of Phosphatidylserine Biosynthesis	74
IV. Enzymes of Phosphatidylethanolamine Biosynthesis	82
V. Regulation of PS and PE Biosynthesis	85
VI. Interorganelle Transport of PS and PE	87
VII. Intramembrane Transport of PS and PE	92
VIII. Biological Functions of PS and PE	95
IX. Future Directions	97
References	98

In this review, the pathways for phosphatidylserine (PS) and phosphatidylethanolamine (PE) biosynthesis, as well as the genes and proteins involved in these pathways, are described in mammalian cells, yeast, and prokaryotes. In mammalian cells, PS is synthesized by a base-exchange reaction in which phosphatidylcholine or PE is substrate for PS synthase-1 or PS synthase-2, respectively. Isolation of Chinese hamster ovary cell mutants led to the cloning of cDNAs and genes encoding these two PS synthases. In yeast and prokaryotes PS is produced by a biosynthetic pathway completely different from that in mammals: from a reaction between CDP-diacylglycerol and serine. The major route for PE synthesis in cultured cells is from the mitochondrial decarboxylation of PS. Alternatively, PE can be synthesized in the endoplasmic reticulum (ER) from the CDP-ethanolamine pathway. Genes and/or cDNAs encoding all the enzymes in these two pathways for PE synthesis have been isolated and characterized.

In mammalian cells, PS is synthesized on the ER and/or mitochondria-associated membranes (MAM). PS synthase-1 and -2 are highly enriched in MAM compared to the bulk of ER. Since MAM are a region of the ER that appears to be in close juxtaposition to the mitochondrial outer membrane, it has been proposed that MAM act as a conduit for the transfer of newly synthesized PS into mitochondria. A similar pathway appears to operate in yeast. The use of yeast

mutants has led to identification of genes involved in the interorganelle transport of PS and PE in yeast, but so far none of the corresponding genes in mammalian cells has been identified.

PS and PE do not act solely as structural components of membranes. Several specific functions have been ascribed to these two aminophospholipids. For example, cell-surface exposure of PS during apoptosis is thought to be the signal by which apoptotic cells are recognized and phagocytosed. Translocation of PS from the inner to outer leaflet of the plasma membrane of platelets initiates the blood-clotting cascade, and PS is an important activator of several enzymes, including protein kinase C. Recently, exposure of PE on the cell surface was identified as a regulator of cytokinesis. In addition, in *Escherichia coli*, PE appears to be involved in the correct folding of membrane proteins; and in *Drosophila*, PE regulates lipid homeostasis via the sterol response element-binding protein. © 2003 Elsevier Science

I. Membrane Phospholipids

A fundamental problem in biology is to understand the mechanisms by which biological membranes are assembled. The major lipids of all cell membranes are phospholipids, which are interspersed with a complex mixture of other lipids such as sterols, sphingolipids, fatty acids, and their derivatives. The large variety of membrane lipids defines the fluidity of the membrane, which in turn is thought to control the activity of membrane proteins. In addition, many membrane lipids or their metabolites, such as eicosanoids, diacylglycerols, phosphatidylinositol bisphosphate, sphingosine, and ceramide, are “second messengers” that are involved in cellular-signaling events.

In eukaryotic cell membranes the most abundant phospholipid is usually phosphatidylcholine (PC¹), although in *Drosophila*, phosphatidylethanolamine (PE) comprises approximately 55% of total membrane phospholipids. In general, prokaryotes do not contain PC, but instead, PE is the most abundant phospholipid. The relatively simple organism, *Escherichia coli*, contains at least 100 different phospholipids, whereas eukaryotic membranes contain more than 1000 distinct phospholipid species. Some of the lipid diversity in membranes is a consequence of the large variety of acyl-chain substituents attached to the glycerol backbone of the various phospholipid classes. In Gram-negative bacteria, the lipid compositions of the inner and outer membranes are quite distinct with the outer membrane containing the unusual complex lipid, Lipid A (1). The different cell types in multicellular organisms also have unique phospholipid compositions. Moreover, the lipid

¹Abbreviations used: CHO, Chinese hamster ovary; ER, endoplasmic reticulum; MAM, mitochondria-associated membranes; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PS, phosphatidylserine; SREBP, sterol response element binding protein.

compositions of individual organelles of eukaryotic cells (e.g., endoplasmic reticulum (ER), mitochondria, lysosomes, peroxisomes, nucleus, Golgi apparatus, and plasma membrane) are distinct.

Not only does the lipid composition differ among different cells and organelle membranes, but the lipids are also not equally distributed across the two leaflets of the membrane bilayer. Particularly striking is the asymmetrical transbilayer distribution of phospholipids in the erythrocyte membrane in which the outer leaflet of the bilayer contains primarily the choline-containing phospholipids, PC, and sphingomyelin, whereas the inner leaflet is highly enriched in PE and phosphatidylserine (PS). Phospholipids are also similarly asymmetrically distributed across the plasma membrane of other eukaryotic cells. Presumably, these unique lipid compositions are required for the specific functions of the membrane. We are only now beginning to understand, at the molecular level, the reasons for this complexity and how these different membrane lipid compositions are established.

PS and PE are important components of prokaryotic membranes, and also constitute 3–10% and 20–40%, respectively, of the phospholipids of mammalian cell membranes. These two aminophospholipids not only serve structural roles in membranes but also are required for other, more specific, functions. For example, PS is required for activation of enzymes such as protein kinase C (2,3), cRaf1 protein kinase (4), Na⁺/K⁺ ATPase (4), dynamin-1 (5), and neutral sphingomyelinase (6). PS has also recently been found to be enriched in lipid rafts (7). In addition, cell-surface exposure of PS initiates the blood-clotting cascade (8,9), and is also thought to be the signal by which apoptotic cells are recognized and phagocytosed by macrophages (10–13). An important, specific role for PE is as a precursor of the glycosylphosphatidylinositol anchors that are covalently attached to many cell-surface proteins (14). In addition, PE has been proposed to regulate the folding and function of membrane proteins in *E. coli* (15), as well as being a regulator of lipid homeostasis via the sterol response element binding protein in *Drosophila* (16,17).

This review will focus on the metabolism of PS and PE in mammalian cells, and will include a comparison with PS and PE metabolism in yeast and prokaryotes.

II. Pathways for Biosynthesis of Phosphatidylserine and Phosphatidylethanolamine

In 1884 Thudicum isolated kephalin and described it as a lipid that contained nitrogen and phosphorus but was different from lecithin (PC).

In 1930 PE was purified from kephalin, thereby setting the stage for elucidation of the structure of PE in 1952 by Baer *et al.* (18). PS was another component of kephalin but not until 1941 did Folch discover that PS was a distinct entity (19). The structure of PS was confirmed by chemical synthesis by Baer and Maurukas in 1955 (20).

In mammalian cells, PS is synthesized by an exchange of L-serine for the ethanolamine or choline head-group of a PE or PC molecule, respectively (Fig. 1B) (21). In contrast, in prokaryotes and yeast, PS is synthesized by reaction of CDP-diacylglycerol with L-serine (Fig. 1A) (22–24). The latter biosynthetic pathway has not been detected in mammalian cells, but recent evidence has revealed that plants have the capacity to make PS by this route, as well as by the base-exchange reaction (25).

The biosynthesis of PE in eukaryotic cells occurs by two major routes—from the CDP-ethanolamine pathway, as originally described by Kennedy and Weiss in 1956 (26), and from the decarboxylation of PS (27) (Fig. 2). PE can also be synthesized from the acylation of lyso-PE, and from a base-exchange reaction in which ethanolamine is exchanged for the serine head-group of PS (Fig. 1B). The latter two pathways are thought to be the minor contributors to PE synthesis (28). The relative importance of the two major pathways for PE synthesis in mammalian cells has not been unambiguously established. However, in many types of cultured mammalian cells, such as BHK cells (29) and Chinese hamster ovary (CHO) cells (30,31), the PS decarboxylation

A.



B.

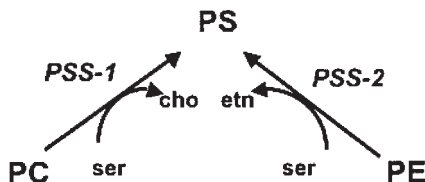


FIG. 1. Biosynthesis of phosphatidylserine. (A) In yeast and prokaryotes, PS is synthesized via a PS synthase that utilizes CDP-diacylglycerol and L-serine. (B) In mammalian cells, PS is made by a base-exchange reaction in which the choline (cho) or ethanolamine (etn) head-groups of PC or PE, respectively, are exchanged for L-serine (ser) in reactions catalyzed by phosphatidylserine synthase-1 (PSS1) or phosphatidylserine synthase-2 (PSS2).

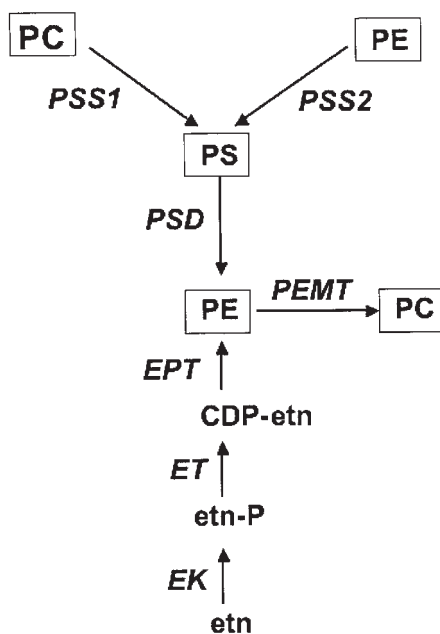


FIG. 2. Interrelationships between PS and PE biosynthetic pathways in mammalian cells. PS is made from PC or PE by base-exchange reactions that are catalyzed by PS synthase-1 (PSS1) or PS synthase-2 (PSS2). PS is decarboxylated to PE by PS decarboxylase (PSD). PE can also be synthesized from the CDP-ethanolamine pathway in which ethanolamine (etn) is phosphorylated to phosphoethanolamine by ethanolamine kinase (EK). Phosphoethanolamine is next converted into CDP-ethanolamine by CTP:phosphoethanolamine cytidylyltransferase (ET) and, subsequently, ethanolaminephosphotransferase (EPT) converts CDP-ethanolamine into PE. PE can be methylated to PC via PE *N*-methyltransferase (PEMT).

pathway is thought to be responsible for the synthesis of at least 80% of PE, even when these cells are supplied with exogenous ethanolamine, and therefore have the capacity to make PE from the alternative CDP-ethanolamine pathway. It is likely that the quantitative importance of the two pathways depends on the type of cell. The CDP-ethanolamine pathway has been reported to be the predominant pathway for PE synthesis in rat liver/hepatocytes (32,33) and hamster heart (34). However, other studies have indicated that the PS decarboxylation pathway is the major route for PE formation in rat liver (35). One caveat to all these studies is that there are difficulties in measurement of pool sizes of precursors. Moreover, radio-labeling of lipid precursors might not be homogeneous; therefore interpretation of the data is not straightforward.

Several types of cells, for example rat mammary carcinoma cells (36), hybridoma cells (37), and human keratinocytes (38), have an absolute

requirement for ethanolamine for growth and/or proliferation (39). However, the ethanolamine requirement of these cells is not necessarily related to its requirement for PE derived from the CDP-ethanolamine pathway. Whether one or the other of these PE biosynthetic pathways is specifically required for survival, or for specific physiological functions in a whole animal, will likely be clarified when mice are generated in which key genes in each of these pathways have been eliminated.

Yeast possess genes encoding enzymes involved in PE synthesis from both the PS decarboxylation and CDP-ethanolamine pathways. In contrast, in prokaryotes, PE is not produced from the CDP-ethanolamine pathway but is made entirely from PS decarboxylation (24).

III. Enzymes of Phosphatidylserine Biosynthesis

A. Mammalian Phosphatidylserine Synthase-1

In mammalian cells, PS is synthesized via a calcium-dependent base-exchange reaction (21) in which a phospholipid head-group is replaced by L-serine (Fig. 1B). The genes and proteins involved in PS biosynthesis have only relatively recently been identified, primarily because of difficulties in purification of these membrane-bound enzymes to homogeneity. In mammalian cells, both PE and PC are substrates for serine exchange catalyzed by PS synthase activity. An enzyme catalyzing serine exchange with PE, but not PC, was partially purified from rat brain (40) providing the first indication that mammalian cells might contain two distinct PS synthases with different substrate specificities. Subsequently, the isolation of mutant CHO-K1 cells exhibiting defects in PS synthesis confirmed the existence of (at least) two distinct mammalian PS synthases: PS synthase-1 uses PC for serine exchange (41,42), whereas PS synthase-2 uses PE (43,44) (Fig. 1B and Fig. 2).

Two similar CHO-K1 mutants, M9.1.1 (41) and PSA-3 (42), have been isolated in the laboratories of Voelker and Nishijima, respectively. In these mutant cells the synthesis of PS, and the production of PE from PS, were reduced by 35–55% compared to parental cells, and the masses of PS and PE were similarly decreased. Importantly, in both M9.1.1 and PSA-3 cell lines, the choline-exchange activity was eliminated providing evidence that the PC-dependent synthesis of PS, catalyzed by PS synthase-1, was the primary lesion in both of these mutant cell lines.

The M9.1.1 and PSA-3 mutants, however, appear not to be identical. For example, in PSA-3 cells the ethanolamine-exchange activity is 55% less than in parental CHO-K1 cells (42), whereas this activity is *increased* by 85% in

M9.1.1 cells compared to CHO-K1 cells (41). This observation suggests that some type of a compensatory regulation might be operating in the M9.1.1 cells. Another difference between the M9.1.1 and PSA-3 cells is that although supplementation of the growth medium with ethanolamine, PS, or lyso-PE supports normal growth of M9.1.1 cells, PSA-3 cells grow normally only when PS or PE, but not ethanolamine (31,45), are added to the culture medium. The differences between these two cell lines have not yet been completely explained.

A cDNA-encoding PS synthase-1 in CHO-K1 cells was cloned by complementation of the PS auxotrophy of the PSA-3 cells (46). Expression of the PS synthase-1 cDNA in PSA-3 cells restored the choline-exchange activity and normalized the phospholipid composition (46). That this cDNA-encoded PS synthase-1 was confirmed when PS synthase-1 activity in a CHO-K1 cell lysate was eliminated upon immunoprecipitation with an antibody raised against a peptide corresponding to the N-terminus of the putative PS synthase-1 (47).

Based on the cDNA sequence of PS synthase-1 from CHO-K1 cells, a cDNA-encoding murine PS synthase-1 was cloned and expressed (48). The protein encoded by this cDNA is predicted to have a molecular weight of 55,617 and to contain 473 amino acids. The PS synthase-1 sequences from murine liver, CHO-K1 cells, and human myeloblasts are approximately 80% identical, and the deduced amino acid sequences are >90% identical. However, the sequence of the murine liver PS synthase-1 bears no homology to the PS synthases of *E. coli*, *Saccharomyces cerevisiae*, or *Bacillus subtilis* (22,24,49,50). This lack of homology was not unexpected since, as shown in Fig. 1, the pathway used by yeast and bacteria for PS synthesis (in which CDP-diacylglycerol is a substrate) is completely different from the base-exchange reaction used by mammalian cells. The protein motifs required for enzymatic activity have not yet been identified in any of the PS synthases.

When the murine PS synthase-1 cDNA was heterologously expressed in M9.1.1 cells, the growth defect of the cells was reversed and a normal phospholipid composition was restored in the absence of exogenously supplied ethanolamine or phospholipids (48). Murine PS synthase-1 cDNA was expressed in McArdle 7777 rat hepatoma cells resulting in the serine-, choline-, and ethanolamine-exchange activities, as measured *in vitro* in cell extracts, being increased several fold, and the incorporation of [³H]serine into PS and PE in the cells being increased by three- and two-fold, respectively (48). Thus, overexpression of PS synthase-1 activity in rat hepatoma cells increased the rate of PS synthesis as well as the rate of PE synthesis from PS decarboxylation. However, the levels of PS and PE were not altered. Instead, the cells were apparently able to maintain constant levels of PS and PE when

confronted with a higher rate of synthesis of PS and PE. In addition to an increased decarboxylation of PS to PE, the rates of degradation of PS, and/or PS-derived PE, were also increased (48). Moreover, the synthesis of PE from the CDP-ethanolamine pathway was reciprocally inhibited (see Fig. 2) in cells overexpressing PS synthase-1 activity. These data indicate that PS and PE homeostasis is tightly controlled, and that the metabolism of these two lipids is coordinately regulated. Similar observations have demonstrated that PC homeostasis is also tightly controlled by coordination of the rates of synthesis and degradation of PC (51).

The murine PS synthase-1 gene has been cloned from a genomic DNA library contained in bacterial artificial chromosomes (52). The gene, which comprises 13 exons and 12 introns, resides on mouse chromosome 13 in the region B-C1; the human PS synthase-1 gene is located on chromosome 8. The organization of the murine and human PS synthase-1 genes is similar. For example, the exon-intron boundaries in the human PS synthase-1 gene are at the same positions as in the murine gene and, consequently, the sizes of the exons are the same in the two species (52).

B. Mammalian Phosphatidylserine Synthase-2

The studies described in the previous section for mutant CHO-K1 cells strongly indicated that a second enzyme (designated PS synthase-2) catalyzes ethanolamine and serine, but not choline, exchange. In an attempt to isolate the cDNA encoding the putative PS synthase-2, the PS synthase-1 cDNA sequence from CHO-K1 cells was compared with known EST sequences. A human EST was found that had approximately 30% sequence identity to PS synthase-1. Based on this information, a PS synthase-2 cDNA was isolated from CHO-K1 cells (43). This cDNA encoded a protein of 474 amino acid residues, corresponding to a calculated molecular weight of 55,003. Consistent with the predicted activity of PS synthase-2, expression of this cDNA in CHO-K1 cells resulted in a six-fold increase in serine-exchange activity, and a 10-fold increase in ethanolamine-exchange activity, without any increase in choline-exchange activity (43). Thus, the isolated PS synthase-2 cDNA encoded the enzyme that was responsible for the residual serine-exchange activity in PSA-3 and M9.1.1 cells. Moreover, these studies clearly established that PS synthase-1 and -2 are encoded by different genes. A cDNA encoding the murine PS synthase-2 was subsequently isolated (53), and alignment of the PS synthase-2 sequences from CHO-K1 cells and murine liver revealed that PS synthase-2 is highly conserved across species.

Mutagenesis of PSA-3 cells (which are deficient in PS synthase-1) resulted in the generation of a cell line, called PSB-2, in which the amount of PS synthase-2 mRNA was reduced by ~80% compared to that in PSA-3 cells and parental CHO-K1 cells, and in which PS synthase-1 was absent (44). When

PSB-2 cells were grown in medium that lacked supplementary phospholipids, the content of PS and PE was reduced by 80% and 50%, respectively, compared to that in CHO-K1 cells, and by ~70% compared to that in PSA-3 cells (44). The serine- and ethanolamine-exchange activities of PSB-2 cells were dramatically reduced (by ~90%) in comparison to wild-type CHO-K1 cells, and similarly reduced (by ~90%) compared to PSA-3 cells. Although the PSB-2 cells were able to grow in medium supplemented with PS, they were unable to grow in medium supplemented with PE, consistent with the idea that PS synthase-2 uses PE as a substrate for PS synthesis. Transfection of PSB-2 cells with PS synthase-2 cDNA restored a normal rate of growth in the presence of exogenous PE, whereas transfection with PS synthase-1 cDNA did not permit the cells to utilize PE for PS synthesis (44). In complementary experiments, the expression of murine PS synthase-2 cDNA in M9.1.1 cells (53) increased the serine- and ethanolamine-exchange activities by 4- and 12-fold, respectively, without increasing the choline-exchange activity. However, the PE level was not normalized in these cells.

These combined observations indicate that in intact cells PS synthase-1 uses PC, but not PE, as a substrate for PS synthesis. On the other hand, PS synthase-2 synthesizes PS from PE, but not from PC. For reasons not completely understood, in *in vitro* assays of cell lysates PS synthase-1 can use either PC or PE for the serine-exchange reaction.

C. Metabolic Differences Between PS Synthase-1 and -2

Several studies indicate that overexpression of PS synthase-1 and -2 activities differentially affects phospholipid metabolism. First, in hepatoma cells that overexpress PS synthase-1 activity (by three-fold according to *in vitro* serine-exchange assays), the rate of PS synthesis (as monitored by the incorporation of [³H]serine into PS) is correspondingly increased three-fold (48). In contrast, in cells that overexpress PS synthase-2 activity (by 2.5-fold according to *in vitro* serine-exchange assays), the rate of PS synthesis is not increased compared to that in parental hepatoma cells (53). The reason why similar levels of overexpression of PS synthase-1 and -2 differentially modulate PS synthesis is not clear. One possible explanation is that PS synthesis via murine PS synthase-2, but not PS synthase-1, is regulated by end-product inhibition (see Section V.A). An alternative explanation is that the amount of PS synthase-1, but not PS synthase-2, might be limiting for PS synthesis. Therefore, an increase in the expression of PS synthase-1 would be expected to increase the rate of PS production. In contrast, if PS synthase-2 were already present in excess, the amount of PS synthase-2 would not be limiting for PS production. Thus, an increase in PS synthase-2 activity would not be expected to increase the rate of PS synthesis.

A second difference in the way in which phospholipid metabolism is differentially altered by overexpression of the two PS synthases is that an increase in PS synthase-1, but not PS synthase-2, activity in hepatoma cells inhibits the CDP-ethanolamine pathway for PE synthesis (53). Perhaps only when the rate of PS synthesis in the cells is increased (i.e., in cells overexpressing PS synthase-1 activity) is the amount of PE made from PS decarboxylation increased. Only under these circumstances of increased production of PE, therefore, would the CDP-ethanolamine pathway for PE synthesis be inhibited.

The PS synthase-2 gene, which resides on chromosome 11 in humans, has not been characterized in detail. The reason why mammalian cells contain two different enzymes for the synthesis of PS is not understood, although there are many parallel examples in which enzymatic activities involved in lipid metabolism are apparently duplicated and are encoded by different genes (54). An important question is: does this duality merely represent redundancy (i.e., a backup mechanism) or do the different isoforms provide pools of lipids that are specifically used for different biological functions?

One indication that PS synthase-1 and -2 might provide pools of PS for distinct functions is that the distribution of PS synthase-1 and -2 mRNAs in murine tissues is different (52,53,55). Northern blotting (53,55), reverse transcriptase-mediated PCR (52), and direct enzymatic assays (52) have revealed that PS synthase-1 is ubiquitously expressed in most tissues but is most abundant in kidney, brain, and liver. The choline-exchange activity, presumably contributed by PS synthase-1, is highest in the brain (52). On the other hand, PS synthase-2 has a more restricted tissue distribution, being most highly expressed in testis and with a lower level of expression in brain and kidney (52,55). The PS synthase-2 gene is particularly highly expressed in the Sertoli cells of the testis and the Purkinje cells of the brain, as well as in brown adipose tissue (55). Whether or not the two PS synthase isoforms perform unique *in vivo* functions in whole animals is not yet clear.

As a step toward understanding the functions of PS synthase-1 and -2 in whole animals, "knockout" mice were recently generated in which the PS synthase-2 gene was disrupted (55). PS synthase-2^{-/-} mice are viable but the testis weight of the mice is modestly reduced when compared to that of wild-type littermates. In addition, some of the male PS synthase-2^{-/-} mice are infertile. The defect in the testis presumably is related to the finding that the principal site of expression of PS synthase-2 in mice is the Sertoli cells of the testis. These cells normally secrete inhibin B, which negatively regulates serum levels of follicle-stimulating hormone (56). In PS synthase-2-deficient mice, the serum level of follicle-stimulating hormone is increased compared to that in wild-type littermates, supporting the idea that PS synthase-2^{-/-} mice have a mild dysfunctioning of their Sertoli cells (56). Interestingly, in all

tissues of the PS synthase-2-deficient mice examined (testis, liver, brain, and heart), the serine-exchange activity was dramatically reduced by 65–98% and the ethanolamine-exchange activity was reduced in parallel. In contrast, the choline-exchange activity was increased two- to three-fold. However, no compensatory increase in the expression of PS synthase-1 mRNA was detected (55).

In light of the massive reduction of serine-exchange activity in tissues of the PS synthase-2^{-/-} mice, particularly in the testis, it was surprising that the phospholipid composition, including the content of PS, was normal. It appears, therefore, that in wild-type mice serine-exchange activity is present in great excess of its requirement for maintaining normal levels of PS and PE. Consequently, the residual serine-exchange activity in PS synthase-2-deficient mice, presumably imparted by PS synthase-1, apparently generates sufficient PS to maintain normal levels of PS and PE. It remains to be determined whether the rate of degradation of PS and PE is compensatorily diminished in PS synthase-2-deficient cells/tissues, and whether PE synthesis from the CDP-ethanolamine pathway is increased, as mechanisms for maintaining PS and PE homeostasis.

It should be noted that *in vitro* assays of serine-exchange activity are usually performed under optimal conditions so that substrates and cofactors are present in excess. The enzyme activity, as measured *in vitro*, therefore, reflects the amount of enzyme but does not necessarily indicate the rate of PS synthesis in an intact cell, in which a factor other than the amount of enzyme might be limiting for PS synthesis. As noted earlier, in cells overexpressing PS synthase-2 activity, no increase in the incorporation of radiolabeled serine into PS was observed (53). This observation is consistent with the hypothesis that PS synthase-2 activity is normally present in excess of its requirement for PS synthesis.

Since PS synthase-1 null mice have not been generated, we do not yet know if the PS synthase-1 gene is essential, or how elimination of PS synthase-1 activity would affect phospholipid homeostasis in a whole animal.

D. Subcellular Location of Mammalian PS Synthases

Serine-exchange activity is present in microsomes and in ER membranes (57–60). Mammalian PS synthase-1 lacks a typical N-terminal signal sequence for targeting the protein to the ER but contains five putative membrane-spanning regions, according to hydropathy plot analysis (43). In addition, the C-terminus contains a Lys–Lys motif that has been proposed to be an ER membrane retention signal (61). The hydropathy plot profile of PS synthase-2 is very similar to that of PS synthase-1 (43,62), suggesting that PS synthase-2 also possesses several membrane-spanning regions. In addition, PS synthase-2 contains at its amino terminus a double Arg motif, which is a

known ER targeting signal (63). These findings, as well as the observation that immunoreactive PS synthase-1 is detectable in microsomes of CHO-K1 cells (47), are consistent with PS synthase-1 and -2 being primarily localized to ER membranes. In addition, rat liver Golgi membranes contain some serine-exchange activity (57).

Although purified mitochondria do not contain PS synthase activity, this activity is robust in a membrane fraction called the "mitochondria-associated membranes" (MAM) that have been isolated from rat liver (64), murine liver (65), and CHO-K1 cells (59). MAM are ER-like membranes that coisolate with mitochondria but can be separated from the latter by centrifugation of crude mitochondria on a Percoll gradient (64). MAM have many, but not all, properties of the ER. One difference is that the specific activity of an enzyme traditionally used as a marker of the ER, NADPH:cytochrome *c* reductase, in the MAM is only 20–30% of that in purified ER (64).

MAM are enriched, compared to the bulk of the ER, in several lipid biosynthetic enzymes such as acyl-CoA:cholesterol acyltransferase (66), diacylglycerol acyltransferase (66), acyl-CoA synthetase-4 (67), as well as enzymes involved in the biosynthesis of glycosylphosphatidylinositol anchors of proteins (68). Interestingly, the specific activity for serine exchange is several-fold higher in MAM than in purified ER fractions (59,64,65). A unique marker protein for MAM is PE *N*-methyltransferase-2 (69), which is an isoform of the methyltransferase that converts PE to PC in liver (70) (Fig. 2). Although the specific activity for PE methylation is similar in the ER and MAM, an antibody raised against a peptide sequence of PE methyltransferase-2 recognizes a protein in the MAM that is not detectable in purified ER (69).

The function of MAM has not been unequivocally established. One proposal is that these membranes constitute a specific region of the ER that is in close juxtaposition with mitochondria (64). Because of this apparent linkage, or close apposition, between MAM and mitochondria, and since PS synthase activity is concentrated in MAM, it has been suggested that MAM mediate the import of newly made PS into mitochondria (see Section VI) (64,65,71,72). Some experimental evidence has been provided in support of this hypothesis (59,71).

On the basis of the observation that both MAM and ER possess PS synthase activity, we hypothesized that the two PS synthase isoforms might be topologically separated in the cell, with one isoform residing in the MAM, the other in the ER. Such a spatial separation would, therefore, be able to produce independent pools of PS that could be used for different functions. The subcellular localization of PS synthase-1 and -2 was, therefore, investigated using immunofluorescence confocal microscopy of liver subcellular fractions as well as *myc*-tagged murine PS synthase cDNAs stably transfected into

hepatoma cells. PS synthase-1 and -2 were generally found to colocalize with each other and also with the resident ER protein, calnexin (65). Previous immunoblotting experiments had revealed that in CHO-K1 cells PS synthase-1 is present in MAM and also, to a lesser extent, in microsomes (which comprises ER membranes as well as MAM) (47). Further immunoblotting studies of murine liver subcellular membrane fractions revealed that PS synthase-1 is abundant in the MAM (which also contains the marker protein, PE *N*-methyltransferase-2) but is undetectable in the ER (which lacks PE *N*-methyltransferase-2) (65). Enzymatic assays demonstrated that choline-exchange activity (presumably contributed by PS synthase-1) is highly enriched in MAM and barely detectable in the ER (65). Consistent with these studies we concluded that in murine liver, PS synthase-1 is located in MAM but is excluded from the ER.

Our hypothesis was, therefore, that PS synthase-2 would be present in ER membranes. Since antibodies that recognized murine PS synthase-2 were not available, the subcellular location of PS synthase-2 was examined indirectly by stable expression of *myc*-tagged PS synthase-2 in hepatoma cells, followed by subcellular fractionation and immunoblotting experiments. Contrary to our expectations, we found that the expressed PS synthase-2 was exclusively located in MAM and was not detectable in microsomes (65). Thus, *both* PS synthase-1 and -2 appear to be localized to the MAM and excluded from the bulk of the ER.

The reason why the two PS synthase isoforms are so highly enriched in MAM is not clear. One speculation is that since essentially all mitochondrial PE is normally derived from PS decarboxylation in the mitochondria (59), a robust synthesis of PS in MAM, which appear to be in close proximity to mitochondrial outer membranes, would provide an efficient mechanism by which PS could be imported into mitochondria for conversion to PE. An intriguing question remains: if both PS synthase-1 and -2 are restricted to the MAM, which protein accounts for the serine- and ethanolamine-exchange activities present in the bulk of the ER?

E. Phosphatidylserine Synthase in Yeast and Prokaryotes

As shown in Fig. 1A, CDP-diacylglycerol serves as the donor of the phosphatidyl moiety for PS biosynthesis in *E. coli* (24) and yeast (22,49,73). In both of these organisms, PS is a key precursor of the other major phospholipids (Fig. 2). In *S. cerevisiae* the primary pathway for the biosynthesis of PC, the most abundant phospholipid, involves the methylation of PE derived from PS decarboxylation. Only a single PS synthase gene apparently exists in yeast and *E. coli*.

PS synthase has been purified from *S. cerevisiae* (73) and the structural gene has been isolated (23). This gene encodes a 276-amino acid protein with

a calculated molecular weight of 30,804. Yeast mutants have been isolated that are deficient in PS synthase activity (23,74). In these mutants, although PE cannot be made from the PS decarboxylation pathway, PE can be made by the CDP-ethanolamine pathway.

The PS synthase from *E. coli* has also been purified to homogeneity (75) and extensively characterized (75–77). In addition, PS synthases from several other bacteria (e.g., *B. subtilis* (24), *Salmonella typhimurium* (78), and *Clostridium perfringens* (79,80)) have been studied. Surprisingly, the PS synthase from *E. coli* exhibits almost no homology to the yeast PS synthase, although the yeast enzyme shows much closer sequence resemblance to the PS synthase from *B. subtilis* (81).

IV. Enzymes of Phosphatidylethanolamine Biosynthesis

A. The CDP-Ethanolamine Pathway

The CDP-ethanolamine pathway for PE biosynthesis (Fig. 2) was elucidated by Kennedy and Weiss in 1956 (26). This pathway is used for PE biosynthesis in mammalian cells, as well as in lower eukaryotes such as yeast. Since animals do not have the capacity to synthesize ethanolamine *de novo*, this PE precursor must be supplied in the diet. Alternatively, small amounts of ethanolamine can be generated in mammalian cells and yeast from the breakdown of sphingolipids via the action of sphingosine phosphate lyase (82,83). The direct formation of ethanolamine from the decarboxylation of serine has not been observed in mammalian cells but, interestingly, this reaction has recently been detected in the plant *Arabidopsis thaliana* (84).

The first step in PE biosynthesis via the CDP-ethanolamine pathway is the phosphorylation of ethanolamine by the cytosolic enzyme ethanolamine kinase (Fig. 2). Several mammalian ethanolamine kinase isoforms have been identified, some of which also possess choline kinase activity (85–87). cDNAs encoding ethanolamine kinase activity have been cloned from human cells (87,88) as well as rat and mouse liver (85,89). In *Drosophila*, choline kinase and ethanolamine kinase activities were shown to reside on distinct gene products because when ethanolamine kinase activity was eliminated in the easily shocked (*eas*) mutant, choline kinase activity was normal (90). In these mutant flies, the amount of PE was only slightly reduced (from 56% to 53% of total phospholipids). Presumably, the PS decarboxylation pathway was able to compensate for disruption of the CDP-ethanolamine pathway so that an almost normal level of PE was produced. The *Drosophila eas* mutants exhibited an interesting phenotype: they underwent transient paralysis after

a mechanical shock. It is not clear if or how the slight decrease in the PE content of the mutant produced such a dramatic phenotype. One speculation is that a small, specific pool of PE that is required for normal neuronal function in *Drosophila* had been depleted. A gene-encoding ethanolamine kinase from *S. cerevisiae* has also been cloned (91,92) and shown to have extensive sequence homology to the mammalian ethanolamine/choline kinases (93,94).

The second step of the CDP-ethanolamine pathway is catalyzed by the cytosolic enzyme CTP:phosphoethanolamine cytidylyltransferase (Fig. 2) (95) which generates CDP-ethanolamine. This enzyme is widely expressed in mammalian tissues. A human cDNA encoding this enzyme has been cloned and shown to have extensive regions of sequence homology to the mammalian CTP:phosphocholine cytidylyltransferase (96). The human CTP:phosphoethanolamine cytidylyltransferase exhibits a high degree (39%) of amino acid similarity to the yeast enzyme (97). Interestingly, the yeast and human CTP:phosphoethanolamine cytidylyltransferases contain regions in which the sequence of the catalytic domain has been duplicated. In contrast to the corresponding enzyme of the CDP-choline pathway for PC synthesis, for which three isoforms are encoded by two separate genes (98), a single gene apparently encodes the cytidylyltransferase that generates CDP-ethanolamine.

When a cDNA encoding the human CTP:phosphoethanolamine cytidylyltransferase was expressed in rat hepatoma cells, the *in vitro* enzymatic activity was increased by ~three-fold. However, the incorporation of [³H]ethanolamine into PE was only modestly increased and the PE content of the cells was unaltered (S. Stone and J. E. Vance, unpublished data). Interestingly, this overexpression of CTP:phosphoethanolamine cytidylyltransferase activity neither inhibited PS synthesis nor reciprocally inhibited the production of PE from the PS decarboxylation pathway (S. Stone and J. E. Vance, unpublished data).

The final reaction of the CDP-ethanolamine pathway (Fig. 2) is catalyzed by CDP-ethanolamine:1,2-diacylglycerol ethanolaminephosphotransferase. This enzyme is an integral membrane protein found primarily on ER membranes, although some enzymatic activity has also been detected on Golgi membranes (57). A yeast gene (*EPT1*) has been cloned that encodes both choline- and ethanolaminephosphotransferase activities. This enzyme is thought to be the one that is primarily involved in PE synthesis from CDP-ethanolamine in yeast (99). A related yeast gene (*CPT1*) encodes cholinephosphotransferase activity and appears not to be involved in the CDP-ethanolamine pathway (100). Based on sequence homology to the yeast gene, a human choline/ethanolaminephosphotransferase cDNA was isolated (101,102). The mRNA corresponding to this cDNA is ubiquitously

expressed in human tissues. It is likely that this cDNA corresponds to the enzyme that was purified from bovine liver and that exhibited both ethanolamine- and cholinephosphotransferase activity (103). In addition, another human cDNA, with 60% sequence identity to the cDNA encoding the human choline/ethanolaminephosphotransferase, has been isolated and shown to be specific for CDP-choline (102). Taken together, these data suggest that in mammalian cells the dual specificity ethanolamine/cholinephosphotransferase gene is responsible for the ethanolaminephosphotransferase activity.

B. The PS Decarboxylation Pathway

The PS decarboxylase pathway for PE synthesis (Fig. 2) operates in prokaryotic, yeast, and mammalian cells. PS decarboxylase enzymatic activity was first described by Kanfer and Kennedy in 1964 (104). The *E. coli* PS decarboxylase, which is located on the cytoplasmic membrane, was purified by Dowhan *et al.* (105) and the gene was subsequently cloned (106).

In mammalian cells, PS decarboxylase is an integral membrane protein whose active site is located on the outer aspect of mitochondrial inner membranes (60,107). The mammalian PS decarboxylase plays an important role in the generation of mitochondrial PE. Indeed, the majority of mitochondrial PE is derived from PS decarboxylation rather than being imported from the ER (the site of the final step of the CDP-ethanolamine pathway) (59). PE that is synthesized in mitochondria is efficiently exported to other organelles including the plasma membrane and the ER (108,109). Although the decarboxylase has not been purified to homogeneity from any mammalian source, mutant CHO-K1 cells deficient in this enzymatic activity were generated which permitted the cloning of a PS decarboxylase cDNA (110). All PS decarboxylase activity in mammalian cells appears to be encoded by a single gene.

In contrast to mammalian cells, yeast contains two distinct PS decarboxylase genes, *PSD1* and *PSD2*, whose sequences are only 19% homologous (111,112). The yeast PS decarboxylase-1, like the mammalian PS decarboxylase, is located in mitochondria and accounts for 95% of PS decarboxylase activity. The remaining 5% of PS decarboxylase activity in yeast has been ascribed to PS decarboxylase-2, which is not mitochondrial but is located in the vacuole and Golgi apparatus (111,112). Yeast with null mutations in either *PSD1* or *PSD2* are viable. However, yeast strains lacking both *PSD1* and *PSD2* are ethanolamine auxotrophs, presumably because they require ethanolamine for the synthesis of PE from the CDP-ethanolamine pathway (111,112).

PS decarboxylase belongs to a small family of decarboxylases that includes histidine decarboxylase (113,114). Instead of containing the more common

pyridoxal phosphate prosthetic group, this group of enzymes contains a pyruvoyl prosthetic group. The existence of the pyruvoyl moiety in PS decarboxylase was first demonstrated for the *E. coli* enzyme (106) but it is widely believed that PS decarboxylases from a wide variety of organisms, including mammalian cells, also contain the pyruvoyl group. The mechanism by which the pyruvoyl group is formed involves the initial synthesis of a 36-kDa precursor protein which is proteolytically cleaved to generate α and β subunits of 7.3 and 28.6 kDa, respectively (106). The proteolysis occurs autocatalytically and involves a serinolysis reaction of Ser-254, resulting in cleavage of the protein between Gly-253 and Ser-254. During the proteolysis, Ser-254 is converted into a pyruvoyl moiety that forms the amino terminus of the α subunit (106). A LGST motif in the bacterial enzyme has been identified as the site of cleavage. This motif is conserved in both the mammalian PS decarboxylase and the yeast PS decarboxylase-1 (115). The yeast PS decarboxylase-2 contains a similar GGST motif, instead of the LGST, and is also thought to undergo the same type of endoproteolytic processing (111).

V. Regulation of PS and PE Biosynthesis

A. Phosphatidylserine

The mechanisms of regulation of the synthesis of PS and PE in mammalian cells have been investigated in surprisingly very few studies. Some experiments from Kanfer *et al.* have suggested that the choline-, serine-, and ethanolamine-exchange activities in rat brain are regulated by phosphorylation-dephosphorylation reactions (116).

A feedback control mechanism has also been identified for regulation of PS synthesis in mammalian cells. The addition of PS to CHO-K1 cells resulted in a striking ($\sim 70\%$) suppression of endogenous PS synthesis (117). Moreover, serine-exchange activity was inhibited when PS was included in *in vitro* enzymatic assays of cell extracts (117). Upon further investigation of this phenomenon, a CHO-K1 cell mutant, called mutant 29, was found to have an increased rate of PS synthesis compared to that of wild-type CHO-K1 cells. The mutant 29 cells were also resistant to inhibition of PS biosynthesis by exogenously added PS (118). Furthermore, addition of PS to a membrane fraction from mutant 29 cells failed to inhibit the *in vitro* serine-exchange activity. Nishijima and coworkers subsequently identified a mutation in the PS synthase-1 gene in mutant 29 cells in which Arg-95 was changed to a Lys residue. This mutation caused the mutant 29 cells to be resistant to end-product inhibition by PS (118,119).

The finding that PS synthesis in PSA-3 cells, which lack PS synthase-1, was similarly inhibited by exogenously added PS suggested that PS synthase-2 might also be regulated by end-product inhibition (120). PS synthase-1 and -2 in CHO-K1 cells have 38% amino acid sequence identity (43,62). PS synthase-2 contains an Arg residue at position 97, corresponding to the Arg-95 that was identified as the crucial residue for regulation of PS synthase-1 activity by PS. Therefore, Arg-97 of PS synthase-2 was mutated to a Lys residue and, consequently, the end-product inhibition of PS synthase-2 by PS was abolished. Thus, Arg-97 of PS synthase-2 and Arg-95 of PS synthase-1 are required for end-product inhibition of PS synthase activity by PS (120).

In yeast, the major mechanism identified for regulation of PS synthase activity, in concert with the regulation of other phospholipid biosynthetic genes, is transcriptional repression by *myo*-inositol and choline (73,121–124). The yeast PS synthase activity is also regulated during the cell cycle with enzyme activity being maximal during the exponential phase of cell growth and declining by 75% upon entry into the stationary phase (124). The mechanism responsible for this regulation of PS synthase activity has not been elucidated but is apparently not related to changes in the amount of PS synthase protein. In addition, the yeast PS synthase is phosphorylated at a single serine residue by cAMP-dependent protein kinase, resulting in a 60–70% reduction in enzymatic activity (125).

B. Phosphatidylethanolamine

Regulation of the PS decarboxylation pathway is poorly understood in mammalian cells. The rate-limiting step in the production of PE from this pathway is the transport of newly made PS to the site of the decarboxylase on the outer aspect of the mitochondrial inner membrane (126).

All available evidence indicates that the CDP-ethanolamine pathway for PE synthesis is regulated independently of the CDP-choline pathway for PC synthesis. Nevertheless, in parallel with the synthesis of PC via the CDP-choline pathway (127), the CTP:phosphoethanolamine cytidylyltransferase is generally thought to catalyze the rate-limiting step of the CDP-ethanolamine pathway (128,129). Under most metabolic conditions the primary mode of regulation of the CDP-choline pathway is thought to be the translocation of the CTP:phosphocholine cytidylyltransferase between a soluble form (an inactive reservoir of the enzyme) and a membrane-bound form (the active form of the enzyme) (130,131). Although some evidence indicates that the predominantly cytosolic CTP:phosphoethanolamine cytidylyltransferase can also partially associate with ER membranes (132,133), there is no evidence that the enzyme activity is regulated by translocation between the cytosol and membranes. Moreover, in contrast to CTP:phosphocholine cytidylyltransferase, the activity of the CTP:phosphoethanolamine cytidylyltransferase

is not regulated by lipids (133). One potential mode of regulation of the CTP:phosphoethanolamine cytidyltransferase is by a phosphorylation–dephosphorylation mechanism, although it is not yet known whether or not the protein is phosphorylated *in vivo*. The amount of diacylglycerol, which is a substrate for ethanolaminephosphotransferase and is probably controlled by the supply of fatty acids, can also limit the rate of PE biosynthesis via the CDP-ethanolamine pathway under some metabolic conditions (129,134).

The ability of the various enzymatic steps to regulate the CDP-ethanolamine pathway might differ among tissues. For example, in hamster heart the activity of ethanolamine kinase has been proposed to regulate flux through the pathway (135). The same conclusion was reached from studies in which human ethanolamine kinase activity was expressed at a high level in COS-7 cells (87). In contrast, however, other studies have indicated that an increased expression of the yeast or the mammalian ethanolamine kinase does not increase the rate of PE synthesis via the CDP-ethanolamine pathway (85,86).

VI. Interorganelle Transport of PS and PE

A. In Mammalian Cells

Our understanding of the molecular mechanisms by which lipids are transported between organelles in eukaryotic cells is still rudimentary. The intermembrane transport of PS and PE could, potentially, be mediated by soluble protein carriers, vesicular transport, or a close apposition or fusion of donor and acceptor membranes. It is likely that different mechanisms, or a combination of these mechanisms, operate for different lipids and for individual pairs of donor and acceptor membrane. The mechanism of transport of PS from its sites of synthesis in the ER and MAM to its site of decarboxylation on the mitochondrial inner membrane has been investigated in yeast and mammalian cells (136). In many of these experiments, PS was pulse-labeled with [^3H]serine and the transport of PS to mitochondria was monitored by observing the conversion of [^3H]PS to [^3H]PE. The slow step in the conversion of PS to PE has been shown to be the interorganelle transport step since PS is rapidly decarboxylated upon its arrival at the mitochondrial inner membrane (126). Using this type of experimental approach, the conversion of PS to PE has been exploited to determine the mechanism of transport of PS from the ER/MAM to mitochondria (136).

Several distinct transport steps are involved in the conversion of newly synthesized PS to PE. First, PS travels from its sites of synthesis (ER/MAM) to

the outer leaflet of the mitochondrial outer membrane. Next, the PS undergoes transbilayer movement across the outer membrane. Finally, the PS is transported to the outer aspect of the mitochondrial inner membrane, the site of PS decarboxylase. The latter step might involve movement of PS across the intermembrane space or, alternatively, PS on the inner leaflet of the mitochondrial outer membrane might be directly accessible to the decarboxylase on the inner membrane, perhaps at sites of contact between outer and inner membranes (137). Contact sites between mitochondrial outer and inner membranes are thought to participate in the import of proteins into mitochondria (138) and perhaps this is also the case for phospholipids, such as PS. Consistent with this idea, an ER-like membrane fraction, most likely corresponding to MAM, was coisolated with a mitochondrial contact site-enriched fraction of murine hepatic mitochondria, suggesting that an association between MAM and mitochondrial membranes provides an efficient mechanism for the import of PS into mitochondria (139,140). Indeed, many morphological studies using electron microscopy have indicated that there are zones of apposition or contact between the ER and mitochondria (141–149). Dinitrophenol, which apparently reduces the number of contact sites between mitochondrial inner and outer membranes, was shown to inhibit the uptake of PS by mitochondria (150). Moreover, in permeabilized mammalian cells the import of PS into mitochondria was inhibited by adriamycin, an agent that disrupts the import of proteins into mitochondria (137,151).

The mechanism of import of newly synthesized PS into mitochondria has been extensively investigated in several experimental systems: intact mammalian and yeast cells (64,111,112,126,152–154), reconstituted systems comprising isolated mitochondria and microsomes/MAM (59,71,126), and permeabilized mammalian (137,155,156) and yeast (157,158) cells. Several independent experiments have indicated that, in contrast to earlier predictions, cytosolic phospholipid transfer proteins are not required for this transport process. For example, the conversion of PS to PE occurs efficiently in permeabilized CHO-K1 cells (152,156) and yeast (157) that have been depleted of essentially all of their cytosolic proteins. Moreover, the addition of cytosol does not significantly stimulate the process (152,156). In addition, in cell-free reconstituted systems, comprising microsomes and mitochondria, the movement of PS into mitochondria was shown to be independent of the presence of cytosolic proteins (126,154). In addition, in a mutant CHO-K1 cell line lacking the nonspecific lipid transfer protein (also known as sterol carrier protein-2), which had been considered to be a likely candidate for mediating the translocation of PS into mitochondria, PS is imported normally into mitochondria (159). These data are consistent with other studies showing that sterol carrier protein-2

functions primarily in peroxisomes (160), and that in genetically modified mice lacking this protein, the phospholipid composition of the membranes is normal (161).

However, in contrast to data showing that cytosolic proteins do not regulate PS transport to the mitochondria, a recent study by Nishijima and coworkers has found that a cytosolic protein from bovine brain stimulates the transport-dependent decarboxylation of PS in permeabilized CHO-K1 cells without increasing the activity of PS decarboxylase (162). The active protein was identified as a member of a family of calcium-binding proteins. The authors speculate that this protein regulates either the PS transport machinery or increases the number of contact sites between the ER and MAM.

The concerted synthesis and decarboxylation of PS also occur in rat liver mitochondria that have been isolated in tight association with MAM (64,154), suggesting that a physical association between MAM and mitochondria provides a conduit for the import of PS into mitochondria. The concept that an association between MAM and mitochondria mediates the import of PS into mitochondria is supported by several other findings. First, newly synthesized PS, rather than preexisting PS, is preferentially imported into mitochondria for decarboxylation to PE (154,163). Second, an experiment was devised in which permeabilized CHO-K1 cells were disrupted by shearing, and a population of "donor" cells was prepared in which PS had been radiolabeled. These cells were poisoned with hydroxylamine so that PS decarboxylase was inactivated. "Acceptor" cells, in which the ability to decarboxylate PS had been retained, were then mixed with the donor cells. The radiolabeled PS was not decarboxylated (155). These experiments provide strong evidence that there is a channeling that results in a restricted movement of PS from its site of synthesis into mitochondria. Moreover, these observations suggest the existence of a physical association between mitochondria and the membranes that synthesize PS. A likely candidate for this membrane linked to mitochondria is the MAM. A similar linkage between MAM and mitochondrial membranes has been proposed to be involved in the import of phospholipids into mitochondria in yeast (164,165).

In intact mammalian cells, ATP is required for the transport of newly synthesized PS to mitochondria (108,126,152), whereas in a reconstituted system comprising isolated microsomes and mitochondria, there is no requirement of ATP for the synthesis, transport, or decarboxylation of PS (152). This difference between intact cells and the reconstituted system of isolated organelles suggests that *in vivo* the ATP-dependent step precedes the events that are reconstituted in the *in vitro* reconstituted system. One potential role for ATP is in increasing the accessibility of newly synthesized PS to the mitochondrial outer membrane.

In yeast, however, ATP appears not to be required for PS import into mitochondria (157). The difference between the ATP requirement for PS transport in intact mammalian cells (108) and yeast might be a consequence of differences in the accessibility of PS to the translocation machinery, since in yeast PS is made by a completely different pathway (from CDP-diacylglycerol) from that operating in mammalian cells (a base-exchange reaction) (Fig. 1).

Although genes involved in the import of PS into mitochondria have not yet been identified, a CHO-K1 cell mutant, designated R-41, has been isolated that is defective in the conversion of PS to PE (166). In these cells, the enzymatic activities of PS synthase and PS decarboxylase are unaffected, and the transport of proteins into mitochondria occurs normally. However, the transport of PS from the mitochondrial outer membrane to the inner membrane is reduced by 60% compared to that in wild-type cells (166). The mutant gene responsible for this defect in PS transport has not yet been identified.

The collective experimental observations summarized here indicate that PS is channeled into mitochondria via the MAM. Some evidence supporting this model comes from pulse-chase studies in which CHO-K1 cells were labeled with [³H]serine, then subjected to subcellular fractionation. Newly made [³H]PS was found in microsomes and MAM (59). However, when ATP was rapidly depleted immediately after the pulse-labeling, [³H]PS accumulated in the MAM and, consequently, was prevented from reaching the mitochondria. This finding suggests that PS traverses the MAM en route to the mitochondria and that ATP is required for the movement of PS out of the MAM. Reconstitution experiments with trypsin-treated MAM and/or mitochondria have indicated that the import of PS into mitochondria requires a protein that is present in the mitochondrial outer membrane, but does not require a protein in the MAM (71).

Almost nothing is known about the molecular basis for the interorganelle transport of PE in mammalian cells. Early studies indicated that the transport of PE from the ER to mitochondria in hepatocytes was ~80-fold slower than for PC (167). As noted earlier, the majority of PE present in mitochondria (an organelle that is enriched in PE compared to other organelles) is made *in situ* from PS decarboxylation, in preference to being imported from the ER (59).

PS-derived PE is also rapidly exported from the mitochondria to other organelle membranes, such as the ER (109,152) and the plasma membrane (109). The transport of PE derived from either CDP-ethanolamine (in the ER), or from the decarboxylation of PS (in mitochondria), to the plasma membrane is insensitive to brefeldin A (109), an agent that disrupts the Golgi apparatus and inhibits the transport of proteins that pass through the Golgi

(168,169). Consequently, PE derived from each of its biosynthetic origins appears not to traverse the Golgi apparatus *en route* to the plasma membrane. Other experiments have demonstrated that the transport of ER-derived PE to the plasma membrane is not inhibited by metabolic poisons that impair vesicle transport or disrupt the cytoskeleton (170). Thus, the movement of PE from the ER to the plasma membrane appears to occur via a mechanism that is distinct from the vesicle-mediated transport processes involved in the trafficking of proteins to the plasma membrane or for secretion. The data available so far are consistent with the transport of PE to the plasma membrane being mediated by either a soluble carrier protein or by membrane juxtaposition.

B. In Yeast Cells

Similar studies have been performed in yeast to elucidate the mechanisms of PS and PE transport. In yeast, PS import into mitochondria, for decarboxylation to PE, does not require ATP (157). As in mammalian cells, the rate of import of PS into mitochondria of yeast limits the rate of PE formation from PS decarboxylation (171). Moreover, only small amounts of PE are imported from the ER into the mitochondria of yeast (112,172,173). Instead, mitochondrial PE is derived primarily from *in situ* synthesis in mitochondria, as was demonstrated in a yeast mutant lacking PS decarboxylase-1 (the mitochondrial form of PS decarboxylase) in which the content of PE in mitochondria was markedly reduced (112). As in mammalian cells, contact sites between mitochondrial inner and outer membranes in yeast appear to participate in the import of PS into mitochondria (164,174,175). Moreover, a membrane fraction in yeast, resembling the MAM isolated from mammalian cells, is tightly associated with mitochondria and contains PS synthase activity (165,176). The coordinated synthesis and decarboxylation of PS have also been reconstituted in an *in vitro* system of yeast MAM and mitochondria (176).

In yeast, PE that has been made in mitochondria can be returned to the MAM/ER for conversion to PC via the methylation pathway (Fig. 2) (176). Permeabilized yeast cells have been used to study the concerted synthesis of PS, the decarboxylation of PS to PE, the return of PE to the MAM/ER, and the subsequent methylation of PE to PC (157). These studies showed that although the import of PS into mitochondria does not require ATP, the export of PE from mitochondria is inhibited when ATP is depleted.

Recent advances have been made in identifying genes involved in the intracellular transport of PS and PE in yeast. Voelker and coworkers have generated a series of mutant yeast strains (136) from which genes involved in the interorganelle transport of PS and PE have been identified. Yeast that

harbor a null mutation in *PSD1*, which encodes the mitochondrial isoform of PS decarboxylase, are unable to synthesize or export mitochondrial PE (153). Moreover, yeast deficient in *PSD2*, encoding the Golgi/vacuole form of PS decarboxylase, cannot synthesize/export Golgi/vacuole PE (111,112). Using these yeast strains, further mutations were made and strains of yeast with defects specifically in the transport of PS to the Golgi/vacuole have been generated (158,171,177). Analysis of these mutants suggests that as many as 19 different genes might play roles in the transport of PS and PE in and out of the Golgi/vacuolar compartments. One mutant that is defective in the transport of PS from the ER to the Golgi/vacuole has a mutation in phosphatidylinositol-4-kinase (171). The precise role of this kinase in PS transport has not yet been elucidated. Another yeast strain that is unable to convert PS to PE via PS decarboxylase-2 exhibits profoundly altered phospholipid metabolism and accumulates PS in the Golgi. The defective gene in this mutant was shown to be homologous to the *SEC14* gene, which encodes a PC/phosphatidylinositol transfer protein (177). These findings are consistent with studies in which polyphosphoinositides have been implicated in several aspects of vesicular trafficking (178). Future studies with other yeast mutants are expected to yield exciting new information on the genes that mediate the import of PS into mitochondria.

VII. Intramembrane Transport of PS and PE

Phospholipids are asymmetrically arranged in the lipid bilayer so that the choline-containing phospholipids (PC and sphingomyelin) are enriched on the outer leaflet of the plasma membrane and the topologically equivalent luminal surface of intracellular membranes (179). In contrast, the amino-phospholipids, PS and PE, are highly enriched on the cytoplasmic surface of these membranes (reviewed in Ref. (72)). For example, in the erythrocyte membrane, approximately 80% of PE resides on the inner monolayer as does essentially all of the PS. Since the rate of spontaneous transbilayer movement of phospholipids is very slow with a $T_{1/2}$ of hours to days due to the polar nature of the phospholipid head-groups, proteins are thought to be required for the transbilayer movement of phospholipids (reviewed in Ref. (72)).

Several transport proteins that mediate the transbilayer of lipids have been identified that differ in their substrate specificity, energy requirement, and direction of transport. One class of proteins that move PS and PE across the membrane bilayer is the "scramblases." These proteins translocate PS and PE bidirectionally across the plasma membrane in an ATP-independent fashion and eventually equalize the distribution of these lipids on both sides of the

bilayer. Scramblase activity was originally detected in platelets and erythrocytes that externalized PS in response to an increase in the intracellular calcium concentration during cell activation (180). Scramblase activity is now also known to be present in many other types of cells, including lymphocytes (181,182). Scramblase is a 35-kDa transmembrane protein (183,184) that is palmitoylated (185) and contains a proline-rich N-terminus. The cytosolic domain of the protein contains a calcium-binding motif (185,186). Scramblase activity has recently been shown to be enriched in lipid rafts where the protein interacts with the epidermal growth factor receptor (187). Moreover, the integrity of lipid rafts has been reported to be essential for the outward movement of PS across the plasma membrane in a human megakaryoblastic cell line (188).

Four members of the human scramblase gene family have been identified and these genes are differentially expressed in tissues (189). The activity of scramblase is probably regulated by phosphorylation–dephosphorylation since the protein is phosphorylated by protein kinase C delta (190) and the protein kinase, c-Abl (191). Scramblase activity shows no preference for the type of phospholipid that is transported.

Platelets from patients with a rare inherited disease, Scott syndrome, have an impaired transbilayer migration of PS that results in a bleeding disorder (192). PS that is exposed on the surface of platelet membranes forms the prothrombinase complex with factor Xa and factor Va, thus generating thrombin from prothrombin during blood coagulation (8,9,193). The defect in PS movement across platelet membranes of Scott-disease patients has been proposed to be due to an impaired interaction between calcium and the scramblase, and is probably not due to a defect in the scramblase gene per se (181,194). A similar defect has been reported in dogs (195). Scramblase also appears to play a critical role in the externalization of PS during apoptosis (see Section VIII.A).

A different type of activity that mediates the unidirectional inward transbilayer movement of PS and PE across the plasma membrane has been described and has been designated as a “flippase” (reviewed in Ref. (196)). This plasma membrane transporter is ATP- and magnesium-dependent (197,198) and exhibits a high degree of selectivity toward PS; PE is transported at approximately one-tenth of the rate of PS (198–200). Most data suggest, therefore, that the flippase is a PS-stimulated Mg^{++} -ATPase. This protein has been proposed to prevent PS accumulation on the outside of the plasma membrane. A wide variety of cells express PS flippase activity including erythrocytes (197,198,201), platelets (198), fibroblasts (170,202), hepatocytes (203), and aortic endothelial cells (204). Several candidate proteins have been suggested to be the flippase including members of the P-type ATPase family (205–207). In addition, a protein-catalyzing PS flippase

activity has been partially purified (196). Nevertheless, the mammalian PS/PE flippase in the plasma membrane has yet to be unambiguously identified.

In yeast, PE is translocated in both directions across the plasma membrane. Mutants defective in the inward translocation of a fluorescent short-chain analog of PE were found to be defective in the transcriptional regulators *PDR1* and *PDR3* (208). The precise relationship between these transcriptional regulators and the transbilayer movement of PE is not yet clear but one possibility is that these transcriptional regulators mediate the expression of an ATP-binding cassette protein. The aminophospholipid flippase activity in yeast has been proposed to be encoded by the *DRS2* gene, which is a member of the P-type ATPase family. Several *drs2* mutants were isolated that were partially defective in the inward translocation of PS across the plasma membrane. However, *drs2* null mutants exhibit normal transbilayer movement of PS (209). Thus, the aminophospholipid flippase in yeast remains to be identified.

A third activity mediating the transbilayer movement of PE and PS is designated as a “floppase” activity that catalyzes the outward movement of these lipids across the bilayer. This class of proteins includes the P-type multidrug resistance proteins (MDR and MRP), which are members of the ATP-binding cassette (ABC) family of transporters. One member of the ABC transporter family, ABC-R, is defective in Stargardt’s macular dystrophy (210). This transporter catalyzes the transbilayer movement of a complex comprising retinaldehyde and PE. Consequently, these lipids accumulate in rod outer segment discs in mice with a null allele for ABC-R (211).

Another ABC transporter, ABC-A1, has been implicated in the externalization of PS across the plasma membrane (212). This transporter is defective in patients with the genetic disorder, Tangier disease, which is characterized by extremely low levels of plasma high-density lipoproteins (213–215). The defect in production of high-density lipoproteins has been attributed to a defect in the ABC-A1 gene, resulting in an impaired efflux of PC and cholesterol to apolipoprotein A1, the major protein of high-density lipoproteins. The precise mechanism by which ABC-A1 mediates the efflux of cholesterol and PC from cells has not been established. In particular, it is not known if ABC-A1 mediates the transbilayer movement/efflux of cholesterol and/or PC directly or indirectly. However, ABC-A1 has been shown to catalyze the transbilayer movement of PS from the inner to the outer leaflet of the plasma membrane (212). Consequently, Chimini and coworkers have proposed that ABC-A1 dissipates the transbilayer gradient of PS across the plasma membrane and thereby alters the lipid microenvironment on the cell surface. According to this model, the binding of apoprotein A1 to the cell surface is thereby facilitated by alteration of the membrane structure so that cholesterol and PC can more readily be effluxed from the cells (212). Recent

studies have, however, argued that the amount of cell-surface PS is only slightly modulated by ABC-A1. Therefore, the degree of alteration in the transbilayer distribution of PS by ABC-A1 is probably not sufficient to increase the binding of apoprotein A1 to the cell surface, or to influence the amount of cholesterol and PC effluxed from the cells (216).

The transbilayer movement of lipids also occurs in *E. coli*. A clue to the mechanism by which lipids traverse the *E. coli* inner membrane and are transported to the outer membrane of Gram-negative bacteria, was the finding that the *MsbA* gene, which encodes an essential ABC transporter, mediates the export of newly synthesized lipopolysaccharide and phospholipids, including PE, across the *E. coli* inner membrane (217,218).

VIII. Biological Functions of PS and PE

A. Functions of Phosphatidylserine

Several proteins have been identified that require PS for optimal activity (219). For example, the C2 domain of the conventional forms of protein kinase C interacts with PS on the inner leaflet of the plasma membrane which is normally enriched in PS (219,220). A disruption of the asymmetrical distribution of PS across the plasma membrane decreases the effective concentration of PS on the inner monolayer and thereby reduces the amount of PS available for interaction with protein kinase C. One would predict, therefore, that such a reorganization of the lipids in the membrane would induce profound changes in signal transduction pathways.

Another important function of PS is in the blood-clotting cascade. When platelets are activated at sites of vascular injury, PS and PE move to the outside surface of the membrane and induce the binding of the prothrombinase complex, thereby initiating the blood-clotting cascade by the subsequent generation of thrombin (9). This process is essential for normal blood clotting.

Another well-characterized role for PS is in the clearance of apoptotic cells from the circulation by macrophages (10–13). This process is critical during tissue remodeling and for the resolution of inflammation. Normally, in circulating blood cells PS and PE reside on the inner leaflet of the plasma membrane. During apoptosis, however, intracellular calcium levels increase which results in a decrease in the activity of the aminophospholipid translocase and an increase in the bidirectional scrambling of phospholipids across the membrane. Consequently, the asymmetrical distribution of PS across the plasma membrane is dissipated and PS becomes exposed on the outside of the apoptotic cells (10). Other studies have shown that the

exposure of PE on the cell surface is also increased during apoptosis (221). The externalization of PS is an early event of the apoptotic pathway and can be dissociated from the nuclear changes that occur (13). When PS becomes exposed on the surface of apoptotic cells, the PS is recognized by, and binds to, cell-surface receptors on macrophages that are highly specific for PS (11). The apoptotic cells are subsequently engulfed and destroyed by the macrophages. Thus, these cells are rapidly and specifically cleared from the blood (222,223). Similarly, the phagocytosis of influenza virus-infected cells by macrophages is thought to be mediated, at least in part, by an increased exposure of PS on the surface of the infected cells (224). In addition, the synthesis of PS is increased during apoptosis induced by UV irradiation or camptothecin in U937 human leukemia cells (225).

A perturbation in the asymmetrical distribution of PS in the plasma membrane of erythrocytes has also been reported in sickle cell disease (226,227), stroke (228), and diabetes (229). Moreover, the cell-surface exposure of PS, which likely occurs independently of apoptosis, has been shown to be a normal event in myotube formation during muscle development in embryonic mice (230).

B. Functions of Phosphatidylethanolamine

Several fascinating and unexpected roles have recently been proposed for PE. For example, PE becomes exposed on the cell surface of mammalian cells at the cleavage furrow during cytokinesis (231,232). When cell-surface PE was immobilized by binding specifically to a peptide, actin-filament disassembly at the cleavage furrow, and subsequent membrane fusion and cell division, were blocked. The blockage was reversed by removal of the PE-binding peptide, suggesting an important role for PE in cytokinesis (231). In addition, in a mutant CHO-K1 cell line, in which the PE level was decreased by 50%, cytokinesis was inhibited (232). Addition of PE or ethanolamine to these cells restored normal amounts of PE on the cell surface at the cleavage furrow and cytokinesis resumed. These studies provide evidence that surface exposure of PE at the cleavage furrow is crucial for cytokinesis in mammalian cells.

In *Drosophila*, PE is the major phospholipid comprising ~55% of total phospholipids, whereas in most mammalian cells, PC is the major phospholipid and PE constitutes ~20% of the total phospholipids. Lipid homeostasis in *Drosophila* has recently been shown to be transcriptionally regulated by the PE content of the cells by a mechanism parallel to the exquisite regulation of lipid homeostasis by cholesterol in mammalian cells (16,17). Although *Drosophila* do not make sterols (233), these flies express an ortholog of the mammalian sterol response element binding protein (SREBP),

as well as the SREBP-cleavage activating protein and the two proteases that proteolytically cleave the ER-bound form of SREBP into its truncated form that acts as a nuclear transcription factor (16,17). In mammalian cells, SREBP senses the sterol content of the cells and regulates cholesterol and fatty-acid homeostasis accordingly (234). In contrast, SREBP processing in *Drosophila* is not regulated by sterols but instead is regulated by the amount of PE or a closely related metabolite (16,17). It is not yet known, however, whether or not PE regulates any aspect of SREBP processing in mammalian cells.

These observations led to the suggestion that cholesterol in mammalian cells, and PE in *Drosophila*, act similarly in regulating the processing of SREBP. Brown and coworkers have proposed that PE and cholesterol induce similar physical changes in the ER membrane that determine whether or not SREBP is proteolytically cleaved into its transcriptionally active form (16). One feature common to PE and cholesterol is that both of these lipids can promote the formation of nonbilayer, hexagonal II phases in membranes. PE itself can form hexagonal II phases whereas cholesterol can induce the formation of hexagonal II phases in membranes that are rich in PE (reviewed in Ref. (235)). Thus, by virtue of their ability to induce hexagonal II phases in membranes, it has been suggested that PE and cholesterol act in parallel to regulate SREBP processing in *Drosophila* and mammalian cells, respectively (16).

In relation to the studies on PE in *Drosophila*, it is noteworthy that in yeast PE performs an unidentified but essential function that is independent of the ability of PE to form hexagonal II-phase structures in membranes (173).

A novel function for PE has been described in *E. coli*. PE normally accounts for 70–80% of total glycerophospholipids of *E. coli*. A mutant *E. coli* strain was generated in which the amount of PE was drastically reduced to 0.007% of total phospholipids but the cells were viable when supplied with divalent cations (236). Thus, *E. coli* cells were able to survive in an almost complete absence of PE suggesting that PE might not be essential for *E. coli*. However, the role of PE in membrane protein assembly and function has been further investigated in *E. coli*. In cells severely depleted of PE, the membrane protein lactose permease has been found to be improperly folded/assembled in the membrane (15,237). Thus, PE appears to play an important role in facilitating the folding of membrane proteins, suggesting that PE can act as a molecular chaperone.

IX. Future Directions

Cloning of the genes encoding enzymes involved in phospholipid metabolism and transport has lagged behind that of many other genes

primarily because of difficulties in purification of integral membrane proteins. However, many of these proteins have now been isolated in pure form, and/or the corresponding cDNAs and genes have been isolated. The use of genetic tools and the generation of mutant cell lines have greatly facilitated the identification of the genes involved in the synthesis and intracellular transport of PS and PE. In the near future, most of the remaining unidentified genes are expected to be isolated and characterized.

With the widespread availability of techniques for generating genetically modified mice, some of these genes will be disrupted in mouse models and the physiological functions of the proteins involved in the metabolism of PS and PE in whole animals will be elucidated. Several mutant yeast and mammalian cell lines have already been generated in which genes of PS and PE metabolism have been eliminated. It will also be important to study the functions of the mammalian orthologs of these genes in whole animals since the interaction of lipid-metabolizing enzymes among different tissues of a whole animal might reveal unexpected roles for these genes of PE and PS metabolism and transport.

A major frontier in this area of research is to understand how the biosynthesis of PS and PE is regulated in mammalian cells since almost nothing is known about this topic. These studies are now feasible and should yield important information once the respective genes have been isolated and characterized.

Another area in which detailed knowledge of the genes and proteins involved is lacking is the interorganelle trafficking of PS and PE in mammalian cells. This will be a major challenge in cell biology over the next decade. Studies that have already been initiated in yeast and CHO-K1 cell mutants will provide important starting points for discovery of the corresponding genes involved in the transport of PS and PE in mammalian cells. Furthermore, identification of the genes involved in the transbilayer movement of lipids, several of which are involved in disease processes, is an important goal.

REFERENCES

1. C. R. Raetz, Regulated covalent modifications of lipid A, *J. Endotoxin Res.* **7**, 73–78 (2001).
2. Y. Nishizuka, Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C, *Science* **258**, 607–614 (1992).
3. L. Bittova, R. V. Stahelin and W. Cho, Roles of ionic residues of the C1 domain in protein kinase C- α activation and the origin of phosphatidylserine specificity, *J. Biol. Chem.* **276**, 4218–4226 (2001).
4. Y. Nagai, J. Aoki, T. Sato, K. Amano, Y. Matsuda, H. Arai and K. Inoue, An alternative splicing form of phosphatidylserine-specific phospholipase A1 that exhibits lysophosphatidylserine-specific lysophospholipase activity in humans, *J. Biol. Chem.* **274**, 11053–11059 (1999).

5. K. A. Powell, V. A. Valova, C. S. Malladi, O. N. Jensen, M. R. Larsen and P. J. Robinson, Phosphorylation of dynamin I on Ser-795 by protein kinase C blocks its association with phospholipids, *J. Biol. Chem.* **275**, 11610–11617 (2000).
6. K. Hofmann, S. Tomiuk, G. Wolff and W. Stoffel, Cloning and characterization of the mammalian brain-specific, Mg^{2+} -dependent neutral sphingomyelinase, *Proc. Natl. Acad. Sci. USA* **97**, 5895–5900 (2000).
7. L. J. Pike, X. Han, K. N. Chung and R. W. Gross, Lipid rafts are enriched in arachidonic acid and plasmenylethanolamine and their composition is independent of caveolin-1 expression: a quantitative electrospray ionization/mass spectrometric analysis, *Biochemistry* **41**, 2075–2088 (2002).
8. A. J. Schroit and R. F. A. Zwaal, Transbilayer movement of phospholipids in red cell and platelet membranes, *Biochim. Biophys. Acta* **1071**, 313–329 (1991).
9. E. Bevers, P. Comfurius, J. van Rijn, H. Hemker and R. Zwaal, Generation of prothrombin-converting activity and the exposure of phosphatidylserine at the outer surface of platelets, *Eur. J. Biochem.* **122**, 429–436 (1982).
10. V. A. Fadok, A. de Cathelineau, D. L. Daleke, P. M. Henson and D. L. Bratton, Loss of phospholipid asymmetry and surface exposure of phosphatidylserine is required for phagocytosis of apoptotic cells by macrophages and fibroblasts, *J. Biol. Chem.* **276**, 1071–1077 (2001).
11. V. A. Fadok, D. L. Bratton, D. M. Rose, A. Pearson, R. A. B. Ezekewitz and P. M. Henson, A receptor for phosphatidylserine-specific clearance of apoptotic cells, *Nature* **405**, 85–90 (2000).
12. V. A. Fadok, D. R. Voelker, P. A. Campbell, J. J. Cohen, D. L. Bratton and P. M. Henson, Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages, *J. Immunol.* **148**, 2207–2216 (1992).
13. D. L. Bratton, V. A. Fadok, D. A. Richter, J. M. Kailey, L. A. Guthrie and P. M. Henson, Appearance of phosphatidylserine on apoptotic cells requires calcium-mediated nonspecific flip-flop and is enhanced by loss of the aminophospholipid translocase, *J. Biol. Chem.* **272**, 26159–26165 (1997).
14. A. K. Menon and V. L. Stevens, Phosphatidylethanolamine is the donor of the ethanolamine residue linking a glycosylphosphatidylinositol anchor to protein, *J. Biol. Chem.* **267**, 15277–15280 (1992).
15. M. Bogdanov, J. Sun, H. R. Kaback and W. Dowhan, A phospholipid acts as a chaperone in assembly of a membrane transport protein, *J. Biol. Chem.* **271**, 11615–11618 (1996).
16. I. Y. Dobrosotskaya, A. C. Seegmiller, M. S. Brown, J. L. Goldstein and R. B. Rawson, Regulation of SREBP processing and membrane lipid production by phospholipids in *Drosophila*, *Science* **296**, 879–883 (2002).
17. A. C. Seegmiller, I. Dobrosotskaya, J. L. Goldstein, Y. K. Ho, M. S. Brown and R. B. Rawson, The SREBP pathway in *Drosophila*: regulation by palmitate, not sterols, *Dev. Cell* **2**, 229–238 (2002).
18. E. Baer, J. Maurukas and M. Russell, Synthesis of enantiomeric α -cephalins, *J. Am. Chem. Soc.* **74**, 152–157 (1952).
19. J. Folch, Isolation of phosphatidylserine from brain cephalin, and identification of the serine component, *J. Biol. Chem.* **139**, 973–974 (1941).
20. F. Baer and J. Maurukas, Phosphatidylserine, *J. Biol. Chem.* **212**, 25–38 (1955).
21. H. G. Hübscher, R. R. Dils and W. F. R. Pover, Studies on the biosynthesis of phosphatidylserine, *Biochim. Biophys. Acta* **36**, 518–525 (1959).
22. V. A. Letts, L. S. Klig, M. Bae-Lee, G. M. Carman and S. A. Henry, Isolation of the yeast structural gene for the membrane-associated enzyme phosphatidylserine synthase, *Proc. Natl. Acad. Sci. USA* **80**, 7279–7283 (1983).

23. J.-I. Nikawa, Y. Tsukagoshi, T. Kodaki and S. Yamashita, Nucleotide sequence and characterization of the yeast PSS gene encoding phosphatidylserine synthase, *Eur. J. Biochem.* **167**, 7–12 (1987).
24. M. Okada, H. Matsuzaki, I. Shibuya and K. Matsumoto, Cloning, sequencing and expression in *Escherichia coli* of the *Bacillus subtilis* gene for phosphatidylserine synthase, *J. Bacteriol.* **176**, 7456–7461 (1994).
25. E. Delhaize, D. M. Hebb, K. D. Richards, J.-M. Lin, P. R. Ryan and R. C. Gardner, Cloning and expression of a wheat (*Triticum aestivum* L.) phosphatidylserine synthase cDNA. Overexpression in plants alters the composition of phospholipids, *J. Biol. Chem.* **274**, 7082–7088 (1999).
26. E. P. Kennedy and S. B. Weiss, The function of cytidine coenzymes in the biosynthesis of phospholipides, *J. Biol. Chem.* **222**, 193–214 (1956).
27. L. F. Borkenhagen, E. P. Kennedy and L. Fielding, Enzymatic formation and decarboxylation of phosphatidylserine, *J. Biol. Chem.* **236**, 28–32 (1961).
28. R. Sundler, B. Akesson and A. Nilsson, Quantitative role of base exchange in phosphatidylethanolamine synthesis in isolated rat hepatocytes, *FEBS Lett.* **43**, 303–307 (1974).
29. D. R. Voelker, Phosphatidylserine functions as the major precursor of phosphatidylethanolamine in cultured BHK-21 cells, *Proc. Natl. Acad. Sci. USA* **81**, 2669–2673 (1984).
30. M. A. Miller and C. Kent, Characterization of the pathways for phosphatidylethanolamine biosynthesis in Chinese hamster ovary mutant and parental cell lines, *J. Biol. Chem.* **261**, 9753–9761 (1986).
31. O. Kuge, M. Nishijima and Y. Akamatsu, Phosphatidylserine biosynthesis in cultured Chinese hamster ovary cells. II. Isolation and characterization of phosphatidylserine auxotrophs, *J. Biol. Chem.* **261**, 5790–5794 (1986).
32. R. Sundler and B. Akesson, Biosynthesis of phosphatidylethanolamines and phosphatidylcholines from ethanolamine and choline in rat liver, *Biochem. J.* **146**, 309–315 (1975).
33. L. B. M. Tijburg, M. J. H. Geelen and L. M. G. Van Golde, Biosynthesis of phosphatidylethanolamine via the CDP-ethanolamine route is an important pathway in isolated rat hepatocytes, *Biochem. Biophys. Res. Commun.* **160**, 1275–1280 (1989).
34. T. A. Zelinski and P. C. Choy, Phosphatidylethanolamine biosynthesis in isolated hamster heart, *Can. J. Biochem.* **60**, 817–823 (1982).
35. S. K. F. Yeung and A. Kuksis, Utilization of L-serine in the in vivo biosynthesis of glycerophospholipids by rat liver, *Lipids* **11**, 498–505 (1976).
36. T. Kano-Sueoka, D. M. Cohen, Z. Yamaizumi, S. Nishimura, M. Mori and H. Fujiki, Phosphoethanolamine as a growth factor of a mammary carcinoma cell line of rat, *Proc. Natl. Acad. Sci. USA* **76**, 5741–5744 (1979).
37. H. Murakami, H. Masui, G. H. Sato, N. Sueoka, T. P. Chow and T. Kano-Sueoka, Growth of hybridoma cells in serum free medium: ethanolamine is an essential component, *Proc. Natl. Acad. Sci. USA* **79**, 1158–1162 (1982).
38. T. Kano-Sueoka, J. E. Errick, D. King and L. A. Walsh, Phosphatidylethanolamine synthesis in ethanolamine-responsive and nonresponsive cells in culture, *J. Cell Physiol.* **117**, 109–115 (1983).
39. H. Sasaki, H. Kume, A. Nemoto, S. Narisawa and N. Takahashi, Ethanolamine modulates the rate of rat hepatocyte proliferation in vitro and in vivo, *Proc. Natl. Acad. Sci. USA* **94**, 7320–7325 (1997).
40. T. T. Suzuki and J. N. Kanfer, Purification and properties of an ethanolamine-serine base exchange enzyme of rat brain microsomes, *J. Biol. Chem.* **260**, 1394–1399 (1985).
41. D. R. Voelker and J. L. Frazier, Isolation and characterization of a Chinese hamster ovary cell line requiring ethanolamine or phosphatidylserine for growth and exhibiting defective phosphatidylserine synthase activity, *J. Biol. Chem.* **261**, 1002–1008 (1986).

42. O. Kuge, M. Nishijima and Y. Akamatsu, Isolation of a somatic cell mutant defective in phosphatidylserine biosynthesis, *Proc. Natl. Acad. Sci. USA* **82**, 1926–1930 (1985).
43. O. Kuge, K. Saito and M. Nishijima, Cloning of a Chinese hamster ovary (CHO) cDNA encoding phosphatidylserine synthase (PSS) II, overexpression of which suppresses the phosphatidylserine biosynthetic defect of a PSS I-lacking mutant of CHO-K1 cells, *J. Biol. Chem.* **272**, 19133–19139 (1997).
44. K. Saito, M. Nishijima and O. Kuge, Genetic evidence that phosphatidylserine synthase II catalyzes the conversion of phosphatidylethanolamine to phosphatidylserine in Chinese hamster ovary cells, *J. Biol. Chem.* **273**, 17199–17205 (1998).
45. O. Kuge, M. Nishijima and Y. Akamatsu, Phosphatidylserine biosynthesis in cultured Chinese hamster ovary cells. III. Genetic evidence for utilization of phosphatidylcholine and phosphatidylethanolamine as precursors, *J. Biol. Chem.* **261**, 5795–5798 (1986).
46. O. Kuge, M. Nishijima and Y. Akamatsu, A Chinese hamster cDNA encoding a protein essential for phosphatidylserine synthase I activity, *J. Biol. Chem.* **266**, 24184–24189 (1991).
47. K. Saito, O. Kuge, Y. Akamatsu and M. Nishijima, Immunochemical identification of the pssA gene product as phosphatidylserine synthase I of Chinese hamster ovary cells, *FEBS Lett.* **395**, 262–266 (1997).
48. S. J. Stone, Z. Cui and J. E. Vance, Cloning and expression of mouse liver phosphatidylserine synthase-1 cDNA: overexpression in rat hepatoma cells inhibits the CDP-ethanolamine pathway for phosphatidylethanolamine biosynthesis, *J. Biol. Chem.* **273**, 7293–7302 (1998).
49. J. Nikawa, Y. Tsukagoshi, T. Kodaki and S. Yamashita, Nucleotide sequence and characterization of the yeast PSS gene encoding phosphatidylserine synthase, *Eur. J. Biochem.* **167**, 7–12 (1987).
50. C. R. H. Raetz, T. J. Larson and W. Dowhan, Gene cloning for the isolation of enzymes of membrane lipid synthesis: phosphatidylserine synthase overproduction in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **74**, 1412–1416 (1977).
51. C. J. Walkey, G. B. Kalmar and R. B. Cornell, Overexpression of rat liver CTP:phosphocholine cytidyltransferase accelerates phosphatidylcholine synthesis and degradation, *J. Biol. Chem.* **269**, 5742–5749 (1994).
52. B. Sturbois-Balcerzak, S. J. Stone, A. Sreenivas and J. E. Vance, Structure and expression of the murine phosphatidylserine synthase-1 gene, *J. Biol. Chem.* **276**, 8205–8212 (2001).
53. S. J. Stone and J. E. Vance, Cloning and expression of murine liver phosphatidylserine synthase (PSS)-2: differential regulation of phospholipid metabolism by PSS1 and PSS2, *Biochem. J.* **342**, 57–64 (1999).
54. J. E. Vance, Eukaryotic lipid-biosynthetic enzymes: the same but not the same, *Trends Biochem. Sci.* **23**, 423–428 (1998).
55. M. O. Bergo, B. J. Gavino, R. Steenbergen, B. Sturbois, A. F. Parlow, D. A. Sanan, W. C. Skarnes, J. E. Vance and S. G. Young, Defining the importance of phosphatidylserine synthase 2 (Ptdss2) in mice, *J. Biol. Chem.* **277**, 47701–47708 (2002).
56. S. J. Meachem, E. Nieschlag and M. Simoni, Inhibin B in male reproduction: pathophysiology and clinical relevance, *Eur. J. Endocrinol.* **145**, 561–571 (2001).
57. J. E. Vance and D. E. Vance, Does rat liver Golgi have the capacity to synthesize phospholipids for lipoprotein secretion?, *J. Biol. Chem.* **263**, 5898–5908 (1988).
58. C. J. Jelsema and D. J. Morré, Distribution of phospholipid biosynthetic enzymes among cell components of rat liver, *J. Biol. Chem.* **253**, 7960–7971 (1978).
59. Y.-J. Shiao, G. Lupo and J. E. Vance, Evidence that phosphatidylserine is imported into mitochondria via a mitochondria-associated membrane and that the majority of mitochondrial phosphatidylethanolamine is derived from decarboxylation of phosphatidylserine, *J. Biol. Chem.* **270**, 11190–11198 (1995).

60. L. M. G. van Golde, J. Raben, J. J. Batenburg, B. Fleischer, F. Zambrano and S. Fleischer, Biosynthesis of lipids in Golgi complex and other subcellular fractions from rat liver, *Biochim. Biophys. Acta* **360**, 179–192 (1974).
61. M. R. Jackson, T. Nilsson and P. A. Peterson, Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum, *EMBO J.* **9**, 3153–3162 (1990).
62. O. Kuge and M. Nishijima, Phosphatidylserine synthase I and II of mammalian cells, *Biochim. Biophys. Acta* **1348**, 151–158 (1997).
63. M.-P. Schutze, P. A. Peterson and M. R. Jackson, An N-terminal double-arginine motif maintains type II membrane proteins in the endoplasmic reticulum, *EMBO J.* **13**, 1696–1705 (1994).
64. J. E. Vance, Phospholipid synthesis in a membrane fraction associated with mitochondria, *J. Biol. Chem.* **265**, 7248–7256 (1990).
65. S. J. Stone and J. E. Vance, Phosphatidylserine synthase-1 and -2 are localized to mitochondria-associated membranes, *J. Biol. Chem.* **275**, 34534–34540 (2000).
66. A. E. Rusñol, Z. Cui, M. H. Chen and J. E. Vance, A unique mitochondria-associated membrane fraction from rat liver has a high capacity for lipid synthesis and contains pre-Golgi secretory proteins including nascent lipoproteins, *J. Biol. Chem.* **269**, 27494–27502 (1994).
67. T. M. Lewin, J. H. Kim, D. A. Granger, J. E. Vance and R. A. Coleman, Acyl-CoA synthetase isoforms 1, 4, and 5 are present in different subcellular membranes in rat liver and can be inhibited independently, *J. Biol. Chem.* **276**, 24674–24679 (2001).
68. J. Vidugiriene, D. K. Sharma, T. K. Smith, N. A. Baumann and A. K. Menon, Segregation of glycosylphosphatidylinositol biosynthetic reactions in a subcompartment of the endoplasmic reticulum, *J. Biol. Chem.* **274**, 15203–15212 (1999).
69. Z. Cui, J. E. Vance, M. H. Chen, D. R. Voelker and D. E. Vance, Cloning and expression of a novel phosphatidylethanolamine N-methyltransferase, *J. Biol. Chem.* **268**, 16655–16663 (1993).
70. D. E. Vance, C. J. Walkey and Z. Cui, Phosphatidylethanolamine N-methyltransferase from liver, *Biochim. Biophys. Acta* **1348**, 142–150 (1997).
71. Y.-J. Shiao, B. Balcerzak and J. E. Vance, A mitochondrial membrane protein is required for translocation of phosphatidylserine from mitochondria-associated membranes to mitochondria, *Biochem. J.* **331**, 217–223 (1998).
72. D. R. Voelker, in: "Lipid Assembly into Cell Membranes," p. 599, Amsterdam, Elsevier Science, 2002.
73. M. A. Carson, K. D. Atkinson and C. J. Waechter, Properties of particulate and solubilized phosphatidylserine synthase activity from *Saccharomyces cerevisiae*. Inhibitory effect of choline in the growth medium, *J. Biol. Chem.* **257**, 8115–8121 (1982).
74. K. Atkinson, S. Fogel and S. A. Henry, Yeast mutant defective in phosphatidylserine synthesis, *J. Biol. Chem.* **255**, 6653–6661 (1980).
75. M. S. Bae-Lee and G. M. Carman, Phosphatidylserine synthesis in *Saccharomyces cerevisiae*. Purification and characterization of membrane-associated phosphatidylserine synthase, *J. Biol. Chem.* **259**, 10857–10862 (1984).
76. M. A. Poole, A. S. Fischl and G. M. Carman, Enzymatic detection of phospholipid biosynthetic enzymes after electroblotting, *J. Bacteriol.* **161**, 772–774 (1985).
77. K. Kuchler, G. Daum and F. Paltauf, Subcellular and submitochondrial localization of phospholipid-synthesizing enzymes in *Saccharomyces cerevisiae*, *J. Bacteriol.* **165**, 901–910 (1986).
78. A. Dutt and W. Dowhan, Intracellular distribution of enzymes of phospholipid metabolism in several gram negative bacteria, *J. Bacteriol.* **132**, 159–165 (1977).
79. J. J. Cousminer, A. S. Fischl and G. M. Carman, Partial purification and properties of phosphatidylserine synthase from *Clostridium perfringens*, *J. Bacteriol.* **151**, 1372–1379 (1982).

80. G. M. Carman and D. S. Wiczorek, Phosphatidylglycerophosphate synthase and phosphatidylserine synthase activities in *Clostridium perfringens*, *J. Bacteriol.* **142**, 262–267 (1980).
81. K. Matsumoto, Phosphatidylserine synthase from bacteria, *Biochim. Biophys. Acta* **1348**, 214–227 (1997).
82. J. D. Saba, F. Nara, A. Bielawska, S. Garrett and Y. A. Hannun, The BST1 gene of *Saccharomyces cerevisiae* is the sphingosine-1-phosphate lyase, *J. Biol. Chem.* **272**, 26087–26090 (1997).
83. Y. A. Hannun, C. Luberto and K. M. Argraves, Enzymes of sphingolipid metabolism: from modular to integrative signaling, *Biochemistry* **40**, 4893–4903 (2001).
84. D. Rontein, I. Nishida, G. Tashiro, K. Yoshioka, W. I. Wu, D. R. Voelker, G. Basset and A. D. Hanson, Plants synthesize ethanolamine by direct decarboxylation of serine using a pyridoxal phosphate enzyme, *J. Biol. Chem.* **276**, 35523–35529 (2001).
85. K. Ishidate, Choline/ethanolamine kinase from mammalian tissues, *Biochim. Biophys. Acta* **1348**, 70–78 (1997).
86. K. Ishidate, K. Iida, K. Tadokoro and Y. Nakazawa, Evidence for the existence of multiple forms of choline (ethanolamine) kinase in rat tissues, *Biochim. Biophys. Acta* **833**, 1–8 (1985).
87. A. Lykidis, J. Wang, M. A. Karim and S. Jackowski, Overexpression of a mammalian ethanolamine-specific kinase accelerates the CDP-ethanolamine pathway, *J. Biol. Chem.* **276**, 2174–2179 (2001).
88. K. Hosaka, S. Tanaka, J. Nikawa and S. Yamashita, Cloning of a human choline kinase cDNA by complementation of the yeast *cki* mutation, *FEBS Lett.* **304**, 229–232 (1992).
89. T. Uchida and S. Yamashita, Molecular cloning, characterization, and expression in *Escherichia coli* of a cDNA encoding mammalian choline kinase, *J. Biol. Chem.* **267**, 10156–10162 (1992).
90. P. Pavlidis, M. Ramaswami and M. A. Tanouye, The *Drosophila* easily shocked gene: a mutation in a phospholipid synthetic pathway causes seizure, neuronal failure and paralysis, *Cell* **79**, 23–33 (1994).
91. K. Hosaka, T. Kodaki and S. Yamashita, Cloning and characterization of the yeast CKI gene encoding choline kinase and its expression in *Escherichia coli*, *J. Biol. Chem.* **264**, 2053–2059 (1989).
92. K. Kim, K. H. Kim, M. K. Storey, D. R. Voelker and G. M. Carman, Isolation and characterization of the *Saccharomyces cerevisiae* EK11 gene encoding ethanolamine kinase, *J. Biol. Chem.* **274**, 14857–14866 (1999).
93. S. Yamashita and K. Hosaka, Choline kinase from yeast, *Biochim. Biophys. Acta* **1348**, 63–69 (1997).
94. K. Hosaka, S. Tanaka, J. Nikawa and S. Yamashita, Cloning of a human choline kinase cDNA by complementation of the yeast *cki* mutation, *FEBS Lett.* **304**, 229–232 (1992).
95. B. A. Bladergroen and L. M. G. van Golde, CTP:phosphoethanolamine cytidyltransferase, *Biochim. Biophys. Acta* **1348**, 91–99 (1997).
96. A. Nakashima, K. Hosaka and J.-I. Nikawa, Cloning of a human cDNA for CTP:phosphoethanolamine cytidyltransferase by complementation in vivo of a yeast mutant, *J. Biol. Chem.* **272**, 9567–9572 (1997).
97. R. Minseok, Y. Kawamata, H. Nakamura, A. Ohta and M. Takagi, Isolation and characterization of ECT1 gene encoding CTP:phosphoethanolamine cytidyltransferase of *Saccharomyces cerevisiae*, *J. Biochem.* **120**, 1040–1047 (1996).
98. A. Lykidis, K. G. Murti and S. Jackowski, Cloning and characterization of a second human CTP:phosphocholine cytidyltransferase, *J. Biol. Chem.* **273**, 14022–14029 (1998).
99. R. H. Hjelmstad and R. M. Bell, sn-1,2-diacylglycerol choline and ethanolaminephosphotransferases in *Saccharomyces cerevisiae*. Nucleotide sequence of the EPT1 gene and comparison of the CPT1 and EPT1 gene products, *J. Biol. Chem.* **266**, 5094–5103 (1991).

100. R. H. Hjelmstad and R. M. Bell, The sn-1,2-diacylglycerol cholinephosphotransferase of *Saccharomyces cerevisiae*. Nucleotide sequence, transcriptional mapping, and gene product analysis of the CTP1 gene, *J. Biol. Chem.* **265**, 1755–1764 (1990).
101. A. L. Henneberry and C. R. McMaster, Cloning and expression of a human choline/ethanolaminephosphotransferase: synthesis of phosphatidylcholine and phosphatidylethanolamine, *Biochem. J.* **339**, 291–298 (1999).
102. A. L. Henneberry, G. Wistow and C. R. McMaster, Cloning, genomic organization, and characterization of a human cholinephosphotransferase [In Process Citation], *J. Biol. Chem.* **275**, 29808–29815 (2000).
103. A. Mancini, F. Del Rosso, R. Roberti, P. Orvietani, L. Coletti and L. Binaglia, Purification of ethanolaminephosphotransferase from bovine liver microsomes, *Biochim. Biophys. Acta* **1437**, 80–92 (1999).
104. J. N. Kanfer and E. P. Kennedy, Metabolism and function of bacterial lipids. II. Biosynthesis of lipids in *Escherichia coli*, *J. Biol. Chem.* **239**, 1720–1726 (1964).
105. W. Dowhan, W. T. Wickner and E. P. Kennedy, Purification and properties of phosphatidylserine decarboxylase from *Escherichia coli*, *J. Biol. Chem.* **249**, 3079–3084 (1974).
106. Q.-X. Li and W. Dowhan, Structural characterization of *Escherichia coli* phosphatidylserine decarboxylase, *J. Biol. Chem.* **263**, 11516–11522 (1988).
107. J. Zborowski, A. Dygas and L. Wojtczak, Phosphatidylserine decarboxylase is located on the external side of the inner mitochondrial membrane, *FEBS Lett.* **157**, 179–182 (1983).
108. D. R. Voelker, Disruption of phosphatidylserine translocation to the mitochondria in baby hamster kidney cells, *J. Biol. Chem.* **260**, 14671–14676 (1985).
109. J. E. Vance, E. J. Aasman and R. Szarka, Brefeldin A does not inhibit the movement of phosphatidylethanolamine from its sites of synthesis to the cell surface, *J. Biol. Chem.* **266**, 8241–8247 (1991).
110. O. Kuge, M. Nishijima and Y. Akamatsu, A cloned gene encoding phosphatidylserine decarboxylase complements the phosphatidylserine biosynthetic defect of a Chinese hamster ovary cell mutant, *J. Biol. Chem.* **266**, 6370–6376 (1991).
111. P. J. Trotter, J. Pedretti, R. Yates and D. R. Voelker, Phosphatidylserine decarboxylase 2 of *Saccharomyces cerevisiae*. Cloning and mapping of the gene, heterologous expression and creation of the null allele, *J. Biol. Chem.* **270**, 6071–6080 (1995).
112. P. J. Trotter and D. R. Voelker, Identification of a non-mitochondrial phosphatidylserine decarboxylase activity (PSD2) in the yeast *Saccharomyces cerevisiae*, *J. Biol. Chem.* **270**, 6062–6070 (1995).
113. E. E. Snell, Pyruvate-containing enzymes, *Trends Biochem. Sci.* **2**, 131–135 (1977).
114. D. R. Voelker, Phosphatidylserine decarboxylase, *Biochim. Biophys. Acta* **1348**, 236–244 (1997).
115. D. R. Voelker, Phosphatidylserine decarboxylase, *Biochim. Biophys. Acta* **1348**, 236–244 (1997).
116. J. N. Kanfer, D. McCartney and H. Hattori, Regulation of the choline, ethanolamine and serine base exchange enzyme activities of rat brain microsomes by phosphorylation and dephosphorylation, *FEBS Lett.* **240**, 101–104 (1988).
117. M. Nishijima, O. Kuge and Y. Akamatsu, Phosphatidylserine biosynthesis in cultured Chinese hamster ovary cells. I. Inhibition of de novo phosphatidylserine biosynthesis by exogenous phosphatidylserine and its efficient incorporation, *J. Biol. Chem.* **261**, 5784–5789 (1986).
118. K. Hasegawa, O. Kuge, M. Nishijima and Y. Akamatsu, Isolation and characterization of a Chinese hamster ovary cell mutant with altered regulation of phosphatidylserine biosynthesis, *J. Biol. Chem.* **264**, 19887–19892 (1989).

119. O. Kuge, K. Hasegawa, K. Saito and M. Nishijima, Control of phosphatidylserine biosynthesis through phosphatidylserine-mediated inhibition of phosphatidylserine synthase I in Chinese hamster ovary cells, *Proc. Natl. Acad. Sci. USA* **95**, 4199–4203 (1998).
120. O. Kuge, K. Saito and M. Nishijima, Control of phosphatidylserine synthase II activity in Chinese hamster ovary cells, *J. Biol. Chem.* **274**, 23844–23849 (1999).
121. M. A. Poole, M. J. Homann, M. S. Bae-Lee and G. M. Carman, Regulation of phosphatidylserine synthase from *Saccharomyces cerevisiae* by phospholipid precursors, *J. Bacteriol.* **168**, 668–672 (1986).
122. T. Kodaki, J. Nikawa, K. Hosaka and S. Yamashita, Functional analysis of the regulatory region of the yeast phosphatidylserine synthase gene, PSS, *J. Bacteriol.* **173**, 7992–7995 (1991).
123. A. M. Bailis, M. A. Poole, G. M. Carman and S. A. Henry, The membrane-associated enzyme phosphatidylserine synthase is regulated at the level of mRNA abundance, *Mol. Cell. Biol.* **7**, 167–176 (1987).
124. M. J. Homann, M. A. Poole, P. M. Gaynor, C. T. Ho and G. M. Carman, Effect of growth phase on phospholipid biosynthesis in *Saccharomyces cerevisiae*, *J. Bacteriol.* **169**, 533–539 (1987).
125. A. J. Kinney and G. M. Carman, Phosphorylation of yeast phosphatidylserine synthase in vivo and in vitro by cyclic AMP-dependent protein kinase, *Proc. Natl. Acad. Sci. USA* **85**, 7962–7966 (1988).
126. D. R. Voelker, Reconstitution of phosphatidylserine import into rat liver mitochondria, *J. Biol. Chem.* **264**, 8019–8025 (1989).
127. D. E. Vance and P. C. Choy, How is phosphatidylcholine biosynthesis regulated?, *Trends Biochem. Sci.* **4**, 145–148 (1979).
128. R. Sundler and B. Akesson, Regulation of phospholipid biosynthesis in isolated rat hepatocytes. Effect of different substrates, *J. Biol. Chem.* **250**, 3359–3367 (1975).
129. L. B. M. Tjburg, M. Houweling, M. J. H. Geelen and L. M. G. van Golde, Stimulation of phosphatidylethanolamine synthesis in isolated rat hepatocytes by phorbol 12-myristate 13-acetate, *Biochim. Biophys. Acta* **922**, 184–190 (1987).
130. S. L. Pelech and D. E. Vance, Regulation of phosphatidylcholine biosynthesis, *Biochim. Biophys. Acta* **779**, 217–251 (1984).
131. C. Kent, Regulation of phosphatidylcholine biosynthesis, *Prog. Lipid Res.* **29**, 87–105 (1990).
132. J. J. van Hellemond, J. W. Slot, M. J. Geelen, L. M. van Golde and P. S. Vermeulen, Ultrastructural localization of CTP:phosphoethanolamine cytidyltransferase in rat liver, *J. Biol. Chem.* **269**, 15415–15418 (1994).
133. P. S. Vermeulen, L. B. M. Tjburg, M. J. H. Geelen and L. M. G. van Golde, Immunological characterization, lipid dependence, and subcellular localization of CTP:phosphoethanolamine cytidyltransferase purified from rat liver, *J. Biol. Chem.* **268**, 7458–7464 (1993).
134. L. B. M. Tjburg, M. Houweling, M. J. H. Geelen and L. M. G. Van Golde, Inhibition of phosphatidylethanolamine synthesis by glucagon in isolated rat hepatocytes, *Biochem. J.* **257**, 645–650 (1989).
135. C. R. McMaster and P. C. Choy, Newly-imported ethanolamine is preferentially utilized for phosphatidylethanolamine biosynthesis in the hamster heart, *Biochim. Biophys. Acta* **1124**, 13–16 (1992).
136. D. R. Voelker, Interorganelle transport of aminoglycerolipids, *Biochim. Biophys. Acta* **1486**, 97–107 (2000).
137. D. R. Voelker, Adriamycin disrupts phosphatidylserine import into the mitochondria of permeabilized CHO-K1 cells, *J. Biol. Chem.* **266**, 12185–12188 (1991).
138. M. Eilers, T. Endo and G. Schatz, Adriamycin, a drug interacting with acidic phospholipids, blocks import of precursor proteins by isolated yeast mitochondria, *J. Biol. Chem.* **264**, 2945–2950 (1989).

139. D. Ardail, F. Lerme and P. Louisot, Involvement of contact sites in phosphatidylserine import into liver mitochondria, *J. Biol. Chem.* **266**, 7978–7981 (1991).
140. D. Ardail, F. Gasnier, F. Lermé, C. Simonot, P. Louisot and O. Gateau-Roesch, Involvement of mitochondrial contact sites in the subcellular compartmentalization of phospholipid biosynthetic enzymes, *J. Biol. Chem.* **268**, 25985–25992 (1993).
141. D. J. Morr  , W. D. Merritt and C. A. Lembi, Connections between mitochondria and endoplasmic reticulum in rat liver and onion stem, *Propolasma* **73**, 43–49 (1971).
142. S. Arnaudeau, W. L. Kelley, J. V., Jr. and Demaurex, N. Walsh, Mitochondria recycle $\text{Ca}(2+)$ to the endoplasmic reticulum and prevent the depletion of neighboring endoplasmic reticulum regions, *J. Biol. Chem.* **276**, 29430–29439 (2001).
143. R. Rizzuto, P. Pinton, W. Carrington, F. S. Fay, K. E. Fogarty, L. M. Lifshitz, R. A. Tuft and T. Pozzan, Close contacts with the endoplasmic reticulum as determinants of mitochondrial $\text{Ca}2+$ responses, *Science* **280**, 1763–1765 (1998).
144. G. C. Shore and J. R. Tata, Two fractions of rough endoplasmic reticulum from rat liver, *J. Cell Biol.* **72**, 714–725 (1977).
145. C. B. Pickett, D. Montisano, D. Eisner and J. Cascarno, The physical association between rat liver mitochondria and rough endoplasmic reticulum, *Exp. Cell Res.* **128**, 343–352 (1980).
146. J. Katz, P. A. Wals, S. Golden and L. Rajman, Mitochondrial-reticular cytostructure in liver cells, *Biochem. J.* **214**, 795–813 (1983).
147. W. W. Franke and J. Kartenbeck, Outer mitochondrial membrane continuous with endoplasmic reticulum, *Protoplasma* **73**, 35–41 (1971).
148. P. J. Meier, M. A. Spycher and U. A. Meyer, Isolation and characterization of rough endoplasmic reticulum associated with mitochondria from normal rat liver, *Biochim. Biophys. Acta* **646**, 283–297 (1981).
149. L. A. Lewis and J. R. Tata, A rapidly sedimenting fraction of rat liver endoplasmic reticulum, *J. Clin. Sci.* **13**, 447–459 (1973).
150. R. Hovius, B. Faber, B. Brigot, K. Nicolay and B. de Kruijff, On the mechanism of the mitochondrial decarboxylation of phosphatidylserine, *J. Biol. Chem.* **267**, 16790–16795 (1992).
151. M. Eilers, T. Endo and G. Schatz, Adriamycin, a drug interacting with acidic phospholipids, blocks import of precursor proteins by isolated yeast mitochondria, *J. Biol. Chem.* **264**, 2945–2950 (1989).
152. D. R. Voelker, Phosphatidylserine translocation to the mitochondrion is an ATP-dependent process in permeabilized animal cells, *Proc. Natl. Acad. Sci. USA* **86**, 9921–9925 (1989).
153. P. J. Trotter, J. Pedretti and D. R. Voelker, Phosphatidylserine decarboxylase from *Saccharomyces cerevisiae*: isolation of mutants, cloning of the gene, and creation of a null allele, *J. Biol. Chem.* **268**, 21416–21424 (1993).
154. J. E. Vance, Newly made phosphatidylserine and phosphatidylethanolamine are preferentially translocated between rat liver mitochondria and endoplasmic reticulum, *J. Biol. Chem.* **266**, 89–97 (1991).
155. D. R. Voelker, The ATP-dependent translocation of phosphatidylserine to the mitochondria is a process that is restricted to the autologous organelle, *J. Biol. Chem.* **268**, 7069–7074 (1993).
156. D. R. Voelker, Characterization of phosphatidylserine synthesis and translocation in permeabilized animal cells, *J. Biol. Chem.* **265**, 14340–14346 (1990).
157. G. Achleitner, D. Zweytick, P. J. Trotter, D. R. Voelker and G. Daum, Synthesis and intracellular transport of aminoglycerophospholipids in permeabilized cells of the yeast, *Saccharomyces cerevisiae*, *J. Biol. Chem.* **270**, 29836–29842 (1995).
158. W. I. Wu and D. R. Voelker, Characterization of phosphatidylserine transport to the locus of phosphatidylserine decarboxylase 2 in permeabilized yeast, *J. Biol. Chem.* **276**, 7114–7121 (2001).

159. G. P. H. van Heusden, K. Bos, C. R. H. Raetz and K. W. A. Wirtz, Chinese hamster ovary cells deficient in peroxisomes lack the nonspecific lipid transfer protein (Sterol Carrier Protein 2), *J. Biol. Chem.* **265**, 4105–4110 (1990).
160. U. Seedorf, P. Brysch, T. Engel, K. Schrage and G. Assmann, Sterol carrier protein X is peroxisomal 3-oxoacyl coenzyme A thiolase with intrinsic sterol carrier and lipid transfer activity, *J. Biol. Chem.* **269**, 21277–21283 (1994).
161. U. Seedorf, M. Raaabe, P. Ellinghaus, F. Kannenberg, M. Fobker, T. Engel, S. Denis, F. Wouters, K. W. A. Wirtz, R. J. A. Wanders, N. Maeda and G. Assman, Defective peroxisomal catabolism of branched fatty acyl coenzyme A in mice lacking the sterol carrier protein-2/sterol carrier protein-X gene function, *Genes Dev.* **12**, 1189–1201 (1998).
162. O. Kuge, Y. Yamakawa and M. Nishijima, Enhancement of transport-dependent decarboxylation of phosphatidylserine by S100B protein in permeabilized Chinese hamster ovary cells, *J. Biol. Chem.* **276**, 23700–23706 (2001).
163. K. S. Bjerve, The biosynthesis of phosphatidylserine and phosphatidylethanolamine from L-[3-14C]serine in isolated rat hepatocytes, *Biochim. Biophys. Acta* **833**, 396–405 (1985).
164. R. Simbeni, L. Pon, E. Zinser, F. Paltauf and G. Daum, Mitochondrial membrane contact sites of yeast. Characterization of lipid components and possible involvement in intra-mitochondrial translocation of phospholipids, *J. Biol. Chem.* **266**, 10047–10049 (1991).
165. B. Gaigg, R. Simbeni, C. Hrastnik, F. Paltauf and G. Daum, Characterization of a microsomal subfraction associated with mitochondria of the yeast, *Saccharomyces cerevisiae*. Involvement in synthesis and import of phospholipids into mitochondria. *Biochim. Biophys. Acta* **1234**, 214–220 (1995).
166. K. Emoto, O. Kuge, M. Nishijima and M. Umeda, Isolation of a Chinese hamster ovary cell mutant defective in intramitochondrial transport of phosphatidylserine, *Proc. Natl. Acad. Sci. USA* **96**, 12400–12405 (1999).
167. M. P. Yaffe and E. P. Kennedy, Intracellular phospholipid movement and the role of phospholipid transfer proteins in animal cells, *Biochemistry* **22**, 1497–1507 (1983).
168. Y. Misumi, Y. Misumi, K. Miki, A. Takatsuki, G. Tamura and Y. Ikehara, Novel blockade by brefeldin A of intracellular transport of secretory proteins in cultured rat hepatocytes, *J. Biol. Chem.* **261**, 11398–11403 (1986).
169. T. Fujiwara, K. Oda, S. Yokota, A. Takatsuki and Y. Ikehara, Brefeldin A causes disassembly of the Golgi complex and accumulation of secretory proteins in the endoplasmic reticulum, *J. Biol. Chem.* **263**, 18545–18552 (1988).
170. O. C. Martin and R. E. Pagano, Transbilayer movement of fluorescent analogs of phosphatidylserine and phosphatidylethanolamine at the plasma membrane of cultured cells. Evidence for a protein mediated and ATP-dependent process (ES), *J. Biol. Chem.* **262**, 5890–5898 (1987).
171. P. J. Trotter, W.-I. Wu, J. Pedretti, R. Yates and D. R. Voelker, A genetic screen for aminophospholipid transport mutants identifies phosphatidylinositol 4-kinase, Stt4p, as an essential component in phosphatidylserine metabolism, *J. Biol. Chem.* **273**, 13189–13196 (1998).
172. R. Birner, M. Burgermeister, R. Schneiter and G. Daum, Roles of phosphatidylethanolamine and of its several biosynthetic pathways in *Saccharomyces cerevisiae*, *Mol. Biol. Cell* **12**, 997–1007 (2001).
173. M. K. Storey, K. L. Clay, T. Kutateladze, R. C. Murphy, M. Overduin and D. R. Voelker, Phosphatidylethanolamine has an essential role in *Saccharomyces cerevisiae* that is independent of its ability to form hexagonal phase structures, *J. Biol. Chem.* **276**, 48539–48548 (2001).
174. R. Simbeni, F. Paltauf and G. Daum, Intramitochondrial transfer of phospholipids in the yeast, *Saccharomyces cerevisiae*, *J. Biol. Chem.* **265**, 281–285 (1990).

175. R. Simbeni, K. Tangemann, M. Schmidt, C. Ceolotto, F. Paltauf and G. Daum, Import of phosphatidylserine into isolated yeast mitochondria, *Biochim. Biophys. Acta* **1145**, 1–7 (1993).
176. G. Achleitner, B. Gaigg, A. Krasser, E. Kainersdorfer, S. D. Kohlwein, A. Perktold, G. Zellnig and G. Daum, Association between the endoplasmic reticulum and mitochondria of yeast facilitates interorganelle transport of phospholipids through membrane contact, *Eur. J. Biochem.* **264**, 545–553 (1999).
177. W.-I. Wu, S. Routt, V. A. Bankaitis and D. R. Voelker, A new gene involved in the transport-dependent metabolism of phosphatidylserine, PSTB2/PDR17, shares sequence similarity with the gene encoding the phosphatidylinositol/phosphatidylcholine transfer protein, SEC14, *J. Biol. Chem.* **275**, 14446–14456 (2000).
178. P. De Camilli, S. D. Emr, P. S. McPherson and P. Novick, Phosphoinositides as regulators in membrane traffic, *Science* **271**, 1533–1539 (1996).
179. M. S. Bretscher, Membrane structure: some general principles, *Science* **181**, 622–629 (1973).
180. F. Basse, J. G. Stout, P. J. Sims and T. Wiedmer, Isolation of an erythrocyte membrane protein that mediates Ca^{2+} -dependent transbilayer movement of phospholipid, *J. Biol. Chem.* **271**, 17205–17210 (1996).
181. P. Williamson, A. Christie, T. Kohlin, R. A. Schlegel, P. Comfurius, M. Harmsma, R. F. A. Zwaal and E. M. Bevers, Phospholipid scramblase activation pathways in lymphocytes, *Biochemistry* **40**, 8065–8072 (2001).
182. G. A. Wurth and A. Zweifach, Evidence that cytosolic calcium increases are not sufficient to stimulate phospholipid scrambling in human T-lymphocytes, *Biochem. J.* **362**, 701–708 (2002).
183. J. Zhao, Q. Zhou, T. Wiedmer and P. J. Sims, Level of expression of phospholipid scramblase regulates induced movement of phosphatidylserine to the cell surface, *J. Biol. Chem.* **273**, 6603–6606 (1998).
184. Q. Zhou, J. Zhao, J. G. Stout, R. A. Luhm, T. Wiedmer and P. J. Sims, Molecular cloning of human plasma membrane phospholipid scramblase: a protein mediating transbilayer movement of plasma membrane phospholipids, *J. Biol. Chem.* **272**, 18240–18244 (1997).
185. J. Zhao, Q. Zhou, T. Wiedmer and P. J. Sims, Palmitoylation of phospholipid scramblase is required for normal function in promoting Ca^{2+} -activated transbilayer movement of membrane phospholipids, *Biochemistry* **37**, 6361–6366 (1998).
186. Q. Zhou, P. J. Sims and T. Wiedmer, Identity of a conserved motif in phospholipid scramblase that is required for Ca^{2+} -accelerated transbilayer movement of membrane phospholipids, *Biochemistry* **37**, 2356–2360 (1998).
187. J. Sun, M. Nanjundan, L. J. Pike, T. Wiedmer and P. J. Sims, Plasma membrane phospholipid scramblase 1 is enriched in lipid rafts and interacts with the epidermal growth factor receptor, *Biochemistry* **41**, 6338–6345 (2002).
188. C. Kunzelmann-Marche, J. M. Freyssinet and M. C. Martinez, Loss of plasma membrane phospholipid asymmetry requires raft integrity. Role of transient receptor potential channels and ERK pathway, *J. Biol. Chem.* **277**, 19876–19881 (2002).
189. T. Wiedmer, Q. Zhou, D. Y. Kwoh and P. J. Sims, Identification of three new members of the phospholipid scramblase gene family, *Biochim. Biophys. Acta* **1467**, 244–253 (2000).
190. S. C. Frasch, P. M. Henson, J. M. Kailey, D. A. Richter, M. S. Janes, V. A. Fadok and D. L. Bratton, Regulation of phospholipid scramblase activity during apoptosis and cell activation by protein kinase C δ , *J. Biol. Chem.* **275**, 23065–23073 (2000).
191. J. Sun, J. Zhao, M. A. Schwartz, J. Y. J. Wang, T. Wiedmer and P. J. Sims, c-Abl tyrosine kinase binds and phosphorylates phospholipid scramblase 1, *J. Biol. Chem.* **276**, 28984–28990 (2001).

192. E. M. Bevers, T. Wiedmer, P. Comfurius, J. Zhao, E. F. Smeets, R. A. Schlegel, A. J. Schroit, H. J. Weiss, P. Williamson, R. F. A. Zwaal and P. J. Sims, The complex of phosphatidylinositol 4,5-bisphosphate and calcium ions is not responsible for the Ca^{++} -induced loss of phospholipid asymmetry in the human erythrocyte: a study in Scott syndrome, a disorder of calcium-induced phospholipid scrambling, *Blood* **86**, 1983–1991 (1995).
193. R. Majumder, G. Weinreb, X. Zhai and B. R. Lentz, Soluble phosphatidylserine triggers assembly in solution of a prothrombin-activating complex in the absence of a membrane surface, *J. Biol. Chem.* **277**, 29765–29773 (2002).
194. J. G. Stout, F. Basse, R. A. Luhm, H. J. Weiss, T. Wiedmer and P. J. Sims, Scott syndrome erythrocytes contain a membrane protein capable of mediating Ca^{2+} dependent transbilayer migration of membrane phospholipids, *J. Clin. Invest.* **99**, 2232–2238 (1997).
195. M. B. Brooks, J. L. Catalfamo, H. A. Brown, P. Ivanova and J. Lovaglio, A hereditary bleeding disorder of dogs caused by a lack of platelet procoagulant activity, *Blood* **99**, 2434–2441 (2002).
196. D. L. Daleke and J. V. Lyles, Identification and purification of aminophospholipid flippases, *Biochim. Biophys. Acta* **1486**, 108–127 (2000).
197. M. Seigneuret and P. F. Devaux, ATP-dependent asymmetric distribution of spin-labeled phospholipids in the erythrocyte membrane: relation to shape changes, *Proc. Natl. Acad. Sci. USA* **81**, 3751–3755 (1984).
198. D. L. Daleke and W. H. Huestis, Incorporation and translocation of aminophospholipids in human erythrocytes, *Biochemistry* **24**, 5406–5416 (1985).
199. A. Zachowski, E. Favre, S. Cribier, P. Hervé and P. F. Devaux, Outside-inside translocation of aminophospholipids in the human erythrocyte membrane is mediated by specific enzyme, *Biochemistry* **25**, 2585–2590 (1986).
200. M. Bitbol and P. F. Devaux, Measurement of outward translocation of phospholipids across human erythrocyte membrane, *Proc. Natl. Acad. Sci. USA* **85**, 6783–6787 (1988).
201. J. Connor, C. Bucana, I. J. Fidler and A. J. Schroit, Differentiation-dependent expression of phosphatidylserine in mammalian plasma membrane: quantitative assessment of outer leaflet lipid by prothrombinase complex formation, *Proc. Natl. Acad. Sci. USA* **86**, 3184–3188 (1989).
202. R. G. Sleight and R. E. Pagano, Transbilayer movement of a fluorescent phosphatidylethanolamine analogue across the plasma membranes of cultured mammalian cells, *J. Biol. Chem.* **260**, 1146–1154 (1985).
203. P. Muller, T. Pomorski, S. Porwoli, R. Tauber and A. Herrmann, Transverse movement of spin-labeled phospholipids in the plasma membrane of a hepatocytic cell line (HepG2): implications for biliary lipid secretion, *Hepatology* **24**, 1497–1503 (1996).
204. M. Julien, J. F. Tournier and J. F. Tocanne, Differences in the transbilayer and lateral motions of fluorescent analogs of phosphatidylcholine and phosphatidylethanolamine in the apical plasma membrane of bovine aortic endothelial cells, *Exp. Cell Res.* **208**, 387–397 (1993).
205. Y. Romsicki and F. J. Sharom, Phospholipid flippase activity of the reconstituted P-glycoprotein multidrug transporter, *Biochemistry* **40**, 6937–6947 (2001).
206. J. Ding, Z. Wu, B. P. Crider, Y. Ma, X. Li, C. Slaughter, L. Gong and X. S. Xie, Identification and functional expression of four isoforms of ATPase II, the putative aminophospholipid translocase. Effect of isoform variation on the ATPase activity and phospholipid specificity, *J. Biol. Chem.* **275**, 23378–23386 (2000).
207. A. Pohl, H. Lage, P. Muller, T. Pomorski and A. Herrmann, Transport of phosphatidylserine via MDR1 (multidrug resistance 1) P-glycoprotein in a human gastric carcinoma cell line, *Biochem. J.* **365**, 259–268 (2002).

208. L. S. Kean, A. M. Grant, C. Angeletti, Y. Mahe, K. Kuchler, R. S. Fuller and J. W. Nichols, Plasma membrane translocation of fluorescent-labeled phosphatidylethanolamine is controlled by transcription regulators, PDR1 and PDR3, *J. Cell Biol.* **138**, 255–270 (1997).
209. A. Siegmund, A. Grant, C. Angeletti, L. Malone, J. W. Nichols and H. K. Rudolph, Loss of Drs2p does not abolish transfer of fluorescence-labeled phospholipids across the plasma membrane of *Saccharomyces cerevisiae*, *J. Biol. Chem.* **273**, 34399–34405 (1998).
210. R. Allikmets, A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy, *Nat. Genet.* **17**, 122 (1997).
211. J. Weng, N. L. Mata, S. M. Azarian, R. T. Tzekov, D. G. Birch and G. H. Travis, Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from hte phenotype in aber knockout mice, *Cell* **98**, 13–23 (1999).
212. Y. Hamon, C. Broccardo, O. Chambenoit, M. F. Luciani, F. Toti, S. Chaslin, J. M. Freyssinet, P. F. Devaux, J. McNeish, D. Marguet and G. Chimini, ABC1 promotes engulfment of apoptotic cells and transbilayer redistribution of phosphatidylserine, *Nat. Cell Biol.* **2**, 399–406 (2000).
213. A. Brooks-Wilson, M. Marcil, S. M. Clee, L. H. Zhang, K. Roomp, M. van Dam, L. Yu, C. Brewer, J. A. Collins, H. O. Mollhuizen *et al.*, Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat. Genet.* **22**, 336–345 (1999).
214. M. Bodzioch, E. Orso, J. Klucken, T. Langmann, A. Bottcher, W. Diederich, W. Drobnik, S. Barlage, C. Buchler, M. Porsch-Ozcurumez, *et al.*, The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease, *Nat. Genet.* **22**, 347–351 (1999).
215. S. Rust, M. Rosier, H. Funke, J. Real, Z. Amoura, J. C. Piette, J. F. Deleuze, H. B. Brewer, N. Duverger, P. Deneffe, *et al.*, Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1, *Nat. Genet.* **22**, 352–355 (1999).
216. J. D. Smith, C. Waelde, A. Horwitz and P. Zheng, Evaluation of the role of phosphatidylserine translocase activity in ABCA1-mediated lipid efflux, *J. Biol. Chem.* **277**, 17797–17803 (2002).
217. W. T. Doerrler, M. C. Reedy and C. R. Raetz, An *Escherichia coli* mutant defective in lipid export, *J. Biol. Chem.* **276**, 11461–11464 (2001).
218. W. T. Doerrler and C. R. Raetz, ATPase activity of the MsbA lipid flippase of *Escherichia coli*, *J. Biol. Chem.* **277**, 36697–36705 (2002).
219. Y. Nishizuka, Studies and perspectives of protein kinase C, *Science* **233**, 305–312 (1986).
220. L. Bittova, R. V. Stahelin and W. Cho, Roles of ionic residues of the C1 domain in protein kinase C- α activation and the origin of phosphatidylserine specificity, *J. Biol. Chem.* **276**, 4218–4226 (2001).
221. K. Emoto, N. Toyamasorimachi, H. Karasuyama, K. Inoue and M. Umeda, Exposure of phosphatidylethanolamine on the surface of apoptotic cells, *Exp. Cell Res.* **232**(2), 430–434 (1997).
222. J. Connor, C. C. Pak and A. J. Schroit, Exposure of phosphatidylserine in the outer leaflet of human red blood cells, *J. Biol. Chem.* **269**, 2399–2404 (1994).
223. A. J. Schroit, J. W. Madsen and Y. Tanaka, In vivo recognition and clearance of red blood cells containing phosphatidylserine in their plasma membranes, *J. Biol. Chem.* **260**, 5131–5138 (1985).
224. Y. Watanabe, A. Shiratsuchi, K. Shimizu, T. Takizawa and Y. Nakanishi, Role of phosphatidylserine exposure and sugar chain desialylation at the surface of influenza virus-infected cells in efficient phagocytosis by macrophages, *J. Biol. Chem.* **277**, 18222–18228 (2002).
225. A. Y. Yu, D. M. Byers, N. D. Ridgway, C. R. McMaster and H. W. Cook, Preferential externalization of newly synthesized phosphatidylserine in apoptotic U937 cells is dependent on caspase-mediated pathways, *Biochim. Biophys. Acta* **1487**, 296–308 (2000).

- 226. D. Chiu, B. Lubin, B. Roelofsen and L. L. van Deenen, Sickled erythrocytes accelerate clotting in vitro: an effect of abnormal membrane lipid asymmetry, *Blood* **58**, 398–401 (1981).
- 227. B. Lubin, D. Chiu, J. Bastacky, B. Roelofsen and L. L. Van Deenen, Abnormalities in membrane phospholipid organization in sickled erythrocytes, *J. Clin. Invest.* **67**, 1643–1649 (1981).
- 228. M. Ciavatti, G. Michel and M. Dechavanne, Platelet phospholipid in stroke, *Clin. Chim. Acta* **84**, 347–351 (1978).
- 229. M. J. Wilson, K. Richter-Lowney and D. L. Daleke, Hyperglycemia induces a loss of phospholipid asymmetry in human erythrocytes, *Biochemistry* **32**, 11302–11310 (1993).
- 230. S. M. van den Eijnde, M. J. van den Hoff, C. P. Reutelingsperger, W. L. van Heerde, M. E. Henfling, C. Vermeij-Keers, B. Schutte, M. Borgers and F. C. Ramaekers, Transient expression of phosphatidylserine at cell-cell contact areas is required for myotube formation, *J. Cell Sci.* **114**, 3631–3642 (2001).
- 231. K. Emoto, T. Kobayashi, A. Yamaji, H. Aizawa, I. Yahara, K. Inoue and M. Umeda, Redistribution of phosphatidylethanolamine at the cleavage furrow of dividing cells during cytokinesis, *Proc. Natl. Acad. Sci. USA* **93**, 12867–12872 (1996).
- 232. K. Emoto and M. Umeda, An essential role for a membrane lipid in cytokinesis: regulation of contractile ring disassembly by redistribution of phosphatidylethanolamine, *J. Cell Biol.* **149**, 1215–1224 (2000).
- 233. A. J. Clark and K. Bloch, The absence of sterol synthesis in insects, *J. Biol. Chem.* **234**, 2578–2588 (1950).
- 234. J. D. Horton, J. L. Goldstein and M. S. Brown, SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver, *J. Clin. Invest.* **109**, 1125–1131 (2002).
- 235. W. Dowhan and M. Bogdanov, in: “Functional Roles of Lipids in Membranes,” pp. 599, Amsterdam, Elsevier Science, 2002.
- 236. A. DeChavigny, P. N. Heacock and W. Dowhan, Sequence and inactivation of the pss gene of *Escherichia coli*. Phosphatidylethanolamine may not be essential for cell viability, *J. Biol. Chem.* **266**, 5323–5332 (1991).
- 237. M. Bogdanov, M. Umeda and W. Dowhan, Phospholipid-assisted refolding of an integral membrane protein. Minimum structural features for phosphatidylethanolamine to act as a molecular chaperone, *J. Biol. Chem.* **274**, 12339–12345 (1999).

This Page Intentionally Left Blank

DNA-Protein Interactions during the Initiation and Termination of Plasmid pT181 Rolling-Circle Replication

SALEEM A. KHAN

*Department of Molecular Genetics and
Biochemistry, University of Pittsburgh
School of Medicine, Pittsburgh, PA
15261, USA*

I. Introduction	114
II. Origins of Replication of the Plasmids of the pT181 Family	116
III. Initiator Proteins of the Plasmid pT181 Family	122
IV. Mechanism of Inactivation of the Plasmid pT181 Family	
Rep Proteins	126
V. Role of Host Proteins in pT181 RC Replication	129
VI. Replication of the Lagging Strand of Plasmid pT181	132
VII. Conclusions and Future Directions	133
Acknowledgments	133
References	133

Initiation of DNA replication requires the generation of a primer at the origin of replication that can be utilized by a DNA polymerase for DNA synthesis. This can be accomplished by several means, including the synthesis of an RNA primer by a DNA primase or RNA polymerase, by nicking of one strand of the DNA to generate a free 3'-OH end that can be used as a primer, and by the utilization of the OH group present in an amino acid such as serine within an initiation protein as a primer. Furthermore, some single-stranded DNA genomes can utilize a snap-back 3'-OH end generated due to self-complementarity as a primer for DNA replication. The different modes of initiation require the generation of highly organized DNA-protein complexes at the origin that trigger the initiation of replication. A large majority of small, multicopy plasmids of Gram-positive bacteria and some of Gram-negative bacteria replicate by a rolling-circle (RC) mechanism (for previous reviews, see Refs. (1-10)). More than 200 rolling-circle replicating (RCR) plasmids have so far been identified and, based on sequence homologies in their replication regions, can be grouped into approximately seven families (Refs. (2,5,9,10), and http://www.essex.ac.uk/bs/staff/osborn/DPR_home.htm). This review will focus on plasmids of the pT181 family that replicate by an RC mechanism. So far, approximately 25 plasmids have been identified as belonging to this family

based on the sequence homology in their double-strand origins (*dsos*) and the genes encoding the initiator (Rep) proteins. This review will highlight our current understanding of the structural features of the origins of replication, and the DNA–protein and protein–protein interactions that result in the generation of a replication–initiation complex that triggers replication. It will discuss the molecular events that result in the precise termination of replication once the leading-strand DNA synthesis has been completed. This review will also discuss the various biochemical activities of the initiator proteins encoded by the plasmids of the pT181 family and the mechanism of inactivation of the Rep activity after supporting one round of leading-strand replication. Finally, the review will outline the mechanism of replication of the lagging strand of the pT181 plasmid as well as the limited information that is available on the role of host proteins in pT181 leading- and lagging-strand replication. © 2003 Elsevier Science

Keywords: Plasmids; Rolling-circle replication; DNA–protein interactions; Origins of replication; Initiator proteins; DNA helicase.

Abbreviations

RC	rolling-circle
RCR	rolling-circle replicating
ds	double stranded
ss	single stranded
SC	supercoiled
<i>dso</i>	double-strand origin
<i>sso</i>	single-strand origin

I. Introduction

Our current understanding of the steps involved in the rolling-circle (RC) replication of plasmids of the pT181 family is summarized in Fig. 1 (for previous reviews, see Refs. (1–10)). The pT181 DNA contains two important *cis*-acting regions required for stable plasmid replication. The double-strand origin (*dso*) is required for leading-strand synthesis, while the single-strand origin (*sso*) plays a role in the conversion of the ssDNA released after the leading-strand synthesis to a double-stranded (ds) form. The *dso*, which includes sequences that promote DNA bending and cruciform extrusion, is present in a unique structural configuration. The plasmid-encoded initiator protein, Rep, binds to the *dso* through a sequence-specific interaction and induces further structural changes at the origin. The Rep protein, which acts as a dimer, then recruits the PcrA helicase to the *dso* through a specific protein–protein interaction. This is followed by nicking of

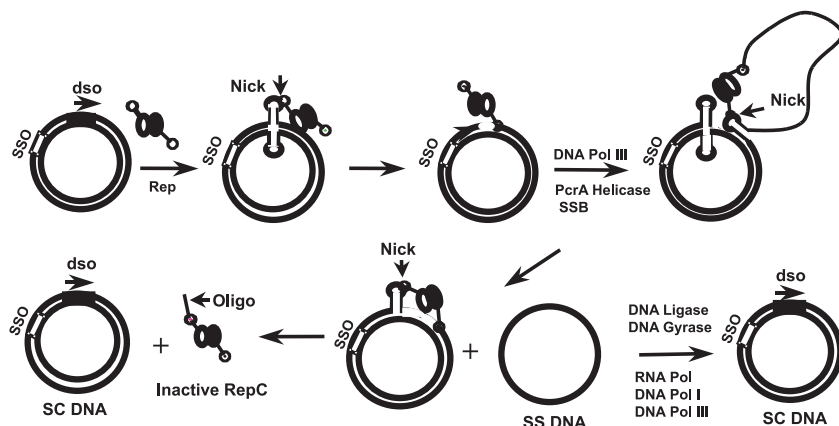


FIG. 1. A model for the rolling-circle replication of the plasmids of the pT181 family. Details are described in the text. *dso*, double-strand origin; *sso*, single-strand origin; SC, supercoiled; SSB, single-strand DNA-binding protein.

the *dso* by the Rep and its covalent attachment to the 5' phosphate of the DNA through a phosphotyrosine bond involving the active tyrosine residue of one monomer of the Rep protein (11,12). This is followed by unwinding of the DNA by PcrA, binding of the single-strand DNA-binding protein (SSB) to the displaced leading strand, and utilization of the Rep-generated nick by the DNA polymerase III as a primer for leading-strand DNA synthesis. Once the replication fork reaches the termination site, i.e., the regenerated *dso*, DNA synthesis proceeds to approximately 10 nucleotides (nt) beyond the Rep nick site (Fig. 1). The second, free monomer of the Rep protein then cleaves the displaced ssDNA. A series of relegation and cleavage events follow, resulting in the release of a circular dsDNA containing the newly synthesized leading strand and an ssDNA representing the displaced leading strand. The dsDNA is then supercoiled (SC) by the host DNA gyrase. The Rep protein in which the active tyrosine of one monomer is covalently attached to the 10-mer oligonucleotide is released (Fig. 1). This form of Rep, termed RepC/RepC* for pT181, which has catalyzed one round of leading-strand replication is inactive in further replication. The displaced ssDNA is converted to the ds form utilizing solely the host proteins. This ss \rightarrow ds conversion, termed lagging-strand synthesis, initiates from the plasmid *sso* that is strand- and orientation-specific and contains extensive secondary structure. The various steps involved in the initiation and termination of pT181 replication will be discussed in the context of the model shown in Fig. 1.

II. Origins of Replication of the Plasmids of the pT181 Family

A. Domains of the Double-Strand Origins

The leading- or double-strand origin, *dso*, of rolling-circle replicating (RCR) plasmids contains the sequence that is recognized by the plasmid-encoded initiator protein and includes the start site of DNA synthesis. The *dsos* are usually less than 100 base pair (bp) in size and contain regions that are required for both the initiation and termination of plasmid RC replication. The *dsos* usually contain structural features that facilitate DNA unwinding, DNA bending, and the generation of hairpin structures that promote extrusion of *dsos* as a cruciform. Presumably, these structural features are critical for recruiting essential proteins to the origin and the formation of a replication-initiation complex. The *dsos* of RCR plasmids usually contain two discrete domains that are critical for replication: the *bind* sequence involved in specific interaction with the plasmid-encoded Rep protein and *nick*, containing the Rep nick site (Fig. 1). Although the *bind* sequences of RCR plasmids belonging to a particular family share homology, they interact stably with only their cognate Rep proteins (9,13–15). On the other hand, the nick sequences are identical or very similar in all plasmids belonging to the same family (5). Thus, the Rep proteins encoded by different plasmids of the same family can nick-close the *dsos* of all the plasmids belonging to the same family (11,16–18). The pT181 family contains approximately 25 plasmids, including pT181, pC221, and pS194 (9). Plasmid pT181 comprises 4440 bp and contains 69.8% AT and 30.2% GC pairs (19). In comparison with the rest of the plasmid, the pT181 *dso* is relatively GC rich. The optimal *dso* of pT181 comprises approximately 70 bp and is 40% GC, whereas the minimal *dso* comprises 40 bp and contains 52% GC nucleotides (20,21). The *dso* of the pT181 plasmid contains three inverted repeat (IR) elements, IRI, IRII, and IRIII (Fig. 2). Deletion of IRI has minimal or no effect on *dso* function, and therefore does not appear to be important in replication (21). The IRII and IRIII elements on the other hand are critical for the *dso* function. Footprinting studies have shown that the RepC protein of pT181 interacts with an approximately 32-bp *bind* region corresponding to pT181 nt 36–68 (Fig. 2 and Refs. (22,23)). Within this region, the major RepC-binding site corresponds to nt 37–60 that includes IRIII. The *bind* sequence is specific for individual members of the pT181 family and is responsible for highly specific interaction between their *dsos* and cognate initiator proteins (Fig. 3). It has been shown that exchanging the 17-bp IRIII sequence between the *dsos* of the plasmids of the pT181 family switches their Rep binding and replication specificities (24). Generally, there is little cross-interaction between the *bind* sequence and the

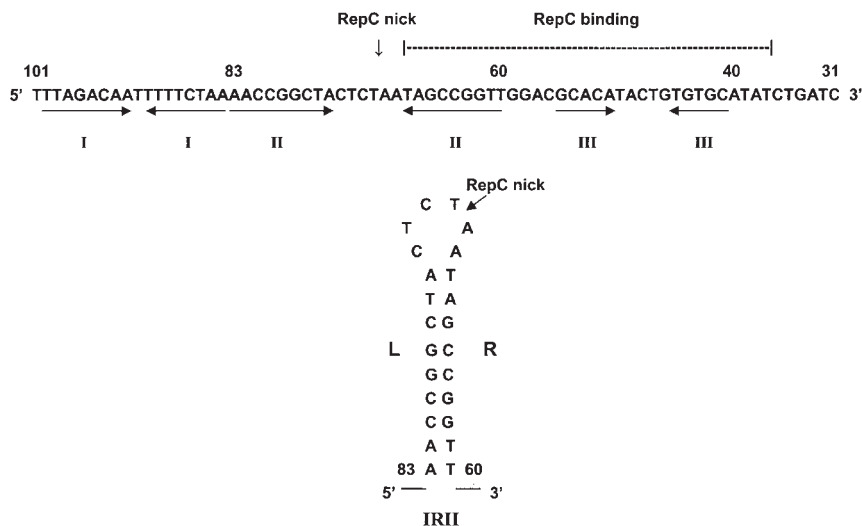


FIG. 2. Nucleotide sequence of the plasmid pT181 origin of replication containing the three inverted repeat elements, IRI, IRII, and IRIII. Only the bottom strand of the DNA containing the RepC nick site is shown. Replication of pT181 proceeds in a rightward direction. The predicted structure of the IRII region is also shown. L and R represent the left and right arms of IRII, respectively.

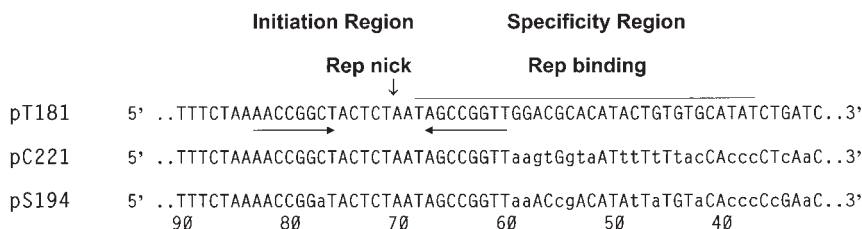


FIG. 3. Sequence alignment of the replication origins of the pT181, pC221, and pS194 plasmids. Numbering corresponds to the pT181 sequence. Nucleotides in pC221 and pS194, which are identical to pT181, are shown in capitals. The initiation region containing the RepC nick site is almost identical, while the specificity region containing the specific initiator-binding site (overlined) is less well conserved. The inverted arrows represent the two arms of the IRII structure.

heterologous Rep proteins belonging to plasmids of the pT181 family (16,25). However, in the absence of the cognate Rep protein, the *bind* sequences can interact weakly with heterologous Rep proteins provided they are over-expressed, thereby resulting in limited cross-replication. The *dsos* of pT181 and related plasmids contains a static bend and this bending is enhanced upon

binding of the Rep protein (26). The *bind* region of the pT181 *dso* also includes a 21-bp region containing alternating pyrimidine–purine nucleotides (nt 37–57 in Fig. 2) that can potentially result in the generation of localized z-DNA structure (22). Furthermore, the IRII region of the *dso* is extruded as a cruciform in SC pT181 DNA upon Rep binding (13). These structural features of the *dso* play an important role in their recognition and nicking by the Rep, and possibly interaction with host proteins that promote the initiation of replication.

The *bind* and *nick* sequences of the plasmids of the pT181 family are located adjacent to each other (Figs. 2 and 3, and Refs. (22,27)). The *nick* region comprises a 24-bp element termed IRII. IRII contains a stem that comprises 9 bp and a 6-nt long loop (Fig. 2). The IRII sequence is identical or nearly identical in all the plasmids of the pT181 family. The Rep nick site is located in the loop of IRII and the initiators encoded by plasmid pT181 family members can nick-close all such plasmids. However, as discussed later, DNA nicking is not sufficient for the initiation of replication, which also requires stable noncovalent Rep–*dso* interactions that are plasmid specific. The Rep protein binds to the *bind* region noncovalently in a stable manner and generates a nick at the adjacent *nick* sequence during the initiation step (18,22,27). The spacing between the IRII and IRIII regions is critical since insertion of sequences that alter this arrangement result in inactivation of the *dso* function (24). This indicates that the Rep proteins are positioned on the *dso* in a specific configuration, and any changes in the overall structure of the *dso* interfere with the assembly of a replication–initiation complex.

Both *in vivo* and *in vitro* studies have shown that binding of the Rep protein to the *dso* melts the IRII region (13,14,23). This results in the exposure of the Rep nick site in an ss form in the loop of IRII (Fig. 2). As stated earlier, the IRII region is extruded as a cruciform in SC pT181 DNA upon Rep binding (13). While this event plays an important role, it is not absolutely essential for the initiation of replication. Mutational analyses have shown that removal/alteration of the left arm of IRII distal to the Rep-binding site (Fig. 2), which is expected to disrupt the IRII stem and cruciform formation, does not abolish initiation (14). However, plasmids with point mutations in the Rep-binding sequence that also lack the ability to extrude cruciform are inactive in replication (14). The presence of wild-type (wt) IRII that can be extruded by Rep binding in these Rep-binding mutants allows initiation to occur. Thus, cruciform extrusion is important when Rep binding to the IRIII region is suboptimal. These studies show that there is some redundancy in the *dso* function, and the loss of some elements can be tolerated.

As discussed earlier, the *dso* has a modular structure containing both Rep *bind* and *nick* sequences. Both of these regions are required for the initiation

of leading-strand replication. Mutant *dsos* containing only the RepC nick sequences are nicked by the initiator but are unable to support replication (21,28,29). Similarly, mutant *dsos* that lack the Rep nick site are inactive in initiation. It is likely that for productive initiation to occur, Rep must first bind stably to the *dso* and recruit host factors such as the PcrA helicase and SSB. Subsequently, nicking at the IRII region of the *dso* is expected to trigger the initiation of replication. In the absence of stable binding, the Rep proteins may transiently nick-close the DNA *in vitro* but this does not result in the assembly of a stable replication-initiation complex (11). Several studies have shown that distinct domains of the *dso* are involved in the initiation and termination steps. While initiation requires both the *bind* and *nick* sequences, the nick region alone is sufficient for the termination of replication, both *in vivo* and *in vitro* (28–30). As summarized in Fig. 1, the Rep protein nicks SC pT181 DNA during initiation. Furthermore, initiation requires the assembly of a highly specific DNA-protein complex that includes the *dso*, Rep, PcrA helicase, and other host proteins. The full *dso* sequence including both the *bind* and *nick* regions are required for this event (28,30). On the other hand, the Rep protein is already bound to the DNA and is presumably present as a component of the replisome as the termination site (i.e., the regenerated *dso*) is approached (Fig. 1). Thus, Rep can recognize the IRII region alone and cleave the DNA even in the absence of its specific *bind* sequence. This postulate is consistent with the results of electrophoretic mobility shift assays (EMSA) showing that Rep can bind weakly to IRII, and this interaction appears to be sufficient for the termination of replication (31).

B. Initiation and Termination of Plasmid pT181 Replication

The sequence requirements for the initiation and termination of pT181 leading-strand replication have been studied utilizing recombinant plasmids containing two copies of an asymmetrically placed *dso* sequence (29,30,32). Replication of a plasmid containing two copies of the wt *dso* *in vivo* results in the generation of two miniplasmids representing initiation and termination at either of the two *dsos* (29,30). *In vitro* replication of such plasmids in the presence of rifampicin (which blocks lagging-strand synthesis) generates two ssDNAs representing initiation and termination at each of the two *dsos* (29,32). No unit-length ssDNA is usually observed demonstrating that termination at the wt *dso* is highly efficient. Plasmids containing one wt and one mutant *dso* were used to study the sequence requirements for the initiation and termination of pT181 replication (28,29). These studies showed that both the Rep binding and nicking sequences were required for appropriate initiation to occur since this step requires stable Rep-*dso* interaction as well as *dso* nicking. Fig. 4 summarizes the results of studies using various IRII derivatives to identify the sequence requirements for the

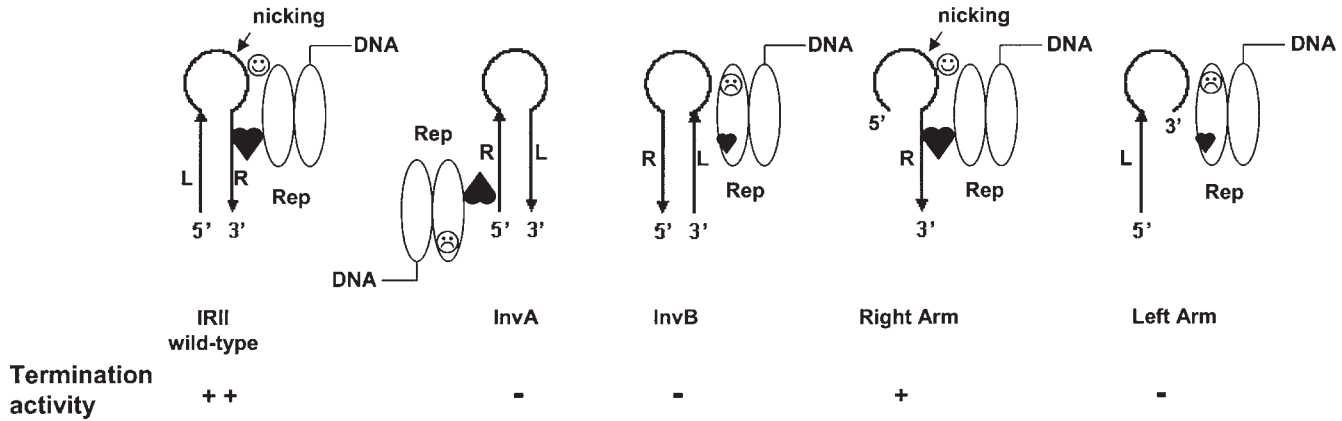


FIG. 4. Activity of IRII and its derivatives in the termination of replication of pT181-type plasmids. Plasmids containing one copy of the wt origin and one copy of IRII or its derivatives were used to test their termination activity *in vivo* and *in vitro* (29). Only the leading (bottom) strand of the DNA is shown and the replication fork approaches from the left. L and R represent the left and right arms of IRII. One monomer of the Rep protein is attached to the displaced parental leading strand. ☺ indicates the DNA-binding domain of Rep, while ☹ corresponds to its nicking domain. Both of these domains have to be properly oriented for a sequence-specific interaction between Rep and the right arm of IRII, and nicking in the loop of IRII. In InvA, the left and right arms of IRII have been switched, while in InvB the sequence as well as the polarity of the L and R arms has been switched (29). The Right Arm and Left Arm constructs also contain the loop sequence.

termination of pT181 replication. Interestingly, IRII by itself was sufficient to promote termination, and several IRII mutants including a smaller 16-bp subregion of IRII representing pT181 nt 78–62 (Fig. 2) were capable of promoting termination (28,29). A sequence containing the right arm and loop of IRII also supported termination (29). Modified IRII derivatives containing the normal loop sequence, but in which either the sequences of left and right arms were switched or their polarity was reversed, were inactive in termination (29). These studies revealed that the IRII structure is not absolutely essential for the termination of replication. Furthermore, the right arm and loop sequences including the RepC nick site were essential for termination (29). These data are consistent with the observation that the Rep protein interacts with the right arm sequence, albeit weakly (31). Another conclusion that can be drawn from these studies is that the Rep protein has a specific orientation with respect to the IRII region, and the right arm sequence must be present downstream of the nick site for the arrest of the replication fork followed by the termination of replication (Fig. 4). Studies have shown that the leading-strand synthesis of pT181 proceeds beyond the regenerated RepC nick site until the full IRII sequence has been synthesized (up to pT181 position 60, Fig. 2) (33). This results in the synthesis of a unit-length leading strand plus approximately 11 extra nucleotides representing the sequences downstream of the Rep nick site to the bottom of IRII. It is postulated that after the right arm of IRII has been synthesized, its interaction with RepC may block the movement of the replication fork and displace the PcrA helicase from the DNA. Following this, RepC would cleave at the regenerated nick site present 11-nt upstream of the 3' end of the newly synthesized DNA (Fig. 1). The earlier studies show that while initiation of pT181 RC replication requires the generation of a highly specific DNA-protein complex involving the full *dso* sequence, requirements for termination are less stringent and a relatively small subregion of the *dso* containing the Rep nick sequence is recognized as an effective signal for termination. As discussed earlier, the IRII region of pT181 is conserved in all the plasmids of this family (5,9). Thus, the termination sequence for all the plasmids of the pT181 family is identical and therefore not plasmid specific. In contrast, the *bind* sequences that are critical for initiation are specific to individual plasmids of the pT181 family (5,9).

An absolute requirement for termination by the Rep protein is its ability to cleave the displaced leading-strand DNA containing the regenerated nick sequence (Fig. 1). However, as discussed later, the efficiency of cleavage by the Rep is not directly related to the efficiency of termination, i.e., sequences that are cleaved efficiently by the Rep do not necessarily have a greater termination activity. Rather, it is likely that the efficiency at which the

replication fork is stalled as well as the efficiency of ligation of the displaced ssDNA may determine the extent of termination.

C. Enhancer of pT181 Replication

The pT181 plasmid contains an enhancer of replication, termed *cmp*, which appears to stimulate replication (34–36). The *cmp* region comprises approximately 100 bp and is located 1 kilobase upstream of the pT181 *dso* (36). The *cmp* region has been shown to contain an intrinsic bent in the DNA (36). Deletion of the *cmp* region does not impair replication in an otherwise wt plasmid (37). However, a plasmid deleted in *cmp* competes poorly in replication with a wt plasmid (37). The Rep proteins encoded by the plasmids of the pT181 family are specific for supporting replication from their cognate origins (25). However, when overexpressed, they can recognize the *dso* of other plasmids of the family and support replication (25). This recognition is dependent upon the presence of the *cmp* locus in *cis* on the plasmid containing the heterologous *dso* (37). The *cmp* region is bound by a host protein, CBF1, which appears to distort the bent DNA in the *cmp* element (38). Taken together, the earlier studies support a role for *cmp* in RepC–*dso* interaction. It is possible that interaction between CBF1 and *cmp* promotes a *dso* conformation that is utilized more efficiently for the initiation of replication.

III. Initiator Proteins of the Plasmid pT181 Family

The RepC protein of pT181 and RepD protein of pC221 have been purified and extensively studied for their biochemical properties. These proteins have origin-binding and origin-nicking activities, and they also promote replication of their cognate plasmids in the presence of cell-free extracts (39,40). The RepC initiator protein encoded by the pT181 plasmid comprises 314 amino acids, and the initiators encoded by other members of this plasmid family also have a very similar size (39). An alignment of the Rep proteins of the pT181 family reveals that in pair-wise comparison they generally have approximately 75–85% homology (9). The two highly divergent regions of these initiators correspond to the amino-terminal end of Rep that is encoded by the *dso* and the carboxyl-terminal region that contains the sequence-specific DNA-binding domain of the Rep proteins (41,42). The domain structure of the initiator proteins of the pT181 family is depicted in Fig. 5. The central regions of the initiator proteins contain their nicking domain that includes the catalytically active tyrosine residue. An alignment of the Rep proteins of the pT181 family shows that their nicking domains containing the active tyrosine residue are nearly identical (Fig. 6).

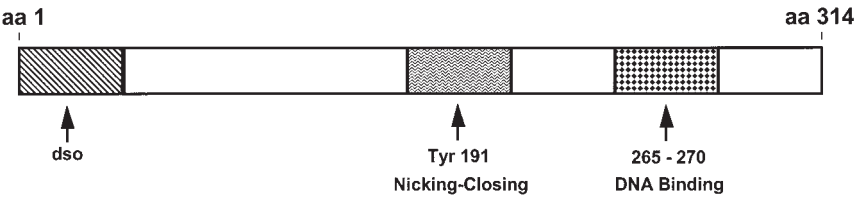


FIG. 5. Domain structure of the Rep proteins of the pT181 family. Amino acid numbering corresponds to that of the RepC protein of pT181. The amino-terminal end of the Rep proteins (encoded by the *dso*) and the carboxyl-terminal region containing the sequence-specific DNA-binding domain is much less conserved as compared to the rest of the protein.

	185						191									200			203
	
pT181	N	R	F	I	R	I	<u>Y</u>	N	K	K	Q	E	R	K	D	N	A	D	A
pC221	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V
pC223	-	-	-	-	-	-	-	-	-	K	-	-	-	E	-	-	-	-	V
pS194	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I
pUB112	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V

FIG. 6. Amino acid sequence of the region containing the active sites of the Rep proteins encoded by the pT181, pC221, pC223, pS194, and pUB112 plasmids. The numbering corresponds to that of the RepC protein of pT181. Tyrosine at position 191 of RepC represents the active tyrosine residue.

The divergent carboxyl-terminal regions of the Rep proteins contain their sequence-specific DNA-binding domain. This domain of the Rep proteins is much less homologous amongst the plasmid pT181 family Rep proteins (Fig. 7) and is responsible for promoting sequence-specific binding and replication from their cognate origins (41–43). The biochemical activities of the plasmid pT181 family initiator proteins are described in the next section.

A. Biochemical Activities of the Rep Proteins

1. SEQUENCE-SPECIFIC DNA-BINDING ACTIVITY OF THE INITIATOR PROTEINS

Although the Rep proteins of the pT181 family are quite homologous, there is considerable variation in their amino acid sequence near the carboxyl-terminal ends (Fig. 7). Studies have shown that this region of the initiator

	250	260	270	280
pT181 K S	I V F M L L S D E E E W G K L H R N S R T K Y K N L I K E I S P V D L T D L M			
pC221 - -	M - Y L - - N E - G T - - - - E - H A K Y - - - Q - - - - - I - - - E - -			
pC223 - -	M - Y L - - H E - S K - - E - - - - - R - - - Q I - Q - - - S I - - - - -			
pS194 - L	M I Y - - I H E - S T - - - - E - R T K N - - R E M L - S - - E I - - - - -			
pUB112	M - Y T - - H E - S M - - - - S K - T K - - F - K M - R - - - - I - - - E - - - -			

FIG. 7. Amino acid sequences of the carboxyl-terminal regions (DNA-binding domain) of the initiator proteins encoded by various plasmid pT181 family members. The top line corresponds to the amino acid sequence of RepC encoded by pT181. Amino acids that are conserved in all the other Rep proteins are indicated by dashes. Amino acids that are not totally conserved in all the proteins are indicated by the single letter code. Amino acids corresponding to positions 265–270 are critical for sequence-specific DNA binding.

proteins contains their origin-specific binding domains (17,41,42). Mutational analyses have shown that amino acids 267–270 of RepC are critical for its sequence-specific DNA binding and replication activities (41,42). Amino acids in the Rep proteins of the pT181 family corresponding to positions 265–270 of RepC are highly divergent (Fig. 7). It has been shown that exchanging these six amino acids between the Rep proteins of the pT181 family switches their sequence-specific DNA binding and replication specificities (41,42). Thus, a RepC derivative containing the corresponding six amino acids from RepD binds to and replicates pC221 DNA instead of pT181 and vice versa (41,42). A synthetic 25-amino acid peptide centered on positions 265–270 of RepC is unable to bind to the DNA, suggesting that the DNA-binding domain corresponds to a larger region of the initiator protein (unpublished data). The availability of the three-dimensional structures of the Rep proteins in the future should be valuable for a better understanding of their biological properties.

2. DNA CLEAVAGE AND LIGATION ACTIVITIES OF THE INITIATOR PROTEINS

Using purified proteins, it has been shown that RepC and RepD can nick-close several plasmids belonging to the pT181 family quite efficiently (16,17). This is due to the fact that the nicking domains of the initiators of the pT181 family and the corresponding *nick* sequences in the plasmid *dsos* are identical (5). Tyr-191 of RepC and Tyr-188 of the RepD protein have been shown to be involved in plasmid nicking (11,12). Mutants of RepC and RepD, in which these active tyrosine residues have been changed to phenylalanine or serine, are inactive in *dso* nicking and replication (11). Furthermore, Tyr-188 of RepD has been directly demonstrated to be covalently attached to the 5' phosphate of the DNA at the nick site (12). All Rep proteins encoded by

the plasmids of the pT181 family contain a tyrosine residue at a similar position (Fig. 6).

In vitro, nicking of SC pT181 DNA by RepC results in the generation of a series of covalently closed topoisomers resembling the products generated by DNA topoisomerase I (27). The sequence requirements for nicking-closing by the RepC and RepD proteins have been investigated in detail. Recombinant plasmids containing the pT181 *dso* that includes both IRII and IRIII are efficiently relaxed by the RepC protein (28). Plasmids containing only the IRII sequence are also relaxed by RepC but this occurs at considerably reduced efficiency (28). Thus, the Rep-binding sequence (IRIII) present adjacent to the IRII region contributes to efficient DNA nicking by the initiator. This is presumably due to the fact that Rep binds stably to IRIII, an event that is expected to facilitate cleavage at IRII. Binding of the Rep protein to IRIII also promotes melting/extrusion of IRII, thus exposing the Rep nick site in the loop. Studies with the RepD protein have also shown that it nicks the cognate pC221 plasmid at a much higher rate than the heterologous pT181 plasmid (17). This is consistent with the fact that the IRIII sequences are specific and allow stable interaction only with their cognate initiator proteins.

The nicking of SC plasmid DNA by the Rep protein generates a free 3'-OH end that is utilized as a primer for DNA replication (Fig. 1). However, the Rep proteins must catalyze additional cleavage/ligation events during the termination step (Fig. 1). During termination, the Rep protein must cleave the displaced, parental leading strand that is present in an ss form (Fig. 1). The Rep protein must also cleave the newly synthesized strand after the leading-strand synthesis has been completed. *In vitro* studies have shown that the RepC and RepD proteins can efficiently cleave ss oligonucleotides containing the Rep nicking sequence located in the bottom strand of the plasmid DNAs (Fig. 2). Single-stranded oligonucleotides (24 mers) representing the bottom strand of IRII are efficiently cleaved by the RepC and RepD proteins (29,44). Subregions of IRII containing 16 nt are also cleaved by the Rep proteins. As with SC DNA, after nicking the Rep proteins become covalently attached to the 5' phosphate of the downstream oligonucleotide (31,45,46). Using mutant oligonucleotides it has also been shown that the loop sequence of IRII by itself is not sufficient for cleavage by Rep (31). Further studies showed that oligonucleotides containing the right (downstream) arm of IRII (Fig. 2) along with the loop sequence are efficiently cleaved by the RepC and RepD proteins (29). This is consistent with the observation that RepC weakly interacts with the right arm of IRII but not with the sequences in the left arm (22,31). The ability of the Rep proteins to cleave ss DNA is consistent with the cleavage reactions; they must perform during the termination of RC replication.

In addition to DNA cleavage during termination, the Rep proteins must also catalyze additional cleavage and rejoining events that result in the release

of a circularized leading strand and a dsDNA containing the newly synthesized leading strand (Fig. 8). The religation activities of the RepC and RepD proteins have also been investigated in considerable detail. These studies utilized derivatives of the Rep proteins, termed RepC* or RepD*, containing an oligonucleotide attached to their active tyrosine residues (15,31,45). Cleavage of an oligonucleotide representing the bottom strand of IRII results in the covalent attachment of an 11-nt long sequence (pT181 nt 70–60) to the active tyrosine residue of the Rep proteins. Since the Rep proteins act as dimers, this generates a derivative protein, termed RepC/RepC* or RepD/RepD*, in which one monomer is normal while the other monomer is attached to the oligonucleotide (Fig. 9A). Oligonucleotides representing sequences present upstream of the Rep nick site (nt 83–71) were incubated with the RepC/RepC* and RepD/RepD* proteins to study the religation properties of these initiator derivatives. These studies showed that an oligonucleotide containing the left arm of IRII and the loop sequence up to the nick site (nt 83–71) can be efficiently religated to the oligonucleotide (position 70–60) attached to the Rep protein (15,31,45). This results in the release of the regenerated IRII sequence along with the active RepC/RepC or RepD/RepD homodimers as depicted in Fig. 9B. This can be demonstrated by the ability of RepC, recovered after religation, to efficiently nick-close SC plasmid DNA (31). Further studies showed that a 5-mer oligonucleotide (nt 75–70) was also religated by the RepC/RepC* protein, although at a lower efficiency (31). These studies demonstrated that the left arm sequence that is complementary to the right arm covalently attached to the RepC/RepC* protein stimulates its religation activity. Thus, religation during termination may involve a pairing of the complementary left and right arm sequences of IRII such that the 5' and 3' ends of the DNA to be joined are in close proximity. The results of the cleavage and religation activities of Rep are also consistent with the ability of the IRII region to promote the termination of plasmid replication (28–30). We have shown that sequences that were both cleaved and religated by the RepC protein promoted termination to a higher extent than sequences that were efficiently cleaved, but poorly ligated, by the Rep (Ref. (29), and Zhao and Khan, unpublished data).

IV. Mechanism of Inactivation of the Plasmid pT181 Family Rep Proteins

As discussed earlier, the initiator proteins of the plasmid pT181 family are known to be inactivated after supporting one round of replication by the attachment of a ~11-nt long sequence to their active tyrosine residues

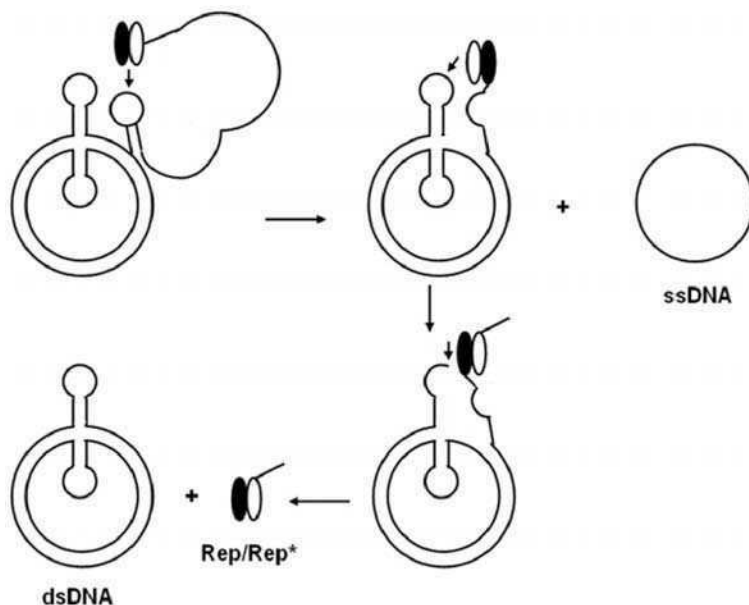


FIG. 8. DNA cleavage and religation events catalyzed by the Rep protein during the termination of rolling-circle replication. Once the replication fork reaches the termination site, i.e., the regenerated *dso*, DNA synthesis proceeds to approximately 10 nt beyond the Rep nick site. At this stage, the first (unfilled) monomer is covalently attached to the 5'-phosphate end of the displaced, parental leading strand. The second (filled) monomer cleaves the regenerated IRII in the displaced leading strand and becomes covalently attached to its 5'-phosphate end. The first monomer subsequently catalyzes a ligation event that results in the release of the circular ssDNA (leading strand). The free, first monomer of Rep subsequently nicks the regenerated *dso* present in the dsDNA generating a free 3'-OH end and becomes covalently attached to the 5' phosphate of the 10-nt long sequence present immediately downstream of the Rep nick site. The second (filled) monomer bound to the newly synthesized strand then catalyzes the religation of this strand generating covalently closed dsDNA. The inactive Rep/Rep* protein is released.

(Fig. 8 and Refs. (33,47,48)). The replication of the plasmids of the pT181 family is regulated by the levels of the Rep protein synthesized in the cell (8,46). The copy number of pT181 is approximately 22 per cell, and studies suggest that approximately one dimer of RepC is synthesized per plasmid copy (8,46). Copy mutants of pT181 synthesize correspondingly higher levels of the RepC protein and, in general, increasing the levels of RepC in cells results in a corresponding increase in plasmid copy number. The earlier data are consistent with the likely possibility that one Rep dimer is utilized per round of leading-strand replication (lagging-strand synthesis does not require the Rep protein). The earlier observations suggest that the Rep protein must be inactivated after it has been utilized in plasmid replication. This was shown to

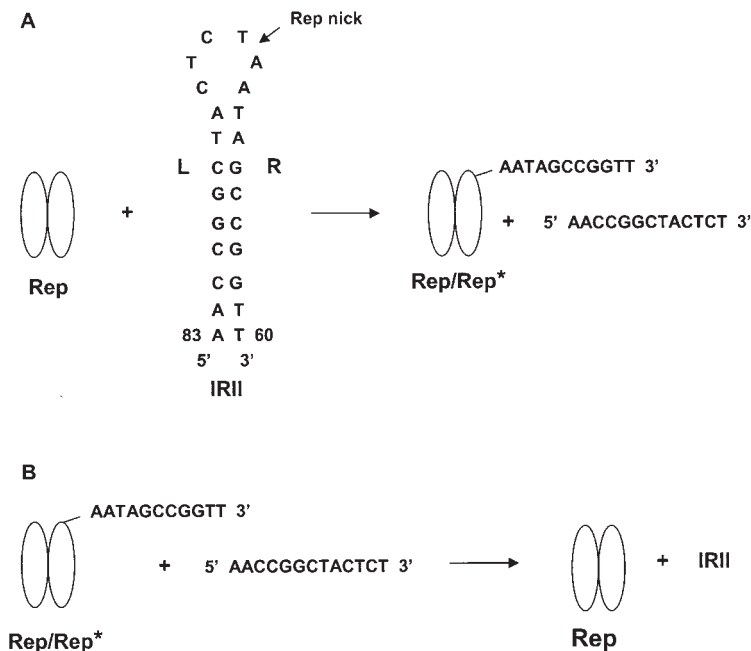


FIG. 9. Cleavage and religation of oligonucleotides by the Rep protein *in vitro*. (A) Generation of the modified Rep/Rep* proteins upon cleavage of the IRII bottom strand by the Rep protein. (B) Religation of the upstream oligonucleotide (containing the right arm and partial loop sequence) to the downstream oligonucleotide attached to the Rep/Rep* protein.

be the case by the discovery of the inactive RepC/RepC* for pT181 (33). More than 95% of the initiator protein present in the cell corresponds to this form (33,46). Two important issues arise from these observations: first, what is the molecular basis for the inactivity of the Rep protein and second, whether the inactive protein could potentially interfere with the activity of the unused, active form of the Rep protein in the cell. A series of studies have been carried out utilizing the RepC/RepC* and RepD/RepD* proteins to answer these questions. The inactive form of the RepC protein has been isolated from *Staphylococcus aureus* cells containing the pT181 plasmid as well as generated *in vitro* by incubating RepC or RepD proteins with the bottom strand of the IRII DNA (Fig. 9A). This results in the cleavage of the ssDNA followed by the attachment of the 5' end of the cleaved oligo to the catalytic Tyr residues of the initiator proteins. Interestingly, the RepC/RepC* and RepD/RepD* proteins nick SC plasmid DNA very poorly (15,31). The modified initiators also bind to the *dso* with reduced affinity and are unable to promote the extrusion of IRII in SC plasmid DNA (23). Thus,

the RepC/RepC* and RepD/RepD* proteins are not only defective in promoting initiation, but are also unlikely to have any significant inhibitory activity in replication since they are unable to compete effectively with the native, unused RepC for *dso* binding.

V. Role of Host Proteins in pT181 RC Replication

A. Rep-PcrA Helicase Interactions and Plasmid pT181 Replication

A key event during the initiation of plasmid RC replication is the recruitment of a helicase to the *dso* to promote DNA unwinding from the Rep-generated nick (Fig. 1). This is accomplished by the PcrA helicase. Studies by Iordanescu *et al.* identified the *pcrA* gene and demonstrated that it was required for the RC replication of the pT181 plasmid in *S. aureus* (49–51). The *pcrA* gene encodes a protein of 730 amino acids in *S. aureus*, and this gene has been identified in all the Gram-positive organisms whose genomes have so far been sequenced. PcrA is a monomeric helicase that belongs to the superfamily I (SF1) of DNA and RNA helicases (52,53). It contains seven conserved motifs that are found in other members of the SFI family (reviewed in Ref. (52)). The PcrA helicase of *Bacillus stearothermophilus* has been crystallized and its three-dimensional structure determined (54–56). Inactivation of the *pcrA* gene is lethal and it has been firmly established in both *S. aureus* and *Bacillus subtilis* that *pcrA* is an essential gene (49,57). Furthermore, plasmids of the pC194 and pE194 families also require PcrA for their RC replication (57,58). A mutant, *pcrA3*, containing a Thr61-Ile change in conserved motif Ia allows cell survival but is defective in supporting the replication of the pT181 plasmid (51). Nicked open-circular pT181 DNA accumulates in the *pcrA3* mutant, suggesting that this helicase may be required for the unwinding of the DNA that had been nicked at the *dso* by the initiator protein (51). These studies also showed that replication of the plasmids of the pC194 and pE194 families is not affected in the *pcrA3* mutant (57). Furthermore, compensatory mutants in the pT181-encoded RepC protein have been isolated that allow pT181 replication in the *pcrA3* strain suggesting an interaction between PcrA and RepC (50). The UvrD helicase of *Escherichia coli*, which belongs to the SFI family and has approximately 40% identity with PcrA, has been shown to be involved in the replication of RCR plasmids in *E. coli* (58).

The biochemical activities of the *B. stearothermophilus* and *S. aureus* PcrA helicases have been studied. These helicases have ATPase as well as NTPase activities which is stimulated by ssDNA and Rep-nicked pT181

DNA (59–61). These studies have shown that PcrA has both $5' \rightarrow 3'$ and $3' \rightarrow 5'$ helicase activities (59–61). The RepD initiator encoded by the pC221 plasmid is known to interact with PcrA at the plasmid origin and increases the processivity of this helicase (60). Furthermore, PcrA is capable of unwinding SC pT181 DNA that has been nicked by RepC and contains initiator covalently attached to the 5'-phosphate end (61). Supercoiled DNA that has been nicked by RepC but from which the initiator has been removed by proteinase digestion and phenol extraction is not unwound by the PcrA helicase, implying that Rep is necessary to target PcrA to the nicked *dso* (61). The *S. aureus* PcrA has also been shown by “pull-down assays” to physically interact with the RepC protein of pT181 (61). Further insight into the recruitment of PcrA to the pT181 *dso* have come from studies utilizing mutant initiators that are defective in *dso* binding (61). These RepC mutants are able to nick-close SC pT181 DNA through a transient interaction but do not stably bind to the *dso* in a sequence-specific manner (11,43). In the presence of a DNA-binding mutant of RepC and pT181 DNA, PcrA is unable to unwind the DNA (61). These studies further suggest that recruitment of PcrA to the *dso* involves a specific PcrA–RepC interaction.

Cell-free extracts made from the *S. aureus* *pcrA3* mutant were found to be inactive in pT181 replication in the presence of the RepC protein (61). However, addition of purified PcrA restored replication in the *pcrA3* extracts, directly demonstrating a role for this helicase in pT181 RC replication (61). The following role for PcrA in plasmid RC replication can be envisioned based on the results of *in vivo* and *in vitro* studies. The plasmid initiator proteins bind to the *dso* through a sequence-specific interaction and nick the DNA to generate a free 3'-OH end. The PcrA helicase is recruited to the *dso* through a specific protein–protein interaction with the Rep protein. This is followed by unwinding of the DNA starting from the nick site by the PcrA helicase. Since PcrA has both $5' \rightarrow 3'$ and $3' \rightarrow 5'$ helicase activities, the direction of its movement during plasmid RC replication is not evident. Most replicative helicases that unwind DNA during theta-type replication have a $3' \rightarrow 5'$ helicase activity, and thus translocate in a $3' \rightarrow 5'$ direction on the template strand ahead of the replication fork (61). PcrA, on the other hand, may either translocate in a $3' \rightarrow 5'$ direction on the template strand or in a $5' \rightarrow 3'$ direction on the displaced leading strand during plasmid RC replication (Fig. 1). The Rep protein is covalently bound to the 5' phosphate of the displaced strand and is expected to be in close proximity to the replication fork (since it is required to terminate replication). Thus, it is possible that PcrA could translocate on either of the two strands to unwind the DNA during movement of the replication fork. Future studies should reveal the nature of movement of this helicase during RC replication.

An issue that remains unresolved concerns the molecular events that promote the termination of plasmid RC replication once the leading strand has been fully synthesized. In the case of chromosomes and plasmids that replicate by a theta-type mechanism, the host replication terminator protein (RTP) binds to the *ter* sequence located approximately at an equal distance from the origin (62). This interaction blocks the movement of the replicative DnaB helicase, displacing it from the DNA and promoting termination. The RTPs of *E. coli* and *B. subtilis* have been biochemically characterized (62). Since RCR plasmids replicate unidirectionally, the *dso* also serves as the termination site. These plasmids are not known to contain any *ter*-like sequences and the movement of the PcrA helicase is not arrested at a *ter* sequence cloned onto the pT181 plasmid (63). What then promotes the termination of plasmid RC replication? It is postulated that the plasmid-encoded Rep protein plays an important role in this step. The Rep protein and PcrA are expected to be in close proximity during the replication fork movement as discussed earlier. Once the leading strand of pT181 has been fully synthesized, the replication fork progresses to approximately 11 nt beyond the Rep nick site regenerating the full IRII sequence (Figs. 1 and 8). It is hypothesized that the free monomer of the dimeric Rep would interact with the right arm of IRII through its sequence-specific DNA-binding domain, presumably blocking the movement of the replication fork. The nicking domain of the free monomer is then expected to cleave the displaced leading strand with the concomitant displacement of the PcrA helicase from the DNA. Another series of DNA cleavage and religation events promoted by the Rep protein would then complete leading-strand synthesis (Fig. 8). Experiments to test this model for the termination of pT181 RC replication are currently in progress.

B. Other Host Proteins Involved in pT181 RC Replication

Limited information is available on the role of other host proteins in the replication of RCR plasmids of the pT181 family. Evidence suggests that DNA polymerase III utilizes the RepC generated nick as a primer for pT181 leading-strand synthesis (64). Recent studies have shown that Gram-positive bacteria, including *S. aureus*, *B. subtilis*, and *Streptococcus pyogenes*, contain two different DNA polymerases, PolC and DnaE (65–67). Both of these polymerases are required for the replication of the chromosome, and further studies have suggested that PolC is required for leading-strand synthesis while DnaE is important for lagging-strand replication (66). Whether this is also the case for the replication of pT181 and other RCR plasmids in Gram-positive bacteria is not known and should be the subject of future investigations. As discussed in the next section, the host RNA

polymerase synthesizes the RNA primer for the lagging-strand replication of pT181.

VI. Replication of the Lagging Strand of Plasmid pT181

The parental, leading strand of pT181 is displaced after one round of leading-strand synthesis has been completed (Fig. 1). This ssDNA is converted to the ds form utilizing solely the host proteins (68,69). The ss \rightarrow ds conversion initiates at the plasmid single-strand origin, termed *ssoA* (70). The *ssos* are usually located slightly upstream of the *dsos* such that the ss \rightarrow ds conversion does not initiate until after the leading-strand synthesis has been completed (Fig. 1). Thus, the synthesis of plasmid leading and lagging strands is uncoupled. Four types of *ssos*, *ssoA*, *ssoT*, *ssoU*, and *ssoW* have been found in RCR plasmids (2,5). The *ssoA*-type origins found in pT181 as well as in many related and unrelated RCR plasmids have a conserved secondary structure although there is considerable variation in their primary sequence (2,5). The structure of *ssoA* has been shown to be critical for its function. The *ssos* contain two conserved sequences, RS_B and CS-6 (69–72). While plasmids of the pT181 family show a high degree of homology in their *dso* and initiator proteins, their *ssos* are not necessarily similar and closely related *ssos* can be found in many unrelated plasmids belonging to other plasmid families (2,5). The *ssos*, including the *ssoA* of pT181 are strand and orientation specific (2,5,70). *In vitro* and *in vivo* studies have shown that rifampicin treatment inhibits the lagging-strand synthesis of pT181 and other RCR plasmids, and subsequent studies have shown that the host RNA polymerase synthesizes the RNA primers required for lagging-strand synthesis (68,69,72,73). Although the pT181 *ssoA* has not been well studied, analysis of the related *ssos* of the pE194 and pLS1 plasmids has shown that the conserved RS_B sequence contains the RNA polymerase-binding site (72,73). This region includes sequences that resemble the canonical –10 and –35 sequences, and RS_B is involved in the initiation of primer RNA synthesis (71–73). The conserved CS-6 region that is present in the central loop of the *ssoA* structure contains the termination site of primer RNA synthesis (69,72). The length of the primer RNA synthesized from the pT181 *ssoA* is not known, but based on its close resemblance to the pE194 and pLS1 *ssos*, it is postulated that it may be approximately 20 nt in length (72,73). Current evidence suggests that DNA PolI utilizes the RNA primer to carry out limited DNA synthesis, and subsequently the more processive DNA PolC (or DnaE) may replicate the lagging strand (72,74). After the ss \rightarrow ds synthesis has been completed, DNA ligase is expected to join the free 3' and 5' ends and the DNA is expected to be converted to the SC form by DNA gyrase (Fig. 1).

VII. Conclusions and Future Directions

Considerable information has been obtained on the molecular mechanisms involved in the initiation and termination of replication of the plasmids of the pT181 family. Sequences critical in the leading- and lagging-strand synthesis have been identified, and a fair understanding of structural features that are important in their activities exists. The biochemical activities of the initiator proteins that are critical in their initiation and termination activities have been investigated, and the biochemical basis for their specificity in the replication of their cognate plasmids is well understood. The role of *ssoA* in pT181 lagging-strand replication has been investigated, but much remains to be done for a better understanding of this process. The role of the PcrA helicase in leading-strand replication and its interaction with the plasmid initiator proteins has been studied, but little is known about the molecular events involved in the elongation and termination of pT181 RC replication. The DNA polymerases involved in the leading- and lagging-strand replication have not been conclusively identified, and biochemical studies are necessary for a better understanding of this process. Finally, determination of the crystal structures of the Rep proteins of the pT181 family will be critical for a comprehensive understanding of the DNA-protein interactions, critical for the initiation and termination of leading-strand replication.

ACKNOWLEDGMENTS

I thank past and present members of the laboratory for helpful discussions. The work in the author's laboratory has been supported by grant GM31685 from the National Institutes of Health.

REFERENCES

1. G. H. del Solar, M. Moscoso and M. Espinosa, Rolling-circle replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls, *Mol. Microbiol.* **8**, 789–796 (1993).
2. G. del Solar, R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa and R. Diaz-Orejas, Replication and control of circular bacterial plasmids, *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).
3. A. Gruss and S. D. Ehrlich, The family of highly interrelated single-stranded deoxyribonucleic acid plasmids, *Microbiol. Rev.* **53**, 231–241 (1989).
4. S. A. Khan, Mechanism of replication and copy number control of plasmids in Gram-positive bacteria, in: J. K. Setlow (ed.), "Genetic Engineering." Vol. 18, pp. 183–201. New York, Plenum Press, 1996.
5. S. A. Khan, Rolling-circle replication of bacterial plasmids, *Microbiol. Mol. Biol. Rev.* **61**, 442–455 (1997).
6. S. A. Khan, Plasmid rolling-circle replication: recent developments, *Mol. Microbiol.* **37**, 477–484 (2000).

7. R. P. Novick, Staphylococcal plasmids and their replication, *Annu. Rev. Microbiol.* **43**, 537–565 (1989).
8. R. P. Novick, Contrasting lifestyles of rolling-circle phages and plasmids, *TIBS* **23**, 434–438 (1998).
9. S. J. Projan and R. P. Novick, Comparative analysis of five related staphylococcal plasmids, *Plasmid* **19**, 203–221 (1988).
10. T. V. Ilyina and E. V. Koonin, Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria, *Nucleic Acids Res.* **20**, 3279–3285 (1992).
11. L. A. Dempsey, P. Birch and S. A. Khan, Uncoupling of the DNA topoisomerase and replication activities of an initiator protein, *Proc. Natl. Acad. Sci. USA* **89**, 3083–3087 (1992).
12. C. D. Thomas, D. F. Balson and W. V. Shaw, Identification of the tyrosine residue involved in bond formation between replication origin and the initiator protein of plasmid pC221, *Biochem. Soc. Trans.* **16**, 758–759 (1988).
13. P. Noirot, J. Bargonetti and R. P. Novick, Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin, *Proc. Natl. Acad. Sci. USA* **87**, 8560–8564 (1990).
14. R. Jin and R. P. Novick, Role of the double-strand origin cruciform in pT181 replication, *Plasmid* **46**, 95–105 (2001).
15. R. Jin, M.-E. Fernandez-Beros and R. P. Novick, Why is the initiation nick site of an AT-rich rolling circle plasmid at the tip of a GC-rich cruciform?, *EMBO J.* **16**, 4456–4466 (1997).
16. J. M. Zock, P. Birch and S. A. Khan, Specificity of RepC protein in plasmid pT181 DNA replication, *J. Biol. Chem.* **265**, 3484–3488 (1990).
17. C. D. Thomas, T. T. Nikiforov, B. A. Connolly and W. V. Shaw, Determination of sequence specificity between a plasmid replication initiator protein and the origin of replication, *J. Mol. Biol.* **254**, 381–391 (1995).
18. C. D. Thomas, D. F. Balson and W. V. Shaw, In vitro studies of the initiation of staphylococcal plasmid replication, *J. Biol. Chem.* **265**, 5519–5530 (1990).
19. S. A. Khan and R. P. Novick, Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from *Staphylococcus aureus*, *Plasmid* **10**, 251–259 (1983).
20. S. A. Khan, G. K. Adler and R. P. Novick, Functional origin of replication of pT181 plasmid DNA is contained within a 168 base-pair segment, *Proc. Natl. Acad. Sci. USA* **79**, 4580–4584 (1982).
21. M. L. Gennaro, S. Iordanescu, R. P. Novick, R. W. Murray, T. R. Steck and S. A. Khan, Functional organization of the plasmid pT181 replication origin, *J. Mol. Biol.* **205**, 355–362 (1989).
22. R. R. Koepsel, R. W. Murray and S. A. Khan, Sequence-specific interaction between the replication initiator protein of plasmid pT181 and its origin of replication, *Proc. Natl. Acad. Sci. USA* **83**, 5484–5488 (1986).
23. R. Jin, X. Zhou and R. P. Novick, The inactive pT181 initiator heterodimer, RepC/C*, binds but fails to induce melting of the plasmid replication origin, *J. Biol. Chem.* **271**, 31086–31091 (1996).
24. P.-Z. Wang, S. J. Projan, V. Henriquez and R. P. Novick, Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element, *EMBO J.* **12**, 45–52 (1993).
25. S. Iordanescu, Specificity of the interaction between the Rep proteins and the origin of replication of *Staphylococcus aureus* plasmids pT181 and pC221, *Mol. Gen. Genet.* **217**, 481–487 (1989).
26. R. R. Koepsel and S. A. Khan, Static and initiator protein-enhanced bending of DNA at a replication origin, *Science* **233**, 1316–1318 (1986).

27. R. R. Koepsel, R. W. Murray, W. D. Rosenblum and S. A. Khan, The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities, *Proc. Natl. Acad. Sci. USA* **82**, 6845–6849 (1985).
28. A. C. Zhao and S. A. Khan, An 18-bp sequence is sufficient for termination of rolling-circle replication of plasmid pT181, *J. Bacteriol.* **24**, 5222–5228 (1996).
29. A. C. Zhao and S. A. Khan, Sequence requirements for the termination of rolling-circle replication of plasmid pT181, *Mol. Microbiol.* **24**, 535–544 (1997).
30. S. Iordanescu and S. J. Projan, Replication terminus for staphylococcal plasmids: plasmids pT181 and pC221 cross-react in the termination process, *J. Bacteriol.* **170**, 3427–3434 (1988).
31. A. C. Zhao, R. A. Ansari, M. C. Schmidt and S. A. Khan, An oligonucleotide inhibits oligomerization of a rolling-circle initiator protein at the pT181 origin of replication, *J. Biol. Chem.* **273**, 16082–16089 (1998).
32. R. W. Murray, R. R. Koepsel and S. A. Khan, Synthesis of single-stranded plasmid pT181 DNA *in vitro*: initiation and termination of DNA replication, *J. Biol. Chem.* **264**, 1051–1057 (1989).
33. A. Rasooly and R. P. Novick, Replication-specific inactivation of the pT181 plasmid initiator protein, *Science* **262**, 1048–1050 (1993).
34. M. L. Gennaro, Genetic evidence for replication enhancement from a distance, *Proc. Natl. Acad. Sci. USA* **90**, 5529–5533 (1993).
35. M. L. Gennaro and R. P. Novick, *cmp*, a *cis*-acting plasmid locus that increases interaction between replication origin and initiator protein, *J. Bacteriol.* **168**, 160–166 (1986).
36. V. Henriquez, V. Milisavljevic, J. D. Kahn and M. L. Gennaro, Sequence and structure of *cmp*, the replication enhancer of the *Staphylococcus aureus* plasmid pT181, *Gene* **134**, 93–98 (1993).
37. D. Colombo, S. Iordanescu and M. L. Gennaro, Replication enhancer requirement for recognition of heterologous replication origin by an initiator protein, *Plasmid* **33**, 232–234 (1995).
38. Q. Zhang, S. S. De Olivera, R. Colangeli and M. L. Gennaro, Binding of a novel host factor to the pT181 plasmid replication enhancer, *J. Bacteriol.* **179**, 684–688 (1997).
39. R. R. Koepsel, R. W. Murray, W. D. Rosenblum and S. A. Khan, Purification of pT181-encoded RepC protein required for the initiation of plasmid replication, *J. Biol. Chem.* **260**, 8571–8577 (1985).
40. S. A. Khan, S. M. Carleton and R. P. Novick, Replication of plasmid pT181 DNA *in vitro*: requirement for a plasmid-encoded product, *Proc. Natl. Acad. Sci. USA* **78**, 4902–4906 (1981).
41. L. A. Dempsey, P. Birch and S. A. Khan, Six amino acids determine the sequence-specific DNA binding and replication specificity of the initiator proteins of the pT181 family, *J. Biol. Chem.* **267**, 24538–24543 (1992).
42. P.-Z. Wang, S. J. Projan, V. Henriquez and R. P. Novick, Specificity of origin recognition by replication initiator protein in plasmids of the pT181 family is determined by a six amino acid residue element, *J. Mol. Biol.* **223**, 145–158 (1992).
43. T.-L. Chang, M. G. Kramer, R. A. Ansari and S. A. Khan, Role of individual monomers of a dimeric initiator protein in the initiation and termination of plasmid rolling circle replication, *J. Biol. Chem.* **275**, 13529–13534 (2000).
44. R. R. Koepsel and S. A. Khan, Cleavage of single-stranded DNA by plasmid pT181-encoded RepC protein, *Nucleic Acids Res.* **15**, 4085–4097 (1987).
45. C. D. Thomas and L. J. Jennings, RepD/D*: a protein-DNA adduct arising during plasmid replication, *Biochem. Soc. Trans.* **23**, 442S (1995).
46. R. Jin, A. Rasooly and R. P. Novick, *In vitro* inhibitory activity of RepC/C*, the inactivated form of the pT181 plasmid initiation protein, RepC, *J. Bacteriol.* **179**, 141–147 (1997).

47. A. Rasooly, S. J. Projan and R. P. Novick, Plasmids of the pT181 family show replication-specific initiator protein modification, *J. Bacteriol.* **176**, 2450–2453 (1994).
48. A. Rasooly, P.-Z. Wang and R. P. Novick, Replication-specific conversion of the *Staphylococcus aureus* pT181 initiator protein from an active homodimer to an inactive heterodimer, *EMBO J.* **13**, 5245–5251 (1994).
49. S. Iordanescu, Identification and characterization of *Staphylococcus aureus* chromosomal gene *pcrA*, identified by mutations affecting plasmid pT181 replication, *Mol. Gen. Genet.* **241**, 185–192 (1993).
50. S. Iordanescu, Plasmid pT181-linked suppressors of the *Staphylococcus aureus* *pcrA3* chromosomal mutation, *J. Bacteriol.* **175**, 3916–3917 (1993).
51. S. Iordanescu and R. Basheer, The *Staphylococcus aureus* mutation *pcrA3* leads to the accumulation of pT181 replication initiation complexes, *J. Mol. Biol.* **221**, 1183–1189 (1991).
52. M. C. Hall and S. W. Matson, Helicase motifs: the engine that powers DNA unwinding, *Mol. Microbiol.* **34**, 867–877 (1989).
53. L. E. Bird, S. Subramanya and D. B. Wigley, Helicases: a unifying structural theme?, *Curr. Opin. Struct. Biol.* **8**, 14–18 (1998).
54. H. S. Subramanya, L. E. Bird, J. A. Branningan and D. B. Wigley, Crystal structure of a DExx box DNA helicase, *Nature* **384**, 379–383 (1996).
55. S. S. Velankar, P. Soultanas, M. S. Dillingham, S. Subramanya and D. B. Wigley, Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism, *Cell* **97**, 75–84 (1999).
56. P. Soultanas, M. S. Dillingham, P. Wiley, M. R. Webb and D. B. Wigley, Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism, *EMBO J.* **19**, 3799–3810 (2000).
57. M.-A. Petit, E. Dervyn, M. Rose, K.-D. Entian, S. McGovern, S. D. Ehrlich and C. Bruand, PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication, *Mol. Microbiol.* **29**, 261–273 (1998).
58. C. Bruand and S. D. Ehrlich, UvrD-dependent replication of rolling-circle plasmids in *Escherichia coli*, *Mol. Microbiol.* **35**, 204–210 (2000).
59. M. S. Dillingham, P. Soultanas and D. B. Wigley, Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation, *Nucleic Acids Res.* **27**, 3310–3317 (1999).
60. P. Soultanas, M. S. Dillingham, F. Papadopoulos, S. E. V. Phillips, C. D. Thomas and D. B. Wigley, Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase, *Nucleic Acids Res.* **27**, 1421–1428 (1999).
61. T.-L. Chang, A. Naqvi, S. P. Anand, M. G. Kramer, R. Munshi and S. A. Khan, Biochemical characterization of the *Staphylococcus aureus* PcrA helicase and its role in plasmid rolling-circle replication, *J. Biol. Chem.* **277**, 45880–45886 (2002).
62. D. E. Bussiere and D. Bastia, Termination of DNA replication of bacterial and plasmid chromosomes, *Mol. Microbiol.* **31**, 1611–1618 (1999).
63. S. Kaul, B. K. Mohanty, T. Sahoo, I. Patel, S. A. Khan and D. Bastia, The replication terminator protein of the Gram-positive bacterium *B. subtilis* functions as a polar contrahelicase in Gram-negative *E. coli*, *Proc. Natl. Acad. Sci. USA* **91**, 11143–11147 (1994).
64. S. Majumder and R. P. Novick, Intermediates in plasmid pT181 DNA replication, *Nucleic Acids Res.* **16**, 2897–2912 (1988).
65. R. Inoue, C. Kaito, M. Tanabe, K. Kamura, N. Akimitsu and K. Sekimizu, Genetic identification of two distinct DNA polymerases, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus*, *Mol. Gen. Genomics* **266**, 564–571 (2001).

66. E. Dervyn, C. Suski, R. Daniel, C. Bruand, J. Chapuis, J. Errington, L. Janniere and S. Dusko Ehrlich, Two essential DNA polymerases at the bacterial replication fork, *Science* **294**, 1716–1719 (2001).
67. I. Bruck and M. O'Donnell, The DNA replication machine of a Gram-positive organism, *J. Biol. Chem.* **275**, 28971–28983 (2000).
68. P. Birch and S. A. Khan, Replication of single-stranded plasmid pT181 DNA in vitro, *Proc. Natl. Acad. Sci. USA* **89**, 290–294 (1992).
69. L. A. Dempsey, A. C. Zhao and S. A. Khan, Localization of the start sites of lagging-strand replication of rolling-circle plasmids from Gram-positive bacteria, *Mol. Microbiol.* **15**, 679–687 (1995).
70. A. Gruss, H. F. Ross and R. P. Novick, Functional analysis of a palindromic sequence required for normal replication of several staphylococcal plasmids, *Proc. Natl. Acad. Sci. USA* **84**, 2165–2169 (1987).
71. M. G. Kramer, S. A. Khan and M. Espinosa, Lagging strand replication from *ssoA* origin of plasmid pMV158 in *Streptococcus pneumoniae*: In vivo and in vitro influences of mutations in two conserved *ssoA* regions, *J. Bacteriol.* **180**, 83–89 (1998).
72. M. G. Kramer, S. A. Khan and M. Espinosa, Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement of DNA polymerase I for lagging strand synthesis, *EMBO J.* **16**, 5784–5795 (1997).
73. M. G. Kramer, M. Espinosa, T. K. Misra and S. A. Khan, Lagging strand replication of rolling-circle plasmids: specific recognition of the *ssoA*-type origins in different Gram-positive bacteria, *Proc. Natl. Acad. Sci. USA* **95**, 10505–10562 (1998).
74. A. Diaz, S. A. Lacks and P. Lopez, Multiple roles for DNA polymerase I in establishment and replication of the promiscuous plasmid pLS1, *Mol. Microbiol.* **14**, 773–783 (1994).

This Page Intentionally Left Blank

Herpes Simplex Virus Type-1: A Model for Genome Transactions

PAUL E. BOEHMER,* AND
GIUSEPPE VILLANI,†

**Department of Biochemistry and
Molecular Biology, University of Miami
School of Medicine, P.O. Box 016129,
Miami, FL 33101-6129, USA; and*

*†Institut de Pharmacologie et de Biologie
Structurale, CNRS-UMR 5089, 205 route
de Narbonne, 31077 Toulouse Cedex 4,
France*

I. Introduction to <i>Herpesviridae</i>	140
II. HSV-1 Life Cycle, Genome Structure, and Replication Genes	140
III. Initiation of Replication	144
IV. Elongation	151
V. Interaction of the HSV-1 DNA Replication Machinery with DNA Damage	155
VI. Homologous Recombination	159
VII. Concluding Remarks	163
Acknowledgments	165
References	165

In many respects, HSV-1 is the prototypic herpes virus. However, HSV-1 also serves as an excellent model system to study genome transactions, including DNA replication, homologous recombination, and the interaction of DNA replication enzymes with DNA damage. Like eukaryotic chromosomes, the HSV-1 genome contains multiple origins of replication. Replication of the HSV-1 genome is mediated by the concerted action of several virus-encoded proteins that are thought to assemble into a multiprotein complex. Several host-encoded factors have also been implicated in viral DNA replication. Furthermore, replication of the HSV-1 genome is known to be closely associated with homologous recombination that, like in many cellular organisms, may function in recombinational repair. Finally, recent data have shed some light on the interaction of essential HSV-1 replication proteins, specifically its DNA polymerase and DNA helicases, with damaged DNA. © 2003 Elsevier Science

I. Introduction to *Herpesviridae*

The *Herpesviridae* family encompasses a group of large double-stranded (ds) DNA viruses that are ubiquitous to vertebrate species. Herpes viruses share the ability to establish latent infections that lead to life-long persistence of the virus and associated symptoms. To date, eight human herpes viruses have been identified (Human Herpes Virus (HHV) 1–8) and classified into three subfamilies, *alpha*-, *beta*-, and *gamma-herpesvirinae*. Common to all human herpes viruses are basic similarities in their virion structure and certain aspects of their life cycle. However, they differ in specific biological properties and in the symptoms they cause, depending on the cell type in which they replicate and establish latency (see Ref. (1), for general review on herpes viruses). Herpes virus infections are a major public health problem, given their prevalence in the population. Primary infections with most herpes viruses are asymptomatic with the exception of genital herpes, mononucleosis, and chicken pox that are caused by primary infections with Herpes Simplex Virus (HSV) type-2, Epstein-Barr Virus (EBV), and Varicella-Zoster Virus (VZV), respectively. Most other symptoms are the result of recurrent infections from reemerging latent viruses including the characteristic “cold sores” caused by HSV-1. Reactivation of latent Cytomegalovirus causes symptoms that are a primary cause of morbidity and mortality in immunocompromised individuals such as those suffering from AIDS or organ-transplant patients. Finally, EBV and HHV-8 are associated with neoplastic transformations including Burkitt’s lymphoma and Kaposi’s sarcoma, respectively.

II. HSV-1 Life Cycle, Genome Structure, and Replication Genes

In many respects, HSV-1 is the prototypic herpes virus. Thus, it serves as a model for studying viral lytic replication (including gene expression and DNA replication), viral latency, and viral pathogenesis. However, as will become apparent in this review, HSV-1 also serves as an excellent model to study genome transactions, including DNA replication, homologous recombination, and the interaction of DNA replication enzymes with DNA damage. In this regard, HSV-1 serves as a model system similar to several others that have been used to investigate these processes including bacteriophages (e.g., T4 and T7), eukaryotic viruses (e.g., simian virus 40 (SV40), or papillomaviruses), prokaryotes such as *Escherichia coli*, and eukaryotes including yeast and

humans. Similarities between HSV-1 replication components and their counterparts in other systems have been discussed previously (2).

Viral genome transactions occur in the nuclei of infected cells. Capsids that are released into the cytosol following fusion of the viral envelope with the plasma membrane are transported toward the nucleus along tubulin microtubules by the dynein/dynactin motor complex (3). Capsids are too large to enter the nucleus via the nuclear pore complex. Instead, viral genomes are released into the nucleus in a process that requires importin- β (4).

The HSV-1 capsid contains a single copy of the viral genome that comprises linear duplex DNA of ~ 152 kbp encoding ~ 75 genes. An HSV-1 sequence database that annotates each gene may be found at <http://www.stdgen.lanl.gov>. A linear representation of the genome is shown in Fig. 1. Unique DNA sequences, unique long (U_L), and unique short (U_S), containing the majority of open-reading frames, are flanked by inverted repeats a , b , and c . The viral a sequences mediate obligatory circularization of the genome shortly (within ~ 30 min) after infection. This process is thought to involve cellular factors including the regulator of chromosome condensation (RCC1), and has been postulated to proceed either by recombination or even direct ligation (5–7). Thus, the linear genome possesses an asymmetrically partitioned terminal a sequence, with single 3'-nucleotide extensions. Circularization reconstitutes a complete a sequence (8). The a sequences also contain the *cis*-acting

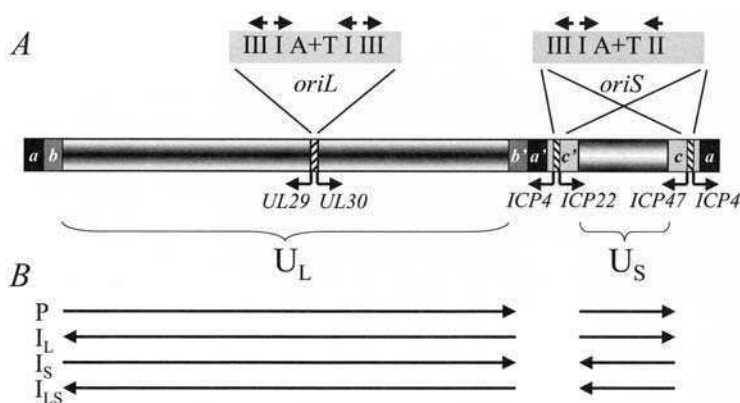


FIG. 1. Structure of the HSV-1 genome. (A) Diagrammatic representation of the HSV-1 genome showing the relative positions of the unique regions, U_L and U_S , the repeats, a , b , and c , and the location of the origins, $oriL$ and $oriS$. Flanking the origins are divergently transcribed genes $UL29$ and $UL30$ ($oriL$), and $ICP4$ and $ICP22/47$ ($oriS$). ' denotes a region present in an inverted orientation. Multiple copies of the a sequence may be present. The inverted arrows within the origins denote the relative orientations of the $UL9$ -binding sites flanking the $A+T$ -rich region ($A+T$). (B) The arrows indicate the relative orientation of U_L and U_S found in the four genomic isoforms P, I_L , I_S , and I_{LS} . Not to scale.

sites (*pac*) that are required for cleavage and packing of full-length genomes (9,10).

In addition to circularization, several other viral processes are initiated immediately following infection. Thus, VP16, a late viral protein that is packaged into the tegument layer of the virion and therefore available immediately upon infection, acts in concert with cellular factors (Oct-1 and HCF) to activate viral immediate early (IE) gene expression (11). This initiates the cascade of viral gene expression in which IE gene products, including ICP0, ICP4, ICP22, and ICP27, are responsible for regulating delayed early (DE) and late gene expression (reviewed in Ref. (1)). All viral transcripts are synthesized by cellular RNA polymerase II and are processed to be 5' capped and 3' polyadenylated. With a few exceptions, the majority of transcripts are intronless. ICP4, ICP22, and ICP27 act by a variety of means to activate, but in some instances to repress, viral gene expression by interacting with components of the cellular transcription machinery (1). ICP0 on the other hand appears to have a more enigmatic mechanism by which it stimulates viral gene expression (12). It has been shown to mediate proteasome-dependent degradation of a variety of cellular targets by a mechanism that presumably involves its ubiquitin ligase activity (13,14). An additional IE gene product, ICP47, elicits a mechanism to enable the virus to evade immune recognition by blocking the transport of peptides for antigen presentation, thereby preventing MHC complex formation (15). Similarly, the virion host shut-off protein (Vhs), a further tegument protein, also contributes toward optimizing viral gene expression by mediating the degradation of cellular transcripts probably by functioning as an RNase (16).

Parental viral genomes localize adjacent to nuclear domain 10 (ND10), subnuclear structures that are defined by immunostaining to specific proteins including Sp100 and promyelocytic leukemia protein (PML), at sites that develop into replication compartments containing viral DNA replication factors where replication occurs (17,18). The specific role of ND10 in this regard is unclear. Nevertheless, following viral infection, ND10 are destroyed in an ICP0-dependent manner, possibly to disrupt a structure or component that is inhibitory to viral gene expression and/or replication, or to liberate a necessary component (19,20).

While VP16-dependent IE gene expression has been shown to rely on cycle-dependent protein kinase (cdk) activity, HSV-1 replication per se is independent of cell cycle functions (21). This is in contrast to other DNA viruses that replicate in the nuclei of infected cells (e.g., *papovaviridae*). In fact, HSV-1 has been shown to stop cells in G1, thereby preventing cells from going into S phase, and to prevent mitosis (22,23).

The overall DNA replication strategy adopted by HSV-1 is as follows. Following circularization, replication is thought to proceed bidirectionally

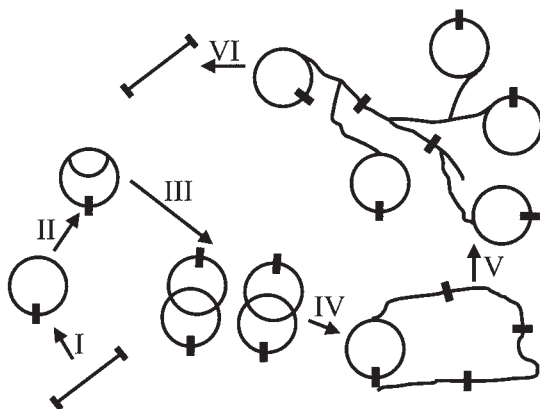


Fig. 2. Schematic representation of the HSV-1 life cycle. I, circularization of parental viral genome. II, initiation of origin-dependent DNA replication. III, theta-type replication generates catenated unit-length circular products. IV, sigma-type replication generates head-to-tail concatamers. V, replication-dependent recombination generates branched replication intermediates. VI, cleavage generates linear unit-length genomes that are available for packing into preformed capsids. The black boxes represent terminal repeats that delineate unit-length genomes.

from one or multiple origins. Despite the lack of clear evidence for theta-type DNA replication intermediates, the existence of specific sites that act as origins, and a protein with properties consistent with those of an initiator protein, indicate that replication initially proceeds by a theta-type mechanism. The products of such a reaction are catenated circular daughter molecules. Later in replication, viral DNA comprises head-to-tail concatamers, suggestive of a sigma or rolling-circle mode of replication. In addition, late replication is accompanied by the formation of highly branched DNA intermediates resulting in large networks. The general replication strategy is outlined in Fig. 2 and has recently been reviewed in Ref. (1).

HSV-1 encodes seven genes that are essential for viral origin-dependent DNA replication (24). These genes encode a set of proteins whose activities would be sufficient to initiate and sustain DNA replication from an origin. The identity of these genes and their functions are summarized in Table I. They include an initiator protein (origin-binding protein with intrinsic DNA helicase activity), a processive DNA polymerase, a replicative DNA helicase with associated primase activity, and a single-strand DNA-binding protein (SSB). The properties and functions of these proteins will be discussed in the appropriate context in the following sections. Apart from these virus-encoded proteins, several host factors have been implicated in viral DNA replication either by their association with viral factors or by binding to *cis*-acting sites in the viral genome. These include an origin-binding factor, OF-1, DNA

TABLE I
HSV-1 DNA REPLICATION PROTEINS

Gene	Function	Size (kDa)
<i>UL9</i>	Initiator (origin-binding protein/3'-5' DNA helicase)	94
<i>UL30</i>	DNA polymerase/RNase H/3'-5' exonuclease	136
<i>UL42</i>	DNA polymerase clamp	51
<i>UL29</i>	SSB/recombinase	128
<i>UL5</i>	Helicase ^a (5'-3' DNA helicase/NTase)	99
<i>UL52</i>	Primase ^a	114
<i>UL8</i>	Helicase-primase-loading protein ^b	80

^aMotifs characteristic of each activity are present in the respective polypeptides but both activities are detectable only in the two subunit complex.

^bHelicase-primase-loading protein is a third subunit of the helicase-primase complex and interacts with the other DNA replication proteins.

polymerase α -primase, molecular chaperones (heat-shock proteins 40 (DnaJ) and 70 (DnaK)), topoisomerase II, and factors that interact with Sp1 transcription factor binding sites and a glucocorticoid response element. Their roles will also be discussed in the appropriate context later.

Viral DNA replication generates head-to-tail concatamers that are the substrate for cleavage and packaging into preformed capsids. Cleavage and packaging depends on at least seven HSV-1 genes including a putative terminase that cleaves the DNA at appropriately oriented *pac* sites to generate unit-length genomes (1,25,26).

III. Initiation of Replication

A. Origins of Replication

The viral genome contains two classes of origins of replication (*ori*) that exhibit significant homology. *oriL* is located roughly in the middle of U_L , while *oriS* is situated in the *c* repeats and is therefore diploid (27,28) (Fig. 1). Both classes of origin are located between divergent transcriptional units. Thus, *oriL* is situated between two DE genes (*UL29* and *UL30*), while *oriS* is situated between two IE genes (ICP4 and ICP22/47) (Fig. 1). The minimal origins, comprising ~ 80 bp palindromes, contain a central A + T-rich region of ~ 18 bp that is flanked by 10-bp inverted repeats (5' CGTTCGCACT) that function as binding sites for the virus-encoded origin-binding protein (Fig. 1) (29). While *oriL* possesses two inverted repeats on either side of the A + T-rich region, *oriS* lacks the second inverted repeat on the right side of the A + T-rich region. The origins are redundant since viruses with a single origin

of either class replicate with wild-type kinetics (30,31). Consequently, it is unclear why multiple and distinct origins exist. It is possible that their location in transcriptional units that are active at different times during infection (IE vs. DE) may contribute to their regulation. In this regard, it has recently been shown that *oriS* can adopt a discrete secondary structure (*oriS**) that involves intrastrand hairpin formation (32). Such a structure may be formed as a result of negative supercoiling induced by divergent transcription from flanking promoters. Indeed, origin activity has been shown to increase when *oriS* and *oriL* are in the context of longer DNA sequences that include their respective divergent promoters (33,34). Sp1 transcription factor-binding sites in proximity to the origins have been shown to contribute to this effect (35). Unique to *oriL* is a glucocorticoid-response element that increases origin activity in neuronal cells, which may be related to enhanced reactivation of latent virus by glucocorticoid treatment (36). Additional sites within the minimal origins that have been shown to affect origin activity include a binding site for a cellular factor (OF-1) that is required for efficient origin activity (37). Its role will be expanded upon in Section III.D.

B. Origin-Binding Protein, UL9

A search for the viral initiator revealed a protein that bound to the inverted repeats in *oriS* (38). Subsequently, it was found that this protein was encoded by the *UL9* gene, which is essential for viral replication (39). UL9 is an 851-amino acid phosphoprotein of ~94 kDa that is dimeric in solution (40,41). It is not clear what role phosphorylation plays. Its N-terminal two-thirds (residues 1–534) possesses motifs present in the superfamily II of DNA helicase, such as *E. coli* UvrB. Indeed, purified recombinant UL9_(1–534) exhibits ssDNA-binding, ATPase, and DNA helicase activities (42). It has been proposed that this domain also contains a leucine zipper that is involved in the dimerization of UL9 (43). The C-terminal two-thirds (residues 535–851) contains a helix-turn-helix motif (43). This part of UL9, which is monomeric in solution, is sufficient for its origin-binding activity (44,45). Although UL9 is dimeric in solution and when bound to its target sequence, only one protomer is in contact with the DNA (46). Molecular modeling has shown that UL9 interacts with its binding site via the major groove (29). The spacing between the inverted repeats orientates UL9 bound at each site to the same face of the DNA (29). In *oriS*, UL9 binds to the inverted repeat to the left of the A + T-rich region (box I) with the highest affinity (k_a and k_d of 0.1 and 0.3 nM, respectively), followed by ~10-fold lower-affinity binding to box II and essentially nonspecific binding to box III (29/38). While these measurements are on the basis of interaction of UL9 with isolated-binding sites, in the context of the entire origin, cooperative interactions occur and binding is detected at all sites leading to the formation of a nucleoprotein

complex (44). Recently, it has been observed that intrastrand base pairing between boxes I and III gives rise to a novel origin conformation designated *oriS** that is efficiently recognized by UL9 to form a stable complex (32). It is possible that this structure is an intermediate in the initiation process.

UL9 exhibits low-level basal ATPase activity that is presumably due to its low affinity for ATP in the absence of DNA (47–49). The ATPase is strongly stimulated by ssDNA ($k_{\text{cat}} \sim 0.3 \text{ s}^{-1}$), which is reflected in an increased but yet relatively weak affinity for ATP ($K_{\text{m}} \sim 0.5 \text{ mM}$) (49). For ATP hydrolysis, the K_{m} for ssDNA is $\sim 1\text{-}\mu\text{M}$ nucleotide (50). The rate of ATP hydrolysis for UL9 increases with increasing length of ssDNA template indicating that the enzyme translocates along the DNA (48,51). Indeed, on nonspecific DNA templates it has been shown to unwind DNA in the 3'–5' direction, in a manner that requires a ssDNA-loading site (40,52). UL9 catalyzed DNA unwinding appears to be quite feeble, requiring high concentrations of enzymes, and is relatively distributive, as judged by the length of products (up to ~ 200 nucleotides) and its sensitivity to template challenge (50–52). Nevertheless, examination by electron microscopy revealed that single complexes of UL9, consistent with the mass of dimers, could traverse regions of duplex DNA ~ 1500 nucleotides from the ssDNA-loading site (53). Consequently, the lack of long unwinding products generated by UL9 may be due to spontaneous rewinding of the DNA following passage of the enzyme. Although UL9 was seen to translocate through regions of duplex DNA, it presumably makes contact with only one strand of the DNA. This is supported by its apparent polarity and by the finding that duplex DNA does not significantly stimulate its ATPase activity (40,47,48,52). Moreover, UL9 helicase activity is only impeded by DNA lesions located on the strand on which the enzyme translocates (see Section V.B) (54). UL9 helicase may be considered to act stoichiometrically rather than catalytically considering that both the rate and extent of unwinding are concentration dependent (52). Of further relevance to the mechanism of action of UL9 is its stimulation by reducing agents and its sensitivity to modification by the sulfhydryl-specific reagent *N*-ethylmaleimide (49). These observations are specific for the ssDNA-dependent ATPase and helicase activities of UL9 and implicate cysteines in its mechanism of action, specifically in coupling ssDNA binding to ATP binding and hydrolysis (49). The rate of DNA unwinding by UL9 has been estimated to be 75 bp min^{-1} (52). However, a detailed presteady state kinetic analysis is necessary to determine several parameters including its exact rate of unwinding, its step-size (i.e., number of bases per translocation event), the efficiency of unwinding (i.e., number of ATP hydrolyzed per base pair unwound), and its exact processivity. Such studies would also allow description of a kinetic scheme for its mechanism of DNA unwinding and

determination of the stoichiometry of the active enzyme (i.e., dimeric vs. multimeric or monomeric).

Although the nonspecific DNA helicase activity of UL9 has been studied as a representative member of its class of DNA helicases, given that it is an origin-binding protein, presumably its role is to unwind the DNA at the origin to generate ss template DNA. This activity is discussed in Section III.E.

Finally, it should be noted that overexpression of UL9 or its origin-binding domain has a dominant negative effect on viral replication. It is possible that in these cases, high occupancy of the origins blocks replication fork progression and thereby inhibits viral replication. In this regard, a transcript corresponding to the C-terminal origin-binding domain of UL9, designated as *UL8.5*, which is expressed at later times during infection, may act as a regulator of viral initiation by nonproductively occupying the origin at later times of the life cycle (55).

C. UL9 Interacting Proteins

ICP8, the viral SSB, specifically interacts with UL9 in a manner that is dependent on the C-terminal 27 amino acids of UL9 (56,57). However, it is not known whether these residues are sufficient for the interaction. Moreover, there is as yet no data pertaining to the strength of this interaction. The finding that the interaction can be disrupted with increasing salt suggests that it is mediated by charged residues. Through this association, ICP8 greatly enhances the ssDNA- and origin-dependent ATPase activities as well as the DNA helicase activity of UL9 (51,52,58). One characteristic of the effect of ICP8 is that it confers increased processivity on the UL9 helicase in as much as it increases the length of DNA unwound (>2 kb) and induces resistance to template challenge (51,52). The functional stimulation has also provided data related to the stoichiometry of the UL9-ICP8 complex. Thus, maximal stimulation is observed at a 1:1 ratio, indicating that the active species is an equimolar complex and that the stimulation is a consequence of specific protein-protein interaction (51). This is confirmed by the finding that heterologous SSBs (e.g., *E. coli* SSB or RP-A) do not stimulate UL9 (48,51,52). Given that UL9 is dimeric in solution, the ICP8-UL9 complex probably comprises two molecules of each component. Deletion of the C-terminal 27 residues of UL9 prevents its interaction with ICP8 and makes this mutant functionally unresponsive to ICP8 (57). Interestingly, however, this deletion mutant exhibits both ssDNA-dependent ATPase and helicase activities that are greatly elevated compared to wild-type UL9 (50,57). In fact, the properties of this mutant resemble those of ICP8-stimulated UL9 (increased turnover and processivity) and led us to conclude that the function of ICP8 is to neutralize an inhibitory domain of

UL9 (the C-terminal tail) that normally keeps the enzyme in an ineffective ground state (50). This mechanism therefore resembles an allosteric-type mechanism rather than the model that was proposed in which the ssDNA-binding activity of ICP8 tethers UL9 to the substrate to increase its processivity (50,51). In fact, ICP8 that has been modified to prevent it from interacting with the DNA substrate can nevertheless stimulate UL9, confirming that the stimulation is independent of the ssDNA-binding activity of ICP8 (50).

The physiological importance of the ICP8–UL9 interaction comes from experiments that examined viral origin-dependent DNA replication in cell culture. Here, it was found that the C-terminal deletion mutant of UL9 that fails to interact with ICP8 exhibited greatly reduced replication efficiency (57). The complex of ICP8 and UL9 has been proposed to be the active helicase at the origin during the initiation of replication (see Section III.E).

In a search for cellular factors that are involved in the initiation of origin-dependent replication, additional UL9 interacting proteins were identified. One of these is the large, catalytic subunit of DNA polymerase α -primase whose DNA polymerase activity is also increased as a consequence of this association (59). It is possible that DNA polymerase α -primase participates in primer synthesis and elongation at the origin, as is the case with other eukaryotic DNA viruses. This is despite the fact that HSV-1 encodes its own DNA polymerase and primase (Table I). It should be noted, however, that there are arguments against a role for DNA polymerase α -primase. Thus, HSV-1 DNA polymerase mutants that are resistant to aphidicolin and 9- β -D-arabinofuranosyladenosine can replicate in the presence of these inhibitors that block both the DNA polymerase and primase activities of DNA polymerase α -primase, respectively (reviewed in Ref. (60)). Examining the replication of HSV-1 in a temperature-sensitive DNA polymerase α -primase cell line should shed more light on the role of this enzyme in viral replication.

Recently, it was found that UL9 also interacts with molecular chaperones. A two-hybrid analysis with UL9 as the bait identified hTid-1, a DnaJ/Hsp40 family member, in a HeLa cell cDNA library (61). In a separate study, HDJ-1, another human DnaJ/Hsp40 member was shown to functionally interact with UL9 (58). In both cases, human DnaJ increased the association of UL9 with the origin (58,61). This effect was also seen with *E. coli* DnaJ indicating a conserved mechanism of action, and was augmented by the chaperone Hsp70/DnaK (58). Moreover, it was shown that the chaperones also stimulated the origin-dependent ATPase activity of UL9 but had no effects on its origin-independent activities (i.e., DNA-independent ATPase, ssDNA-dependent ATPase, and nonspecific DNA helicase activities) (58).

Taken together, these data indicate that chaperones increase the interaction of UL9 with the origin without affecting its other activities. These findings are in agreement with other DNA replication systems in which chaperones have been shown to increase the interaction of initiator proteins with origins, in most cases by modulating the oligomerization state of the initiator protein (62,63). In the case of UL9, it is not clear what the mechanism of stimulation is, but preliminary data from our laboratories indicate that the chaperones may act by converting aggregates of UL9 into more active, presumably dimeric species (Tanguy Le Gac and Boehmer, unpublished observations).

It should be noted that the above-mentioned two-hybrid screen identified two additional UL9 interacting proteins, a neural F Box 42-kDa protein and the dynactin complex component p150^{Glued} (61). However, the relevance of these associations remains to be established.

Several of the HSV-1 DNA replication proteins also interact with UL9. These associations may serve to recruit these factors to the origins in order to assemble a replisome. This will be discussed in Section III.E.

D. The Role of Host Proteins

Additional proteins that may be involved in the initiation of origin-dependent replication are implicated by their association with ICP8. Alternatively, factors that interact with the origin or affect origin structure may also contribute to this process. Thus, in the former category, we identified high mobility group protein-1 (HMG-1) as a cellular protein that interacts with ICP8 (Boehmer, unpublished observation). HMG-1 binds with high affinity to the minor groove of A+T-rich DNA and induces structural alterations such as DNA bending and DNA unwinding at these sites (64). Consequently, in the context of origin-specific DNA unwinding, it is possible that HMG-1 complements the action of UL9 and ICP8 in binding and distorting the A+T region at the HSV-1 origins.

OF-1 is a cellular factor whose binding sites overlap with those of UL9 that has been implicated as being required for viral replication (37). OF-1 is a tetramer composed of two heterodimers of 73 and 90 kDa, the smaller of which exhibits DNA-binding activity. This factor was shown to bind tightly to *oriS* box I (K_d of 24 nM) (65). OF-1 and UL9 appear to compete for binding in as much as when both proteins are present, the affinity of UL9 for box I decreases while that of OF-1 increases (65). Moreover, OF-1 exhibits a 2.5-fold higher affinity toward the lower strand of box I (65). These findings suggest an interesting scenario in the context of *oriS** that may regulate the initiation process. Thus, UL9 could form a stable complex with the upper strand while the lower strand is bound by OF-1.

The role of transcription factors and the mechanics of transcription in generating negative supercoiling have been discussed earlier.

E. Model for Initiation

We are interested in the mechanism by which UL9 targets the origins, unwinds the origins, and leads to the assembly of a replisome to initiate DNA replication. The recognition of origin sequences by UL9 was described in Section III.A, as was its nonspecific DNA helicase activity.

Here we discuss the interaction of UL9 with the origin to generate single-stranded template DNA. Thus, UL9 by itself has been shown to induce structural changes in the origin in an ATP-independent manner, including the introduction of nuclease and chemical sensitive sites in the A + T-rich region (66). This action of UL9 requires supercoiled DNA that may be generated by divergent transcription from promoters flanking the origins, and is stimulated by molecular chaperones (Hsp40/Hsp70) (58). Examination of structures formed at the origin by electron microscopy shows that the DNA at the origin is bent by $\sim 90^\circ$ presumably due to protein-protein interactions between origin-bound UL9 dimers on either side of the A + T-rich region (67). A further indication that UL9 plays an active role at the origin is shown by the stimulation of its ATPase activity by origin-containing DNA (58,68). The only evidence that UL9 promotes a reaction consistent with origin unwinding is provided by electron microscopy in which UL9 was seen to generate large (up to 1 kilobase) stem loops extruded from the base of a UL9 complex (67). Subsequent studies revealed that the viral SSB, ICP8, plays a major role in destabilizing the origin. Thus, in the absence of ATP, the complex of UL9 and ICP8 initially interacts with the A + T-rich region that may be destabilized by the action of ICP8 while UL9 targets the flanking recognition sites (69). Although these data indicate that ICP8 participates very early in the initiation process, we believe that its specificity is nevertheless governed by the origin-binding activity of UL9.

Initial distortion of the origin may occur in the absence of ATP hydrolysis. However, actual unwinding would be expected to require the DNA helicase activity of UL9 and ATP hydrolysis. In this regard, UL9 and ICP8 have been shown to unwind model origin-containing substrates in which the A + T-rich region was replaced with ssDNA or by noncomplementary sequences to create a bubble-type substrate (70,71). More recently, it was shown that UL9 and ICP8 could promote ATP-dependent unwinding of a linear minimal origin that proceeds via the putative *oriS** intermediate (72). Consistent with these studies, electron microscopic evidence illustrates how UL9 and ICP8 can promote efficient unwinding of supercoiled origin containing plasmids (73).

Unwinding of the DNA at the origin is the first step in the initiation process. As described earlier, in HSV-1 this is achieved by the action of UL9 and ICP8. Subsequent steps involve recruitment of other replication factors to assemble a replisome. In HSV-1 we anticipate that this is guided by specific protein–protein interactions. Thus, UL9 bound at the origin may be able to recruit the viral helicase–primase via an interaction with the UL8 subunit (74). Loading of one molecule of helicase–primase on each side of the A+T-rich region would establish a replication fork. The helicase would unwind the DNA ahead of the polymerase, while the primase would initially prime both leading- and lagging-strand DNA synthesis and subsequently lagging-strand synthesis for each DNA replication fork. As mentioned in Section III.C, the interaction between UL9 and DNA polymerase α -primase may recruit an alternative primase. The viral DNA polymerase may be recruited to the origin by the interaction of the UL30 catalytic subunit with UL8 or by the interaction of the DNA polymerase clamp (UL42) with UL9 (75,76).

A model for origin-dependent initiation is shown in Fig. 3. For simplicity, the involvement of some factors (e.g., OF-1, DNA polymerase α -primase, or HMG-1) discussed in the preceding section has not been considered. Moreover, we have not accurately depicted the formation of *oriS** in our model. Rather, we show that divergent transcription results in extrusion of the origin as a cruciform structure that may be efficiently recognized by UL9 (upper strand) and OF-1 (lower strand). Finally, it should be noted that the products of theta replication are catenated circles. These would have to be resolved by the action of a type II topoisomerase. In this context, studies with a topoisomerase II inhibitor have shown that this activity is required for viral replication in cell culture, presumably to resolve catenated products of theta replication (77).

IV. Elongation

A. Replication Fork Enzymes

All herpes viruses encode a conserved set of proteins that are active at the DNA replication fork. As previously mentioned, these are similar in function to enzymes that promote DNA synthesis in other prokaryotic and eukaryotic systems. The properties of the HSV-1 DNA replication fork proteins are summarized in Table I, and have previously been reviewed extensively (2,60).

The viral DNA polymerase is a heterodimer of the UL30 catalytic subunit and the UL42 processivity factor. UL30 also exhibits 3′–5′ proofreading

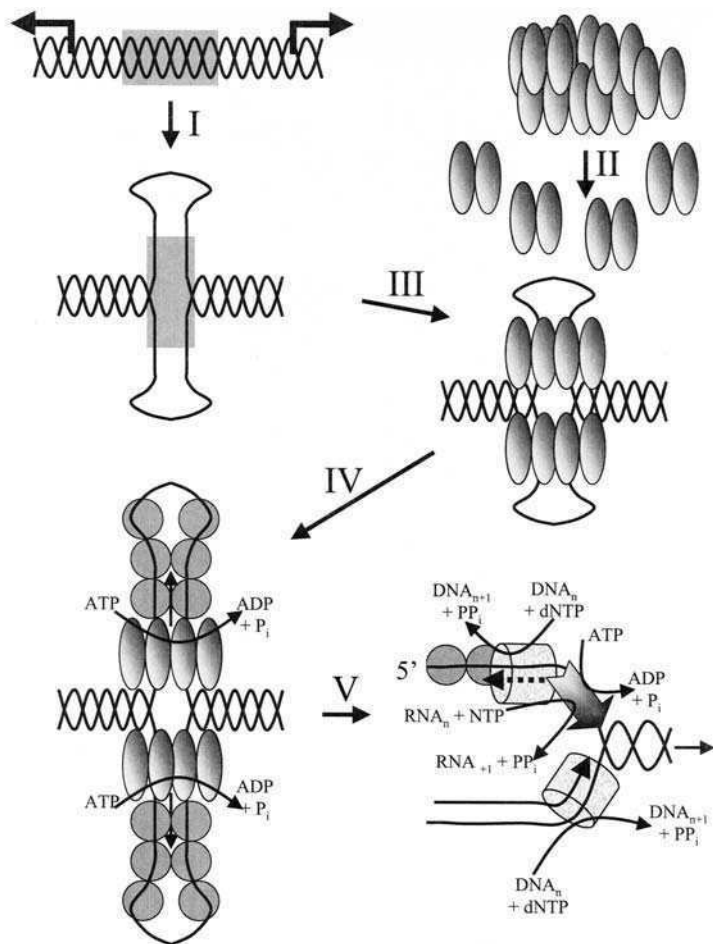


Fig. 3. A model for origin-dependent initiation of DNA replication in HSV-1. I, divergent transcription from promoters flanking the origin (denoted by the divergent arrowheads) generates supercoiling that results in the extrusion of the origin as a cruciform-type structure (*oriS**). II, molecular chaperones convert UL9 aggregates into dimeric UL9. III, high-affinity binding of dimeric UL9 to its recognition sites in *oriS**. IV, ICP8 is recruited to *oriS** via its interaction with UL9 to stabilize distorted regions of the origin. ATP-dependent helicase action of the UL9-ICP8 complex generates ssDNA loops that are stabilized by ICP8 binding. V, collapse of the ssDNA loops formed by the action of the UL9-ICP8 complex and assembly of a replisome. Only one-half of a bidirectional fork is shown. Recruitment of the HSV-1 DNA polymerase and helicase-primase is guided by protein-protein interactions with UL9 and ICP8. UL9, ICP8, HSV-1 DNA polymerase, and helicase-primase are denoted by the ovals, circles, cylinder, and arrow, respectively. Host factors involved in this process have been omitted. The gray shaded region depicts the UL9-binding site/origin. ATP hydrolysis, DNA (solid line), and RNA (dashed line) synthesis are as indicated.

exonuclease and RNaseH activities. UL30 interacts with UL42 via its C-terminal domain, where deletion of as little as 35 amino acids disrupts the interaction (78). The site of interaction in UL42 is less well defined. UL42 stimulates the polymerase by decreasing the dissociation of the heterodimer from the primer template as well as by slightly increasing its association rate (79,80). The net effect is to increase the processivity of DNA synthesis from ~ 1600 nucleotides for UL30 alone to ~ 5000 nucleotides for the holoenzyme (80). In doing so, UL42 has no effect on the translocation rate ($\sim 30 \text{ s}^{-1}$) of the polymerase (79). It is interesting to note that UL42 exhibits dsDNA-binding activity (81). Yet, it functions as a sliding clamp, facilitating processive DNA synthesis (81). In this regard, UL42 behaves similarly to other DNA polymerase processivity factors such as eukaryotic PCNA (82). In fact, UL42 shows structural homology to one protomer of the PCNA toroid, but since it functions as a monomer, it would not be able to encircle a DNA duplex (83). The mechanism of action of UL42 is not solely dependent on tethering the polymerase to the template since the DNA polymerase holoenzyme–DNA complex is significantly more resistant to disruption by salt than complexes formed with the individual subunits (84). Unlike other DNA polymerase clamps, such as PCNA, UL42 is not loaded onto the DNA by elaborate, ATP-dependent clamp loader assemblies (82). Its mechanism may therefore be similar to how *E. coli* thioredoxin increases the processivity of T7 DNA polymerase.

ICP8 is a multifunctional protein that interacts functionally and/or physically with the remaining viral DNA replication proteins. Its primary function is that of an SSB. Its role in the initiation of origin-dependent replication and association with UL9 was discussed in Section III. ICP8 interacts nonspecifically with ssDNA, with a binding site size of ~ 10 nucleotides and an association constant of $\sim 0.5 \times 10^6 \text{ M}^{-1}$ (85). Compared with other SSBs, ICP8 exhibits relatively weak binding and poor cooperativity (ω of ~ 15) (85). These properties may be related to its role as a recombinase (see Section VI). In this regard, it should also be noted that ICP8 stretches DNA (85).

The helicase–primase is a heterotrimeric complex although only UL5 and UL52 are required for catalytic activity (86). The enzyme promotes DNA unwinding in the 5′–3′ direction at a rate of $\sim 60 \text{ bases s}^{-1}$ with concomitant ATP hydrolysis (87). Its helicase activity, like that of UL9, is stimulated by ICP8 (87–89). However, rather than conferring increased processivity, ICP8 appears to recruit the helicase–primase to the template via specific protein–protein interactions that are mediated by UL8 (88–90). The primase activity synthesizes short ~ 10 -nucleotide RNA primers (91). This activity is stimulated by the UL8 subunit, which is also required for primase activity on ICP8-coated templates (88,89,92). Taken together, it appears that UL8

acts as a loading protein for the helicase–primase, particularly on ICP8-coated DNA, similar to the functions performed by *E. coli* DnaC and T4 gp59 (88,89).

B. *In Vitro* Reconstitution of Viral DNA Replication

Several successful attempts have been made to reconstitute the rolling-circle phase of HSV-1 replication *in vitro*. Early attempts focused on DNA substrates based on M13 DNA, and nicked plasmids using fractionated extracts from HSV-1-infected cells that were enriched in the viral DNA polymerase, helicase–primase, and ICP8 (93,94). These studies provided evidence for long leading- and lagging-strand DNA synthesis corresponding to multiple unit-length substrates. This was also supported by visualization of long products by electron microscopy (93,94). Similar studies were also performed with a partially purified complex comprising the DNA polymerase, helicase–primase, and ICP8 fractionated from insect cells infected with recombinant baculoviruses (95). More recently, it has been possible to reconstitute HSV-1 rolling-circle replication with purified proteins using a minicircular DNA substrate (96). In this system, both leading- and lagging-strand DNA replication is detected although the efficiency of primer synthesis is low, giving rise to long Okazaki fragments (~ 3 kilobases) (96). It is possible that inefficient priming may be attributed to the small size (70 bases) of the minicircular DNA substrates used (97). Interestingly, in this system ICP8 is dispensable and even inhibitory at high concentration, although its effect could be reversed by supplementing with UL8 subunit which has previously been reported to mediate the interaction of the helicase–primase with ICP8-coated DNA (96,88,89).

The data suggest a simple model in which two molecules of the HSV-1 DNA polymerase (UL30 and UL42) catalyze leading- and lagging-strand DNA synthesis. Replication fork progression is determined by the helicase activity of the helicase–primase complex (UL5, UL52, and UL8). In contrast to other systems, where the DNA polymerase affects the translocation rate of the helicase, there appears to be no functional interaction between the DNA polymerase and helicase–primase (87). ICP8 is present at the fork to melt secondary structure and to stabilize ssDNA, and to stimulate DNA unwinding by the helicase–primase. RNA primers synthesized by the helicase–primase are presumably removed by the action of the RNaseH that is associated with UL30. This model is depicted in the final step of Fig. 3. Two necessary activities that are not encoded by the virus are DNA ligase and topoisomerase. These are presumably provided by the host.

V. Interaction of the HSV-1 DNA Replication Machinery with DNA Damage

The sole enzymes encoded by HSV-1 that play somewhat of a role in DNA repair are a uracil DNA glycosylase (UDG) and a deoxyuridine triphosphatase (dUTPase) (reviewed in Ref. (60)). Both enzymes presumably contribute to the fidelity of DNA replication by preventing the occurrence of uracil in newly replicated DNA. UDG accomplishes this task by removing uracil from DNA, either as a consequence of misincorporation or from spontaneous deamination of cytosine, while the dUTPase reduces the pool of dUTP that would be available for incorporation by the DNA polymerase. It is not known how abasic sites that remain after the action of the glycosylase are repaired. Presumably, this is accomplished by the host base excision repair machinery since the virus lacks such enzymes. In case such sites remain unrepaired, they may either stall advancing replication forks, or lead to the introduction of single-strand breaks as a result of β -elimination. These assumptions are consistent with the structure of replicating viral DNA where single-stranded regions and nicks are prevalent (98,99). Interruptions in the continuity of the DNA chain can result in dsDNA breaks that ultimately become substrates for repair by homologous recombination as described in Section VI.

As an alternative to repairing DNA damage, it is plausible that HSV-1 has adapted its replication strategy to cope with certain types of damage. Consequently, we have examined the ability of the viral replication machinery to interact with sites of DNA damage. Most of our studies have focused on DNA modified with the antitumor drug *cis*-diamminedichloroplatinum (II) or cisplatin. Cisplatin is used in the treatment of a variety of tumors, notably ovarian and testicular cancers, as well as for primary and recurrent chemotherapy of certain brain tumors. It is activated into a potent electrophile at intracellular chloride concentrations, thereby allowing it to react with bases in DNA among others. The most common adduct ($\sim 65\%$) formed by the interaction of cisplatin with DNA is the 1–2 d(GpG) intrastrand cross-link in which two adjacent guanines are cross-linked. This lesion induces severe structural alterations in the DNA, bending the DNA backbone by $\sim 40^\circ$ (100). Consequently, DNA damaged in this manner may be considered as a model for DNA structure distorting lesions. The most important biological consequences of this lesion are to inhibit both replication and transcription, and to induce programmed cell-death. However, replication past this adduct increases mutation frequency that has been reported to play a role in the induction of secondary tumors (101).

A. Effect of DNA Damage on the HSV-1 DNA Polymerase

DNA polymerases can be classified into six main families based upon phylogenetic relationships with *E. coli* pol I (Class A), *E. coli* pol II (Class B), *E. coli* pol III (Class C), euryarchaeotic pol II (Class D), human pol β (Class X), and *E. coli* pol UmuC/DinB and eukaryotic RAD 30/XP-V (Class Y) (102). The B family includes replicative DNA polymerases such as HSV-1 polymerase, bacteriophage T4 polymerase, and eukaryotic pol α , δ , and ϵ . *In vitro* elongation catalyzed by replicative DNA polymerases appears to be blocked by helix-distorting or bulky DNA lesions, such as the cisplatin-induced 1–2 d(GpG) intrastrand cross-link (Pt-d(GpG) adduct) (101).

As expected for a DNA polymerase belonging to the B family, UL30 elongates ssDNA templates containing a unique Pt-d(GpG) adduct only to the base preceding the first G of the lesion (103) (Fig. 4A). However, substitution of the divalent metal ion Mg^{2+} , which is probably used for catalysis *in vivo*, with Mn^{2+} allows UL30 to promote quantitative translesion DNA synthesis (Fig. 4B). The ability of Mn^{2+} ions to confer bypass of the cisplatin lesion is

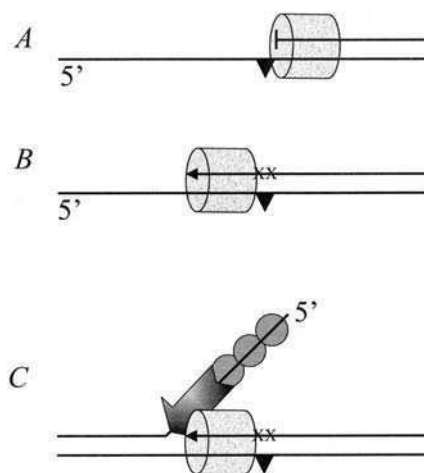


Fig. 4. Interaction of HSV-1 replication proteins with damaged DNA. (A) In the presence of Mg^{2+} as catalytic divalent cation, UL30 DNA polymerase stalls at the base preceding the cisplatin cross-linked d(GpG) dinucleotide. (B) In the presence of Mn^{2+} as catalytic divalent cation, UL30 DNA polymerase promotes translesion synthesis, resulting in misincorporation opposite the lesion. (C) Model for helicase-mediated bypass of a lesion. Protein–protein interactions between SSB (ICP8), helicase (helicase–primase), and DNA polymerase (UL30) results in mutagenic translesion synthesis. UL30 DNA polymerase, ICP8, and helicase–primase are denoted by the cylinder, circles, and arrow, respectively. The position of the Pt-d(GpG) lesion is indicated by the triangle. XX and the arrowhead denote misincorporation opposite the cisplatin cross-linked d(GpG) dinucleotide and the direction of DNA synthesis, respectively.

not observed with other replicative DNA polymerases of the B family, such as bacteriophage T4 or polymerases δ and ε . Nevertheless, for these enzymes, Mn^{2+} induces the incorporation of one nucleotide opposite the first (3') guanine of the Pt-d(GpG) adduct. Translesion replication by UL30 leads to the incorporation of mismatched bases, with the preferential incorporation of dAMP opposite the 3' guanine of the lesion.

To gain insight into the mechanism of translesion synthesis, we observed that substitution of $MgCl_2$ with $MnCl_2$ greatly inhibits the 3'-5' proofreading exonuclease of UL30 but has a far lesser effect on that of T4 DNA polymerase. Mn^{2+} was also observed to induce a conformational change in the structure of UL30 bound to the cisplatin-damaged substrate (103). These results suggest that Mn^{2+} can influence the translesion synthesis capacity of the catalytic subunit of the HSV-1 replicative polymerase, both by reducing its proofreading activity and by inducing conformational changes in its structure when bound to DNA. The availability of UL30 deficient in its 3'-5' exonuclease should help to establish the role of the proofreading activity in the damage bypass capacity of the HSV-1 DNA polymerase. In this regard, it should be noted that the DNA polymerases of the Y family, which can replicate through structure-distorting lesions even in presence of Mg^{2+} *in vitro* (104), not only lack an associated 3'-5' exonuclease but also have a structure harboring a more "open" active site compared to the replicative polymerases of the B family. Therefore, the specific effect of Mn^{2+} on UL30 may be the consequence of increased accessibility to Mn^{2+} of its catalytic and exonuclease domains compared to those of the other polymerases of the B family. The crystal structure of UL30 should help to verify such a possibility.

The HSV-1 DNA polymerase processivity factor UL42 does not increase the capacity of the enzyme to replicate past the Pt-d(GpG) adduct (Villani *et al.*, unpublished data). Preliminary results from our laboratories indicate that UL42 appears to inhibit translesion synthesis even in the presence of Mn^{2+} . This is surprising considering that UL42 reportedly stabilizes the association of UL30 with the primer template and would therefore be expected to facilitate translesion synthesis, assuming that the lesion induces dissociation of the polymerase. This finding is also in contrast to the effect of other DNA polymerase clamps. Thus, PCNA was found to induce translesion DNA synthesis by DNA polymerase δ (105).

B. Effect of DNA Damage on the HSV-1 DNA Helicases

Among the components of the replication, repair, and recombination apparatus, the first that may encounter a site of DNA damage are the DNA helicases. Thus, a complete understanding of the effect of DNA lesions on genome transactions requires a biochemical analysis of their interaction with this class of proteins. It was found that a Pt-d(GpG) adduct significantly

reduced, but did not abolish, the helicase activity of the HSV-1 UL9 helicase, but only when it was present on the strand along which the protein translocated (54). There was also concomitant inhibition of the DNA-dependent ATPase activity of UL9, indicating that the lesion induced dissociation of the helicase rather than merely stalling its progress along the DNA (54). Addition of the viral SSB, ICP8 greatly stimulated the capacity of the helicase to unwind cisplatin-damaged DNA (54). Furthermore, the stimulation appeared to be the result of the functional and physical interaction that is known to exist between UL9 and ICP8, and not because of a preferential binding of ICP8 at the site of the adduct (54). Results from a subsequent study showed that ICP8 stimulates the DNA helicase activity of UL9 by increasing its processivity, thus preventing its dissociation and facilitating its translocation along DNA and through regions of secondary structure (51). Based on these findings, it is tempting to speculate that ICP8 enables UL9 to bypass the Pt-d(GpG) lesion by tethering it to the DNA substrate, thereby preventing its dissociation. However, a later study showed that the increase in processivity by ICP8 is mediated by protein-protein interactions rather than a tethering mechanism (50).

The helicase activity of UL9 has also been examined on a partial DNA duplex complexed with bis-peptide nucleic acid (PNA) that results in a stable clamp within the ssDNA-loading site of the DNA helicase. These experiments showed that UL9 was inhibited by the presence of the PNA clamp indicating that it forms a structure that either prevents binding of UL9 to the substrate or inhibits its translocation to enable subsequent unwinding of the duplex region (106). In any case, the results indicate that PNA presumably form stable complexes with target regions in duplex DNA that may prevent helicase action.

DNA unwinding promoted by the viral replicative DNA helicase-primase is also inhibited by a site-specific Pt-d(GpG) adduct, but only when located on the strand along which the helicase translocates (90). Similar to the situation with UL9, ICP8 also enhances the bypass of Pt-d(GpG) adducts by the helicase-primase. However, in contrast to what was found for UL9, ICP8 stimulates helicase-primase by recruiting the enzyme to the DNA rather than by increasing its processivity (90). Thus, these studies suggest that specific protein-protein interactions between an SSB and two viral helicases with distinct roles in DNA replication mediate substantial unwinding of substrates containing a bulky intrastrand DNA lesion.

C. A Model for Lesion Bypass

Specific protein-protein interactions have been demonstrated between the viral DNA helicases, DNA polymerase, and SSB, and concomitant replication of the leading and lagging DNA strands requires the coordinate

action of these enzymes (see Section IV.B). Therefore, the capacity of a replicative DNA helicase to unwind past a lesion, in conjunction with the SSB, while the associated polymerase cannot, would result either in stalling of the whole replication complex or in the uncoupling of leading- and lagging-strand DNA synthesis. To prevent this, it may be possible that the capacity of a helicase to unwind damaged DNA allows some lesion bypass by the cognate polymerase (helicase-mediated bypass). This model is depicted in Fig. 4C, where the ICP8-stimulated replicative helicase–primase allows the DNA polymerase to bypass the lesion in a process that is dictated by protein–protein interactions. The establishment of an experimental system with the HSV-1 replication proteins and appropriate damaged DNA templates *in vitro* may shed some light on this possibility.

VI. Homologous Recombination

A. Recombination in HSV-1

During the viral life cycle, the HSV-1 genome undergoes several recombination reactions. Thus, circularization of the viral genome is an early event that appears to be independent of virus-encoded functions (5,7). In contrast, isomerization of the viral genome, in which the U_L and U_S subgenomic elements invert with respect to each other, and the occurrence of branched DNA intermediates are recombination events that are linked to viral DNA replication and potentially mediated by viral factors (107,108).

Inversion of U_L and U_S with respect to each other generates equimolar amounts of four isoforms: P (prototype), I_{LS} (inversion of L and S), I_L (inversion of L), I_S (inversion of S) (Fig. 1). The significance of this phenomenon is unclear since there appears to be no growth advantage to any one form and viruses with “frozen” genome configurations are equally viable as wild-type virus (109).

While the P and I_{LS} isoforms may be generated by cleavage at alternate L–S junctions, an explanation for the origin of the other two forms is not simple. One appealing hypothesis states that homologous recombination between repeats at L–S junctions generate a “double rolling-circle” replication template in which the inversions have already taken place. Replication of such templates would therefore yield the four isoforms (109). Regardless of the precise mechanism, it is thought that the viral *a* sequences play a role in this process by mediating a homologous recombination event. In this regard, it was recently shown that a host-encoded endonuclease, endonuclease G, cleaves within *a* sequences, thereby introducing dsDNA breaks at these sites and conferring recombination hot-spot activity on

a sequences (110). Alternative dsDNA-break sites may function in a similar fashion to induce genome isomerization.

The initial phase of homologous recombination involves processing of duplex DNA to generate ssDNA tails that are utilized in the subsequent-pairing reaction. It is easier to envisage this processing occurring at dsDNA breaks although other more complicated scenarios may exist. dsDNA breaks may arise as a result of direct DNA damage, collapse of DNA replication forks at sites of DNA damage (e.g., nicks), or may be introduced by endonucleolytic cleavage such as that mediated by endonuclease G at *a* sequences (110). Multiple mechanisms exist whereby ssDNA may be generated. Thus, helicase action may generate ssDNA tails. In HSV-1, this may be accomplished by the action of UL9, or helicase–primase. However, both UL9 and helicase–primase require ssDNA-loading sites to promote DNA unwinding and may therefore not be appropriate for this task by themselves. The helix-destabilizing activity of ICP8, either by itself or in conjunction with one of the aforementioned helicases, is an alternative to generating ssDNA tails. Finally, exonucleolytic processing at dsDNA breaks, such as by the viral *UL12*-encoded nuclease or even the host Mre11/Rad50/Xrs2 complex (111), may also generate ssDNA tails.

B. Strand Invasion

Invasion of a homologous duplex by ssDNA followed by establishment of a replication fork at that site would certainly account for the highly branched intermediates observed at later times during viral replication. In other systems, strand invasion is promoted by recombinases that include *E. coli* RecA, bacteriophage T4 UvsX, and eukaryotic Rad51 (see Ref. (112), for a review on *E. coli* RecA). These proteins form nucleoprotein filaments on ssDNA in an ATP-dependent manner (albeit without requiring ATP hydrolysis) followed by a search for homology- and polar-strand invasion into a homologous duplex, resulting in the formation of displacement loop (D-loop) (112). A scheme depicting this reaction is shown in Fig. 5. We have recently demonstrated that in HSV-1 this role is performed by ICP8, the viral SSB (113). This is interesting in as much as ICP8 is primarily an SSB, albeit it may be considered as a multifunctional protein on the basis of its interaction with other viral replication proteins. Perhaps this reflects the need for the virus to condense multiple functions into limited-coding sequence. Although the product of ICP8-catalyzed strand invasion is the same as those formed by RecA-type recombinases, ICP8 achieves this without a requirement for ATP. In this regard it is interesting to note that several thermodynamic parameters governing the interaction of ICP8 with ssDNA are closer to those of RecA and other recombinases than they are to other SSBs (85). Thus, ICP8 stretches ssDNA, possibly to facilitate the search for

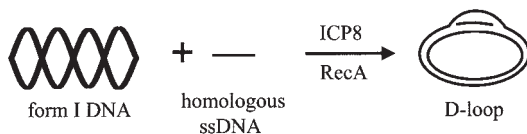


Fig. 5. Formation of a displacement loop. Schematic representation of strand invasion promoted by ICP8 or RecA in which assimilation of ssDNA into homologous form I DNA results in the formation of a D-loop.

homology, forms complexes with ssDNA that are stable at high salt concentrations, and exhibits both weaker and lower cooperativity ssDNA binding than other SSBs (85). It should be noted that ICP8-mediated strand invasion depends on the free energy of supercoiling within the acceptor DNA. During viral replication, this may be provided in transcriptionally active regions of the viral genome. Since there is no apparent bias toward the assimilation of a particular end (i.e., 3' vs. 5'), ICP8-catalyzed D-loop formation would result in at least a fraction of structures in which an invading 3' end would function as a primer for DNA synthesis. Consequently, ICP8-mediated strand invasion would allow reestablishment of a replication fork at the site of assimilation, as has been proposed for the restart of DNA synthesis in *E. coli* following RecA-mediated strand invasion and PriA-mediated primosome assembly (114). A proposed model of how this occurs in HSV-1 is described in Section VI.D.

C. Strand Exchange

A further means of promoting homologous recombination is by strand exchange in which strands from homologous DNA molecules are swapped. As with strand invasion, this reaction initiates with nucleoprotein filament formation on ssDNA, which then searches for homology in an acceptor DNA molecule followed by intermolecular pairing between complementary strands (heteroduplex DNA formation). As shown in Fig. 6, this process may involve exchange of strand between a duplex donor and an ssDNA acceptor (three-stranded reaction) or between two duplex molecules (four-stranded reaction). Strand exchange is catalyzed by RecA-type recombinases although extensive heteroduplex DNA formation also requires branch migration that is catalyzed by specialized motor-protein complexes. In *E. coli*, this task is performed by the RuvAB complex in which a RuvA tetramer binds to the crossover point (Holliday junction), and RuvB hexamers on either side of it to translocate the DNA to generate heteroduplex DNA (115). In HSV-1, strand exchange appears to be mediated by a mechanism similar to that adopted by bacteriophage T7 involving the SSB and replicative helicase that are also essential for DNA replication (116). The reaction follows

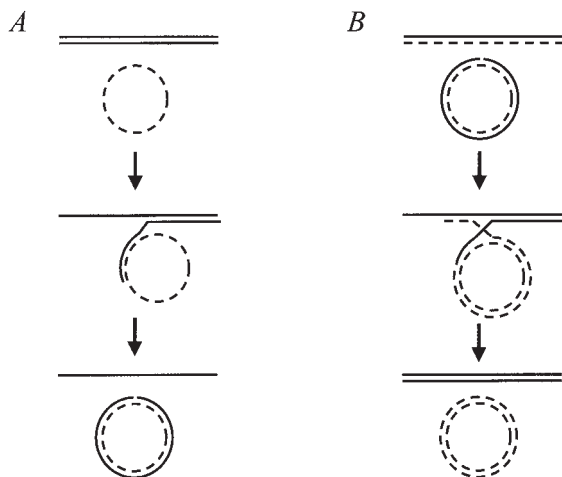


Fig. 6. Strand exchange reactions. Diagrammatic representation of three (A) and four (B) stranded reactions. In each case, a complementary strand from the linear duplex is transferred onto the circular molecule. In the four-stranded reaction, this involves transfer of complementary strands from each substrate molecule. The resulting molecules are nicked circular and linear ss or dsDNA in the three- and four-stranded reactions, respectively. Dashed lines have been added to distinguish DNA strands.

a single-strand annealing and helicase-mediated heteroduplex extension mechanism. In HSV-1, ICP8 uses its helix-destabilizing activity to generate ssDNA ends. Subsequently, intermolecular pairing of homologous DNA is promoted by the reannealing activity of ICP8. Heteroduplex DNA intermediates formed in this fashion are further processed by helicase-mediated branch migration catalyzed by the replicative DNA helicase-primase (117). Although the individual steps of this reaction can be dissected, this process appears to be tightly coupled, presumably due to the known association between ICP8 and helicase-primase (88,89). It should be noted that, as in the case of strand invasion, ssDNA may also be generated by exonucleolytic processing of DNA ends such as by the *UL12*-encoded nuclease. A model of strand exchange involving HSV-1 ICP8 and helicase-primase is shown in Fig. 7.

D. Model for Recombination-Mediated Replication

To account for the prevalence of branched replication intermediates, we propose a model in which ICP8-mediated strand invasion is coupled to DNA synthesis as depicted in Fig. 8. The salient features of the model as shown are: introduction of a dsDNA break, such as by the action of endonuclease G at *a* sequences (110); generation of ssDNA tails by the helix-destabilizing activity

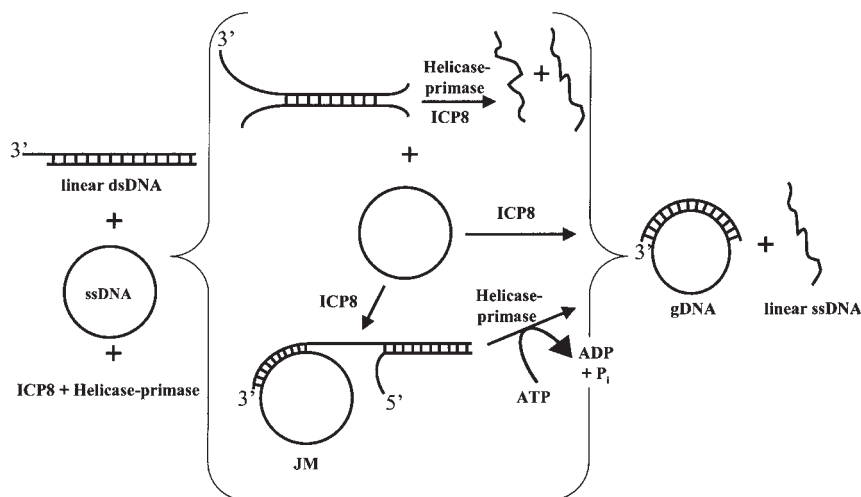


Fig. 7. A model for strand exchange promoted by ICP8 and helicase-primase. The overall reaction converts linear duplex and circular ssDNA substrates into gapped circular (gDNA) and linear ssDNA products. This may be achieved via certain intermediates as indicated in the bracket. Thus, ICP8 interacts with the donor duplex DNA to promote partial unwinding. This allows it to pair the complementary strand to the circular ssDNA acceptor to form a joint-molecule (JM) intermediate. Displacement of the noncomplementary 5'-terminal strand acts as a loading site for the helicase-primase. ATP-dependent helicase action and concomitant pairing results in branch migration, which completes strand exchange. Alternatively, in the presence of helicase-primase, complete unwinding of the donor dsDNA allows it to pair directly with the acceptor DNA without the need for branch migration.

of ICP8, possibly in conjunction with either UL9 or helicase-primase, or by exonucleolytic processing of the ends (117); assimilation of ssDNA into a locally supercoiled homologous duplex, resulting in the formation of a D-loop (113); assembly of a replication fork at the site of assimilation guided by protein-protein interactions or by as yet unidentified affinity of particular viral replication factors for D-loops. This assembly may be driven by the interaction of ICP8 with the helicase-primase (via UL8) which in turn could recruit the viral DNA polymerase via the UL8-UL30 interaction (75,88,89). The net effect is the establishment of leading- and lagging-strand DNA synthesis that would result in branched DNA replication intermediates.

VII. Concluding Remarks

The aim of this review was to provide insight into the details of DNA transactions in HSV-1, to illustrate its use as a model system to study these

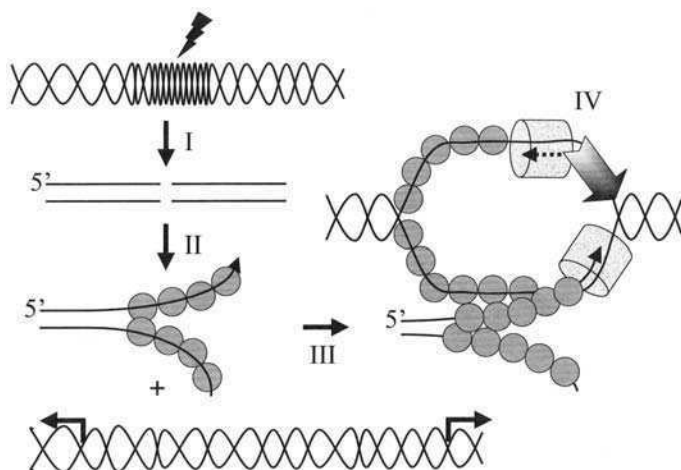


Fig. 8. A model for recombination-mediated DNA replication in HSV-1. I, generation of dsDNA break by the action of endonuclease G at *a* sequences. II, ICP8-mediated destabilization of the ends of the broken DNA. III, invasion of a 3'-terminal strand (denoted by the arrowhead) into a locally supercoiled (denoted by the divergent arrowheads) homologous duplex. IV, assembly of a replisome guided by protein-protein interactions or by structure-specific recognition of the D-loop. The invading 3'-terminal strand (denoted by the arrowhead) acts as a primer for leading-strand DNA synthesis, while helicase-primase synthesizes a primer for lagging-strand DNA synthesis. ICP8, HSV-1 DNA polymerase, and helicase-primase are denoted by the circles, cylinder, and arrow, respectively. The bolt, dashed line, and arrowhead denote endonuclease G cleavage, RNA primer, and the direction of DNA synthesis, respectively. The anisomorphous DNA structure at the *a* sequence is depicted by the tighter-wound helix.

processes, and to highlight some of the specific questions that have been applied to this system. Thus, the relative simplicity of the system has allowed it to be used to: (a) study the mechanism by which DNA replication initiates an origin, (b) examine the mechanism of nascent DNA chain elongation, (c) examine the interaction of essential replication proteins with damaged DNA, and (d) investigate the possible roles of recombination in reestablishing a replication fork.

It is noteworthy that HSV-1 is an amenable system in which to study recombination-mediated DNA replication, particularly in light of the finding that almost every organism uses a recombination process at some stage of its replication cycle, usually as a repair mechanism (118). Likewise, HSV-1 has proven to be an excellent system to examine the interaction of essential replication factors with damaged DNA, particularly to study translesion DNA synthesis. This is due, in part, to the fact that the viral helicase-primase is the replicative helicase, an enzyme that is still elusive in eukaryotic cell systems.

In future, the availability of large amounts of all the essential HSV-1 DNA replication and recombination proteins could provide a useful tool for structural and mechanistic studies of genome transactions on both undamaged and damaged DNA.

ACKNOWLEDGMENTS

Work in the authors' laboratories is supported by grants GM62643 from the National Institutes of Health and BM 022 from the Florida Biomedical Research Program (to P. E. B.) and 4373 from the Association pour la Recherche sur le Cancer (to G. V.).

REFERENCES

1. P. E. Boehmer and A. V. Nimonkar, Herpes virus replication, *IUBMB Life* **55**, 13–22 (2003).
2. I. R. Lehman and P. E. Boehmer, Replication of herpes simplex virus DNA, *J. Biol. Chem.* **274**, 28059–28060 (1999).
3. K. Dohner, A. Wolfstein, U. Prank, C. Echeverri, D. Dujardin, R. Vallee and B. Sodeik, Function of dynein and dynactin in herpes simplex virus capsid transport, *Mol. Biol. Cell* **13**, 2795–2809 (2002).
4. P. M. Ojala, B. Sodeik, M. W. Ebersold, U. Kutay and A. Helenius, Herpes simplex virus type 1 entry into host cells: reconstitution of capsid binding and uncoating at the nuclear pore complex in vitro, *Mol. Cell. Biol.* **20**, 4922–4931 (2000).
5. K. Umene and T. Nishimoto, Replication of herpes simplex virus type 1 DNA is inhibited in a temperature-sensitive mutant of BHK-21 cells lacking RCC1 (regulator of chromosome condensation) and virus DNA remains linear, *J. Gen. Virol.* **77**, 2261–2270 (1996).
6. D. A. Garber, S. M. Beverley and D. M. Coen, Demonstration of circularization of herpes simplex virus DNA following infection using pulsed field gel electrophoresis, *Virology* **197**, 459–462 (1993).
7. X. D. Yao, M. Matecic and P. Elias, Direct repeats of the herpes simplex virus a sequence promote nonconservative homologous recombination that is not dependent on XPF/ERCC4, *J. Virol.* **71**, 6842–6849 (1997).
8. E. S. Mocarski and B. Roizman, Structure and role of the herpes simplex virus DNA termini in inversion, circularization and generation of virion DNA, *Cell* **31**, 89–97 (1982).
9. N. D. Stow, E. C. McMonagle and A. J. Davison, Fragments from both termini of the herpes simplex virus type 1 genome contain signals required for the encapsidation of viral DNA, *Nucleic Acids Res.* **11**, 8205–8220 (1983).
10. F. L. Homa and J. C. Brown, Capsid assembly and DNA packaging in herpes simplex virus, *Rev. Med. Virol.* **7**, 107–122 (1997).
11. S. LaBoissiere and P. O'Hare, Analysis of HCF, the cellular cofactor of VP16, in herpes simplex virus-infected cells, *J. Virol.* **74**, 99–109 (2000).
12. R. D. Everett, ICP0, a regulator of herpes simplex virus during lytic and latent infection, *BioEssays* **22**, 761–770 (2000).
13. R. Hagglund, C. Van Sant, P. Lopez and B. Roizman, Herpes simplex virus 1-infected cell protein 0 contains two E3 ubiquitin ligase sites specific for different E2 ubiquitin-conjugating enzymes, *Proc. Natl. Acad. Sci. USA* **99**, 631–636 (2002).

14. C. Boutell, S. Sadis and R. D. Everett, Herpes simplex virus type 1 immediate-early protein ICP0 and its isolated RING finger domain act as ubiquitin E3 ligases in vitro. *J. Virol.* **76**, 841–850 (2002).
15. F. Momburg and H. Hengel, Corking the bottleneck: the transporter associated with antigen processing as a target for immune subversion by viruses, *Curr. Top. Microbiol. Immunol.* **269**, 57–74 (2002).
16. D. N. Everly, Jr., P. Feng, I. S. Mian and G. S. Read, mRNA degradation by the virion host shutoff (Vhs) protein of herpes simplex virus: genetic and biochemical evidence that Vhs is a nuclease, *J. Virol.* **76**, 8560–8571 (2002).
17. J. Burkham, D. M. Coen, C. B. Hwang and S. K. Weller, Interactions of herpes simplex virus type 1 with ND10 and recruitment of PML to replication compartments, *J. Virol.* **75**, 2353–2367 (2001).
18. G. Sourvinos and R. D. Everett, Visualization of parental HSV-1 genomes and replication compartments in association with ND10 in live infected cells, *EMBO J.* **21**, 4989–4997 (2002).
19. G. G. Maul, H. H. Guldner and J. G. Spivack, Modification of discrete nuclear domains induced by herpes simplex virus type 1 immediate early gene 1 product (ICP0), *J. Gen. Virol.* **74**, 2679–2690 (1993).
20. R. D. Everett, P. Freemont, H. Saitoh, M. Dasso, A. Orr, M. Kathoria and J. Parkinson, The disruption of ND10 during herpes simplex virus infection correlates with the Vmw110- and proteasome-dependent loss of several PML isoforms, *J. Virol.* **72**, 6581–6591 (1998).
21. E. K. Flemington, Herpesvirus lytic replication and the cell cycle: arresting new developments, *J. Virol.* **75**, 4475–4481 (2001).
22. G. L. Ehmann, T. I. McLean and S. L. Bachenheimer, Herpes simplex virus type 1 infection imposes a G(1)/S block in asynchronously growing cells and prevents G(1) entry in quiescent cells, *Virology* **267**, 335–349 (2000).
23. R. D. Everett, W. C. Earnshaw, J. Findlay and P. Lomonte, Specific destruction of kinetochore protein CENP-C and disruption of cell division by herpes simplex virus immediate-early protein Vmw110, *EMBO J.* **18**, 1526–1538 (1999).
24. C. A. Wu, N. J. Nelson, D. J. McGeoch and M. D. Challberg, Identification of herpes simplex virus type 1 genes required for origin-dependent DNA synthesis, *J. Virol.* **62**, 435–443 (1988).
25. K. M. Koslowski, P. R. Shaver, J. T. Casey, 2nd, T. Wilson, G. Yamanaka, A. K. Sheaffer, D. J. Tenney and N. E. Pederson, Physical and functional interactions between the herpes simplex virus UL15 and UL28 DNA cleavage and packaging proteins, *J. Virol.* **73**, 1704–1707 (1999).
26. P. M. Beard, N. S. Taus and J. D. Baines, DNA cleavage and packaging proteins encoded by genes U(L)28, U(L)15, and U(L)33 of herpes simplex virus type 1 form a complex in infected cells, *J. Virol.* **76**, 4785–4791 (2002).
27. N. D. Stow, Localization of an origin of DNA replication within the TRs/IRs repeated region of the herpes simplex virus type 1 genome, *EMBO J.* **1**, 863–867 (1982).
28. S. K. Weller, A. Spadaro, J. E. Schaffer, A. W. Murray, A. M. Maxam and P. A. Schaffer, Cloning, sequencing, and functional analysis of oriL, a herpes simplex virus type 1 origin of DNA synthesis, *Mol. Cell. Biol.* **5**, 930–942 (1985).
29. D. J. Hazuda, H. C. Perry, A. M. Naylor and W. L. McClements, Characterization of the herpes simplex virus origin binding protein interaction with OriS, *J. Biol. Chem.* **266**, 24621–24626 (1991).
30. M. Polvino-Bodnar, P. K. Orberg and P. A. Schaffer, Herpes simplex virus type 1 oriL is not required for virus replication or for the establishment and reactivation of latent infection in mice, *J. Virol.* **61**, 3528–3535 (1987).

31. K. Igarashi, R. Fawl, R. J. Roller and B. Roizman, Construction and properties of a recombinant herpes simplex virus 1 lacking both S-component origins of DNA synthesis, *J. Virol.* **67**, 2123–2132 (1993).
32. A. Aslani, B. Macao, S. Simonsson and P. Elias, Complementary intrastrand base pairing during initiation of Herpes simplex virus type 1 DNA replication, *Proc. Natl. Acad. Sci. USA* **98**, 7194–7199 (2001).
33. S. W. Wong and P. A. Schaffer, Elements in the transcriptional regulatory region flanking herpes simplex virus type 1 oriS stimulate origin function, *J. Virol.* **65**, 2601–2611 (1991).
34. M. A. Hardwicke and P. A. Schaffer, Cloning and characterization of herpes simplex virus type oriL comparison of replication and protein-DNA complex formation by oriL and oriS, *J. Virol.* **69**, 1377–1388 (1995).
35. A. T. Nguyen-Huynh and P. A. Schaffer, Cellular transcription factors enhance herpes simplex virus type 1 oriS-dependent DNA replication, *J. Virol.* **72**, 3635–3645 (1998).
36. M. A. Hardwicke and P. A. Schaffer, Differential effect of nerve growth factor and dexamethasone on herpes simplex virus type 1 oriL- and oriS-dependent DNA replication in PC12 cells, *J. Virol.* **71**, 3580–3587 (1997).
37. C. E. Dabrowski, P. J. Carmillo and P. A. Schaffer, Cellular protein interactions with herpes simplex virus type 1 oriS, *Mol. Cell. Biol.* **14**, 2545–2555 (1994).
38. P. Elias and I. R. Lehman, Interaction of origin binding protein with an origin of replication of herpes simplex virus 1, *Proc. Natl. Acad. Sci. USA* **85**, 2959–2963 (1988).
39. P. D. Olivo, N. J. Nelson and M. D. Challberg, Herpes simplex virus DNA replication: the UL9 gene encodes an origin-binding protein, *Proc. Natl. Acad. Sci. USA* **85**, 5414–5418 (1988).
40. D. S. Fierer and M. D. Challberg, Purification and characterization of UL9, the herpes simplex virus type 1 origin-binding protein, *J. Virol.* **66**, 3986–3995 (1992).
41. J. A. Isler and P. A. Schaffer, Phosphorylation of the herpes simplex virus type 1 origin binding protein, *J. Virol.* **75**, 628–637 (2001).
42. A. P. Abbotts and N. D. Stow, The origin-binding domain of the herpes simplex virus type 1 UL9 protein is not required for DNA helicase activity, *J. Gen. Virol.* **76**, 3125–3130 (1995).
43. S. Deb and S. P. Deb, A 269-amino-acid segment with a pseudo-leucine zipper and a helix-turn-helix motif codes for the sequence-specific DNA-binding domain of herpes simplex virus type 1 origin-binding protein, *J. Virol.* **65**, 2829–2838 (1991).
44. P. Elias, C. M. Gustafsson, O. Hammarsten and N. D. Stow, Structural elements required for the cooperative binding of the herpes simplex virus origin binding protein to oriS reside in the N-terminal part of the protein, *J. Biol. Chem.* **267**, 17424–17429 (1992).
45. H. M. Weir, J. M. Calder and N. D. Stow, Binding of the herpes simplex virus type 1 UL9 gene product to an origin of viral DNA replication, *Nucleic Acids Res.* **17**, 1409–1425 (1989).
46. S. S. Lee and I. R. Lehman, The interaction of herpes simplex type 1 virus origin-binding protein (UL9 protein) with Box I, the high affinity element of the viral origin of DNA replication, *J. Biol. Chem.* **274**, 18613–18617 (1999).
47. R. C. Bruckner, J. J. Crute, M. S. Dodson and I. R. Lehman, The herpes simplex virus 1 origin binding protein: a DNA helicase, *J. Biol. Chem.* **266**, 2669–2674 (1991).
48. M. S. Dodson and I. R. Lehman, The herpes simplex virus type 1 origin binding protein. DNA-dependent nucleoside triphosphatase activity, *J. Biol. Chem.* **268**, 1213–1219 (1993).
49. D. A. Sampson, M. E. Arana and P. E. Boehmer, Cysteine 111 affects coupling of single-strand dna-binding to ATP hydrolysis in the herpes simplex virus type-1 origin-binding protein, *J. Biol. Chem.* **275**, 2931–2937 (2000).
50. M. E. Arana, B. Haq, N. Tanguy Le Gac and P. E. Boehmer, Modulation of the herpes simplex virus type-1 UL9 DNA helicase by its cognate single-strand DNA-binding protein, ICP8, *J. Biol. Chem.* **276**, 6840–6845 (2001).

51. P. E. Bohemer, The herpes simplex virus type-1 single-strand DNA-binding protein, ICP8, increases the processivity of the UL9 protein DNA helicase, *J. Biol. Chem.* **273**, 2676–2683 (1998).
52. P. E. Bohemer, M. S. Dodson and I. R. Lehman, The Herpes Simplex Virus type-1 origin binding protein: DNA helicase activity, *J. Biol. Chem.* **268**, 1220–1225 (1993).
53. A. M. Makhov, P. E. Bohemer, I. R. Lehman and J. D. Griffith, Visualization of the unwinding of long DNA chains by the herpes simplex virus type 1 UL9 protein and ICP8, *J. Mol. Biol.* **258**, 789–799 (1996).
54. G. Villani, M. J. Pillaire and P. E. Bohemer, Effect of the major adduct of the antitumor drug *cis*-diamminedichloroplatinum II on the activity of a helicase essential for DNA replication, the herpes simplex virus type-1 origin-binding protein, *J. Biol. Chem.* **269**, 21676–21681 (1994).
55. K. Baradaran, C. E. Dabrowski and P. A. Schaffer, Transcriptional analysis of the region of the herpes simplex virus type 1 genome containing the ULS, UL9, and UL10 genes and identification of a novel delayed-early gene product, OBPC, *J. Virol.* **68**, 4251–4261 (1994).
56. P. E. Bohemer and I. R. Lehman, Physical interaction between the Herpes Simplex Virus 1 origin-binding protein and single-stranded DNA-binding protein ICP8, *Proc. Natl. Acad. Sci. USA* **90**, 8444–8448 (1993).
57. P. E. Bohemer, M. C. Craigie, N. D. Stow and I. R. Lehman, Association of origin-binding protein and single-strand DNA binding protein, ICP8, during Herpes Simplex Virus Type-1 DNA replication *in vivo*, *J. Biol. Chem.* **269**, 29329–29334 (1994).
58. N. Tanguy Le Gac and P. E. Bohemer, Activation of the herpes simplex virus type-1 origin binding protein (UL9) by heat shock proteins, *J. Biol. Chem.* **277**, 5660–5666 (2002).
59. S. S. Lee, Q. Dong, T. S. Wang and I. R. Lehman, Interaction of herpes simplex virus 1 origin-binding protein with DNA polymerase alpha, *Proc. Natl. Acad. Sci. USA* **92**, 7882–7886 (1995).
60. P. E. Bohemer and I. R. Lehman, Herpes simplex virus DNA replication, *Annu. Rev. Biochem.* **66**, 347–384 (1997).
61. C. Y. Eom and I. R. Lehman, The human DnaJ protein, hTid-1, enhances binding of a multimer of the herpes simplex virus type 1 UL9 protein to *oris*, an origin of viral DNA replication, *Proc. Natl. Acad. Sci. USA* **99**, 1894–1898 (2002).
62. I. Konieczny and M. Zylicz, Role of bacterial chaperones in DNA replication, *Genet. Eng.* **21**, 95–111 (1999).
63. J. S. Liu, S. R. Kuo, A. M. Makhov, D. M. Cyr, J. D. Griffith, T. R. Broker and L. T. Chow, Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication, *J. Biol. Chem.* **273**, 30704–30712 (1998).
64. M. Stros, DNA bending by the chromosomal protein HMG1 and its high mobility group box domains. Effect of flanking sequences, *J. Biol. Chem.* **273**, 10355–10361 (1998).
65. R. O. Baker, L. B. Murata, M. S. Dodson and J. D. Hall, Purification and characterization of OF-1, a host factor implicated in herpes simplex replication, *J. Biol. Chem.* **275**, 30050–30057 (2000).
66. A. Koff, J. F. Schwedes and P. Tegtmeyer, Herpes simplex virus origin-binding protein (UL9) loops and distorts the viral replication origin, *J. Virol.* **65**, 3284–3292 (1991).
67. A. M. Makhov, P. E. Bohemer, I. R. Lehman and J. D. Griffith, The herpes simplex virus type 1 origin-binding protein carries out origin specific DNA unwinding and forms unwound stem-loop structures, *EMBO J.* **15**, 1742–1750 (1996).
68. L. B. Murata and M. S. Dodson, The herpes simplex virus type 1 origin-binding protein sequence-specific activation of adenosine triphosphatase activity by a double-stranded DNA containing box I, *J. Biol. Chem.* **274**, 37079–37086 (1999).

69. X. He, An initial ATP-independent step in the unwinding of a herpes simplex virus type I origin of replication by a complex of the viral origin-binding protein and single-strand DNA-binding protein, *Proc. Natl. Acad. Sci. USA* **98**, 3024–3028 (2001).
70. S. S. Lee and I. R. Lehman, Unwinding of the box I element of a herpes simplex virus type 1 origin by a complex of the viral origin binding protein, single-strand DNA binding protein, and single-stranded DNA, *Proc. Natl. Acad. Sci. USA* **94**, 2838–2842 (1997).
71. X. He and I. R. Lehman, Unwinding of a herpes simplex virus type 1 origin of replication (Ori(S)) by a complex of the viral origin binding protein and the single-stranded DNA binding protein, *J. Virol.* **74**, 5726–5728 (2000).
72. A. Aslani, M. Olsson and P. Elias, ATP-dependent unwinding of a minimal origin of DNA replication by the origin binding protein and the single-strand DNA binding protein ICP8 from herpes simplex virus type 1, *J. Biol. Chem.* **275**, 41204–41212 (2002).
73. A. M. Makhov, S. S. Lee, I. R. Lehman and J. D. Griffith, Origin specific unwinding of herpes simplex virus type 1 DNA by the viral UL9 and ICP8 proteins: visualization of a specific pre-unwinding complex. *Proc. Natl. Acad. Sci. USA* **100**, 898–903 (2003).
74. G. W. McLean, A. P. Abbotts, M. E. Parry, H. S. Marsden and N. D. Stow, The herpes simplex virus type 1 origin-binding protein interacts specifically with the viral UL8 protein, *J. Gen. Virol.* **75**, 2699–2706 (1994).
75. H. S. Marsden, G. W. McLean, E. C. Barnard, G. J. Francis, K. MacEachran, M. Murphy, G. McVey, A. Cross, A. P. Abbotts and N. D. Stow, The catalytic subunit of the DNA polymerase of herpes simplex virus type 1 interacts specifically with the C terminus of the UL8 component of the viral helicase-primase complex, *J. Virol.* **71**, 6390–6397 (1997).
76. S. J. Monahan, L. A. Grinstead, W. Olivieri and D. S. Parris, Interaction between the herpes simplex virus type 1 origin-binding and DNA polymerase accessory proteins, *Virology* **241**, 122–130 (1998).
77. O. Hammarsten, X. Yao and P. Elias, Inhibition of topoisomerase II by ICRF-193 prevents efficient replication of herpes simplex virus type 1, *J. Virol.* **70**, 4523–4529 (1996).
78. P. Digard, W. R. Bebrin, K. Weissbart and D. M. Coen, The extreme C terminus of herpes simplex virus DNA polymerase is crucial for functional interaction with processivity factor UL42 and for viral replication, *J. Virol.* **67**, 398–406 (1993).
79. K. Weissbart, C. S. Chow and D. M. Coen, Herpes simplex virus processivity factor UL42 imparts increased DNA-binding specificity to the viral DNA polymerase and decreased dissociation from primer-template without reducing the elongation rate, *J. Virol.* **73**, 55–66 (1999).
80. M. Chaudhuri, L. Song and D. S. Parris, The herpes simplex virus type 1 DNA polymerase processivity factor increases fidelity without altering pre-steady-state rate constants for polymerization or excision. *J. Biol. Chem.* **278**, 8996–9004 (2003).
81. J. C. Randell and D. M. Coen, Linear diffusion on DNA despite high-affinity binding by a DNA polymerase processivity factor, *Mol. Cell* **8**, 911–920 (2001).
82. D. Jeruzalmi, M. O'Donnell and J. Kuriyan, Clamp loaders and sliding clamps, *Curr. Opin. Struct. Biol.* **12**, 217–224 (2002).
83. H. J. Zuccola, D. J. Filman, D. Coen and J. M. Hogle, The crystal structure of an unusual processivity factor, herpes simplex virus UL 42, bound to the C terminus of its cognate polymerase, *Mol. Cell* **75**, 267–278 (2000).
84. M. Chaudhuri and D. S. Parris, Evidence against a simple tethering model for enhancement of herpes simplex virus DNA polymerase processivity by accessory protein UL42, *J. Virol.* **76**, 10270–10281 (2002).
85. A.-S. Gourves, N. Tanguy Le Gac, G. Villani, P. E. Boehmer and N. P. Johnson, Equilibrium binding of single stranded DNA with HSV-1 coded single stranded DNA binding protein. ICP8, *J. Biol. Chem.* **275**, 10864–10869 (2000).

86. J. M. Calder and N. D. Stow, Herpes simplex virus helicase-primase: the UL8 protein is not required for DNA-dependent ATPase and DNA helicase activities, *Nucleic Acids Res.* **18**, 3573–3578 (1990).
87. M. Falkenberg, P. Elias and I. R. Lehman, The herpes simplex virus type 1 helicase-primase. Analysis of helicase activity, *J. Biol. Chem.* **273**, 32154–32157 (1998).
88. N. Tanguy Le Gac, G. Villani, J.-S. Hoffmann and P. E. Boehmer, The UL8 subunit of the herpes simplex virus type-1 DNA helicase-primase optimizes utilization of DNA templates covered by the homologous single strand DNA-binding protein, ICP8, *J. Biol. Chem.* **271**, 21645–21651 (1996).
89. M. Falkenberg, D. A. Bushnell, P. Elias and I. R. Lehman, The UL8 subunit of the heterotrimeric herpes simplex virus type 1 helicase-primase is required for the unwinding of single strand DNA-binding protein (ICP8)-coated DNA substrates, *J. Biol. Chem.* **272**, 22766–22770 (1997).
90. N. Tanguy Le Gac, G. Villani and P. E. Boehmer, Herpes simplex virus type-1 single-strand DNA-binding protein (ICP8) enhances the ability of the viral DNA helicase-primase to unwind cisplatin-modified DNA, *J. Biol. Chem.* **273**, 13801–13807 (1998).
91. K. A. Ramirez-Aguilar, N. A. Low-Nam and R. D. Kuchta, Key role of template sequence for primer synthesis by the herpes simplex virus 1 helicase-primase, *Biochemistry* **41**, 14569–14579 (2002).
92. G. Sherman, J. Gottlieb and M. D. Challberg, The UL8 subunit of the herpes simplex virus helicase-primase complex is required for efficient primer utilization, *J. Virol.* **66**, 4884–4892 (1992).
93. S. D. Rabkin and B. Hanlon, Herpes simplex virus DNA synthesis at a preformed replication fork in vitro, *J. Virol.* **64**, 4957–4967 (1990).
94. R. Skaliter, A. M. Makhov, J. D. Griffith and I. R. Lehman, Rolling circle DNA replication by extracts of herpes simplex virus type 1-infected human cells, *J. Virol.* **70**, 1132–1136 (1996).
95. R. Skaliter and I. R. Lehman, Rolling circle DNA replication in vitro by a complex of herpes simplex virus type 1-encoded enzymes, *Proc. Natl. Acad. Sci. USA* **91**, 10665–10669 (1994).
96. M. Falkenberg, I. R. Lehman and P. Elias, Leading and lagging strand DNA synthesis in vitro by a reconstituted herpes simplex virus type 1 replisome, *Proc. Natl. Acad. Sci. USA* **97**, 3896–3900 (2000).
97. F. A. Kadyrov and J. W. Drake, Characterization of DNA synthesis catalyzed by bacteriophage T4 replication complexes reconstituted on synthetic circular substrates, *Nucleic Acids Res.* **30**, 4387–4397 (2002).
98. J. Shlomai, A. Friedmann and Y. Becker, Replication intermediates of herpes simplex virus DNA, *Virology* **69**, 647–659 (1976).
99. I. Hirsch, G. Cabral, M. Patterson and N. Biswal, Studies on the intracellular replicating DNA of herpes simplex virus type 1, *Virology* **81**, 48–61 (1977).
100. P. M. Takahara, A. C. Rosenzweig, C. A. Frederick and S. J. Lippard, Crystal structure of double stranded DNA containing the major adduct of the anticancer drug cisplatin, *Nature* **377**, 649–652 (1995).
101. G. Villani, N. Tanguy le Gac and Jean-Sébastien Hoffmann, Replication of platinated DNA and its mutagenic consequences, in: “Chemistry and Biochemistry of Leading Anticancer Drug,” pp. 135–157. Wiley-VCH ed, 1999.
102. P. M. Burgers, E. V. Koonin, E. Bruford, L. Blanco, K. C. Burtis, M. F. Christman, W. C. Copeland, E. C. Friedberg, F. Hanaoka, D. C. Hinkle, C. W. Lawrence, M. Nakanishi, H. Ohmori, L. Prakash, S. Prakash, A. Reynaud, A. Sugino, T. Todo, Z. Wang, J. C. Weill and R. Woodgate, Eukaryotic DNA polymerases: proposal for a revised nomenclature, *J. Biol. Chem.* **276**, 43478–43490 (2001).

103. G. Villani, N. Tanguy le Gac, L. Wasungu, D. Burnouf, R. P. Fuchs and P. E. Boehmer, Effect of manganese on *in vitro* replication of damaged DNA catalyzed by the herpes simplex virus type-1 DNA polymerase, *Nucleic Acids Res.* **30**, 3323–3332 (2002).
104. E. Friedberg, R. Wagner and M. Radman, Specialized DNA polymerases, cellular survival and the genesis of mutations, *Science* **296**, 1627–1630 (2002).
105. C. L. O'Day, P. M. Burgers and J. S. Taylor, PCNA-induced DNA synthesis pas cis-syn and trans-syn thymine dimers by calf thymus DNA polymerase δ *in vitro*, *Nucleic Acids Res.* **20**, 5403–5406 (1992).
106. L. Bastide, P. E. Boehmer, G. Villani and B. Lebleu, Inhibition of a DNA-helicase by peptide nucleic acids, *Nucleic Acids Res.* **27**, 551–554 (1999).
107. R. E. Dutch, R. C. Bruckner, E. S. Mocarski and I. R. Lehman, Herpes simplex virus type 1 recombination: role of DNA replication and viral a sequences, *J. Virol.* **66**, 277–285 (1992).
108. A. Severini, D. G. Scraba and D. L. Tyrrell, Branched structures in the intracellular DNA of herpes simplex virus type 1, *J. Virol.* **70**, 3169–3175 (1996).
109. B. Slobedman, X. Zhang and A. Simmons, Herpes simplex virus genome isomerization: origins of adjacent long segments in concatemeric viral DNA, *J. Virol.* **73**, 810–813 (1999).
110. K. J. Huang, B. V. Zemelman and I. R. Lehman, Endonuclease G, a candidate human enzyme for the initiation of genomic inversion in herpes simplex type 1 virus, *J. Biol. Chem.* **277**, 21071–21079 (2002).
111. L. S. Symington, Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair, *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).
112. S. L. Lusetti and M. M. Cox, The bacterial RecA protein and the recombinational DNA repair of stalled replication forks, *Annu. Rev. Biochem.* **71**, 71–100 (2002).
113. A. V. Nimonkar and P. E. Boehmer, The herpes simplex virus type-1 single-strand DNA-binding protein (ICP8) promotes strand invasion, *J. Biol. Chem.* **278**, 9678–9682 (2003).
114. K. J. Mariani, PriA-directed replication fork restart in *Escherichia coli*, *Trends Biochem. Sci.* **25**, 185–189 (2000).
115. S. C. West, Processing of recombination intermediates by the RuvABC proteins, *Annu. Rev. Genet.* **31**, 213–244 (1997).
116. D. Kong and C. C. Richardson, Single-stranded DNA binding protein and DNA helicase of bacteriophage T7 mediate homologous DNA strand exchange, *EMBO J.* **15**, 2010–2019 (1996).
117. A. V. Nimonkar and P. E. Boehmer, *In vitro* strand-exchange promoted by the Herpes Simplex Virus Type-1 single-strand DNA-binding protein (ICP8) and DNA helicase-primase, *J. Biol. Chem.* **277**, 15182–15189 (2002).
118. A. Kuzminov, DNA replication meets genetic exchange: chromosomal damage and its repair by homologous recombination, *Proc. Natl. Acad. Sci. USA* **98**, 8461–8468 (2001).

This Page Intentionally Left Blank

Enzymology and Molecular Biology of Glucocorticoid Metabolism in Humans

ANDREAS BLUM AND
EDMUND MASER

*Institute of Experimental Toxicology,
Universitäts Klinikum Schleswig –
Holstein, Campus Kiel, Brunswiker
Strasse 10, D-24105 Kiel, Germany*

I. Introduction	174
II. Glucocorticoid Action	175
III. The Free Hormone Hypothesis	177
IV. Glucocorticoid Receptors	178
V. Nongenomic Action of Glucocorticoids	179
VI. Glucocorticoid Resistance	180
VII. 11 β -Hydroxysteroid Dehydrogenase (11 β -HSD) as Preceptor Control	181
VIII. Hydroxysteroid Dehydrogenases: Physiological Role	182
IX. The History of Two Isoforms: 11 β -HSD Type 1 and 11 β -HSD Type 2	182
X. A (Patho)physiological Role for 11 β -HSD 1 and 11 β -HSD 2	183
XI. 11 β -HSD 2 and the Syndrome of "Apparent Mineralocorticoid Excess" (AME)	184
XII. 11 β -Hydroxysteroid Dehydrogenase Type 1 (11 β -HSD 1)	187
XIII. Conclusions	203
References	203

Glucocorticoids (GCs) are a vital class of steroid hormones that are secreted by the adrenal cortex and that are regulated by ACTH largely under the control of the hypothalamic-pituitary-adrenal axis. GCs mediate profound and diverse physiological effects in vertebrates, ranging from development, metabolism, neurobiology, anti-inflammation and programmed cell death to many other functions. Multiple factors "downstream" of GC secretion, such as glucocorticoid receptor (GR) number and the abundance of plasma binding proteins have originally been considered as modulators of GC action. However, in the last decade the role of tissue-specific GC activating and inactivating enzymes have been identified as additional determinants in GC signalling pathways. On the cellular level, they function as important pre-receptor regulators by acting as "molecular switches" for receptor-active and receptor-inactive GC hormones. According to their biologic activity to catalyze the interconversion of C₁₁-hydroxyl and C₁₁-oxo GCs these enzymes have been named 11 β -hydroxysteroid dehydrogenase (11 β -HSD; EC 1.1.1.146). Two isoforms of 11 β -HSD have been cloned and characterized so far. 11 β -HSD type 1 is found in a wide range of

tissues, acts predominantly as a reductase in intact cells and tissues by regenerating active cortisol from cortisone, and has been described to regulate GC access to the GR. 11β -HSD type 2 is found mainly in mineralocorticoid target tissues such as kidney and colon, acts only as a dehydrogenase by producing inactive cortisone, and has been found to protect the mineralocorticoid receptor from high levels of receptor-active cortisol. Recently, 11β -HSD 1 has become highly topical due to the finding that 11β -HSD 1 plays a pivotal role in the pathogenesis of central obesity and the appearance of the metabolic syndrome. This review provides an overview on the components involved in GC signalling of 11β -HSD type 1 as an important pre-receptor control enzyme that modulates activation of the GR.

Abbreviations

11β -HSD	11β -hydroxysteroid dehydrogenase
ACTH	adenocorticotrophic hormone
AME	apparent mineralocorticoid excess
CBP	cortisol-binding protein
ER	endoplasmic reticulum
GC	glucocorticoid
GH	growth hormone
GR	glucocorticoid receptor
GRE	glucocorticoid response element
HPA	hypothalamic–pituitary–adrenal axis
HSD	hydroxysteroid dehydrogenase
HSP	heat-shock protein
MR	mineralocorticoid receptor
SDR	short-chain dehydrogenase/reductase

I. Introduction

The capacity of individual tissues within the intact organism to participate in integrated and complex metabolic pathways depends on the activity of the endocrine system. The actions of hormones secreted by endocrine organs are diverse and important, influencing cell differentiation, development and growth, as well as metabolism in most tissues of higher animals. Despite the diversity of chemical structure, all hormones share important characteristics in terms of their mode of action. Most are present in the circulation in low concentration, are delivered to target tissues, and interact with specific high-affinity receptors to exert their actions. Peptide hormones bind to and activate cell-surface receptors, resulting in transduction of a signal to the interior of the cell by a complex system of second messengers. In contrast, thyroid, steroid, and related hormones exert their

actions directly on target gene expression by binding to specific, structurally similar, nuclear receptors which function as hormone-inducible transcription factors.

Glucocorticoids (GCs) are a vital class of steroid hormones that mediate profound and diverse physiological effects in vertebrates from fish to man. Produced and secreted by the adrenal cortex, levels of circulating GCs are regulated by adenocorticotrophic hormone (ACTH) largely under the control of the hypothalamic–pituitary–adrenal axis (HPA). Although named for their role in glucose haemostasis, GCs are also eminently important throughout physiology, with regulatory roles in development, metabolism, neurobiology, programmed cell death, and many other functions. It is also common to regard GC release as the entire organismal response to stress.

In addition to these far-reaching physiological roles, corticosteroids are among the most widely prescribed class of drugs in the world. The pharmacological benefits of GCs are primarily anti-inflammatory and immunosuppressive. They also find widespread use in chemotherapeutic regimes in patients with leukemias, lymphomas, and other cancers due to their critical role in the induction of apoptosis.

However, multiple factors “downstream” of GC secretion, such as glucocorticoid receptor (GR) number, the number of the abundance of plasma-binding proteins, or GC-activating and -inactivating enzymes that act as prereceptor regulative on the cellular level may be regulated to alter the organismal response to GCs. This review provides an overview on the components involved in GC signaling and, in particular, emphasizes the enzymology, physiology, and significance of 11 β -hydroxysteroid dehydrogenase type 1 (11 β -HSD 1) (EC 1.1.1.146) as an important prereceptor control enzyme that modulates activation of the GC receptor.

II. Glucocorticoid Action

A. Molecular Mechanism of Glucocorticoid Action

Glucocorticoids enter the cell from circulating fluids and exert their effects following binding with their inactive intracellular GR, a hormone activated, dual zinc finger transcription factor. The inactivated ligand-free GR is found in the cytoplasm as a hetero-oligomer complex with other proteins, such as e.g., heat-shock protein (HSP) 90 (1). Binding of the GC to its receptor results in conformational changes in the ligand/receptor complex which allows the release of HSP-90 and unmasking

of the domains responsible for dimerization, nuclear localization, DNA binding, and transactivation (2). The hormone-bound GR then translocates to the nucleus, where it transiently associates with another HSP, HSP-56, and later dissociates from HSP-56 and binds as a dimer to conserved palindromic DNA sequences, named the GC response elements (GREs).

The GRE was the first hormone-responsive element described and was found in the long-terminal repeat of the mouse mammary tumor virus (3). Similar GREs were identified in the promoters of the human metallothionein IIA, human growth hormone, and other GC-regulated target genes. Comparison of the nucleotide sequences of each of these elements led to the proposal of a consensus GRE motif considered to be optimal in sequence and configuration for GR binding to DNA (reviewed in Ref. (4)).

The consensus GRE comprises two boxes spaced by three variable nucleotides (GGTACAnnnTGTTCT). The hexamers exhibit a dyad symmetry in a palindromic arrangement (4), and each box interacts with one of the GR zinc fingers (5). GREs are located in variable copy numbers and are found at variable distances from the TATA box in the promoter region of GCs-responsive genes, including cytokine genes (6).

The ligand-activated GR interacts with a multitude of transcription factors such as c-jun, nuclear factor- κ B (NF- κ B), the TFIID complex, STAT5, as well as a host of coactivators where they are known to act on the function of these signaling molecules. In addition, the GR interacts with numerous cytosolic proteins including chaperones, kinases, phosphatases, nuclear-shuttling proteins, and the proteasome (reviewed in Ref. (7)). Depending on the target gene, ligand-activated GRs may stimulate (transactivation) or alternatively inhibit (transrepression) gene transcription. The former is exemplified by the capacity of GCs to upregulate the expression of the specific NF- κ B inhibitor, I κ B (8). The latter is evidenced by the well-documented capability of GCs in inhibiting the expression of IL-2 and other cytokine genes (9).

B. Glucocorticoid Negative Action

A major molecular pathway, via which GCs exert part of their antigrowth and anti-inflammatory effects, appears to be the formation of intranuclear complexes of GC-bound receptors with apolipoprotein (AP) 1 transcription factor (c-Fos–c-Jun heterodimer), which prevents this transcription factor from exerting its ubiquitous-activating and growth-promoting effects on its DNA-responsive element, the AP-1 site (10). Reduction in transcription factor availability and/or function, in turn, results in downstream reduction in, and arrest of, transcriptional activity in target genes. In antagonizing transcription

factors, GR does not modulate the correct assembly of the preinitiation complex (TFII–DAB complex), hence localizing its effect upstream of the TATA box (11). Several mechanisms were postulated for GCs antagonism of NF- κ B. These included induction of the synthesis of the NF- κ B inhibitor, I κ B (8), a protein–protein interaction model, which proposes that GR-repressed NF- κ B activation and/or function either by blocking its access to its DNA (κ B) site (12), or by forming a complex with NF- κ B which loses DNA capacity (11) and/or by competition with NF- κ B for nuclear coactivators (13).

C. Glucocorticoids and the Immune Response

Glucocorticoids are clinically used in treating disorders of heightened immunity, including transplantation rejection and autoimmunity (14). They exert their anti-inflammatory and antiproliferative effects principally by inhibiting the expression of cytokines and adhesion molecules at the transcriptional and posttranscriptional levels (15). Glucocorticoids inhibit the expression of proinflammatory and immunoregulatory cytokines, including IL-1 to IL-6, IL-8, IL-11, IL-12, IL-15, IL-16, interferon- γ , tumor necrosis factor- α (TNF- α), and the colony-stimulating factors (CSFs) macrophage (M)-CSF, granulocyte (G)-CSF, CSF-1, and granulocyte-macrophage (GM)-CSF (reviewed in Ref. (16)).

Originally, it was postulated that this inhibition involved binding of the hormone-activated GC–GR complex with the GREs in the promoter of GC-regulated cytokine genes (16). Recent evidence suggested an additional mechanism, namely antagonism of nuclear transcription factor- κ B (NF- κ B), which resulted in attenuation or arrest of transcriptional activities (17). However, it is very likely that GCs may utilize more than one mechanism in exerting their antiproliferative effect (14).

III. The Free Hormone Hypothesis

The principal binding protein for GCs in the circulation is the cortisol-binding protein (CBP). One molecule of CBP binds one molecule of cortisol with high affinity. Also albumin binds cortisol but with low affinity. The third major protein transporting steroid is sex hormone-binding globulin (SHBG) which, however, only binds cortisol with low affinity and is without physiological significance with regard to GCs (18).

According to the free hormone hypothesis, steroid bound to plasma-binding proteins is unavailable to tissues, whereas the “free” (unbound) hormone is the biologically active fraction, able to enter cells, activate intracellular or membrane receptors, and also be available for metabolism in

the liver (19). The free hormone hypothesis predicts that the biological activity of cortisol is proportional to the concentration of free hormone and not to the total concentration including the protein-bound fraction. Although this hypothesis may not be valid for all steroid hormones in every organ, there seems to be good evidence to suggest that it accounts for cortisol action in most situations. The function of steroid-binding proteins therefore is to act as a buffer reservoir of steroids present throughout the body that can readily be made available to the pool of free hormone by simple dissociation (20). Therefore, the localization of CBP to tissues may serve to locally enhance or inhibit GC action in that tissue.

IV. Glucocorticoid Receptors

The concept that the actions of steroid hormones occur within the nucleus by hormone activation of specific intracellular receptors was initially proposed almost 30 years ago. Early ligand-binding studies demonstrated the presence of high-affinity binding sites for steroid hormones, predominantly in the cytoplasm. The cloning of GR and the estrogen receptor cDNAs led to the recognition of sequence homologies between these receptors at both nucleotide and amino acid levels (21). Later recognition of similarity between the GR, estrogen receptor, and receptors for the structurally related androgen, progesterone, and mineralocorticoid hormones (22) led to the concept of a superfamily of steroid hormone receptors (23). Natural GCs and their synthetic derivatives work through the GR. The GR is a member of the nuclear hormone receptor superfamily of ligand-activated transcription factors. Binding of hormone-activated GR to GRE elements results in the transactivation of downstream target genes.

Comparison of predicted amino acid structures of receptor proteins indicated the presence of distinct regions that share close structural homology, and of other regions in which specific differences between receptors occur but which are conserved between isoform variants of the same receptor. These domains correspond to the variable sized, nonhomologous N-terminal region, the zinc finger-containing DNA-binding domain (60–70 amino acids), and the C-terminal ligand-binding region (ca. 250 amino acids). The organization of each domain is conserved in all superfamily members. The N-terminus is involved in transactivation; DNA-binding domains recognize the sequence and anatomy of DNA-binding sites and may participate in receptor dimerization. The C-terminus mediates several receptor activities, including ligand binding, nuclear localization and HSP binding, receptor homo- and heterodimerization, and interaction with other nuclear proteins to participate in transactivation.

The precise molecular mechanism of hormone-dependent target gene activation or repression by steroid hormone receptors still requires elucidation. Nevertheless, biophysical and structural studies over the last half decade, most notably with the estrogen receptor, have led to an understanding of steroid hormone action and antagonism at the atomic scale (24,25). Structure models predict both subtle and drastic changes in ligand-induced receptor conformations that can direct or impede specific protein interactions responsible for downstream biological effects.

To date there is only one known GR gene which is located on chromosome 5q11–q13 (26). By having more than one promoter sequence, it is thought that GR protein expression can be differentially regulated under diverse signaling environments. Alternative splicing also gives rise to numerous GR isoforms. When cloned in 1985, two human GR gene-splicing products were identified and termed hGR α and hGR β . hGR α was recognized as the classical GR which is the primary mediator of GC action (26). The hGR β isoform is transcriptionally inactive and is unable to bind agonists or antagonists in all systems tested to date. However, it is the dominant negative effect of hGR β on hGR α -mediated transactivation that is most intriguing, and potentially physiologically important. Recent studies investigating the mechanism of hGR β -dominant negative activity suggest that it is the formation of hGR α / β heterodimers, incapable of binding coactivators as a ligand-activated hGR α homodimer would, that causes attenuated hGR α response (27). Recently, proinflammatory cytokines such as IL-8, TNF α , and IL-1 were shown to increase hGR β expression (28). Changing the ratio of hGR β to hGR α , where hGR β predominates, may promote GC resistance (28).

V. Nongenomic Action of Glucocorticoids

A notable paradox of GC biology is the rapid release of the hormone during a stress response (seconds) and the perceived delay (hours) in eliciting a genomic response consistent with the classical action of steroid and other nuclear hormone receptors. In the past several years, there has been an increased interest in nongenomic and/or membrane-associated actions of steroids and their receptors (29). Distinct GR forms could presumably mediate the rapid actions of GCs. One possibility is the presence of a membrane-bound or associated form of the GR (30). This receptor may either be a unique gene product, as proposed for the progesterone receptor (31), or a modified version of the classical GR capable of integrating into the plasma membrane. Another possibility is that a cytosolic subset of GR mediates the rapid actions of GCs by participating in signal-transduction pathways usually associated with membrane receptor-signaling events (32). Whether

this receptor turns out to be a unique gene product, specific isoform, or merely a subset of the classical receptor that alters cellular phenotype through protein-protein interactions with factors such as NF- κ B etc. remains to be determined.

VI. Glucocorticoid Resistance

The physiological response and sensitivity to GCs varies among species, individuals, tissues, cell types, and even during the cell cycle. For example, the guinea pig is a well-known example of a cortico-resistant species (33) as are several New World primates, such as the squirrel monkey and the marmoset (34). Given the complex array of life-sustaining functions, it is of no surprise that complete inability of GCs to exert their effects on their target tissues would be incompatible with life and that only syndromes of partial or incomplete GC resistance would exist.

The molecular basis of GC resistance varies widely and is not completely understood. Generalized inherited GC resistance is classified by a failure of synthetic GCs to elicit a normal physiological response. Although abnormalities in GC responsiveness can be attributed in many cases to specific inherited mutations, epigenic factors are also involved, because in many diseases long-term GC treatment can result in GC resistance (35). A number of kindred and individual patients suffering from partial forms of GC resistance have been described to date. Recent advances in our understanding of the molecular mechanism of action of GCs have allowed a first glimpse into the molecular pathophysiology of GC resistance (23).

A. Pathophysiology

A complex negative-feedback system exists in the human central nervous system (CNS) that regulates GC homeostasis (36). Glucocorticoids exert their negative-feedback effects on the secretion of both hypothalamic corticotrophic-releasing hormone (CRH) and arginine vasopressin (AVP) and, in addition, inhibit the secretion of pituitary ACTH itself. Furthermore, GCs influence the activity of the suprahypothalamic centers that control the activity of CRH and AVP neurones, such as the hippocampus and amygdala. This complex autoregulatory system is activated in states of generalized GC resistance, resulting in compensatory increases in ACTH and cortisol secretion. Although adequate compensation is achieved by elevated cortisol concentrations in the great majority of the patients described, excess ACTH secretion results also in increased production of adrenal steroid with salt-retaining (mineralocorticoid) activity, such as deoxycorticosterone and corticosterone (37), and enhanced secretion of adrenal androgens, such as

Δ^4 -androstendione, dehydroepiandrosterone (DHEA), and DHEA sulfate (38). Generally, the pathophysiological manifestations of GC resistance appear to arise from the compensatory elevations of steroids with mineralocorticoid and androgen activity.

B. Familial Glucocorticoid Resistance

Familial GC resistance, for example, results from partial inability of GCs to exert their effects on their target tissues throughout the organism. The manifestations of GC resistance vary from asymptomatic to chronic fatigue, to varying degrees of hypertension and/or hypokalemic alkalosis and hyperandrogenism. The latter can be manifested in woman as acne, hirsutism, menstrual irregularity, oligoanovulation, and infertility, in men as infertility, and in children as precocious puberty. Different molecular defects of the highly conserved GR gene, altering its concentration and functional characteristics, appear to cause the syndrome of familial GC resistance. Depending on the molecular defect, this syndrome is transmitted by an autosomal dominant or recessive trait. Several reviews on GC resistance discuss the complexity of the syndrome (39,40).

VII. 11 β -Hydroxysteroid Dehydrogenase (11 β -HSD) as Prereceptor Control

Sometimes it appears too simple to measure levels of a hormone in a disease process and to conclude on the basis of “normal” circulating concentrations that such a hormone cannot be involved in the pathogenesis of that condition. It will become clear that many diverse clinical conditions can arise from abnormalities of a tissue-specific receptor, an enzyme defect, or enzyme hyperactivity, and that this commonly occurs independent of circulating hormone levels. In case of steroid hormones, it has long been assumed that the intracellular concentration of the corresponding receptor represents the primary mechanism for regulating the activity of a particular steroid hormone. However, in the last decade the role of tissue-specific hydroxysteroid dehydrogenases (HSDs) that control the amount of active steroid in target cells has been recognized as another important regulatory mechanism (41–46). For GCs, a major challenge is now to consider how endocrinopathies such as hirsutism, obesity, infertility, and osteoporosis could relate to defects in tissue-specific cortisol metabolism.

VIII. Hydroxysteroid Dehydrogenases: Physiological Role

Hydroxysteroid dehydrogenases are commonly thought to play pivotal roles in the biosynthesis and inactivation of steroid hormones. Steroid hormones act by binding to receptor proteins in target cells, which lead to transcriptional regulation of different gene products and the desired physiological response. In target tissues, however, HSDs interconvert potent steroid hormones to their cognate inactive metabolites (and vice versa) and thus regulate the occupancy of steroid hormone receptors. The reactions they catalyze are positional and stereospecific, and usually involve the interconversion of a carbonyl with a hydroxyl group on either the steroid nucleus or side chain (47). In normal cell physiology, HSDs therefore function as important prereceptor regulators of signaling pathways by acting as “molecular switches” for receptor-active and receptor-inactive hormone (48–50). Because the enzymes have much greater tissue specificity than the receptors, they have emerged as attractive targets for the design of potent and selective drugs that combat steroid-related disorders (cf. Section XII.J of this review).

IX. The History of Two Isoforms: 11 β -HSD Type 1 and 11 β -HSD Type 2

The fact that the biological activity of GCs resides with the C₁₁-hydroxyl group, and that the conversion of this group to a C₁₁-oxo group inactivates the steroid, has been recognized since the 1950s. 11 β -HSD activity was first described by Amelung *et al.* (51) 50 years ago. Although subsequently described in a broad range of tissues, e.g., kidney (52), placenta (53), lung, and testis (54), it needed more than 30 years since an 11 β -HSD enzyme was purified from rat liver (55). The purified enzyme was NADPH dependent and exhibited only dehydrogenase activity, converting the active GC corticosterone into the corresponding inactive form dehydrocorticosterone. So many questions remained open, due to the fact that in some tissues, e.g., kidney and placenta, the dehydrogenase activity predominated, whereas in liver and lung the oxoreductase activity was found.

Evidence for a second 11 β -HSD isoform came shortly thereafter. Immunohistochemical studies of 11 β -HSD in the rat kidney, by using antibodies specific for the liver form, showed that the enzyme did not colocalize with the mineralocorticoid receptor (MR), but instead was found in the proximal tubule, whereas aldosterone-binding sites are located in the distal

nephron (56–58). Although immunoreactivity was apparently absent in the distal nephron, elegant microdissection studies on the rabbit nephron indicated enzyme activity within the distal nephron and collecting ducts (59). And indeed, an NAD⁺-dependent 11 β -HSD was soon identified in isolated rabbit cortical collecting duct cells (60), and later confirmed by enzyme-activity studies (61) and cytochemical *in situ* techniques in the rat kidney (62). Subsequent cloning and expression in mammalian cells and *Xenopus* oocytes of human (63) and sheep (64) 11 β -HSD kidney isoforms confirmed the hypothesis that a different isoform distributes in the kidney which differs in Km-value and cofactor dependency (NAD⁺ instead of NADPH) from that enzyme found in liver (Fig. 1). Therefore, the kidney enzyme was named 11 β -hydroxysteroid dehydrogenase type 2 (11 β -HSD 2) in contrast to the liver form, which was named 11 β -hydroxysteroid dehydrogenase type 1 (11 β -HSD 1) (Table I).

X. A (Patho)physiological Role for 11 β -HSD 1 and 11 β -HSD 2

Owing to the fact that congenital defects in 11 β -HSD 2 expression and activity have been observed very early and shown to account for the syndrome of apparent mineralocorticoid excess (AME) (56,65) (cf. Section XI of this review), the major focus of research in the last decade was on 11 β -HSD 2. However, the reverse defect has also been described in a woman who was unable to convert cortisone to cortisol (66). Since cortisol levels were low, cortisol secretion increased accordingly to maintain normal circulating concentrations, but at the expense of ACTH-mediated androgen excess. This woman presented, therefore, with the clinical features of hyperandrogenemia, i.e., hirsutism and oligomenorrhoea. A syndrome associated with 11 β -HSD 1 deficiency has only recently been reported to result in an increased ACTH-driven androgen production (67).

Moreover, 11 β -HSD 1 has become highly topical due to an interesting finding on an important endocrinological role of 11 β -HSD 1. It has already previously been speculated that GCs have a pivotal role in the pathogenesis of central obesity (68,69). Recently, Masuzaki *et al.* (70) showed that transgenic mice, harboring an additional gene in adipocytes coding for rat liver 11 β -HSD 1, show all symptoms of central obesity (cf. Section XII.I of this review). It is now well established that 11 β -HSD 1 has a crucial role in modulating corticosteroid hormone action at the tissue level (Fig. 2), and that alterations in enzyme activity may explain a diverse array of disease processes.

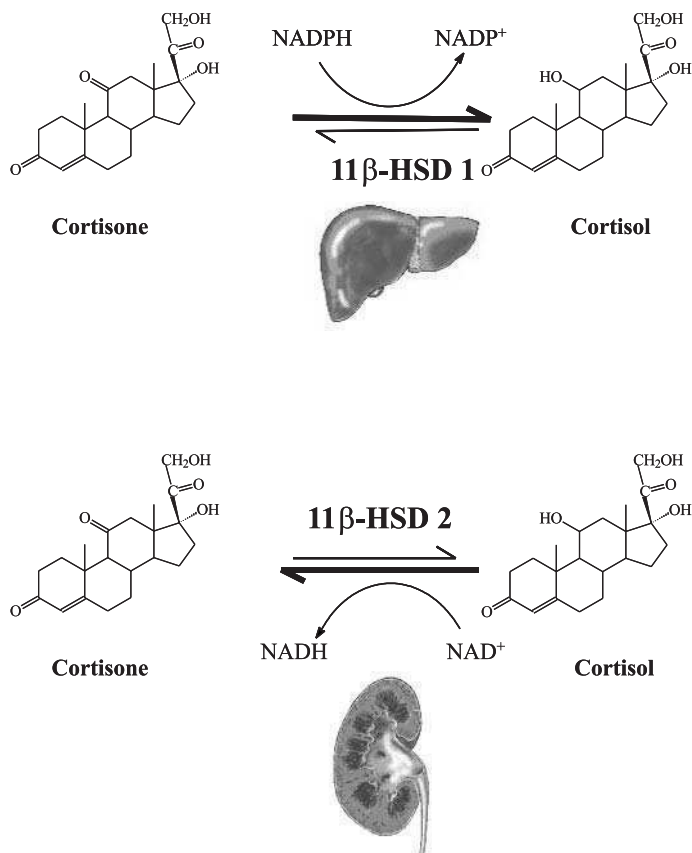


FIG. 1. Predominant *in vivo* reaction direction of 11 β -HSD 1 and 11 β -HSD 2. 11 β -HSD 1 distributes ubiquitously with major site of expression in liver and adipose tissues. It acts as an oxoreductase and generates receptor-active cortisol from inactive cortisone. By catalyzing inactivation of cortisol to cortisone, 11 β -HSD 2 protects the MR in aldosterone target tissues.

XI. 11 β -HSD 2 and the Syndrome of "Apparent Mineralocorticoid Excess" (AME)

Intracellular concentration of ligand is of particular importance when considering the actions of GCs and mineralocorticoids. Mineralocorticoid receptors (MRs) and GRs lack GC-binding specificity and both bind cortisol with high affinity, but only the MR can bind aldosterone. Thus, in tissues where the MR is expressed, specificity of action is provided by coexpression of the enzyme 11 β -HSD 2. Type 2 11 β -HSD is a high affinity, unidirectional, NAD-dependent dehydrogenase with an apparent K_m for cortisol in the

TABLE I
CHARACTERISTICS OF 11 β -HSD 1 AND 11 β -HSD 2

	11 β -HSD 1	11 β -HSD 2
Distribution	Ubiquitous (liver, adipose tissue)	Aldosterone target tissues, placenta
Reaction direction	Bidirectional (<i>in vitro</i>) Reductase (<i>in vivo</i>)	Unidirectional (dehydrogenase)
Cofactor	NADP(H)	NAD
Kinetics for GCs	Low affinity ($K_m \sim \mu M$) Michaelis–Menten (for oxidation) Cooperativity (for reduction)	High affinity ($K_m \sim nM$) Michaelis–Menten (for oxidation) –
Molecular mass	34 kDa	40 kDa
Oligomeric forms	Dimer	?
Physiological role	Regulates GC access to GR	Protects MR

Abbreviations: GC, glucocorticoids; GR, glucocorticoid receptor; MR, mineralocorticoid receptor.

nanomolar range (63,71,72). This isozyme is found principally in mineralocorticoid target tissues such as the kidney and colon, where it (by inactivating GCs) protects the MR from cortisol excess, thereby ensuring mineralocorticoid specificity of aldosterone (56,65). A defect in the 11 β -HSD 2 gene leads to binding of cortisol to the unprotected MR, resulting in a form of hypertension called AME (73).

The condition was first described in the early 1970s, although the association of mineralocorticoid hypertension with defective peripheral cortisol metabolism was not established until the pioneering studies of Ulick *et al.* several years later (74). Despite an extensive search for the mineralocorticoid responsible for the clinical feature of AME, mineralocorticoid bioassays failed to detect activity in the plasma or urine of such patients. In 1983 it was first suggested that hydrocortisone, but not aldosterone, acted as a mineralocorticoid in a patient with AME, and a defect in the renal MR was hypothesized to account for this (75). During a series of metabolic imbalance studies, it became clear that cortisol was indeed the mineralocorticoid responsible for the condition. Children presented with symptoms of hypokalemia, i.e., thirst, polyuria, failure to thrive, and weakness. The hypokalemic nephropathy frequently resulted in nephrocalcinosis and renal cysts. Rare manifestations included bone disease (76). The hypertension was severe, with a 10% mortality rate, secondary to intracranial hemorrhage, or complications of hypokalemia (77).

More recently, the so-called type 2 variant of AME has been reported in children (78). These children presented as having AME, but had notable differences in cortisol metabolism, i.e., a normal THF + allo-THF:THE ratio, and the principal abnormality appeared to be a generalized impairment of cortisol A ring reduction.

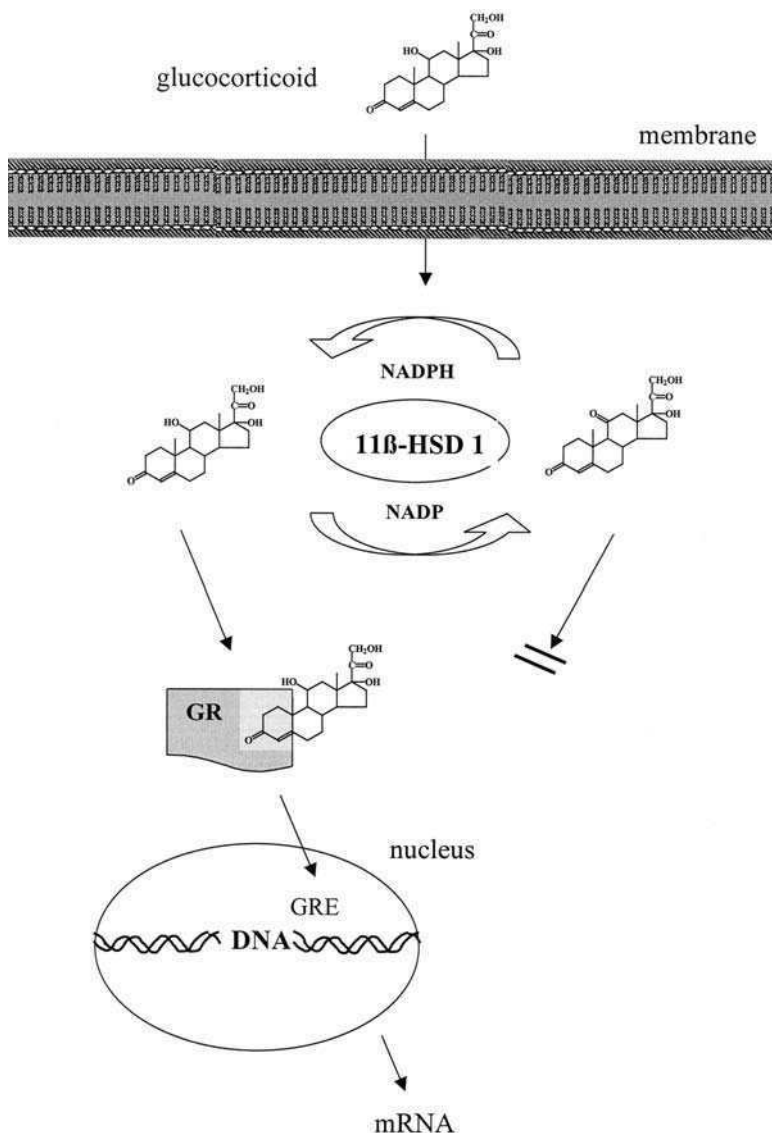


FIG. 2. 11β -HSD 1 as prereceptor regulative in intracellular GC signaling. Glucocorticoids enter the cell by diffusion. Within the cell, the enzyme 11β -HSD 1 acts as prereceptor-control device in that it allows active cortisol to bind to the GR. Upon binding of cortisol to the receptor, the steroid-receptor complex dimerizes (not shown) and translocates into the nucleus to mediate target gene transcription. By inactivating cortisol to cortisone, 11β -HSD 1 prevents receptor activation. Accordingly, the function of 11β -HSD 1 can be considered as enzymatic regulator in the signaling of GC hormones.

XII. 11 β -Hydroxysteroid Dehydrogenase Type 1 (11 β -HSD 1)

A. 11 β -HSD 1, a Dehydrogenase or Oxoreductase?

11 β -HSD 1 is widely expressed, most notably in liver, lung, testis, adipose tissue, vasculature, ovary, and the CNS. Highest amounts of this isoform have been found in rodent and human liver, and it exhibits exclusive reductase activity in primary cultures of rat and human hepatocytes (79,80). Preferential oxoreductase activities were confirmed by gene deletion experiments in the mouse (81) and perfusion studies in the rat (82). In human and rat liver, 11 β -HSD 1 is localized centripetally, with maximum expression around the central vein (83,84). The first purified 11 β -HSD 1 from rat liver exhibited dehydrogenase activity only. This has led to some initial controversy over whether 11 β -HSD 1 oxidase and reductase activity reside in one or two different proteins (55,85). However, studies on the purified enzyme from mouse and human liver proved the concept of a unique, reversible GC oxidoreductase (86,87).

These results, together with the finding that 11 β -HSD 1 in testis acts above all as a dehydrogenase, underline the fact that 11 β -HSD 1 is a bidirectional enzyme. In principle, the preferential reaction of 11 β -HSD 1 may be dependent on the cellular context, but also translational modifications and the subcellular environment, e.g., redox state or cosubstrate availability, may come into account. However, the subcellular environment could be ruled out as a determinant for the preferred reaction direction, since manipulation of hepatocyte culture media regarding pH, as well as metabolic inhibitors and altered cellular NADP⁺/NADPH ratios, had no effects on 11 β -dehydrogenase or reductase activities (79). In rat testis, 11 β -HSD activity is localized uniquely in Leydig cells (88), displaying a preference for NADP (89). Together with immunological studies (90), this indicates that 11 β -HSD 1 is acting as a dehydrogenase in testis. Hence, a putative third isoform of 11 β -HSD, which would explain the reaction direction opposite to that of liver and which was postulated (91–93), could at present be excluded. However, the existence of a further NADP-dependent 11 β -HSD isoform should not be ruled out.

B. Regulation of Reaction Direction

The conversion of active cortisol to inactive cortisone prevents the GC-mediated inhibition of steroidogenesis in Leydig cells (94–97) and inhibition of the testicular 11 β -HSD 1 suppresses testosterone production in Leydig cells (98). Ge and Hardy (99) showed that 11 β -HSD 1 oxidative and reductive

activities in rat Leydig cells are regulated inversely by two separate intracellular pathways. Protein kinase C signaling stimulated the 11β -HSD 1 dehydrogenase while inhibiting the reductase activity, maintaining a predominance of 11β -HSD 1 oxidation over reduction. In contrast, calcium-dependent signaling had the opposite effect, stimulating 11β -HSD 1 reductase, while inhibiting the dehydrogenase activity. The underlying mechanisms or determinants that direct the two described pathways can only be speculated. Two scenarios seem possible in case of protein kinase C and the calcium-signaling pathway. First, an indirect mechanism in that testosterone itself acts as inhibitor of 11β -HSD 1 dehydrogenase activity (100). Both, protein kinase C and calcium signaling have been shown to regulate testosterone production (101,102) and these, in turn, could affect 11β -HSD 1 activity via the action of testosterone. Second, a direct mechanism may come into question in that the two pathways separately phosphorylate 11β -HSD 1 serine/threonine residues near the basic residues, arginine and lysine, resulting in opposite 11β -HSD 1 activities. Two putative protein kinase C motifs have been identified in rat 11β -HSD 1. Phosphorylation of these amino acid residues could regulate cofactor binding and/or protein dimerization, as has been speculated by the authors. In case of cofactor binding, if phosphorylation of amino acid residues promotes binding of the NADP^+ cofactor, the oxidative activity of the enzyme would predominate over the reductive one. If this phosphorylation determines dehydrogenase or reductase activities of 11β -HSD 1, the enzyme in Leydig cell extracts or isolated microsomes should exhibit a preferred reaction direction. But these fractions behave like extracts obtained from liver regarding 11β -HSD 1 activity and kinetic constants in that both dehydrogenase as well as reductase activities were observed (Blum *et al.*, unpublished results).

11β -Hydroxylated steroid derivatives are efficient inhibitors of 11β -HSD 1 dehydrogenase activity, whereas 11 -keto steroid derivatives are competent inhibitors of oxoreductase activity (103). These compounds are formed endogenously through 11β -hydroxylase, an enzyme that contributes to GC biosynthesis, catalyzing the 11β -hydroxylation at the steroid nucleus and thus producing 11β -hydroxylated steroid derivatives. By performing RT-PCR, immunohistochemistry, and western-blot analysis, it was shown that 11β -hydroxylase is also expressed in testis, which supports the speculation that 11β -HSD 1, and hence the active/inactive GC equilibrium, within the testis is regulated through local biosynthesis of endogenous steroid inhibitors. These results were supported by investigations with rat and human hepatocytes, revealing that thyroid hormone and progesterone exert species-specific effects on the 11β -HSD 1 activities, while having no effect on the mRNA levels (83). Furthermore, these results are in agreement with the indirect hypothesis provided by Ge and Hardy (99).

A switch in dehydrogenase to reductase activity of 11 β -HSD 1 has also been reported upon differentiation of human omental adipose stromal cells (ASCs) (104). In undifferentiated omental ASCs, 11 β -HSD 1 acted primarily as a dehydrogenase, whereas in mature adipocytes oxoreductase activity predominated.

In conclusion, these findings reveal the importance of determining the major control parameters that govern the reaction direction of 11 β -HSD 1 *in vivo*, since shifting the balance between reduction and dehydrogenation might be an alternative approach to manipulating tissue GC levels.

C. 11 β -HSD 1 is a Dimeric Enzyme and Exhibits Cooperativity in Glucocorticoid Activation

In contrast to 11 β -HSD 2, the functional role of 11 β -HSD 1 in cell physiology has long been discussed. This was partly due to discrepancies in low levels of circulating GC hormones and high-binding affinities of GCs to the GR (both in low-nanomolar concentrations) on the one side, and the relatively low affinities of GC substrates to the 11 β -HSD 1 enzyme (in micromolar concentrations) on the other. It is clear that homogenously purified enzymes are a prerequisite to elucidate their physiological functions, but 11 β -HSD 1 is located within the membranes of the endoplasmic reticulum (ER) and many attempts failed to purify this enzyme in active form from human tissues. In previous studies, 11 β -HSD 1 has been purified from rat (55) and mouse (86) liver. However, whereas the mouse enzyme showed both oxidase and reductase activities in purified form, rat liver 11 β -HSD 1 failed to exhibit reductive activities toward its GC substrates, which is remarkable since the physiological function of this enzyme has been proposed to be the 11-oxoreduction of GCs. This failure has later been discussed in terms of the sensitivity of this enzyme to changes in its three-dimensional architecture, as a consequence of alterations within the lipid or detergent environment during the isolation procedure (41). Recently, 11 β -HSD 1 from human liver microsomes was isolated in an active state by a purification procedure that afforded a gentle solubilization method as well as providing a favorable detergent surrounding during the various chromatographic steps (87).

Most of the short-chain dehydrogenase/reductase (SDR) enzymes are active as either homodimers or -tetramers (47,105–109). By gel-permeation chromatography, it was shown that 11 β -HSD 1 occurs as dimeric enzyme in human liver (87). Already in previous studies, a second band corresponding to approximately 68 kDa has been observed in immunoblots with total protein from rat (90) and human (83) liver, suggesting that 11 β -HSD 1 may exist as a dimer in this tissue. In addition, results from nonreducing gel electrophoresis

and mutation of a nonconserved cysteine residue in recombinant 11 β -HSD 1 led Walker *et al.* (110) to conclude that 11 β -HSD 1 in human liver may be dimeric, where Cys272 is involved in dimerization via disulfide bonds between adjacent polypeptide chains of the enzyme.

The kinetic data of 11 β -HSD 1 published so far demonstrated an enzyme with a micromolar affinity for GCs (86,111–115). This hardly seemed compatible with a role of 11 β -HSD 1 to regulate access of GCs to the GR since both endogenous-“free” GC levels, as well as the dissociation constants of the GR for GCs, range in low-nanomolar concentrations. This enigma was solved by functional studies and detailed kinetic analyses with purified 11 β -HSD 1 from human liver (87). When Michaelis–Menten kinetics were considered with the purified enzyme, the K_m values for GC oxidoreduction ranged in low micromolar concentrations. However, oligomeric enzymes often display kinetics other than that described by the Michaelis–Menten equation. When applying the sigmoidal dose-response kinetics, cooperative kinetics of 11 β -HSD 1 with cortisone and dehydrocorticosterone (glucocorticoid 11-oxoreducing activity) were observed (87). Accordingly, the 11 β -HSD 1 enzyme is obviously able to adapt to low (nanomolar) as well as to high (micromolar) substrate concentrations.

Endogenous GC profiles are highly variable (ranging between 0.1 and 100 nm) in that they are triggered by the circadian rhythm and influenced by external challenges (e.g., stress). On the other hand, access of active GCs to the GR (K_d for GCs in the low-nanomolar region) requires some fine-tuning which could be best achieved by an enzyme that responds to these wide variations by dynamically adapting to low as well as to high endogenous GC levels. The fact that 11 β -HSD 1 shows this cooperative kinetics for 11-oxoreduction but not for 11-dehydrogenation of GCs is compatible with the concept of this enzyme to have its real physiological role as glucocorticoid 11-oxoreductase.

D. 11 β -HSD 1 Structural Determinants

11 β -HSD 1 belongs to the short-chain dehydrogenase/reductase (SDR) protein superfamily, in which, at present, about 3000 primary structures are annotated in sequence databases, and corresponding sources represent all forms of life (116,117). All the enzymes of this protein superfamily have some common structural features, although the amino acid sequence identity is very low (only 15–30%) (106). One of these structural features of the tertiary structure has become known as the *Rossmann fold*, a sequence of alternating α -helices and β -strands, which has been found in all tertiary structures of SDR enzymes solved so far. These β -strands form a four- or five-stranded parallel β -sheet with two or three α -helices residing on either side, which is characteristic of many NADH- and NADPH-binding domains. A comparison

of the conformation of five SDR structures revealed that, although there are only 11 fully conserved residues common to the five structures, the three-dimensional conformation is highly conserved (118). These findings were confirmed by newly solved three-dimensional structures of SDR enzymes. At present, there are 27 crystal structures of SDR enzymes available (119). Furthermore, SDR enzymes have a catalytically active triad comprising Ser, Tyr, and Lys residues, and a similar conserved sequence including three glycine residues (GxxGxG) at a comparable region in their sequences, which forms a turn between a β -strand and α -helix that borders on the cofactor-binding site (120). Both motifs have also been found in 11 β -HSD 1.

A single disulfide bond formation between Cys77 and Cys212 has been described for rabbit 11 β -HSD 1 (121). However, bacterial expression of human 11 β -HSD 1 in *Escherichia coli* strains, which are unable to form disulfide bonds, indicated that intermolecular disulfide bonds are not essential for enzyme activity (110). Computer-based simulation and modeling programs together with site-directed mutagenesis experiments (122) gave some evidence that a modeled structure of 11 β -HSD 1, which was based on an alignment of the primary sequence of 11 β -HSD 1 with those of the known structures of 3 α ,20 β -HSD from *Streptomyces hydrogenans* (123) and Tropinone reductase 1 from *Datura stramonium* (124) may be a good approach to the three-dimensional structure of 11 β -HSD 1. Interestingly, glycyrrhizic acid and carbenoxolone, two very potent inhibitors of both 11 β -HSD enzymes, also act as competitive inhibitors of 3 α ,20 β -HSD (125). This finding suggests similarities of the catalytic center of 3 α ,20 β -HSD with that of the 11 β -HSD enzymes, a fact which eventually may be extended to the complete three-dimensional structure of 11 β -HSD 1.

Nevertheless, a key step for understanding the interaction of 11 β -HSD 1 with substrates and cofactors and the modeling of new specific inhibitors will be the crystallization and elucidation of the three-dimensional structure of 11 β -HSD 1.

E. 11 β -HSD 1 Gene Organization

The human gene of 11 β -HSD 1 has been cloned and its chromosomal localization determined. It comprises six exons, spanning approximately 9 kb, and maps to chromosome 1 (126). Differential splicing and promoter usage give rise to the transcription of three different 11 β -HSD 1 proteins, designated 11 β -HSD 1A, B, and C (127,128) (Fig. 3). 11 β -HSD 1A represents the full-length protein, whereas the 11 β -HSD 1B form is shortened by the peptide sequence encoded by the entire first exon and which was first described by Monder and Lakshmi (90) in rat brain. 11 β -HSD 1C lacks the peptide sequence encoded by exon 5 and has been described in sheep

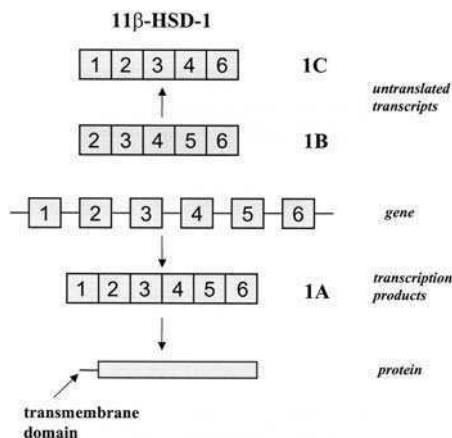


FIG. 3. Genomic organization of the human type 1 11β-HSD gene. The human gene of 11β-HSD 1 comprises six exons (boxed), spanning approximately 9 kb, and maps to chromosome 1 (126). Differential splicing and promoter usage give rise to the transcription of three different 11β-HSD 1 proteins, designated 11β-HSD 1A, B, and C (127–129). 11β-HSD 1A represents the full-length protein, whereas the 11β-HSD 1B form is shortened by the peptide sequence encoded by the entire first exon. 11β-HSD 1C lacks the peptide sequence encoded by exon 5. The corresponding 11β-HSD 1C protein has lost 48 amino acid residues, including the critical tyrosine needed for enzyme activity.

tissues (128) as being inactive. A cDNA bearing a deletion exactly matching exon 5 of the human 11β-HSD 1 gene and thus corresponding to the 11β-HSD 1C form has recently been described in pregnant uterus (129). The corresponding protein has lost 48 amino acid residues, including the critical tyrosine needed for enzyme activity. The authors hypothesized that such a protein might play a role as an alternative form of GC-binding protein without enzyme activity.

F. Truncated 11β-HSD 1B: is this Form of Physiological Relevance?

11β-HSD 1B was subject to many investigations in the early nineties. It was speculated that this protein might represent the second 11β-HSD activity, which was found in the kidney and which, in contrast to hepatic 11β-HSD 1, exhibited only dehydrogenase activity and was specific for NAD^+ (60,61). However, *in vitro* transcription/translation experiments and expression of rat 11β-HSD 1B in Chinese hamster ovary (CHO) (130) and COS (127) cells led to recombinant proteins that failed to exhibit any activity. The state of science regarding the question of whether an 11β-HSD 1B enzyme does exist and whether such a protein could play a physiological role was summarized by Stewart and Krozowski (131). Since then the interest on this topic faded, due to

the fact that expression experiments led to inactive enzyme forms only, and that a truncated 11 β -HSD 1B enzyme has not been detected in human tissues yet. On the other hand, human and sheep 11 β -HSD 2 had meanwhile been cloned, expressed, and characterized in mammalian cells and *Xenopus* oocytes (63,64).

The peptide of 11 β -HSD 1, corresponding to the first exon of the 11 β -HSD 1 gene serves as a membrane anchor which is responsible for the retention of the enzyme in the ER membrane with luminal orientation (121,132,133). Alignment with other SDR enzymes revealed that the SDR-core region of 11 β -HSD 1 is extended N-terminally by the peptide encoded by the first exon. The fact that the second exon starts with a codon for methionine raised the possibility that this codon could serve as an alternative start codon to physiologically produce the 11 β -HSD 1B variant.

In an ongoing study, to solve the structure of 11 β -HSD 1B, the earlier finding was another reason for expression experiments with 11 β -HSD 1B and the aim to obtain a soluble form of this membrane-bound enzyme (134). Deletion of the first 30 amino acid residues, however, did not improve the solubility of 11 β -HSD 1 expressed in *E. coli*. In contrast, the truncated enzyme expressed in the yeast *P. pastoris* still remained attached to the microsomal fraction, although lacking the membrane-spanning domain (134). This led to the hypothesis that there must exist at least one additional domain within the 11 β -HSD 1-core enzyme which is responsible for membrane attachment of the truncated form. Microsomal fractions containing this truncated human 11 β -HSD 1B form did not exhibit 11 β -HSD 1 activity. This activity, however, is prevented by attachment to the ER and is seen only upon release of 11 β -HSD 1B from the membrane by detergent solubilization. Unfortunately, solubilized 11 β -HSD 1B turned out to be unstable, such that the activity declines within a few hours. Computer-based simulations and modeling of the 11 β -HSD 1 structure, combined with "wet chemistry" (site-directed mutagenesis, heterologous expression, and activity determination), led to the identification of a hydrophobic domain which could interact with the ER membrane and which may be responsible for additional membrane attachment of truncated 11 β -HSD 1 forms (122). Combined, despite its occurrence in mammalian tissues 11 β -HSD 1B has obviously no physiological role, since it is either inactive while being attached to the ER or it is rapidly losing activity once being released from intracellular membranes.

The mechanism of inactivation of the truncated enzyme caused by membrane attachment remains speculative. Two explanations seem possible: first, the substrate and/or cofactor do not have access to the active site of 11 β -HSD 1B due to a sterical hindrance of this domain through the ER membrane; second, the truncated enzyme is tightly fixed to the ER membrane, such that the protein cannot undergo the conformational changes

required for the catalytic process. Although hydrophobic substrates like the GCs should be able to freely cross the ER membrane, the membrane-bound form of 11 β -HSD 1B was not found to exhibit any GC oxidoreduction activity. This observation leads to the assumption that, if the first hypothesis is correct, the failure of cofactor binding to the active site is critical for activity. On the other hand, the second hypothesis would be in agreement with results published recently, where 11 β -HSD 1 was shown to be active as a homodimer that exhibits enzyme cooperativity (87) (cf. Section XII.C of this review).

Walker *et al.* (110) reported first that a construct of 11 β -HSD 1, shortened by the first 23 amino acid residues, is still active. Therefore the question arose, whether the seven amino acid residues 24–30 on the N-terminus determine stability and activity of the catalytic domain of human 11 β -HSD 1. A hydrophobicity plot revealed that this peptide could be divided into two parts. The first part comprises amino acid residues with hydrophobic character, whereas the second part is represented by hydrophilic amino acid residues (135). Further work with deletion and point mutations within the 11 β -HSD 1 primary structure revealed that the hydrophobic N-terminal part has the function to anchor 11 β -HSD 1 in the ER membrane, whereas the hydrophilic part stabilizes the catalytic domain of 11 β -HSD 1 and acts as a spacer to prevent membrane attachment of the catalytic domain.

G. Glycosylation Status has No Impact on 11 β -HSD 1 Activity

Examination of the 11 β -HSD 1 primary structure of different species revealed the presence of putative N-linked glycosylation sites. The number of glycosylation sites within 11 β -HSD 1 differs between various species. In human and rabbit (121), three glycosylation sites have been found (Fig. 4), whereas in rat two (112,136) and in guinea pig only one exist (114). Alignment of the location of the glycosylation sites within the 11 β -HSD 1 primary structure from various species revealed that only one is highly conserved. Nevertheless, deglycosylation experiments of purified human and rabbit 11 β -HSD 1 enzymes, as well as expression of 11 β -HSD 1 in *E. coli* and the yeast *Pichia pastoris* (110,121,135,137) gave clear evidence that glycosylation is not necessary for proper enzyme function or correct protein folding.

H. Regulation of 11 β -HSD 1

1. IDENTIFICATION OF *CIS*-ACTING ELEMENTS

The promoter region of the rat 11 β -HSD 1 gene has been cloned (138) and several putative transcription factor binding sites have been identified, although the promoter apparently is lacking a TATA-box. A CCAAT sequence at –73 to –69 as well as a GCAAT sequence (CP-1-binding site) in the inverse

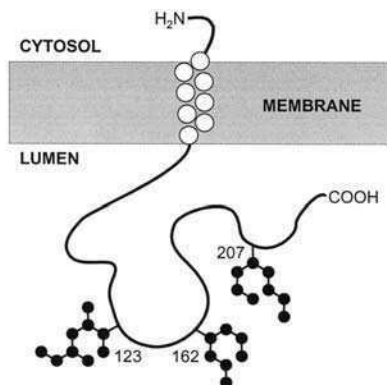


FIG. 4. Schematic representation of 11 β -HSD 1. 11 β -HSD 1 is anchored in the ER with luminal orientation of its active site. Open circles represent the amino acid residues of the N-terminal hydrophobic domain spanning the ER membrane. The three glycosylation sites are indicated at Asn 123, Asn 162, and Asn 207.

orientation between -63 and -67 has been found. Furthermore, six CACCC box elements at -476 , -466 , -427 , -290 , -282 , and -82 and a GC box at -299 , which is the consensus binding site for the transcription factor SP1, have been identified. Sequences corresponding to the consensus half-site GC-responsive element TGT^T/C²CT at $+61$ and $+250$, as well as the pentanucleotide TGACC, common in estrogen- and thyroid-responsive elements, at positions -839 , $+640$, and $+1300$ have also been reported.

Comparison of the sequence of the mouse and rat 11 β -HSD 1 gene 5'-flanking DNA region revealed that it is highly conserved (139). In the upstream region of the mouse 11 β -HSD 1 gene, several C/EBP-binding sites, a HNF 1 and HNF 3 site, together with binding sites for the nuclear receptor SF 1, the general transcription factors (activator protein) AP-1, AP-2 as well as a CACCC box are located (140). Members of both the HNF 1 and HNF 3 families of transcription factors have been implicated in the regulation of liver-specific genes in glucose metabolism (141) and, furthermore, members of the HNF 1 gene family may be involved in the induction of liver-specific genes by steroids (142). In addition, a binding site corresponding to a sequence recognized by members of the nuclear receptor family was found. This site is almost identical to those recognized by transcription factor SF 1. At present, for none of the above-mentioned *trans*-acting factors, any interactions with the promoter region or an influence on transcription of the 11 β -HSD 1 gene has been evidenced, with the exception of the members of the C/EBP transcription factor family (see next section). Like in the rat gene, no TATA-box is found in the mouse 11 β -HSD 1 gene, whereas a CAAT-box is located at -74 .

A TATA-box is also absent in the human 11 β -HSD 1 gene (126) and a CAAT sequence is located 76-bp upstream of the transcriptional start site. Furthermore, a palindromic sequence, CTGTACAG, resembling a GRE is present at -188, although some investigations in rat tissues suggested that GCs do not alter the level of 11 β -HSD 1 gene expression in liver, lung, or kidney (143).

2. INVOLVEMENT OF TRANS-ACTING ELEMENTS

The transcription factor C/EBP α is a potent enhancer of 11 β -HSD 1 gene transcription in hepatic cells (144). Interestingly, transcription factor C/EBP β , which alone is a relatively weak activator of 11 β -HSD 1 gene transcription, acts as a strong inhibitor of C/EBP α in 11 β -HSD 1 transcription. This result was confirmed *in vivo* in mice deficient in either C/EBP α or C/EBP β . In striking contrast to C/EBP α -deficient mice, which showed a dramatic decrease in hepatic expression of 11 β -HSD 1 mRNA, mice lacking C/EBP β showed an increased expression of 11 β -HSD 1 mRNA in their livers.

There are six known members of the C/EBP transcription factor family. Two, C/EBP α and C/EBP β , are expressed at significant levels in liver under basal conditions (145). C/EBP α is a central regulator of energy metabolism (146) and loss of C/EBP α in knockout mice dramatically alters energy metabolism (147). The involvement of C/EBP family members, particularly C/EBP α and C/EBP β and perhaps other members, in the regulation of the hepatic 11 β -HSD 1 gene suggests that C/EBP indirectly regulates the level of active intracellular GCs in the liver by governing the transcription of hepatic 11 β -HSD 1. On the other hand, GCs are important regulators of C/EBP transcription, markedly inducing C/EBP β and C/EBP δ mRNA (148–150), and with differing temporal and tissue-specific effects on C/EBP α mRNA (151,152). Under basal conditions in liver, the ratio of C/EBP α to C/EBP β will be high, triggering high levels of 11 β -HSD 1 transcription and hence local production of active GCs. Treatments that decrease the ratio of C/EBP α to C/EBP β would be predicted to decrease 11 β -HSD 1 transcription. This explanation could be accounted for by the reduced 11 β -HSD 1 activity found in some studies *in vitro* (153,154) and *in vivo* (155,156).

3. LXR

Recently, it could be shown that liver X receptors (LXR), which belong to the nuclear receptor superfamily of ligand-activated transcription factors with distinct oxysterols as identified endogenous activators (157,158), down-regulate 11 β -HSD 1 expression and activity (159) in adipocytes and liver. Activation of LXRs by endogenous and synthetic ligands causes down-regulation of 11 β -HSD 1 gene expression *in vitro* and *in vivo*, but the mechanism of down regulation is still unclear. Experiments in the presence of

cycloheximide revealed that ongoing protein synthesis is required for the effect of LXR agonist on 11 β -HSD 1 mRNA production, thus pointing to an indirect mechanism. The inhibition of 11 β -HSD 1 gene expression by LXR agonists seems to be very similar to the down regulation of 11 β -HSD 1 gene expression caused by peroxisome proliferator-activated receptor- γ (PPAR γ) ligands (160,161). Stulnig *et al.* (159) suggested that the action of PPAR γ ligands is mediated via LXR α because LXR α , the major LXR in adipose tissue and liver (162), is a direct target gene of PPAR γ (163–165). Furthermore, protein synthesis inhibition due to cycloheximide treatment abolished the effect of PPAR γ ligands on 11 β -HSD 1 gene expression, indicating that PPAR γ acts not directly on 11 β -HSD 1 promoter.

4. PPAR γ

C/EBP α and PPAR γ have been described as the most important transcription factors in adipogenesis. Gene-knockout studies demonstrated that PPAR γ is essential for adipocyte differentiation *in vivo* (166,167), whereas loss-of-function studies *in vivo* and *in vitro* have shown that C/EBP α is also required for adipogenesis (168,169). Both proteins positively regulate each other's expression. Cells lacking PPAR γ expressed greatly reduced levels of C/EBP α (166,167,170). Similarly, fibroblasts lacking C/EBP α have reduced adipogenic potential, and expressed reduced levels of PPAR γ (169). Importantly, adding PPAR γ back to C/EBP α $-/-$ fibroblasts with a retroviral vector restored their capacity to accumulate lipid and active markers of adipogenesis (reviewed in Refs. (171,172)).

5. GLUCOCORTICOIDS

Further factors like GCs, insulin, interleukin-1 β (IL-1 β), TNF α , growth hormone, thyroid and sex hormones, which have influence on 11 β -HSD 1 expression and activity, have been described in the literature. Combining all these studies reveals that regulation of 11 β -HSD 1 expression and activity is rather complicated and happens in a sex- (173,174), development- (175,176), and tissue- (155) dependent manner.

Glucocorticoid-induced 11 β -HSD 1 gene expression and activity has been reported, both in the liver *in vivo* (177,178) and in cell culture *in vitro* (79,179,180). Jamieson *et al.* (82) described a temporal-specific regulation of 11 β -HSD 1 gene expression and activity by GCs *in vivo* in rat liver and hippocampus. In the 2S FAZA rat hepatoma cell line, reductase activity is inhibited by insulin and stimulated by dexamethasone, a synthetic GC (153). 11 β -HSD 1-promoter/luciferase assays gave some evidence that dexamethasone acts directly on the proximal-promoter region of the 11 β -HSD 1 gene. This result is supported by investigations concerning the regulation of GR alpha (GR α) and beta (GR β) and 11 β -HSD 1 expression

in human skeletal muscle cells (181,182). The data indicate that in human skeletal myoblasts, 11 β -HSD 1 expression and activity is sensitively upregulated by physiological concentrations of cortisol in a dose-dependent manner. Furthermore, abolition of the GC-dependent induction of 11 β -HSD 1 due to treatment with RU38486, a GR α antagonist, showed that this effect is mediated exclusively by GR α . Two isoforms of GR have been described that comprise splice variants of the same gene (40). GR α is able to bind ligand, whereas the truncated β isoform (GR β), which is unable to bind ligand, is thought to act as a dominant negative inhibitor of GC hormone action through heterodimerization with GR α (40,183) (cf. Section IV of this review). Moreover, the expression of GR α is regulated by its own ligand, such that cortisol induces down regulation of GR α mRNA expression and stability, and increases the posttranslational turnover of GR α protein (183,184). The published data suggest that regulation of basal expression of GR α , its regulation by cortisol, and the GC-dependent regulation of 11 β -HSD 1 expression in human skeletal myoblasts are mediated exclusively by binding of cortisol to its receptor, and that the conversion of cortisone to cortisol by 11 β -HSD 1 activity may represent a mechanism of autoregulation of 11 β -HSD 1 activity by the product of its own catalytic activity.

Nevertheless, the last proof that GCs modulate their own metabolism, to be shown by *in vitro* transcription experiments or protein/DNA-binding assays determining the binding site for the GR on the 11 β -HSD 1 gene promoter, is still lacking. It should be noted, however, that GCs are potent antagonists of insulin action and, when in excess, can promote insulin resistance and obesity (185).

6. CYTOKINES

Glucocorticoids exert anti-inflammatory effects via the ligand-activated GR signaling pathway in inflamed tissues, to antagonize the induction of the proinflammatory transcription factor NF- κ B (186,187). Many investigations suggested that sites of inflammation express relatively high levels of 11 β -HSD 1, which presumably promotes ligand access to GR through its 11-oxoreductive catalytic activity. Cytokines like IL-1 β and TNF α have been reported as inducers of 11 β -HSD 1 gene expression in various cell types. IL-1 β upregulates 11 β -HSD 1 in cultured rat glomerular mesangial cells (188), ovarian granulosa cells (189), and human ovarian-surface epithelial cells (190), while TNF α upregulates 11 β -HSD 1 in human ASCs (191). Interestingly, Tomlinson *et al.* (192) reported in human subcutaneous and omental ASCs an upregulation of 11 β -HSD 1 expression and activity, whereas in hepatocytes, TNF α and IL-1 β had no effect on 11 β -HSD 1-oxoreductase

activity, indicating a tissue-specific regulation of 11 β -HSD 1 gene expression. TNF α , acting through the p60 receptor, signals either via mitogen-activated protein kinase and NF- κ B or through activator protein 1 (AP-1) (193). AP-1 transcription factor-binding sites have been identified in the 5'-promoter region of the mouse (139) and human (194) 11 β -HSD 1 gene (cf. Section XII.H.1 of this review). IL-1 β and TNF α also enhance 11 β -HSD 1 expression in human aortic smooth muscle cells. These results suggest that increased 11 β -HSD 1-reductase activity would generate a negative feedback upon inflammation and may be a generalized response to inflammation that ensures maximal accessibility of anti-inflammatory GCs to sites of tissue injury.

7. GROWTH HORMONE AND IGF-1

Growth hormone (GH) inhibits 11 β -HSD 1 via insulin-like growth factor-I (IGF-I) (195). In addition, IGF-I exhibits a tissue-specific influence on 11 β -HSD 1 activity (192), which is opposite to that of TNF α . Treatment with IGF-I in subcutaneous and omental ASC reduces 11 β -HSD 1 activity, whereas treatment of cultures of human hepatocytes has no effect. IGF-I acts via either inositol trisphosphate- or mitogen-activated protein kinase. The specific intracellular pathway that is activated can have profound effects on adipocyte development to promote either proliferation or differentiation (196). However, the precise mechanism of action of IGF-I to modulate 11 β -HSD 1 activity in ASC has not yet been determined. Interestingly, adult patients with endogenous GH deficiency are known to have a series of metabolic defects, including insulin resistance and obesity (197). Furthermore, it is known that obese patients show an increased metabolic-clearance rate of cortisol and a secondary enhanced ACTH-cortisol secretion (198,199). It could be shown that the reason for this increased clearance of cortisol in obesity appears to be secondary to inhibition of 11 β -HSD 1 in the liver, as reflected by a reduction in the THF/THE ratio and impaired generation of circulating cortisol following an oral dose of cortisone acetate (69,200). A change in cortisol metabolism has also been reported in patients with active acromegaly by treatment with the GH receptor-antagonist pegvisomant (201).

In conclusion, the puzzle of 11 β -HSD 1 gene regulation and expression is very incomplete and much work has to be done to elucidate the molecular mechanisms behind the 11 β -HSD 1 enzyme. Tissue-specific, differential expression of 11 β -HSD 1 has to be taken into account, due to the observation that shift in the set-point of cortisone to cortisol toward cortisone increases cortisol clearance and causes compensatory stimulation of the HPA axis to maintain normal circulating cortisol levels. Furthermore, it is known that

TNF α , the levels of which are elevated in obese subjects, fails to regulate 11 β -HSD 1 expression in human hepatocytes, whilst simultaneously stimulating enzyme expression in human omental adipose tissue (192). Knowledge about the mechanisms of 11 β -HSD 1 gene expression, on the other hand, will open new possibilities to affect 11 β -HSD 1 activity in the course of a therapeutic regimen.

1. 11 β -HSD 1 and the Metabolic Syndrome

1. A ROLE FOR CORTISOL IN CENTRAL OBESITY?

Given its clinical similarities to Cushing's syndrome, many investigators have considered cortisol as a pathophysiological mediator in idiopathic obesity. Although cortisol-secretion rate is enhanced in obesity (202,203), plasma-cortisol levels are not consistently elevated. Metabolic-clearance rate for cortisol is also increased in obesity (204). One possible explanation for this finding was given by the observation that reactivation of cortisone to cortisol by 11 β -HSD 1 in liver is impaired, and inactivation of cortisol by 5 α -reductase enzymes is enhanced (200,205). These changes in metabolism predict lower plasma-cortisol levels, which, by negative feedback, could explain a compensatory rise in ACTH and cortisol secretion. But this hypothesis could not be confirmed by clinical studies published by Rask *et al.* (69). In contrast, the data obtained from this study supported the hypothesis that increased 11 β -HSD 1 in omental adipose tissues is responsible for central obesity in man (81).

2. ADIPOSE TISSUE-SPECIFIC GLUCOCORTICOID ACTIVATION BY 11 β -HSD 1

Convincing data reported by Mazuzaki *et al.* (70) emphasized the pivotal role of 11 β -HSD 1 in central obesity. Mice overexpressing 11 β -HSD 1 in adipocytes not only showed an increase in enzyme activity, but also exhibited two- to three-fold higher local concentrations of corticosterone in mesenchymal adipose tissue. Most importantly, serum corticosterone concentration in these transgenic mice were similar to those in their nontransgenic littermates, which is consistent with findings in most adipose tissues in humans who are lacking systemic hypercortisolism. The body weight of transgenic animals receiving a high-fat diet exceeded that of nontransgenic littermates by 25–35%. Interestingly, the increase in weight was mostly confined to mesenchymal fat tissue. The quantity of visceral fat has long been identified as the best predictor of morbidity emerging from obesity (206). The overexpression did not lead to a higher number of adipose cells but induced

a more than threefold adipocyte volume, an observation similarly found in obese humans.

3. ELEVATED 11 β -HSD 1 LEVELS IN HUMAN ADIPOSE TISSUE

An *in situ* hybridization study with adipose tissue from obese patients showed that the 11 β -HSD 1 mRNA level was elevated (207), which is in line with results reported by Rask *et al.* (208). They found that in adipose tissue of obese women 11 β -HSD 1 activity was increased, a fact which has also been observed in obese males (69,209). The hypothesis that an altered 11 β -HSD 1 activity in adipocytes is responsible for obesity is supported by a DNA-microarray study which monitored the expression profile of 3T3-L1 fibroblasts during differentiation to mature adipocytes (210). In this study, it could be shown that 11 β -HSD 1 expression is induced by adipocyte differentiation. Furthermore, the observation that an overexpression of 11 β -HSD 1 may explain part of the GC-induced metabolic disorders linked to obesity and may therefore promote visceral fat deposition, is supported by the description of a patient with Cushing's disease with no obvious development of a Cushingoid phenotype (211). The authors postulate that the absence of a Cushingoid phenotype is due to a partial defect of 11 β -HSD 1 activity. The defect in cortisone to cortisol conversion increases the clearance of cortisol and thus protects the patient from the effects of cortisol excess. Similar effects have also been observed in 11 β -HSD 1-null mice (81,212). These 11 β -HSD 1-null mice exhibited attenuated gluconeogenic responses upon stress and resisted hyperglycemia induced by chronic high-fat feeding, and 11 β -HSD 1 deficiency may therefore oppose the pathogenic lipid and lipoprotein profile found in disease states such as the metabolic syndrome.

That this topic is rather complicated revealed a clinical study with paired omental and subcutaneous fat biopsies obtained from 32 women (213). This study shows that there is no evidence for overexpression of 11 β -HSD 1 in obesity in whole adipose tissue, adipocytes, or preadipocytes.

J. Inhibitors

11 β -HSD 1 can modulate GC action in key sites controlling metabolic-energy utilization. Enhanced 11 β -HSD 1 is important in increasing local GC action and promoting adverse metabolic effects. In the face of impaired 11 β -HSD 1 activity in liver, Zucker obese rats show selectively enhanced activity of 11 β -HSD 1 in omental adipose tissue (214). Other studies confirmed the same tissue-specific pattern of dysregulation of 11 β -HSD 1 (i.e., impaired in liver, enhanced in adipose tissue) in human obesity (69,209). Insulin sensitivity of human volunteers was significantly improved by

pharmacological inhibition of 11β -HSD 1-reductase activity through carbenoxolone administration (215). Unfortunately, carbenoxolone, the hemisuccinate derivative of 18β -glycyrrhetic acid and a potent inhibitor of 11β -HSD 1, and 18β -glycyrrhetic acid itself display poor selectivity between the two isoforms of 11β -HSD (216).

The medicinal properties of liquorice have been appreciated since ancient Chinese times. However, the association of liquorice and mineralocorticoid side effects came only in the middle of the nineties with hypertension and hypokalemia occurring in patients given liquorice for dyspepsia. The active mineralocorticoid in liquorice is the aglycone derivative of glycyrrhizic acid, glycyrrhetic acid. The carbenoxolone derivative of glycyrrhetic acid proved to be a very successful antiulcer drug. Its use, however, also results in hypertension, salt and water retention, and hypokalemia in up to 50% of patients (217). The cause of these mineralocorticoid side effects had been thought to be secondary to a direct action on the MR, with binding studies indicating a weak affinity of glycyrrhizic acid, glycyrrhetic acid, and carbenoxolone for the MR (218). Stewart *et al.* (219) showed that licorice acts as a mineralocorticoid via inhibition of 11β -HSD 2.

On the other hand, several other endo- and xenobiotics have been found to inhibit 11β -HSD 1 *in vivo* and *in vitro*, such as bile acids (in cholestasis), estrogens, progestagens or hydroxyprogesterones (in pregnancy or upon usage of oral contraceptives), naringenin (a bioflavonoid of grapefruit), pharmacological drugs such as dexamethasone and furosemide, or ethanol (reviewed in Ref. (220)).

But all these compounds have the disadvantage of a poor selectivity between the two isoforms of 11β -HSD. Selective inhibition of 11β -HSD 1 using arylsulfonamidothiazoles has been reported recently (221), and these compounds are considered as promising antidiabetic drugs. Treatment of hyperglycemic and -insulinemic mice with some arylsulfonamidothiazoles resulted in lowering of blood glucose and serum-insulin concentrations (222). Diederich *et al.* (223) reported how the steroid scaffold must be modified to affect 11β -HSD activities and, furthermore, to influence the reaction direction of the enzymes. They could show that 11β -HSD 2, an enzyme which has been reported to act only as a dehydrogenase, prefers the reductase rather than the dehydrogenase direction when using 9α -fluorinated GCs as substrates (224,225). Like that of 11β -HSD 2, dehydrogenase activity of 11β -HSD 1 was completely abolished using 6α - or 9α -fluorinated GCs, whereas a methyl group in 16α -position of the GC scaffold only decreased the oxidation rate by 11β -HSD 1. Reduction direction of 11β -HSD 1 was impaired by 16α -methyl, 16β -methyl, 2α -methyl, and 2-chloro substituents. The information given by this study could be very useful to design a synthetic steroid that acts exclusively as an inhibitor of 11β -HSD 1 but not of 11β -HSD 2.

Finally, it should be mentioned that some antidiabetic drugs, namely thiazolidinedione and nonthiazolidinedione PPAR γ agonists have been reported to markedly decreasing 11 β -HSD 1 activity and expression (161).

XIII. Conclusions

During the last two decades significant advances have been made toward understanding the precise mode of action of the GCs, and it now appears to be multifaceted involving local regulation of GR activation by prereceptor control enzymes. Clearly, HSDs like 11 β -HSD 1 are much more than simple dehydrogenases that convert inactive into active steroid hormones. In contrast, they act as molecular switches and are part of a dynamic and intricate network, linking and integrating steroid-signaling pathways, and regulatory processes in cells, tissues, and whole organisms. It was apparent that the level of complexity in hormone action arose from the diversity of receptor isoforms themselves. However, nature additionally favored the evolution of different HSD isoforms as natural selective hormone modulators that provide more finely tuned stimuli, or cell- or tissue-specific signaling networks. This does not only apply for 11 β -HSD 1 in GC signaling but also for a series of other HSDs such as 3 α - and 17 β -HSDs in sex hormone action. Thus, the paradox of how one steroid hormone can elicit such an enormous variety of cell- and tissue-specific responses is explained not only by the diversity of receptor isoforms, but also by the consequence of the action of HSDs as a prereceptor-modulating device.

REFERENCES

1. F. C. Dalman, L. C. Scherrer, L. P. Taylor, H. Akil and W. B. Pratt, Localization of the 90-kDa heat shock protein-binding site within the hormone-binding domain of the glucocorticoid receptor by peptide competition, *J. Biol. Chem.* **266**, 3482–3490 (1991).
2. W. B. Pratt, Glucocorticoid receptor structure and the initial events in signal transduction, *Prog. Clin. Biol. Res.* **322**, 119–132 (1990).
3. C. Scheidereit, S. Geisse, H. M. Westphal and M. Beato, The glucocorticoid receptor binds to defined nucleotide sequences near the promoter of mouse mammary tumour virus, *Nature* **304**, 749–752 (1983).
4. M. A. Carson-Jurica, W. T. Schrader and B. W. O'Malley, Steroid receptor family: structure and functions, *Endocr. Rev.* **11**, 201–220 (1990).
5. J. M. Berg, DNA binding specificity of steroid receptors, *Cell* **57**, 1065–1068 (1989).
6. W. Y. Almawi and O. K. Melemedjian, Negative regulation of nuclear factor-kappaB activation and function by glucocorticoids, *J. Mol. Endocrinol.* **28**, 69–78 (2002).
7. M. R. Yudt and J. A. Cidlowski, The glucocorticoid receptor: coding a diversity of proteins and responses through a single gene, *Mol. Endocrinol.* **16**, 1719–1726 (2002).

8. N. Auphan, J. A. DiDonato, C. Rosette, A. Helmberg and M. Karin, Immunosuppression by glucocorticoids: inhibition of NF-kappa B activity through induction of I kappa B synthesis, *Science* **270**, 286–290 (1995).
9. A. Mori, O. Kaminuma, M. Suko, S. Inoue, T. Ohmura, A. Hoshino, Y. Asakura, K. Miyazawa, T. Yokota, Y. Okumura, K. Ito, H. Okudaira, Two distinct pathways of interleukin-5 synthesis in allergen-specific human T-cell clones are suppressed by glucocorticoids, *Blood* **89**, 2891–2900 (1997).
10. C. Jonat, H. J. Rahmsdorf, K. K. Park, A. C. Cato, S. Gebel, H. Ponta and P. Herrlich, Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone, *Cell* **62**, 1189–1204 (1990).
11. R. M. Nissen and K. R. Yamamoto, The glucocorticoid receptor inhibits NFkappaB by interfering with serine-2 phosphorylation of the RNA polymerase II carboxy-terminal domain, *Genes Dev.* **14**, 2314–2329 (2000).
12. R. Newton, L. A. Hart, D. A. Stevens, M. Bergmann, L. E. Donnelly, I. M. Adcock and P. J. Barnes, Effect of dexamethasone on interleukin-1beta-(IL-1beta)-induced nuclear factor-kappaB (NF-kappaB) and kappaB-dependent transcription in epithelial cells, *Eur. J. Biochem.* **254**, 81–89 (1998).
13. Z. Zhang, S. Jones, J. S. Hagood, N. L. Fuentes and G. M. Fuller, STAT3 acts as a co-activator of glucocorticoid receptor signaling, *J. Biol. Chem.* **272**, 30607–30610 (1997).
14. W. Y. Almawi, D. A. Hess and M. J. Rieder, Multiplicity of glucocorticoid action in inhibiting allograft rejection, *Cell Transplant.* **7**, 511–523 (1998).
15. W. Y. Almawi, M. M. Abou Jaoude and X. C. Li, Transcriptional and post-transcriptional mechanisms of glucocorticoid antiproliferative effects, *Hematol. Oncol.* **20**, 17–32 (2002).
16. W. Y. Almawi, H. N. Beyhum, A. A. Rahme and M. J. Rieder, Regulation of cytokine and cytokine receptor expression by glucocorticoids, *J. Leukoc. Biol.* **60**, 563–572 (1996).
17. K. De Bosscher, M. L. Schmitz, W. Vanden Berghe, S. Plaisance, W. Fiers and G. Haegeman, Glucocorticoid-mediated repression of nuclear factor-kappaB-dependent transcription involves direct interference with transactivation, *Proc. Natl. Acad. Sci. USA* **94**, 13504–13509 (1997).
18. J. F. Dunn, B. C. Nisula and D. Rodbard, Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding globulin in human plasma, *J. Clin. Endocrinol. Metab.* **53**, 58–68 (1981).
19. C. W. Breuner and M. Orchinik, Plasma binding proteins as mediators of corticosteroid action in vertebrates, *J. Endocrinol.* **175**, 99–112 (2002).
20. W. Rosner, The functions of corticosteroid-binding globulin and sex hormone-binding globulin: recent advances, *Endocr. Rev.* **11**, 80–91 (1990).
21. S. M. Hollenberg and R. M. Evans, Multiple and cooperative trans-activation domains of the human glucocorticoid receptor, *Cell* **55**, 899–906 (1988).
22. O. M. Conneely, W. P. Sullivan, D. O. Toft, M. Birnbaumer, R. G. Cook, B. L. Maxwell, T. Zarucki-Schulz, G. L. Greene, W. T. Schrader, B. W. O'Malley, Molecular cloning of the chicken progesterone receptor, *Science* **233**, 767–770 (1986).
23. R. M. Evans, The steroid and thyroid hormone receptor superfamily, *Science* **240**, 889–895 (1988).
24. J. M. Wurtz, W. Bourguet, J. P. Renaud, V. Vivat, P. Chambon, D. Moras and H. Gronemeyer, A canonical structure for the ligand-binding domain of nuclear receptors, *Nat. Struct. Biol.* **3**, 206 (1996).
25. M. Gangloff, M. Ruff, S. Eiler, S. Duclaud, J. M. Wurtz and D. Moras, Crystal structure of a mutant hERalpha ligand-binding domain reveals key structural features for the mechanism of partial agonism, *J. Biol. Chem.* **276**, 15059–15065 (2001).

26. S. M. Hollenberg, C. Weinberger, E. S. Ong, G. Cerelli, A. Oro, R. Lebo, E. B. Thompson, M. G. Rosenfeld and R. M. Evans, Primary structure and expression of a functional human glucocorticoid receptor cDNA, *Nature* **318**, 635–641 (1985).
27. R. H. Oakley, C. M. Jewell, M. R. Yudt, D. M. Bofetiado and J. A. Cidlowski, The dominant negative activity of the human glucocorticoid receptor beta isoform. Specificity and mechanisms of action, *J. Biol. Chem.* **274**, 27857–27866 (1999).
28. J. C. Webster, R. H. Oakley, C. M. Jewell and J. A. Cidlowski, Proinflammatory cytokines regulate human glucocorticoid receptor gene expression and lead to the accumulation of the dominant negative beta isoform: a mechanism for the generation of glucocorticoid resistance, *Proc. Natl. Acad. Sci. USA* **98**, 6865–6870 (2001).
29. M. Wehling, Specific, nongenomic actions of steroid hormones, *Annu. Rev. Physiol.* **59**, 365–393 (1997).
30. F. Chen, C. S. Watson and B. Gametchu, Association of the glucocorticoid receptor alternatively-spliced transcript 1A with the presence of the high molecular weight membrane glucocorticoid receptor in mouse lymphoma cells, *J. Cell Biochem.* **74**, 430–446 (1999).
31. E. Falkenstein, C. Meyer, C. Eisen, P. C. Scriba and M. Wehling, Full-length cDNA sequence of a progesterone membrane-binding protein from porcine vascular smooth muscle cells, *Biochem. Biophys. Res. Commun.* **229**, 86–89 (1996).
32. T. Simoncini, A. Hafezi-Moghadam, D. P. Brazil, K. Ley, W. W. Chin and J. K. Liao, Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase, *Nature* **407**, 538–541 (2000).
33. M. C. Keightley and P. J. Fuller, Cortisol resistance and the guinea pig glucocorticoid receptor, *Steroids* **60**, 87–92 (1995).
34. P. D. Reynolds, S. J. Pittler and J. G. Scammell, Cloning and expression of the glucocorticoid receptor from the squirrel monkey (*Saimiri boliviensis boliviensis*), a glucocorticoid-resistant primate, *J. Clin. Endocrinol. Metab.* **82**, 465–472 (1997).
35. A. G. Hillmann, J. Ramdas, K. Multanen, M. R. Norman and J. M. Harmon, Glucocorticoid receptor gene mutations in leukemic cells acquired in vitro and in vivo, *Cancer Res.* **60**, 2056–2062 (2000).
36. L. Jacobson and R. Sapolsky, The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis, *Endocr. Rev.* **12**, 118–134 (1991).
37. G. P. Chrousos, A. Vingerhoeds, D. Brandon, C. Eil, M. Pugeat, M. DeVroede, D. L. Loriaux and M. B. Lipsett, Primary cortisol resistance in man. A glucocorticoid receptor-mediated disease, *J. Clin. Invest.* **69**, 1261–1269 (1982).
38. S. W. Lamberts, D. Poldermans, M. Zweens and F. H. de Jong, Familial cortisol resistance: differential diagnostic and therapeutic aspects, *J. Clin. Endocrinol. Metab.* **63**, 1328–1333 (1986).
39. M. Bronnegard, P. Stiern and C. Marcus, Glucocorticoid resistant syndromes—molecular basis and clinical presentations, *J. Neuroendocrinol.* **8**, 405–415 (1996).
40. C. M. Bamberger, H. M. Schulte and G. P. Chrousos, Molecular determinants of glucocorticoid receptor function and tissue sensitivity to glucocorticoids, *Endocr. Rev.* **17**, 245–261 (1996).
41. C. Monder, Corticosteroids, receptors, and the organ-specific functions of 11 beta-hydroxysteroid dehydrogenase, *FASEB J.* **5**, 3047–3054 (1991).
42. R. Benediktsson and C. R. Edwards, 11 beta-Hydroxysteroid dehydrogenases: tissue-specific dictators of glucocorticoid action, *Essays Biochem.* **31**, 23–36 (1996).
43. J. W. Funder, Glucocorticoid and mineralocorticoid receptors: biology and clinical relevance, *Annu. Rev. Med.* **48**, 231–240 (1997).

44. T. M. Penning, M. E. Burczynski, J. M. Jez, C. F. Hung, H. K. Lin, H. Ma, M. Moore, N. Palackal and K. Ratnam, Human 3 α -hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones, *Biochem. J.* **351**, 67–77 (2000).
45. J. Adamski and F. J. Jakob, A guide to 17 β -hydroxysteroid dehydrogenases, *Mol. Cell. Endocrinol.* **171**, 1–4 (2001).
46. M. E. Baker, Evolution of 17 β -hydroxysteroid dehydrogenases and their role in androgen, estrogen and retinoid action, *Mol. Cell. Endocrinol.* **171**, 211–215 (2001).
47. T. M. Penning, Molecular endocrinology of hydroxysteroid dehydrogenases, *Endocr. Rev.* **18**, 281–305 (1997).
48. C. Monder and P. C. White, 11 β -hydroxysteroid dehydrogenase, *Vitam. Horm.* **47**, 187–271 (1993).
49. F. Labrie, V. Luu-The, S. X. Lin, C. Labrie, J. Simard, R. Breton and A. Belanger, The key role of 17 β -hydroxysteroid dehydrogenase in sex steroid biology, *Steroids* **62**, 148–158 (1997).
50. T. M. Penning, J. E. Pawlowski, B. P. Schlegel, J. M. Jez, H. K. Lin, S. S. Hoog, M. J. Bennett and M. Lewis, Mammalian 3 α -hydroxysteroid dehydrogenases, *Steroids* **61**, 508–523 (1996).
51. D. Amelung, H. J. Huebner, L. Roka and G. Meyerheim, Conversion of cortisone to compound F, *J. Clin. Endocrinol. Metab.* **13**, 1125–1129 (1953).
52. V. B. Mahesh and F. J. Ulrich, Metabolism of cortisol and cortisone by various tissues and subcellular particles, *J. Biol. Chem.* **235**, 356–360 (1960).
53. P. A. Osinski, Steroid 11-dehydrogenase in human placenta, *Nature* **187**, 777 (1960).
54. D. R. Koerner, 11 β -hydroxysteroid dehydrogenase of lung and testis, *Endocrinology* **79**, 935–938 (1966).
55. V. Lakshmi and C. Monder, Purification and characterization of the corticosteroid 11 β -dehydrogenase component of the rat liver 11 β -hydroxysteroid dehydrogenase complex, *Endocrinology* **123**, 2390–2398 (1988).
56. C. R. Edwards, P. M. Stewart, D. Burt, L. Brett, M. A. McIntyre, W. S. Sutanto, E. R. de Kloet and C. Monder, Localisation of 11 β -hydroxysteroid dehydrogenase—tissue specific protector of the mineralocorticoid receptor, *Lancet* **2**, 986–989 (1988).
57. R. Castello, R. Schwarting, C. Muller and K. Hierholzer, Immunohistochemical localization of 11-hydroxysteroid dehydrogenase in rat kidney with monoclonal antibody, *Renal. Physiol. Biochem.* **12**, 320–327 (1989).
58. S. E. Rundle, J. W. Funder, V. Lakshmi and C. Monder, The intrarenal localization of mineralocorticoid receptors and 11 β -dehydrogenase: immunocytochemical studies, *Endocrinology* **125**, 1700–1704 (1989).
59. J. P. Bonvalet, I. Doignon, M. Blot Chabaud, P. Pradelles and N. Farman, Distribution of 11 β -hydroxysteroid dehydrogenase along the rabbit nephron, *J. Clin. Invest.* **86**, 832–837 (1990).
60. A. Naray-Fejes-Toth, C. O. Watlington and G. Fejes-Toth, 11 β -Hydroxysteroid dehydrogenase activity in the renal target cells of aldosterone, *Endocrinology* **129**, 17–21 (1991).
61. B. R. Walker, J. C. Campbell, B. C. Williams and C. R. Edwards, Tissue-specific distribution of the NAD(+) -dependent isoform of 11 β -hydroxysteroid dehydrogenase, *Endocrinology* **131**, 970–972 (1992).
62. W. R. Mercer and Z. S. Krozowski, Localization of an 11 β hydroxysteroid dehydrogenase activity to the distal nephron. Evidence for the existence of two species of dehydrogenase in the rat kidney, *Endocrinology* **130**, 540–543 (1992).
63. A. L. Albiston, V. R. Obeyesekere, R. E. Smith and Z. S. Krozowski, Cloning and tissue distribution of the human 11 β -hydroxysteroid dehydrogenase type 2 enzyme, *Mol. Cell. Endocrinol.* **105**, R11–R17 (1994).

64. A. K. Agarwal, T. Mune, C. Monder and P. White, NAD⁺-dependent isoform of 11 β -hydroxysteroid dehydrogenase. Cloning and characterization of cDNA from sheep kidney, *J. Biol. Chem.* **269**, 25959–25962 (1994).
65. J. W. Funder, P. T. Pearce, R. Smith and A. I. Smith, Mineralocorticoid action: target specificity is enzyme, not receptor mediated, *Science* **242**, 583–585 (1988).
66. G. Phillipou and B. A. Higgins, A new defect in the peripheral conversion of cortisone to cortisol, *J. Steroid Biochem.* **22**, 435–436 (1985).
67. A. Jamieson, A. M. Wallace, R. Andrew, B. S. Nunez, B. R. Walker, R. Fraser, P. C. White and J. M. Connell, Apparent cortisone reductase deficiency: a functional defect in 11 β -hydroxysteroid dehydrogenase type 1, *J. Clin. Endocrinol. Metab.* **84**, 3570–3574 (1999).
68. I. J. Bujalska, S. Kumar and P. M. Stewart, Does central obesity reflect “Cushing’s disease of the omentum”?, *Lancet* **349**, 1210–1213 (1997).
69. E. Rask, T. Olsson, S. Soderberg, R. Andrew, D. E. Livingstone, O. Johnson and B. R. Walker, Tissue-specific dysregulation of cortisol metabolism in human obesity, *J. Clin. Endocrinol. Metab.* **86**, 1418–1421 (2001).
70. H. Masuzaki, J. Paterson, H. Shinyama, N. M. Morton, J. J. Mullins, J. R. Seckl and J. S. Flier, A transgenic model of visceral obesity and the metabolic syndrome, *Science* **294**, 2166–2170 (2001).
71. R. W. Brown, K. E. Chapman, C. R. Edwards and J. R. Seckl, Human placental 11 β -hydroxysteroid dehydrogenase: evidence for and partial purification of a distinct NAD-dependent isoform, *Endocrinology* **132**, 2614–2621 (1993).
72. P. M. Stewart, B. A. Murry and J. I. Mason, Human kidney 11 β -hydroxysteroid dehydrogenase is a high affinity nicotinamide adenine dinucleotide-dependent enzyme and differs from the cloned type I isoform, *J. Clin. Endocrinol. Metab.* **79**, 480–484 (1994).
73. R. C. Wilson, S. Dave-Sharma, J. Q. Wei, V. R. Obeyesekere, K. Li, P. Ferrari, Z. S. Krozowski, C. H. Shackleton, L. Bradlow, T. Wiens, M. I. New, A genetic defect resulting in mild low-renin hypertension, *Proc. Natl. Acad. Sci. USA* **95**, 10200–10205 (1998).
74. S. Ulick, L. S. Levine, P. Gunczler, G. Zanconato, L. C. Ramirez, W. Rauh, A. Rosler, H. L. Bradlow and M. I. New, A syndrome of apparent mineralocorticoid excess associated with defects in the peripheral metabolism of cortisol, *J. Clin. Endocrinol. Metab.* **49**, 757–764 (1979).
75. S. E. Oberfield, L. S. Levine, R. M. Carey, F. Greig, S. Ulick and M. I. New, Metabolic and blood pressure responses to hydrocortisone in the syndrome of apparent mineralocorticoid excess, *J. Clin. Endocrinol. Metab.* **56**, 332–339 (1983).
76. M. C. Batista, B. B. Mendonca, C. E. Kater, I. J. Arnhold, A. Rocha, W. Nicolau and W. Bloise, Spironolactone-reversible rickets associated with 11 β -hydroxysteroid dehydrogenase deficiency syndrome, *J. Pediatr.* **109**, 989–993 (1986).
77. C. H. Shackleton, J. Rodriguez, E. Arteaga, J. M. Lopez and J. S. Winter, Congenital 11 β -hydroxysteroid dehydrogenase deficiency associated with juvenile hypertension: corticosteroid metabolite profiles of four patients and their families, *Clin. Endocrinol. (Oxf)* **22**, 701–712 (1985).
78. F. Mantero, R. Tedde, G. Opocher, F. P. Dessi, G. Arnaldi and S. Ulick, Apparent mineralocorticoid excess type II, *Steroids* **59**, 80–83 (1994).
79. P. M. Jamieson, K. E. Chapman, C. R. Edwards and J. R. Seckl, 11 β -hydroxysteroid dehydrogenase is an exclusive 11 β -reductase in primary cultures of rat hepatocytes: effect of physicochemical and hormonal manipulations, *Endocrinology* **136**, 4754–4761 (1995).
80. M. L. Ricketts, K. J. Shoesmith, M. Hewison, A. Strain, M. C. Eggo and P. M. Stewart, Regulation of 11 β -hydroxysteroid dehydrogenase type 1 in primary cultures of rat and human hepatocytes, *J. Endocrinol.* **156**, 159–168 (1998).

81. Y. Kotelevtsev, M. C. Holmes, A. Burchell, P. M. Houston, D. Schmoll, P. Jamieson, R. Best, R. Brown, C. R. W. Edwards, J. R. Seckl, J. J. Mullins, 11 β -Hydroxysteroid dehydrogenase type 1 knockout mice show attenuated glucocorticoid-inducible responses and resist hyperglycemia on obesity or stress, *Proc. Natl. Acad. Sci. USA* **94**, 14924–14929 (1997).
82. P. M. Jamieson, B. R. Walker, K. E. Chapman, R. Andrew, S. Rossiter and J. R. Seckl, 11 beta-hydroxysteroid dehydrogenase type 1 is a predominant 11 beta-reductase in the intact perfused rat liver, *J. Endocrinol.* **165**, 685–692 (2000).
83. M. L. Ricketts, J. M. Verhaeg, I. Bujalska, A. J. Howie, W. E. Rainey and P. M. Stewart, Immunohistochemical localization of type 1 11beta-hydroxysteroid dehydrogenase in human tissues, *J. Clin. Endocrinol. Metab.* **83**, 1325–1335 (1998).
84. P. S. Brereton, R. R. van Driel, F. Suhaimi, K. Koyama, R. Dille and Z. Krozowski, Light and electron microscopy localization of the 11beta-hydroxysteroid dehydrogenase type I enzyme in the rat, *Endocrinology* **142**, 1644–1651 (2001).
85. C. Monder and V. Lakshmi, Evidence for kinetically distinct forms of corticosteroid 11 beta-dehydrogenase in rat liver microsomes, *J. Steroid Biochem.* **32**, 77–83 (1989).
86. E. Maser and G. Bannenberg, The purification of 11 β -hydroxysteroid dehydrogenase from mouse liver microsomes, *J. Steroid Biochem. Mol. Biol.* **48**, 257–263 (1994).
87. E. Maser, B. Volker and J. Friebertshauser, 11 Beta-hydroxysteroid dehydrogenase type 1 from human liver: dimerization and enzyme cooperativity support its postulated role as glucocorticoid reductase, *Biochemistry* **41**, 2459–2465 (2002).
88. D. M. Phillips, V. Lakshmi and C. Monder, Corticosteroid 11 beta-dehydrogenase in rat testis, *Endocrinology* **125**, 209–216 (1989).
89. H. B. Gao, L. X. Shan, C. Monder and M. P. Hardy, Suppression of endogenous corticosterone levels in vivo increases the steroidogenic capacity of purified rat Leydig cells in vitro, *Endocrinology* **137**, 1714–1718 (1996).
90. C. Monder and V. Lakshmi, Corticosteroid 11 beta-dehydrogenase of rat tissues: immunological studies, *Endocrinology* **126**, 2435–2443 (1990).
91. E. P. Gomez-Sanchez, V. Ganjam, Y. J. Chen, D. L. Cox, M. Y. Zhou, S. Thanigaraj and C. E. Gomez-Sanchez, The sheep kidney contains a novel unidirectional, high affinity NADP(+)-dependent 11 beta-hydroxysteroid dehydrogenase (11 beta-HSD-3), *Steroids* **62**, 444–450 (1997).
92. R. S. Ge, H. B. Gao, V. L. Nacharaju, G. L. Gunsalus and M. P. Hardy, Identification of a kinetically distinct activity of 11beta-hydroxysteroid dehydrogenase in rat Leydig cells, *Endocrinology* **138**, 2435–2442 (1997).
93. A. E. Michael, M. Evagelatou, D. P. Norgate, R. J. Clarke, J. W. Antoniow, B. A. Stedman, A. Brennan, R. Welsby, I. Bujalska, P. M. Stewart, B. A. Cooke, Isoforms of 11beta-hydroxysteroid dehydrogenase in human granulosa-lutein cells, *Mol. Cell. Endocrinol.* **132**, 43–52 (1997).
94. T. H. Bambino and A. J. Hsueh, Direct inhibitory effect of glucocorticoids upon testicular luteinizing hormone receptor and steroidogenesis in vivo and in vitro, *Endocrinology* **108**, 2142–2148 (1981).
95. D. C. Cumming, M. E. Quigley and S. S. Yen, Acute suppression of circulating testosterone levels by cortisol in men, *J. Clin. Endocrinol. Metab.* **57**, 671–673 (1983).
96. D. B. Hales and A. H. Payne, Glucocorticoid-mediated repression of P450_{sc} mRNA and de novo synthesis in cultured Leydig cells, *Endocrinology* **124**, 2099–2104 (1989).
97. C. Monder, Y. Miroff, A. Marandici and M. P. Hardy, 11 beta-Hydroxysteroid dehydrogenase alleviates glucocorticoid-mediated inhibition of steroidogenesis in rat Leydig cells, *Endocrinology* **134**, 1199–1204 (1994).
98. C. Monder, M. P. Hardy, R. J. Blanchard and D. C. Blanchard, Comparative aspects of 11 β -hydroxysteroid dehydrogenase. Testicular 11 β -hydroxysteroid dehydrogenase: development

- of a model for the mediation of Leydig cell function by corticosteroids, *Steroids* **59**, 69–73 (1994).
99. R. S. Ge and M. P. Hardy, Protein kinase C increases 11 β -hydroxysteroid dehydrogenase oxidation and inhibits reduction in rat Leydig cells, *J. Androl.* **23**, 135–143 (2002).
100. H. B. Gao, R. S. Ge, V. Lakshmi, A. Marandici and M. P. Hardy, Hormonal regulation of oxidative and reductive activities of 11 β -hydroxysteroid dehydrogenase in rat Leydig cells, *Endocrinology* **138**, 156–161 (1997).
101. H. Nikula, Z. Naor, M. Parvinen and I. Huhtaniemi, Distribution and activation of protein kinase C in the rat testis tissue, *Mol. Cell. Endocrinol.* **49**, 39–49 (1987).
102. T. Gudermann, C. Nichols, F. O. Levy, M. Birnbaumer and L. Birnbaumer, Ca²⁺ mobilization by the LH receptor expressed in *Xenopus* oocytes independent of 3',5'-cyclic adenosine monophosphate formation: evidence for parallel activation of two signaling pathways, *Mol. Endocrinol.* **6**, 272–278 (1992).
103. G. M. Wang, R. S. Ge, S. A. Latif, D. J. Morris and M. P. Hardy, Expression of 11 β -hydroxylase in rat Leydig cells, *Endocrinology* **143**, 621–626 (2002).
104. I. J. Bujalska, E. A. Walker, M. Hewison and P. M. Stewart, A switch in dehydrogenase to reductase activity of 11 β -hydroxysteroid dehydrogenase type 1 upon differentiation of human omental adipose stromal cells, *J. Clin. Endocrinol. Metab.* **87**, 1205–1210 (2002).
105. M. Krook, D. Ghosh, W. Duax and H. Jornvall, Three-dimensional model of NAD(+)-dependent 15-hydroxyprostaglandin dehydrogenase and relationships to the NADP(+)-dependent enzyme (carbonyl reductase), *FEBS Lett.* **322**, 139–142 (1993).
106. H. Jörnvall, B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery and D. Ghosh, Short-chain dehydrogenases/reductases (SDR), *Biochemistry* **34**, 6003–6013 (1995).
107. I. Tsigelny and M. E. Baker, Structures stabilizing the dimer interface on human 11 β -hydroxysteroid dehydrogenase types 1 and 2 and human 15-hydroxyprostaglandin dehydrogenase and their homologs, *Biochem. Biophys. Res. Commun.* **217**, 859–868 (1995).
108. I. Tsigelny and M. E. Baker, Structures important in mammalian 11 β - and 17 β -hydroxysteroid dehydrogenases, *J. Steroid Biochem. Mol. Biol.* **55**, 589–600 (1995).
109. C. Grimm, E. Maser, E. Mobus, G. Klebe, K. Reuter and R. Ficner, The crystal structure of 3 α -hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonas testosteroni* shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family, *J. Biol. Chem.* **275**, 41333–41339 (2000).
110. E. A. Walker, A. M. Clark, M. Hewison, J. P. Ride and P. M. Stewart, Functional expression, characterization, and purification of the catalytic domain of human 11 β -hydroxysteroid dehydrogenase type 1, *J. Biol. Chem.* **276**, 21343–21350 (2001).
111. S. Diederich, B. Hanke, P. Burkhardt, M. Muller, M. Schoneshofer, V. Bahr and W. Oelkers, Metabolism of synthetic corticosteroids by 11 β -hydroxysteroid-dehydrogenases in man, *Steroids* **63**, 271–277 (1998).
112. A. K. Agarwal, M. T. Tusie Luna, C. Monder and P. C. White, Expression of 11 β -hydroxysteroid dehydrogenase using recombinant vaccinia virus, *Mol. Endocrinol.* **4**, 1827–1832 (1990).
113. C. C. Moore, S. H. Mellon, J. Murai, P. K. Siiteri and W. L. Miller, Structure and function of the hepatic form of 11 β -hydroxysteroid dehydrogenase in the squirrel monkey, an animal model of glucocorticoid resistance, *Endocrinology* **133**, 368–375 (1993).
114. X. Pu and K. Yang, Guinea pig 11 β -hydroxysteroid dehydrogenase type 1: primary structure and catalytic properties, *Steroids* **65**, 148–156 (2000).
115. E. Maser and G. Bannenberg, 11 β -Hydroxysteroid dehydrogenase mediates reductive metabolism of xenobiotic carbonyl compounds, *Biochem. Pharmacol.* **47**, 1805–1812 (1994).

116. Y. Kallberg, U. Oppermann, H. Jornvall and B. Persson, Short-chain dehydrogenases/reductases (SDRs), *Eur. J. Biochem.* **269**, 4409–4417 (2002).
117. Y. Kallberg, U. Oppermann, H. Jornvall and B. Persson, Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes, *Protein Sci.* **11**, 636–641 (2002).
118. W. L. Duax, J. F. Griffin and D. Ghosh, The fascinating complexities of steroid-binding enzymes, *Curr. Opin. Struct. Biol.* **6**, 813–823 (1996).
119. U. C. T. Oppermann, C. Filling, M. Hult, N. Shafqat, X. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson and H. Jörnvall, Short-chain dehydrogenases/reductases (SDR): the 2002 update, *Chem. Biol. Interact.* **143–144**, 247–253 (2003).
120. W. L. Duax, D. Ghosh and V. Pletnev, Steroid dehydrogenase structures, mechanism of action, and disease, *Vitam. Horm.* **58**, 121–148 (2000).
121. J. Ozols, Luminal orientation and post-translational modifications of the liver microsomal 11 β -hydroxysteroid dehydrogenase, *J. Biol. Chem.* **270**, 2305–2312 (1995).
122. A. Blum, A. Raum and E. Maser, Functional Characterization of the human 11 β -hydroxysteroid dehydrogenase 1B (11 β -HSD 1B) variant, *Biochemistry* **42**, 4108–4117 (2003).
123. D. Ghosh, Z. Wawrzak, C. M. Weeks, W. L. Duax and M. Erman, The refined three-dimensional structure of 3 α ,20 β -hydroxysteroid dehydrogenase and possible roles of the residues conserved in “short chain” dehydrogenases, *Structure* **2**, 629–640 (1994).
124. K. Nakajima, A. Yamashita, H. Akama, T. Nakatsu, H. Kato, T. Hashimoto, J. Oda and Y. Yamada, Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold, *Proc. Natl. Acad. Sci. USA* **95**, 4876–4881 (1998).
125. D. Ghosh, M. Erman, W. Pangborn, W. L. Duax and M. E. Baker, Inhibition of *Streptomyces hydrogenans* 3 α ,20 β -hydroxysteroid dehydrogenase by licorice-derived compounds and crystallization of an enzyme-cofactor-inhibitor complex, *J. Steroid Biochem. Mol. Biol.* **42**, 849–853 (1992).
126. G. M. Tannin, A. K. Agarwal, C. Monder, M. I. New and P. C. White, The human gene for 11 β -hydroxysteroid dehydrogenase. Structure, tissue distribution, and chromosomal localization, *J. Biol. Chem.* **266**, 16653–16658 (1991).
127. W. Mercer, V. Obeyesekere, R. Smith and Z. Krozowski, Characterization of 11 β -HSD1B gene expression and enzymatic activity, *Mol. Cell. Endocrinol.* **92**, 247–251 (1993).
128. K. Yang, M. Yu and V. K. Han, Identification and tissue distribution of a novel variant of 11 β -hydroxysteroid dehydrogenase 1 transcript, *J. Steroid Biochem. Mol. Biol.* **55**, 247–253 (1995).
129. E. Caramelli, P. Strippoli, T. Di Giacomini, C. Tietz, P. Carinci and R. Pasquali, Lack of mutations of type 1 11 β -hydroxysteroid dehydrogenase gene in patients with abdominal obesity, *Endocr. Res.* **27**, 47–61 (2001).
130. J. Obeid, K. M. Curnow, J. Aisenberg and P. C. White, Transcripts originating in intron 1 of the HSD11 (11 β -hydroxysteroid dehydrogenase) gene encode a truncated polypeptide that is enzymatically inactive, *Mol. Endocrinol.* **7**, 154–160 (1993).
131. P. M. Stewart and Z. Krozowski, Lessons from Appenberg: 11 β -hydroxysteroid dehydrogenase 1B or 2?, *J. Endocrinol.* **141**, 191–193 (1994).
132. A. Odermatt, P. Arnold, A. Stauffer, B. M. Frey and F. J. Frey, The N-terminal anchor sequences of 11 β -hydroxysteroid dehydrogenases determine their orientation in the endoplasmic reticulum membrane, *J. Biol. Chem.* **274**, 28762–28770 (1999).
133. H. Mziaut, G. Korza, A. R. Hand, C. Gerard and J. Ozols, Targeting proteins to the lumen of endoplasmic reticulum using N-terminal domains of 11 β -hydroxysteroid dehydrogenase and the 50-kDa esterase, *J. Biol. Chem.* **274**, 14122–14129 (1999).

134. A. Blum, A. Raum, H. J. Martin and E. Maser, Human 11 β -hydroxysteroid dehydrogenase 1/carbonyl reductase: additional domains for membrane attachment?, *Chem. Biol. Interact.* **130–132**, 749–759 (2001).
135. A. Blum and E. Maser, The critical role of the N-terminus of 11beta-hydroxysteroid dehydrogenase type 1, as being encoded by exon 1, for enzyme stabilization and activity, *Chem. Biol. Interact.* **143–144**, 469–480 (2003).
136. A. K. Agarwal, T. Mune, C. Monder and P. White, Mutations in putative glycosylation sites of rat 11 β -hydroxysteroid dehydrogenase, *Biochem. Biophys. Acta* **1248**, 70–74 (1995).
137. A. Blum, H. J. Martin and E. Maser, Human 11beta-hydroxysteroid dehydrogenase type 1 is enzymatically active in its nonglycosylated form, *Biochem. Biophys. Res. Commun.* **276**, 428–434 (2000).
138. M. P. Moisan, C. R. Edwards and J. R. Seckl, Differential promoter usage by the rat 11 beta-hydroxysteroid dehydrogenase gene, *Mol. Endocrinol.* **6**, 1082–1087 (1992).
139. M. W. Voice, J. R. Seckl and K. E. Chapman, The sequence of 5' flanking DNA from the mouse 11 beta-hydroxysteroid dehydrogenase type 1 gene and analysis of putative transcription factor binding sites, *Gene* **181**, 233–235 (1996).
140. C. DeVack, B. Lupp, M. Nichols, E. Kowenz-Leutz, W. Schmid and G. Schutz, Characterization of the nuclear proteins binding the CACCC element of a glucocorticoid-responsive enhancer in the tyrosine aminotransferase gene, *Eur. J. Biochem.* **211**, 459–465 (1993).
141. F. P. Lemaigre and G. G. Rousseau, Transcriptional control of genes that regulate glycolysis and gluconeogenesis in adult liver, *Biochem. J.* **303**(Pt 1), 1–14 (1994).
142. V. De Simone, V. and R. Cortese, Transcription factors and liver-specific genes, *Biochim. Biophys. Acta* **1132**, 119–126 (1992).
143. Z. Krozowski, S. Stuchbery, P. White, C. Monder and J. W. Funder, Characterization of 11 beta-hydroxysteroid dehydrogenase gene expression: identification of multiple unique forms of messenger ribonucleic acid in the rat kidney, *Endocrinology* **127**, 3009–3013 (1990).
144. L. J. Williams, V. Lyons, I. MacLeod, V. Rajan, G. J. Darlington, V. Poli, J. R. Seckl and K. E. Chapman, C/EBP regulates hepatic transcription of 11beta-hydroxysteroid dehydrogenase type 1. A novel mechanism for cross-talk between the C/EBP and glucocorticoid signaling pathways, *J. Biol. Chem.*, 30232–30239 (2000).
145. J. Lekstrom-Himes and K. G. Xanthopoulos, Biological role of the CCAAT/enhancer-binding protein family of transcription factors, *J. Biol. Chem.* **273**, 28545–28548 (1998).
146. S. L. McKnight, M. D. Lane and S. Gluecksohn-Waelsch, Is CCAAT/enhancer-binding protein a central regulator of energy metabolism?, *Genes Dev.* **3**, 2021–2024 (1989).
147. N. D. Wang, M. J. Finegold, A. Bradley, C. N. Ou, S. V. Abdelsayed, M. D. Wilde, L. R. Taylor, D. R. Wilson and G. J. Darlington, Impaired energy homeostasis in C/EBP alpha knockout mice, *Science* **269**, 1108–1112 (1995).
148. F. Matsumo, S. Chowdhury, T. Gotoh, K. Iwase, H. Matsuzaki, K. Takatsuki, M. Mori and M. Takiguchi, Induction of the C/EBP beta gene by dexamethasone and glucagon in primary-cultured rat hepatocytes, *J. Biochem. (Tokyo)* **119**, 524–532 (1996).
149. T. Gotoh, S. Chowdhury, M. Takiguchi and M. Mori, The glucocorticoid-responsive gene cascade. Activation of the rat arginase gene through induction of C/EBPbeta, *J. Biol. Chem.* **272**, 3694–3698 (1997).
150. D. R. Breed, L. R. Margraf, J. L. Alcorn and C. R. Mendelson, Transcription factor C/EBPdelta in fetal lung: developmental regulation and effects of cyclic adenosine 3',5'-monophosphate and glucocorticoids, *Endocrinology* **138**, 5527–5534 (1997).
151. O. A. MacDougald, P. Cornelius, F. T. Lin, S. S. Chen and M. D. Lane, Glucocorticoids reciprocally regulate expression of the CCAAT/enhancer-binding protein alpha and delta

- genes in 3T3-L1 adipocytes and white adipose tissue, *J. Biol. Chem.* **269**, 19041–19047 (1994).
152. R. A. Ramos, Y. Nishio, A. C. Maiyar, K. E. Simon, C. C. Ridder, Y. Ge and G. L. Firestone, Glucocorticoid-stimulated CCAAT/enhancer-binding protein alpha expression is required for steroid-induced G1 cell cycle arrest of minimal-deviation rat hepatoma cells, *Mol. Cell. Biol.* **16**, 5288–5301 (1996).
153. M. W. Voice, J. R. Seckl, C. R. Edwards and K. E. Chapman, 11 beta-hydroxysteroid dehydrogenase type 1 expression in 2S FAZA hepatoma cells is hormonally regulated: a model system for the study of hepatic glucocorticoid metabolism, *Biochem. J.* **317**, 621–625 (1996).
154. A. Napolitano, M. W. Voice, C. R. Edwards, J. R. Seckl and K. E. Chapman, 11Beta-hydroxysteroid dehydrogenase 1 in adipocytes: expression is differentiation-dependent and hormonally regulated, *J. Steroid Biochem. Mol. Biol.* **64**, 251–260 (1998).
155. P. M. Jamieson, E. Fuchs, G. Flugge and J. R. Seckl, Attenuation of hippocampal 11beta-hydroxysteroid dehydrogenase type 1 by chronic psychosocial stress in the tree shrew, *Stress* **2**, 123–132 (1997).
156. P. M. Jamieson, K. E. Chapman and J. R. Seckl, Tissue- and temporal-specific regulation of 11beta-hydroxysteroid dehydrogenase type 1 by glucocorticoids in vivo, *J. Steroid Biochem. Mol. Biol.* **68**, 245–250 (1999).
157. J. M. Lehmann, S. A. Kliewer, L. B. Moore, T. A. Smith-Oliver, B. B. Oliver, J. L. Su, S. S. Sundseth, D. A. Winegar, D. E. Blanchard, T. A. Spencer, T. M. Willson, Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway, *J. Biol. Chem.* **272**, 3137–3140 (1997).
158. B. A. Janowski, P. J. Willy, T. R. Devi, J. R. Falck and D. J. Mangelsdorf, An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha, *Nature* **383**, 728–731 (1996).
159. T. M. Stulnig, U. Oppermann, K. R. Steffensen, G. U. Schuster and J. A. Gustafsson, Liver X receptors downregulate 11beta-hydroxysteroid dehydrogenase type 1 expression and activity, *Diabetes* **51**, 2426–2433 (2002).
160. G. U. Schuster, P. Parini, L. Wang, S. Alberti, K. R. Steffensen, G. K. Hansson, B. Angelin and J. A. Gustafsson, Accumulation of foam cells in liver X receptor-deficient mice, *Circulation* **106**, 1147–1153 (2002).
161. J. Berger, M. Tanen, A. Elbrecht, A. Hermanowski-Vosatka, D. E. Moller, S. D. Wright and R. Thieringer, Peroxisome proliferator-activated receptor-gamma ligands inhibit adipocyte 11beta-hydroxysteroid dehydrogenase type 1 expression and activity, *J. Biol. Chem.* **276**, 12629–12635 (2001).
162. J. J. Repa and D. J. Mangelsdorf, Nuclear receptor regulation of cholesterol and bile acid metabolism, *Curr. Opin. Biotechnol.* **10**, 557–563 (1999).
163. G. Chinetti, S. Lestavel, V. Bocher, A. T. Remaley, B. Neve, I. P. Torra, E. Teissier, A. Minnich, M. Jaye, N. Duverger, H. B. Brewer, J. C. Fruchart, V. Clavey, B. Staels, PPAR-alpha and PPAR-gamma activators induce cholesterol removal from human macrophage foam cells through stimulation of the ABCA1 pathway, *Nat. Med.* **7**, 53–58 (2001).
164. K. A. Tobin, H. H. Steineger, S. Alberti, O. Spydevold, J. Auwerx, J. A. Gustafsson and H. I. Nebb, Cross-talk between fatty acid and cholesterol metabolism mediated by liver X receptor-alpha, *Mol. Endocrinol.* **14**, 741–752 (2000).
165. A. Chawla, W. A. Boisvert, C. H. Lee, B. A. Laffitte, Y. Barak, S. B. Joseph, D. Liao, L. Nagy, P. A. Edwards, L. K. Curtiss, R. M. Evans, P. Tontonoz, A PPAR gamma-LXR-ABCA1 pathway in macrophages is involved in cholesterol efflux and atherogenesis, *Mol. Cell* **7**, 161–171 (2001).

166. Y. Barak, M. C. Nelson, E. S. Ong, Y. Z. Jones, P. Ruiz-Lozano, K. R. Chien, A. Koder and R. M. Evans, PPAR gamma is required for placental, cardiac, and adipose tissue development, *Mol. Cell* **4**, 585–595 (1999).
167. E. D. Rosen, P. Sarraf, A. E. Troy, G. Bradwin, K. Moore, D. S. Milstone, B. M. Spiegelman and R. M. Mortensen, PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro, *Mol. Cell*, 611–617 (1999).
168. A. K. El Jack, J. K. Hamm, P. F. Pilch and S. R. Farmer, Reconstitution of insulin-sensitive glucose transport in fibroblasts requires expression of both PPARgamma and C/EBPalpha, *J. Biol. Chem.* **274**, 7946–7951 (1999).
169. Z. Wu, E. D. Rosen, R. Brun, S. Hauser, G. Adelmant, A. E. Troy, C. McKeon, G. J. Darlington and B. M. Spiegelman, Cross-regulation of C/EBP alpha and PPAR gamma controls the transcriptional pathway of adipogenesis and insulin sensitivity, *Mol. Cell* **3**, 151–158 (1999).
170. N. Kubota, Y. Terauchi, H. Miki, H. Tamemoto, T. Yamauchi, K. Komeda, S. Satoh, R. Nakano, C. Ishii, T. Sugiyama, K. Eto, Y. Tsubamoto, A. Okuno, K. Murakami, H. Sekihara, G. Hasegawa, M. Naito, Y. Toyoshima, S. Tanaka, K. Shiota, T. Kitamura, T. Fujita, O. Ezaki, S. Aizawa, R. Nagai, K. Tobe, S. Kimura and T. Kadowaki, PPAR gamma mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance, *Mol. Cell* **4**, 597–609 (1999).
171. E. D. Rosen, C. J. Walkey, P. Puigserver and B. M. Spiegelman, Transcriptional regulation of adipogenesis, *Genes Dev.* **14**, 1293–1307 (2000).
172. T. M. Willson, M. H. Lambert and S. A. Kliewer, Peroxisome proliferator-activated receptor gamma and metabolic disease, *Annu. Rev. Biochem.* **70**, 341–367 (2001).
173. S. C. Low, S. N. Assaad, V. Rajan, K. E. Chapman, C. R. Edwards and J. R. Seckl, Regulation of 11 beta-hydroxysteroid dehydrogenase by sex steroids in vivo: further evidence for the existence of a second dehydrogenase in rat kidney, *J. Endocrinol.* **139**, 27–35 (1993).
174. S. C. Low, K. E. Chapman, C. R. Edwards, T. Wells, I. C. Robinson and J. R. Seckl, Sexual dimorphism of hepatic 11 β -hydroxysteroid dehydrogenase in the rat: the role of growth hormone patterns, *J. Endocrinol.* **143**, 541–548 (1994).
175. K. Yang, C. L. Smith, D. Dales, G. L. Hammond and J. R. Challis, Cloning of an ovine 11 beta-hydroxysteroid dehydrogenase complementary deoxyribonucleic acid: tissue and temporal distribution of its messenger ribonucleic acid during fetal and neonatal development, *Endocrinology* **131**, 2120–2126 (1992).
176. M. P. Moisan, C. R. Edwards and J. R. Seckl, Ontogeny of 11 beta-hydroxysteroid dehydrogenase in rat brain and kidney, *Endocrinology* **130**, 400–404 (1992).
177. S. C. Low, M. P. Moisan, J. M. Noble, C. R. Edwards and J. R. Seckl, Glucocorticoids regulate hippocampal 11 beta-hydroxysteroid dehydrogenase activity and gene expression in vivo in the rat, *J. Neuroendocrinol.* **6**, 285–290 (1994).
178. B. R. Walker, B. C. Williams and C. R. Edwards, Regulation of 11 beta-hydroxysteroid dehydrogenase activity by the hypothalamic-pituitary-adrenal axis in the rat, *J. Endocrinol.* **141**, 467–472 (1994).
179. M. M. Hammami and P. K. Siiteri, Regulation of 11 beta-hydroxysteroid dehydrogenase activity in human skin fibroblasts: enzymatic modulation of glucocorticoid action, *J. Clin. Endocrinol. Metab.* **73**, 326–334 (1991).
180. Y. Takeda, I. Miyamori, T. Yoneda, Y. Ito and R. Takeda, Expression of 11 beta-hydroxysteroid dehydrogenase mRNA in rat vascular smooth muscle cells, *Life Sci.* **54**, 281–285 (1994).
181. C. B. Whorwood, S. J. Donovan, P. J. Wood and D. I. Phillips, Regulation of glucocorticoid receptor alpha and beta isoforms and type I 11beta-hydroxysteroid dehydrogenase

- expression in human skeletal muscle cells: a key role in the pathogenesis of insulin resistance?, *J. Clin. Endocrinol. Metab.* **86**, 2296–2308 (2001).
182. C. B. Whorwood, S. J. Donovan, D. Flanagan, D. I. Phillips and C. D. Byrne, Increased glucocorticoid receptor expression in human skeletal muscle cells may contribute to the pathogenesis of the metabolic syndrome, *Diabetes* **51**, 1066–1075 (2002).
 183. R. H. Oakley and J. A. Cidlowski, Homologous down regulation of the glucocorticoid receptor: the molecular machinery, *Crit. Rev. Eukaryot. Gene Expr.* **3**, 63–88 (1993).
 184. K. L. Burnstein, C. M. Jewell, M. Sar and J. A. Cidlowski, Intragenic sequences of the human glucocorticoid receptor complementary DNA mediate hormone-inducible receptor messenger RNA down-regulation through multiple mechanisms, *Mol. Endocrinol.* **8**, 1764–1773 (1994).
 185. D. N. Brindley, Role of glucocorticoids and fatty acids in the impairment of lipid metabolism observed in the metabolic syndrome, *Int. J. Obes. Relat. Metab. Disord.* **19 Suppl. 1**, S69–S75 (1995).
 186. B. B. van der and P. T. van der Saag, Nuclear factor-kappa-B/steroid hormone receptor interactions as a functional basis of anti-inflammatory action of steroids in reproductive organs, *Mol. Hum. Reprod.* **2**, 433–438 (1996).
 187. L. I. McKay and J. A. Cidlowski, Molecular control of immune/inflammatory responses: interactions between nuclear factor-kappa B and steroid receptor-signaling pathways, *Endocr. Rev.* **20**, 435–459 (1999).
 188. G. Escher, I. Galli, B. S. Vishwanath, B. M. Frey and F. J. Frey, Tumor necrosis factor alpha and interleukin 1beta enhance the cortisone/cortisol shuttle, *J. Exp. Med.* **186**, 189–198 (1997).
 189. M. Tetsuka, L. C. Haines, M. Milne, G. E. Simpson and S. G. Hillier, Regulation of 11beta-hydroxysteroid dehydrogenase type 1 gene expression by LH and interleukin-1beta in cultured rat granulosa cells, *J. Endocrinol.* **163**, 417–423 (1999).
 190. P. Y. Yong, C. Harlow, K. J. Thong and S. G. Hillier, Regulation of 11beta-hydroxysteroid dehydrogenase type 1 gene expression in human ovarian surface epithelial cells by interleukin-1, *Hum. Reprod.* **17**, 2300–2306 (2002).
 191. K. Handoko, K. Yang, B. Strutt, W. Khalil and D. Killinger, Insulin attenuates the stimulatory effects of tumor necrosis factor alpha on 11beta-hydroxysteroid dehydrogenase 1 in human adipose stromal cells, *J. Steroid Biochem. Mol. Biol.* **72**, 163–168 (2000).
 192. J. W. Tomlinson, J. Moore, M. S. Cooper, I. Bujalska, M. Shahmanesh, C. Burt, A. Strain, M. Hewison and P. M. Stewart, Regulation of expression of 11beta-hydroxysteroid dehydrogenase type 1 in adipose tissue: tissue-specific induction by cytokines, *Endocrinology* **142**, 1982–1989 (2001).
 193. E. C. Ledgerwood, J. S. Pober and J. R. Bradley, Recent advances in the molecular basis of TNF signal transduction, *Lab. Invest.* **79**, 1041–1050 (1999).
 194. E. A. Walker, A. Li, M. Hewison and P. M. Stewart, Characterization of the proximal promoter region of the human type 1 11beta-hydroxysteroid dehydrogenase gene. 81st Annual Meeting of The Endocrine Society, San Diego, CA, Abstract P-555, 1999.
 195. J. S. Moore, J. P. Monson, G. Kaltsas, P. Putignano, P. J. Wood, M. C. Sheppard, G. M. Besser, N. F. Taylor and P. M. Stewart, Modulation of 11beta-hydroxysteroid dehydrogenase isozymes by growth hormone and insulin-like growth factor: in vivo and in vitro studies, *J. Clin. Endocrinol. Metab.* **84**, 4172–4177 (1999).
 196. T. Petley, K. Graff, W. Jiang, H. Yang and J. Florini, Variation among cell types in the signaling pathways by which IGF-I stimulates specific cellular responses, *Horm. Metab. Res.* **31**, 70–76 (1999).
 197. S. V. Gelding, N. F. Taylor, P. J. Wood, K. Noonan, J. U. Weaver, D. F. Wood and J. P. Monson, The effect of growth hormone replacement therapy on cortisol-cortisone

- interconversion in hypopituitary adults: evidence for growth hormone modulation of extrarenal 11 beta-hydroxysteroid dehydrogenase activity, *Clin. Endocrinol. (Oxf)* **48**, 153–162 (1998).
198. S. S. Dunklemann, B. Fairhurst, J. Plager and C. Waterhouse, Cortisol metabolism in obesity, *J. Clin. Endocrinol. Metab.* **24**, 832–841 (1964).
199. C. J. Migeon, O. C. Green and J. P. Eckert, Study of adrenocortical function in obesity, *Metabolism* **12**, 718–739 (1963).
200. P. M. Stewart, A. Boulton, S. Kumar, P. M. Clark and C. H. Shackleton, Cortisol metabolism in human obesity: impaired cortisone \rightarrow cortisol conversion in subjects with central adiposity, *J. Clin. Endocrinol. Metab.* **84**, 1022–1027 (1999).
201. P. J. Trainer, W. M. Drake, L. A. Perry, N. F. Taylor, G. M. Besser and J. P. Monson, Modulation of cortisol metabolism by the growth hormone receptor antagonist pegvisomant in patients with acromegaly, *J. Clin. Endocrinol. Metab.* **86**, 2989–2992 (2001).
202. P. Bjorntorp, G. Holm and R. Rosmond, Hypothalamic arousal, insulin resistance and Type 2 diabetes mellitus, *Diabet. Med.* **16**, 373–383 (1999).
203. G. W. Strain, B. Zumoff, J. J. Strain, J. Levin and D. K. Fukushima, Cortisol production in obesity, *Metabolism* **29**, 980–985 (1980).
204. G. W. Strain, B. Zumoff, J. Kream, J. J. Strain, J. Levin and D. Fukushima, Sex difference in the influence of obesity on the 24 hr mean plasma concentration of cortisol, *Metabolism* **31**, 209–212 (1982).
205. R. Andrew, D. I. Phillips and B. R. Walker, Obesity and gender influence cortisol secretion and metabolism in man, *J. Clin. Endocrinol. Metab.* **83**, 1806–1809 (1998).
206. C. T. Montague, I. S. Farooqi, J. P. Whitehead, M. A. Soos, H. Rau, N. J. Wareham, C. P. Sewter, J. E. Digby, S. N. Mohammed, J. A. Hurst, C. H. Cheetham, A. R. Earley, A. H. Barnett, J. B. Prins, S. O'Rahilly, Congenital leptin deficiency is associated with severe early-onset obesity in humans, *Nature* **387**, 903–908 (1997).
207. O. Paulmyer-Lacroix, S. Boullu, C. Oliver, M. C. Alessi and M. Grino, Expression of the mRNA coding for 11beta-hydroxysteroid dehydrogenase type 1 in adipose tissue from obese patients: an in situ hybridization study, *J. Clin. Endocrinol. Metab.* **87**, 2701–2705 (2002).
208. E. Rask, B. R. Walker, S. Soderberg, D. E. Livingstone, M. Eliasson, O. Johnson, R. Andrew and T. Olsson, Tissue-specific changes in peripheral cortisol metabolism in obese women: increased adipose 11beta-hydroxysteroid dehydrogenase type 1 activity, *J. Clin. Endocrinol. Metab.* **87**, 3330–3336 (2002).
209. J. R. Katz, V. Mohamed-Ali, P. J. Wood, J. S. Yudkin and S. W. Coppack, An in vivo study of the cortisol-cortisone shuttle in subcutaneous abdominal adipose tissue, *Clin. Endocrinol. (Oxf)* **50**, 63–68 (1999).
210. B. A. Jensen and G. J. Stevens, Expression profiling during adipocyte differentiation of 3T3-L1 fibroblasts, *Gene* **299**, 95–100 (2002).
211. J. W. Tomlinson, N. Draper, J. Mackie, A. P. Johnson, G. Holder, P. Wood and P. M. Stewart, Absence of Cushingoid phenotype in a patient with Cushing's disease due to defective cortisone to cortisol conversion, *J. Clin. Endocrinol. Metab.* **87**, 57–62 (2002).
212. N. M. Morton, M. C. Holmes, C. Fievet, B. Staels, A. Tailleux, J. J. Mullins and J. R. Seckl, Improved lipid and lipoprotein profile, hepatic insulin sensitivity, and glucose tolerance in 11beta-hydroxysteroid dehydrogenase type 1 null mice, *J. Biol. Chem.* **276**, 41293–41300 (2001).
213. J. W. Tomlinson, B. Sinha, I. Bujalska, M. Hewison and P. M. Stewart, Expression of 11beta-hydroxysteroid dehydrogenase type 1 in adipose tissue is not increased in human obesity, *J. Clin. Endocrinol. Metab.* **87**, 5630–5635 (2002).
214. D. E. Livingstone, G. C. Jones, K. Smith, P. M. Jamieson, R. Andrew, C. J. Kenyon and B. R. Walker, Understanding the role of glucocorticoids in obesity: tissue-specific

- alterations of corticosterone metabolism in obese Zucker rats, *Endocrinology* **141**, 560–563 (2000).
215. B. R. Walker, A. A. Connacher, R. M. Lindsay, D. J. Webb and C. R. Edwards, Carbenoxolone increases hepatic insulin sensitivity in man: a novel role for 11-oxosteroid reductase in enhancing glucocorticoid receptor activation, *J. Clin. Endocrinol. Metab.* **80**, 3155–3159 (1995).
216. S. Diederich, C. Grossmann, B. Hanke, M. Quinkler, M. Herrmann, V. Bahr and W. Oelkers, In the search for specific inhibitors of human 11 β -hydroxysteroid-dehydrogenases (11 β -HSDs): chenodeoxycholic acid selectively inhibits 11 β -HSD-I, *Eur. J. Endocrinol.* **142**, 200–207 (2000).
217. A. G. Turpie and T. J. Thomson, Carbenoxolone sodium in the treatment of gastric ulcer with special reference to side-effects, *Gut* **6**, 591–594 (1965).
218. D. Armanini, I. Karbowiak, Z. Krozowski, J. W. Funder and W. R. Adam, The mechanism of mineralocorticoid action of carbenoxolone, *Endocrinology* **111**, 1683–1686 (1982).
219. P. M. Stewart, R. Valentino, A. M. Wallace, D. Burt, C. H. L. Shackleton and C. R. W. Edwards, Mineralocorticoid activity of liquorice: 11 β -hydroxysteroid dehydrogenase deficiency comes of age, *Lancet* **ii**, 821–824 (1987).
220. E. Maser, Stress, hormonal changes, alcohol, food constituents and drugs: factors that advance the incidence of tobacco smoke-related cancer?, *Trends Pharmacol. Sci.* **18**, 270–275 (1997).
221. T. Barf, J. Vallgarda, R. Emond, C. Haggstrom, G. Kurz, A. Nygren, V. Larwood, E. Mosialou, K. Axelsson, R. Olsson, L. Engblom, N. Edling, Y. Ronquist-Nii, B. Ohman, P. Alberts, L. Abrahmsen, Arylsulfonamidothiazoles as a new class of potential antidiabetic drugs. Discovery of potent and selective inhibitors of the 11 β -hydroxysteroid dehydrogenase type 1, *J. Med. Chem.* **45**, 3813–3815 (2002).
222. P. Alberts, L. Engblom, N. Edling, M. Forsgren, G. Klingstrom, C. Larsson, Y. Ronquist-Nii, B. Ohman and L. Abrahmsen, Selective inhibition of 11 β -hydroxysteroid dehydrogenase type 1 decreases blood glucose concentrations in hyperglycaemic mice, *Diabetologia* **45**, 1528–1532 (2002).
223. S. Diederich, E. Eigendorff, P. Burkhardt, M. Quinkler, C. Bumke-Vogt, M. Rochel, D. Seidelmann, P. Esperling, W. Oelkers, V. Bahr, 11 β -hydroxysteroid dehydrogenase types 1 and 2: an important pharmacokinetic determinant for the activity of synthetic mineralo- and glucocorticoids, *J. Clin. Endocrinol. Metab.* **87**, 5695–5701 (2002).
224. S. Diederich, B. Hanke, W. Oelkers and V. Bähr, Metabolism of dexamethasone in the human kidney: nicotinamide adenine dinucleotide-dependent 11 β -reduction, *J. Clin. Endocrinol. Metab.* **82**, 1598–1602 (1997).
225. K. X. Z. Li, V. R. Obeyesekere, Z. S. Krozowski and P. Ferrari, Oxoreductase and dehydrogenase activities of the human and rat 11 β -hydroxysteroid dehydrogenase type 2 enzyme, *Endocrinology* **138**, 2948–2952 (1997).

Dancing with Complement C4 and the *RP-C4-CYP21-TNX* (RCCX) Modules of the Major Histocompatibility Complex

C. YUNG YU,^{*,†,‡} ERWIN
K. CHUNG,^{*,†} YAN YANG,^{*,†}
CAROL A. BLANCHONG,^{*}
NATALIE JACOBSEN,^{*}
KAPIL SAXENA,^{*} ZHENYU YANG,^{*,‡}
WEBB MILLER,[§] LILIAN VARGA,^{||}
AND GEORGE FUST^{||}

^{*}*Center for Molecular and Human
Genetics, Columbus Children's Research
Institute, 700 Children's Drive,
Columbus, OH 43205-2696, USA*

[†]*Department of Molecular Virology,
Immunology and Medical Genetics,
The Ohio State University, Columbus,
OH, USA*

[‡]*Department of Pediatrics; Molecular,
Cellular and Developmental Biology
Graduate Program, The Ohio State
University, Columbus, OH, USA*

[§]*Department of Computer Science and
Engineering, University Park,
Pennsylvania State University, PA 16802,
USA; and*

^{||}*Third Department of Internal Medicine,
Semmelweis Medical University,
Budapest, Hungary*

I. Introduction	219
II. The Sophisticated Genetic Diversities of Human C4 and RCCX Modules	221
III. Structural Diversities of the Human and Mouse C4 Proteins	232
IV. Nucleotide Polymorphisms in Human <i>C4A</i> and <i>C4B</i> Genes	245
V. Expression of <i>C4A</i> and <i>C4B</i> Transcripts and Proteins	255
VI. The Endogenous Retrovirus that Mediates the Dichotomous Size Variation of C4 Genes	260
VII. Evolution of Complement C3/C4/C5 and RCCX Modules in the MHC . .	267
VIII. Afterthoughts: Eleven Outstanding Issues to be Addressed	274
Acknowledgments	276
References	276

The number of the complement component *C4* genes varies from 2 to 8 in a diploid genome among different human individuals. Three quarters of the *C4* genes in Caucasian populations have the endogenous retrovirus, HERV-K(C4), in the ninth intron. The remainder does not. The *C4* serum proteins are highly polymorphic and their concentrations vary from 100 to $\sim 1000 \mu\text{g/ml}$. There are two distinct classes of *C4* protein, *C4A* and *C4B*, which have diversified to fulfill (a) the opsonization/immunoclearance purposes and (b) the well-known complement function in the killing of microbes by lysis and neutralization, respectively. Many infectious and autoimmune diseases are associated with complete or partial deficiency of *C4A* and/or *C4B*. The adverse effects of high *C4* gene dosages, however, are just emerging, as the concepts of human *C4* genetics are revised and accurate techniques are applied to distinguish partial deficiencies from differential expression caused by unequal *C4A* and *C4B* gene dosages and gene sizes. This review attempts to dissect the sophisticated genetics of complement *C4A* and *C4B*. The emphases are on the qualitative and quantitative diversities of *C4* genotypes and phenotypes. The many allotypic variants and the processed products of human and mouse *C4* proteins are described. The modular variation of *C4* genes together with the serine/threonine nuclear kinase gene *RP*, the steroid 21-hydroxylase *CYP21*, and extracellular matrix protein *TNX* (RCCX modules) are investigated for the effects on homogenization of *C4* protein polymorphisms, and on the unequal genetic crossovers that knocked out the functions of *CYP21* and/or *TNX*. Furthermore, the influence of the endogenous retrovirus HERV-K(C4) on *C4* gene expression and the dispersal of HERV-K(C4) family members in the human genome are discussed. © 2003 Elsevier Science

Keywords: Complement *C4*; Diversity; Endogenous retrovirus; Evolution; Gene dosage; Polygenic variation; Gene size; Partial gene duplication; Polymorphism; Recombination; Steroid 21-hydroxylase; Tenascin-X.

Abbreviations

CAH	congenital adrenal hyperplasia
Ch	Chido blood-group antigen
HERV-K(C4)	human endogenous retrovirus firstly found in the intron 9 of the long <i>C4</i> gene
MHC	major histocompatibility complex (also known as the HLA in humans)
NTR	netrin repeat
RCCX	serine/threonine nuclear protein kinase RP, complement component <i>C4</i> , steroid 21-hydroxylase <i>CYP21</i> , extracellular matrix protein tenascin <i>TNX</i>
Rg	Rodgers blood-group antigen
SLE	systemic lupus erythematosus
Slp	mouse sex-limited protein

I. Introduction

C4, the fourth component of the complement system, was discovered by Gordon *et al.* in 1926. The hemolytic activity of this protein in serum was shown to withstand heating up to 56 degrees, was not affected by mixing with yeast cells, but could be inactivated by treatment with ammonia or hydrazine (1,2).

Together with many other proteins involved in adaptive immunity, complement component C4 emerged in the vertebrates as part of a *big bang* (3,4). From the nurse shark to amphibians to birds or to mice and to humans, the complement C4 gene is linked to, or located in, the major histocompatibility complex (MHC) (5–10). Among different individuals in the human populations, the number and size and sequences of the C4A and C4B genes vary, so do the serum concentrations of C4A and C4B proteins (11,12).

The activated C4 is a subunit of the C3 and C5 convertases, the enzymatic complexes that activate C3 and C5 of the classical and lectin complement-activation pathways. Products of these convertases include the activated components C3b and C5b, and the anaphylatoxins C3a and C5a. The association of C3b to C4b2a alters the C3 convertase specificity to C5 convertase. The generation of C5b initiates the assembly of the membrane-attack complex (13–17). In other words, complement component C4 is positioned at the pivotal point by which the activation of the classical pathway and the lectin pathway is accomplished. Downstream of C4 activation includes the production of C3b and its split products that facilitate the opsonization and immunoclearance processes. In a process that probably intertwines with C3, C4 is involved in the communication with other branches of the immune system to achieve immune tolerance and to potentiate the humoral immune response (18,19). However, deficiency or excessive amounts of C4 could adversely affect an individual's immune system (20). C4 deficiency may lead to defective processing of immune complexes, impairment of B-cell memory, and/or persistence of bacterial and viral infections (19,21,22). Excessive production of complement C4 could potentially cause overactivation of the complement pathways and exacerbate the inflammatory response at the local tissues.

Human complement C4 exists as two isotypes, C4A and C4B. Although sharing >99% sequence identity, they have different hemolytic activities, serological activities, and covalent-binding affinities to antigens and immune complexes. After activation, the metastable thioester in C4B tends to form a covalent ester bond with hydroxyl group containing substrates very rapidly and efficiently. Therefore, C4B is important in the propagation of the activation pathways that lead to the formation of the membrane-attack complex against the foreign antigens in a precise and well-regulated fashion (23–31). On the other hand, the activated product of C4A forms a covalent amide bond with

amino group containing antigens effectively. The reaction rate of activated C4A was estimated to be >20-fold slower than that of C4B (32–34). These differences infer that C4A may be advantageous for ensuring the solubilization of antibody–antigen aggregates (or inhibition of immune precipitation) through binding to IgG or to antigens of immune complexes (IC), and the clearance of IC through binding to complement receptor CR1.

The difference in C4A- versus C4B-binding affinities lies within the C4d region of the α -chain. This region also determines the Rodgers (Rg) versus Chido (Ch) antigenic determinants (35). The C4A isotype demonstrates an association with Rg blood-group antigens. The C4B isotype manifests an association with Ch blood-group antigens (36–38). Based on electrophoretic mobilities due to gross electric charge differences, >41 allotypes for C4A and C4B proteins have been documented (39). The nucleotide substitutions that contribute to the C4A/C4B isotypes and the major Rg/Ch antigenic determinants have been defined (40).

The *C4* gene is located within a genetic module of four consecutive genes known as the RCCX, which encodes for the serine/threonine nuclear protein kinase RP (also known as STK19), complement component C4, steroid 21-hydroxylase CYP21, and extracellular matrix protein tenascin TNX (41,42). These four genes are always duplicated together as a module in the MHC (9,43–47). When duplicated into a bimodular structure, an additional functional gene is generated for complement component C4, but the other three concurrently duplicated constituents located between the *C4* genes are usually nonfunctional. The integration of the endogenous retrovirus into intron 9 of the long genes created an additional level of *C4* genetic diversity: a dichotomous gene size variation of *C4A* and *C4B* (40,44,48–51). The remarkably frequent variation in gene dosage and gene size, diversities in the protein primary structures, expression levels, and functions of human C4A and C4B are probably some of the mechanisms developed by the immune system to overcome a vast variety of microbial infections. As a double-edged sword, however, such inborn genetic diversity has its price, as shown by the associations of many diseases with C4 deficiency, and recombinations of the constituent genes in the RCCX modules. While examining the susceptibility of an individual to microbial infections (52) or autoimmune disease, or managing an organ transplant recipient (53), or considering immunotherapies such as the application of complement inhibitors (54,55), knowledge on the status of *C4A* and *C4B* gene dosage, gene size, and protein polymorphisms may be relevant.

Remarkable parallels in structural diversities and protein polymorphisms are present in the mouse C4 and Slp. The structural and possibly functional divergence of mouse Slp from C4 offers an informative and natural system for further in-depth physiological studies of C4.

The complex, qualitative and quantitative diversities of *C4* genes and *C4* proteins have been a well-guarded secret in immunology. It is the purpose of this article to unveil the many “faces” of complement *C4* in humans and in mice.

II. The Sophisticated Genetic Diversities of Human *C4* and *RCCX* Modules

A. The Genetic Study of Human *C4*: A Historic Background

Since 1969, when Rosenfeld *et al.*, using antibody–antigen crossed electrophoresis, illustrated complement *C4* polymorphisms in hemolytic activity and electrophoretic mobility (56), the progress toward understanding the sophisticated genetics of human *C4* has taken a somewhat meandering course. It appears that there are multiple allotypes of *C4* protein (39). These protein allotypes may be grouped into two classes: the acidic *C4A* with lower hemolytic activities and the basic *C4B* with higher hemolytic activities (31). There is a wide spectrum of *C4A* and *C4B* protein levels in the population: from complete absence of *C4A* or *C4B*, more *C4A* than *C4B*, more *C4B* than *C4A*, to equal quantities of *C4A* and *C4B* (57,58). A single locus with codominant expression of the fast- and the slow-migrating forms of *C4* was first proposed (59) but was soon replaced by a two-locus model with one *C4A* and one *C4B* gene (60). The differential levels of *C4A* and *C4B* proteins in the plasma were explained by the presence of null or silent alleles in either the *C4A* or the *C4B* locus. The model was widely supported by results from independent laboratories (61,62). However, there were abundant exceptions to a rigid *C4A–C4B*, two-locus model of *C4* genes in the MHC class III region of chromosome 6 (9,59). Family-segregation studies showed considerable frequencies of single-locus haplotypes coding for *C4A* or *C4B* proteins, and haplotypes with three loci coding for one *C4A* and two different *C4B* allotypes (63,64). There was also mounting evidence for the homoexpression of *C4A* proteins from two consecutive *C4* loci (e.g., *C4A3*, *C4A2*) (65–67). The phenomenon was observed when each *C4* gene coded for a polymorphic variant distinguishable in an allotyping gel (31,68) (Table I).

Molecular cloning and sequencing of various *C4A* and *C4B* genes from individuals with defined phenotypes allowed identification of amino acid residues that contributed to the differential chemical reactivities of the activated *C4A* and *C4B*, and also their associated Rg and Ch blood-group antigens (9,34,37,69,70). It enabled the creation of molecular biologic techniques to determine the presence and number of *C4A* and *C4B* genes in the MHC (71,72), and led to the elucidation of several dichotomic features of

TABLE I
A COMPARISON OF THE BIOCHEMICAL PROPERTIES BETWEEN HUMAN C4A AND C4B PROTEINS

	C4A	C4B
Electrophoretic mobility		
Agarose gel	Fast	Slow
SDS-PAGE (α -chain)	M_r 96,000	M_r 94,000
Thioester reactivity		
Hemolytic activity*	Lower	Higher
Relative covalent-binding affinities		
Amino group	Higher	Lower
Hydroxyl group	Lower	Higher
Half-life of C4b	>10 s	<1 s
Binding to complement receptor CR1*	Higher	Lower
Inhibition of immunoprecipitation*	Higher	Lower
Association with blood-group antigens ⁺	Rg: 1, 2	Ch: 1, 2, 3, 4, 5, 6

*Demonstrated when purified proteins were used in *in vitro* assays or when autologous-regulatory proteins were absent.
⁺With exceptions such as C4A1 and C4B5.

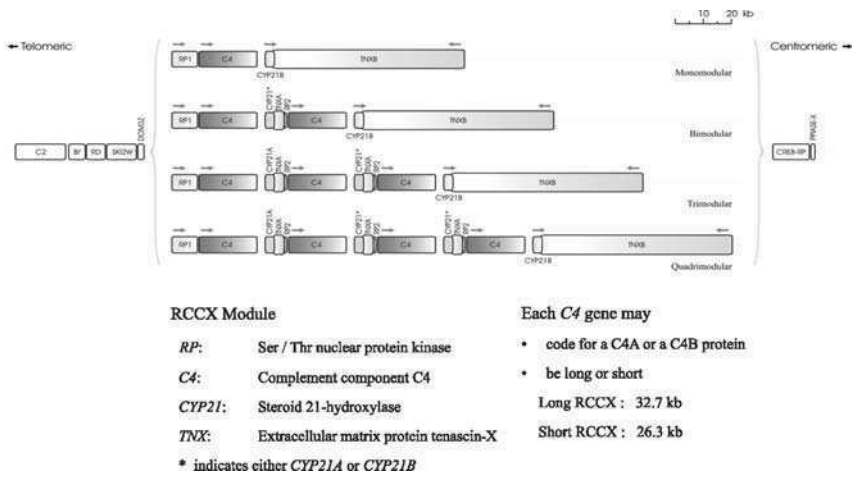


FIG. 1. The 1-2-3-4 loci concept of human *C4* and RCCX modular variations. This map of the human MCGC region shows the physical length variants among the monomodular, bimodular, trimodular, and quadrimodular RCCX structures. The MHC class I genes are located at the telomeric end; the MHC class II genes are located at the centromeric end (modified from Ref. (77)). (See Color Insert).

C4 and other constituent genes of the RCCX modules (Fig. 1). First, there is a long form and a short form of the *C4* gene, being 20.6 and 14.2 kb in size, respectively. The long *C4* gene has an endogenous retrovirus of 6.36 kb integrated into its ninth intron (50,73). The size of intron 9 in the short

C4 gene is only 416 bp. Second, a *C4* gene may code for a *C4A* protein or a *C4B* protein. The basis for this variation is 5-nt substitutions in exon 26 leading to four isotype-specific amino acid residues, i.e., PCPVLD 1101-6 in *C4A* and LSPVIH 1101-6 in *C4B* (51,69). In haplotypes with two or more *C4* genes, duplication is discretely modular, with modules of 32.7 (long, L) or 26.3 kb (short, S) that include an intact *C4A* or *C4B* gene, a gene for the steroid *CYP21* that is frequently the pseudogene *CYP21A*, and partially duplicated gene fragments *TNXA* and *RP2* (42,47).

Initial structural characterization of a limited number of clones revealed that the *C4* gene linked to factor B at the 5' region (i.e., located at the telomeric end) is a long gene coding for *C4A*, and a short *C4B* gene is present at the centromeric region linked to the functional *CYP21B* gene (9,74,75). However, population studies using an informative or definitive RFLP analysis coupled with phenotypic characterization of *C4A* and *C4B* proteins revealed a great repertoire of length variants for the *RCCX* modules with different combinations of genotypic features for the *RCCX* constituents (11,47,72,76,77).

B. Determining the 1, 2, 3, or 4 Long or Short *C4* Loci in an MHC Haplotype Coding for *C4A* or *C4B* Proteins

To facilitate elucidation of the sophisticated complexity of *C4* genetics, a combination of diagnostic techniques have been created to accurately determine the gene dosage of human complement *C4A* and *C4B* and *RCCX* length variants. These techniques were derived from our current knowledge on the gene organizations and the sequence data of the HLA class III region (7,11,41,78–82). After a series of trials, it appears that the polygenic and gene size variations of *C4A* and *C4B* are best elucidated by concurrent genotypic analysis using RFLPs and phenotypic analysis using immunochemical methods (Figs. 2 and 3).

For genomic RFLPs, we chose the restriction enzymes that yield the maximum information, and deliberately designed length and locations of the hybridization probes so that the band intensities for each set of restriction fragments on the autoradiographs or phosphorimages faithfully reflect the relative gene dosage of the corresponding genes or gene segments. The *TaqI* genomic Southern blots using *RP*, *CYP21*, and *TNX* probes yield information relating to the length variants of the *RCCX* modules, the total number of long and short *C4* genes linked to *RP1* or to *RP2*, the relative dosage of *CYP21B* and *CYP21A*, and the relative dosage of *TNXB* and *TNXA*. Results of *RP-PshAI* Southern blots further provide independent information on the relative dosage of *RP1* and *RP2* that essentially yields the number of *RCCX* modules present in the diploid genome of an individual. A *PshAI-PvuII* genomic

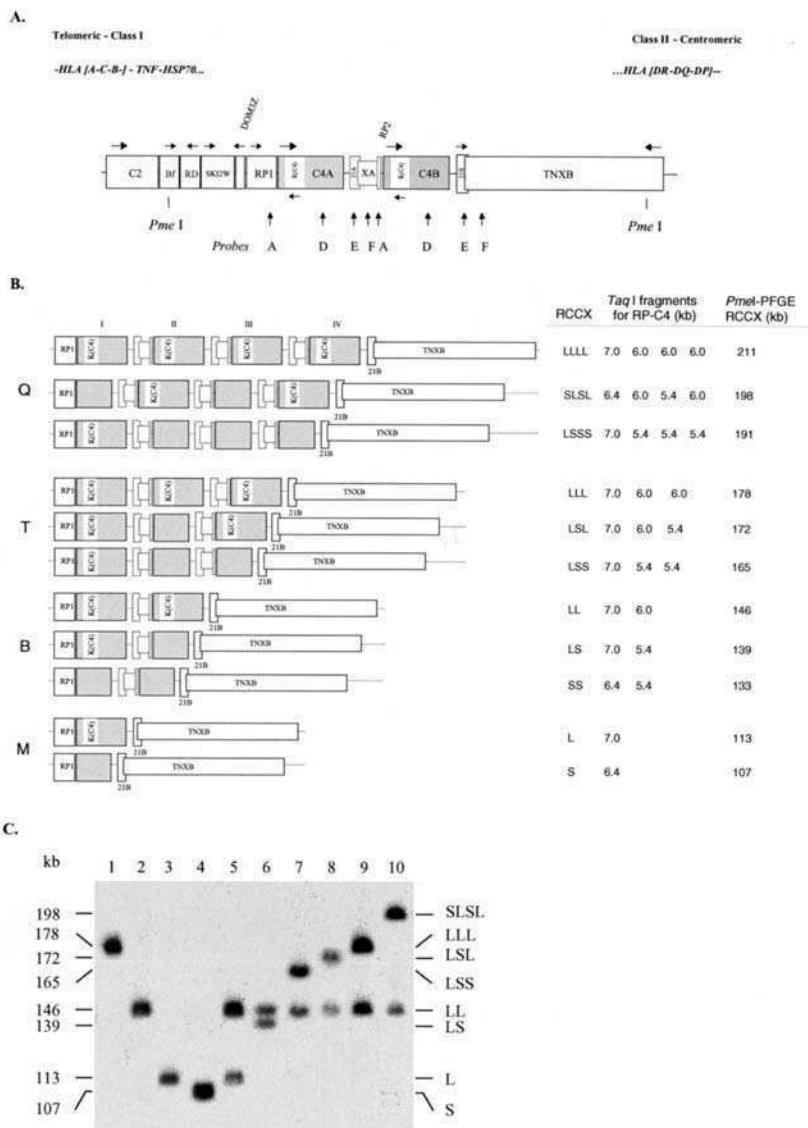


FIG. 2. Detection of the human RCCX length variants. (A) A molecular map of the MCGC with bimodular LL RCCX modules. The locations of the two *PmeI*-restriction sites are shown. A horizontal arrow represents the direction of transcription of a gene. A vertical arrow represents the position at which probes A, D, E, or F hybridizes to the DNA. The position of probe 22–25 is 3 exons upstream of probe D. (B) An alignment of 11 unique RCCX length variants at the telomeric end, which are grouped into Q, T, B, or M haplotypes. The sizes of the *TaqI*- and *PmeI*-restriction fragments for each haplotype are shown on the right. (C) Elucidation of the RCCX length variants in 10 different individuals by pulsed field gel electrophoresis of *PmeI*-digested DNA. Lanes 1–4 are from homozygous with trimodular LLL, bimodular LL, monomodular L, and monomodular S structures, respectively. Lanes 5–10 are from individuals with heterozygous RCCX modules. Key: K(C4), endogenous retrovirus HERV-K(C4); 21A, *CYP21A*; 21B, *CYP21B*; XA, *TNXA*; Q, quadrimodular; T, trimodular; B, bimodular; M, monomodular (modified from Ref. (72)).

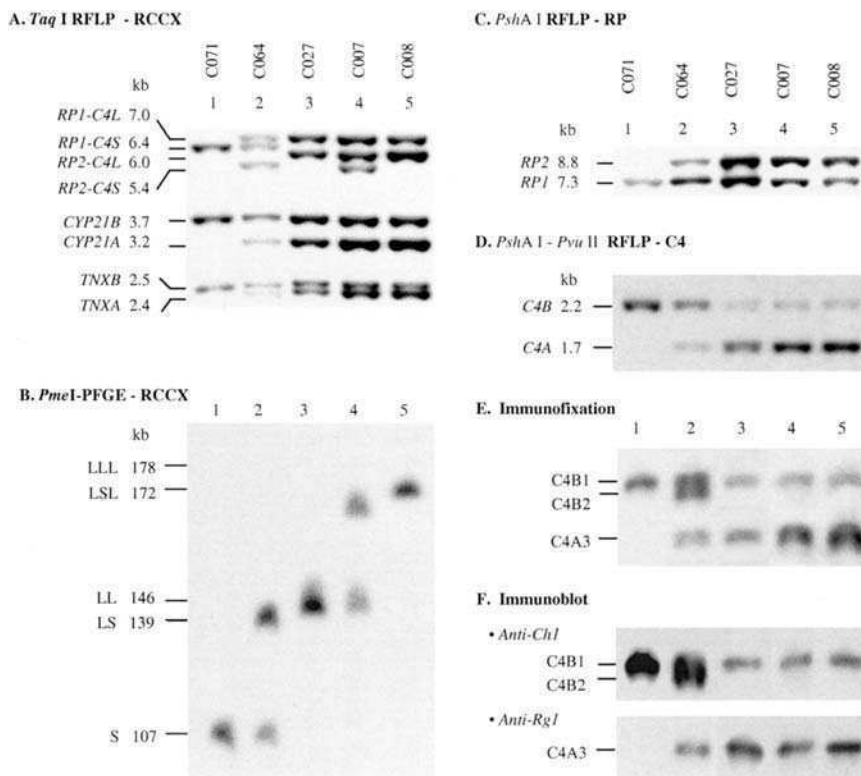


FIG. 3. Genotypic and phenotypic analyses of RCCX structures and complement C4A and C4B from five individuals with 2, 3, 4, 5, and 6 *C4* genes, respectively. (A) Restriction patterns of *Taq*I RFLP upon simultaneous hybridization of probes specific for 3' *RP*, *CYP21*, and 3' *TNX*. (B) *Pme*I-PFGE to show the RCCX length variants of the same five individuals. (C) *Psh*AI RFLP hybridized to a 3' *RP* probe to show the relative gene dosages of *RP1* and *RP2*. (D) *Psh*AI-*Pvu*II Southern blots are hybridized with probes 22–25 to show the relative *C4A* and *C4B* gene dosages. (E) Detection of C4A and C4B protein polymorphism by high-voltage agarose gel electrophoresis (HVAGE) of EDTA-plasma samples and immunofixation using goat antisera against human C4. (F) Immunoblot analyses of the plasma samples resolved by HVAGE using anti-Ch1 mAb (upper panel) and anti-Rg1 mAb (lower panel) (modified from Ref. (72)).

Southern hybridized with a C4d probe gives the results on the relative dosage of *C4A* and *C4B* genes (Fig. 3).

While the *Taq*I and *Psh*AI RFLPs give excellent account of the number of long and short *C4A* and *C4B* genes present in a diploid genome, these results are not necessarily segregated unless family data are available. A long-range mapping method employing pulsed field gel electrophoresis of high molecular-weight genomic DNA digested by *Pme*I restriction enzyme for

Southern-blot analysis is applied to determine the variation of *C4* gene number. The recognition and cleavage site of *PmeI* has no CpG sequence, which is common in most other restriction enzymes used for long-range mapping by PFGE. The *PmeI*-digested fragments are predictable, homogeneous, independent of the tissue origins of the genomic DNA that may have different methylation patterns, and reflect the actual size of the DNA fragments with the recognition sequences. Therefore, the *PmeI*-PFGE allows a definitive and unambiguous elucidation of the number and size of the RCCX modules or *C4* genes that are present in haplotypes. The application of these techniques has helped the demonstration of eight common and three rare RCCX length variants (Fig. 2) (11,47,72,77).

The limitation of the genomic RFLPs to determine *C4A* and *C4B* gene dosage and RCCX variants is the requirement of relatively large quantities of genomic DNA, which may pose constraints in multiloci genetic studies. Two PCR-based methods to determine *C4A* and *C4B* gene dosages have been developed to facilitate the genotyping when quantities of genetic materials are limited. The module-specific PCR yields information on the total number of RCCX modules and therefore the total number of *C4* genes that are present in a diploid genome. A refined "hot-stop" PCR (83) coupled with *PshAI* or *XcmI* RFLP allows elucidation of the ratio of *C4A* and *C4B* genes, and the *C4* genes with Rg1 or with Ch1 antigenic determinants, respectively (72).

C. RCCX Modular and *C4* Polygenic Variations in Caucasians

To establish the genetic variations of complement *C4* in Caucasians, we applied definitive molecular biologic techniques to determine the diversities of the constituents of RCCX modules, i.e., *RP1* and *RP2*, long and short genes for *C4A* and *C4B*, *CYP21A* and *CYP21B*, and *TNXA* and *TNXB* in 150 healthy Caucasian females in Columbus Ohio, and in 128 Hungarian Caucasians.

Seventy-six percent of the *C4* genes in the Columbus Caucasian population have the retrovirus integrated, 24% do not. In the Columbus study population, seven RCCX length variants were identified, including monomodular L, monomodular S, bimodular LL, bimodular LS, and trimodular structures LLL, LSS, and LSL.

Although the two-loci model with a *C4A* gene and a *C4B* gene in the human MHC is a deeply ingrained concept in the genetics of human *C4* (60–62), it is important to note that such *C4A*–*C4B* configuration only accounts for 55% of the RCCX haplotypes in the Columbus Caucasian population. The *C4A*–*C4A* and *C4B*–*C4B* configurations comprise an additional 13.3% and 0.67% for the bimodular RCCX structures, respectively. The remaining 31% of the RCCX haplotypes are trimodular, with three *C4* genes, or monomodular, with one *C4* gene. The monomodular structures have a

frequency of 17%, among them 12% code for C4B and 5% code for C4A. The frequencies of the trimodular structures are almost equally split between LLL and LSS. Two-thirds of the LLL structures have *C4A-C4A-C4B* haplotypes, and two-thirds of the LSS structures have *C4A-C4B-C4B* haplotypes. From another point of view, the number of *C4* genes present in a diploid genome (*C4* gene dosage) varies between two and six in the population. About 52% of the population has a *C4* gene dosage of four. The frequencies for *C4* gene dosages of 2, 3, 5, and 6 in Caucasians are 2, 25.3, 17.3, and 3.3%, respectively (Fig. 4A). Such a high degree of variation in gene number is unprecedented in human populations. The high frequencies of the mono-, bi-, and trimodular RCCX structures in the population prompted a revision of the human *C4* genetics from a fixed two-locus model to the dynamic 1–2–3 *loci* system (11).

The RCCX and *C4* genotypic and phenotypic variations have been investigated in an independent, central European population. A definitive genotypic analysis of the polygenic *C4A* and *C4B* genes together with the RCCX modules was performed concurrent with phenotypic characterizations of the polymorphic variants for C4A and C4B isotypes, measurement of serum C4 concentrations, and assay of hemolytic activities in 128 healthy individuals from Budapest, Hungary. In this population, 69% were females, 31% were males.

Table II summarizes the similarities and differences between the Columbus and Budapest Caucasian populations. It is noted that the

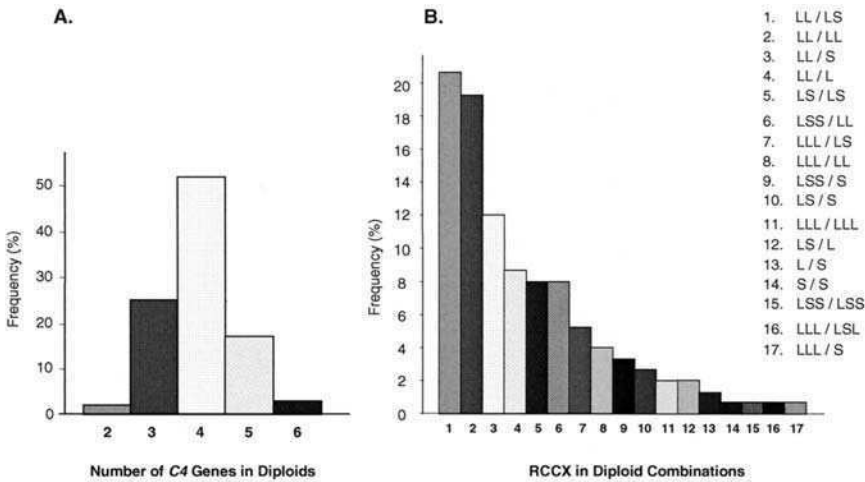


FIG. 4. Variations of *C4* gene dosages and RCCX modules in a Caucasian population. (A) Frequencies of individuals with 2–6 *C4* genes in a diploid genome ($n = 150$). (B) Frequencies of individuals with 17 different combinations of RCCX length variants (modified from Ref. (11)).

Table II
A COMPARISON OF RCCX LENGTH VARIANTS AND *C4* GENE FREQUENCIES BETWEEN OHIOAN AND HUNGARIAN CAUCASIANS

	Ohioan Caucasian		Hungarian Caucasian	
	<i>n</i>	Frequency	<i>n</i>	Frequency
Number of L genes	450	0.763	384	0.774
Number of S genes	141	0.237	112	0.226
Total	591		496	
RCCX modules				
LL	138	0.460	128	0.500
LS	69	0.230	74	0.289
S	33	0.110	22	0.086
L	18	0.060	12	0.047
LSS	19	0.063	7	0.027
LLL	22	0.073	10	0.039
LLS	1	0.003	1	0.004
Lx	0	0	2	0.008
Total	300		256	
Number of <i>C4</i> genes in diploids				
2	3	0.020	2	0.016
3	38	0.253	27	0.211
4	78	0.520	84	0.656*
5	26	0.173	15	0.117
6	5	0.033	0	0
Total	150		128	
<i>C4</i> gene index	3.94		3.875	
Number of <i>C4A</i> genes				
0	1	0.007	2	0.016
1	30	0.201	26	0.203
2	69	0.463	73	0.570*
3	44	0.295	24	0.190*
4	5	0.034	3	0.023
5	1	0.007	0	0
Total	150		128	
Gene index	2.167		2.0	
Number of <i>C4B</i> genes				
0	3	0.020	1	0.008
1	48	0.320	26	0.200*
2	80	0.533	90	0.700*
3	18	0.120	10	0.080
4	1	0.007	1	0.008
Total	150		128	
Gene index	1.773		1.875	

*Marked difference between the two populations: Lx, undefined bimodular haplotype.

frequencies of the long and short *C4* genes are remarkably similar, i.e., 76–77% long and 23–24% short. A similar distribution pattern of RCCX modular variation is also observed between these two Caucasian populations. When we compared the *C4* gene dosage in diploid genomes, it appears that the Hungarian Caucasians have a higher frequency of four *C4* genes (i.e., 65.6%) than that in Ohio Caucasians (i.e., 52%). The Hungarian Caucasians also have slightly higher frequencies of two *C4A* and two *C4B* genes in a genome, but lower frequencies of three *C4A* and one *C4B* genes than those in Ohio Caucasians.

D. Heterozygosity in RCCX Length Variants Promotes Genetic Exchange or Unequal Crossovers

1. HETEROZYgosITIES OF RCCX LENGTH VARIANTS

At least 22 different haplotypes of RCCX with different combinations of long and short genes for *C4A* and *C4B*, *CYP21A* and *CYP21B*, *TNXA* and *TNXB* have been detected (Fig. 4B). In the diploid genomes, 17 combinations of RCCX length variants have been shown. Of these, 69.4% of the population have heterozygous combinations of the number of RCCX modules and/or the size of the *C4* genes. The high frequency of trimodular and monomodular structures and the heterozygous combinations may promote “misalignments” of the homologous chromosome 6 during meiosis. Alignments of heterozygous RCCX length variants at either centromeric or telomeric ends would lead to various degrees of mispairings and exchanges of sequences. There are two possible outcomes. First, the polymorphic residues among the *C4A* and *C4B* variants in the population are homogenized (84). The reverse association of the Rg blood-group antigenic determinant with *C4B5* (37,69) and the Ch blood-group antigenic determinants with *C4A1* are the prototypic examples (69,85,86). Second, it may generate HLA haplotypes that predispose to the development of disease states. This is because the recombination may lead to the acquisition of deleterious mutations from gene fragments or pseudogenes to the respective functional genes, or the deletion of structural genes in or close to the RCCX modules that would be nonproductive or lethal. A typical example of unequal, interchromosomal crossovers at the RCCX modules leading to an MHC-associated disease is the steroid 21-hydroxylase deficiency in congenital adrenal hyperplasia (CAH).

2. RECOMBINATIONS LEADING TO STEROID 21-HYDROXYLASE DEFICIENCY

There is high sequence identity (98.7%) between *CYP21B* and its pseudogene *CYP21A*. This pseudogene carries 58 point mutations, 8 mini-insertions, and 2 mini-deletions in the 5.14-kb sequence that includes 5'- and 3'-flanking regions (87,88). Among the mutations in *CYP21A*, three are

particularly deleterious: an 8-bp deletion in exon 3, a single-nucleotide insertion in exon 7, and a nonsense mutation in exon 8.

The heterozygosity of a bimodular RCCX with a monomodular or a trimodular structure may create an interchromosomal pairing of a *CYP21B* gene with a *CYP21A* gene during meiosis. Exchange of genetic information between the *CYP21B* and *CYP21A* may reduce or abrogate the functionality of the *CYP21B* gene, depending on the site and the extent of exchanges. Specifically, the *CYP21B* gene in the monomodular-long haplotype (mono-L) would be more vulnerable to crossover or conversion in heterozygous RCCX individuals due to its intrinsic misalignments with the *CYP21A* in a bimodular or trimodular chromosome. Sequencing of the mono-L *CYP21A* gene from the HLA B47 DR7 C4A1 haplotype revealed a hybrid gene with marker mutations for *CYP21A* at one end and specific sequence for *CYP21B* at the other end, suggesting a past recombination between *CYP21A* and *CYP21B* (89).

In a small study population of 22 CAH patients, it was found that 20.5% of the salt-losing CAH patients had mono-L with the pseudogene *CYP21A*, and 22.7% of the CAH haplotypes had the *CYP21A*–*CYP21A* configuration in bimodular structures (11) (Fig. 5). This bimodular, disease haplotype could be the result of recombinations between a trimodular chromosome with *CYP21A*–*CYP21A*–*CYP21B* configuration and a regular bimodular chromosome (at or after the second *CYP21A* gene of the trimodular chromosome).

The absence of the *CYP21B* gene may also be caused by an unequal crossover between *TNXA* and *TNXB* from heterozygous RCCX modules. The recombination would lead to the deletion of *CYP21B*–*C4B*–*RP2* and result in the formation of a *TNXB*–*XA* hybrid with a 120-bp deletion between exon 36 and intron 36 (47,90). This *TNXB*–*XA* hybrid would produce a truncated *TNX* protein without the carboxyl fibrinogen domain (Fig. 5C). In two independent CAH patients (E1 and AE) from Ohio, such *TNXB*–*XA* recombinants were discovered and characterized. Similar recombination products were also detected in CAH patients with the connective tissue disorder Ehlers Danlos syndrome (EDS) (91), raising the possible association of *TNX* mutation with EDS. In a study of 39 Dutch families with steroid 21-hydroxylase deficiency, it was found that ~10% of the patients are carriers of *TNXB* deficiency, who probably had the *TNXB*–*XA* recombinant similar to that in E1 (Fig. 5C) (92).

In a patient with juvenile rheumatoid arthritis (JRA-L1), the reciprocal recombination product with two *CYP21B* genes and a *TNXA*–*XB* hybrid in a haplotype was discovered. The recombination breakpoint regions from E1 and L1 were both cloned and sequenced (47,90). Informative sites for *TNXA* (pink) and *TNXB* (blue) specific sequences were switched in the recombinants. A sequence motif resembling the bacterial recombination hotspot χ is present at the site of recombination region (lower panel, Fig. 5C).

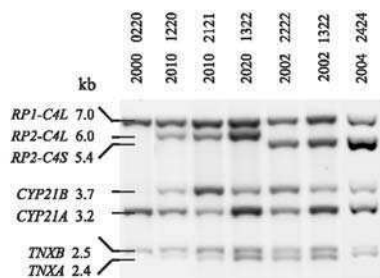
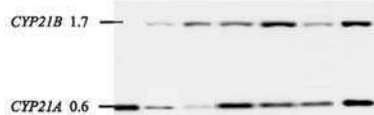
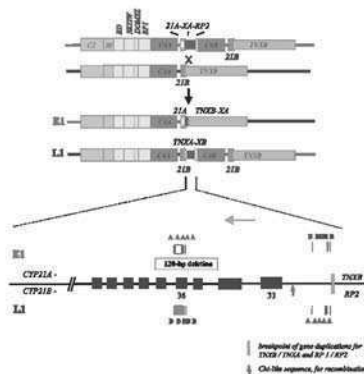
A. *TaqI* RFLP - RCCXB. *BsaI* RFLP - *CYP21*C. Recombination between *TNXB* and *TNXA*

FIG. 5. RCCX modular variations in classical CAH patients. (A) *TaqI* RFLP on seven selected CAH patients to show the RCCX diploid combinations. The eight-digit number above each lane represents the relative dosage of the 7.0, 6.4, 6.0, and 5.4 kb fragments for RP-C4, the 3.7, 3.2, 2.5, and 2.4 kb fragments for *CYP21B*, *CYP21A*, *TNXB*, and *TNXA*, respectively. (B) *BsaI* RFLP to determine the ratio of the *CYP21A* pseudogene and the mutant *CYP21B* genes in the same patients. (C) Unequal crossover between bimodular and monomodular RCCX chromosomes as a mechanism for gene deletion. The upper panel shows the pairing between homologous chromosomes and recombination between *TNXA* and *TNXB* to yield the chimeric 5'-*TNXB*-XA-3' with *CYP21B* deletion in E1, or the chimeric 5'-*TNXA*/XB-3' with *CYP21B* duplication in L1, respectively. The lower panel shows the reciprocal association of *TNXA* markers (A) and *TNXB* markers (B) in a patient with CAH (E1) (47), and a patient with pauciarticular juvenile rheumatoid arthritis (L1) (90). (See Color Insert.)

3. *De Novo* RECOMBINATIONS OR GENE CONVERSION-LIKE EVENTS AT THE RCCX

While the recombinants in *C4*, *CYP21*, and *TNX* present in the population are most likely remnants of genetic recombinations that occurred generations ago, *de novo* recombinations or exchange of specific polymorphic or deleterious sequences between constituents of the RCCX have been documented in the following cases:

- (i) Exchange or conversion of the five isotype-specific nucleotide sequences in *C4* exon 26, thereby switching a *C4A3a* gene to a *C4B5-like* gene in a Finnish individual with HLA A19 B15 C3 DR5 (93);
- (ii) A recombination between bimodular LS haplotype from HLA A30 B13 DR7 with a monomodular S haplotype from HLA A1 B8 DR3 leading to the generation of a novel monomodular L structure with *CYP21A* was found in a hybrid haplotype HLA A30 B13 DR3 of a 21-hydroxylase deficiency patient (94); and

- (iii) A probable recombination between bimodular LL haplotype (coding for C4A3 C4B1) and a monomodular S (for C4AQ0 C4B1) to generate a monomodular S haplotype with a C4AQ0 C4BQ0 phenotype. This recombinant was discovered in a pediatric patient with recurrent sinopulmonary infections. However, the basis for the generation of C4BQ0 phenotype was not elucidated (95).

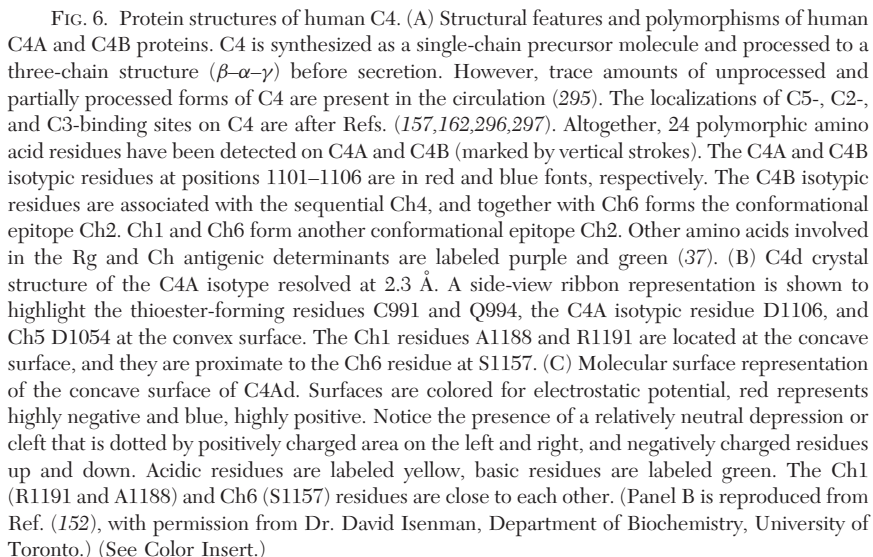
4. ANCESTRAL HAPLOTYPES AND POLYMORPHIC FROZEN BLOCKS

The presence of multiple RCCX length variants and the homogenization of polymorphic sequences between homologous alleles would suggest that the MHC class III region, particularly at the RCCX modules, has/had been under a dynamic state of genetic exchanges or recombination during primate evolution. However, the paradox is that the human MHC is also well known for linkage disequilibrium of class I and class II alleles that flank the class III region where the RCCX modules are located. MHC haplotypes, which may extend over much of the length of the MHC, were observed by Dawkins and coworkers and termed “ancestral” haplotypes (AHs) (96–98). It was proposed that AHs are sequences that diverged radically at some time, but which have remained conserved during the human expansion. AHs contain several regions or blocks of several hundred kilobases which are characterized by high levels of polymorphism, but little or no recombination. These blocks, termed “polymorphic frozen blocks” appear to have been maintained from ancient ancestral sequences through a “freezing” of recombination. The mechanism leading to this paradoxical phenomenon is yet to be established. Whether the physical length variations of the RCCX modules and also the nearby *DRB9* to *DRB1* genomic regions among different individuals would suppress *productive* recombinations during meiosis has not been critically analyzed.

III. Structural Diversities of the Human and Mouse C4 Proteins

A. Tyrosine Sulfation at the Carboxyl End of the α -Chain

C4 is synthesized as a single-chain precursor molecule ~ 200 kDa in size and processed to a three-chain structure linked by disulfide bonds before secretion (Fig. 6) (99–104). The precursor C4 molecule has 1725 amino acid residues that are preceded by a 19-amino acid leader peptide characteristic of a secretory protein (51,82). Besides the formation of the internal thioester bond (105–109), the mature C4 protein undergoes a series of



post-translational modifications including sulfations, glycosylations, and proteolyses. The presence of tyrosine-O-sulfation is probably a unique feature of C4 among the complement proteins. Protein sulfation is a modulator of extracellular protein-protein interactions (reviewed in Ref. (110)). The sites

of sulfation are located at residues 1198, 1199b, and 1200 that are present at a region surrounded by many negatively charged amino acids (111). Sulfation of C4 probably plays a role on the activation of C4 by the C1 complex. Nonsulfated C4, secreted by HepG2 cells cultured in media with sulfation inhibitors, has a hemolytic activity about 50% of normal C4. This decreased activity may be due to an impaired protein-protein interaction between the nonsulfated C4 and C1s of the C1 complex, as a 10-fold increase in concentration of C1s is required to yield the equivalent cleavage of sulfated C4 (112).

The derived amino acid sequence from a cDNA clone corresponding to the C4A4 allotype was shown to have a three-residue deletion (EDY) at positions 1399a, b, c. This deletion contains one of the three sulfation sites in human C4 (113). However, whether the deletion was a cloning artifact or a genuine polymorphism has yet to be clarified.

B. Structural Variation Created by Difference in Glycosylation

Carbohydrate accounts for ~7% of the molecular weight of the human protein (114). Four N-linked glycosylation sites are present in human C4, among them N207 is present on the β -chain, and N843, N1309, and N1372 are present on the α -chain (113). The oligosaccharide structures of human C4 synthesized by the hepatoma cell HepG2 were studied in detail (115). The single glycosylation on the β -chain belongs to the high-mannose oligosaccharide type. Digestion of the plasma C4 with endo-H reduces the 78-kDa β -chain by ~5 kDa. The three glycosylation sites on the α -chain have the biantennary, complex-fucosylated oligosaccharides. The heterogeneity of glycosylation at each site does contribute to structural variation of the C4A and C4B allotypes. One of the essential procedures in C4 allotyping is to reduce such variation by neuraminidase treatment (31). It is not known if glycosylation plays a role on the protein function of human C4A and C4B.

The sites of glycosylation in the mouse Slp and C4 protein are slightly different from those of human C4. There is a glycosylation site at N724 in the anaphylatoxin domain of both mouse C4 and Slp. The α -chain molecular weight of mouse Slp and C4 are 105 kDa and 94–98 kDa, respectively. The lower molecular weight for mouse C4 protein is probably due to the absence of the glycosylation site at position 843, which is present both in mouse Slp and in human C4A and C4B (116). The mouse C4 proteins from the *w7*, *w16* and *w19* haplotypes have low hemolytic activity and their α -chains are 94 kDa instead of 98 kDa. It was shown that there is one glycosylation site C-terminal to the thioester of the α -chain (117,118).

A comparison of the mouse C4 amino acid sequences from the *H-2^{w7}* haplotype (119) with those from the *H-2^{FM}* (120), *H-2^d* (121), and *H-2^b* (122) haplotypes with high hemolytic efficiency reveals the N1305K substitution. This substitution abrogates the N-linked glycosylation site at the C4d region and probably accounts for the 4-kDa difference in the molecular weight and the four-fold reduced hemolytic efficiency (Table III) (121).

TABLE III
POLYMORPHISMS OF MOUSE C4 AND Slp PROTEINS

(A) Allelic variation of mouse C4					
Amino acid no.*	Mouse haplotype/C4 sequence				<i>Slp sequence</i> ⁺
	<i>b</i>	<i>d</i>	<i>w7</i>	<i>FM</i>	
113	F	–	F	Y	F/Y
308	G	–	G	E	G
551	Q	–	Q	E	Q
585	M	M	T	T	T
819	P	P	P	R	P
1305 ⁺⁺	N	N	K	N	N
1423	R	R	R	K	R
1434	V	V	A	V	A
1517	E	Q	E	E	E
1592	T	A	A	A	A

(B) Allelic variation of mouse Slp			
Amino acid no.*	Haplotype/Slp sequence		<i>C4 sequence</i>
	<i>w7</i>	<i>FM</i>	
57	S	L	S
86	H	Y	H
113	F	Y	F/Y
134	P	R	R
276	R	Q	R
354	K	T	K
355	R	A	R
356	H	Q	H
1222	A	R	T
1417	Q	E	Q

Database accession numbers for mouse C4 and Slp protein sequences are: C4 (*H-2^b*), AAC05279; C4 (*H-2^d*), CAA28936; C4 (*H-2^{w7}*), NP_033910; C4 (*H-2^{FM}*), AAA39557; Slp (*H-2^{w7}*), AAA40117; and Slp (*H-2^{FM}*), AAA39683.

*Amino acid numbering is based on the N-terminal residue of the mouse C4 β -chain (K) as number 1.
⁺Mouse C4 and Slp proteins share 95% sequence identities with 88 substitutions in 1738 residues. In addition, there is a three-residue deletion (EED) at amino acids 742–744 in Slp.
⁺⁺N1305 is a glycosylation site in both mouse C4 and Slp.

C. Structural Variation Created by Incomplete Processing of the C4 Protein

Four basic residues are present between the β - α and the α - γ chain junctions, which are normally removed before secretion (123). The plasma form of the α -chain is further reduced by 5 kDa as the 26 amino acids from the carboxyl terminus are cleaved, presumably by an extracellular elastase-like protease (124). However, the described proteolytic processings are incomplete as 5–7% of the precursor molecule remains unprocessed as single chained pro-C4, or partially processed with two-chained structures (β - α + γ or β + α - γ), and 8% of C4 molecules remain as the secretory form. In addition, the removal of the four basic residues at the carboxyl termini of the β -chain and the α -chain are also imperfect (68). These irregularities at the C-termini are readily demonstrated by earlier phenotyping experiments using the immunofixation technique, as each C4 allotype is manifested as one major and two minor mobility variants (61). Although no difference in the hemolytic activity has been demonstrated in the completely or incompletely processed forms of human C4, these structural variants add another dimension of structural diversity to the C4A and C4B proteins (125).

D. Protein Sequence Polymorphism and Functional Variation of C4A and C4B

Substantial efforts were made to determine the primary structures of human C4 by protein sequencing, which yielded approximately 28% of the amino acid sequence of the whole protein (114,126–130). This facilitated the cloning of the C4 cDNA and gene (9,113,131). The complete primary structures for C4A4 and C4A3a were deduced and shown to code for 1744 amino acids with a 19-amino, acid leader peptide characteristic of a secretory protein (51,82,113).

A compilation was performed on known polymorphic sequences in C4A and C4B alleles. The data is derived from 19 full-length or partial cDNA and genomic DNA sequences (7,51,69,82,132–134). Besides the seven synonymous changes that do not alter the amino acid sequences, a total of 24 polymorphic amino acid residues have been elucidated in the C4 protein sequences. Six of these polymorphic residues were detected by amino acid sequencing of C4 proteins isolated from pooled blood sera (124,126) and are yet to be shown in specific C4 allotypes. Among the polymorphic residues, five are located on the β -chain, 18 on the α -chain, and one on the γ -chain (Fig. 3). These polymorphic sites can be categorized into four groups:

- (i) the four C4A/C4B isotypic residues at positions 1101, 1102, 1105, and 1106;

- (ii) the Rg and Ch antigenic determinants at positions 1054, 1101, 1102, 1105, 1106, 1157, 1188, and 1191;
- (iii) the C5-binding site at residues 458 and 459; and
- (iv) private or specific allelic residues.

1. ISOTYPIC RESIDUES OF C4A AND C4B

The C4A/C4B isotypic residues PCPVLD 1101-6 LSPVIH are responsible for the differences in chemical reactivities toward amino group and hydroxyl group containing antigens (51,69,135). Site-directed mutagenesis experiments suggested that D1106 is the critical residue affecting the favorable binding of C4A toward IgG immune aggregates (70), and H1106 is responsible for the catalytic binding of C4B thioester carbonyl group to hydroxyl groups (34). Therefore, it is generally accepted that the defining feature for C4A and C4B rests on D/H 1106.

The evolution of the P1101L, C1102S, L1105I, and D1106H, C4A/C4B isotypic residues is of special interest. The mouse *Slp* and *C4* genes both have the hybrid combination P, C, I, and H at the orthologous positions (136–139). The gorilla has the identical C4A/C4B isotypic residues as humans but the other Old World primates have different combinations of P/L, C/S, L/I, and D/H at positions 1101, 1102, 1105, and 1106, respectively (140,141). Most, if not all, Old World primates have two or more *C4* loci that evolved to C4A and C4B proteins. Sheep and cattle also have two or more *C4* loci that diversify to encode C4A- and C4B-like proteins with PFPVMD and PCPVIH at the respective isotypic positions from 1101 to 1106 (142). Most other mammals have single *C4* loci that code for C4B-like proteins (reviewed in Ref. (143)). Therefore, C4A probably evolved from C4B after gene duplication. In the normal Caucasian population, the gene frequency of *C4A* (55%) is slightly more than that of *C4B* (45%) (11). Deficiency of C4A appears to be a common denominator in human lupus among different ethnic groups (144), suggesting C4A has acquired a new and relevant role in immunity.

2. THE PRESENCE OF ACTIVATED C4 ON CELL SURFACES: THE RODGERS (RG) AND CHIDO (CH) BLOOD-GROUP ANTIGENS

The Rg and Ch blood-group antigens are formed by the deposition of the activated C4, C4b, on the membranes of erythrocytes. In the presence of CR1 and factor I, C4b is cleaved twice on the α -chain to yield the soluble C4c and the membrane-bound C4d (35,36,145,146). Alloantibodies against complement C4 are formed in blood-transfusion recipients who have C4A or C4B protein deficiency, or if there are mismatches in the variants of C4A or C4B between the donor and the recipient. Rodgers antibodies are formed in C4A-deficient recipients. Chido antibodies are formed in C4B-deficient recipients.

The generation of such alloantibodies could cause adverse-transfusion reactions (147,148). Radioligand-binding assays using C4d monoclonal antibody established that the number of C4d molecules on the surface of a red cell was 90 ± 30 and 670 ± 693 from normal and SLE patients, respectively (149). The latter would infer higher consumption or metabolic rates of C4 in the lupus patients. The mean number of C3d on the red cell membrane in the normal and lupus patients are 4–6 fold less than that of C4d.

The red cell membranes are not the only surface loaded with activated C4 or C4d fragments. A specific membrane marker for mouse follicular dendritic cells, FDC-M2, has been implicated to be the activated C4 (150). The presence of the processed C4 products on the membrane of human red cells and mouse FDC are indicative that these cells have circulated through sites of complement activations or microbial infections. However, it is not known if the membrane-bound C4d also plays a direct physiological function in these cells.

Two Rg (Rg1 and Rg2) and six Ch (Ch1 to Ch6) epitopes were defined from a battery of alloantibodies against C4. Through structural characterization of C4d regions in C4A1 and C4B5 (69), which exhibit reversed associations with the major Ch1 and Rg1 determinants (151), respectively, and several other informative C4A and C4B alleles, a model with continuous (or sequential) and discontinuous (or conformational) epitopes for Rg and Ch determinants was proposed (37,69). The major determinants Rg1 and Ch1 are defined by VDLL 1188-91 and ADLR 1188-91, respectively. The C4B isotypic residues LSPVIH 1101-6 are associated with Ch4. It was deduced that the Ch6 determinant (S1057) and the Ch1 determinant constitute the conformational epitope Ch3, and that the Ch5 determinant (G1054) and Ch4 determinant constitute the conformational determinant Ch2 (37). Interestingly, the C4A isotypic residues PCPVLD 1101-6 and the antithetic residue for Ch5, D1054, are not associated with an Rg antigenic determinant (Fig. 6A).

The structural model of sequential and conformational Rg and Ch determinants was strongly supported by serological data from 325 families (36). It was extended to include another conformational epitope WH that involves Rg1 and Ch6 (38). The recent C4d X-ray crystal structure of human C4d is highly consistent with the model (152–154).

3. THE X-RAY CRYSTAL STRUCTURE OF HUMAN C4d

A fusion protein containing human C4d protein sequence with the Ch-5 D1054, C4A isotypic residues PCPVLD 1101-6, Ch6 S1157, and Ch1 ADLR1188-91 was produced in bacteria, purified, subjected to V8 protease cleavage, repurified by chromatography, and crystallized under reducing conditions. The latter ensured monomeric structure of C4d that would

otherwise dimerize by disulfide linkage through C991, which normally forms the thioester bond with Q994. Hollow rod-shaped crystals were obtained, which allowed X-ray diffraction with a resolution of 2.3 Å. A three-dimensional structure that represented 79.5% of the entire C4d was determined by molecular replacement, using the coordinates from the X-ray crystal structure of the complement C3d fragment resolved at 2.0 Å (152,155,156). With a framework structure that is essentially superimposable with C3d, C4d has an α - α six-barrel fold (Fig. 6B). The structure has six parallel α -helices forming the core of the barrel, surrounded by another set of six parallel helices running antiparallel to the core helices. Attempts to crystallize the B-isotype C4d region were unsuccessful but it was suggested that the C4d structure of C4B should be identical to C4A.

In the open conformation, the thioester residues, C4A isotypic residues and the Ch5 residue are located on the convex surface of the barrel. The Ch1 and Ch6 antigenic determinants are located on the opposite concave surface. The crystal structure confirms several important predictions about the chemical reactivities and serological properties of C4. First, the location of D/H1106 at the convex surface together with the thioester C991 and Q994 support the role of H1106 in the catalysis of the thioester carbonyl group from Q994 in a transesterification reaction to form covalent ester or amide bonds with hydroxyl or amino groups of substrates (34). Second, the proximate locations of D/G1054 to PCPVL D/LSPVIH 1101-6 at the convex surface were as predicted by the presence of the conformational epitope Ch2 between G1054 (Ch5) and LSPVIH 1101-6 (Ch4) (37). Third, the proximate locations of S1157 (Ch6) and ADLR 1188-91 on the concave surface were also as predicted by the conformational epitopes Ch3 and WH (37,38,153).

In contrast to the highly acidic nature of the concave surface in C3d, the central depression area of C4d concave surface is relatively neutral (Fig. 6C). No electron density was detected for a 24-residue loop containing the anchor site for C3b, S1217 (157). This loop appears to extend from a highly acidic surface on the side of the barrel that would be involved in the formation of the C5 convertase (152). Other unresolved regions included 39 residues at the N-terminus and 15 residues at the C-terminus where an N-linked glycosylation site is located.

The crystal structure raised the possibility of a receptor-binding site at the concave surface where the Rg1/Ch1 epitope is located. One proposed candidate is the complement receptor CR1 (152).

4. C4A6 AND THE C5 BINDING SITE

Complement C4A6 is an unusual allotype that shows fast mobility in the electrophoretic typing gel but an almost nonexistent hemolytic activity (158,159). Extensive immunochemical analyses revealed C4A6 retains all

characteristic C4A-specific properties but it has a very low C5 convertase activity, which is caused by the lack of C5 binding to the C4b–C3b or to the C4b–C4b complex. In other words, the high affinity-binding site for C5 is abrogated in the C4A6 (160). Sequencing of the *C4A6* gene revealed the R458W substitution on the β -chain (161) that was confirmed to confer low C5 convertase activity by site-directed mutagenesis, implying the R458 is part of the C5-binding site on the β -chain (162). This observation was further substantiated by the discovery and characterization of a hemolytically inactive C4B1 allotype associated with HLA haplotype A28 B35 *Cw4* DR6, C4A3 C4B1(*hi*), that has the P459L substitution (163), which is adjacent to the polymorphic site on C4A6. The defective C5 convertase activity translates to the inability to generate the powerful anaphylatoxin C5a and to assemble the membrane-attack complex through the classical-activation pathway. Under appropriate conditions, however, C4A6 may complete the complement pathway through the activation of C3 and the amplification loop of the alternative pathway (164).

There is a report suggesting that mouse C4 could be devoid of C5 convertase activity. In keeping with this observation, site-directed mutagenesis experiments by the incorporation of the mouse C4 sequence at 458–459 PL to human C4 abrogated the C5 convertase activity (165).

In the Caucasian study population, C4A6 has an allotype frequency of 4.7% in C4A (or 2.55% among all C4A and C4B allotypes) (11). Previous studies estimated an allotype frequency of C4A6 between 1.5% and 7.9% in Caucasians and African Americans, but it is infrequent in Asian Chinese, Korean, and Japanese (166–168). Intriguingly, allelic frequency of C4A6 is significantly elevated, i.e., 19.4% (169) to 26% (170) in patients with psoriasis. C4A6 is in strong linkage disequilibrium with HLA B17 (or B57). It remains to be determined if A6 is involved in the pathogenesis of the dermatological disorder. In the Polish psoriasis patient population, it was also found that the frequency of another rare C4 allele, C4B3, was significantly increased from 1% in controls to 13.4% in patients ($P=0.0009$) (169).

5. THE ACTIVATION PEPTIDE C4a

The activation of C4 by the C1s of the C1 complex or by MASP of the MBL complex leads to the proteolysis of the N-terminal region of the α -chain. A 77-amino acid peptide, C4a (residues 616–737), is released. C5a, C3a, and C4a are anaphylatoxins (ANAs) (171,172). C5a and C3a are potent inflammatory agents that promote contraction of smooth muscles, increase vascular permeability and platelet aggregation. They also induce the release of histamine by degranulation of mast cells and stimulate the secretion of interleukin-1 from monocytes. C5a is also a powerful chemotactic agent for phagocytes (15,171). The physiologic role of C4a remains as a mystery,

however. It was suggested that C4a was 100- to 1000-fold less active than C5a in the promotion of smooth muscle contraction or the increase of vascular permeability. However, some of the previous observations on C4a activities could be the results of C5a contamination (173). On the other hand, it was suggested that C4a might be playing an anti-inflammatory role by inducing the release from monocytes of a 20-kDa protein that inhibits monocyte chemotaxis (174). This reaction was extremely sensitive as the chemotaxis-inhibitory activity could be achieved at 10^{-16} M of C4a for 5 min at 37 degrees (175). Using guinea-pig macrophages as target cells, human C4a was shown to induce a biphasic Ca^{+2} mobilization characterized by a rapid mobilization from the intracellular Ca^{+2} pool followed by a weak extracellular Ca^{+2} influx. The macrophages thereby became desensitized and did not react to C4a again, but remained responsive to C3a induction. Moreover, human C4a was unable to compete against human C3a on binding to guinea-pig macrophages (176) and had no effects on Ca^{+2} mobilization in neutrophils or cells stably expressing cloned human C3a receptors, suggesting the cellular receptor for C4a is distinct from C3a receptor (173,177). Although the carboxyl-terminal Arg would have been removed, submicrogram levels of C4a are continuously present in the circulation and therefore cells in contact with vascular circulation could be continuously desensitized for C4a induction. Therefore, a careful choice of *naïve* target cells or tissues might be important in testing the human C4a activities. It might shed light on the physiologic roles of C4a if its receptor was identified and characterized.

A P707L polymorphism was detected in human C4a but the possible effect on its activity has not been tested (82).

6. MOBILITY POLYMORPHISM OF THE C4 α -CHAIN AND β -CHAIN

The C4A/C4B isotypic residues may contribute to the apparent molecular-weight difference of the α -chain by 2 kDa, in which the C4A α -chain migrates as a 96-kDa polypeptide, while the C4B α -chain migrates as a 94-kDa polypeptide (62,115). A similar phenomenon of the structural polymorphism detectable by SDS-PAGE on the β -chain was observed, which was described as the H (heavy) and the L (light) forms (178). The H and L variants were observed in many allotypes of C4A and C4B (179,180). Among the five polymorphic residues present in the β -chain, only the Y328S variation is common in both C4A and C4B. Whether this polymorphism leads to the H and L phenotypes has not been established.

7. HETEROGENEITY OF C4A AND C4B ALLOTYPES

Allotyping of C4A and C4B based mainly on gross charge differences in electrophoretic mobilities and in hemolytic activity revealed 41 isoforms (39). Most of them are rare variants and require specific standards for unambiguous

definition. The most common allotypes C4A3 and C4B1, and some less-frequent allotypes such as C4B3 and C4B5, are not homogeneous as each of these allotypes may be split into two or more subtypes by serological typing and DNA sequencing (36,69,82,151,181). The extent of heterogeneity of each C4 allotype has not been systematically analyzed. At this stage only a relatively small number of human *C4A* and *C4B* genes have been sequenced. Elucidation of the structural and functional variations of all *C4A* and *C4B* alleles, and determination of the allelic frequencies in the normal and disease populations, and in different ethnic groups, would yield insights into the physiological significance of the C4A and C4B polymorphism.

E. Divergence and Polymorphisms of Mouse C4 and Slp Proteins

A comparison of three complete and one partial C4 sequences from *H-2* haplotypes b, d, w7, and FM (also designated as (d)), and two complete Slp sequences from the w7 and FM haplotypes revealed a total of 88 amino acid substitutions and a three-residue indel between the mouse C4 and Slp proteins. The mouse C4 and Slp share 95% sequence identities, in contrast to >99.5% identities between human C4A and C4B. Unlike the clustering of changes at the C4d region in human C4A and C4B, the sequence changes between mouse C4 and Slp scatter throughout the entire molecules except the N-terminal region of the β -chain and the C-terminal region of the γ -chain. Besides the specific substitutions leading to the variation in glycosylation sites mentioned earlier, there are remarkable changes that would diverge the functional activities of the C4 and Slp. Notably, substantial alterations including the EED 736–738 deletion and five substitutions are present in a 10-amino acid stretch in Slp that probably abrogate the classical pathway C1s-cleavage site (panel A, Fig. 7). Similarly, there are marked sequence alterations near the carboxyl terminus of the α -chain, particularly at the metalloprotease-cleavage site (panel B, Fig. 7). It is the proteolysis at this site that reduces the plasma form from the secretory form of C4 α -chain by 22 amino acids. This is of interest because it was observed that there is no discernible difference in the molecular weight of the α -chains between the secretory form and the plasma form of Slp (182). It is not clear if mouse C4 and Slp have a similar extent of tyrosine sulfation but the distribution of the tyrosine residues and their adjacent amino acid sequence pattern at positions 1399–1409 are quite different between the two paralogous molecules. In human C4, tyrosine sulfation enhances the C1s activity.

Other relevant changes between the mouse C4 and Slp molecules include the RT 943–944 QM substitutions where one of the factor I-cleavage sites in C4 is present, and the QG 985–986 RS substitution preceding the thioester

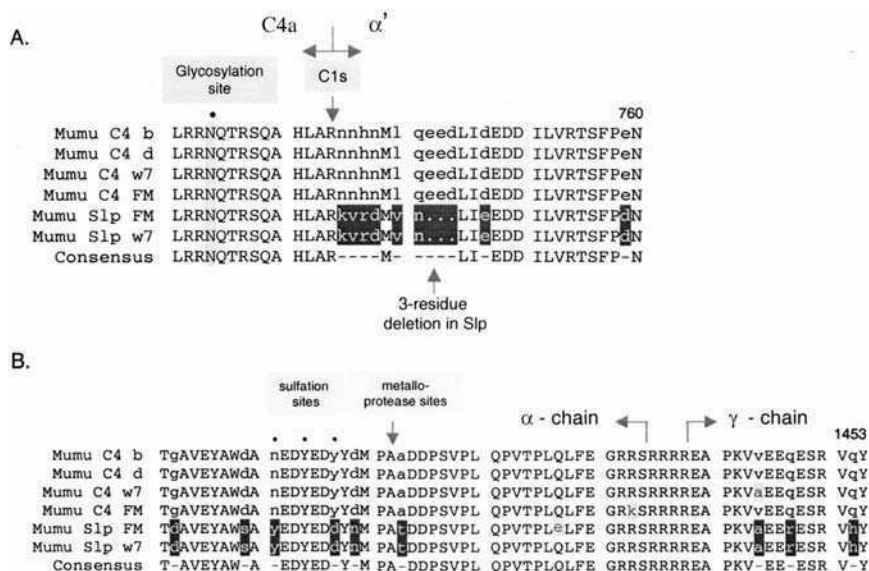


FIG. 7. Diversity and polymorphism of mouse C4 and Slp. An alignment of amino acid sequences around the C1s-cleavage site (panel A) and the α - γ chain junction (panel B). Four C4 sequences from *H-2* haplotypes *b*, *d*, *w7*, and *FM*, and two Slp sequences from *H-2* haplotypes *FM* and *w7* are shown. Conserved sequences are in upper case and variant sequences in lower case. Slp-specific sequences are highlighted in black. Allele-specific substitutions are highlighted in gray. The N-linked glycosylation site at residue 734 is not present in human C4. The eight substitutions and the three-residue deletion between positions 744 and 759 probably abrogate the C1s-cleavage site in Slp. The substitution of A1423 from C4 to T1423 in Slp probably eliminates the metalloprotease-cleavage site that removes a 26-residue peptide from the carboxyl terminus of the α -chain. The V1444A substitution might affect the processing of the α - γ chain junctions in C4^{w7} and in Slp^{w7} and Slp^{FM} (modified from Ref. (78)).

residues CAEQ at positions 987–990. It is not clear if these changes would affect the inactivation by factor I and the thioester activity of Slp.

Ten polymorphic sites are present among the C4 molecules from four mouse haplotypes *H-2^b*, *H-2^d*, *H-2^{w7}*, and *H-2^{FM}*. Four of these polymorphic sites are present on the β -chain. The α -chain and the γ -chain each have three polymorphic sites (Table III).

It was observed that mouse C4 from the *H-2^{w7}* has a large proportion of incomplete processing at the α - γ chain junction, yielding a 128-kDa polypeptide (183,184). There are only two unique amino acid substitutions in C4^{w7}. One is the N1305K that abrogates the glycosylation site. The other is the V1423A substitution that is six residues from the α - γ chain junction at positions 1425–1428 (Table III, Fig. 7). Interestingly, the V1423A substitution is also present in Slp^{w7} (and Slp^{FM}), and a partially processed 137-kDa

polypeptide representing the uncleaved α - γ chain was detected in Slp^{w7} (184). Therefore, it is possible that the V1423A substitution affects the processing of the α - γ chain junction.

Ten polymorphic sites are present between the Slp molecules from *H*-2^{w7} and *H*-2^{FM}. Eight of them are present in the β -chain. The remaining two sites are on the α -chain (Table III).

The mouse erythrocyte and plasma alloantigen C4d.1 from the *H*-2^k and *H*-2^s haplotypes are the result of polymorphism in the C4d region (185–187). The antigenic determinant for C4d.1 and its antithetical antigen C4d.2 was elucidated by sequencing (188) and confirmed by site-directed mutagenesis (189). The C4d.1 determinant is caused by the Q1187R substitution. The orthologous position of C4d.1 in human C4 (i.e., L1191R) is part of the Rg1/Ch1 antigenic determinant for the Rg and Ch blood groups.

F. The Phenotypes of C4-Deficient Mice

C4-deficient mice were generated using gene-targeting technology on mouse strain 129 embryonic stem cells with *H*-2^b genetic background that lacks Slp protein expression. The gene-targeting construct was generated by replacing a 987-bp genomic fragment of mouse *C4* exons 23–29, which contained the essential sequences coding for the thioester residues from exon 24 and the catalytic residues for thioester reactivity from exon 26, by PGK-*neo*. After immunization with a T-dependent antigen, C4-deficient mice had reduced number and size of germinal centers in splenic follicles, diminished primary and secondary antibody responses and failure in Ig class-switching from IgM to IgG (190). Compared with wild-type mice, C4-deficient mice had increased susceptibility to infections and the 50% lethal dose for group B streptococci infection in C4-deficient mice was reduced by 25-fold (191).

More recent studies of the C4 knockout mice offer support to the essential role of complement C4 in tolerance and autoimmunity (192). In the 129/B6 mixed background, C4-deficient mice had splenomegaly, impairment in clearance of immune complexes, and high levels of autoantibodies including antinuclear antibodies and anti-dsDNA with high penetrance in the female mice of 10 months of age (193). The anti-DNA autoreactivity was further shown to penetrate multiple genetic backgrounds including the heterozygous *C4*^{+/-} or “partial” deficient mice (194). Comparing with wild-type mice, the C4-deficient mice had markedly higher interstitial inflammation in the kidney as shown by increased infiltration of interstitial cells and foci of tubular atrophy (195). Similar to the cases in humans, only C4 and not C3 was protective against the lupus disease, as shown by experiments using single and double C4 and C3 knockout mice (196). Moreover, a bone marrow transplant

of *C4*-deficient mice with bone marrows from *C4* wild-type mice could restore the humoral immunity, as it was shown that the transplant mice produced IgG antibodies against immunogens (197).

Most of the observations in *C4* knockout mice are consistent with previous studies in *C4*-deficient guinea pigs (198). These include the roles of *C4* in bacteriocidal and opsonic activities (199), diminished secondary immune response (198,200,201), and elevated levels of autoantibodies including IgM rheumatic factor (202–204). The predisposition to autoimmune diseases, particularly to systemic lupus erythematosus (SLE), is similar to those described in complete *C4A*- and *C4B*-deficient patients, and homozygous or heterozygous *C4A*-deficient individuals (21,205,206).

IV. Nucleotide Polymorphisms in Human *C4A* and *C4B* Genes

To date, complete genomic DNA sequences for two long *C4A* from a bimodular RCCX structure (*C4A3a*, accession numbers: M59816–U07856–M59815; *C4A*, with HLA DR51, NT_034874), one long *C4B* from a monomodular structure (AF019413, allotype not assigned), one long *C4B* with HLA DR51 (NT_034874) (207), and two short *C4B* genes (AL049547; U24578, *C4B1*, from HLA haplotype *A1 B8 DR3*) from monomodular structures are available. Sequences for the long *C4A3a* were performed by manual method (50,82), others were achieved by automated sequencing (134,208,209). The exon–intron structures of a long *C4* gene and a short *C4* gene is shown in Fig. 7. These human *C4* sequences allow a detailed analysis of variations, both in introns and exons. An annotated sequence corresponding to a short *C4B* gene and its upstream region is shown in Fig. 8. Also included in the annotated sequences is the 5′-regulatory region of the *C4* gene in module I that is characterized by the presence of the *RPI* gene.

The sequence spanning from the initiation codon to the poly(A) site for a short *C4* gene is 14,205 bp, while that for a long *C4* gene is 20,574 bp. Analyses of human *C4* genomic cDNA and protein sequences accessible in the public domains or generated in our laboratory revealed the following features (Fig. 9, Table IV):

- (i) At the 5′-upstream regions from the module I of the RCCX, a trinucleotide deletion (CTC) is present in the two long *C4* genes at position –1463. Five single-nucleotide polymorphisms (SNPs) are present at positions –1938, –2291, –2425, –2799 (and –3048, not shown in Fig. 9). Whether the trinucleotide deletion and the upstream

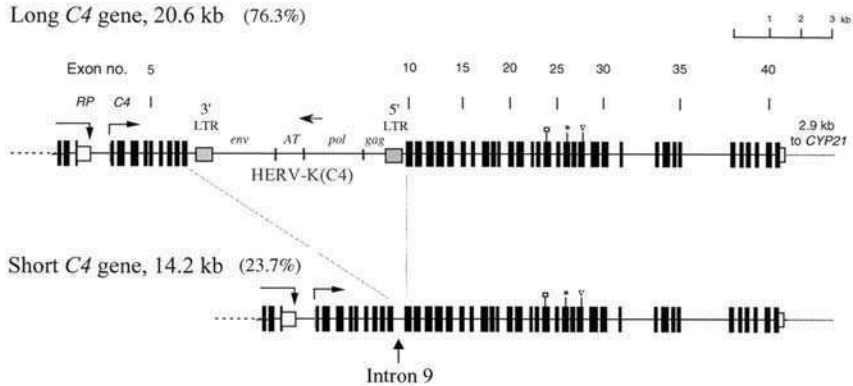


FIG. 8. Exon-intron structures of long and short C4 genes in human. The long and short C4 genes have frequencies of 76.3–77.4% and 22.6–23.7%, respectively, in Caucasian populations (11,12). The endogenous retrovirus HERV-K(C4) integrated into intron 9 of the long C4 genes (modified from Refs. (50,82)).

SNPs would affect the transcriptional efficiency of the C4 genes has not been determined.

- (ii) Among the exon sequences, there are eight synonymous (silent) substitutions and 22 nonsynonymous substitutions leading to polymorphisms at 20 amino acid residues. The D1106H responsible for the C4A/C4B isotypes, and the L1191R involved in the Rg1/Ch1 antigenic determinants each involves 2 nt changes in a codon. In addition to the four amino acid changes, which were detected by protein sequencing only, 24 polymorphic amino acid residues have been detected in the C4 proteins so far.
- (iii) The 6364-bp endogenous retrovirus HERV-K(C4) is integrated into the long C4 gene after nucleotide 2584, in a configuration opposite to the transcriptional orientation of C4. The endogenous retrovirus is flanked by a 6-bp direct repeat "cagaca" at both ends.
- (iv) Twenty-three SNPs and five single-nucleotide insertions or deletions (indels) are present in the intron sequences. Notably, the intron 9 nonretroviral sequences have seven SNPs, of which five appear to be specific for the short or long C4 genes (50). Three other introns appear to be hotspots of sequence mutations: intron 21 has three SNPs, intron 28 has one SNP and three indels, intron 30 has four SNPs and two indels.
- (v) In a 3.45-kb genomic region spanning exon 20 (nt 5781) to intron 30 (nt 9230), a great majority of the synonymous (5/8) and nonsynonymous substitutions (13/20), and indels in introns (5/5) are

FIG. 9. An annotated genomic DNA sequence of a human short *C4* gene from the first RCCX module. All published polymorphic nucleotide and amino acid sequences in C4A and C4B are shown. The A-nucleotide of the *C4* Met translation initiation codon is assigned number 1 for the DNA numbering system. The N-terminal residue of the *C4* mature protein is assigned number 1 for amino acid sequence (113). Amino acid sequences are in one-letter code and translated below genomic sequences. Sequences 3-kb upstream of the *C4* gene in RCCX module I is included to show the 5'-regulatory motifs and the breakpoint of the *RP1/RP2* gene duplication for the RCCX modules. Exon sequences are in upper case; intron sequences are in lower case. A polymorphic DNA sequence is italicized with the variant nucleotide listed and numbered above. A polymorphic amino acid sequence is highlighted in bold and the substituted sequence is listed below and numbered. A synonymous change in the coding sequence is shown with the same amino acid listed below and numbered. Relevant DNA sequence motifs are underlined and annotated. The endogenous retrovirus is present at nt 2585 in the long *C4* genes. The target site DNA sequence that is repeated in the long gene is underlined. An insertion of DNA sequence is marked by a "+" above the nucleotide where insertion sequence follows. A deletion in DNA is shown by the sequence underlined and a "-" above the nucleotide involved. The N-linked glycosylation sites, sulfation sites, β - α and α - γ chain junctions, and the thioester residues are underlined by solid line on the amino acid sequences. The regions where the crystal structure at the C4d region is underlined by dotted line. Demarcation of various protein domains or protease-cleavage sites is marked by vertical strokes below amino acid sequence. The data are mainly derived from the following GenBank entries: long *C4A* genes: M59816-U07856-M59815, NT_034874; long *C4B* genes: AF019413, NT_034874; short *C4B* genes: AL049547, U24578; short intron 9: U07851-U07855 (modified from Ref. (82)). (Continued).

-1000	TGCATGA7GGTGGCTGAGACAGGCTCACCTTGGGAAGGCTTGGGAGCCAGGATGAGTGTGGGCTCTCGTGTGTGCAAAAGTGCAGATGTGACTGCTGCT	-901
-900	GTTTGCCTGGTTTCTGACCCAGTGGTGGGTTTGAGCAATGCTTCTCTGCCCTTCCATGGAAATGAGAACAGAAATGTGTCACAGCTGTGCTGCTTTC	-801
-800	CTTTGCTGTAAATGGTGTGTTTATTACTCTGTCTGAAATAGGAAGTGGGATTTCTGGGAGGCTGTGTGAAGGAGGCGAGGCTCTTTTCTCTACGTG	-701
-700	TCATGTTAAATTGC ^{poly(A) signal} ATAA GTACTCTGCTGTGATattttctggatgtcctttatttactgtgacgtgtgtttgggtgcttgtttagggtaga	-601
-600	ggtagagctgagctttgctcctattcagagaggaaaggggtcaggggttcaactctgacgttcaggccattctccctgtggagtggtgaggggtgacctaa	-501
-500	tctcctaaccacgggaattttctgttagggctcaaaaaagcaaaagcctagatatagttcaatttctgttggaaatgaagtagagacaaggtctcttagaag	-401
-400	cctgtcattgtttttaggggctttaaatactcgtactcgtgggccaatgttggggccctgtacgcccaggtatatacatgagcttgtgtgacactatacc	-301
-300	ctgatacagatatacctgttagggggaggtgtctcaggcactggaatgagaggagtttaacgggggaaggacaggggtatttctgtgggccaagattcagagctt	-201
-200	cccatggg ^{CACCC Box} gagc gggtgtccgggggtgcccccacaactct ^{NF1} gagcctcagagccag ttgtgcacttcttggtctg ^{E-Box Enhancer Core, SPI} cagcgtggttcc ccagctctagcagagctag	-101
-100	gggaggagcaaggctccagagctcaactctgcccgcaggcctagcttggcc ^{* Exon 1 (complement C4)} AGAAGGTAGCAGACAGACAGACAGCATTAACCTCTCTTGGATCCTCCAGCC	-1
1	ATGAGGCTGCTCTGGGGCTGATCTGGGCTACAGCTTCTTCACTTATCTCTGAGAAACCCAGTctctggaggaggatgctgggtgcttggattggg	100
(-19)	M R L L W G L I W A S S F F T L S L Q K P R (3) C4 Leader peptide C4 N-terminus	
101	gcagggtcgtgattcgggaccgattcaggagtgaggagagcaggggtggaggtgtcagagcgagctctgactgctgactgtctgttctccccagGTT	Exon 200
	(4) L	
201	GCTCTTGTCTCTCCTCTCTGTGTTTCATCTGGGGTCCCCCTATCGGTGGGGGTGCAGCTCCAGGATGTGCCCCAGGACAGGTAGTGAAGGATCAGTG	300
	L L F S P S V V H L G V P L S V G V Q L Q D V P R G Q V V K G S V	37
301	TTCTGAGAAACCCATCTGTGAATAATGTCCCTGCTCCCAAGGTGAGACTCACCTTAGCTCAGAAAGACATCTGCACTCTCAGTCTCCAGGtaa	400
	F L R N P S R N N V P C S P K V D F T L S S E R D F A L L S L Q (69) T378 F63	
401	ccagaccatgcccctctgctgcttggggggcctcctgcccctgttcccactgtcttgaagtgtcatcatctcccactggcctctcccctcctcgt	500
501	cttcccacccctggcattctctcccaagcttctcccttgggtctctgctcttcttgggtcagctgtctctgtctgtgaccccgctccctctcccctcctct	600
601	ctcctgacagGTGCCCTTGAAGATGCGAAGAGCTGTGGCTCCATCACTCTCAGAGGCCCTGAGTGTCCAGCTGGTGCCCACTTGGCCATGGCTAAAG	700
	(70) V P L K D A K S C G L H Q L L R G P E V Q L V A H S F W L K	99
701	GACTCTCTGTCCAGAACGACAAACATCCAGGGTATCAACCTGCTCTCTCTCTCGCCGGGGGCTCTTTTTCAGAGGGACAGCCCATTTTACAC	800
	D S L S R T T N T I Q G I N L L F S S R R G H L F L Q C T G Q C C Q P I Y N P	133
801	CTGGCCAGGGGGtgagctctcagcccaggcctcaacctttaacccccccagagccctctcaggatgagtttgggtgcccctcaagtgaagataacctgaa	900
	G Q R V (137)	
901	agaaagtgccacacagaggggtgtctaggaaaattgtcccctgctcccctctgtggagtttgaccccacccctccccttgacataggaccccctgctcacc	1000
1001	tctctcctctcccacccagTTCCGTACCGGGTCTTCTCTCTGGATCAGAAGATGCGCCGAGCACTGACACCATCACAGTCACTGGTGGAdgtgagctcc	1100
	(138) R Y R V F A L D Q K M R P S T D T I T V M V E (160)	
1101	ccgaactctggcctctcctgactctggccactgagtgacctcctgctgtgagcaactctctccccttgagAACCTCTCAGGGCTTCCCGTCCGGAAGAG	1200
	(161) N S H G L R V R K K	170
1201	GAGGTGTACATGCCCTGTCCATCTCTCAGGATGACTTTGTGATCCAGACATCTCAGAGtgagcctcccattgtgggggctgccccacagctacacca	1300
	E V Y M P S S I F Q D D F V I P D I S E (190)	
1301	cccccaattctgtaggctctccacactccacacagagggaagcagctcccagatgcctgacccctcagcctctctgagcctctgggttaacccccacagctctc	1400
1401	ttcccagggaagcaggctgctggtcctccgtgccccactgtacagatggggtgagccccctctctgtccattctcagCCAGGACCTGGAAGATCTCAG	1500
	(191) P G T W K I S A	198
1501	CCCGATTCTCAGATGCCCTGGAATCCAGACAGCACCCAGTTTGAAGTGAAGAAATATgtgagagctggaaactggaggagcaggcagctgcttctct	1600
	R F S D G L E S N S S T Q F E V K K Y V (218) N207 (Glycosylation-high mannose type)	
1601	gaaggaaataagggtggaaggagagactgaggagcagctcagggcaggagagatatgggtgcccacagccctgacagggaggagctctttagctggagtc	1700
1701	tgacctgcatatccctctcaacccgggtcagTCTTCCCACTTTGAGGTGAAGATCACCCCTGGAAAGCCCTACATCTCAGCGTCCAGGCCCATCTTGA	1800
	(219) L P N F E V K I T P G K P Y L I T V P G H L D	241
1801	TGAATGCACTTACACATTCAGGCCAGgtataactcctcccacactctgccccagacaggggctcctgctccctactcagtagaaggggtcctctgc	1900
	E M Q L D I Q A R (250)	
1901	TTCCCTGccctcgggcaattattcccccgaccccttggggccacccctctctctctgcccacagGTACATCTATGGGAAGCCAGTGCAGGGGGTGGCAT	2000
	(251) Y I Y G K P V Q G V A Y	262

Fig. 9. Continued.

2001	ATGTGCGCTTTGGGCTCCTAGATGAGGATGGTAAGAAGACTTTCTTCGGGGGCTGGAGAGTCAGACCAAGGtaggaaggagaataggggctggggaggg	2100
	V R F G L L D E D G K K T F F R G L E S Q T K (285)	
2101	GAAAGGGcaaggagggtgagggtgggaagactcagctctccaccatgtctcgttttcttatgcccacagCTGGTGAATGACAGAGCCACATTTCCTCCTC	2200
	(286) L V N G Q S H I S L S	296
2201	AAAGGCAAGCTTCCAGACGCCCTGGAGAGCTGAATATGGCATTACTGACCTCCAGGGGCTGCCTCTACGTTCTCTGACCCATCATTTGAGTCTCCA	2300
	K A E F Q D A L E K L N M G I T D L Q G L R L Y V A A A I I E S P	329
2301	Gtgggtagctttcccttattgttaaccagacccttgccctctgacctctgagctaacctctgtctccggcaccacaccacccactttctcaactct	2400
	G (330)	
2401	catctcagactcaaaaccaggaaacaccaggagactgggtttctctccaaactctgtctctgactggccctttctccgtgagtttatttatttc	2500
2501	tttgcctgtctgtcattctcttccactccctccagtgagcatgtgtgttcaatgcccctgtgctgggcctcagcatgtacagacgaataacagcagcctg	2600
	c2558 a2564 c2577 g2592	
	HERV-K(C4) integration site for long C4 gene	
2601	c2601 c2651	2700
	cgccctggagactcactgtctgtgggggggaaccactcaagccactccctactgtctcctgtctccctctctttctgggctctgtccccacacctctctc	
2701	tgccctgtgtcttgaggtgggagatgagagggcagagctcACATCTGGTATTTTGTGTATCTCCCTCTCCTTGGATCTTAGCAAGCAACGACGA	2800
	(331) G E M E E A E L T S W Y F V S S P F S L D L S K T K R	357
2801	CACCTTGTGCTGGGGCCCCCTTCTGCTGCAGGttttctccaggggggaaggatgagtagggaggatgtggtagttaggggggctcaggggtctgacac	2900
	H L V P G A P F L L Q (368)	
2901	tctcttttgcctgccctcttttaacctgactagGCTTGGTCCGTGAGATGTTCAGGCTCCACAGCTCTGCGATTCTGCTGAAAGTTTCTGCCACGGTGTG	3000
	(369) A L V R E M S G S P A S G I P V K V S A T V S	391
3001	TTCTCTGGGTCTGTTCCTGAAGTCCAGGACATTTCAGCAAAACACAGACGGGAGCGGCCAAGTCAGCATTCCTCAATAATTATCCCTCAGACCATCTCAGAG	3100
	S P G S V P E V Q D I Q Q N T D G S G Q V S I P I I I P Q T I S E	424
3101	CTGCAGCTCTCagtaggactcctcggaccctcgggagatggtggggggaaggaggagggtgagctggggtcccaaggatccatggcctgacttgggggg	3200
	L Q L S (428)	
3201	AAGGTGGtagcttctgtctgagctactaccctattctgacactgacccctctccagGTATCTGCGAGCTCCCCACATCCAGCGATAGCCAGGCTCATG	3300
	(429) V S A G S P H P A I A C C R L T V	443
3301	TGGAGCCCCACCTTCAGCAAGCCGCCGTTTCTGCTATTCAGGCGCCGATTTCTCGACCTCTCTGCTGTGGGACATCTGGAACCTGAACCTGGAGAC	3400
	A A P P S G G P F L S I E R P D S R P P R V G D T L N L N L R A	476
3401	CTGGGCACTGGGCCACCTTTCTCATTACTACTAGTgtgtgcatgagctggggaggtcacggagggctgggggtgaggggaagagccctctgggtggg	3500
	V G S G A T F S H Y Y N (489)	
3501	GCTGGGGgggttcaaggctgaggctgtcccaatgaagaggcaaccactctgtgcctccctattcttggcccagATCCTATCCCGAGGCAGATCGTGTTC	3600
	(490) I L S R G Q I V F M	499
3601	TGAATCGAGAGCCCAAGAGGACCTCGACCTCGTCTCGTGTCTTGTGGACCATCACCTGGCACCTCTCTACTTTGTGGCTTTCTACTACCATGGAGA	3700
	N R E P K R T L T L S V S V F V D H H L A P S F Y F V A F Y Y H G D	532
3701	CCACCCAGTGGCCAACTCCCTCCGAGTGGATGTCCAGGCTGGGGCTCGGAGGGCAAGgtgagccgggttcaggaaagatggcacttgtgcgaggggggt	3800
	H P V A N S L R V D V Q A G A C E G K (551)	
3801	gaggacagggtgattgccaaacaggcatggatttagcttgggggcagtaggataccgggactgaaggaagctctccactctgaccgcccccacactgcc	3900
3901	gccccctgcccagCTGGAGCTCAGCCTGGAGCGTCCCAAGCAGTACCAGAAACGGGGAGTCCGTGAAGCTCCCTTAGAAGCCGACTCCCTAGCCCTGGTGGC	4000
	(552) L E L S V D G A K Q Y R N G E S V K L H L E T D S L A L V A	581
4001	GCTGGAGGCTTGGACACAGCTCTGTATCTGTCAGGCGCAAGTCCCAAGACCCCTCAACATGGGCAAGgtttgtccagaccctctccacagctctctcc	4100
	L G A L D T A L Y A A G S K S H K F L N M G K (604)	
4101	accctccatggctcactccctgtctccctgagccttggggcagccctggatccactgaggtctccacagctctctccacttggccctgtggt	4200
4201	ctccatctctggtctgtatcctttctatccccctatgtgtgcctctccactgtgccagtgctcagctgcccctcagccacacttggctctta	4300
4301	gcattctgtccttcttctgagCTTTTGAAGCTATGAACAGCTATGACCTGGCTGTGGTCTCTGGGGTGGGACAGTCCCTCTCAGGTGTTCAGGACG	4400
	(605) V F E A M N S Y D L G C G F G G G D S A L Q V F Q A A	631
4401	COGGCTGGCCTTTTCTGATGGAGACCACTGACCTTATCCAGAAAGgtgagacagagaaggaggagggtgggtggcgggaagataaaggaggagga	4500
	G L A F S D G D Q W T L S R K R (647)	

Fig. 9. Continued.

[illegible]

Fig. 9. Continued.

7001	7022T CACAGAGCAGTGGAGACACACTGCTCCCGAGACCAAGGACACAGCCCTGGATCTGATCCAGAAAGttctgggtgcagggaagcagcagggggggccag T E Q W S T L P P E T K D H A V D L I Q K G (1033) L1018	7100
7101	gaaaggacagttacttgaagatggacagcccgagggtcacagagggaagaaagggggccctgatgagatggggagcatggccttgggtcacaacag L1018	7200
7201	Exon 25 cagaagggtgagtgtaoactgagcgggcaacctctctctcccaagGCTACATGCGGATCCAGCAGTTTCCGAAGCGGATGGTTCCTATGCGGCTTGTTG (1034) Y M R I Q Q P R K A P G S Y A A W L V1049	7300 1051
7301	A7308 T7318 TCACGGGACACAGCAGCTGtgtagcttgggagagtggttccaggggtctctgaggggtcaggggtggggcagagctggtatgatgggag S R G S S T W (1058) (Ch5) P1054 T1057	7400
7401	ggtggataaccaggccacctgggggtggggcataatgagaagcaagtccttctcccaacctctctctctccctccagGCTCAGACCCCTTGTTGTA (1059) L T A F V L K Exon 26	7500 1065
7501	A7335 AGSTCTGAGTTTGCCCGCAGGAGCAGTAGGAGGCTGCTGAGAAACTGCAGGAGACATCTAACTGGCTTCTGTCGCCAGCAGCAGCTGACGCGCTGTT V L S L A Q E Q V G G S P E K L Q E T S N W L L S Q Q Q A P G S F G1076 I1090 (protein sequencing)	7600 1098
7601	7699C G T Q C7625 CCAGGACCTTCTCCAGTATACATGAGGACATGCAggtggggcatgctggggctggggcagagaagcagctctggaggactctcttggccctcccc Q D L S P V I H R S M Q (1110) 1101P C L D1106 (C4A/C4B isotypic residues)	7700
7701	Exon 27 ctcctgttgacatctttttctcccttactagGGGGTTTGTGGGCAATGATGAGACTGTGGCACTCAGACGCTTTGTGACCATGCCCTTCATCATGG (1111) G S L V G N D E T V A L T A R F V T I A P H H G L1111	7800 1133
7801	GCTGGCGCTCTCCAGGATGAAGGTCGACAGGCCATTGAAGCAGAGAGTGtaagttcagtggggtttctgctcctctgctggccccccagctctctccctt L A V F Q D E G A E F L K Q R Y (1149)	7900
7901	Exon 28 7977A ttctcaggaaccagggttcaggcccccaagacctctccctgtttctccagGAGCCCTCCATCTCAAGGCAAGCTCATTTTTGGGGGAGAAAGCAA (1150) E A S I S K A S S F L G R K A S N1157 (Rg3*/Ch6)	8000 1165
8001	8052T 8065G 78070 TC8080 GTGCTGGCTCTCTGGGTGCCACGAGCTGCCATCAGCGGCTATGCCCTGACACTGACCAAGGCCCTGCGACCTGCGGGGTGTGTCGCCACAAACCT A G L L G A H A A A I T A Y A L T L T K A P A D L R G V A H N N L 1182S 1186A V1188 L1191 (Rgl/Ch1 determinant)	8100 1198
8101	8140c +c8144 -c8172 -c8182 CATGGCAATGGCCAGGAGCTGAGGtgtaggggtgaggggtctctggcagtagagctgagggccaggggacgttaggactccgtgagtggtccccagggga M A M A Q E T G D (1207)	8200
8201	gaggctggatgaagactcagaaggaaggaatgaagttataagcaggggtgagttggggagagactcagagagccagcaggggggtggctgaagggccagggga L1207	8300
8301	Exon 29 8372+TC C4A20, BQ ccaggctctctccctgcttccctgttactctgttggttcccttcaacttcagATAACTGTACTGGGCTCAGTCACTGGTTTCCAGAGCAATGCCGTG (1208) N L Y W G S V T G S Q S N A V C3b binding site	8400 1222
8401	8412A TCGCCACCCCGCTCTCTCCCAACCCATCCAGCCCATGCCCCAGGCCCGCCCGCTGTGGATTGAACCAACCACTACCGCTGCTGCACTCTCTGTTC S P T P A P R N P S D P M P Q A P A L W I E T T A Y A L H L L L L H P1226	8500 1256
8501	T8533 ACGAGGGCAAGCAGAGATGGACAGCGCTGCGGCTGCTCACCCCTCAGGCGAGCTTCCAAAGGGGATTCCGAGTACCCAAgtagggcgctccc E G K A E M A D Q A A A W L T R Q G S F Q S G F R S T G (1284) S1267 V1281 (protein sequencing)	8600
8601	g8611 cgggctctggcgggggtggtagtctcagacaggaaggggtggttgagtcctgagctcaacctccctagGACACGGTGATTGCCCTGGATGCCCTGCTGTC (1285) D T V I A L D A L S A 1286G G1287 (protein sequencing)	8700 1295
8701	8708T CTACTGAGTTGCTCCACACCACTGAGGAGAGGGGTCTCAATGTGACTCTCAGCTCCACAGCGCGGAATGGGTTCAAGTCCACCGCTGCACTGTAAC Y W I A S H T T E R G L N V T L S S T G R N G F K S H A L Q L N P1298 (glycosylation-BAC) N1299 C4d ends <- factor I cleavage site	8800 1328
8801	8898a AAGCCGAGATTGCGGCTGAGAGGAGCTGCAAGtgtaaccactccctggtagaaccactccctcgctgggtagccaggacacctgggctcgtggcc N R Q I R G L E E E L Q (1340)	8900
8901	a8920 aggccagaagcgtccccacacctccccacgtggaatcccccgagcactcttctctgggtctctcggggaagagactgactctctggctcgtgacactgga L1340	9000
9001	gctctgagcttcagttttctcaacttgtagagtaacatacacagaggtcaccctacagggctgcttagaaggctgaagtagagataattcctgtgtggtata L1340	9100
9101	g9131 +c9139 c9164 aaccttgggaagtggaggtggggagaggggtggtcttggaggagagataagttatggagccgcaaaaaacaggtttgtctgtgccccttcta L1340	9200
9201	+c 9219 Exon 31 acatcgcttctctctctgtgctgaagttttctctGGGAGCAGATCAATGTGAAGTGGAGGAAACAGCAAGGAACCTCTGAAGtgtagggcgag (1341) F S L G S K I N V K V G G N S K G T L K (1360)	9300

Fig. 9. Continued.

9301	ggaaggggtggggccaggacactgtggaggagaggggtgaggagagggcctgtgggacagagacacatggtccgggaaggaggacacacccagggt	9400
9401	tgggtcccgctgcttcgctcctgggtgtttttccccctgcttgccttgcgtctcccccatctctgggaacctgttgtttcctttaoccgctcagtg	9500
9501	ctgggtggctccgaatccccactctcagccaggcctcttccctgaaccatgggcccactctgcccactccccacagcactcagacagggcatgtcccaa	9600
9601	agccctctctcatctctgtctctctgtctgggtggagccctctccagccaggagcccgacactcttagaggccgtgttagtgccctctctcc	9700
9701	aagcctgtctctatgtctcctagtgaactcctctctctctgctccctctgctgctgtggcccttggtgctgcatcctagatctctgtctgagaggcctctctcc	9800
9801	tacctggaaactctctctcactcctgtctccctctgtctgatccactgtccacacggcagtgacactgaaccttccaaaagcccccagagatcagccttgg	9900
9901	ggaaaagactcactccccgctgccccaggctcagatggctgggctctctgcccacccctccggccagacagctctctctgtctacacagatcccccttgccttt	10000
10001	cctgtcctctcctgctctcttggccacaggacaagctctttctctcctccagccttggccagaaagccttctcagcttttcagctccagcctctctcc	10100
10101	agcacagctctggagtgctggcctctgggggcaggccctctctctttacctctctgtctcgcctgacgcctgtggcgaatgtgggtcccaactcgtctgtgt	10200
10201	ggactgtgcagtgacggggaggaaaagggctgaaggctcaaatcctgtagcccaggagagatgcccttaggtatggcacccagagaggtctgtggcctca	10300
10301	catgtccccagctcctctcctgccccttgctgagccaggtCCTTCTGTACCTACAATGTCTTGACATGAAGAACACGACTGCCAGGACCTACAGATAGA	10400
10401	AGTGACAGTCAAAAGCCATCGAGTACAGAGTACAGTgagtggtgggggttggggagccttggggccagggcaggggctggcgccaggagcgggtggccatccc	10500
10501	agccctctcacaatgtctcctctgctgagcagTGGAAAGCAACGAGGACTATGAGGACTATGATGATGATGCTCCAGCCAAAGTATGACCCAGATGCCCT	10600
10601	CTGCAGCCCTGACACCCCTGCACTGTTTGGAGGTCGGAGGAACCGCCGAGGAGGAGCGCCCAAGCTGTGGAGGAGGACGAGTCCAGAGTGCACCT	10700
10701	ACACCGTGTGATCTGTggggccggggagctgcccgtgggcccaggggaggaggaggaggacccagcggggctggggtctgtggagcccgccagggcaga	10800
10801	acctggagcagcagctcacaagctctccacagGCGAAAGCGCAAGGTGGGGCTGTCTGGCATGGCCATCGCGAGCTACCTCTCTGATGGATTCCAGCC	10900
10901	CTGCTGCTGACCTCGAAGAGTGTGTgctcagcaccacggggcaccctcctctgctccaggtactgagccctgtcatgtgcaggccctgtgaccaaactccc	11000
11001	cttttccacagCTGACCTCCCTCTCTGACCGTGTACGTGAGTCACTTTGAGACCGAGGGGCCCCAGCTGCTGTGATTTTGACTCGtgagtgaggggagag	11100
11101	atgaggccaggaaaggactcagatggcacgggtttactgagtatgcgttaggaaggtttctccaggagacagctgtgtcagcggctggctcctcttgagaactc	11200
11201	gtgatgtcatcagagagaagaagaatgtgagcccggtgagacacagacagagtaaggggcagacactcagccaggccaggagacagatgcccagtcagcaggg	11300
11301	accctcagggtttgagagggaagtcttctcctaagtctggttttattcagcttgaggggctgctcttcttttggtagaactcctactctttttttaata	11400
11401	ttaaagcgtattttccttacaagtgatggtggccatagatgatagttgatttgccttttccagcacttatttggctaaaaatagttatcaacctctt	11500
11501	acggctctcaaaacatttttattttatttttagtaaaagacagggtctcgtctctgttgcacagctggcttgaactcccggcctcaagcgtactctctgg	11600
11601	cctaggcctttcaagttacgggatttaccagccagagccaccatgcccggccttcaaaaaaagttttggaacatttactgtaacctctgggagaaaatgt	11700
11701	gagaaggtgtgggtgctgctcattagccagctgttttaggtcaggggagacccctaccagtggtgtgcagagggggccagccccctcagctggggaaagcc	11800
11801	tgctgcacacatctgggttgaacacacatagaaaacacagagccaacagatctccggataggggagctgacgggtgcagcgcctgactcagagagggaacat	11900
11901	ggcacggccagctgtggactgggcccgggtggggccagagagggtcaggcctgggacctgagtcgggggtcaggcagagatgacagaaacctgcagttag	12000
12001	gttgctggcaaaataaaggaggaccagtgatccatcagacaaaagatgaggcccgaggaggggcagtgaggtttgggggcaggcagtgagctcttggaagact	12100
12101	tacaggtcctgcacaaactcctaagtgcagagatggagctgcaagttcagtttgggaatcatcagcctggatgggttgggtggaagccaggaggtggtgtgag	12200
12201	acccccacaggaggagctctcaggaaaggaagttccgaaggagggaacgttaagaatgaccaggtcagaaaccaaagggtggttcagaaagctaaaccttagctt	12300
12301	aggggacagttttcacagagaacacgtccatgatgcaagactctgtcaggggctcggagcagtgaaagctggggcaaggtcaccctctgggaagtgaagtca	12400
12401	ccagagaccttcgggagcagcttttagaggtctctctgagtaggaaggttaacagaatgtgaaggacactggagagaaggccaatagggaagaacaaacaaaca	12500
12501	ggccaaagaaacccagtcaggggggtgcaggggccagggagtgaggtccctctcatctcctccccacgcttgccagagTCCCCACTCTCCCGAGTGGCT	12600
12601	GGCCTTTGAGGCTGTGCAAGAAAGTGCOCGTGGGCTGTGTCAGCCGCCAGCGCAACCTGTATGACTACTACAACCCCGtgagcagctcaggacaccc	12700
12701	tgaattccaggagaactctggcataggtgcccctcctatgggacaaatggacacccgggtagtgagggggacagagaccccggggtcctctgggactgagga	12800

Fig. 9. Continued.

Exon 37	
12801 ggcagaatggaggggctgtgccctaactctctctgttctctcagacgacgagatgttctgtctttttacgggocacccaagtangagcagactcttggccac	12900
(1542) R R C S V F Y G A P S K S R L L A T	1559
Exon 38	
12901 CTTGTGTTCTGCTGAAGTCTCCAGTGTGCTGAGGGtgagactgagggcctggggcggggcagtgaggcgggagtgccggggccccccacacactgtct	13000
L C S A E V C Q C A E G (1571)	
Exon 38	
13001 gatgggttcccccaacttcagGGAAGTGCCTCGCCAGCGTGGCGCCCTGGAGCGGGGTCTGCAAGACGAGGATGGCTACAGGATGAAGTTTGCCTGCTAC	13100
(1572) K C P R Q R R A L E R G L Q D E D G Y R M K F A C Y	1599
c13169	
13101 TACCCCGTGTGAGTACGtgagctcttccccaccaggccctggcctgacccctccctcggggacgggctgtgttttggtctctctgggtgtagcctgtcct	13200
Y P R V E Y G (1604)	
13201 cttacaggtoatgcaacgagcgtctgttctgtctgacaccaaacttctacccctctcagcctcaaaagtaactcacctttccccccttctctccacccccctctta	13300
Exon 39	
13301 gGCTTCCAGGTTAAGGTTCTCCGAGAAACAGCAGAGCTGCTTTCCGCCTCTTTGAGACCAAGATCACCCAAAGTCTTGCACTTCagtatgaagcaaacg	13400
F Q V K V L R E D S R A A F R L F E T K I T Q V L H F T (1632)	
13401 gagagggcgggcagggctggggggagacagggaggctgagggtgtggccgaggacctgacctctctggaagtgtgaaatcccccttgggctgtcagaagcctt	13500
13501 gggcttggccataaatgaggaggcagtgccacctctccatgggggtggcgaaggtgggaatgagaggatctacacagagctccccagcctgggctcaccctg	13600
Exon 40	
13601 caccctctcttcccccttgaccacettttggcgacgtatcccccgagCCAAGGATGTCAAGGCCGCTGTCTAATCAGATGGCACTTCCTGGTTGAGCC	13700
(1633) K D V K A A A N Q M R N F L V R A	1649
13701 TCTGCGCGCTTGTGCTTGGAACTCGGGAAGAATATTGATCACTGGGTCTGGATGGGGCCACCTATGACTCGAGGGACgtgagtcactctgtccccctc	13800
S C R L R L E B F K E Y L I M G L D G A T Y D L E G H T (1676)	
13801 agtctcttctctccccatgcctcgccacctaggccttgccctcagaaagccagatgacctgtgtctctcgtttccacctgccatctctcccgagccctgct	13900
Exon 41	
13901 gactgccccctttgccccctgcagCCCCCAGTACTGTGACTCGAATAGCTGGATCGAGGAGATGCCCTCTGAACGCCTGTGCGGAGCACCCGCCAGC	14000
(1677) P Q Y L L D S N S W I E E M P S L E R C R S T R Q R	1702
14001 GGCAGCCTGTGCCAGCTCAACGACTTCTCTCAGAGATATGGCACTCAGGGGTGCCAGGTGTGAGGGCTGCCCTCCCACTCCGCTGGAGGAACCTGA	14100
A A C A Q L N D F L Q E Y G T Q G C Q V * (1722)	
poly(A) signal. * poly(A) site (C4B)	
14101 ACCTGGGAACCATGAAGCTGGAAGCACTGCTGTGTGTCGGTTTCATGAACACAGCGCTGGGACAGGGCATATTAAAGGCTTTTGGCAGCAAGTGTCACTG	14200
* poly(A) site (C4A)	
14201 TTGGCAGcgaaggtcagtcgtgtgtgtgagggctgagagcagtgccccctgcccgatgcagttctgggcagggcaggttgacataacccttagactctctg	14300

Fig. 9. Continued.

found. There are also nine intronic SNPs. The high GC content of this region, 67.6%, would favor exchange of genetic information by gene conversion-like events.

- (vi) Three different nonsense mutations leading to nonexpression of *C4A* or *C4B* genes have been found so far. These include a deletion at nucleotide C3671 for amino acid residue F522 in exon 13 of the *C4BQ0* gene from HLA A2 *B12 DR6* (210), a nucleotide deletion at nucleotide C5808 for amino acid residue F811 in exon 20 of the *C4AQ0* gene from HLA A30 *B18 DR3* (211), and a 2-bp TC insertion after nucleotide 8372 of amino acid residue S1213 in the *C4AQ0* genes of HLA haplotypes A2 *B12 DR6* (210), A2 *Cw7 B39 DR15* (212), A2 *Cw3 B40 DR6* (211,212), and A2 *Cw3 B60 DR6* (213), and the *C4BQ0* gene of HLA A2 *Cw7 B39 DR15* (212) (Fig. 8).

TABLE IV
A LIST OF NUCLEOTIDE AND AMINO ACID POLYMORPHISMS OR MUTATIONS IN HUMAN
COMPLEMENT C4

Location*	nt		Location*	nt	
	polymorphism/ mutation	a.a. substitution/ <i>remarks</i>		polymorphism/ mutation	a.a. substitution/ <i>remarks</i>
RP1, I-3	g -2799 a	<i>RP1/RP2 breakpoint</i>	E-24	G 7022 T	L 1018
	I-4 g -2425 t		E-25	G 7292 T	A 1049 V
	I-4 g -2291 c		E-25	G 7308 A	G 1054 D
	I-4 t -1938 g		E-25	C 7318 T	T 1057
	E5 -1572		E-26	C 7335 A	G 1076
	I-5 c -1465 del		E-26	nd	S 1090 I
	I-5 t -1464 del		E-26	T 7609 C	L 1101 P
	I-5 c -1463 del		E-26	C 7612 G	S 1102 C
C4, I-1	g 116 a	F 63	E-28	G 7977 A	S 1157 N
	E-2 C 378 T		E-28	C 8052 T	T 1182 S
	E-9 T 2296 A		E-28	C 8065 G	A 1186
	I-9 g 2371 a	<i>ins HERV-K(C4), 6360-bp + 6-bp TSR</i>	E-28	C 8070 T	A 1188 V
	I-9 t 2558 c		E-28	G 8079 T	R 1191 L
	I-9 G 2564 a		E-28	G 8080 C	R 1191 L
	I-9 t 2577 c		I-28	g 8140 c	
	I-9 2585		I-28	c 8144 ins	
			I-28	c 8172 del	
	I-9 a 2592 g		I-28	c 8182 del	
	I-9 C 2601 t		E-29	TC 8372 ins	<i>C4AQ0, C4BQ0</i>
	I-9 t 2651 c		E-29	G 8412 A	P 1226
	E-11 T 3024 C	V 399 A	E-29	G 8533 T	A 1267 S
	I-11 c 3119 t		E-29	nd	R 1281 V
	I-11 a 3127 c		E-30	nd	T 1286 G
	E-12 C 3345 T	R 458 W	E-30	nd	V 1287 G
	E-12 C 3349 T	P 459 L	E-30	A 8708 T	I 1298 F
	I-13 c 3463 a		I 30	g 8898 a	
	E 13 C 3671 del	<i>C4BQ0</i>	I-30	c 8920 a	
	E 15 nd	C 616 S	I-30	a 9131 g	
	E 17 C 5056 T	P 707 L	I-30	c 9139 ins	
	E-20 T 5796 C	V 806	I-30	t 9164c	
	E-20 C 5808 del	<i>C4AQ0</i>	I-30	c 9219 ins	
	E-21 G 6077 T	G 863	I-31	t 10248 c	
	E-21 A 6150 G	T 888 A	E-33	10550-10559 del	EDY 1399a, b, c
	I-21 c 6272 t		E-34	G 10910 T	D1478Y
	I-21 a 6331 g		I-35	t 12028 g	
	I-21 t 6419 c		I-38	t 13169 c	
	I-23 c 6775 t				

a.a., amino acid; del, deletion; ins, insertion; nd, not determined; nt, nucleotide; TSR, target site repeat.
*E-, exon; I-, intron.

V. Expression of C4A and C4B Transcripts and Proteins

A. Human Complement C4 Transcripts are Present in Multiple Tissues

The major site of C4 expression is the liver. However, C4 transcripts/proteins are also expressed in peripheral blood monocytes, skin fibroblasts, and epithelial cells of the intestine and the kidney (reviewed in Refs. (214,215)). The extrahepatic sites of C4 expression may be important for the local protection and inflammatory response (216). To illustrate the expression of human C4 in extrahepatic sites, a northern-blot analysis was performed using human multiple tissue poly(A)⁺ RNA blots and a human *C4d*-specific probe. As shown in Fig. 10, the highest level of C4 transcripts is detectable in the liver. Moderate quantities of C4 transcripts are present in the adrenal cortex, adrenal medulla, thyroid gland, and the kidney. Low levels of C4 mRNA are observed in the heart, ovary, small intestine, thymus, pancreas, and spleen. A similar observation was made by Feucht *et al.* (217).

Table V lists some cell lines in which C4 transcripts were tested for regulation of gene expression. IFN- γ was the cytokine that induced or further

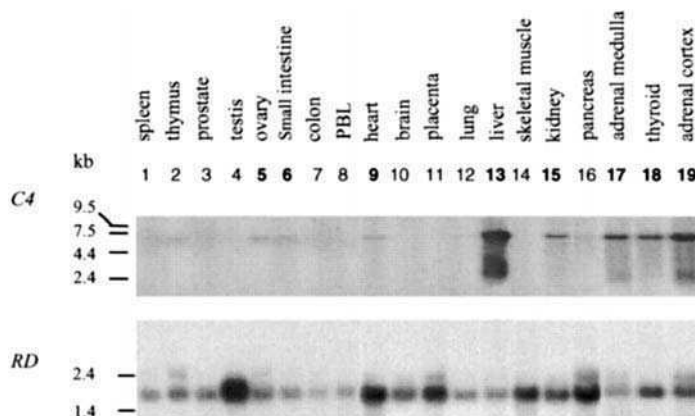


FIG. 10. A northern-blot analysis to show the expression of human complement *C4* transcripts in multiple tissues. Multiple tissue poly(A)⁺ RNA blots (Clontech, Palo Alto, CA) were hybridized with human *C4d*-specific probe (upper panel) and with a human *RD* probe corresponding to a subunit of the negative transcription-elongation factor (80) as positive control (lower panel). The results represent expression of C4 transcripts under normal conditions, i.e., *not* under γ -interferon induction. The liver is the major tissue expressing *C4* (lane 13). Other tissues expressing high levels of *C4* transcripts include the adrenal cortex (lane 19), adrenal medulla (lane 17), the thyroid gland (lane 18), and the kidneys (lane 15). Low levels of *C4* transcripts are detectable in the heart (lane 9), small intestine (lane 6), ovary (lane 5), pancreas (lane 16), thymus (lane 2), and spleen (lane 1) (taken from Ref. (78)).

TABLE V
A LIST OF CYTOKINES/REAGENTS TESTED TO AFFECT HUMAN *C4* GENE EXPRESSION

Inducer	C.E.	IFN- γ	LPS	TNF- α	IL-6	Vitamin D ₃	IFN- γ + LPS	IFN- γ + TNF- α	IFN- γ + vitamin D3	Ref.
<i>Cell type</i>										
Hepatoma, HepG2	+	↑		n.c.						(250)
Blood monocytes	+	↑	↓	n.c.	n.c.		Neutralized			(218)
Monocytic U937	—	↑				n.c.			↑ 2.3 ×	(220)
Skin fibroblasts	+	↑ 2.5 ×	n.c.	n.c.	n.c.			↑ 6 ×		(226)
Intestine epithelial, Coco-2	+	↑ 4.5 ×	n.c.	n.c.	↑ 3 ×					(222)

n.c., no change; neutralized, effects of IFN- γ neutralized.

enhanced expression of *C4* mRNA by 2- to 6-fold in a dose-dependent manner in many different cell types involved in local defense. In peripheral blood monocytes the stimulatory effect of IFN- γ was neutralized by bacteriotoxin lipopolysaccharide (LPS) (218). On the contrary, LPS slightly potentiated the effect of IFN- γ in the monocytic cell line U937 (219). Vitamin D₃ enhanced the effect of IFN- γ by 2.5-fold (220). By *in situ* hybridization experiments, *C4* transcripts were detected in intestinal epithelium and in mast cells (221). The synthesis of *C4* protein in the colonic adenocarcinoma cell line Caco-2 was enhanced by 3- to 4-fold in the presence of IL-6 or IFN- γ (222). In the kidney, *C4* transcripts were detected throughout the cortical tubular epithelium (223). Using glomerular mesangial cells, it was shown that the synthesis of *C4* transcripts and the secretion proteins were upregulated by IFN- γ (224). IFN- γ also induced *C4* biosynthesis in the glomerular epithelial cells (225). Treatment of the skin fibroblasts with TNF- α , LPS, IL-1, or IL-6 alone had no effect on the expression of *C4*. However, TNF- α (20 ng/ml) and IFN- γ (100 U/ml) synergistically increased the level of *C4* expression by 6-fold in skin fibroblasts (226).

B. The Size and Polygenic Variations of *C4* Strongly Correlate with Serum Protein Concentrations and Hemolytic Activities

An in-depth study on the biologic impacts of *C4A* and *C4B* polygenic and gene size variations has been performed in a study population from Budapest. The serum concentrations of *C4* and *C3* were determined immunochemically

by single radial immunodiffusion (RID) (227) using commercial antibodies against C3 and C4. The hemolytic C4 activities were measured by the effective molecule titration method (228). Hemolytic activity of C4 in serum was expressed in CH63 units/ml. The body mass index and cholesterol levels for all participants were also determined.

A strong positive correlation of serum protein concentrations of total C4 (and hemolytic activities) was established, not only with dosages of total *C4* genes but also with short *C4* genes. Besides the expected correlation with their corresponding gene dosages, C4A serum protein concentrations also correlate with the number of long *C4* genes, while C4B serum protein concentrations correlate with the number of short *C4* genes. Individuals containing one or more short *C4* genes consistently have higher serum C4 concentrations (Fig. 11) and C4 hemolytic activities than those with long *C4* genes only (12).

The positive impact of short *C4* genes is far-reaching because it not only affects the C4B and the total C4 protein levels, but also the level of C4A. This can be best illustrated in subjects with bimodular RCCX structures from both haplotypes that contain two *C4A* and two *C4B* genes. Individuals with homozygous LS/LS have serum C4A and C4B protein levels 34% and 40% higher than those from homozygous LL/LL, respectively. The levels of C4A and C4B proteins from individuals with heterozygous LL/LS are higher than those with LL/LL but lower than those with LS/LS. The higher serum C4 protein levels from individuals with LS/LS also translate into higher hemolytic activities. Thus, difference in intrinsic strength of the complement system in human populations may be achieved by variations in serum protein levels that are determined by gene size and gene dosages (12).

C. The Regulatory Motifs of Human C4 Gene Expression

The major start site for *C4* gene transcription is 51 nt upstream of the sequence for its translation initiation codon (51,82,229). No canonical TATA box is present in the 5'-regulatory sequences of the *C4A* and *C4B* genes. By means of reporter gene assays in transient-expression systems using human HepG2 cells, a biphasic promoter with proximal and distal positive-regulatory regions separated by a negative-regulatory region was shown within the 1-kb sequence upstream of the CAP site (230,231). The proximal-regulatory region lies within -178, and the CAP site contains an Sp1 site and three E boxes (232). The 3' E-box between -78 and -73 has been shown responsible for the IFN- γ induction of *C4* gene expression in U937 cells (231). In keeping with the observation that *C4* transcripts are present in the adrenal cortex (see below), considerable levels of human *C4*-promoter activity were demonstrated in transient-expression assay using mouse adrenal cell line Y1 (233).

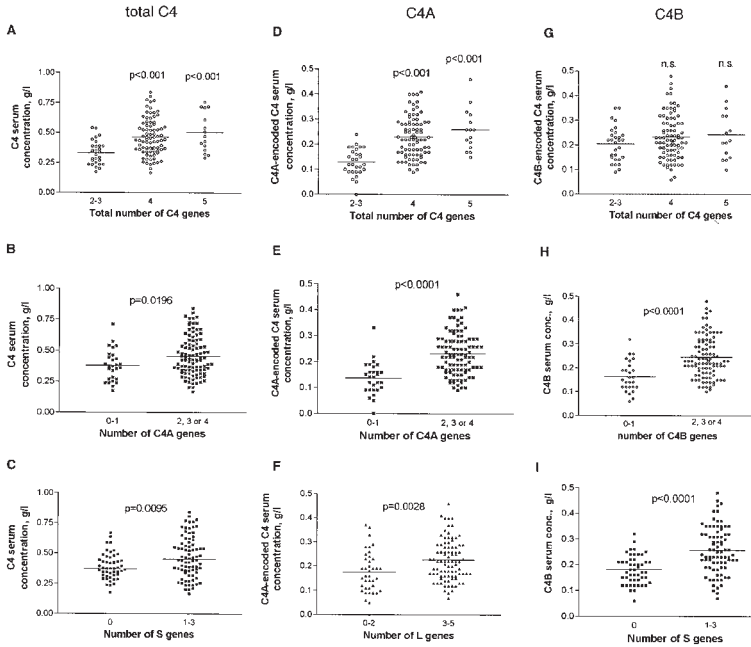
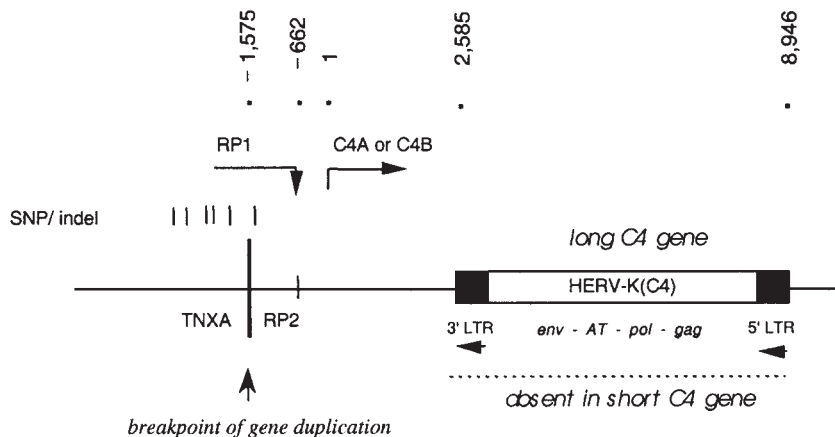


FIG. 11. Correlation of C4 serum concentrations with *C4* gene dosage and gene size variations. Panels A to C show the relationships of *total* serum C4 levels with *C4* polygenic and gene size variations. Panels D to F show the relationships of serum C4A proteins with *C4* polygenic and gene size variations. Panels G to I show the relationships of serum C4B proteins with *C4* polygenic and gene size variations. In panels A, D, and G, the *P* values for Tukey *post hoc* test of one-way ANOVA are indicated. For all other panels, *P* values for unpaired *t*-test are indicated. The data shows a strong correlation of total serum C4 concentration with *C4* gene dosages, the number of *C4A* genes, and the number of short *C4* genes. C4A serum protein levels correlate with total *C4* gene dosage, *C4A* gene dosage, and the number of long *C4* genes. C4B serum protein levels correlate with *C4B* gene dosage and the number of short genes (modified from Ref. (12)).

The possible variation of the promoter activities among different *C4A* and *C4B* genes has not been investigated in the context of SNPs present in the upstream-regulatory sequences. Two additional physical entities could have an impact on controlling the expression levels of human *C4A* and *C4B*. The first is the different genetic environment beyond position -1575 among the *C4* genes from the first RCCX module, and those from the second or third (or fourth) RCCX module (Figs. 9 and 12). The former is characterized by the presence of the *RP1*-specific sequences. The latter are characterized by the *TNXA* sequences. *cis*-Acting elements modulating *C4* gene expression could be present in the *RP1*-specific or the *TNXA*-specific sequences. The second entity is the presence of the endogenous retrovirus HERV-K(C4) in intron 9

module I or monomodular RCCX



module II, III or IV RCCX

FIG. 12. The different physical environments surrounding human *C4A* and *C4B* genes. At the 5' region, the DNA sequences beyond -1575 with respect to the *C4* translation initiation codon may differ totally. In module I it is the *RP1*-specific sequence. In module II, III, or IV it is the *TNXA*-specific sequence. SNPs and indels are present in *RP1*. Downstream of the *C4* promoter, at nucleotide 2585, the 6.36-kb endogenous retrovirus HERV-K(C4) is present in the intron 9 of the long *C4* genes. This large retroviral sequence is not present in the short *C4* genes. The rate to generate transcripts from long *C4* genes (20.6 kb in size) could be slower than that from short *C4* genes (14.2 kb in size). *cis*-Acting motifs regulating the human complement *C4* gene expression might be present in *RP1*, *TNXA*, and HERV-K(C4) (modified from Ref. (78)).

that is located 2585 nt downstream of the *C4* CAP site. This endogenous retrovirus would impose transcription of an additional 6.36-kb retroviral sequence in the *C4* heteronuclear RNA from the long genes. In addition, this retroviral element may have positive or negative-regulatory motifs that could affect the transcription of the *C4* gene. In the mouse MHC, the long-terminal repeats (LTRs) of the IMP present in intron 3 of the *RP1* gene confers the male-specific expression of the Slp protein in certain mouse strains (81,118,234).

D. Variation in Expression Levels of Mouse C4 and Slp Proteins

The levels of mouse plasma C4 and Slp also show great variation (235,236). This may be caused by variation in the gene number or defects in gene expression. For example, the C4 protein level from *H-2^k* is about 20-fold lower than that of *H-2^{w7}*. It was found that there was an insertion of the repetitive DNA element B2 into intron 13 in the *C4* gene of *H-2^k*, which led to

aberrant RNA splicing of liver *C4* transcripts (237). The abnormal *C4* mRNA from *H-2^k* contained a 200-bp insertion with the B2 sequence and part of the intron sequence (238). Unexpectedly, peritoneal macrophages and other organs except liver and lung of *H-2^k* had only the correctly processed *C4* mRNA, suggesting that the production of abnormal *C4* transcripts was tissue specific (239).

The serum level of Slp in *H-2^{w7}* is extraordinarily high and is expressed constitutively in both male and female mice (118,240). Mouse strains with this haplotype have one regular *Slp* gene, one regular *C4* gene, and three copies of hybrid genes that comprise 1–3 kb of *C4* sequence at the 5' region, and *Slp* sequence at the remaining 3' region (241–244). On the contrary, no Slp protein is made in mouse strains with *H-2^b* haplotype. The cause of this nonexpression is due to a single G-nucleotide insertion in exon 24 in which the thioester residues are encoded (81,245). This mutation may lead to frameshift mutation and truncation of the Slp protein.

VI. The Endogenous Retrovirus that Mediates the Dichotomous Size Variation of *C4* Genes

A. HERV-K(*C4*) in the Intron 9 of Long *C4* Genes

The size of intron 9 in a long *C4* gene is 6782 bp. The size of intron 9 in a short *C4* gene is 416 bp. This is because an endogenous retrovirus of 6360 bp in size had integrated into intron 9 of long *C4* genes (50,246). The site of integration is 283-bp downstream of the intron donor site and 133-bp upstream of the intron acceptor site (Fig. 9). The configuration of this endogenous retrovirus is opposite to the transcriptional orientation of *C4*. A direct target site repeat of 6 nt in the *C4* intron, 5'-cagaca-3' (or 5'-tgtctg-3' with respect to the endogenous retrovirus), flanks the endogenous retrovirus. This human endogenous retrovirus in the *C4* gene is named HERV-K(*C4*) because it contains a characteristic primer-binding site (PBS) for the transfer RNA for lysine (i.e., K).

A typical genomic organization is observed in the endogenous retrovirus (Fig. 8) located in the long *C4* genes (50,247,248). The 5' LTR and the 3' LTR are 548 and 550 bp, respectively. An 18-bp PBS that is complementary to the Lys₁₂ tRNA is present 3-bp downstream of the 5' LTR. A polypurine track (PPT) of 23 bp is present 2-bp upstream of the 3' LTR. The PBS and PPT are sequences necessary for DNA synthesis of the minus and the plus strands, respectively, during the retroviral replication. Between the PBS and PPT are the sequences corresponding to *gag*, *pol*, and *env* of 547, 1890, and 1874 bp, respectively. Located 149-bp downstream of the 5' LTR, the *gag* sequence has

one nonsense and four frameshift mutations. The succeeding *pol* sequence can be divided into a region coding for a reverse transcriptase (RT) that is 1440 bp in size, and another region coding for an integrase that is 450 bp long. There are 14 nonsense and four frameshift mutations in the *pol* sequence. Located between the *pol* and *env* regions is distinct sequence of 830 bp that is 72.7% A and T. The *env* region spans 1874 bp and contains sequences corresponding to a leader peptide, an outer membrane region, and a transmembrane region. The *env* sequence contains one frameshift and 23 nonsense mutations. In other words, it is highly unlikely that the retroviral sequence in the C4 gene is capable of generating a live retrovirus, although retrovirus-related transcripts or protein products could be generated (50).

The sequences for the 5' LTR and the 3' LTR were identical at the time of retroviral integration but they were subjected to independent deviations afterwards. The two LTR sequences in the long C4 gene share 89% sequence identities with 40 single-nucleotide substitutions and four indels. In addition, the 5' and 3' LTRs acquired an 11-bp and a 7-bp minideletion, respectively. Therefore, the original length of a LTR would be about 557 ± 4 bp long. Present in each LTR are sequences for a TATA box and an SV40-type transcriptional enhancer. A polyadenylation signal is also present in the 5' LTR but not the 3' LTR. An *in vitro* transcriptional assay showed that the 3' LTR still maintains the capability to direct the transcription of a reporter gene (Fig. 13A) (50).

Sequence analysis of intron 9 from three different long *C4B* genes and two long *C4A* genes revealed the integration of HERV-K(C4) into the identical site. The retroviral sequences among the long C4 genes are >99.7% identical. Moreover, the 5' and 3' LTRs from different RCCX modules or C4 loci share the identical point mutations and indels. Therefore, it is deduced that the presence of HERV-K(C4) in the long C4 genes is the result of a single retroviral integration event. Moreover, the same endogenous retrovirus, ERV-K(C4), is present at precisely the identical integration site in intron 9 of long C4 genes in Old World primates such as the African green monkey, rhesus macaque, and orangutan. Both long and short genes are present in all other Old World primates studied. Only short C4 genes were detected in the New World primate cotton-top tamarin, and in chimpanzee and gorilla. Either the short genes were the ancestral genes that never underwent integration of the endogenous retrovirus ERV-K(C4) and coevolved with the long genes, or the short genes were created by the loss of the ERV-K(C4). The latter could be achieved by a recombination between the 5' and 3' LTRs, leading to the deletion of the intervening genomic sequences *gag-pol-AT-env*, and the formation of a solitary LTR flanked by a copy of the target site repeat at each end. A comparison of the short intron 9 sequences from seven different human C4 genes and six other short genes from nonhuman

primates did not reveal any traces of LTR–LTR recombination, and no solitary LTR sequences flanked with target site repeats in the short *C4* genes were found. Therefore, it is highly likely that the short *C4* genes are the ancestral forms that coexist with the long *C4* genes in the Old World primates except the chimpanzee and gorilla. The long *C4* genes in these apes were probably lost during evolution (50,249).

Molecular genetic analysis of >300 healthy Caucasians revealed that HERV-K(C4) is present in 76% of human *C4* genes. In other words, about 24% of the human *C4* genes are short. Such a size dichotomy poses a question about the evolutionary advantage of maintaining both long and short *C4* genes in primate evolution. It has been suggested that the presence of this retroviral element (with multiple nonsense mutations) in intron 9 of the long *C4* gene would lead to the generation of antisense retroviral sequences whenever the *C4* gene is transcribed. Such antisense sequences could confer intrinsic protection against future retroviral infections. Double-stranded RNA molecules could be formed between the proviral genomic RNA and the antisense retroviral RNA with multiple mismatches, which may trigger the antiviral response through the interferon system and also cellular RNases that cleave mismatches between double-stranded RNA molecules.

This “antisense protection” hypothesis has been elegantly tested by Schneider *et al.* using mouse fibroblast cell line Ltk[−] transfected with human long *C4A* (LC4A) or long *C4B* (LC4B) genes as a source of mRNA for RT-PCR analysis, and for a reporter gene assay (73). First, antisense mRNA specific for human HERV-K(C4) was detected in the mouse transfectants, LC4A and LC4B. The LC4A and LC4B transfectants had been shown to produce human *C4A* and *C4B* transcripts and therefore could also generate specific antisense HERV-K(C4) retroviral RNA molecules (250). Second, β -galactosidase reporter constructs were generated in which specific sequences corresponding to *gag*, *pol*, or *env* of HERV-K(C4) were spliced upstream of β -gal reading frame in the reporter constructs, in the sense and in antisense configurations. A reduction of the β -galactosidase activities was demonstrated in the transfectants with constructs containing the *gag*- and *pol*-specific sequences in the sense configurations, but not the antisense configuration or the control mouse L cells without the human *C4* gene. This phenomenon is indicative of the human *C4*-mediated antisense inhibition on the β -galactosidase activity with the HERV-K(C4) *gag* or *pol* fusion. The reduction of the β -galactosidase activity was further potentiated by γ -interferon induction, which is in keeping with the higher expression level of the human *C4* gene in the presence of the cytokine. Thus, the results support the hypothesis of an antisense strategy mediated by HERV-K(C4) insertion into the long *C4* genes as a possible defense mechanism against retroviral infections (50,73).

The copy number of HERV-K(C4) in the *C4* genes, or the number of long *C4* genes, may vary between 0 and 8 among different individuals. It remains to be demonstrated if there is a dose-dependent effect of the HERV-K(C4) in the protection against retroviral infections, which could also be indirectly involved in the trigger of autoimmune diseases (251,252). A recent study of type I diabetes mellitus patients claimed that there is no preferential transmission of HERV-K(C4) to the affected offspring (253). A precautionary note for such epidemiologic studies is that the relevance of HERV-K(C4) on disease associations would be more on the *quantitative* rather than the *qualitative* aspects.

B. HERV-K(C4) in the Human and Primate Genomes

The application of a HERV-K(C4) LTR probe of an LTR-*env* probe for Southern-blot analysis of restriction enzyme digested genomic DNAs revealed the presence of 30–50 hybridizing fragments in all samples from Old World primates with a species-specific pattern, but not in samples from New World primates (Fig. 13B).

A bioinformatic, genome-wide database search has been performed to determine the presence of HERV-K(C4) sequences in the human genome, which has been sequenced almost to completion. The 550-bp 3' LTR sequence and the 5.26-kb *PBS-gag-pol-AT-env-PPT* sequence were used as the query (or bait) sequences separately in the BLAST research program at the NCBI human genome database (254), and the GCG-FASTA (255) search of the primate sequence database in GenBank. Seven non-*C4* HERV-K(C4) elements with similar structural organizations and high sequence identities are found. These HERV-K(C4)-related elements are assigned as F1, and F3 to F8 (Table VI). The chromosomes 1, 6, and 19 each contain two HERV-K(C4)-related sequences. One copy is present on the Y chromosome. The sequence identities between the prototype sequence from the long *C4* gene and F1, F3 to F8 were plotted by the BLAST-2-Sequences program and shown in Fig. 13. The HERV-K(C4)/F1 on the Y chromosome (gi:16172964) and HERV-K(C4)/F3 on chromosome 6 (gi:15551758) appear to be intact with all hallmark features. The other five have various degrees of insertions or deletions, dissimilarities or interruptions. Notably, HERV-K(C4)/F7 is interrupted by multiple copies of Alu repeats. Amongst these almost full-length retroviral sequences, the intraretroviral LTRs are 90–94% identical and the interretroviral LTR identities with the *C4* prototype are mostly 84–85% identical (Fig. 14).

C. Solitary LTRs of HERV-K(C4) in the Human Genome

Although no evidence for LTR–LTR recombination has been found in the intron 9 of the short *C4* genes, at least 44 copies of the

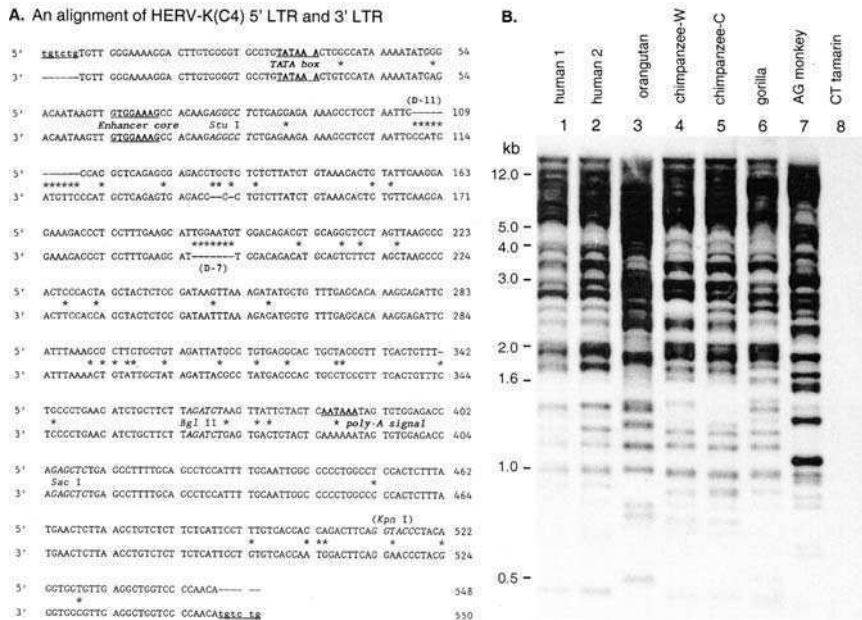


FIG. 13. The LTRs of the HERV-K(C4). (A) An alignment of the 5'- and 3'-LTR sequences from the long C4 gene. Mismatches in sequences are shown by asterisks. The target site repeats preceding the 5' LTR and succeeding the 3' LTR are in lower case and underlined. Distinct regulatory sequence motifs are labeled. The 5' LTR has an 11-bp deletion (D-11). The 3' LTR has a 7-bp deletion (D-7). (B) Genomic Southern-blot analysis of primate genomic DNA digested with *Stu*I-restriction enzyme and hybridized to a probe from the 5' LTR of HERV-K(C4). The hybridization pattern shows multiple copies and species-specific patterns of the ERV-K(C4) LTR-related sequences in various Old World primates. Most of the bands represent the presence of the solitary LTRs in the genomes of the Old World primates (modified from Ref. (50)).

HERV-K(C4) 5' LTR/3' LTR recombination remnants, i.e., solitary LTR flanked by target site repeats, are present in 17 different chromosomes of the human genome (Table VII). Among them, chromosomes 19, 9, and 6 have eight, six, and five copies of HERV-K(C4) solo-LTRs, respectively. All except the one that is interrupted by an *Alu* element have a size between 535 and 580 bp. There is no specific sequence for the integration site but the length of the target site repeats is predominantly 5 or 6 bp long. An additional remnant of intraretroviral recombination or rearrangement is present on chromosome 1, which is characterized by the presence of the first 496 bp of the 5' LTR that is immediately succeeded by the full-length 3' LTR.

Because of the presence of a TATA box, transcriptional enhancer(s), and a poly(A) signal, the solo-LTRs are a rich reservoir of regulatory sequences for

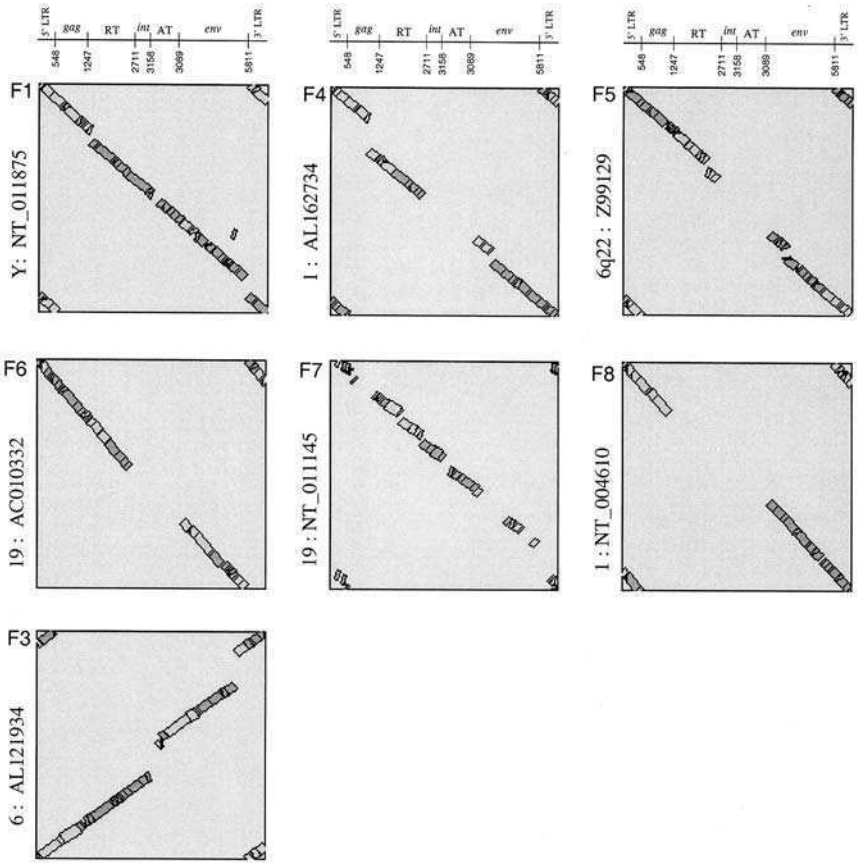


FIG. 14. The family members of HERV-K(C4) in the human genome. Seven members of HERV-K(C4) with structural organization and high sequence identities to that in the long C4 gene are present in the human genome. These genetic elements are named F1, and F3 to F8. The sequence identities between each element (Y-axis) and the prototypic sequence from the long C4 gene (X-axis) are plotted by the Blast-2-Sequences program. The chromosomal location and the GenBank accession number for each element is labeled on the Y-axis. The framework structure with nucleotide number (GenBank accession number: U07856) is shown on the X-axis to highlight the regions with sequence similarities. All entries except F3 (GenBank: AL121934) have sequence orientations opposite to the sense configuration of the endogenous retrovirus.

gene expression. There is evidence of the HERV-K(C4) solo-LTRs' role on the modulation of gene expression of the nearby structural genes. For example, a HERV-K(C4) solo-LTR is present at the genomic region for the vitamin B6 (pyridoxal) kinase, PDXK (AP001052; gi:18644884), in chromosome 21. Novel

TABLE VI
A LIST OF HERV-K(C4) FAMILY MEMBERS IN THE HUMAN GENOME

HERV-K(C4)	Accession no./gi	Ch. no.	Ori* (+/−)	Positions	Target sites	Size (bp)	5′ LTR/3′ LTR (identities)	Remarks
F1	NT_011875 16172964	Y	−	1306601 ← 1313823	TGGATT	7227	558/558 (90%)	Located in the intron 6 of the TAK1-binding protein 2; LOC256944
F3	AL121934 15551758	6	+	21909 → 30668	AAGAT A/C	8760	551/557 (92%)	4793 kb upstream of <i>C4</i>
F4	AL162734 19309425	1	−	34003 ← 41696	ATTAA	7695	545/547 (94%)	14-bp deletion in LTR
F5	Z99129 3281967	6q22	−	102056 ← 109304	GGACAT	7249	551/529 (91%)	
F6	AC010332 13786327	19	−	107155 ← 112785	AAAAC	5630	572/26	Most of 3′ LTR deleted
F7	NT_011145 9558577	19	−	1422484 ← 1432078	Unknown	9594	n.a.	Incomplete matches; 1–177 of 5′ LTR missing
F8	NT_004610 22045511	1	−	191541 ← 197366	Undetermined	5826	551/537 (92%)	Located between phospholipase PLA2G2E and PLA2G2A

*Ori, orientation of the sequence entry with respect to HERV-K(C4).

transcripts containing 412 bp of the HERV-K(C4) LTR sequence with the poly(A) signal at the 3' ends were found in three independent cDNA clones generated from teratocarcinoma cell NT2 (AK056502) and rhabdomyosarcoma muscle tissue (BC021550, BC008008). These transcripts only share one exon of 60 bp with PDXK and would code for a distinct protein of 162 amino acid residues (256,257).

VII. Evolution of Complement C3/C4/C5 and the RCCX Modules in the MHC

A. The α_2 Macroglobulin Protein Family

Together with complement components C3 and C5, C4 belongs to the α_2 macroglobulin protein family, A2M (143,258). α_2 Macroglobulin is a proteinase inhibitor that binds to all classes of proteases and conveys them to a receptor-mediated endocytosis-clearance pathway. Thus, α_2 macroglobulin may potentially contribute to the innate immunity by inactivating the proteinase virulence factors of pathogens (259). Structurally, members of the A2M protein family are characterized by the presence of an intrachain thioester bond that is able to form a covalent amide or ester linkage with a target. In humans there are seven members in the A2M protein family (260,261), which are the α_2 macroglobulin, pregnancy zone protein (PZP) (262), KIAA1283 (263), complement components C4A, C4B, C3, and C5. The human α_2 macroglobulin gene and the PZP gene are closely linked and located on chromosome 12, the KIAA1283 gene and the complement C3 gene are located at chromosome 19p13.1–13.3, C4A and C4B are present in the MHC, and complement C5 gene is located on chromosome 9q33–34. Each of these is a member of paralogous genomic segments with multiple linked genes that duplicated and dispersed to four different chromosomes (264).

The C3/C4/C5 proteins are more closely related to each other as they acquired an anaphylatoxin (ANA) domain of 74–77 amino acids close to the middle of the protein and a netrin domain (NTR) of about 152 residues at the carboxyl terminus of each protein (265). Preceding the ANA domain there are four basic residues which are subjected to proteolysis and therefore give rise to the β and α chains of the proteins. Further distinct structural features added to complement C4 include the basic residues between the α - γ chain junction and a cluster of three tyrosine-sulfation sites close to the carboxyl terminus of the α -chain (Fig. 9) (266). Complement C5 is an odd member of the A2M protein family as it has lost the thioester structure.

TABLE VII
A LIST OF SOLITARY LTRs (SOLO-LTR) OF HERV-K(C4) PRESENT IN THE HUMAN GENOME

Solo-LTR	GenBank Accession no.	Ch no.	Ori (+/−)	Positions	Target site repeats	Size (bp)	Remarks
1	AC 009408	2	(−)	144384 ← 144939	TATAC	555	
2	AC 007868	8	(−)	54534 ← 55088	CTCCC	554	
3	AL713966	6	(+)	46771 → 47326	GGGG	555	
4	AC 005204	19	(−)	16530 ← 17085	GTGTG	555	
5	AP001752	21q	(+)	92139 → 92692	TTTAG	553	
	AP 001052*	21	(+)	70318 → 70871	TTTAG	553	
6	AL662886	X	(−)	59008 ← 59563	GCTCTA	555	
7	AL591682	6	(−)	51927 ← 52482	GAAAC	555	
8	AC105751	3	(−)	9243 ← 9802	CTGCTG	559	
9	AJ305027	17p13.1	(−)	679 ← 1234	CCAGG	555	
10	AL35479	9q31.3–33.2	(+)	167520 → 168080	CAAAG	560	
11	AL645665	X	(+)	66421 → 66965	CTTAAG	544	
12	AL157827	9q21.33–22.31	(−)	43615 ← 44169	T/A ATTGT	554	
13	AC092305	19	(−)	27246 ← 27784	GTCTC	538	
14	HS425C14	6q22	(−)	102056 ← 102606	GGACA	550	
15	AC019252	8	(+)	64520 → 65070	C/G CC T/G	550	
16	AC022143	19	(−)	81655 ← 82207	TGTAGG	552	
17	AC104090	4	(+)	42764 → 43322	GG T/C G A	558	
18	AC010311	19	(+)	78408 → 78952	GTCTC	544	
19	AC104825	4	(−)	122282 ← 122835	CTTTTA	553	
20	AL391383	13	(+)	54238 → 54790	CCTGTT	552	
21	AC113618	2	(+)	86383 → 86938	AC G/A G T	555	
22	AL160270	9	(−)	124475 ← 125018	GGCATG	543	
23	AC92619	2	(+)	157626 → 158181	AC G/A GT	555	
24	AL132657	10	(+)	51633 → 52185	GAATC	552	
25	ACO74331	19	(−)	152418 ← 152967	GTAAG	549	
26	AL008718	22q13.2–13.3	(+)	10399 → 10945	TGCAAC	546	

27	AL158151	9	(+)	8770 → 9315	GGGGAA	545	
28	AL121932	6	(+)	2394 → 2933	CATGT	539	
	AL671879*	6	(+)	34396 → 34935	CATGT	539	
29	ACO74138	19	(-)	88839 ← 89386	GAAC	547	
30	AC010128	Y	(+)	10293 → 10838	GTCCAA A/G	545	
31	AC093726	7	(-)	11125 ← 11667	ACTCCA	542	
32	AC006376	Y	(+)	2860 → 3405	C A/G T/C AGC	545	
33	AL159997	9q34.12–34.3	(+)	77116 → 77660	CTCT C/T A	544	
34	AC002470	22q11.2	(-)	54638 ← 55190	TAAGA	552	
	AC007308	22q11	(-)	125790 ← 126342	TAAGA	552	
35	AC016773	4	(+)	85430 → 86240	ATGGGG	810	Interrupted by Alu
	M63005	4	(-)	22492 ← 23301	ATGGGG	810	Interrupted by Alu
36	AP002533	1q22–23	(+)	59712 → 60288	GTCCAA	576	
37	AL590398	6	(+)	57555 → 57889; 64267 → 64483	CCCTAT	(552)	Interrupted
38	AC008056	14	(+)	142682 → 143262	TTGGTC	580	
39	AC007785	19	(+)	28627 → 29202	CAAGGT	575	
40	AC003973	19	(-)	65544 ← 66121	ATAA T/C A	577	
41	AL353572	9	(-)	82148 ← 82724	ATGGC	576	
42	AL357500	1	(-)	92531 ← 93106	GTTAC	575	
	AC004236	1	(+)	38815 → 39390	GTAC	575	
43	AL117327	Xq22.1–23	(-)	41471 ← 42006	C A/G GGAG	535	
44	AC104576	8	(+)	8966 → 9535	CATTTC	569	198–258 deleted
	AC093368*	8	(-)	59037 ← 59606	CATTTC	569	198–258 deleted
	AC104566*	8	(+)	99163 → 99732	CATTTC	569	198–258 deleted
45**	AL645860	1	(-)	4770 → 5325	TATGG	555	Complete 3' LTR; partial 5' LTR
				5325 ← 5821		496	

*Probably identical to the previous entry due to redundant data deposition or multiple sequences available.

**Continuous sequence with the complete 3' LTR plus nucleotide sequences 1–496 of the 5' LTR. Sequences between the 3' end of the 5' LTR plus the *PBS-gag-pol-AT-env* regions deleted.

Besides the C3, C4, and C5 proteins, the ANA motifs are also present in fibulin-1 (267) and fibulin-2 (268), which are extracellular matrix proteins in close association with microfibrils containing fibronectin or fibrillin.

The NTR module at the carboxyl region of complement C5 has been shown by alanine-scanning site-directed mutagenesis to be the distal binding site of the *classical* pathway C5 convertase, C4b2a3b (269). The physiological roles of the NTR modules in complement C4 and C3 are unknown, and it will be worthwhile to investigate if they form binding sites for the multicomponent proteases in the complement-activation pathway such as the C1 complex, the MBL/MASP complex, and the C3 convertase. NTR modules are also present in type I collagen C-proteinase enhancer protein and tissue inhibitors of metalloproteases (270).

Although components of the classical pathway and the lytic pathway, and the regulatory proteins of the complement system are vertebrate-specific proteins, complement proteins of the lectin pathway and the alternative pathway appeared much earlier in evolution. In *C. elegans* and *Drosophila* there are one and five A2M-like genes, respectively. Four of the *Drosophila* A2M-like proteins, TEP1–TEP4, have the thioester protein signature sequences. The expression levels of these *Tep* genes are markedly upregulated after an immune challenge. Moreover, the increased expression level of *Tep1* is dependent on the Janus kinase gene, *hopscotch*, implying that the TEP proteins could be involved in the innate immune response in the insect (271–273). Sequence analyses suggested that the five A2M-like proteins in *Drosophila* are closely related to α_2 macroglobulin and to KIAA1283.

The presence of complement C3/C4/C5-like protein can be traced back to echinoderms and ascidians (274–276). These complement proteins probably play important roles in the discrimination of the host and parasites through opsonization of parasites for phagocytosis. In different species of bony fishes, 2–5 genes for complement C3 are present. In a phenomenon similar to that of complement C4A and C4B proteins in Old World primates (141,249,277) and in cattle and sheep (143), the teleost C3 proteins diversified to react to substrates with differential affinities (275,278). In particular, one of the two C3 genes in the Japanese medaka fish has the catalytic residue corresponding to His-1106 in human C4B changed to Ala (278). The common carp has the five C3 genes, C3-1 to C3-5, which has Ser, His, His, Gln, and His residues at the described catalytic residue, respectively. Interestingly, the isoform with Ser, C3-1, has the highest hemolytic activity in the artificial assay system (279). The presence of multiple C3 genes in the common carp and multiple C4 genes in humans, old world primates, sheep, and cattle underscores the importance of this protein family to accomplish functional diversities to defend against different parasites in the innate immune response.

B. Similarities and Dissimilarities of the Human and Mouse RCCX in the MHC

A percent-identity plot of the human and mouse genomic regions spanning from complement *C2* to the *Notch4* gene at MHC class III and class II boundary is shown in Fig. 15. The diagram highlights the conserved genomic organizations including gene orders, the exon–intron structures, and also the regulatory sequences between the two mammalian species (41).

In both human and mouse MHC complement gene cluster (MCGC), there are two sets of gene duplications. The first set is the single gene duplication that gives rise to *C2* and factor B (*Bf*). In both human and mouse, *C2* and factor B proteins share 41.3% amino acid sequence identities. The second set is the modular gene duplication for *RP1/RP2*, *C4A/C4B* (*Slp/C4* in mouse), *CYP21A/CYP21B*, and *TNXA/TNXB*. In humans, the duplicated RCCX modules are highly similar (99% identical) in sequence. There is convincing evidence for the homogenization of polymorphic or mutant sequences in the RCCX modules due to misalignments among monomodular, bimodular, and trimodular structures. As described in the previous sections, this homogenization process is not only a driving force for the functional diversities of *C4*, but may also be one of the root causes for MHC disease associations (42,47,280). In mouse, the RCCX modules have more sequence deviations that include selective integrations of numerous repetitive DNAs, retroelements, multiple point mutations, and minideletions. These gross sequence changes would decrease the efficiency of alignments between nonallelic genes from homologous chromosomes. Therefore, homogenizations of the diversified sequences in the duplicated modules would be less frequent in the mouse MCGC than that of human's.

An endogenous retrovirus, IMP, is present in reverse orientation in intron 4 of mouse *RP1*. The 5' LTR of the IMP was shown to contain a glucocorticoid-responsive element that conferred the sex-limited expression of *Slp* in male mice (281,282). Whether the LTRs would bestow a differential expression of *RP1* and/or *DOM3Z* in male and female mice has yet to be determined.

The intergenic sequence between *Slp* and *C4* in the mouse *H2^b* haplotype is 65.5 kb (283). The orthologous region in humans is 10 kb in size (Fig. 16). In humans the duplicated sequence for *RP2* is 913 bp and spans the last two and half exons. The breakpoint of gene duplication is located in exon 5 (Figs. 9 and 16). The *RP1/RP2*-duplicated sequences are identical. In the mouse, the duplicated sequence for *RP2* is 3.4 kb and corresponds to the last five exons of *RP1*. The breakpoint of gene duplication occurs in intron 2. The duplicated DNA sequence in mouse *RP2* contains

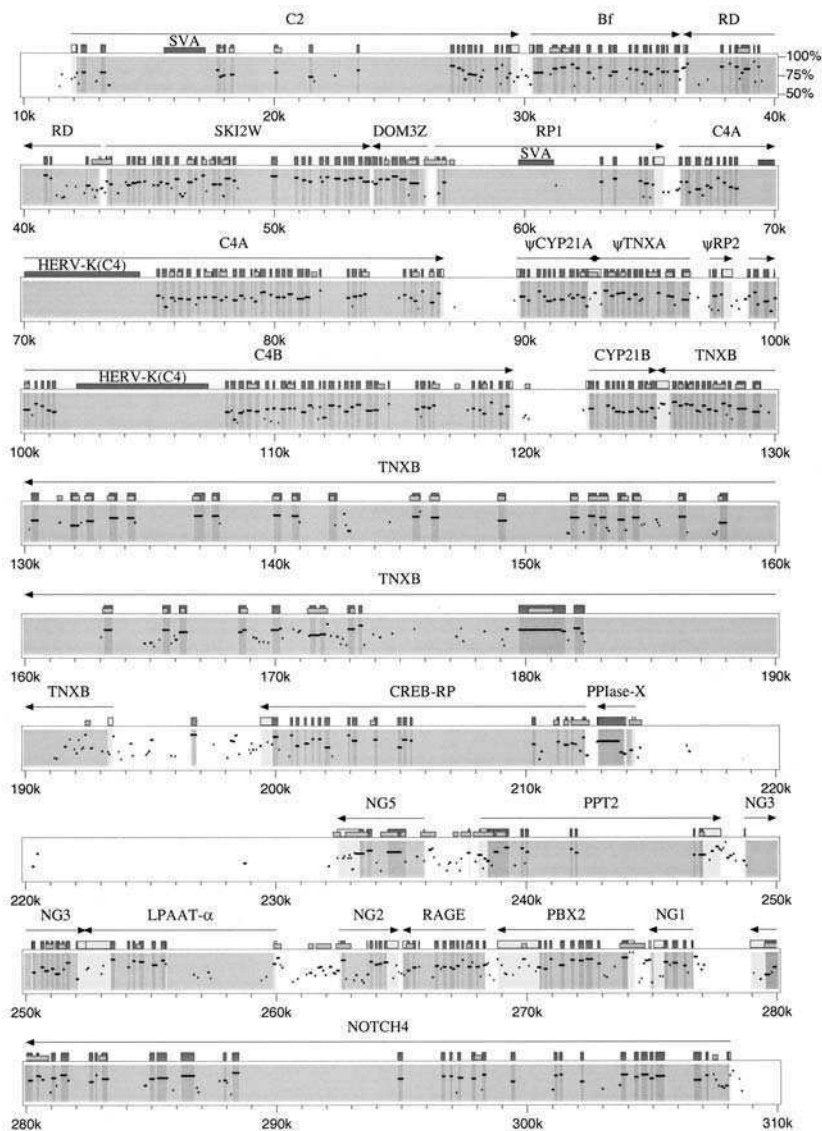


FIG. 15. A molecular map comparing the gene organizations of human and mouse MHC class III regions from complement *C2* to *Notch4*. The mouse sequences with identities from 50 to 100% are plotted below the human sequences. Exons for human genes are shown as solid, purple boxes. The genomic region spanned for each gene is shaded green. The 5'- and 3'-untranslated sequences are in yellow boxes. The conserved coding sequences between human and mouse genes are highlighted in purple. Locations of CpG-rich dinucleotides are shown as orange boxes. Large, repetitive retroelements SVA and endogenous retrovirus HERV-K(C4) are shown in burgundy. Arrows with solid lines represent configurations of structural genes. *RP2* and *TNXA* are partially duplicated gene fragments. *CYP21A* is a pseudogene in human. Pseudogenes and gene fragments are labeled with ψ in front of gene names. Genomic DNA sequences are obtained from the following accession numbers: human, U89335–U89337, AF019413, M59815, M59816, L26260–L26263, U07856, AF059675, and AF077974; mouse, AF030001, AF049850, and AF109906. Numberings below box represents length in human genomic DNA in kb (2k=2 kb) (modified from Ref. (41)). (See Color Insert.)

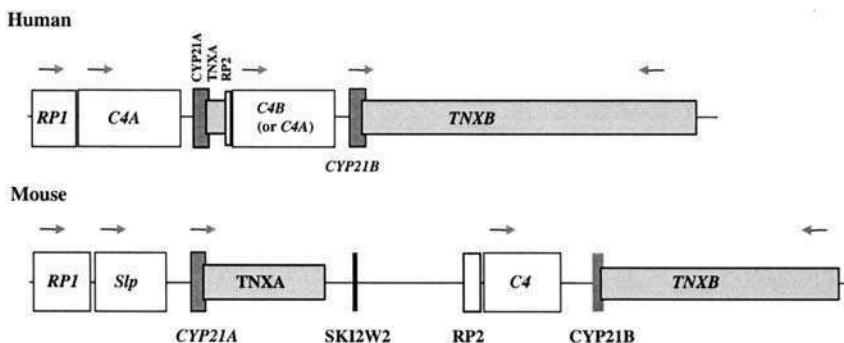


FIG. 16. The structures of *RCCX* modules in humans and in mice. Only the bimodular *RCCX* haplotype in human is shown. Arrows represent transcriptional configuration of structural genes. Pseudogenes or gene segments are labeled in gray without arrows. Note that the intergenic distance between human *C4A* and *C4B*, and between mouse *Slp* and *C4* are markedly different (modified from Ref. (41)).

multiple point mutations and minideletions, and is therefore unlikely to be a functional gene.

The duplicated region for mouse *TNXA* is 10 kb with 16 exons, compared with 4.5 kb and 13 exons for human *TNXA*. Again, there are extensive sequence variations/mutations in the mouse *TNXA* gene fragment. Unexpectedly, a partially duplicated gene fragment of 1.8 kb for *SKI2W* is present between *RP2* and *TNXA*, which is not observed in the human MHC. In other words, there is a considerable divergence in the organizations of the human and mouse *RCCX*.

It is possible that there was a single primordial duplication event involving *RP*, *C4*, *CYP21*, and *TNX*. Afterwards, the duplicated four-gene modules in human and in mouse took different routes of secondary recombinations, deletions, insertions, and point mutations that led to the current divergence in the *RP2* and *TNXA* gene fragments, and different mechanisms to inactivate one of the duplicated *CYP21* genes. It is also of interest to note that in some mouse haplotypes there are multiple copies of *Slp/C4* and *CYP21* genes (241,284,285), a phenomenon reminiscent of the *C4* polygenic variation in humans.

It had been presumed that there was one *C4* gene in the rat genome. However, recent advances in the characterization of the rat MHC have conclusively shown the presence of a duplicated genomic segment with *RP*, complement *C4*, and *CYP21* at the genomic region between the MHC class II and class III (286), in addition to the *RP-C4-CYP21-TNX* genes at the orthologous genomic positions between *DOM3Z* and *CREB-BP* (41,81). Moreover, the Old World primates have been shown to possess a similar

degree of complexity to humans with polygenic and size variations of the RCCX (79,84,141,249,287–289). Regardless of the mechanism(s) involved during the evolutionary process, the discoveries of an additional *C4-like* gene in the rat and multiple *C4A* and *C4B* genes in Old World primates are consistent with the hypothesis that mammalian C4 is under continuous selection pressure to achieve qualitative and quantitative diversities to deal with a variety of parasitic antigens effectively.

VIII. Afterthoughts: Eleven Outstanding Issues to be Addressed

(i) *Roles of C4 on Tolerance and Autoimmunity*

The involvement of C4 in autoimmunity may be achieved by its role on immunoclearance of immune complexes and apoptotic materials. However, it is puzzling why C3 deficiency does not manifest an autoimmune disease phenotype similar to C4 deficiency. It appears that C4 plays a role in the switching of IgM to IgG, immunotolerance and deletion of autoreactive B cells (reviewed in Refs. (290,291)). However, little is known about the pathways or the underlying mechanisms. Moreover, in many ethnic groups of the human population, complete and partial deficiencies of C4A appears to be predisposing risk factors of SLE. It is not known whether there are specific cellular factors interacting with C4A to protect against autoimmune disorder.

(ii) *Partial C4A Deficiency and Disease Associations*

Most of the previous epidemiologic studies of C4 and disease association were based on an assumption that there were two *C4A* and two *C4B* genes in a diploid genome. The partial deficiency of an isotype was assigned based on a difference in band intensities of the C4A and C4B proteins. This assumption is faulty because such phenotypes only represent unequal expression of C4A and C4B proteins that is a combined result of unequal gene dosages, long and short genes, and possibly differential responses to immunoregulatory stimulus such as induction by γ -interferon. In light of the unprecedented *C4* polygenic and gene size variations, many previous studies on the associations of C4A deficiencies with SLE or other diseases require careful reinterpretations, and preferably new prospective investigations.

(iii) *Polygenic and Gene Size Variations of C4 in Different Races*

To date, detailed and accurate information is only available in two Caucasian populations (11,12). An accurate account of the quantitative and qualitative variations of *C4A* and *C4B* genes and proteins in other races and ethnic groups is not available. The information may be relevant to

epidemiologic studies of infectious and autoimmune diseases, and for the therapeutic applications of complement inhibitors in various clinical intervention including transplantation and myocardial infarction.

(iv) Structural and Functional Basis of C4 Polymorphisms

There may be more than polymorphic 40 variants of C4 proteins in human populations. However, only a handful of them have been fully characterized. The structural basis for 75% of the C4A or C4B allotypes remains unknown. It is not known whether C4 polymorphisms lead to functional diversities as predicted (20). Elucidation of the polymorphic residues will not only help development of efficient DNA-typing techniques that would circumvent the heavy reliance on well-preserved plasma samples for allotyping experiments, it will also help in locating protein regions for in-depth physiological studies.

(v) Local Expression of C4A and C4B Proteins and Polygenic Variation

The impact on the expression of *C4A* and *C4B* transcripts and protein expression by *C4* gene dosage and gene size variation, and the response to LPS, γ -interferon, and other cytokines in various cell types have not been directly investigated.

(vi) The Three-Dimensional Structure of the C4 Protein

The crystal structure of C4d helps explaining the role of C4A and C4B isotypic residues in modulating the thioester activities, and yields support to the conformational epitopes of the Rg and Ch blood-group antigens. Crystal structures on other C4 protein regions such as the β -chain, the γ -chain, and possibly the native, activated, or inactivated forms of C4 would shed light on explaining how C4 interacts with the C1 complex, the MBL complex, C2, C3, C5, receptors of C4, DAF, factor I and cofactor proteins DAF, MCP, and C4b-binding protein.

(vii) The Function and Receptor of the Activation Peptide C4a

While C4a may be a useful marker of complement activation, disease activity (292), infections, and transplant rejection (53), the physiologic roles of C4a remain a mystery. Whether C4a exerts an inhibitory role on monocyte chemotaxis (175) deserves further exploration and critical analysis. By the same token, it would be rewarding if the C4a receptor is characterized and cloned.

(viii) Post-translational Processing of the C4 Protein

Little is known about the importance of the biantennary glycosylations on the α -chain and the high mannose-type glycosylation of the β -chain on the reactivities of human C4A and C4B. It is of interest to determine whether high mannose-type glycosylation on C4 plays a role on opsonization by "alienating" the antigens. Similarly, little is known about the physiological significance of the tyrosine sulfations, and the roles or the fate of the 26 amino acids removed from the carboxyl terminus of the α -chain.

(ix) *Receptor(s) for C4d?*

The physiological relevance of C4d on erythrocytes and possibly follicular dendritic cells is unknown. Whether there is a receptor for, or a protein interacting with, human C4d is a matter of interest. Actually there is no knowledge of membrane receptor(s) for human complement C4 other than CR1. Many novel genes have been discovered in the regulator for complement-activation (RCA) complex on chromosome 1, which include CRIL1, MCPL1, C4BPAL1, C4BPAL2 (293,294). It is possible some of these gene products are important on the C4 protein function.

(x) *Why are RCCX Modules Located in the MHC?*

Is the physical linkage and the modular duplication of *RP*, *C4*, *CYP21*, and *TNX* in the MHC by chance or by design? Why are these four genes expressing at high levels in the adrenal glands?

(xi) *C4 in MHC-Associated Diseases*

Many variants of C4A and C4B belong to different ancestral haplotypes. Strong associations between these haplotypes and autoimmune diseases were described. Little is known, however, about the mechanism of these associations. In other words, the role(s) of the central part of the MHC (class III) including *C4* genes in the susceptibility for autoimmune diseases need to be vigorously investigated.

ACKNOWLEDGMENTS

We are indebted to Drs. Frank Arnett (Houston, TX), Roger Dawkins (Perth, Australia), Joann M. Moulds (Philadelphia, PA), Barry Myones (Houston, TX), Maisa Lokki (Helsinki, Finland), Daniel Birmingham, Lee Hebert, Gloria Higgins, Robert Rennebohm, Brad Rovin, Juan Sotos, and Bill Zipf for collaborations, former laboratory members Brad Baker, Drew Dangel, Anna Mendoza, Kristi Rupert, Salih Sanglioglu, and Liming Shen for contributions relevant to this review. We are grateful to Ms Charlene Cameron and Ms Carrie Schreiner Worthington for enthusiastic secretarial assistance.

This work was supported by NIDDK grant P01-55546, NIAMS grant R01 AR43969, Columbus Children's Research Institute pilot grants, the March of Dimes Basil O'Connor Starter Scholar Award (5-FY93-0887), the Central Ohio Diabetes Association, and Pittsburgh Supercomputing Center through NIH Center for Research Resources Cooperative Agreement grant 1P41 RR06009 (C. Y. Y.).

REFERENCES

1. J. Gordon, H. R. Whitehead and A. Wormald, The action of ammonia on complement. The fourth complement, *Biochem. J.* **20**, 1028–1035 (1926).
2. J. Klein. In: "Immunology: The Science of Self-Nonself Discrimination", pp. 310–346. New York, John Wiley & Sons, 1982.

3. L. Du Pasquier and G. W. Litman. In: "Origin and Evolution of the Vertebrate Immune System". Berlin, Springer, 2000.
4. L. A. Rached, M. F. McDermott and P. Pontarotti, The MHC big bang, *Immunol. Rev.* **167**, 33–44 (1999).
5. Y. Ohta, M. Nonaka, M. F. Flajnik, Linkage analysis of MHC genes in nurse shark. 7th International Workshop on MHC Evolution and VIIIth Complement Genetics Workshop and Conferences A36, 2002.
6. R. Mo, Y. Kato, M. Nonaka, K. Nakayama and M. Takahashi, Fourth component of *Xenopus laevis* complement: cDNA cloning and linkage analysis of the frog MHC, *Immunogenetics* **43**, 360–369 (1996).
7. The MHC Sequencing Consortium, Complete sequence and gene map of a human major histocompatibility complex, *Nature* **401**, 921–923 (1999).
8. J. Kaufman, S. Milne, T. W. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob and S. Beck, The chicken B locus is a minimal essential major histocompatibility complex, *Nature* **401**, 923–925 (1999).
9. M. C. Carroll, R. D. Campbell, D. R. Bentley and R. R. Porter, A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B, *Nature* **307**, 237–241 (1984).
10. D. Chaplin, D. Woods, A. Whitehead, G. Goldberger, H. Colten and J. Seidman, Molecular map of the murine S region, *Proc. Natl. Acad. Sci. USA* **80**, 6947–6951 (1983).
11. C. A. Blanchong, B. Zhou, K. L. Rupert, E. K. Chung, K. N. Jones, J. F. Sotos, R. M. Rennebohm and C. Y. Yu, Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in Caucasians: the load of RCCX genetic diversity on MHC-associated disease, *J. Exp. Med.* **191**, 2183–2196 (2000).
12. Y. Yang, E. K. Chung, B. Zhou, C. A. Blanchong, C. Y. Yu, G. Füst, M. Kovács, A. Vatay, C. Szalai, I. Karádi, L. Varga, Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities and body mass indexes. *J. Immunol.* Submitted.
13. K. B. M. Reid and R. R. Porter, The proteolytic activation systems of complement, *Ann. Rev. Biochem.* **50**, 433–464 (1981).
14. H. J. Muller-Eberhard, Molecular organization and function of the complement system, *Ann. Rev. Biochem.* **57**, 321–347 (1988).
15. S. K. A. Law and K. B. M. Reid. In: "Complement". Oxford, IRL Press at Oxford University Press, 1995.
16. M. J. Walport, Complement—part I, *New Engl. J. Med.* **344**, 1058–1066 (2001).
17. M. J. Walport, Complement—part II, *New Engl. J. Med.* **344**, 1140–1144 (2001).
18. M. B. Pepys, Role of complement in the induction of antibody production in vivo. Effect of cobra venom factor and other C3-reactive agents on thymus-dependent and thymus-independent antibody responses, *J. Exp. Med.* **140**, 126–145 (1974).
19. M. C. Carroll, The role of complement and complement receptors in induction and regulation of immunity, *Annu. Rev. Immunol.* **16**, 545–568 (1998).
20. R. R. Porter, Complement polymorphism, the major histocompatibility complex and associated diseases: a speculation, *Mol. Biol. Med.* **1**, 161–168 (1983).
21. G. Hauptmann, G. Tappeiner and J. Schifferli, Inherited deficiency of the fourth component of human complement, *Immunodeficiency Rev.* **1**, 3–22 (1988).
22. H. R. Colten and F. S. Rosen, Complement deficiencies, *Annu. Rev. Immunol.* **10**, 809–834 (1992).
23. J. A. Schifferli, Y. C. Ng and D. K. Peters, The role of complement and its receptor in the elimination of immune complexes, *New Engl. J. Med.* **315**, 488–495 (1986).

24. J. B. Cornacoff, L. A. Hebert, W. L. Smead, M. E. VanAman, D. J. Birmingham and F. J. Waxman, Primate erythrocyte-immune complex-clearing mechanism, *J. Clin. Invest.* **71**, 236–247 (1983).
25. R. D. Campbell, A. W. Dodds and R. R. Porter, The binding of human complement component C4 to antibody-antigen aggregates, *Biochem. J.* **189**, 67–80 (1980).
26. N. Kishore, D. Shah, V. M. Skanes and R. P. Levine, The fluid-phase binding of human C4 and its genetic variants, C4A3 and C4B1, to immunoglobulins, *Mol. Immunol.* **25**, 811–819 (1988).
27. J. A. Schifferli, G. Hauptmann and J. Pierre-Paccaud, Complement-mediated adherence of immune complexes to human erythrocytes. Difference in the requirements for C4A and C4B, *FEBS Lett.* **213**, 415–418 (1987).
28. L. Paul, V. M. Skanes, J. Mayden and R. P. Levine, C4-mediated inhibition of immune precipitation and differences in inhibitory action of genetic variants, C4A3 and C4B1, *Complement* **5**, 110–119 (1988).
29. P. A. Gatenby, J. E. Barbosa and P. J. Lachmann, Differences between C4A and C4B in the handling of immune complexes: the enhancement of CR1 binding is more important than the inhibition of immunoprecipitation, *Clin. Exp. Immunol.* **79**, 158–163 (1990).
30. J. A. Schifferli, G. Steiger, J.-P. Paccaud and A. G. Sjöholm, Difference in the biological properties of the two forms of the fourth component of human complement (C4), *Clin. Exp. Immunol.* **63**, 473–477 (1986).
31. Z. L. Awdeh and C. A. Alper, Inherited structural polymorphism of the fourth component of human complement, *Proc. Natl. Acad. Sci. USA* **77**, 3576–3580 (1980).
32. D. E. Isenman and J. R. Young, The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotypes of human complement component C4, *J. Immunol.* **132**, 3019–3027 (1984).
33. S. K. A. Law, A. W. Dodds and R. R. Porter, A comparison of the properties of two classes, C4A and C4B, of the human complement component C4, *EMBO J.* **3**, 1819–1823 (1984).
34. A. W. Dodds, X.-D. Ren, A. C. Willis and S. K. A. Law, The reaction mechanism of the internal thioester in the human complement component C4, *Nature* **379**, 177–179 (1996).
35. C. A. Tilley, D. G. Romans and M. C. Crookston, Localization of Chido and Rodgers determinants to the C4d fragment of human C4, *Nature* **276**, 713–715 (1978).
36. C. M. Giles, B. Uring-Lambert, J. Goetz, G. Hauptmann, A. H. L. Fielder, W. Ollier, C. Rittner and T. Robson, Antigenic determinants expressed by human C4 allotypes: a study of 325 families provides evidence for the structural antigenic model, *Immunogenetics* **27**, 442–448 (1988).
37. C. Y. Yu, R. D. Campbell and R. R. Porter, A structural model for the location of the Rodgers and the Chido antigenic determinants and their correlation with the human complement C4A/C4B isotypes, *Immunogenetics* **27**, 399–405 (1988).
38. J. M. Moulds, S. L. Roberts and T. D. Wells, DNA sequence analysis of the C4 antigen WH: evidence for two mechanisms of expression, *Immunogenetics* **44**, 104–107 (1996).
39. G. Mauff, B. Luther, P. M. Schneider, C. Rittner, B. Strandmann-Bellinghausen, R. Dawkins and J. M. Moulds, Reference typing report for complement component C4, *Exp. Clin. Immunogenet.* **15**, 249–260 (1998).
40. C. Y. Yu and R. D. Campbell, Definitive RFLPs to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles, *Immunogenetics* **25**, 383–390 (1987).
41. C. Y. Yu, Z. Yang, C. A. Blanchong and W. Miller, The human and mouse MHC class III region: a parade of the centromeric segment with 21 genes, *Immunol. Today* **21**, 320–328 (2000).
42. C. Y. Yu, Molecular genetics of the human MHC complement gene cluster, *Exp. Clin. Immunogenet.* **15**, 213–230 (1998).

43. P. C. White, D. Grossberger, B. J. Onufer, M. I. New, B. DuPont and J. Strominger, Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man, *Proc. Natl. Acad. Sci. USA* **82**, 1089–1093 (1985).
44. M. C. Carroll, A. Palsdottir, K. T. Belt and R. R. Porter, Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region, *EMBO J.* **4**, 2547–2552 (1985).
45. J. Bristow, M. K. Tee, S. E. Gitelman, S. H. Mellon and W. L. Miller, Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene overlapping P450c21B, *J. Cell Biol.* **122**, 265–278 (1993).
46. L. Shen, L.-C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. W. Dangel, M. C. Carroll, W. B. Zipf and C. Y. Yu, Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region: molecular cloning, exon intron structure, composite retroposon, and breakpoint of gene duplication, *J. Biol. Chem.* **269**, 8466–8476 (1994).
47. Z. Yang, A. R. Mendoza, T. R. Welch, W. B. Zipf and C. Y. Yu, Modular variations of HLA class III genes for serine/threonine kinase RP, complement C4, steroid 21-hydroxylase CYP21 and tenascin TNX (RCCX): a mechanism for gene deletions and disease associations, *J. Biol. Chem.* **274**, 12147–12156 (1999).
48. P. M. Schneider, M. C. Carroll, C. A. Alper, C. Rittner, A. S. Whitehead, E. J. Yunis and H. R. Colten, Polymorphism of human complement C4 and steroid 21-hydroxylase genes. Restriction fragment length polymorphisms revealing structural deletions, homoduplications, and size variants, *J. Clin. Invest.* **78**, 650–657 (1986).
49. S. Simon, L. Truedsson, D. Marcus-Bagley, Z. Awdeh, G. S. Eisenbarth, S. J. Brink, E. J. Yunis and C. A. Alper, Relationship between protein complotypes and DNA variant haplotypes: complotypes-RFLP constellations (CRC), *Hum. Immunol.* **57**, 27–36 (1997).
50. A. W. Dangel, A. R. Mendoza, B. J. Baker, C. M. Daniel, M. C. Carroll, L.-C. Wu and C. Y. Yu, The dichotomous size variation of human complement C4 gene is mediated by a novel family of endogenous retroviruses which also establishes species-specific genomic patterns among Old World primates, *Immunogenetics* **40**, 425–436 (1994).
51. K. T. Belt, C. Y. Yu, M. C. Carroll and R. R. Porter, Polymorphism of human complement component C4, *Immunogenetics* **21**, 173–180 (1985).
52. K. A. Joiner, E. J. Brown and M. M. Frank, Complement and bacteria: chemistry and biology in host defense, *Annu. Rev. Immunol.* **2**, 461–491 (1984).
53. P. H. Pfeifer, J. J. Brems, M. Brunson and T. E. Hugli, Plasma C3a and C4a levels in liver transplant recipients: a longitudinal study, *Immunopharmacology* **46**, 163–174 (2000).
54. J. J. Alexander, A. Lim, C. He, R. L. MacDonald, V. M. Holers and R. J. Quigg, Renal, central nervous system and pancreatic overexpression of recombinant soluble Crry in transgenic mice. A novel means of protection from complement-mediated injury, *Immunopharmacology* **42**, 245–254 (1999).
55. B. P. Morgan and C. L. Harris. In: “Complement Regulatory Proteins”. San Diego, Academic Press, 1999.
56. S. I. Rosenfeld, S. Ruddy and K. F. Austin, Structural polymorphism of the fourth component of human complement, *J. Clin. Invest.* **48**, 2283–2292 (1969).
57. J. M. Moulds, F. C. Arnett, C. M. Giles and R. G. Hamilton, A novel immunoassay for the quantitation of human C4 gene products, *Complement Inflamm.* **7**, 95–101 (1990).
58. V. Rebmann, I. Doxiadis, B. S. Kubens and H. Grosse-Wilde, Quantitation of the human component C4: definition of C4 Q0 alleles and C4A duplications, *Vox Sang* **62**, 117–123 (1992).
59. P. Teisberg, I. Akesson, B. Olaisen, T. Gedde-Dahl Jr. and E. Thorsby, Genetic polymorphism of C4 in man and localisation of a structural C4 locus to the HLA gene complex of chromosome 6, *Nature* **264**, 253–254 (1976).

60. G. J. O'Neill, S. Y. Yang and B. DuPont, Two HLA-linked loci controlling the fourth component of human complement, *Proc. Natl. Acad. Sci. USA* **75**, 5165–5169 (1978).
61. Z. L. Awdeh, D. Raum and C. A. Alper, Genetic polymorphism of human complement C4 and detection of heterozygotes, *Nature* **282**, 205–208 (1979).
62. M. H. Roos, E. Mollenhauer, P. Demant and C. Rittner, A molecular basis for the two locus model of human complement component C4, *Nature* **298**, 854–856 (1982).
63. B. Olaisen, P. Teisberg and R. Jonassen, The C4 system: quantitative studies of different genotypes, *Immunobiology* **158**, 82–85 (1980).
64. P. Teisberg, R. Jonassen, B. Mevag, T. Gedde-Dahl, Jr. and B. Olaisen, Restriction fragment length polymorphisms of the complement component C4 loci on chromosome 6: studies with emphasis on the determination of gene number, *Ann. Hum. Genet.* **52**, 77–84 (1988).
65. D. Raum, Z. Awdeh, J. Anderson, L. Strong, J. Granados, L. Teran, E. Giblett, E. J. Yunis and C. A. Alper, Human C4 haplotypes with duplicated C4A or C4B, *Am. J. Hum. Genet.* **36**, 72–79 (1984).
66. C. Rittner, C. M. Giles, C. M. Roos, P. Demant and E. Mollenhauer, Genetics of human C4 polymorphism: detection and segregation of rare and duplicated haplotypes, *Immunogenetics* **19**, 321–333 (1984).
67. D. J. Schendel, R. Wank and G. J. O'Neill, C4 phenotypic variation suggests an unusual class III gene organization, *Vox Sang.* **48**, 110–115 (1985).
68. E. Sim and S. Cross, Phenotyping of human complement component C4, a class III HLA antigen, *Biochem. J.* **239**, 763–767 (1986).
69. C. Y. Yu, K. T. Belt, C. M. Giles, R. D. Campbell and R. R. Porter, Structural basis of the polymorphism of human complement component C4A and C4B: gene size, reactivity and antigenicity, *EMBO J.* **5**, 2873–2881 (1986).
70. M. C. Carroll, D. M. Fathallah, L. Bergamaschini, E. M. Alicot and D. E. Isenman, Substitution of a single amino acid (aspartic acid for histidine) converts the functional activity of human complement C4B to C4A, *Proc. Natl. Acad. Sci. USA* **87**, 6868–6872 (1990).
71. C. Y. Yu, C. A. Blanchong, E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou and J. M. Moulds. In: N. R. Rose, R. G. Hamilton and B. Detrick, (eds.), "Manuals of Clinical Laboratory Immunology," 6th ed. pp. 117–131. Washington, DC, ASM Press, 2002.
72. E. K. Chung, Y. Yang, K. L. Rupert, K. N. Jones, R. M. Rennebohm, C. A. Blanchong and C. Y. Yu, Determining the one, two, three or four long and short loci of human complement C4 in a major histocompatibility complex haplotype encoding for C4A or C4B proteins, *Am. J. Hum. Genet.* **71**, 810–822 (2002).
73. P. M. Schneider, K. Witzel-Schlomp, C. Rittner and L. Zhang, The endogenous retroviral insertion in the human complement C4 gene modulates the expression of homologous genes by antisense inhibition, *Immunogenetics* **53**, 1–9 (2001).
74. M. C. Carroll, T. Belt, A. Palsdottir and R. R. Porter, Structure and organization of the C4 genes, *Philos. Trans. R. Soc. Lond.* **306**, 379–388 (1984).
75. M. C. Carroll, K. T. Belt, A. Palsdottir and Y. Yu, Molecular genetics of the fourth component of human complement and the steroid 21-hydroxylase, *Immunol. Rev.* **87**, 39–60 (1985).
76. P. M. Schneider, C4 DNA RFLP reference typing report, *Complement Inflamm.* **7**, 218–224 (1990).
77. E. K. Chung, Y. Yang, R. M. Rennebohm, M. L. Lokki, G. C. Higgins, K. N. Jones, B. Zhou, C. A. Blanchong and C. Y. Yu, Genetic sophistication of human complement C4A and C4B and RP-C4-CYP21-TNX (RCCX) modules in the major histocompatibility complex (MHC), *Am. J. Hum. Genet.* **71**, 823–837 (2002).
78. C. A. Blanchong, E. K. Chung, K. L. Rupert, Y. Yang, Z. Yang, B. Zhou and C. Y. Yu, Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologs, Slp and C4, *Int. Immunopharmacol.* **1**, 365–392 (2001).

79. L. M. Shen, L. C. Wu, S. Sanlioglu, R. Chen, A. R. Mendoza, A. Dangel, M. C. Carroll, W. Zipf and C. Y. Yu, Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and C4B genes in the HLA class III region: molecular cloning, exon-intron structure, composite retroposon and breakpoint of gene duplication, *J. Biol. Chem.* **269**, 8466–8476 (1994).
80. Z. Yang, L. Shen, A. W. Dangel, L. C. Wu and C. Y. Yu, Four ubiquitously expressed genes, *RD* (D6S45)–*SKI2W* (SKIV2L)–*DOM3Z*–*RP1* (D6S60E), are present between complement component genes factor B and C4 in the class III region of the HLA, *Genomics* **53**, 338–347 (1998).
81. Z. Yang and C. Y. Yu, Organization and gene duplications of the human and mouse MHC complement gene clusters, *Exp. Clin. Immunogenet.* **17**, 1–17 (2000).
82. C. Y. Yu, The complete exon-intron structure of a human complement component C4A gene: DNA sequences, polymorphism, and linkage to the 21-hydroxylase gene, *J. Immunol.* **146**, 1057–1066 (1991).
83. H. Uejima, M. P. Lee, H. Cui and A. P. Feinberg, Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios, *Nat. Genet.* **25**, 375–376 (2000).
84. H. Kawaguchi, M. Golubic, F. Figueroa and J. Klein, Organization of the chimpanzee *C4-CYP21* region: implications for the evolution of humans genes, *Eur. J. Immunol.* **20**, 739–745 (1990).
85. D. A. Wiginton, D. J. Kaplan, J. C. States, A. L. Akesson, C. M. Perme, I. J. Bilyk, A. J. Vaughn, D. L. Lattier and J. J. Hutton, Complete sequence and structure of the gene for human adenosine deaminase, *Biochemistry* **25**, 8234–8244 (1986).
86. N. Martinez-Quiles, E. Paz-Artal, M. A. Moreno-Pelayo, J. Longas, S. Ferre-Lopez, M. Rosal and A. Arnaiz-Villena, C4d DNA sequences of two infrequent human allotypes (C4A13 and C4B12) and the presence of signal sequences enhancing recombination, *J. Immunol.* **161**, 3438–3443 (1998).
87. Y. Higashi, H. Yoshioka, M. Yamane, O. Gotoh and Y. Fujii-Kuriyama, Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene, *Proc. Natl. Acad. Sci. USA* **83**, 2841–2845 (1986).
88. P. C. White and P. W. Speiser, Congenital adrenal hyperplasia due to 21-hydroxylase deficiency, *Endocr. Rev.* **21**, 245–291 (2000).
89. X. Chu, L. Braun-Heimer, C. Rittner and P. M. Schneider, Identification of the recombination site within the steroid 21-hydroxylase gene (CYP21) of the HLA-B47, DR7 haplotype, *Exp. Clin. Immunogenet.* **9**, 80–85 (1992).
90. K. L. Rupert, R. M. Rennebohm and C. Y. Yu, An unequal crossover between the RCCX modules of the human MHC leading to the presence of a *CYP21B* gene and a tenascin *TNXB/TNXA-RP2* recombinant between *C4A* and *C4B* genes in a patient with juvenile rheumatoid arthritis, *Exp. Clin. Immunogenet.* **16**, 81–97 (1999).
91. J. Schalkwijk, M. Zweers, P. M. Steijen, W. B. Dean, G. Taylor, I. M. van Vlijmen, B. van Haren, W. L. Miller and J. Bristow, A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X deficiency, *New Engl. J. Med.* **345**, 1167–1175 (2001).
92. P. F. J. Koppens, T. Hoogenboezem and H. J. Degenhart, Carriership of a defective tenascin-X gene in steroid 21-hydroxylase deficiency patients: *TNXB-TNXA* hybrids in apparent large-scale gene conversions, *Hum. Mol. Genet.* **11**, 2581–2590 (2002).
93. T. Jaatinen, M. Eholouto, T. Laitinen and M.-L. Lokki, Characterization of a de novo conversion in human complement C4 gene producing a C4B5-like protein, *J. Immunol.* **168**, 5652–5658 (2002).
94. P. Sinnott, S. Collier, C. Costigan, P. A. Dyer, R. Harris and T. Strachan, Genesis by meiotic unequal crossover of a de novo deletion that contributes to steroid 21-hydroxylase deficiency, *Proc. Natl. Acad. Sci. USA* **87**, 2107–2111 (1990).

95. M. B. Fasano, J. A. Winkelstein, T. LaRosa, W. B. Bias and R. H. McLean, A unique recombination event resulting in a C4A*Q0,C4B*Q0 double null haplotype, *J. Clin. Invest.* **90**, 1180–1184 (1992).
96. Z. L. Awdeh, D. Raum, E. J. Yunis and C. A. Alper, Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man, *Proc. Natl. Acad. Sci. USA* **80**, 259–263 (1983).
97. R. L. Dawkins, F. T. Christiansen, P. H. Kay, M. Garlepp, J. McCluskey, P. N. Hollingsworth and P. J. Zilko, Disease associations with complotypes, supratypes and haplotypes, *Immunol. Rev.* **70**, 5–22 (1983).
98. M. Degli-Esposti, A. L. Leaver, F. T. Christiansen, C. S. Witt, L. J. Abraham and R. L. Dawkins, Ancestral haplotypes: conserved population MHC haplotypes, *Hum. Immunol.* **34**, 242–252 (1992).
99. I. Gigli, A single-chain precursor of C4 in human serum, *Nature* **272**, 836–837 (1978).
100. R. D. Schreiber and H. J. Muller-Eberhard, Fourth component of human complement: description of a three polypeptide chain structure, *J. Exp. Med.* **140**, 1324–1335 (1974).
101. R. E. Hall and H. R. Colten, Cell-free synthesis of the fourth component of guinea pig complement (C4): identification of a precursor of serum C4 (pro-C4), *Proc. Natl. Acad. Sci. USA* **74**, 1707–1710 (1977).
102. G. Goldberger and H. Colten, Precursor complement protein (pro-C4) is converted in vitro to native C4 by plasmin, *Nature* **286**, 514–516 (1980).
103. T. Seya, S. Nagasawa and J. P. Atkinson, Location of the interchain disulfide bonds of the fourth component of human complement (C4): evidence based on the liberation of fragments secondary to thiol-disulfide interchange reactions, *J. Immunol.* **136**, 4152–4156 (1986).
104. J. Janatova, Detection of disulfide bonds and localization of interchain linkages in the third (C3) and fourth (C4) components of human complement, *Biochem. J.* **233**, 819–825 (1986).
105. S. K. Law, N. A. Lichtenberg and R. P. Levine, Covalent binding and hemolytic activity of complement proteins, *Proc. Natl. Acad. Sci. USA* **77**, 7194–7198 (1980).
106. J. Janatova and B. F. Tack, Fourth component of human complement: studies of an amine-sensitive site comprised of a thiol component, *Biochemistry* **20**, 2394–2402 (1981).
107. B. F. Tack, R. A. Harrison, J. Janatova, M. L. Thomas and J. W. Prah, Evidence for presence of an internal thiolester bond in third component of human complement, *Proc. Natl. Acad. Sci. USA* **77**, 5764–5768 (1980).
108. D. R. Karp, Post-translational modification of the fourth component of complement. Effect of tunicamycin and amino acid analogs on the formation of the internal thiol ester and disulfide bonds, *J. Biol. Chem.* **258**, 12745–12748 (1983).
109. S. G. Davies and R. B. Sim, Internal general acid catalysis in the binding reactions of α_2 -macroglobulin and complement components C3 and C4, *Biosci. Rep.* **1**, 461–468 (1981).
110. J. W. Kehoe and C. R. Bertozzi, Tyrosine sulfation: a modulator of extracellular protein-protein interactions, *Chem. Biol.* **7**, R53–R61 (2000).
111. G. Hortin, H. Sims and A. W. Strauss, Identification of the site of sulfation of the fourth component of human complement, *J. Biol. Chem.* **261**, 1786–1793 (1986).
112. G. L. Hortin, T. C. Farries, J. P. Graham and J. P. Atkinson, Sulfation of tyrosine residues increase activity of the fourth component of complement, *Proc. Natl. Acad. Sci. USA* **86**, 1338–1342 (1989).
113. K. T. Belt, M. C. Carroll and R. R. Porter, The structural basis of the multiple forms of human complement component C4, *Cell* **36**, 907–914 (1984).
114. I. Gigli, I. VonZabern and R. R. Porter, The isolation and structure of C4, the fourth component of human complement, *Biochem. J.* **165**, 439–446 (1977).
115. A. C. Chan and J. P. Atkinson, Oligosaccharide structure of human C4, *J. Immunol.* **134**, 1790–1798 (1985).

116. D. R. Karp, K. L. Parker, D. C. Shreffler, C. Slaughter and J. Capra, Amino acid sequence homologies and glycosylation differences between the fourth component of murine complement and sex-limited protein, *Proc. Natl. Acad. Sci. USA* **79**, 6347–6349 (1982).
117. D. R. Karp, J. P. Atkinson and D. C. Shreffler, Genetic variation in glycosylation of the fourth component of murine complement. Association with hemolytic activity, *J. Biol. Chem.* **257**, 7330–7335 (1982).
118. D. C. Shreffler, J. P. Atkinson, A. C. Chan, D. R. Karp, C. C. Killion, R. T. Ogata and P. A. Rosa, The C4 and Slp genes of the complement region of the murine H-2 major histocompatibility complex, *Phil. Trans. R. Soc. Lond.* **B 306**, 395–403 (1984).
119. D. S. Sepich, D. J. Noonan and R. T. Ogata, Complete cDNA sequence of the fourth component of murine complement, *Proc. Natl. Acad. Sci. USA* **82**, 5895–5899 (1985).
120. M. Nonaka, K. Nakayama, Y. D. Yeul and M. Takahashi, Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4). Evolutionary aspects, *J. Biol. Chem.* **260**, 10936–10943 (1985).
121. C. Hemenway, M. Kalf, J. Stavenhagen, D. Walthall and D. Robins, Sequence comparison of alleles of the fourth component of complement (C4) and sex-limited protein (Slp), *Nucleic Acids Res.* **14**, 2539–2554 (1986).
122. L. Rowen, S. Qin, S. R. Lasky, C. Loretz, M. Dors, G. Mahairas and L. Hood, Mus musculus major histocompatibility complex locus class III region: complement C4 (C4) and cytochrome P450 hydroxylase A (CYP21OH-A) genes, complete cds; slp pseudogene, complete sequence, NC6, SKI, and complement factor B (Bf) genes, complete cds; and complement factor C2 (C2) gene, partial cds). GenBank, Accession no. AF049850, 1998.
123. A. C. Chan and J. P. Atkinson, Identification and structural characterization of two incompletely processed forms of the fourth component of human complement, *J. Clin. Invest.* **72**, 1639–1649 (1983).
124. S. K. A. Law and J. Gagnon, The primary structure of the fourth component of human complement (C4)-C-terminal peptides, *Biosci. Rep.* **5**, 913–921 (1985).
125. A. C. Chan and J. P. Atkinson, Functional studies on the secreted form of human C4 (C4^s), two incompletely processed two-subunit molecules (β - α + γ and β + α - γ), and pro-C4, *J. Immunol.* **132**, 1967–1971 (1984).
126. D. N. Chakravarti, R. D. Campbell and R. R. Porter, The chemical structure of the C4d fragment of the human complement component C4, *Mol. Immunol.* **24**, 1187–1197 (1987).
127. K. E. Moon, J. P. Gorski and T. E. Hugli, Complete primary structure of human C4a anaphylatoxin, *J. Biol. Chem.* **256**, 8685–8692 (1981).
128. R. D. Campbell, J. Gagnon and R. R. Porter, Amino acid sequence around the thiol and reactive acyl groups of human complement component C4, *Biochem. J.* **199**, 359–370 (1981).
129. E. M. Press and J. Gagnon, Human complement component C4: structural studies on the fragments derived from C4b by cleavage with C3b inactivator, *Biochem. J.* **199**, 351–357 (1981).
130. U. Hellman, G. Eggertsen, A. Lundwall, I. Engstrom and J. Sjoquist, *FEBS Lett.* **170**, 254–258 (1984).
131. M. C. Carroll and R. R. Porter, Cloning of a human complement component C4 gene, *Proc. Natl. Acad. Sci. USA* **80**, 264–267 (1983).
132. N. Martinez-Quiles, E. Paz-Artal, M. A. Moreno-Pelayo, J. Longas, S. Ferre-Lopez, M. Rosal and A. Arnaiz-Villena, C4d DNA sequences of two infrequent human allotypes (C4A13 and C4B12) and the presence of signal sequences enhancing recombination, *J. Immunol.* **161**, 3438–3443 (1998).
133. E. Paz-Artal, A. Corell, P. Varela, P. Morales, J. M. Martin-Villa, O. G. Segurado, M. Alvarez and A. Arnaiz-Villena, New DNA sequences for the human complement gene C4, *Mol. Immunol.* **30**, 515–516 (1993).

134. D. Ulgiati and L. J. Abraham, Comparative analysis of the disease-associated complement C4 gene from the HLA-A1, B8, DR3 haplotype, *Exp. Clin. Immunogenet.* **13**, 43–54 (1996).
135. A. W. Dodds, S.-K.A. Law and R. R. Porter, The purification and properties of some less common allotypes of the fourth component of human complement, *Immunogenetics* **24**, 279–285 (1986).
136. M. Nonaka, K. Nakayama, Y. D. Yeul and M. Takahashi, Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4), *J. Biol. Chem.* **260**, 10936–10943 (1985).
137. M. Nonaka, M. Takahashi, S. Natsume-Sakai, S. Tanaka, A. Shimizu and T. Honjo, Isolation of cDNA clones specifying the fourth component of mouse complement and its isotype, sex-limited protein, *Immunology* **81**, 6822–6826 (1984).
138. R. T. Ogata and D. S. Sepich, Murine sex-limited protein: complete cDNA sequence and comparison with murine fourth complement component, *J. Immunol.* **135**, 4239–4244 (1985).
139. R. T. Ogata, P. A. Rosa and N. E. Zepf, Sequence of the gene for murine complement component C4, *J. Biol. Chem.* **264**, 16565–16572 (1989).
140. H. Kawaguchi, Z. Zaleska-Rutczynska, F. Figueroa, C. O'hUigin and J. Klein, C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for extensive homogenization, *Immunogenetics* **35**, 16–23 (1992).
141. E. Paz-Artal, A. Corell, M. Alvarez, P. Varela, L. Allende, A. Mardrono, M. Rosal and A. Arnaiz-Villena, C4 gene polymorphism in primates: evolution, generation, and Chido and Rodgers antigenicity, *Immunogenetics* **40**, 381–396 (1994).
142. X.-D. Ren, A. W. Dodds and S. K. A. Law, The thioester and isotypic sites of complement component C4 in sheep and cattle, *Immunogenetics* **37**, 120–128 (1993).
143. A. W. Dodds and S. K. Law, The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and α_2 macroglobulin, *Immunol. Rev.* **166**, 15–26 (1998).
144. J. P. Atkinson and P. M. Schneider. In: R. G. Lahita (ed.), “Systemic Lupus Erythematosus”, pp. 91–104. San Diego, Academic Press, 1999.
145. J. P. Atkinson, A. C. Chan, D. R. Karp, C. C. Killion, R. Brown, D. Spinella, D. C. Shreffler and R. P. Levine, Origins of the fourth component of complement related Chido and Rodgers blood-group antigens, *Complement* **5**, 65–76 (1988).
146. G. J. O'Neill, S. Y. Yang, J. Tegoli, R. Berger and B. DuPont, Chido and Rodgers blood groups are distinct antigenic components of human complement C4, *Nature* **273**, 668–670 (1978).
147. P. Lambin, P. Y. Pennec, G. Hauptmann, O. Desaint, B. Habibi and Ch. Salmon, Adverse transfusion reactions associated with a precipitating anti-C4 antibody of anti-Rodgers specificity, *Vox Sang.* **47**, 242–249 (1984).
148. C. M. Giles and J. L. Swanson, Anti-C4 in the serum of a transfused C4-deficient patient with systemic lupus erythematosus, *Vox Sang.* **46**, 291–299 (1984).
149. C. M. Giles, K. A. Davies and M. J. Walport, In vivo and in vitro binding of C4 molecules on red cells: a correlation of numbers of molecules and agglutination, *Transfusion* **31**, 222–228 (1991).
150. P. R. Taylor, M. C. Pickering, M. Kosco-Vilbois, M. J. Walport, M. Botto, S. Gordon and L. Martinex-Pomares, The follicular dendritic cell restricted epitope, FDC-M2, is complement C4; localization of immune complexes in mouse tissues, *Eur. J. Immunol.* **32**, 1888–1896 (2002).
151. M. H. Roos, C. M. Giles, P. Demant, E. Mollenhauer and C. Rittner, Rodgers (Rg) and Chido (Ch) determinants on human C4: characterization of two C4 B5 subtypes, one of which contains Rg and Ch determinants, *J. Immunol.* **133**, 2634–2640 (1984).
152. J. M. H. van den Elsen, A. Martin, V. Wong, L. Clemenza, D. R. Rose and D. E. Isenman, X-ray crystal structure of the C4d fragment of human complement component C4, *J. Mol. Biol.* **322**, 1103–1115 (2002).

153. C. M. Giles and J. W. Jones, A new antigenic determinant from C4 of relatively low frequency, *Immunogenetics* **26**, 392–394 (1987).
154. C. M. Giles, Antigenic determinants of human C4, Rodgers and Chido, *Exp. Clin. Immunogenet.* **5**, 99–114 (1988).
155. B. Nagar, R. G. Jones, R. J. Diefenbach, D. E. Isenman and J. M. Rini, X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2, *Science* **280**, 1277–1281 (1998).
156. G. Szakonyi, J. M. Guthridge, D. Li, K. Young, M. Holers and X. S. Chen, Structure of complement receptor 2 in complex with its C3d ligand, *Science* **292**, 1725–1728 (2001).
157. Y. U. Kim, M. C. Carroll, D. E. Isenman, M. Nonaka, P. Pramoongjago, J. Takeda, K. Inoue and T. Kinoshita, Covalent binding of C3b to C4b within the classical complement pathway C5 convertase, *J. Biol. Chem.* **267**, 4171–4176 (1992).
158. P. Tiesberg, B. Olaisen, R. Nordhagen, E. Thorsby and T. Gedde-Dahl, A hemolytically non-active C4 gene C4 gene product, *Immunobiology* **158**, 91–95 (1980).
159. G. J. O'Neill, P. Minter, M. S. Pollock and B. DuPont, HLA antigen associations for the functionally active and inactive products of the complement F1 allele, *Hum. Immunol.* **1**, 23–30 (1980).
160. T. Kinoshita, A. W. Dodds, S. K. A. Law and K. Inoue, The low C5 convertase activity of the C4A6 allotype of human complement component C4, *Biochem. J.* **261**, 743–748 (1989).
161. M. J. Anderson, C. M. Milner, R. G. H. Cotton and R. D. Campbell, The coding sequence of the hemolytically inactive C4A6 allotype of human complement component C4 reveals that a single arginine to tryptophan substitution at β -chain residue 458 is the likely cause of the defect, *J. Immunol.* **148**, 2795–2802 (1992).
162. R. O. Ebanks, A. S. I. Jaikaran, M. C. Carroll, M. J. Anderson, R. D. Campbell and D. E. Isenman, A single arginine to tryptophan interchange at β -chain residue 458 of human complement component C4 accounts for the defect in classical pathway C5 convertase activity of allotype C4A6, *J. Immunol.* **148**, 2803–2811 (1992).
163. R. H. McLean, G. Niblack, B. Julian, T. Wang, R. Wyatt, J. A. Philips, III, T. S. Collins, J. Wickelstein and D. Valle, Hemolytically inactive C4B complement allotype caused by a proline to leucine mutation in the C5-binding site, *J. Biol. Chem.* **269**, 27727–27731 (1994).
164. S. Meri and M. K. Pangburn, A mechanism of activation of the alternative complement pathway by the classical pathway: protection of C3b from inactivation by covalent attachment to C4b, *Eur. J. Immunol.* **20**, 2555–2561 (1990).
165. R. O. Ebanks and D. E. Isenman, Mouse complement component C4 is devoid of classical pathway C5 convertase subunit activity, *Mol. Immunol.* **33**, 297–309 (1996).
166. K. S. Park, S. Y. Choi, M. H. Park and K. Tokunaga, Allotypes of the fourth component of complement in Korean, *Jpn. J. Hum. Genet.* **37**, 285–292 (1992).
167. G. Mauff, Application of the MHC-class III complement markers to population genetics, *Colloque INSERM* **142**, 143–166 (1986).
168. K. S. Park, M. H. Park, T. Juji and K. Tokunaga, Complement C4A, C4B and BF haplotypes in Koreans, *Tissue Antigens* **47**, 200–205 (1996).
169. M. Cislo, J. Halasa, F. Wasik, P. Nockowski, M. Prussak, M. Manczak and P. Kusnierczyk, Allelic distribution of complement components BF, C4A, C4B, and C3 in Psoriasis vulgaris, *Immunol. Lett.* **80**, 145–149 (2002).
170. R. J. Wyatt, C. Wang, E. C. Hudson, J. M. Jones, P. W. Noah and N. Rosenberg, Bf and C4 phenotypes in patients with psoriasis, *Acta Derm. Venereol. (Stockh)* **146**, 211–213 (1989).
171. T. E. Hugli, Biochemistry and biology of anaphylatoxins, *Complement* **3**, 111–127 (1986).
172. J. Greer, Comparative structural anatomy of the complement anaphylatoxin proteins C3a, C4a, and C5a, *Enzyme* **36**, 150–163 (1986).

173. R. S. Ames, M. A. Tornetta, J. J. Foley, T. E. Hugli and H. M. Sarau, Evidence that the receptor for C4a is distinct from the C3a receptor, *Immunopharmacology* **38**, 87–92 (1997).
174. S. Matsubara, T. Yamamoto, T. Tsuruta, K. Takagi and T. Kambara, Complement C4-derived monocyte-directed chemotaxis-inhibitory factor, *Am. J. Pathol.* **138**, 1279–1291 (1991).
175. T. Tsuruta, T. Yamamoto, S. Matsubara, S. Nagasawa, S. Tanese, J. Tanaka, K. Takagi and T. Kambara, Novel function of C4a anaphylatoxin: release from monocytes of protein which inhibits monocyte chemotaxis, *Am. J. Pathol.* **142**, 1848–1857 (1993).
176. Y. Murakami, T. Yamamoto, T. Imamichi and S. Nagasawa, Cellular responses of guinea-pig macrophages to C4a; inhibition of C3a-induced O₂-generation by C4a, *Immunol. Lett.* **36**, 301–304 (1993).
177. S. Lienenklaus, R. S. Ames, M. A. Tornetta, H. M. Sarau, J. J. Foley, T. Crass, B. Sohns, U. Raffetseder, M. Grove, A. Holzer, A. Klos, J. Kohl, W. Bautsch, Human anaphylatoxin C4a is a potent agonist of the guinea pig but not the human C3a receptor, *J. Immunol.* **161**, 2089–2093 (1998).
178. G. Mauff, M. Steuer, M. Weck and K. Bender, The C4 β -chain: evidence for a genetically determined polymorphism, *Hum. Genet.* **64**, 186–188 (1983).
179. M. Steuer, G. Mauff, C. Adam, M. P. Baur, K. Bender, J. Goetz, S. F. Goldmann, G. Hauptmann, M. Neugebauer, M.-M. Tongio, B. Lambert-Uring, A. Wolph, An estimate on the frequency of duplicated haplotypes and silent alleles of human C4 protein polymorphism. I. Investigations in healthy Caucasoid families, *Tissue Antigens* **33**, 501–510 (1989).
180. T. Robson, R. N. S. Heard and C. M. Giles, An epitope on the C4 β light (L) chains detected by human anti-Rg; its relationship with β -chain polymorphism and MHC associations, *Immunogenetics* **30**, 344–349 (1989).
181. R. H. McLean, C. M. Giles, C. Y. Yu, R. D. Campbell and W. B. Bias, Reversed restriction sites for the C4 isotype and Rodgers (1)/Chido (1) regions on two unique C4 genes, *Tissue Antigens* **35**, 75–81 (1990).
182. D. R. Karp, D. C. Shreffler and J. P. Atkinson, Characterization of the Mr difference between secreted murine fourth component of complement and the major plasma form: evidence for carboxyl-terminal cleavage of the alpha chain, *Proc. Natl. Acad. Sci. USA* **79**, 6666–6670 (1982).
183. D. R. Karp and D. C. Shreffler, S region genetic control of murine C4 biosynthesis: analysis of pro-C4 cleavage, *Immunogenetics* **16**, 171–176 (1982).
184. D. R. Karp, J. D. Capra, J. P. Atkinson and D. C. Shreffler, Structural and functional characterization of an incompletely processed form of murine C4 and Slp, *J. Immunol.* **128**, 2336–2341 (1982).
185. G. Hoecker, O. Pizarro and A. Ramos, Some new antigens and histocompatibility factors in the mouse, *Transplant Bull.* **6**, 407 (1959).
186. A. Ferreira, C. S. David and V. Nussenzweig, The murine H-2.7 specificity is an antigenic determinant of C4d, a fragment of the fourth component of the complement system, *J. Exp. Med.* **151**, 1424–1435 (1980).
187. D. G. Spinella and H. C. Passmore, A newly defined murine alloantigen controlled by the S region of the H-2 complex: molecular association with the fourth component of complement, *J. Immunol.* **130**, 824–828 (1983).
188. P. A. Taillon-Miller and D. C. Shreffler, Structural basis for the C4d.1/C4d.2 serologic allotypes of murine complement component C4, *J. Immunol.* **141**(7), 2382–2387 (1988).
189. P. Mathias, D. C. Shreffler, N. R. Cooper and R. T. Ogata, Conversion of the C4d.2 serologic allotype of murine complement component C4 to the C4d.1 allotype by site-specific mutagenesis, *J. Immunol.* **144**(2), 607–609 (1990).

190. M. B. Fischer, M. Ma, S. Goerg, X. Zhou, J. Xia, O. Finco, S. Han, G. Kelsoe, R. G. Howard, T. L. Rothstein, E. Kremmer, F. S. Rosen, M. C. Carroll, Regulation of the B cell response to T-dependent antigens by classical pathway complement, *J. Immunol.* **157**, 549–556 (1996).
191. M. R. Wessels, P. Butko, M. Ma, H. B. Warren, A. L. Lage and M. C. Carroll, Studies of group B streptococcal infection in mice deficient in complement component C3 or C4 demonstrate an essential role for complement in both innate and acquired immunity, *Proc. Natl. Acad. Sci. USA* **92**, 11490–11494 (1995).
192. A. P. Prodeus, S. Goerg, L. M. Shen, O. O. Pozdnyakova, L. Chu, E. M. Alicot, C. C. Goodnow and M. C. Carroll, A critical role for complement in maintenance of self-tolerance, *Immunity* **9**, 721–731 (1998).
193. Z. Chen, S. B. Koralov and G. Kelsoe, Complement C4 inhibits systemic autoimmunity through a mechanism independent of complement receptors CR1 and CR2, *J. Exp. Med.* **192**, 1339–1351 (2000).
194. E. Paul, O. O. Pozdnyakova, E. Mitchell and M. C. Carroll, Anti-DNA autoreactivity in c4-deficient mice, *Eur. J. Immunol.* **32**, 2672–2679 (2002).
195. T. R. Welch, M. Frenzke, M. C. Carroll and D. P. Whitte, Evidence of a role for C4 in modulating interstitial inflammation in experimental glomerulonephritis, *Clin. Immunol.* **101**, 366–370 (2002).
196. S. Einav, O. O. Pozdnyakova, H. Ma and M. C. Carroll, Complement C4 is protective for lupus disease independent of C3, *J. Immunol.* **168**, 1036–1041 (2002).
197. M. Gadjeva, A. Verschoor, M. A., Brockman, H. Jezak, L. M. Shen, D. M. Knipe and M. C. Carroll, Macrophage-derived complement component c4 can restore humoral immunity in C4-deficient mice, *J. Immunol.* **169**, 5489–5495 (2002).
198. L. Ellman, I. Green, F. Judge and M. M. Frank, In vivo studies in C4-deficient guinea pigs, *J. Exp. Med.* **134**, 162–175 (1971).
199. R. K. Root, L. Ellman and M. M. Frank, Bactericidal and opsonic properties of C4-deficient guinea pig serum, *J. Immunol.* **109**, 477–486 (1972).
200. E. C. Bottger, T. Hoffman, U. Hadding and D. Bitter-Suermann, Influence of genetically inherited complement deficiencies on humoral immune response in guinea pigs, *J. Immunol.* **135**, 4100–4107 (1985).
201. H. D. Ochs, R. J. Wedgwood, M. M. Frank, S. R. Heller and S. W. Hosea, The roles of complement in the induction of antibody responses, *Clin. Exp. Immunol.* **53**, 208–216 (1983).
202. E. C. Bottger, T. Hoffmann, U. Hadding and D. Bitter-Suermann, Guinea pigs with inherited deficiencies of complement components C2 or C4 have characteristics of immune complex disease, *J. Clin. Invest.* **78**, 689–695 (1986).
203. D. Bitter-Suermann and R. Burger, Guinea pigs deficient in C2, C4, C3 or C3a receptor, *Prog. Allergy* **39**, 134–158 (1986).
204. L. Ellman, I. Green and M. Frank, Genetically controlled total deficiency of the fourth component of complement in the guinea pig, *Science* **170**, 74–75 (1970).
205. C. G. Jackson, H. D. Ochs and R. J. Wedgwood, Immune response of a patient with deficiency of the fourth component of complement and systemic erythematosus, *New Engl. J. Med.* **300**, 1124–1129 (1979).
206. J. P. Atkinson and J. A. Schifferli. In: G. M. Kammer and G. C. Tsokos, (eds.), “Lupus, Molecular and Cellular Pathogenesis”, pp. 529–540. Totowa, Humana Press, 1999.
207. NCBI Annotation Project. *Homo sapiens* chromosome 6 genomic contig, alternative haplotype DR51. GenBank, Accession number NT_034874, 2002.
208. L. Rowen, C. Dankers, D. Baskin, J. Faust, C. Loretz, M. E. Ahearn, A. Banta, S. Schwartzell, T. M. Smith, T. Spies and L. Hood, *Homo sapiens* HLA class III region containing tenascin X gene, partial cds, cytochrome P450 21-hydroxylase (CYP21B), complement C4 (C4B), G11,

- helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes, complete cds. GenBank, Accession no. AF019413, 1997.
209. K. Barlow, Human DNA sequences from clone RP1-34f7 on chromosome 6p21.2-21.33. GenBank, Accession no. AL049547, 2000.
 210. K. L. Rupert, J. M. Moulds, Y. Yang, F. C. Arnett, R. W. Warren, J. D. Reveille, B. L. Myones, C. A. Blanchong and C. Y. Yu, The molecular basis of complete C4A and C4B deficiencies in a systemic lupus erythematosus (SLE) patient with homozygous C4A and C4B mutant genes, *J. Immunol.* **169**, 1570–1578 (2002).
 211. G. N. Fredrikson, B. Gullstrand, P. M. Schneider, K. Witzel-Schlomp, A. G. Sjöholm, C. A. Alper, Z. Awdeh and L. Truedsson, Characterization of non-expressed C4 genes in a case of complete C4 deficiency: identification of a novel point mutation leading to a premature stop codon, *Hum. Immunol.* **59**, 713–719 (1998).
 212. M.-L. Lokki, A. Circolo, P. Ahokas, K. L. Rupert, C. Y. Yu and H. R. Colten, Deficiency of complement protein C4 due to identical frameshift mutations in the C4A and C4B genes, *J. Immunol.* **162**, 3687–3693 (1998).
 213. G. Barba, C. Rittner and P. M. Schneider, Genetic basis of human complement C4A deficiency: detection of a point mutation leading to nonexpression, *J. Clin. Invest.* **91**, 1681–1686 (1993).
 214. H. R. Colten and G. Garnier. In: J. E. Volanakis and M. M. Frank, (eds.), “The Human Complement System in Health and Disease”, pp. 217–240. New York, Marcel Dekker, 1998.
 215. J. E. Volanakis, Transcriptional regulation of complement genes, *Annu. Rev. Immunol.* **13**, 277–305 (1995).
 216. H. R. Colten, Drawing a double-edged sword, *Nature* **371**, 474–475 (1994).
 217. H. Feucht, J. Zwirner, D. Bevec, M. Lang, E. Felber, G. Riethmüller and E. Weiss, Biosynthesis of complement C4 messenger RNA in normal human kidney, *Nephron* **53**, 338–342 (1989).
 218. J. Kulics, H. R. Colten and D. H. Perlmuter, Counterregulatory effects of interferon-gamma and endotoxin on expression of the human C4 genes, *J. Clin. Invest.* **85**, 943–949 (1990).
 219. H. Tsukamoto, K. Nagasawa, S. Yoshizawa, Y. Tada, A. Ueda, Y. Ueda and Y. Niho, Synthesis and regulation of the fourth component of complement (C4) in the human monocytic cell line U937: comparison with that of the third component of complement, *Immunology* **75**, 565–569 (1992).
 220. H. Tsukamoto, K. Nagasawa, Y. Ueda, T. Mayumi, I. Furugo, T. Tsuru and Y. Niho, Effects of cell differentiation on the synthesis of the third and fourth component (C3, C4) by the human monocytic cell line U937, *Immunology* **77**, 621–623 (1992).
 221. J. Laufer, I. Goldberg, R. Oren, A. Horwitz, J. Kopolovic, Y. Chowers and J. H. Passwell, Cellular localization of complement C3 and C4 transcripts in intestinal specimens from patients with Crohn's disease, *Clin. Exp. Immunol.* **120**, 30–37 (2000).
 222. A. Andoh, Y. Fujiyama, T. Bamba and S. Hosoda, Differential cytokine regulation of complement C3, C4, and factor B synthesis in human intestinal epithelial cell line, Caco-2, *J. Immunol.* **151**, 4239–4247 (1993).
 223. T. R. Welch, L. S. Beischel and D. P. Witte, Differential expression of complement C3 and C4 in the human kidney, *J. Clin. Invest.* **92**, 1451–1458 (1993).
 224. J. J. Timmerman, C. L. Verweij, D. J. vanGijlswijk-Janssen, F. J. vanderWoude, L. A. vanEs and M. R. Daha, Cytokine-regulated production of the major histocompatibility complex class-III-encoded complement proteins Factor B and C4 by human glomerular mesangial cells, *Hum. Immunol.* **43**, 19–28 (1995).
 225. W. Zhou, R. D. Campbell, J. Martin and S. H. Sacks, Interferon- γ regulation of C4 gene expression in cultured human glomerular epithelial cells, *Eur. J. Immunol.* **23**, 2477–2481 (1993).

226. J. Kulics, A. Circolo, R. C. Strunk and H. R. Colten, Regulation of synthesis of complement protein C4 in human fibroblasts: cell- and gene-specific effects of cytokines and lipopolysaccharide, *Immunology* **82**, 509–515 (1994).
227. P. F. Kohler and H. J. Müller-Eberhardt, Immunochemical quantitation of the third, fourth and fifth components of human complement: concentrations of sera in healthy adults, *J. Immunol.* **99**, 1211–1216 (1967).
228. R. A. J. Nelson, J. M. Jensen, E. Gigli and N. Tamura, Methods for the separation, purification and measurement of nine components of hemolytic complement in guinea pig serum, *Immunochemistry* **3**, 111–135 (1966).
229. K. T. Belt, The structural genes of the human complement component C4. D. Phil. Thesis, University of Oxford, 1984.
230. L. C. Wu, Studies on the expression of the complement genes of the HLA. D. Phil. Thesis, University of Oxford, 1987.
231. D. Ulgiati, L. S. Subrata and L. J. Abraham, The role of Sp family members, basic kruppel-like factor, and E box factors in the basal and IFN- γ regulated expression of the human complement C4 promoter, *J. Immunol.* **164**, 300–307 (2000).
232. A. K. Vaishnav, T. J. Mitchell, J. Rose, M. J. Walport and B. J. Morley, Regulation of transcription of the TATA-less human complement component C4 gene, *J. Immunol.* **160**, 4353–4360 (1998).
233. M. K. Tee, A. A. Thomson, J. Bristow and W. L. Miller, Sequences promoting the transcription of the human XA gene overlapping P450c21A correctly predict the presence of a novel, adrenal-specific, truncated form of tenascin-X, *Genomics* **28**, 171–178 (1995).
234. J. B. Stavenhagen and D. M. Robins, An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene, *Cell* **55**, 247–254 (1988).
235. J. P. Atkinson, D. R. Karp, E. P. Seeskin, C. C. Killion, P. A. Rosa, S. L. Newell and D. C. Shreffler, H-2 S region determined polymorphic variants of the C4, Slp, C2, and B complement proteins: a compilation, *Immunogenetics* **16**, 617–623 (1982).
236. K. L. Parker, J. P. Atkinson, M. H. Roos and D. C. Shreffler, Genetic and structural characterization of H-2-controlled allelic forms of murine C4, *Immunogenetics* **11**, 55–63 (1980).
237. J. H. Zheng, S. Natsuume-Saki, M. Takahashi and M. Nonaka, Insertion of the B2 sequence into intron 13 is the only defect of the H-2^k C4 gene which causes low C4 production, *Nucleic Acids Res.* **20**, 4975–4979 (1992).
238. S. Pattanakitsakul, J.-H. Zheng, S. Natsuume-Sakai, M. Takahashi and M. Nonaka, Aberrant splicing caused by the insertion of the B2 sequence into an intron of the complement C4 gene is the basis for low C4 production in H-2^k mice, *J. Biol. Chem.* **267**, 7814–7820 (1992).
239. J. H. Zheng, M. Takahashi and M. Nonaka, Tissue-specific RNA processing for the complement C4 gene transcript in the H-2^k mouse strain, *Immunogenetics* **37**, 390–393 (1993).
240. J. Klein, A case of no sex-limitation of Slp in the murine H-2 complex, *Immunogenetics* **2**, 297–299 (1975).
241. S. Pattanakitsakul, K. Nakayama, M. Takahashi and M. Nonaka, Three extra copies of a C4-related gene in H-2^{w7} mice are C4/Slp hybrid genes generated by multiple recombinational events, *Immunogenetics* **32**, 431–439 (1990).
242. M. Levi-Strauss, E. Georgatsoy, M. Tosi and T. Meo, Gene-specific probes demonstrate selective duplications of the C4-Slp gene in the H-2S alleles associated with a testosterone-independent expression of this isotype, *Immunogenetics* **21**, 397–401 (1985).
243. M. Levi-Strauss, M. Tosi, M. Steinmetz, J. Klein and T. Meo, Multiple duplications of complement C4 gene correlate with H-2-controlled testosterone-independent expression of its sex-limited isoform, C4-Slp, *Proc. Natl. Acad. Sci. USA* **82**, 1746–1750 (1985).

244. R. D. Miller, Y. Aizawa and D. C. Shreffler, Restriction fragment length polymorphism of the murine *C4* and *Slp* genes: two *C4* groups, *J. Immunol.* **149**, 2926–2934 (1992).
245. L. Rowen, S. Qin, A. Madan, N. Abbasi, R. James, R. Dickhoff, T. Shaffer, A. Ratcliffe, C. Loretz, S. Lasky and L. Hood, Sequence of the mouse major histocompatibility locus class III region. GenBank, Accession nos. AF030001, AF049850, 1997.
246. T. Heidmann, O. Heidmann and J.-F. Nicolas, An indicator gene to demonstrate intracellular transposition of defective retroviruses, *Proc. Natl. Acad. Sci. USA* **85**, 2219–2223 (1988).
247. M. Tassabehji, T. Strachan, M. Anderson, R. D. Campbell, S. Collier and M. Lako, Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement *C4* gene cluster, *Nucleic Acids Res.* **22**, 5211–5217 (1994).
248. X. Chu, C. Rittner and P. M. Schneider, Length polymorphism of the human complement component *C4* gene is due to an ancient retroviral integration, *Exp. Clin. Immunogenet.* **12**, 74–81 (1995).
249. A. W. Dangel, B. J. Baker, A. R. Mendoza and C. Y. Yu, Complement component *C4* gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution, *Immunogenetics* **42**, 41–52 (1995).
250. N. Miura, H. L. Prentice, P. M. Schneider and D. H. Perlmutter, Synthesis and regulation of the two human complement *C4* genes in stable transfected mouse fibroblasts, *J. Biol. Chem.* **262**, 7298–7305 (1987).
251. K. Nakagawa and L. Harrison, The potential roles of endogenous retroviruses in autoimmunity, *Immunol. Rev.* **152**, 193–236 (1996).
252. J. D. Mountz and N. Talal, Retroviruses, apoptosis and autogenes, *Immunol. Today* **14**, 532–536 (1993).
253. M. A. Pani, J. P. Wood, K. Bieda, R. R. Toenjes, K. H. Usadel and K. Badenhop, The variable endogenous retroviral insertion in the human complement *C4* gene: a transmission study in type I diabetes mellitus, *Hum. Immunol.* **63**, 481–484 (2002).
254. S. F. Altschul, A. A. Schäffer, T. L. Madden, J. Zhang, Z. Zhang, D. J. Lipman and W. Miller, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**, 3389–3402 (1997).
255. Genetics Computer Group, GCG Sequence Analysis Software Package: Program Manual, Version 7. Genetics Computer Group, Inc., Madison, 1991.
256. R. Strausberg, Homo sapiens hypothetical protein MGC15873. GenBank, Accession no. NM_032920, 2002.
257. K. Ninomiya, M. Wagatsuma, K. Kanda, H. Kondo, T. Yokoi, H. Kodaira, T. Furuya, M. Takahashi *et al.* Homo sapiens cDNA FLJ31940 (full length insert sequence). GenBank, AK056502, 2002.
258. A. L. Hughes and M. Yeager, Molecular evolution of the vertebrate immune system, *Bioessays* **19**, 777–786 (1997).
259. P. B. Armstrong, The contribution of proteinase inhibitors to immune defense, *Trends Immunol.* **22**, 47–52 (2001).
260. J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural and G. G. Sutton, The sequence of the human genome, *Science* **291**, 1304–1351 (2001).
261. International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921 (2001).
262. K. Devriendt, H. Van den Berghe, J. J. Cassiman and P. Marynen, Primary structure of pregnancy zone protein. Molecular cloning of a full-length PZP cDNA clone by the polymerase chain reaction, *Biochim. Biophys. Acta* **1088**, 95–103 (1991).

263. T. Nagase, K. Ishikawa, R. Kikuno, M. Hirosawa, N. Nomura and O. Ohara, Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro, *DNA Res.* **7**, 337–345 (1999).
264. M. F. Flajnik and M. Kasahara, Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system, *Immunity* **15**, 351–362 (2001).
265. L. Banyai and L. Patthy, The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases, *Protein Sci.* **8**, 1636–1642 (1999).
266. G. Hortin, A. C. Chan, K. F. Rok, A. W. Strause and J. P. Atkinson, Sequence analysis of the COOH terminus of the α -chain of the fourth component of human complement, *J. Biol. Chem.* **261**, 9065–9069 (1986).
267. W. S. Argraves, H. Tran, W. H. Burges and H. Dickerson, Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure, *J. Cell Biol.* **111**, 3146–3155 (1990).
268. R.-Z. Zhang, T.-C. Pan, Z.-Y. Zhang, M.-G. Mattei, R. Timpl and M.-L. Chu, Fibulin (FBLN2): human cDNA sequence, mRNA expression, and mapping of the gene on human and mouse chromosomes, *Genomics* **22**, 425–430 (1994).
269. A. Sandoval, R. Ai, J. M. Ostresh and R. T. Ogata, Distal recognition site for the classical pathway convertase located in the C345C/Netrin module of complement component C5, *J. Immunol.* **165**, 1066–1073 (2000).
270. L. Banyai and L. Patthy, The NTR module: domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases, *Protein Sci.* **8**, 1636–1642 (1999).
271. M. Lagueus, E. Perrodou, E. A. Levashina, M. Capovilla and J. A. Hoffmann, Constitutive expression of a complement-like protein in Toll and JAK gain-of-function mutants of *Drosophila*, *Proc. Natl. Acad. Sci. USA* **21**, 11427–11432 (2001).
272. D. A. Kimbrell and B. Beutler, The evolution and genetics of innate immunity, *Nat. Rev. Genet.* **2**, 256–267 (2001).
273. J. A. Hoffmann, F. C. Kafatos, C. A. Janeway, Jr. and R. A. B. Ezekowitz, Phylogenetic perspectives in innate immunity, *Science* **284**, 1313–1318 (2002).
274. M. Nonaka, K. Azumi, X. Ji, C. Namikawa-Yamada, M. Saasaki, H. Saiga, A. W. Dodds, H. Sekine, M. K. Homma, Y. Matsushita, T. Fujita, Opsonic complement component C3 in the solitary ascidian *Halocynthia roretzi*, *J. Immunol.* **162**, 387–391 (1999).
275. J. O. Sunyer, I. K. Zarkadis and J. D. Lambris, Complement diversity: a mechanism for generating immune diversity, *Immunol. Today* **19**, 519–523 (1998).
276. M. Nonaka. In: L. Du Pasquier and G. W. Litman, (eds.), “Origin and Evolution of the Vertebrate System”, pp. 37–50. Berlin, Springer, 2000.
277. C. Leelayuwat, W. J. Zhang, L. J. Abraham, D. Townsend, S. Gaudieri and R. L. Dawkins, Differences in the central major histocompatibility complex between humans and chimpanzees, *Hum. Immunol.* **38**, 30–41 (1993).
278. N. Kuroda, K. Naruse, A. Shima, M. Nonaka and M. Sasaki, Molecular cloning and linkage analysis of complement C3 and C4 genes of the Japanese medaka fish, *Immunogenetics* **51**, 117–128 (2000).
279. M. Nakao, J. Mutsuro, R. Obo, K. Fujiki, M. Nonaka and T. Yano, Molecular cloning and protein analysis of divergent forms of the complement component C3 from a bony fish, the common carp (*Cyprinus carpio*): presence of variants lacking the catalytic histidine, *Eur. J. Immunol.* **30**, 858–866 (2000).
280. H. Kawaguchi, C. O’hUigin and J. Klein. In: J. Klein and D. Klein, (eds.), “Molecular Evolution of the Major Histocompatibility Complex”, pp. 357–381. Heidelberg, Springer-Verlag, 1991.

281. C. Ramakrisman and D. M. Robins, Steroid hormone responsiveness of a family of closely related mouse proviral elements, *Mamm. Genome* **8**, 811–817 (1998).
282. A. J. Adler, M. Manielsen and D. M. Robins, Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors, *Proc. Natl. Acad. Sci. USA* **89**, 11660–11663 (1992).
283. L. Rowen, G. Mahairas, S. Qin, M. E. Ahearn, C. Dankers, S. Lasky, C. Loretz, S. Schmidt, S. Tipton, R. Traicoff, K. Zackrone and L. Hood, Mus musculus major histocompatibility locus class III region: butyrophilin-like protein gene, partial cds; Notch4, PBX2, RAGE, lysophatidic acid acyl transferase-alpha, palmitoyl-protein thioesterase 2 (PPT2), CREB-RP, and tenascin X (TNX) genes, complete cds; and CYP21OHB pseudogene, complete sequence. GenBank, Accession no. AF030001, 1997.
284. P. A. Rosa, D. S. Sepich, D. C. Shreffler and R. T. Ogata, Mice constitutive for sex limited (Slp) expression contain multiple Slp gene sequences, *J. Immunol.* **135**, 627–631 (1985).
285. M. Levi-Strauss, M. Tosi, M. Steinmetz, J. Klein and T. Meo, Multiple duplications of complement C4 gene correlate with H-2-controlled testosterone-independent expression of its sex-limited isoform, C4-Slp, *Proc. Natl. Acad. Sci. USA* **82**, 1746–1750 (1985).
286. L. Walter, P. Hurt, H. Himmelbauer, R. Sudbrak and E. Gunther, Physical mapping of the major histocompatibility complex class II and class III regions of the rat, *Immunogenetics* **54**, 268–275 (2002).
287. Y. Horiuchi, H. Kawaguchi, F. Figueroa, C. O'hUigin and J. Klein, Dating the primigenial C4-CYP21 duplication in primates, *Genetics* **134**, 331–339 (1993).
288. W. J. Zhang, F. T. Christiansen, X. Wu, L. J. Abraham, M. Giphart and R. L. Dawkins, Organization and evolution of C4 and CYP21 genes in primates: importance of genomic segments, *Immunogenetics* **37**, 170–176 (1993).
289. R. E. Bontrop, L. A. M. Broos, N. Otting and M. J. Jonker, Polymorphism of C4 and CYP21 genes in various primate species, *Tissue Antigens* **37**, 145–151 (1991).
290. S. Georg, The association between systemic lupus erythematosus and deficiencies of the complement system, *Cell. Mol. Biol.* **48**, 237–245 (2002).
291. C. H. Nielsen and R. G. Q. Leslie, Complement's participation in acquired immunity, *J. Leukoc. Biol.* **72**, 249–261 (2002).
292. G. Wild, J. Watkins, A. M. Ward, P. Hughes, A. Hume and N. R. Rowell, C4a anaphylatoxin levels as an indicator of disease activity in systemic lupus erythematosus, *Clin. Exp. Immunol.* **80**, 167–170 (1990).
293. S. R. de Cordoba, O. C. Gacia and P. Sanchez-Corral. In: B. J. Morley and M. J. Walport, (eds.), "The Complement Factsbook", pp. 161–167. San Diego, Academic Press, 2000.
294. D. J. Birmingham and L. A. Hebert, CR1 and CR1-like: the primate immune adherence receptors, *Immunol. Rev.* **180**, 100–111 (2001).
295. A. C. Chan, D. R. Karp, D. C. Shreffler and J. P. Atkinson, The 20 faces of the fourth component of complement, *Immunol. Today* **5**, 200–203 (1984).
296. R. O. Ebanks and D. E. Isenman, Evidence for the involvement of Arginine 462 and the flanking sequence of human C4 β -chain in mediating C5 binding to the C4b subcomponent of the classical complement pathway C5 convertase, *J. Immunol.* **154**, 2808–2820 (1995).
297. Q. Pan, R. O. Ebanks and D. E. Isenman, Two clusters of acidic amino acids near the NH₂ terminus of complement component C4 α' -chain are important for C2 binding, *J. Immunol.* **165**, 2518–2527 (2000).

The Roles and Regulation of Potassium in Bacteria

WOLFGANG EPSTEIN

*Department of Molecular Genetics and
Cell Biology, The University of Chicago,
Chicago, IL 60637, USA*

I. Paths of K^+ Movement	296
II. Regulation of K^+ Transport Systems	305
III. K As An Intracellular Activator	309
IV. Regulation of Internal pH by K^+	311
V. Future Directions	312
References	313

Potassium is the major intracellular cation in bacteria as well as in eucaryotic cells. Bacteria accumulate K^+ by a number of different transport systems that vary in kinetics, energy coupling, and regulation. The Trk and Kdp systems of enteric organisms have been well studied and are found in many distantly related species. The Ktr system, resembling Trk in many ways, is also found in many bacteria. In most species two or more independent saturable K^+ -transport systems are present. The KefB and KefC type of system that is activated by treatment of cells with toxic electrophiles is the only specific K^+ -efflux system that has been well characterized. Pressure-activated channels of at least three types are found in bacteria; these represent nonspecific paths of efflux when turgor pressure is dangerously high. A close homolog of eucaryotic K^+ channels is found in many bacteria, but its role remains obscure.

K^+ transporters are regulated both by ion concentrations and turgor. A very general property is activation of K^+ uptake by an increase in medium osmolarity. This response is modulated by both internal and external concentrations of K^+ . Kdp is the only K^+ -transport system whose expression is regulated by environmental conditions. Decrease in turgor pressure and/or reduction in external K^+ rapidly increase expression of Kdp. The signal created by these changes, inferred to be reduced turgor, is transmitted by the KdpD sensor kinase to the KdpE-response regulator that in turn stimulates transcription of the *kdp* genes.

K^+ acts as a cytoplasmic-signaling molecule, activating and/or inducing enzymes and transport systems that allow the cell to adapt to elevated osmolarity. The signal could be ionic strength or specifically K^+ . This signaling response is probably mediated by a direct sensing of internal ionic strength by each particular system and not by a component or system that coordinates this response by different systems to elevated K^+ . © 2003 Elsevier Science

Potassium (K^+), an element necessary to life, is the major monovalent intracellular cation, reflecting a very ancient decision of living cells to accumulate K^+ and to exclude the smaller sodium (Na^+) ion. These ions act directly as cations since none of them form covalent bonds in aqueous environments. K^+ has four roles in bacteria: (i) an osmotic solute, (ii) an activator of intracellular enzymes, (iii) a regulator of internal pH, and (iv) a second messenger.

The marked diversity of structure and metabolism in the bacterial kingdom is not reflected in a similar diversity of ways bacteria deal with K^+ . Only a few types of specific K^+ -transport systems are found in bacteria, many are widely distributed in eubacteria and some are found in the distantly related archaea. Only for a few species do we have detailed knowledge of the way K^+ is transported and what its specific roles are. Data is most extensive for two enteric organisms, *Escherichia coli* and *Salmonella typhimurium*. Where no species is mentioned, the data are for one of these two closely related organisms. More extensive coverage of the older literature on K^+ can be found in a review (1).

Bacteria can be divided into two groups based on how they respond to growth at elevated osmolarity. A number of halophilic species, including eubacterial and archaeal species, use K^+ and to some extent Na^+ as major cytoplasmic-osmotic solutes regardless of medium osmolarity, behavior referred to as the "salt-in" adaptation (2). These species can grow only in media containing high concentrations of NaCl and have cytoplasmic contents adapted to high ionic strength. A large number of studies have examined the adaptations needed to allow enzymes to function well at high ionic strength (reviewed in Ref. (3)). However, there are few studies of K^+ transport in the halophiles. Most organisms, by contrast, use K^+ as a primary osmotic solute only during growth at low osmolarity. During growth at higher osmolarity, one of a number of neutral, polar molecules, "compatible solutes" is accumulated (4). K^+ appears to act as a second messenger to stimulate accumulation of compatible solutes.

The development of spectroscopic analysis by flame emission and later by atomic absorption in the middle of the last century has made analysis of inorganic cations sensitive and specific. There are also convenient radioactive tracers for virtually all cations. The only generally useful tracer for K^+ is $^{42}K^+$, which has an inconveniently short half-life of only 12.4 h. The larger congener, $^{86}Rb^+$ with a half-life of 18.8 days, is useful for most eucaryotic K^+ -transport systems, but is not a good quantitative tracer for most bacterial K^+ -transport systems since they discriminate against Rb^+ .

An important feature of bacteria with walls is the maintenance of an osmotic pressure difference across the cell envelope, creating a hydrostatic pressure called turgor pressure. Turgor, as used here, refers to both the

pressure across the cell envelope as well as the tension it creates in the envelope. It is generally believed that turgor is essential for growth and is important in determining the shape of bacteria (5). Turgor is both a signal and an end product of K^+ transport: changes in cell K^+ concentration alter turgor while turgor is implicated as the signal-controlling activity of K^+ -transport systems and the expression of one of them, Kdp. Since all of the proteins that create the solute gradients required for turgor, as well as those believed to respond to turgor, are in the inner (cytoplasmic) membrane of enteric and other Gram-negative bacteria, turgor is expected to be exerted across the inner membrane. An alternate view suggests that turgor is expressed across the outer membrane of Gram-negative bacteria (6,7). There is expected to be some turgor across the outer membrane, since fixed anionic charges on periplasmic macromolecules require cations to match, and this creates an electrical Donnan potential. Porins in the outer membrane allow small molecules free access to the periplasm, so ions will distribute themselves across the outer membrane at thermodynamic equilibrium and not at equal concentration. Since ions associated with macromolecules have low osmotic activity (8), the fraction of turgor expected to be expressed across the outer membrane is probably not large. Gram-positive bacteria have no outer membrane, so the equivalent model would have turgor expressed across their cell wall. It is virtually certain that a Donnan potential could not create the high turgor pressures of Gram-positive bacteria, as high as 20 atm in *Bacillus subtilis* (9). Therefore, turgor in Gram-positive bacteria, must be exerted across their cytoplasmic membrane. The suggestion that turgor in Gram-negative bacteria is expressed across the outer membrane implies that Gram-positive bacteria have a radically different way of establishing turgor although they share many of the systems that adjust and respond to turgor with Gram-negative bacteria. A number of other considerations also suggest that turgor for the most part is exerted across the inner membrane in Gram-negative bacteria (10).

The magnitude of turgor is expected to be fairly constant in growing cells if turgor is needed for growth, an expectation supported by the fact that the same degree of osmotic downshock was required to the release of K^+ in cells growing at high as well as low osmolality (11). Recent measurements of cell solutes and water suggest that turgor is much lower in medium of higher osmolality (7). The calculated value of turgor of 3.1 atm at low osmolality fell to about 0.4 atm at an osmolality of 0.5 osM. The implication that bacteria can grow at such low values of turgor seems incompatible with the stress-theory of bacterial growth. The questions about the site where turgor is expressed and its magnitude must be resolved by further studies.

Vesicle systems make it possible to control the composition of the solution on both sides of the membrane, something not possible when whole cells are

used. However, in studying the effects of the solution facing the cytoplasmic face of the membrane, salts of unphysiologic anions such as chlorides and phosphates are often used. While some studies where the physiological anion, glutamate, was also used suggest the choice of anion is not crucial (see Section IV), it would be well to include glutamate since a number of cytoplasmic enzymes have altered the behavior with glutamate as compared to other anions (12).

I. Paths of K^+ Movement

A. Measuring K^+ Transport in Bacteria

The rate of K^+ uptake into cells depends on how it is measured. Bacteria generally regulate their cytoplasmic pool of K^+ rather tightly. The sole K^+ movement in nongrowing cells in the steady state is an exchange of external for internal K^+ , a process whose measurement requires use of a radioactive tracer. Net uptake during growth reflects the low rate needed for growth and therefore does not reflect the capacity of transport systems. Net uptake limited by the capacity of transport systems can be measured in cells that have been depleted of K^+ , or following an increase in medium osmolarity, a procedure referred to as osmotic upshock. Uptake in K^+ -depleted cells is a convenient but artificial condition, while stimulation of net K^+ uptake by osmotic upshock reflects a basic physiological response of many K^+ -transport systems.

Valid data require that the methods used to deplete cells of K^+ do not damage the transport systems under study. A variety of methods has been used to deplete bacteria of K^+ , including treatment with high concentrations of 2,4-dinitrophenol (13), growth into the stationary phase (14), treatment with a divalent cation-chelating agent (15), treatment with a permeant amine at high pH (16), treatment with suitable ionophores (17), and washing with low osmolarity solutions free of K^+ (18,19). The rapid loss of K^+ after treatment with 2,4-dinitrophenol, a classical protonophore, suggests that when present at high concentrations it acts as an ionophore for small, monovalent cations (W. Epstein, unpublished).

B. Influx Systems

1. THE Trk SYSTEM

The system found in the largest number of bacterial species, including several archaeal species, is a multicomponent complex referred to as Trk. These systems have a moderate affinity for K^+ in the vicinity of 1 mM, relatively high specificity for K^+ , are generally expressed constitutively and are probably energized by taking up K^+ in symport with a proton. Our

knowledge of these systems is limited to characterization of the kinetics of transport *in vivo* and identification of the genes necessary to their activity; as yet there are no reports of Trk activity in vesicles.

The Trk system of *E. coli* is the one most extensively studied. The products of three different and unlinked genes are necessary for Trk activity in *E. coli* and similar enterics: *trkA*, *trkE*, and *trkH* (20,21). The corresponding genes in *S. typhimurium* are *sapG*, *sapD*, and *sapJ*, respectively (22). The *trkA* gene encodes a 458-residue peripheral membrane protein, the product of a gene-duplication event. The TrkA protein has a region similar to the dinucleotide fold of dehydrogenases, and TrkA binds NADH and NAD with higher affinity for NADH (23,24). The *trkE* gene, now often referred to as *sapD*, is one of the two ATP-binding components of an ABC transporter called Sap and first identified in *S. typhimurium* (25). This designation arose because the Sap transporter was proposed to transport toxic peptides. Subsequent studies indicate that *sapD* and other genes of the Trk system are required because restoration of internal K^+ is necessary to allow the cell to resist toxic peptides (26). The *trkH* gene is a 483-amino acid protein predicted to span the membrane 10 times, and thus is the component of Trk that forms the transmembrane-path component for K^+ movement (27).

The Trk system prefers K^+ over other monovalent cations. Cs^+ is not measurably transported (28), while Rb^+ is a substrate but with about a 10-fold lower rate of transport (29). External pH has little effect on the affinity of Trk for K^+ , but the affinity for Rb^+ increases about 10-fold as the pH rises from 6.1 to 8.4 (30). Studies of energetics indicate that Trk activity requires both the proton-motive force (PMF) and ATP (31). The current view is that transport is driven by the PMF with one proton entering with one K^+ ion and that ATP has a regulatory role (32). The finding that the *trkE* (*sapD*) gene product is functional in mutant forms that are expected to bind ATP but not to hydrolyze it, supports the idea that ATP is regulatory and does not provide the energy to drive transport via Trk (33).

A complication, unique to some strains of *E. coli* K-12 that carry a defective prophage called *rac*, is the presence of TrkG, a second version of the transmembrane protein component of the Trk system. TrkG is very similar to TrkH, being 41% identical at the amino acid level, but its low GC content (37%) suggests an origin in some distant species of bacteria (34). TrkG replaces TrkH so that either product allows full Trk activity (21). The specificity of the system depends on which of these two membrane-spanning components is involved (27). However, when TrkG is present, Trk activity is only partially dependent on TrkE (21) although ATP is still required (33), suggesting that an ATP-binding component of some other ABC transporter can replace the role of TrkE (SapD) when TrkG is present. It is conceivable, if not very plausible, that TrkA is the component that binds ATP; failure to

detect binding of ATP to TrkA at submicromolar concentrations (24) does not exclude binding of ATP with modest affinity.

The Trk systems in other bacteria that have been studied seem to resemble that of *E. coli*. Introduction of a plasmid encoding the TrkA + TrkH components from *Vibrio alginolyticus* into *E. coli* results in a system that is totally independent of *sapD*, but does require ATP (33,35). Thus, a third component that binds ATP, plausibly from some other ABC transporter, may be required for the *V. alginolyticus* Trk system. The thermophilic *Alicycobacillus acidocaldarius* (formerly *Bacillus acidocaldarius*) has a K^+ -transport system whose kinetic properties resemble those of Trk (18).

K^+ uptake in the extreme halophiles, such as *Halobacterium halobium* and *Halofeax volcanii*, is probably mediated by Trk-type systems, since there are homologs of *trk* genes in several of the archaea. This finding solves the confusion about how this class of organisms takes up and maintains such high internal concentrations of K^+ . It had been suggested that uniport, uptake driven solely by the membrane potential, was sufficient (36), but the extent of accumulation is too high to be achieved by uniport (2,37).

2. THE Ktr SYSTEM

A similar system, found in many bacteria but not seen in those archaea whose genomes have been sequenced, comprises two components, KtrA and KtrB (38). KtrA is a distant relative of TrkA while KtrB is distantly related to TrkH. KtrA is about half the size of TrkA, and thus probably evolved from a precursor common to Ktr and Trk; TrkA arose as a gene duplication of this precursor. KtrA, like TrkA, has a region whose sequence is homologous to the NAD-binding domain of some dehydrogenases. The Ktr system of *Vibrio alginolyticus* is Na^+ dependent (39), suggesting that it is energized by $K^+ + Na^+$ symport instead of the $H^+ + K^+$ symport inferred for the Trk system. Study of other Ktr systems is needed to determine if coupling to Na^+ is generally the case, or whether Ktr in marine organisms uses Na^+ while Ktr in nonmarine forms uses H^+ as the coupling ion.

A study of the KtrA protein from *B. subtilis* and a similar protein from *Methanococcus jannaschi*, both containing a NAD/NADH-binding domain, showed that they form tetramers in solution in the presence of either nucleotide but aggregate in the absence of nucleotide (40). A 136-residue fragment of the *B. subtilis* protein with bound NADH and a 144-residue fragment with bound NAD were crystallized and their structures solved to a resolution of 2.3 Å. NAD was bound somewhat differently from the binding of NADH. The authors suggest a conformational change depending on which form of NAD was bound, but it is also possible that each protein binds NAD or NADH in the same way and that binding to one protein is different from that to the other. Studies of the KtrA protein from *V. alginolyticus* suggest that

ATP binds to this subunit with much higher affinity than do NAD or NADH (M. Willenborg and E. P. Bakker, personal communication). If this is true for other KtrA proteins, they may not be sensing the redox state of the cell.

Plasmids expressing the *ktrA* and *ktrB* genes from *V. alginolyticus* mediate K^+ transport in *E. coli* independent of any *trk* gene of the host. This system transports Rb^+ at the same maximal rate as K^+ , but affinity for Rb^+ is about 5-fold lower. K^+ uptake in *Enterococcus hirae* (formerly *Streptococcus fecalis*) at modest K^+ concentrations is now known to be via one of two Ktr systems in this species (41). K^+ uptake in this organism requires both the PMF and ATP (42). However, since this organism can only generate a PMF from ATP via the F_0F_1 -ATPase, it is unclear whether the ATP is required solely to generate a PMF or also has a direct role in transport.

Bacillus subtilis has two Ktr systems that were discovered through the similarity of their sequences to those in *V. alginolyticus* (19). These have been cloned and derivatives that lack one or both have been constructed. One of these systems, KtrAB, has rather high affinity for K^+ in the range of 1 mM while the KtrCD system has an apparent affinity of about 10 mM. When both are mutated, K^+ concentrations of 25 mM or higher are required to achieve growth. Expression of the systems is independent of medium K^+ concentration, but transcription of each is elevated when the other systems are abolished by mutation. Each system is important for adaptation to elevated osmolarity in media containing 2–4 mM K^+ ; adaptation to high osmolarity is restored if medium K^+ concentration is suitably increased.

3. THE Kdp SYSTEM

The third system found in *E. coli* and many other bacteria is Kdp, an inducible system with high affinity and specificity for K^+ (see Ref. (13), for reviews see Refs. (43,44)). Reports of K^+ transporting ATPases in other bacteria (45,46) are most likely due to Kdp since this system is very widely distributed. Kdp is a P-type ATPase consisting of four subunits, all membrane proteins. KdpF is a small hydrophobic peptide probably important for stability and perhaps assembly but not essential for function (47). KdpA has 557 amino acid residues, crosses the membrane 10 times and is the subunit that binds and moves K^+ across the membrane (48,49). KdpB has 682 residues, is homologous to the large subunit of other P-type ATPases and has the highly conserved phosphorylation site (50). KdpC is a 190-residue protein with one predicted membrane span; one identified function of KdpC is in the assembly of the Kdp complex (51). Like other P-type ATPases, only ATP and a divalent cation are required; transport by Kdp is independent of the PMF (31). Measurements of a partial reaction of Kdp in black lipid membranes indicate

that K^+ uptake via Kdp is not associated with efflux of another ion since the activity is electrogenic (52). Kdp is expressed only when the cell's need for K^+ is not satisfied by other transport systems, and thus serves to scavenge the ion when present at concentration too low to be efficiently taken up by the other K^+ -transport systems. Kdp becomes important for growth at higher K^+ concentrations when other K^+ -transport systems are abolished by mutation. Kdp is the only bacterial K^+ -transport system whose expression is strongly regulated at the transcriptional level, a topic considered in Section IV.

4. OTHER K^+ -UPTAKE SYSTEMS

A constitutive transport system of modest affinity and rate found in some bacteria is Kup, formerly TrkD. It is a single membrane protein of 622 amino acid residues and 12 predicted membrane spans, and has a large C-terminal soluble domain of 182 residues. It probably functions by proton symport (53) and is the only saturable system with reasonable affinity for Cs^+ (28). Kup activity is more important at low pH where its maximum rate exceeds that of Trk (54). Kup activity is inhibited at elevated osmolarity (Ref. (55), W. Epstein, unpublished observations).

Mutants in which all three saturable systems, Trk, Kup, and Kdp, are inactivated by mutation have residual K^+ -uptake activity referred to as TrkF (13). These mutants require about 25-mM K^+ to achieve half-maximal growth, and fail to grow in media containing less than about 10-mM K^+ . The rate of uptake by this system is virtually linear with external K^+ concentration. This activity requires the PMF but not ATP (31). Recent work suggests that this activity is due to illicit transport of K^+ by many different systems whose primary substrates are not K^+ (E. Buurman, J. Naprstek, D. McLaggan and W. Epstein, unpublished). Other examples of what is presumably illicit transport of K^+ have been reported (56,57). Complementation of the transport defect of such mutant strains by plasmids carrying foreign DNA has been used to isolate K^+ -transport systems in other bacteria and some eucaryotes.

A system of modest affinity for K^+ has been identified in *Enterococcus hirae* (58). This system seems to be essential for growth at very alkaline pH (pH 10) and has a modest affinity for K^+ ($K_m = 20$ mM) but a rate comparable to the Ktr systems in this organism. Since this system is saturable and has a rather high rate, it must be mediated by a system specific for K^+ and thus not an example of TrkF-like activity.

The *tetL* gene in *B. subtilis* and the *tetK* gene in a plasmid from *Staphylococcus aureus* encode tetracycline antiporters that can exchange various monovalent cations (59), and similar results have been reported for a divalent cation antiporter (60). When present on a multicopy plasmid, they partially complement mutants of *E. coli* lacking all saturable K^+ -uptake

systems, behavior suggesting that part of their activity represents illicit transport analogous to the TrkF activity in *E. coli*. When these two Tet transporters are expressed in vesicles, they mediate antiporter activities in which K^+ is a substrate (59). *tetA(L)*-deletion mutants of *B. subtilis* have reduced growth yields in media containing less than 10-mM K^+ (61). These data are difficult to reconcile with the observation that when the two Ktr systems of *B. subtilis* are deleted, no growth occurs below 20-mM K^+ (19). Since the loss of the Tet elements in *B. subtilis* reduced growth yields at intermediate K^+ concentrations, their primary role may be other than in K^+ transport. Growth of *B. subtilis* in limiting K^+ concentrations did not alter the kinetics of K^+ transport, implying that this species does not have a K^+ -transport system induced by growth at low K^+ concentrations (W. Epstein, unpublished observations).

A K^+ -stimulated ATPase induced by growth at low K^+ concentrations and inhibited by low concentrations of vanadate has been described in *B. subtilis* (62), but there is as yet no evidence that this activity is involved in K^+ transport.

C. Efflux Systems

Energetic and physiologic considerations predict specific systems mediating efflux of K^+ will be found in almost all bacteria. One argument, an energetic one, is based on the high electrical membrane potential, interior negative, maintained by bacteria growing in neutral and alkaline media. This potential will tend to accumulate K^+ through leakage paths until the K^+ gradient is in equilibrium with membrane potential. This is the presumed mechanism of TrkF activity described above. A membrane potential of -120 mV would concentrate K^+ and other monovalent cations 100-fold; if the external concentration is 20 mM the cell would have 2 M K^+ at equilibrium, a situation not compatible with growth of most bacteria. This type of consideration led Mitchell to predict the existence of proton/cation antiporters to avoid inhibition by excess accumulation of cations (63). Most bacteria grow well in media of elevated K^+ concentration, so an energy-coupled efflux system must be present to prevent excessive accumulation of K^+ . Acidophiles that have a reversed membrane potential, interior positive, by contrast need to protect themselves from anion flooding and thus must have anion-exporting systems.

A second argument is based on the requirement of bacteria to reduce their K^+ content when turgor is dangerously high, as in a reduction of medium osmolarity referred to as osmotic downshock or simply downshock here. Gradual K^+ efflux is seen when medium osmolarity is progressively diluted (64) or when cells at high osmolarity are supplied with compatible solutes whose uptake requires loss of cell K^+ to avoid excessive levels of turgor (65).

Very rapid efflux of K^+ produced by sudden downshock is surely mediated by mechanosensitive channels considered in Section II.C, but paths of K^+ efflux when turgor rises more slowly remain to be identified. Paths more specific than the channels are likely since efflux under these conditions appears to be rather specific.

The only well-characterized K^+ -efflux systems in *E. coli* are the KefB and KefC systems. Homologous systems are found in many Gram-negative bacteria but not in any Gram-positive bacteria (66). The activity of these systems is manifested by the rapid but reversible loss of K^+ when cells were treated with *N*-ethylmaleimide (67,68). The major effect is an exchange of K^+ for H^+ (69). Each of these systems consists of a large protein with an N-terminal region that has 12 predicted membrane spans and a larger C-terminal domain, probably cytoplasmic, that has a nucleotide-binding region (70). The C-terminal domain does not have homology to glutathione-binding proteins as reported earlier; there was an error in the reported sequence of the glutathione-binding protein (I. R. Booth, personal communication). Each of these two systems has, in addition, a smaller ancillary soluble protein, YheR and YabF, each encoded by genes adjacent to *kefB* and *kefC*, respectively. YheR and YabF have sequence similarity to some dehydrogenases. In the absence of the ancillary proteins, the systems have very low activity (71). Regulation of the activity of the KefB and KefC systems by glutathione adducts and potential regulation by the redox state of the cell is discussed in Section II.B.

K^+/H^+ antiporters are ideal candidates to mediate K^+ efflux, and such activity is probably present in virtually all bacteria. K^+/H^+ activity has been demonstrated in *E. coli* (72). This activity has moderate affinity for K^+ , K_m of about 2.5 mM, and low substrate specificity since Na^+ , Li^+ , and Rb^+ compete with K^+ . A mutant lacking such activity has been isolated; its only apparent defect was inability to grow at high osmolarity and high pH (73). A mutant of *E. hirae* has been isolated that is totally defective in K^+/H^+ antiport activity at pH 9.5, does not properly regulate internal pH above 8 and does not grow above pH 9 (74). These data suggest that bacteria may, in general, have two or more antiporters, each operating optimally over a different range of medium pH. This pattern resembles that of the two major Na^+/H^+ antiporters of *E. coli*, which are active over different ranges of pH (75).

An activity that exchanges K^+ for methylammonium and which is competitively inhibited by NH_4^+ and Tl^+ , but noncompetitively inhibited by K^+ has been characterized in *E. coli* (76). Similar activities have been reported in a thermophile (18) and a methanogen (77). Nothing is known about the genetics or physiological function of these exchange activities; they may reflect the activity of an NH_4^+ -transport system.

D. Channels

Bacteria have several types of channels with properties similar to those of eucaryotic cells. Only brief consideration of channels will be dealt with here, since they seem to have only a minor role in the way bacteria deal with K^+ . There have been suggestions that some of the systems considered here under the heading of transport systems are channels, but evidence of channel activity of any of these systems remains to be demonstrated. Bacterial channels are the subjects of a recent review (78).

Two bacterial K^+ channels have been characterized and a gene that encodes a bacterial homolog of eucaryotic channels has been identified. The *Streptomyces lividans* KcsA channel and the *Methanobacterium thermoautotrophicum* MthK channel have been purified, crystallized, and their structures determined with a resolution of 3.2 and 3.3 Å, respectively (79,80). The former channel appears to be gated by protons while the latter is Ca^{2+} gated. Studies of the ionic preference of the KcsA channel show that it has highest conductance for K^+ , somewhat lower conductance for similar ions Rb^+ , Tl^+ , and NH_4^+ , and over 1000-fold lower conductance for Na^+ , Li^+ , and Cs^+ (81). The MthK channel is reported to prefer K^+ to Na^+ (80). The KcsA structure has served as a model for the many ion channels of eucaryotic cells.

The *kch* gene of *E. coli* encodes a protein homologous to eucaryotic K^+ channels (82). There are homologs of this gene in many other bacteria. The Kch protein does not appear to form functional channels in vesicles or lipid bilayers under conditions tested so far since no reports of channel activity of the Kch protein have appeared. Although the *kch* gene is expressed (78), deletion of the gene has no detectable phenotype (*E. Buurman and W. Epstein, unpublished observations*). Overexpression of Kch in wild-type *E. coli* results in growth inhibition in media of moderate K^+ concentration (83). Inhibition is overcome by K^+ or several congeners, NH_4^+ , Cs^+ , or Rb^+ , when present at 30–90 mM concentration. Na^+ does not reverse the inhibition. The Kch protein did not form channels when expressed in oocytes, nor did chimeras in which the pore region of Kch was introduced into the Shaker channel or vice versa. Inhibition at moderate but not high-medium K^+ concentrations implies that Kch does not form a channel of the usual sort. These data are more consistent with an electrically neutral process, symport of K^+ with an anion, or exchange of K^+ for protons or some other cation.

The physiological roles of the bacterial K^+ channels are not known. The frequent openings and closings of the KcsA channel is typical of channels, but what function that would serve is unclear. Brief openings would tend to bring the membrane potential toward the equilibrium potential for K^+ , but

since it also conducts NH_4^+ it could lead to accumulation of the latter. Physiologic studies of strains in which KcsA or MthK are deleted might be informative. Studies of *kch*-deletion mutants have not been informative so far.

The mechanosensitive channels of bacteria open when their threshold pressure is exceeded, allowing efflux of a relatively nonspecific group of molecules, selected on the basis of size and to some extent by charge. Since rapid efflux of K^+ and other osmotic solutes is necessary to avoid cell rupture upon sudden osmotic downshock, these types of channels are an important if nonspecific path for K^+ efflux. Three types of channels have been identified in *E. coli*: MscL, MscS, and MscM, encoding channels of large, medium, and small diameter. The pressure to open the channels increases with the size of the channel opening.

The MscL channel has been crystallized from *Mycobacterium tuberculosis* and found to be pentameric with an estimated diameter in the open state of about 40 Å which should allow proteins of 40,000–70,000 Da to pass (84). There are two MscS channels; the major one, encoded by the *mscS* (former *yggB*) gene, is a 286-amino acid protein with three membrane spans. This MscS channel has also been crystallized to a resolution of 3.9 Å (85). The protein is a heptamer with, as predicted from the sequence, three membrane spans in each monomer. The *kefA* gene encodes an 1120-amino acid protein, of which a region toward the C-terminus is very similar to MscS. Channel activity of KefA, also called MscK, is dependent on external K^+ (86). The wild-type channel opens only by pressure and only when high concentrations of external K^+ are present, while the *kefA2* mutant channel opens spontaneously in the presence of external K^+ but with pressure whether K^+ is present or not. The *kefA2* mutant channel opens at a lower pressure than does the wild type (86,87). Growth of *kefA2* mutants is inhibited in the presence of high external K^+ and betaine (88). Efflux via bacterial-mechanosensitive channels is inhibited by Gd^{3+} (89) as is true of eucaryotic-mechanosensitive channels.

The presumed role of mechanosensitive channels, to protect the cells from cell rupture upon sudden osmotic downshock, has been confirmed. After osmotic downshock, mutants lacking both MscL and MscS have a survival of somewhat under 10% while over 90% of wild-type cells survive (90). Cells lacking only MscL or only MscS survive as well as the wild type after downshock. The MscS and the MscK channels have similar properties *in vitro*, but loss of the MscK channel has little effect and its presence is not protective. Curiously, even in the presence of high external K^+ , the MscK channel is not protective (I. R. Booth, personal communication).

II. Regulation of K^+ Transport Systems

A. Control of Activity of Influx Systems

Over half a century ago, it was reported that increasing the osmolarity of the medium, osmotic upshock, resulted in net K uptake by *E. coli* (91). This response to upshock may be universal since it is seen in diverse species (92–97). This response is crucial to many other effects of upshock discussed in the following sections, since they seem to be triggered by increased internal K^+ or some effect dependent on uptake of K^+ .

Kinetic analyses of the Trk system of *E. coli* show that K^+ uptake at a maximum rate begins without measurable lag (<5 s) after osmotic upshock (Ref. (92), W. Epstein, unpublished observations). Net uptake is due to a stimulation of influx; exchange of external for internal K^+ continues unaltered after upshock (64). The rate of uptake is maximal with osmotic shifts of 0.2 osM and higher (64,92). A similar but quantitatively smaller effect is mediated by the Kdp system. Similar studies have not been reported for the Kup system. In strains that have only the residual TrkF activity, osmotic upshock results in net K^+ uptake, but here the effect appears to be an inhibition of efflux (64). The stimulus that produced net uptake of K^+ in response to osmotic upshock was suggested to be a reduction in turgor (92,98). This model explains the transient nature of the increase in the rate of uptake, since once turgor was restored there would be no net uptake of K^+ . This model also accounts for the maintenance of a constant internal pool of K^+ during growth, in that turgor is assumed to be slightly suboptimal during growth to produce a slight stimulation of net K^+ uptake sufficient to meet the needs during growth. Once growth stops, turgor rises slightly to the point where net K^+ uptake ceases (98).

Each of the K^+ -transport activities identified in *E. coli* is sensitive to the ionic environment. The Trk system is maximally active in cells that are K depleted. The depleted cells have increased cell Na^+ and probably lower internal pH, but neither can account for a high rate of transport, since low pH is inhibitory (99) while internal Na^+ does not appear to be very stimulating (14). Kdp activity is strongly but reversibly inhibited by external K^+ concentrations above 0.1 mM (92). Such inhibition may account at least in part for reduced growth rates in media containing 10–40 mM K^+ of strains expressing only Kdp (16,100). TrkF activity, probably a mixture of paths, exhibits transstimulation, being higher in cells that are loaded with K^+ than in K^+ -depleted cells (92). Transstimulation might be indirect since K^+ -depleted cells may have a somewhat lower membrane potential, the presumed driving force for TrkF activity.

B. Control of Activity of Efflux Systems

The only K^+ -specific efflux systems whose control has been studied are the KefB and KefC systems, each composed of a membrane protein and a soluble protein as described earlier. These systems seem to serve primarily as emergency systems to deal with the onslaught of certain toxic compounds. Compounds that activate these systems react with glutathione to create adducts, and certain of these adducts activate the systems. Once activated, K^+ release by the systems acidifies the cytoplasm. Acidification does not accelerate the inactivation of toxins (reviewed in Ref. (66)). Instead, it appears to reduce the rate at which the toxins create lethal lesions, an effect demonstrated for the attack of DNA by methylglyoxal (101). Both KefB and KefC seem to be inactive in the absence of the toxins that activate them; strains deleted for both are abnormal only in their hypersensitivity to these toxins (65). Mutations that cause them to be active in the absence of toxins identify regions of the proteins important in control of their activity (13,102). The similarity of the nucleotide-binding site of the KefC protein to those of the TrkA/KtrA class of proteins suggests a similar mode of interaction and possible regulation by NAD(H) (40).

C. Control of Expression

The only K^+ -uptake system whose expression is closely regulated at the transcriptional level is Kdp. It appears to be inducible in all strains in which its expression has been examined (103–107). Control is mediated by the KdpD sensor kinase and the KdpE-response regulator proteins. In *E. coli* the two regulators are encoded by an operon that overlaps the operon encoding the structural components of Kdp (108), but different arrangements are found in distantly related bacteria (109). In some bacteria only part of the *kdpD* gene is near to the other *kdp* genes; genes encoding KdpE and the missing fragment of KdpD must be elsewhere. KdpD function can occur by complementation of separately encoded N- and C-terminal fragments of the *E. coli* gene (110) as well as by a chimeric protein combining the 365-residue KdpD protein of *Anabaena* with a large C-terminal fragment of *E. coli* that includes the membrane spans (109).

Control of Kdp in *E. coli* has been extensively studied and is the subject of some controversy. KdpD is a 94-kDa protein with four closely spaced membrane spans separating two cytoplasmic domains of almost equal size (108,111), while KdpE is a 21-kDa soluble protein. *In vitro* studies confirm the role of phosphate in Kdp regulation: KdpD is autophosphorylated on what is probably a histidine residue, the phosphate is transferred to form the phosphorylated form of KdpE, P-KdpE, probably on an aspartate residue, and ultimately KdpD acts as phosphatase to remove phosphate from P-KdpE.

The C-terminal half of KdpD is the kinase domain, while the N-terminal domain is unique. There is an ATP-binding site in the N-terminal domain which seems to have an important role in signaling by KdpD, and a set of basic residues early in the C-terminal domain that are also found in homologs although neither the specific residues nor their precise location are conserved (112). KdpD acts as a dimer (113), and it seems that P-KdpE is also a dimer (M. Lucassin, cited in Ref. (114)). All current ideas about the regulation of Kdp assume that only KdpD and KdpE are involved, since no other mutations that radically alter regulation of Kdp have been reported. Insertion mutations in the *hns* and in *trxB* genes, encoding the H-NS nucleoid protein and thioredoxin reductase, respectively, reduce the threshold concentration below which Kdp is expressed (115). The mutations also reduced expression slightly when tested in a strain with two *kdpD* mutations that made expression constitutive; there was no reduction with a different *kdpD* mutation. These results are consistent with effects on the KdpD and/or KdpE proteins, but also with some steps upstream of these regulatory proteins. A trivial explanation, that these mutations increase uptake of compatible solutes or of K^+ , was not tested. Other proteins may be directly involved in regulation of Kdp expression. However, if these proteins have other functions essential to survival, they would be rarely found in a screen for constitutive or nonexpressing mutants.

It was early noted that Kdp is only expressed when the growth rate of the strain became limited by the K^+ concentration of the medium (13,103). This observation has been confirmed under a variety of conditions (116,117). The few reported exceptions are based on incomplete data, where sufficiently low K^+ concentrations were not tested because complex medium containing about 7 mM K^+ was used (118,119). The specific signal to turn on Kdp was suggested to be low turgor (103). The role of turgor was supported by finding that upshock resulted in a transient expression of Kdp which terminated when the new cellular K^+ concentration was reached which presumably restored normal turgor. Since turgor cannot be measured very well in *E. coli* or most other bacteria, this model cannot be tested rigorously.

A number of alternative models for control of Kdp have been suggested. Some *in vivo* data has been used to suggest a key role for the movement of K^+ across the cell membrane (117), while studies of purified KdpD reconstituted in vesicles suggest control by internal K^+ (summarized in Ref. (114)). None of these models fit very well with the control of Kdp seen *in vivo*. Regulation of Kdp is independent of which K^+ -transport system is present and of any particular K^+ concentration inside or outside, indicating that neither a specific range of K^+ concentrations nor K^+ fluxes through any of the known systems are involved (103,116). A specific role for salts has been suggested, based on the observation that an osmotic challenge by a salt was much more

effective in turning on expression of Kdp than an equivalent challenge with a sugar (120). This effect is due to an unexplained effect of salts, as compared to sugars, on the K^+ -concentration dependence of growth in strains expressing the Trk system (116). Whether in the presence of high concentrations of salts or of sugars, Kdp expression begins at the point where the growth rate becomes limited by medium K^+ concentration.

In an *in-vitro* system in which KdpD is oriented in an inside-out orientation, KdpD activity in phosphorylating KdpE from ATP was stimulated by increased KCl or NaCl on the cytoplasmic surface, but little if at all when these salts were present at the periplasmic surface (121). A somewhat different *in vitro* system, in which the orientation is the same as in intact cells, found inhibition of KdpD phosphorylation by K^+ ; however, inhibition is saturated at concentrations of K^+ lower than are found in the cell (122). Inhibition of Kdp expression by elevated internal K^+ could not explain how Kdp expression is off when growing in media where internal K^+ is low. The analysis of constitutive mutants has suggested that it is dephosphorylation of P-KdpE that is inhibited to turn on Kdp *in vivo*, rather than stimulation of the phosphorylation of KdpD or transfer of the phosphate to KdpE (123).

The four transmembrane segments of KdpD are believed to be crucial to physiological regulation of Kdp. Deletions of all four membranes was reported to result in a moderately active enzyme which did not respond to environmental signals (124) or in an inactive protein (125). A recent report shows that deletions of some or all of the segments impairs but does not abolish regulation totally (126). A deletion of all four membrane segments exhibits high expression in media of low K^+ concentration, which falls by a factor of 4 at high K^+ concentrations. Since the protein lacking membrane spans could not sense turgor, this effect must be mediated by the cytoplasmic parts of KdpD, which could be responding to changes in internal pH, internal K^+ , or other factors. Since the deletion mutant mediates only a 4-fold change in expression, while in the wild type there is a 500- to 1000-fold change (116), the residual regulation expressed in the deletion mutant may not reflect the key factor or factors that mediate regulation *in vivo*. Any model for control ought to allow for an effect of the cytoplasmic environment, including but not limited to K^+ and pH, in modulating the response. The *in vivo* data (103,116) argue very strongly that internal K^+ is not the primary signal controlling Kdp expression.

All of the major K^+ -uptake systems of moderate affinity appear to be expressed constitutively so that control is expressed at the level of activity. One exception to this pattern has been reported. In *B. subtilis* inactivation of one of the two Ktr systems leads to greatly increased expression of the other system (19). This effect is not due to a response to altered K^+ levels in the

cell, since expression of each was not altered when growth became limited by external K^+ .

III. K as an Intracellular Activator

The fact that uptake of K^+ is probably the earliest response to osmotic upshock suggested that a rise in internal K^+ might serve as a signal for osmotic stress sensed as a reduction in turgor (98,127). Osmotic upshock raises internal K^+ concentration, or more correctly K^+ activity, in less than a second by causing water to leave the cell due to the elasticity of the cell wall and at greater degrees of upshock by producing plasmolysis (reviewed in Ref. (98)). Although this effect has been suggested to explain the need for K^+ in the responses to upshock (128), this cannot be correct since K^+ uptake is required. The resulting uptake of K^+ , which requires only a minute or two in some bacteria but is considerably slower in others, allows a partial restoration of cell water and a direct increase in K^+ concentration in the cell. Most responses to upshock are permanent and depend on the final medium osmolality, very different from the effect of upshock on expression of Kdp, which is transient if other transport systems can maintain turgor at the new, higher osmolality.

This chapter will stress only very recent data on the role of K^+ in activating transport systems in response to osmotic stress since this topic and the roles of K^+ in increasing expression of some genes has been the subject of several recent reviews (128–133). This role of K^+ was first demonstrated for the ProU system of *S. typhimurium*, a system whose expression is increased several hundred fold after shift to elevated osmolality (120). Uptake of K^+ was found essential; the increase due to reduced cell water alone did not cause elevated expression of the ProU system. There is general agreement that virtually all the bacterial responses to elevated osmolality require an increase of cytoplasmic K^+ , but until recently the specific way K^+ produced the responses was not clear. A direct mechanistic model has the concentration (activity) of K^+ in the cytoplasm act directly on the system involved to activate or stimulate expression; a more general version would implicate ionic strength in the cytoplasm instead of being specific for K^+ . Alternative models would implicate some “adaptor” function, a protein or other cell component that reacts to increased K^+ and mediates its effects.

In vitro systems, in which purified transport systems for compatible solutes are incorporated into vesicles, has allowed direct tests of the role of K^+ and other solutes in activating transport systems. The BetP glycine betaine transporter of *C. glutamicum* when incorporated into vesicles prepared with *E. coli* phospholipids was stimulated when internal K^+ or Rb^+ or Cs^+ , used

as their phosphate salts, rose to levels above about 220 mM, while choline or NH_4^+ were not stimulatory (96). When K^+ glutamate was used as the internal solvent, results were very similar although the extent of activation was slightly reduced. Specific effects of Na^+ could not be tested since a gradient of this ion is needed to energize BetP. Earlier work in which extramembraneous regions of BetP were deleted showed that loss of 23 residues at the C-terminus abolished response to osmotic stimuli although the protein was active, albeit with lower activity and moderately reduced affinity for Na^+ (134). These results implicate that K^+ itself or a very similar ion interacts with the C-terminal tail of BetP to control its activity. Although the topology of the protein has not been established, it is likely that this C-terminal region is cytoplasmic.

Similar studies of the ProP transporter of *E. coli* have reached somewhat similar conclusions. *In vivo* ProP activity is activated rapidly, achieving and then maintaining its maximal rate within about a minute after osmotic upshock (135). The purified ProP protein was studied after reconstitution into vesicles in the absence of ProQ (136). The activity of ProP is increased about 6-fold in the presence of the soluble ProQ protein *in vivo*. Activation occurred when the vesicles were shrunk by increasing external osmolality with either NaCl or glucose. Half-maximal activation occurred when the internal cation was calculated to have reached about 0.35 M; K-phosphate, NaCl, KCl, CsCl, and LiCl all seemed equally effective in activation. When larger polyethylene glycols were added, the salt concentration for half-maximal activation fell to about 0.25 M. The data confirm that normally monovalent cation concentration is activating. The basis of the effect of polyethylene glycols is not clear. These compounds sequester water so that at least part of their effect is to reduce free water and thereby increase the activity of K^+ well above the calculated value. The study suggests that these compounds do not sequester sufficient water to explain their total effect, but the calculations of sequestration based on literature values of specific volumes may underestimate the sequestering effect. A direct effect of polyethylene glycols on ProP is possible.

The OpuA glycine betaine transporter of *Lactobacillus lactis* is 7-fold more active at high than at low osmolality (137). OpuA is a two-component ABC transporter that was energized in vesicles by incorporating an ATP-generating system within the vesicles where its activity was stimulated 5-fold by high external osmolality (138). The finding, not infrequent for membrane proteins, that the amphipathic anesthetic tetracaine altered the behavior of OpuA led to the suggestion that alterations in the membrane mediated regulation of OpuA by medium osmolality. A subsequent study shows that the activity of OpuA in liposomes depends on the presence of at least 25% anionic lipids and that activation requires elevated internal ionic strength since K^+

and Na^+ salts are equally effective (139). The authors infer that the anionic lipids mediate the effect of internal ionic strength, that they transmit the ionic signal to OpuA.

It is well known that membrane-transport systems and enzymes are sensitive to the lipid environment (reviewed in Refs. (140,141)). The fact that anionic lipids are required by OpuA could be due to a requirement of this lipid for stabilization or proper incorporation of the OpuA transporter in the membrane rather than any direct role in activation. Note that the KcsA channel also requires anionic lipids, yet no one would argue that it is the lipids that recognize or assist K^+ in getting into the channel (142). The data available does not identify the location of the ionic strength sensor for the OpuA system. Studies of OpuA mutants may help in identifying which regions of which components sense internal ionic strength.

IV. Regulation of Internal pH by K^+

To control the pH of the cytoplasm, cells must displace protons from the various buffer anions in the cell. The major buffer anions are the phosphate groups of nucleic acids and those charged residues in proteins whose pK is near the pH of the cytoplasm. To do this, the cell must use some other monovalent cation, and since K^+ is the major cellular cation it has a primary role in this process. Thus, one of the parameters that must control K^+ movements is cytoplasmic pH, but exactly how this is done is not very well understood. Since this topic is the subject of reviews (143,144) only brief consideration is given here.

In the steady state, one assumes that cellular buffer anions are constant so that exchanges of K^+ for H^+ would be the primary way internal pH can be altered. Yet there is only superficial understanding of the K^+/H^+ antiporters in the neutrophilic organisms, *E. coli* and *S. typhimurium*, whose physiology is best understood. In fact, we must turn to the Na^+/H^+ antiporters to get a glimpse of how such activities are regulated (75). There are two such activities in *E. coli*, differing in the pH range over which they are active and in their stoichiometry. The task is divided between one activity active in alkaline environments and another active in acidic environments. These activities are surely important in controlling cellular Na^+ but may be only minor players in maintaining the internal pH of a neutrophile such as *E. coli*.

A K^+/H^+ antiporter in *E. hirae* active in media of alkaline pH has been inferred to reduce internal pH when internal pH was increased by the presence of high concentrations of NH_4^+ (74). This activity has a 1:1 stoichiometry of K^+ for H^+ under the conditions studied, indicating that the outwardly directed gradient of H^+ is sufficient to drive K^+ accumulation.

The role of K^+/H^+ antiporters in regulation of pH when in an alkaline environment is confirmed by the failure of antiporter mutants in *E. coli* and *E. hirae* to grow in media above about pH 9 (73,74).

In *E. coli*, the only well-studied antiporters are the KefB and KefC activities. However, these act to acidify the cytoplasm only when activated by toxins and are probably not involved in regulating internal pH under other conditions. Internal pH in *E. coli* is somewhat dependent on the specific K^+ -transport system present (145). Internal pH was about 7.7 when the Trk system predominated, but rose to about 8.0 when the Kdp system was active. This small difference in internal pH results in increased sensitivity to the toxins that activate the KefB/KefC activities. That a rather small change in internal pH could significantly alter sensitivity to toxins is supported by the fact that when KefB/KefC activities reduce sensitivity to the toxins, they do so by reducing internal pH from 7.7 to 7.4, also a change of only 0.3 pH units (66). The difference in internal pH may be due to the difference in mechanisms of the two systems. Trk should tend to increase internal pH since it is inferred to be a K^+-H^+ symporter while Kdp is an electrogenic ATP-driven system that would tend to displace cytoplasmic H^+ .

V. Future Directions

There are still many important questions about the role of K^+ , not to speak of so many other aspects of bacterial metabolism that remain to be established. Many important systems, such as K^+/H^+ antiporters, are poorly understood and their genes in the commonly studied bacteria have yet to be identified. The hope that complete genome sequences will greatly speed our understanding has not been fully realized. The utility of these sequences are greatest when a gene not previously identified is homologous to genes of known or suspected function in other species. However, there are still many putative genes, now just open-reading frame, whose function is not known.

The function of an open-reading frame may be determined in strains in each of which one putative gene has been deleted or otherwise inactivated. This is potentially a good route, except that unless one has an idea of the function of the gene the screen used may not detect the function that has been lost. Just to cite one example, inactivation of any of the known K^+ -transport systems in *E. coli* would not have been detected by any ordinary screen. An approach that is generally more productive is to start with a mutant in a function, since determining the gene affected is relatively easy given all the tools now available.

Studies of cellular systems in simplified, *in vitro* systems have been essential to a full understanding of enzymes, and today are almost requisite for

a full understanding of transport systems and how they are regulated. It is well to remember, however, that *in vitro* results must always be compared to data of function *in vivo*, since the *in vitro* conditions at best approach but are almost never identical to those *in vivo*.

REFERENCES

1. M. O. Walderhaug, D. C. Dosch and W. Epstein, Potassium transport in bacteria. In B. Rosen, and S. Silver (eds.), "Ion Transport in Prokaryotes," pp. 85–103, New York, Academic Press, 1987.
2. A. Oren, Bioenergetic aspects of halophilism, *Microbiol. Mol. Biol. Rev.* **63**, 334–348 (1999).
3. D. Madern, C. Ebel and G. Zaccai, Halophilic adaptation of enzymes, *Extremophiles* **4**, 91–98 (2000).
4. A. D. Brown, "Microbial Water Stress Physiology: Principles and Perspectives," Chichester, John Wiley, 1990.
5. A. L. Koch, The surface stress theory for the case of *Escherichia coli*: the paradoxes of gram-negative growth, *Res. Microbiol.* **141**, 119–130 (1990).
6. J. B. Stock, B. Rausch and S. Roseman, Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*, *J. Biol. Chem.* **252**, 7850–7861 (1977).
7. D. S. Cayley, H. J. Guttman and M. T. Record, Jr., Biophysical characterization of changes in the amounts and activity of *Escherichia coli* cell and compartment water and turgor pressure in response to osmotic stress, *Biophys. J.* **78**, 1748–1764 (2000).
8. S. Alexandrowicz, Osmotic and Donnan equilibria in poly(acrylic acid)-sodium bromide solutions, *J. Poly. Sci.* **56**, 115–132 (1962).
9. A. M. Whatmore and R. H. Reed, Determination of turgor pressure in *Bacillus subtilis*: a possible role for K^+ in turgor regulation, *J. Gen. Microbiol.* **136**, 2521–2526 (1990).
10. A. L. Koch, The biophysics of the gram-negative periplasmic space, *Crit. Rev. Microbiol.* **24**, 23–59 (1998).
11. W. Epstein and S. G. Schultz, Cation transport in *Escherichia coli*. V. Regulation of cation content, *J. Gen. Physiol.* **49**, 221–234 (1965).
12. S. Leirimo, C. Harrison, D. S. Cayler, R. R. Burgess and M. T. Record, Jr., Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions *in vitro*, *Biochemistry* **26**, 7157–7164 (1987).
13. D. B. Rhoads, F. B. Waters and W. Epstein, Cation transport in *Escherichia coli*. VIII. Potassium transport mutants, *J. Gen. Physiol.* **67**, 325–341 (1976).
14. S. G. Schultz, W. Epstein and A. K. Solomon, Cation transport in *Escherichia coli*. IV. Kinetics of net K uptake, *J. Gen. Physiol.* **47**, 329–346 (1963).
15. E. P. Bakker and W. E. Mangerich, Interconversion of components of the bacterial proton motive force by electrogenic potassium transport, *J. Bacteriol.* **147**, 820–826 (1981).
16. A. J. Roe, D. McLaggan, C. P. O'Byrne and I. R. Booth, Rapid inactivation of the *Escherichia coli* Kdp K^+ uptake system by high potassium concentrations, *Mol. Microbiol.* **35**, 1235–1243 (2000).
17. F. M. Harold and D. Papineau, Cation transport and electrogenesis by *Streptococcus faecalis*, *J. Membr. Biol.* **8**, 27–44 (1972).
18. M. Michels and E. P. Bakker, Low-affinity potassium uptake system in *Bacillus acidocaldarius*, *J. Bacteriol.* **169**, 4335–4341 (1987).

19. G. Holtmann, E. P. Bakker, N. Uozumi and E. Bremer, KtrAB and KtrCD: two K^+ -uptake systems in *Bacillus subtilis* and their role in adaptation to hypertonicity, *J. Bacteriol.* **185**, 1289–1298 (2003).
20. W. Epstein and B. S. Kim, Potassium transport loci in *Escherichia coli* K-12, *J. Bacteriol.* **108**, 639–644 (1971).
21. D. C. Dosch, G. L. Helmer, S. H. Sutton, F. F. Salvacion and W. Epstein, Genetic analysis of potassium transport loci in *Escherichia coli*: evidence for three constitutive systems mediating uptake of potassium, *J. Bacteriol.* **173**, 687–696 (1991).
22. C. Parra-Lopez, R. Lin, A. Aspedon and E. A. Groisman, A *Salmonella* protein that is required for resistance to antimicrobial peptides and transport of potassium, *EMBO J.* **13**, 3964–3972 (1994).
23. D. Bossemeyer, A. Borchard, D. C. Dosch, G. C. Helmer, W. Epstein, I. R. Booth and E. P. Bakker, K^+ -Transport protein TrkA of *Escherichia coli* is a peripheral membrane protein that requires other *trk* gene products for attachment to the cytoplasmic membrane, *J. Biol. Chem.* **264**, 16403–16410 (1989).
24. A. Schlösser, A. Hamann, D. Bossemeyer, E. Schneider and E. P. Bakker, NAD^+ binding to the *Escherichia coli* K^+ -uptake protein TrkA and sequence similarity between TrkA and domains of a family of dehydrogenases suggest a role for NAD^+ in bacterial transport, *Mol. Microbiol.* **9**, 533–543 (1993).
25. C. Parra-Lopez, M. T. Baer and E. A. Groisman, Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*, *EMBO J.* **12**, 4053–4062 (1993).
26. S. Stumpe and E. P. Bakker, Requirement for a large K^+ uptake capacity and of extracytoplasmic protease activity for protamine resistance of *Escherichia coli*, *Arch. Microbiol.* **167**, 126–136 (1997).
27. A. Schlösser, M. W. Meldorf, S. Stumpe, E. P. Bakker and W. Epstein, TrkH and its homolog, TrkG, determine the specificity and kinetics of cation transport by the Trk system of *Escherichia coli*, *J. Bacteriol.* **177**, 1908–1910 (1995).
28. D. Bossemeyer, A. Schlösser and E. P. Bakker, Specific cesium transport via the *Escherichia coli* Kup (TrkD) K^+ uptake system, *J. Bacteriol.* **171**, 2219–2221 (1989).
29. D. B. Rhoads, A. Woo and W. Epstein, Discrimination between Rb^+ and K^+ by *Escherichia coli*, *Biochim. Biophys. Acta* **469**, 45–51 (1977).
30. E. Bakker, pH-dependent transport of rubidium by the constitutive potassium transport system TrkA of *Escherichia coli*, *FEMS Micro. Lett.* **16**, 229–233 (1983).
31. D. B. Rhoads and W. Epstein, Energy coupling to net K^+ -transport in *Escherichia coli* K-12, *J. Biol. Chem.* **252**, 1394–1401 (1977).
32. L. M. D. Stewart, E. P. Bakker and I. R. Booth, Energy coupling to K^+ uptake via the Trk system of *Escherichia coli*: the role of ATP, *J. Gen. Microbiol.* **131**, 77–85 (1985).
33. C. Harms, Y. Domoto, C. Celik, E. Rahe, S. Stumpe, R. Schmid, T. Nakamura and E. P. Bakker, Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K^+ -uptake systems Trk^H and Trk^G from *Escherichia coli* K-12, *Microbiology* **147**, 2991–3003 (2001).
34. A. Schlösser, S. Klüttig, A. Hamann and E. P. Bakker, Subcloning, nucleotide sequence and expression of *trkG*, a gene that encodes an integral membrane protein involved in potassium uptake via the Trk system of *Escherichia coli*, *J. Bacteriol.* **173**, 3170–3176 (1991).
35. T. Nakamura, N. Yamamuro, S. Stumpe, T. Unemoto and E. P. Bakker, Cloning of the *trkAH* gene cluster and characterization of the Trk K^+ -uptake system of *Vibrio alginolyticus*, *Microbiology* **144**, 2281–2289 (1998).
36. G. Wager, R. Hartmann and D. Oesterheld, Potassium uniport and ATP synthesis in *Halobacterium halobium*, *Eur. J. Biochem.* **89**, 169–179 (1978).

37. J. Meury and M. Kohiyama, ATP is required for K^+ active transport in the archaebacterium *Haloferax volcanii*, *Arch. Microbiol.* **151**, 530–536 (1989).
38. T. Nakamura, R. Yuda, T. Unemoto and E. P. Bakker, KtrAB, a new type of bacterial K^+ -uptake system from *Vibrio alginolyticus*, *J. Bacteriol.* **180**, 3491–3494 (1998).
39. N. Tholema, E. P. Bakker, A. Suzuki and T. Nakamura, Change to alanine of one out of four selectivity filter glycines in KtrB causes a two order of magnitude decrease in the affinities for both K^+ and Na^+ of the Na^+ dependent K^+ uptake system KtrAB from *Vibrio alginolyticus*, *FEMS Lett.* **450**, 217–220 (1999).
40. T. P. Roosild, S. Miller, I. R. Booth and S. Choe, A mechanism of regulating transmembrane potassium flux through a ligand-mediated conformational switch, *Cell* **109**, 781–791 (2002).
41. M. Kawano, K. Igarashi and Y. Kakinuma, Two major potassium uptake systems, KtrI and KtrII, in *Enterococcus hirae*, *FEMS Microbiol. Lett.* **176**, 449–453 (1999).
42. E. P. Bakker and F. M. Harold, Energy coupling to potassium transport in *Streptococcus fecalis*: interplay of ATP and the protonmotive force, *J. Biol. Chem.* **255**, 433–440 (1980).
43. K. Altendorf and W. Epstein., The Kdp-ATPase of *Escherichia coli*. In A. G. Lee (ed.), "Biomembranes," Vol. V, pp. 403–420, Greenwich, JAL Press, 1996.
44. K. Altendorf, M. Gassel, W. Puppe, T. Mollenkamp, A. Zeeck, C. Boddien, K. Fendler, E. Bamberg and S. Dröse, Structure and function of the Kdp-ATPase of *Escherichia coli*, *Acta Physiol. Scand.* **163** (Suppl. 643), 137–146 (1998).
45. R. H. Reed, P. Rowell and W. D. P. Stewart, Uptake of potassium and rubidium ions by the cyanobacterium *Anabaena variabilis*, *FEMS Microbiol. Lett.* **11**, 233–236 (1981).
46. S. T. Lim, K^+ -ATPase from *Rhizobium* sp. UMKL 20, *Arch. Microbiol.* **142**, 393–396 (1985).
47. M. Gassel, T. Mollenkamp, W. Puppe and K. Altendorf, The KdpF subunit is part of the K^+ -translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex in vitro, *J. Biol. Chem.* **274**, 37901–37907 (1999).
48. E. T. Buurman, K.-T. Kim and W. Epstein, Genetic evidence for two sequentially occupied K^+ binding sites in the Kdp transport ATPase, *J. Biol. Chem.* **270**, 6678–6685 (1995).
49. S. Dorus, H. Mimura and W. Epstein, Substrate-binding clusters of the K^+ -transporting Kdp ATPase of *Escherichia coli* investigated by amber suppression scanning mutagenesis, *J. Biol. Chem.* **276**, 9590–9598 (2001).
50. J. E. Hesse, L. Wiczorek, K. Altendorf, A. S. Reicin, E. Dorus and W. Epstein, Sequence homology between two membrane transport ATPases, the Kdp-ATPase of *Escherichia coli* and the Ca^{2+} -ATPase of sarcoplasmic reticulum, *Proc. Natl. Acad. Sci. USA* **81**, 4746–4750 (1984).
51. M. Gassel, A. Siebers, W. Epstein and K. Altendorf, Assembly of the Kdp complex, the multi-subunit K^+ -transport ATPase of *Escherichia coli*, *Biochim. Biophys. Acta* **1315**, 77–84 (1998).
52. K. Fendler, S. Dröse, K. Altendorf and E. Bamberg, Electrogenic transport by the Kdp-ATPase of *Escherichia coli*, *Biochemistry* **35**, 8009–8017 (1996).
53. E. Zakharyan and A. Trchounian, K^+ influx by Kup in *Escherichia coli* is accompanied by a decrease in H^+ efflux, *FEMS Microbiol. Lett.* **204**, 61–64 (2001).
54. A. Trchounian and H. Kobayashi, Kup is the major K^+ uptake system in *Escherichia coli* upon hyperosmotic stress at low pH, *FEBS Lett.* **447**, 144–148 (1999).
55. M. Schleyer and E. P. Bakker, Nucleotide sequence and 3'-end deletion studies indicate that the K^+ -uptake protein Kup from *Escherichia coli* is composed of a hydrophobic core linked to a large and partially essential hydrophilic C terminus, *J. Bacteriol.* **175**, 6925–6931 (1993).
56. D. C. Dosch, F. F. Salvacion and W. Epstein, Tetracycline resistance element of pBR322 mediates potassium transport, *J. Bacteriol.* **160**, 1188–1190 (1984).
57. M. Ito, B. Cooperberg and T. A. Krulwich, Diverse genes of the alkaliphilic *Bacillus firmus* OF4 that complement K^+ -uptake-deficient *Escherichia coli* including an *ftsH* homologue, *Extremophiles* **1**, 22–28 (1995).

58. M. Kawano, K. Igarashi and Y. Kakinuma, Isolation of *Enterococcus hirae* mutant deficient in low-affinity potassium uptake at alkaline pH, *Biosci. Biotechnol. Biochem.* **66**, 1597–1600 (2002).
59. A. A. Guffanti, J. Cheng and T. A. Krulwich, Electrogenic antiport activities of the Gram-positive Tet proteins include a $\text{Na}^+(\text{K}^+)/\text{K}^+$ mode that mediates net K^+ uptake, *J. Biol. Chem.* **273**, 26447–26454 (1998).
60. A. A. Guffanti, Y. Wei, S. Y. Rood and T. A. Krulwich, An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K^+ and H^+ , *Mol. Microbiol.* **45**, 145–153 (2002).
61. W. Wang, A. A. Guffanti, Y. Wei, M. Ito and T. A. Krulwich, Two types of *Bacillus subtilis* tetA(L) deletion strains reveal the physiological importance of TetA(L) in K^+ acquisition as well as in Na^+ , alkali and tetracycline resistance, *J. Bacteriol.* **182**, 2088–2095 (2000).
62. J. Šebestian, Z. Petrmichlová, Š. Šebestianová, J. Náprstek and J. Svobodová, Osmoregulation in *Bacillus subtilis* under potassium limitation: a new inducible K^+ -stimulated VO_4^{3-} -inhibited ATPase, *Can. J. Microbiol.* **47**, 1116–1125 (2001).
63. P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism, *Nature* **191**, 144–148 (1961).
64. J. Meury, A. Robin and P. Monnier-Champeix, Turgor-controlled K^+ fluxes and their pathways in *Escherichia coli*, *Eur. J. Biochem.* **151**, 613–619 (1985).
65. E. P. Bakker, I. R. Booth, U. Dinnbier, W. Epstein and A. Gajewska, Evidence for multiple K^+ export systems in *Escherichia coli*, *J. Bacteriol.* **169**, 3743–3749 (1987).
66. G. P. Ferguson, Protective mechanisms against toxic electrophiles in *Escherichia coli*, *Trends Microbiol.* **7**, 242–247 (1999).
67. J. Meury, S. Lebail and A. Kepes, Opening of potassium channels in *Escherichia coli* membranes by thiol reagents and the recovery of potassium tightness, *Eur. J. Biochem.* **113**, 33–38 (1980).
68. J. Meury and A. Kepes, Glutathione and the gated potassium channels of *Escherichia coli*, *EMBO J.* **1**, 339–343 (1982).
69. E. P. Bakker and W. E. Mangerich, N-ethylmaleimide induces K^+-H^+ antiport activity in *Escherichia coli* K-12, *FEBS Lett.* **140**, 177–180 (1982).
70. A. W. Munro, G. Y. Ritchie, A. J. Lamb, R. M. Douglas and I. R. Booth, The cloning and DNA sequence of the gene for the glutathione-regulated potassium-efflux system KefC of *Escherichia coli*, *Mol. Microbiol.* **5**, 607–616 (1991).
71. S. Miller, L. S. Ness, C. M. Wood, B. C. Fox and I. R. Booth, Identification of an ancillary protein, YapF, required for the activity of the KefC glutathione-gated potassium efflux system in *Escherichia coli*, *J. Bacteriol.* **182**, 6536–6540 (2000).
72. R. N. Brey, B. P. Rosen and E. N. Sorensen, Cation/proton antiport systems in *Escherichia coli*: properties of the potassium/proton antiporter, *J. Biol. Chem.* **255**, 39–44 (1980).
73. R. H. Plack, Jr. and B. P. Rosen, Cation/proton antiport systems in *Escherichia coli*: absence of potassium/proton antiporter activity in a pH-sensitive mutant, *J. Biol. Chem.* **255**, 3824–3825 (1980).
74. Y. Kakinuma and K. Igarashi, Isolation and properties of *Enterococcus hirae* mutants defective in the potassium/proton antiport system, *J. Bacteriol.* **181**, 4103–4105 (1999).
75. E. Padan, M. Venturi, Y. Gerchman and N. Dover, Na^+/H^+ antiporters, *Biochim. Biophys. Acta* **1505**, 144–157 (2001).
76. A. Jayakumar, W. Epstein and E. M. Barnes, Jr., Characterization of ammonium (methylammonium)/potassium antiport in *Escherichia coli*, *J. Biol. Chem.* **260**, 7528–7532 (1985).
77. G. D. Sprott, K. M. Shaw and K. F. Jarrell, Ammonia/potassium exchange in methanogenic bacteria, *J. Biol. Chem.* **259**, 12602–12608 (1984).

78. I. R. Booth, Bacterial channels. In J. K. Setlow (ed.), "Genetic Engineering," Vol. XXV, pp. 91–111, New York, Kluwer Academic/Plenum Publishing, 2003.
79. D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. L. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait and R. MacKinnon, The structure of the potassium channel: molecular basis of K^+ conduction and selectivity, *Science* **280**, 69–77 (1998).
80. Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait and R. MacKinnon, Crystal structure and mechanism of a calcium-gated potassium channel, *Nature* **417**, 515–522 (2002).
81. M. LeMasurier, L. Heginbotham and C. Miller, KcsA: it's a potassium channel, *J. Gen. Physiol.* **118**, 303–313 (2001).
82. R. Milkman, An *Escherichia coli* homologue of eucaryotic potassium channel proteins, *Proc. Natl. Acad. Sci. USA* **91**, 3510–3514 (1994).
83. T. S. Munsey, A. Hohindra, S. P. Yusaf, A. Grainge, M.-H. Wang, D. Wray and A. Sivaprasadarao, Functional properties of Kch, a prokaryotic homologue of eucaryotic potassium channels, *Biochem. Biophys. Res. Commun.* **297**, 10–16 (2002).
84. G. Chang, R. H. Spencer, A. T. Lee, M. T. Barclay and D. C. Rees, Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel, *Science* **282**, 2220–2226 (1998).
85. R. B. Bass, P. Strop, M. Barclay and D. C. Rees, Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel, *Science* **298**, 1582–1587 (2002).
86. Y. Li, P. C. Moe, S. Chandrasekaran, I. R. Booth and P. Blount, Ionic regulation of MscK, a mechanosensitive channel from *Escherichia coli*, *EMBO J.* **21**, 1–8 (2002).
87. C. Cui and J. Adler, Effect of mutation of potassium-efflux system, KefA, on mechanosensitive channels in the cytoplasmic membrane of *Escherichia coli*, *J. Membr. Biol.* **150**, 143–152 (1996).
88. D. McLaggan, M. A. Jones, G. Gouesbet, N. Levina, S. Lindey, W. Epstein and I. R. Booth, Analysis of the *kefA2* mutation suggests that KefA is a cation-specific channel involved in osmotic adaptation in *Escherichia coli*, *Mol. Microbiol.* **43**, 521–536 (2002).
89. C. A. Berrier, A. Coulombe, I. Szabó, M. Zoratti and A. Ghazi, Gadolinium ion inhibits loss of metabolites induced by osmotic shock and large stretch-activated channels in bacteria, *Eur. J. Biochem.* **206**, 559–565 (1992).
90. N. Levina, S. Töttemeyer, N. R. Stokes, P. Louis, M. A. Jones and I. R. Booth, Protection of *Escherichia coli* cells against extreme turgor by activation of the MscS and MscL mechanosensitive channels: identification of the genes required for MscS activity, *EMBO J.* **18**, 1730–1737 (1999).
91. S. L. Ørskov, Experiments in active and passive permeability of *Bacillus coli communis*, *Acta Path. Microbiol. Scand.* **25**, 277–283 (1948).
92. D. B. Rhoads and W. Epstein, Cation transport in *Escherichia coli*. IX. Regulation of K^+ transport, *J. Gen. Physiol.* **72**, 283–295 (1978).
93. E. Blumwald, R. J. Mehlhorn and L. Packer, Ionic osmoregulation during salt adaptation of the cyanobacterium *Synechococcus* 6311, *Plant Physiol.* **73**, 377–380 (1983).
94. A. M. Whatmore, J. A. Chudek and R. H. Reed, The effects of osmotic upshock on the intracellular solute pools of *Bacillus subtilis*, *J. Gen. Microbiol.* **136**, 2527–2535 (1990).
95. S. P. Cummings, M. P. Williamson and D. J. Gilmour, Turgor regulation in a novel *Halomonas* species, *Arch. Microbiol.* **160**, 319–323 (1993).
96. R. Rübenhagen, S. Morbach and R. Krämer, The osmoreactive betaine carrier BetP from *Corynebacterium glutamicum* is a sensor for cytoplasmic K^+ , *EMBO J.* **20**, 5412–5420 (2001).
97. D. D. Martin, R. A. Ciulla, P. M. Robinson and M. F. Roberts, Switching osmolyte strategies: response of *Methanococcus thermolithotrophicus* to changes in external NaCl, *Biochim. Biophys. Acta* **1524**, 1–10 (2000).

98. W. Epstein, Osmoregulation by potassium transport in *Escherichia coli*, *FEMS Microbiol. Rev.* **39**, 73–78 (1986).
99. E. P. Bakker and W. E. Mangerich, The effects of weak acids on potassium uptake by *Escherichia coli* K-12. Inhibition by low cytoplasmic pH, *Biochim. Biophys. Acta* **730**, 379–386 (1983).
100. J. S. Frymier, T. D. Reed, S. A. Fletcher and L. N. Csonka, Characterization of transcriptional regulation of the *kdp* operon of *Salmonella typhimurium*, *J. Bacteriol.* **179**, 3061–3063 (1997).
101. G. P. Ferguson, J. R. Battista, A. T. Lee and I. R. Booth, Protection of the DNA during exposure of *Escherichia coli* cells to a toxic metabolite: the role of the KefB and KefC potassium channels, *Mol. Microbiol.* **35**, 113–122 (2000).
102. S. Miller, R. M. Douglas, P. Carter and I. R. Booth, Mutations in the glutathione-gated KefC K^+ efflux system of *Escherichia coli* that cause constitutive activation, *J. Biol. Chem.* **272**, 24942–24947 (1997).
103. L. A. Laimins, D. B. Rhoads and W. Epstein, Osmotic control of *kdp* operon expression in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **78**, 464–468 (1981).
104. E. P. Bakker, A. Borchard, M. Michels, K. Altendorf and A. Siebers, High-affinity potassium uptake system in *Bacillus acidocaldarius* showing immunological cross-reactivity with the Kdp system from *Escherichia coli*, *J. Bacteriol.* **169**, 4342–4348 (1987).
105. J. Hafer, A. Siebers and E. P. Bakker, The high-affinity K^+ -translocating ATPase complex from *Bacillus acidocaldarius* consists of three subunits, *Mol. Microbiol.* **3**, 487–495 (1989).
106. A. Treuner-Lange, A. Kuhn and P. Durre, The *kdp* system of *Clostridium acetobutylicum*: cloning, sequencing, and transcriptional regulation in response to potassium concentration, *J. Bacteriol.* **179**, 4501–4512 (1997).
107. A. Alahari, A. Ballal and S. Apte, Regulation of potassium-dependent Kdp-ATPase expression in the nitrogen-fixing cyanobacterium *Anabaena torulosa*, *J. Bacteriol.* **183**, 5778–5781 (2001).
108. M. O. Walderhaug, J. W. Polarek, P. Voelkner, J. M. Daniel, J. E. Hesse, K. Altendorf and W. Epstein, KdpD and KdpE, proteins that control expression of the *kdpABC* operon, are members of the two-component sensor-effector class of regulators, *J. Bacteriol.* **174**, 2152–2159 (1992).
109. A. Ballal, R. Heermann, K. Jung, M. Gassel, S. K. Apte and K. Altendorf, A chimeric *Anabaena/Escherichia coli* KdpD protein (Anacoli KdpD) functionally interacts with *E. coli* KdpE and activates *kdp* expression in *E. coli*, *Arch. Microbiol.* **178**, 141–148 (2002).
110. R. Heermann, K. Altendorf and K. Jung, The hydrophilic N-terminal domain complements the membrane-anchored C-terminal domain of the sensor kinase KdpD of *Escherichia coli*, *J. Biol. Chem.* **275**, 17080–17085 (2000).
111. P. Zimmann, W. Puppe and K. Altendorf, Membrane topology analysis of the sensor kinase KdpD of *Escherichia coli*, *J. Biol. Chem.* **270**, 28282–28288 (1995).
112. K. Jung and K. Altendorf, Individual substitutions of clustered arginine residues of the sensor kinase KdpD of *Escherichia coli* modulate the ratio of kinase to phosphatase activity, *J. Biol. Chem.* **273**, 26415–26420 (1998).
113. R. Heermann, K. Altendorf and K. Jung, The turgor sensor KdpD of *Escherichia coli* is a homodimer, *Biochim. Biophys. Acta* **1415**, 114–124 (1998).
114. K. Jung and A. Altendorf, Towards an understanding of the molecular mechanisms of stimulus perception and signal transduction by the KdpD/KdpE system of *Escherichia coli*, *J. Mol. Microbiol. Biotechnol.* **4**, 223–228 (2002).
115. A. A. Sardesai and J. Gowrishankar, *trans*-acting mutations in loci other than *kdpDE* that affect *kdp* operon regulation in *Escherichia coli*: effects of cytoplasmic thiol oxidation status and nucleoid protein H-NS on *kdp* expression, *J. Bacteriol.* **183**, 86–93 (2001).

116. R. Malli and W. Epstein, Expression of the Kdp ATPase is consistent with regulation by turgor pressure, *J. Bacteriol.* **180**, 5102–5108 (1998).
117. H. Asha and J. Gowrishankar, Regulation of *kdp* operon expression in *Escherichia coli*: evidence against turgor as a signal for transcriptional control, *J. Bacteriol.* **175**, 4528–4537 (1993).
118. A. Sugiura, K. Hirokawa, K. Nakashima and T. Mizuno, Signal-sensing mechanisms of the putative osmosensor KdpD in *Escherichia coli*, *Mol. Microbiol.* **14**, 929–938 (1994).
119. I. Stallkamp, K. Altendorf and K. Jung, Amino acid replacements in transmembrane domain 1 influence osmosensing but not K^+ sensing by the sensor kinase KdpD of *Escherichia coli*, *Arch. Microbiol.* **178**, 525–530 (2002).
120. L. Sutherland, J. Cairney, M. J. Elmore, I. R. Booth and C. F. Higgins, Osmotic regulation of transcription: induction of the *proU* transport gene is dependent on accumulation of intracellular potassium, *J. Bacteriol.* **168**, 804–814 (1986).
121. K. Nakashima, A. Sugiura and T. Mizuno, Functional reconstitution of the putative *Escherichia coli* osmosensor, KdpD, into liposomes, *J. Biochem.* **114**, 615–621 (1993).
122. K. Jung, M. Veen and K. Altendorf, K^+ and ionic strength directly influence the autophosphorylation activity of the putative turgor sensor KdpD of *Escherichia coli*, *J. Biol. Chem.* **275**, 40142–40147 (2000).
123. L. Brandon, S. Dorus, W. Epstein, K. Altendorf and K. Jung, Modulation of KdpD phosphatase implicated in the physiological expression of the Kdp ATPase of *Escherichia coli*, *Mol. Microbiol.* **38**, 1086–1092 (2000).
124. K. Nakashima, A. Sugiura, K. Kanamaru and T. Mizuno, Signal transduction between the two regulatory components involved in the regulation of the *kdpABC* operon in *Escherichia coli*: phosphorylation-dependent functioning of the positive regulator, KdpE, *Mol. Microbiol.* **7**, 109–116 (1993).
125. W. Puppe, P. Zimmann, K. Jung, M. Lucassen and K. Altendorf, Characterization of truncated forms of the KdpD protein, the sensor kinase of the K^+ -translocating Kdp system of *Escherichia coli*, *J. Biol. Chem.* **271**, 25027–25034 (1996).
126. R. Heermann, A. Fohrmann, K. Altendorf and K. Jung, The transmembrane domains of the sensor kinase KdpD of *Escherichia coli* are not essential for sensing K^+ limitation, *Mol. Microbiol.* **47**, 839–848 (2003).
127. I. R. Booth and C. F. Higgins, Enteric bacteria and osmotic stress: intracellular potassium glutamate as a secondary signal of osmotic stress, *FEMS Microbiol. Lett.* **75**, 239–246 (1990).
128. R. D. Sleator and C. Hill, Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence, *FEMS Microbiol. Rev.* **26**, 49–71 (2001).
129. L. N. Csonka and W. Epstein, Osmoregulation. In F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger (ed.), “*Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology,” Vol. I, pp. 1210–1223, Washington, DC, ASM Press (1996).
130. B. Kempf and E. Bremer, Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments, *Arch. Microbiol.* **170**, 319–330 (1998).
131. J. M. Wood, E. Bremer, L. N. Csonka, R. Kraemer, B. Poolman, T. van der Heide and L. T. Smith, Osmosensing and osmoregulatory compatible solute accumulation by bacteria, *Comp. Biochem. Physiol. A* **130**, 437–460 (2001).
132. B. Poolman, P. Blount, J. H. A. Folgering, R. H. E. Friesen, P. C. Moe and T. van der Heide, How do membrane proteins sense water stress?, *Mol. Microbiol.* **44**, 889–902 (2002).
133. S. Morbach and R. Krämer, Body shaping under water stress: osmosensing and osmoregulation of solute transport in bacteria, *ChemBioChem.* **3**, 384–397 (2002).

134. H. Peter, A. Burkovski and R. Krämer, Osmo-sensing by N- and C-terminal extensions of the glycine betaine uptake system BetP of *Corynebacterium glutamicum*, *J. Biol. Chem.* **273**, 2567–2574 (1998).
135. J. L. Milner, S. Grothe and J. M. Wood, Proline porter II is activated by a hyperosmotic shift in both whole cells and membrane vesicles of *Escherichia coli*, *J. Biol. Chem.* **263**, 14900–14905 (1990).
136. D. E. Culham, J. Henderson, R. A. Crane and J. M. Wood, Osmosensor ProP of *Escherichia coli* responds to the concentration, chemistry and molecular size of osmolytes in the proteoliposome lumen, *Biochemistry* **42**, 410–420 (2003).
137. T. van der Heide and B. Poolman, Glycine betaine transport in *Lactococcus lactis* is osmotically regulated at the level of expression and translocation activity, *J. Bacteriol.* **182**, 203–206 (2000).
138. T. van der Heide and B. Poolman, Osmoregulated ABC-transport system of *Lactococcus lactis* senses water stress via changes in the physical state of the membrane, *Proc. Natl. Acad. Sci. USA* **97**, 7102–7106 (2000).
139. T. van der Heide, M. C. A. Stuart and B. Poolman, On the osmotic signal and osmosensing mechanism of an ABC transport system for glycine betaine, *EMBO J.* **20**, 7022–7032 (2001).
140. M. Bogdanov and W. Dowhan, Lipid-assisted protein folding, *J. Biol. Chem.* **274**, 36827–36830 (1999).
141. G. I. Veld, A. J. Driesen and W. N. Konings, Bacterial solute transport proteins in their lipid environment, *FEMS Microbiol. Rev.* **12**, 293–314 (1993).
142. L. Heginbotham, L. Kolmakova-Partensky and C. Miller, Functional reconstitution of a prokaryotic K⁺ channel, *J. Gen. Physiol.* **111**, 741–749 (1998).
143. I. R. Booth, Regulation of cytoplasmic pH in bacteria, *Microbiol. Rev.* **49**, 359–378 (1985).
144. I. R. Booth, The regulation of intracellular pH in bacteria, *Novartis Found. Symp.* **221**, 19–28 (1999).
145. G. P. Ferguson, A. D. Chacko, C. Lee and I. R. Booth, The activity of the high-affinity potassium uptake system Kdp sensitises cells of *Escherichia coli* to methylglyoxal, *J. Bacteriol.* **178**, 3957–39661 (1996).

Conformational Polymorphism of d(A-G)_n and Related Oligonucleotide Sequences

NINA G. DOLINNAYA,¹ AND
JACQUES R. FRESCO

*Department of Molecular Biology,
Princeton University, Princeton,
NJ 08544, USA*

I. Introduction	321
II. α -DNA Helix \rightleftharpoons Acid-Dependent Duplex Equilibrium	323
III. Acid-Dependent Parallel Duplex \rightleftharpoons pH-Independent Parallel Duplex Equilibrium	332
IV. pH-Independent Parallel Duplex \rightleftharpoons Single-Hairpin Duplex \rightleftharpoons Two-Hairpin Tetraplex Equilibria	334
V. Concluding Remarks	342
Acknowledgments	344
References	344

I. Introduction

Throughout mammalian genomes, the repetitive sequence d(A-G)_n occurs with greater random frequency, especially in gene control elements (1,2). Moreover, a number of proteins have been isolated that bind to one such strand of a duplex rather than to the intact duplex, although the conformational determinants for such binding are unknown (3,4). It is perhaps for this reason that over the last decade a number of investigations addressed the structure of the repetitive homopurine-strand element d(A-G)_n (5–7). The initial impression gleaned from those studies was of results at variance with one another, or possibly that each revealed properties characteristic of a different structural form.

Our own efforts in this regard were aimed at defining the conformational potential of d(A-G)_n and the factors that may alter the equilibria between different conformers. Thus, we delineated the properties of five conformations for this polymorphic sequence element (see Table I and Fig. 1),

¹Present address: Department of Chemistry, Lomonosov Moscow State University, 119992 Moscow, Russia.

TABLE I
SOME GENERAL FEATURES OF CONFORMATIONAL VARIANTS OF d(A-G)_n

Conformation	Salient structural features	Stability conditions				
		pH	Ionic strength	Temperature	Length	Oligomer concentration
α -DNA (single-stranded helix)	Base-backbone: A-N1H ⁺ ... ⁻ O-P, A-N6H ... O=P; no dA-dG stacking; left-handed twist; 12Å helix diameter	<6, ≥ 4; <pH 4 it denatures as Gs are protonated	≤ 0.01 M Na ⁺ ; inverse <i>T</i> _m - dependence; stabilized by ethanol	Maximum <i>T</i> _m ~ 40°C	>(A-G) ₅ - (A-G) ₃₀	Stability concentration- independent
Acid- dependent parallel duplex	Stacked A ⁺ · A ⁺ and G · G base pairs; interstrand A ⁺ ... ⁻ O-P and A-N6H ... O=P	<6, ≥ 4	≤ 0.01 M Na ⁺ ; inverse <i>T</i> _m - dependence; stabilized by ethanol	Maximum <i>T</i> _m ~ 40°C	>(A-G) ₁₅	Stability concentration- dependent
pH- independent parallel duplex	Stacked A · A, G · G base pairs	>5.5	≥ 0.1 M Na ⁺ ; positive <i>T</i> _m - dependence	Maximum <i>T</i> _m ~ 40°C	>(A-G) ₁₅	Stability concentration- dependent
Hairpin (antiparallel) duplex	Probably A · A and G · G base pairs	~ 7	0.1-0.2 M Na ⁺ ; positive <i>T</i> _m - dependence	Maximum <i>T</i> _m ~ 40°C	>(A-G) ₁₀	Stability concentration- independent; favored over tetraplex at low concentration
Two- hairpin tetraplex	Partially H-bonded, intercalated A tetrads; fully H-bonded G tetrads	~ 7	>0.6 M Na ⁺ or 0.005 M Mg ²⁺ ; positive <i>T</i> _m -dependence	Maximum <i>T</i> _m ~ 40°C	>(A-G) ₁₀	Favored over hairpin at high concentration

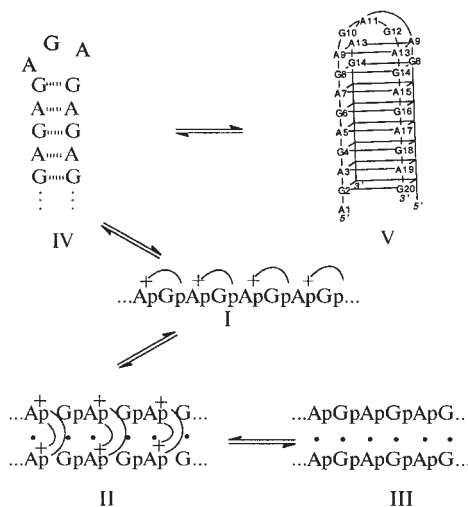


FIG. 1. Schematic representation of secondary structures formed by $d(A-G)_n$ ($6 \leq n \leq 30$) under different conditions of pH, ionic strength, temperature, and oligomer concentration. (I) α -DNA (single-stranded) helix; (II) Acid-dependent (parallel) duplex; (III) pH-independent (parallel) duplex; (IV) Hairpin (antiparallel) duplex; (V) Two-hairpin tetraplex.

whose occurrence depends on particular combinations of pH, ionic strength, types of cations, length of repeat, oligomer concentration, and the presence of adjacent helix-stabilizing residues. The conformations identified include a single-stranded helix (8–11), two different parallel-stranded linear duplexes (12), a hairpin (antiparallel) duplex (13), and a two-hairpin tetraplex (13).

It is our primary goal in this report to describe and rationalize the several equilibria that underlie the conformational polymorphism of $d(A-G)_n$. We describe each of the equilibria along with the properties of their participating conformers. In addition, as various of these $d(A-G)_n$ conformations seem to share physical-chemical properties with naturally occurring irregular, i.e., nonrepetitive, homopurine sequences in both DNA and RNA, we use the more readily interpretable properties of the regular $d(A-G)_n$ repeat to extrapolate structure for those irregular sequences and vice versa.

II. α -DNA Helix \leftrightarrow Acid-Dependent Duplex Equilibrium

A novel type of single-stranded helix (8–11) has been observed for the repeating sequence $d(A-G)_n$ between pH 6 and pH 4 and 0.01–0.001 M

Na^+ (structure I in Fig. 1). This conformation is stabilized not by helically wound stacks of bases or base pairs, but instead by an unusual combination of ionic and hydrogen bonds between dA^+ residues (electropositive hydrogens $-\text{N6}-\text{H}$ and $\text{N1}-\text{H}^+$) and the phosphodiester backbone such that the dA residues do not overlap their dG nearest neighbors. Increasing the length of $\text{d}(\text{A-G})_n$ from $n = 6$ to 10 raises the pK_a of the coil \Rightarrow helix transition (9), i.e., stabilizes the helix at higher pH, and therefore farther away from the intrinsic pK_a of the dA residues. In contrast, lengthening the oligomer to $n = 20$ –30 favors transformation of this single-stranded helical structure, designated α -DNA (10), to a double-stranded one (12). Between 0.001 M and 0.01 M Na^+ below pH 6.0, both oligomers $[\text{d}(\text{A-G})_{20,30}]$ form an acid-dependent parallel-stranded duplex with $\text{A}^+ \cdot \text{A}^+$ and $\text{G} \cdot \text{G}$ base pairs (structure II in Fig. 1). Along with base pairing, the important interactions that contribute to parallel duplex stability are ionic and associated hydrogen bonds between the protonated dA residues and the backbone phosphates of opposing strands (12). Under slightly more acidic conditions, the acid-dependent duplex is in equilibrium with the single-stranded α -DNA conformation taken up by $\text{d}(\text{A-G})_{6,10}$. In effect, the intramolecular hydrogen bonds and ionic bonds of α -DNA become interstrand bonds in the acid-dependent duplex.

It has been convenient to study these equilibria by a combination of CD and UV spectroscopy. While both acid-dependent structures (8,12) are characterized by intense similar but distinguishable CD spectra, only the duplex conformation stabilized by helically wound and stacked base pairs displays a large cooperative thermally induced hyperchromic transition. By superimposing normalized plots of CD intensity and UV absorption as a function of pH, it was possible to delineate the pH ranges over which α -DNA and the acid-dependent duplex are stable for $\text{d}(\text{A-G})_{10}$, $\text{d}(\text{A-G})_{20}$, and $\text{d}(\text{A-G})_{30}$ in 0.01 M and 0.001 M Na^+ . These phase-like diagrams (Fig. 2), together with the results of our earlier studies, reveal several significant trends.

- $\text{d}(\text{A-G})_{10}$ and shorter sequences do not form the acid duplex at least to the 2-mM residue concentration used for ultraviolet resonance Raman (UVRR) spectral measurements (11), presumably for entropic reasons.
- At Na^+ concentrations < 0.015 M, as pH rises from 5.0 to 5.5, the alternative strongly chiral-protonated helical structures [α -DNA for $\text{d}(\text{A-G})_{10}$, acid-dependent duplex for $\text{d}(\text{A-G})_{20,30}$] undergo a major conformational transition to unstructured single strands (12).
- The pK_a values for those transitions to unstructured single strands drop as the Na^+ concentration is raised from 0.001 to 0.01 M. The drop is more pronounced for α -DNA than for the acid-dependent duplex. This could be explained by a relatively greater electrostatic contribution to

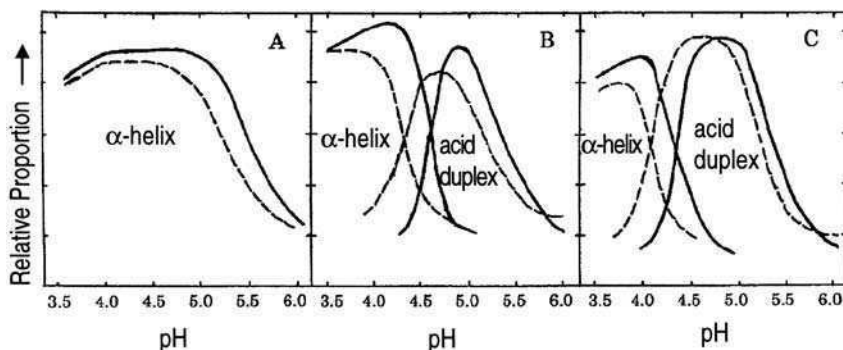


FIG. 2. Changes in the relative proportion (in arbitrary units) of α -DNA helix and acid-dependent parallel duplex for $d(A-G)_{10}$ (A), $d(A-G)_{20}$ (B), $d(A-G)_{30}$ (C) as a function of pH in 0.001 M Na^+ (—) and 0.01 M Na^+ (---) (after Ref. (12)).

the stability of α -DNA since in that structure there is no need to reduce interstrand backbone repulsion by counterion shielding, as is the case for the acid-dependent duplex.

- At $pH \geq 4.0$, $d(A-G)_{20,30}$ undergo a transition from α -DNA to the base-stacked acid-dependent duplex. Moreover, the tendency to form α -DNA is less for $d(A-G)_{30}$ than for $d(A-G)_{20}$, probably because the number of residues available for cooperative interstrand base pairing is greater (12).
- In 0.01 M Na^+ , the acid-dependent duplex requires a lower pH to attain maximum stability ($pH \sim 4.6$) than in 0.001 M Na^+ ($pH \sim 5.0$) (12).

From these various observations, it becomes apparent that the predominant structural form of $d(A-G)_n$ in acid medium depends on oligomer length, pH, and ionic strength.

A. α -DNA

A novel nucleic acid secondary structure, exemplified by $d(A^+-G)_{10}$, was shown in our studies to be formed by an intramolecular cooperative, acid-induced coil \Rightarrow helix transition. In the absence of base pairing and base stacking, it is maintained by hydrogen and ionic bonds between A^+ and the phosphodiester backbone (8,9). Formation of the $d(A^+-G)_{10}$ helix is accompanied by cooperative uptake of nine protons by the nine dA residues that can form ionic bonds with all the available distal phosphates (9). A well-resolved doublet ^{31}P -NMR spectrum (one peak for the A^+pG segments with phosphates bridged by ionic and hydrogen bonds to distal dA^+ residues, and

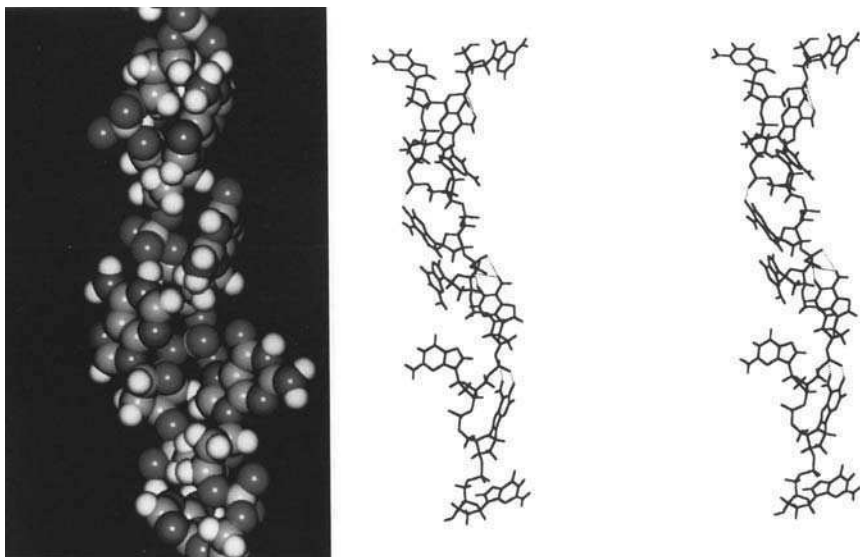


FIG. 3. Helical secondary structure of single-stranded α -DNA. Five center residues of $d(A-G)_5$, i.e., $5'-d(pA^+pGpA^+pGpA^+)-3'$, showing ionic and hydrogen bonds as dashed lines. All structures are oriented $5' \rightarrow 3'$ from top to bottom. The space-filling model on the left is rotated 180° relative to the orientation of the stereoview wire model on the right (10).

the other peak for the GpA^+ segments with noninteracting backbone phosphates) confirms two distinct phosphate environments in the α -DNA helix (10). UVRR spectroscopy reveals that both dG and dA residues of the helix are in the *anti* conformation (11). UVRR- and solvent-perturbation spectroscopy as well as the melting profile of $d(A^+-G)_{10}$ at 280 nm indicate that the dG residues of the helix are extrahelical, unstacked, and accessible to bulk solvent (10). They become hypochromic due to development of stacking with neighboring dA residues only as α -DNA is thermally disrupted (9). In the UVRR spectrum, a shift in the exocyclic amino scissors mode of the dA^+ but not the dG residues provides direct evidence for hydrogen bonds between A-N6-H and the electronegative oxygens of the phosphates.

From the foregoing, it becomes apparent that the basic repeating element of this helix is the dinucleotide $d(A^+pGp)$ with dA^+p and dGp in very different conformations. Additionally, energy-minimized molecular modeling (10) reveals a left-handed helical structure of ~ 12 Å in diameter, with unstacked dA^+ residues oriented toward the helix axis and linked to the $n-1$ phosphate oxygen, somewhat like the $-C=O \cdots H-N$ longitudinal hydrogen-bonding interactions in a protein α -helix (Fig. 3); hence the designation α -DNA.

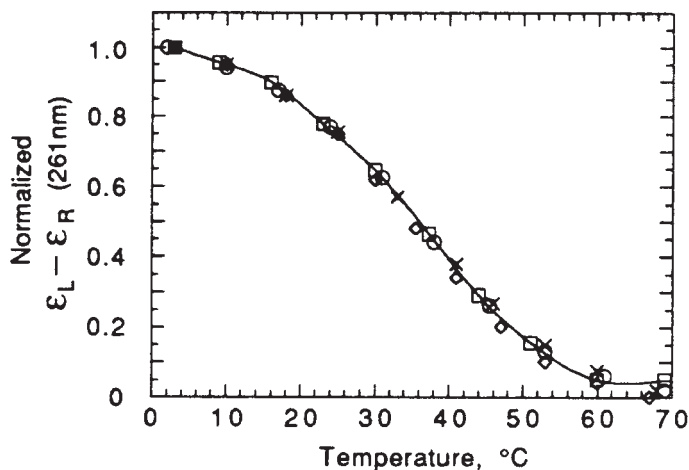


FIG. 4. CD melting profile of d(A-G)₁₀ in 0.001 M Na⁺, pH 4.0. Data points were taken from profiles measured at three oligomer concentrations varying from 0.82×10^{-3} to 0.64×10^{-4} M [per monomer unit] (8).

The thermal stability (T_m) of α -DNA as a function of n for d(A-G)_n was analyzed by CD-monitored melting profiles under conditions favorable for the single-stranded helical conformation — 0.001 M Na⁺, pH 4.0. The profile in Fig. 4 was unaffected by a 12.8-fold decrease in oligonucleotide concentration, consistent with the intramolecular structure (8). The most stable α -DNA formed by d(A-G)_{10,20} melts with T_m of $\sim 37^\circ\text{C}$ under these conditions. There is a slight decrease of T_m for d(A-G)₃₀, reflecting competition due to the overlapping equilibrium of α -DNA with the acid-intermolecular duplex. On increasing the salt concentration to 0.01 M, the observed biphasic melting makes this competition more obvious (Fig. 5, insert). Note that the low-temperature transition, which corresponds to the conversion of the duplex structure to the α -DNA helix, is relatively sharp compared to that for transition 2 at higher temperature, which is due to α -DNA melting. This observation is consistent with reports on the relative transition breadths and stabilities associated with B-like duplexes and intermolecular hairpin structures (14). Salt-dependent shifting of the α -DNA \rightleftharpoons acid-dependent duplex equilibrium reflects the higher charge density of the low-temperature ordered state (duplex) compared with the high-temperature state (α -DNA).

Our findings were confirmed by Vorličkova *et al.* (7), who reported that the related sequence, d(G-A)₁₀, undergoes a rapid (shorter than minutes) reversible and cooperative acid-induced transition into some ordered conformation. This transition, which is accompanied by an increase in CD intensity, nevertheless results in a loss in UV hypochromism, reflecting a loss

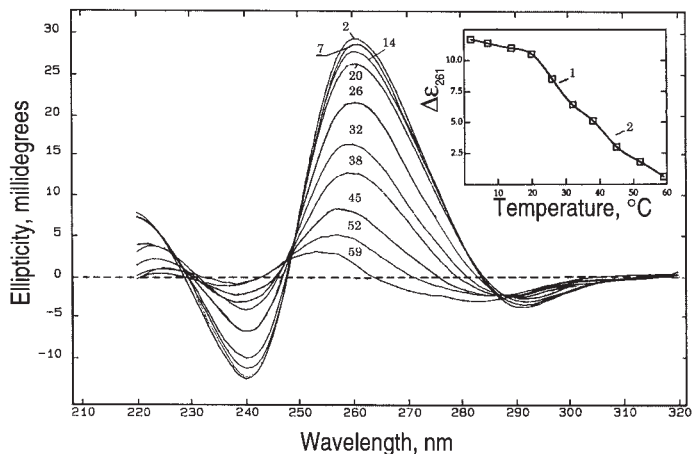


FIG. 5. CD spectra for $d(A-G)_{30}$ at pH 4.0 in 0.01 M Na^+ as a function of temperature. [Inset: Melting profile monitored by CD. Note the sequential transitions, 1 for the acid duplex and 2 for the α -helix. Melting is biphasic as there is a temperature boundary between the two conformations at this pH and ionic strength (after Ref. (12)).]

of base stacking. In contrast, the chimeric $d(G-A)_3(T-A)_4(G-A)_3$ and “mixed” $d(G-A-T-A)_5$ oligonucleotides of the same length manifest only very small acid-induced CD changes (15), suggesting that thymine is not an acceptable substitute for guanine in the α -DNA structure.

The uptake of three protons by $d(A-G)_{10}$ probably enables the minimum number of Coulombic interactions required to nucleate the intramolecular helix (9). It is apparent that additional protonation for sequences of $d(A-G)_n$ between $n=4$ and 10 helps to stabilize and extend the single-strand helical structure within the oligomer chain. But with longer sequences, there is an unexpected transformation of α -DNA to the acid-dependent duplex outside a narrow range of extremely low pH and ionic strength (12). Such transformation of an intramolecular structure to a duplex is not unprecedented. It is well known that single-stranded poly(A) is transformed into a parallel-stranded duplex when the adenines are protonated at N1 (16–18). In this case, duplex formation is also limited to sequences greater than $n=6$ –8 (19,20). It is not clear how this chain-length requirement relates to helix nucleation and propagation of this acid-dependent duplex with $A^+ \cdot A^+$ base pairs and $N6-H \cdots O=P$ hydrogen bonds.

Upon addition of ethanol to single-stranded $d(G-A)_{10}$ in low salt at pH 7, changes have been reported in the CD spectra that are very similar to those observed when pH is lowered (7). The intramolecular transition is cooperative and complete at about 40% ethanol, and UV hyperchromic changes occur

similar to those accompanying acid stabilization of α -DNA. In effect, Vorlíčková *et al.* (7) view the structure stabilized by ethanol as a variant of the α -DNA structure, which, due to the *presumed* absence of protonation of dA residues, lacks the ionic bonds between A^+ and $P-O^-$. Additional factors may help to stabilize this structure. Base-backbone hydrogen bonding could be stronger due to somewhat reduced competition from water for dA-residue hydrogen bonding. Also, the tendency for base stacking could be weakened due to the enhanced base solvation. The absence of ionic interactions between $A-N1-H^+$ and $P-O^-$, and the simultaneous effect of these factors could possibly compensate to provide the necessary stabilizing force for the formation of an ordered single-stranded conformer resembling that of α -DNA. Perhaps, unstacked alternating dA and dG residues cause the deoxyribose-phosphate backbone to take up a regular conformation, allowing effective longitudinal adenine-phosphate interaction (Fig. 3). In any case, indicative of the dominant contribution of the $A^+ \cdots P-O^-$ ionic bonds to the stability of α -DNA, the stability of the ethanol-induced α -DNA-like structure is much less than that of the acid-induced one (7,8). Notwithstanding these considerations (7), the possibility that the dA residues are weakly protonated at neutrality due to a large upward pK_a shift induced by the reduced dielectric constant of the solvent in the presence of ethanol cannot be excluded.

As might be expected, the pK_a for the α -DNA \Rightarrow duplex transition shows length dependence (12). The transition pK_a for d(A-G)₂₀ of 4.6 shifts to 4.4 for d(A-G)₃₀ in 0.001 M Na^+ and from 4.35 to 4.1, respectively, in 0.01 M Na^+ (Fig. 2). These changes are consistent with the finding of a reduced tendency to form α -DNA with increasing length of oligomer, i.e., a preference for the acid-dependent duplex. In contrast, the pK_a values for the transition from α -DNA to unstructured single strands are shifted toward neutrality with increasing oligomer length, cf. pK_a of 4.8 for d(A-G)₆ versus 5.57 for d(A-G)₁₀ in 0.01 M Na^+ and 4.5 for d(A-G)₆ versus 5.3 for d(A-G)₁₀ in 0.01 M Na^+ (9). This length dependence probably reflects the difference in the nucleation of the two types of helix, i.e., intra- versus intermolecular, requiring, as noted, just three sequential base-backbone interactions for α -DNA and many more for the duplex (see next section). It is also noteworthy that for all oligomers studied, d(A-G)_{6,10,20,30}, the upper pH limit of α -DNA stability is shifted toward neutrality on decreasing the ionic strength, consistent with stronger ionic bonding in the face of less charge shielding.

B. Acid-Dependent Parallel Duplex

The following features of this structure should be noted.

- The acid duplex is favored over the single-stranded α -DNA helix by increasing n in the sequence d(A-G)_n. A minimum length of $10 < n < 20$ is

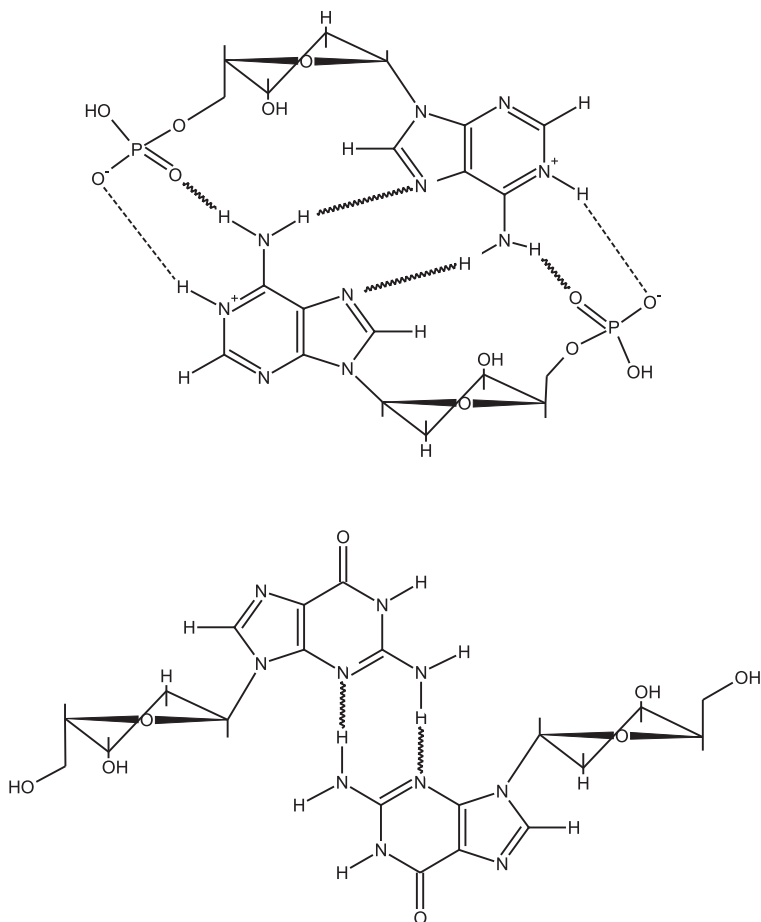


FIG. 6. Hydrogen- and ionic-bond schemes for the base pairs of the parallel homoduplex of $d(A^+-G)_n$. In the comparable homoduplex of $d(A-G)_n$, the interbase hydrogen bonds are the same, but the base-backbone hydrogen and ionic bonds for the A·A base pair are absent (12).

- necessary for acid-duplex formation (12), so that for $n \leq 10$, α -DNA prevails. However, when short C termini are present, shorter $d(G-A)$ repeats do form the parallel duplex in preference to α -DNA. These terminal dC residues probably provide additional stability to the potential duplex by forming $C^+ \cdot C$ base pairs (21,22) at its end.
- The parallel-stranded acid-dependent duplex contains $A^+ \cdot A^+$ and $G \cdot G$ base pairs (Fig. 6). The type of base pairing involved was elucidated using the chemical probes diethylpyrocarbonate (DEPC)

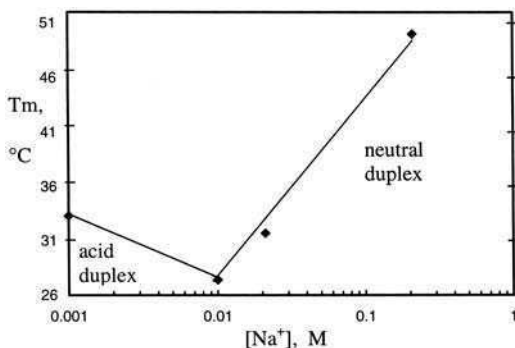


FIG. 7. Inversion of the ionic strength-dependence of the stability of d(A-G)₃₀ at pH 5.0, as determined from UV thermal melting profiles measured at 260 nm. Below 0.01 M Na⁺, the dependence is negative because increasing counterion concentration weakens the ionic bonds by shielding the negative phosphate charges and suppressing the pK_a of the dA residues. The ionic strength-dependence of T_m turns positive above 0.01 M Na⁺ as the A·A base pairs lose their charges, becoming the same as that for a Watson–Crick duplex at pH 6–7, with dT_m/dlog[Na⁺] = 18°C (after Ref. (12)).

and dimethylsulfate (DMS) (5,12). G·A heteropairs could be dismissed because protonation of A would be prohibited by the involvement of N1 of A as a hydrogen acceptor. Presumably, extensive interstrand stacking between A and G bases in GpA steps provides much of the necessary stabilizing force for the structure (21,22). Similar structural features have been revealed by NMR analysis (22) for the parallel acid duplex formed by d(C-G-A-G-A-G-A), i.e., all residues are oriented *anti* about the glycosyl bond; there are two interbase hydrogen bonds in the G·G and A⁺·A⁺ base pairs, and reciprocal interstrand hydrogen bonds from the N6–H of each A base to the phosphate oxygen at the same level on the opposite strand (Fig. 6). Additionally, in this case, reciprocal interstrand hydrogen bonds are claimed between the N1 of G and the O4' of the deoxyribose of the A immediately above or below. Given their presence in α -DNA (10) and in the acid poly(A⁺·A⁺) helices (16–18,23,24), base·backbone hydrogen bonds involving protonated A bases are not unexpected. Our studies confirm that d(A-G)_{20,30} can form stable acid-dependent duplexes at low pH and low salt, *but also pH-independent duplexes at Na⁺ concentration* ≥ 0.2 M (see Section III).

- In contrast to the behavior of Watson–Crick duplexes, the acid-dependent duplexes formed by d(A-G)_{20,30} display inverse dependence of T_m on Na⁺ concentration below 0.01 M (Fig. 7) (12). Thus, as in the case of the acid duplex poly(A⁺·A⁺), electrostatic attractions between protonated dA residues and phosphate groups (Fig. 6) in the

pH-dependent stacked base-paired structure apparently contribute more to helix stability than does the low salt concentration to the reduction of backbone-strand repulsion. Obviously, for single-stranded α -DNA, the inverse dependence of T_m on Na^+ concentration is more pronounced (9) than it is for the acid-dependent duplex.

- Increasing length of oligomer favors the structure, e.g., in 0.001 M Na^+ $T_m = 23.5^\circ\text{C}$ for $\text{d}(\text{A}^+ - \text{G})_{20}$ and 33.4°C for $\text{d}(\text{A}^+ - \text{G})_{30}$ (12).
- At constant pH, higher Na^+ concentration shifts the equilibrium to the pH-independent duplex (12).
- On going from 0 to 5% ethanol in 0.01 M Na^+ at pH 5.0, T_m of $\text{d}(\text{A} - \text{G})_{20}$ increases from 21.8 to 24°C , confirming that ionic interactions between protonated dA residues and the $\text{P}-\text{O}^-$ of the backbone phosphate moieties contribute more to stabilization of the acid-dependent duplex than does the resulting weakened base stacking to its destabilization (12). It is obvious, however, that the stabilization by ethanol of the $\text{d}(\text{A} - \text{G})_{20}$ duplex is not so pronounced as it is in the case of the single-stranded structure because the base stacking, increasingly advantageous for the double helix, is weakened in the lower dielectric medium.

III. Acid-Dependent Parallel Duplex \rightleftharpoons pH-Independent Parallel Duplex Equilibrium

At higher Na^+ concentration, the acid-dependent parallel-stranded duplex formed by $\text{d}(\text{A} - \text{G})_{20,30}$ has a counterpart in which the A bases are no longer protonated (12). The absence of protons on N1 of the dA residues eliminates their base \cdot backbone hydrogen bonds as well as the salt bridges, but the same interbase hydrogen bonds between stacked $\text{A} \cdot \text{A}$ and $\text{G} \cdot \text{G}$ pairs are maintained at the higher ionic strength as in the acid-dependent duplex. Undoubtedly, the reduced interstrand electrostatic repulsion due to the greater Debye–Huckel shielding compensates for the lost ionic bonds. The equilibrium between the two types of duplex are dependent on an interesting interplay of ionic strength and pH, that becomes readily apparent upon comparing the properties of the acid-dependent duplex described earlier with those of the pH-independent duplex.

A. pH-Independent Parallel Duplex

- These duplexes, which are formed by $\text{d}(\text{A} - \text{G})_{20,30}$ in 0.21 M Na^+ and stabilized by $\text{A} \cdot \text{A}$ and $\text{G} \cdot \text{G}$ base pairs (cf. Fig. 6) (structure III in Fig. 1), display complete pH-independence of T_m above pH 5.5, where



FIG. 8. Native PAGE assays for complexes formed by d(A-G)₂₀ and d(A-G)₃₀ at different oligomer concentrations. Gel is 12% polyacrylamide; running and sample buffer: 0.21 M Na⁺, pH 5.0; *T* = 4°C. Lane 1, control 40-bp duplex; lane 2, d(A-G)₂₀: 2.1×10^{-6} M; lane 3, 2.1×10^{-5} M; lane 4, 1×10^{-4} M; lane 5, 5.2×10^{-4} M; lane 6, control 60-bp duplex; lane 7, d(A-G)₃₀: 2.1×10^{-5} M; lane 8, 1×10^{-4} M; lane 9, 5.2×10^{-4} M; lane 10, 2.6×10^{-3} M (after Ref. (12)).

dA residues are no longer protonated (12). The pH-independent duplex formed by d(A-G)₂₀ undergoes cooperative melting with concomitant large CD and UV absorbance changes. Both the strandedness and the absence of a hairpin-helix conformation were confirmed by native gel-electrophoresis mobility assays (Fig. 8) and insensitivity to digestion by S₁ nuclease. Over a 10³-fold range of oligonucleotide concentration, d(A-G)₂₀ and d(A-G)₃₀ have mobilities comparable to the corresponding Watson-Crick duplexes, indicating homopurine · homopurine duplex formation in moderate salt.

The pH-independent parallel duplex formed by d(A-G)_{20,30} is essentially the same as the parallel-stranded duplex described by Jovin and coworkers (5,25). They confirmed the parallel orientation of d(G-A)_n strands by showing that the fluorescence of strands end-labeled with the fluor pyrene is quenched (5). However, our proposed structure with G_{anti} · G_{anti} base pairs, based upon the NMR data of Robinson *et al.* (22), differs from that of Rippe *et al.* (5), who described a model of the parallel-stranded d(A-G)_n duplex with G_{syn} · G_{syn} base pairs, which are energetically less favorable.

- These pH-independent duplexes are favored over their acid-dependent counterparts by both higher Na⁺ concentration and higher pH.
- Consistent with this behavior, these duplexes show positive ionic strength dependence of melting not unlike that for DNA duplexes of similar length (12) (Fig. 7).
- Also like Watson-Crick duplexes, these double-stranded helices are destabilized by decreasing the dielectric constant of the solvent.
- In contrast to poly(A⁺ · A⁺), which does not survive at neutral pH (16–18,23,24), neutral duplexes of d(A-G)_{20,30} derive their unique stability from the alternating presence of the G · G base pairs, which

NMR and model-building studies (21,22) have shown to stack especially well with their cross-strand nearest-neighbor dA residues.

It is noteworthy that some oligonucleotides which lack a strict alternation of dA and dG residues also form double-stranded self-structures (26,27), albeit with reduced stability in comparison to that of repetitive d(G-A)_n sequences (25). The thermal stability of these homoduplexes depend strongly on their length and the number of GpA steps, along with extrinsic factors such as counterions and pH. A plausible model consistent with the observed data is that of a parallel-stranded double helix. Since homopurine oligomers with an unequal content of G and A residues form stable duplexes, it follows that they can only contain hydrogen bonded A · A and G · G but not A · G base pairs, and that the strands can only be oriented parallel. In fact, oligonucleotides with such irregular homopurine sequences apparently adopt duplex structures very similar to that of oligonucleotides with strictly alternating sequences. This similarity provides independent support for the hydrogen-bonding schemes and chain orientation deduced for homopurine duplexes with alternating A and G residues. It is also worth noting that the symmetry inherent in a strictly alternating homopurine sequence means that such strands have the potential for more conformational possibilities than do irregular homopurine sequences, i.e., hairpin duplexes, concatemers, etc.

The foregoing examples notwithstanding, there are cases of enforced antiparallel A · G pairing. For example, at neutral pH, similar CD spectra are displayed not only by d(A-G)_n and d(G-A)_n, but also by d(G-A)₃(T-A)₄(G-A)₃, which is the same length as d(G-A)₁₀, but with the four central G residues replaced by thymines to give a central (T-A)₄ core (15). Although the orientation of chains in the d(G-A)₃(T-A)₄(G-A)₃ homodimer was not determined, an antiparallel arrangement with A · T and A · G base pairs seems reasonable, with the resultant central A · T base pairs enforcing antiparallel A · G pairing on the ApG/ApG tandem repeats. Such antiparallel A · G pairs have been shown to occur as well in DNA hairpins formed by repeating A-G sequences flanked by Watson-Crick base pairs (6,28).

IV. pH-Independent Parallel Duplex ⇌ Single-Hairpin Duplex ⇌ Two-Hairpin Tetraplex Equilibria

Regular sequences of d(A-G)_n have the potential of forming hairpin duplex structures. Native gel assays and S1 nuclease digestion provided evidence that in moderate ionic strength at neutrality and low temperature, there are just two equilibrium forms of the *short* oligomer d(A-G)₁₀, a single-hairpin duplex (structure IV in Fig. 1) and a two-hairpin tetraplex (structure V

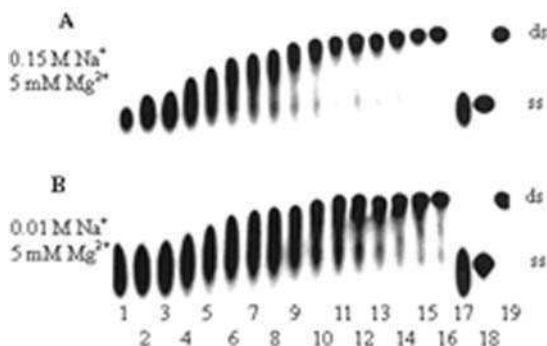


FIG. 9. Equilibrium between single-hairpin duplex and two-hairpin tetraplex, revealed by 16% nondenaturing PAGE of varying concentrations of d(A-G)₁₀ at pH 7.4 in 0.15 Na⁺/5 mM Mg²⁺ (A) or 0.01 M Na⁺/5 mM Mg²⁺ (B). Concentrations of d(A-G)₁₀: 6×10^{-4} mM (lanes 1 and 17); 8×10^{-4} mM (lane 2); 1, 2, 4, 6, and 8×10^{-3} mM (lanes 3–7); 1, 2, 4, 6, and 8×10^{-2} mM (lanes 8–12); 0.1, 0.2, 0.4, and 0.6 (lanes 13–16). Lane 18 contains single-stranded d(C-T)₁₀ at 0.6 mM and lane 19 contains the duplex d(A-G)₁₀ · d(C-T)₁₀ at 0.6 mM (13).

in Fig. 1), *but no linear duplex*. The short-hairpin duplex is probably entropically favored over the alternative longer linear duplex because the latter conformation is much more difficult to nucleate. The bimolecular tetraplex is favored over the single-hairpin duplex by both higher ionic strength (Na⁺ > 0.6 M or 5 mM Mg²⁺) and higher oligonucleotide concentration (13).

A. Single-Hairpin (Antiparallel) Duplex

A species of d(A-G)₁₀ at neutrality in moderate Na⁺ (0.15 M), migrating faster under nondenaturing gel-electrophoresis conditions than a random coil of identical chainlength, was interpreted as a compact hairpin duplex (Fig. 9) (13). The diffuse band migrating between the fastest species and the single-stranded random coil suggests a dynamic interconversion between intermediate conformations, i.e., a broad range of variously folded hairpins differing in the number of base pairs and unpaired overhangs, on a timescale comparable with their electrophoretic mobilities. A hairpin conformation is also consistent with results obtained with S1 nuclease, which under conditions previously shown by PAGE to stabilize the fully developed monomer hairpin, i.e., with no overhangs, preferentially cleaves d(A-G)₁₀ at the center of the chain. Some type of folded structure was also suggested for the repetitive sequence d(T₄-G-A-G-A)₄, but not for d(T₄-A-G-A-G)₄ (29). Apparently, whether a dA or dG residue occurs on the 5' side is significant (see below).

Under physiological conditions, longer d(A-G)_n oligonucleotides tend to form an intermolecular parallel-stranded duplex instead of a hairpin duplex

(12,30). Nevertheless, there is a wealth of data revealing the propensity of d(G-A)₂₀ to undergo either inter- or intramolecular duplex formation at pH values from 4.6 to 8.3 and moderate ionic strength (28). Several gel bands are observed with electrophoretic mobilities corresponding to hairpin helices, suggesting that various intermolecular associations are possible. However, at pH 7.0, significant proportions of hairpins are not observed, being replaced by linear duplexes. Such pH-dependence of helical form may reflect particular advantages of different charge distributions. As might be expected, at higher oligomer concentrations, the proportions of hairpin drop radically, in favor of intermolecular duplexes (28).

Notwithstanding the examples cited of parallel duplex formation by nonrepetitive sequences with homopurine base pairs (26,27), A·G pairing can be sterically accommodated within antiparallel structures in which there is alternation of dA and dG residues. Such an antiparallel-stranded duplex with A·G base pairs has, in fact, been shown to occur for d(G-A)₁₅ at neutrality, when forced by a run of flanking Watson–Crick pairs (6,28). It has been proposed (6) that the structure is stabilized by alternating $A_{anti} \cdot G_{anti}$ and $G_{anti} \cdot A_{syn}$ pairs since adenines located on the 3' side of the hairpin stem are significantly less reactive with DEPC (indicating that A–N7 participates in the interaction with G, as occurs in the $G_{anti} \cdot A_{syn}$ pairing scheme) than those located on the 5' side of the stem. At the same time, all the guanines of the d(G-A)₁₅ sequence are accessible to methylation by DMS, indicating that N7 of G is not involved in hydrogen bonding. Further, a combination of gel electrophoresis, chemical probing, and base substitution was used to characterize the folded structure, i.e., intramolecular hairpin duplex or two-hairpin tetraplex, formed by d(T₄-G-A-G-A)₄ (29). Substitution of G by 7-deazaG does not prevent formation of such a structure. Since a tetraplex would involve hydrogen bonding to N7 of G, this finding eliminates such a conformation. On the other hand, substitution of A residues by 7-deazaA does prevent the folded conformation from being adopted, suggesting that A–N7 is somehow involved in the base-pairing interactions of the hairpin (29). As a plausible hydrogen-bonding scheme between A and G residues consistent with these observations (6,29), we suggest one with sheared G·A base pairs (31), in which A–N6–H is hydrogen bonded to G–N3 and G–N2–H is hydrogen bonded to A–N7 (Fig. 10). The special geometry of sheared (side-by-side) tandem (i.e., alternating) G·A and A·G base pairs results in extensive interstrand stacking (for review, see Ref. (32)). In fact, tandem G·A and A·G mismatches have been found to be very stable in many different sequence contexts, both in short DNA (33–36) or RNA duplexes (37,38) and long DNA duplex fragments (39). In contrast, an isolated G·A pair in any of the other three possible configurations— $G_{anti} \cdot A_{anti}$ (40–42), $G_{anti} \cdot A_{syn}$ (43,44), and $G_{syn} \cdot A_{anti}^+$ (45,46) destabilizes a B-DNA duplex to varying extents. Additional

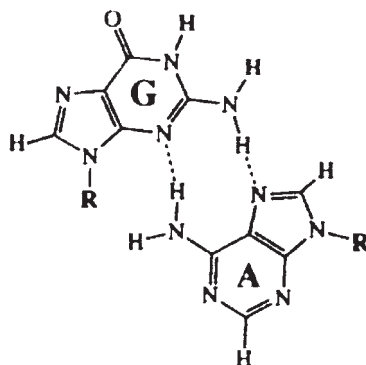


FIG. 10. Hydrogen-bonding scheme for a sheared side-by-side A·G base pair (31).

stabilization in hairpins of $d(A-G)_n$ may derive from two-residue (A·G or G·A) turns closed by a sheared Purine·Purine pair. Hirao *et al.* (47) have described a short hairpin with an A·A turn closed by a sheared G·A base pair. A similar turn motif, 5'-G-X-Y-A-3', was studied by NMR spectroscopy (48). The identity of the turn residues, X and Y, was found to play a negligible role in determining the stability of hairpin structures (48). However, the polarity of sheared pairs is very important for the stabilization of a hairpin turn. Thus, in the 5'-G-A-A-A-3' turn sequence, with a sheared G·A base pair closing the two-residue turn, there is a strong tendency to form a stable hairpin when it is flanked by only two Watson-Crick base pairs (47). In contrast, at least eight Watson-Crick base pairs were found necessary to maintain the hairpin in the case of the turn sequence 5'-A-A-A-G-3' (49), which can be considered as a prototype of the $d(A-G)_{10}$ sequence.

In view of these findings, we suggest as a reasonable hairpin structure for the short repetitive $d(A-G)_n$ or related $d(G-A)_n$ sequences, one with a two-residue turn closed by an A·G sheared base pair, stabilized by additionally sheared alternating A·G and G·A base pairs which are comparable in stability to complementary A·T base pairs (32,50). The minimal length of stem should be at least eight base pairs. In the case of the hairpin helix formed by $d(A-G)_{10}$, a single-hairpin duplex contains nine alternating A·G and G·A base pairs without terminal-unpaired residues.

Alternatively, from the point of view of forming a two-hairpin tetraplex directly from two preexisting single-hairpin duplexes, those hairpins would have to be stabilized by Hoogsteen-type G·G and probably A·A base pairs (51). The hydrogen-bonding potential of the bases of each of these Purine·Purine pairs would not be entirely satisfied (except by bonding to water), as they would still have the capacity to form two additional hydrogen

bonds along the unused edge to form the two-hairpin tetraplex (52). Moreover, the spatial distribution of these hydrogen-bonding positions in the case of G (N7-acceptor, O6-acceptor, N1-donor, N2-donor) is self-complementary, which would tend to promote self-dimerization to form all-G tetrads and indirectly partially hydrogen-bonded and intercalated all-A tetrads.

In the absence of residues interrupting the run of alternating A and G residues (as are present in telomeres), an all-purine triplet sequence provides a new turn motif (31,53). Remarkably, single-residue turns closed by sheared Purine · Purine pairs have been found possible, in fact extremely stable (31). The extraordinary hairpin-promoting ability of the sheared A · G, A · A, and G · G base pairs to close single-residue hairpin turns is due to strong stacking of a center purine base on the sheared Purine · Purine pair responsible for closing the turns (31,54), as revealed by NMR spectroscopy (31,53,54). This turn structure is made possible mainly by a change in the backbone phosphate torsion angles at the turn. All deoxyriboses in this turn motif adopt the C2'-endo pucker conformation, permitting unusual stacking of the furanose ring of the turn residue on the neighboring purine base to take place.

Both single-hairpin models described are reasonable. Interestingly, G · A pairs were shown to contribute importantly to the stability of an intramolecular hairpin formed by the sequence d(G-G-A)₁₀ flanked by Watson-Crick pairs, while G · G and A · A pairs stabilize the similar DNA hairpin containing d(G-G-G-A) repeats (55). However, the model with hydrogen bonding between dA and dG residues appears to be inconsistent with the finding of wavelength dispersion of melting of d(A-G)₁₀. Since the unstacking of dA residues is preferentially observed at 260 nm, and that of dG residues at 280 nm (56), one would expect a structure with A · G base pairs to display coincident-normalized melting profiles at those two wavelengths, but they are not. Instead, differential melting at those two wavelengths shows that the dA residues unstack at a lower temperature than the dG residues; in other words, reflecting differential unstacking of A · A and G · G base pairs under conditions corresponding to coexistence of single- and double-hairpin structures (13).

B. Two-Hairpin Tetraplex

A four-stranded complex stabilized by alternating G and A tetrads formed by polymeric d(A-G)_n sequences was proposed by Lee *et al.* (57–59). However, the melting data and model building they presented provided no convincing evidence that the structures studied were, in fact, tetraplexes rather than parallel-stranded duplexes.

The first direct evidence for antiparallel tetraplex formation was obtained by UVRR spectroscopy (60) and buttressed by gel electrophoresis and chemical probing of the deoxyoligonucleotide containing 10 A–G repeats (13).

Using selective excitation of dA and dG residues, it was found that this novel hairpin dimer is stabilized at higher Na^+ concentration and neutral pH by G tetrads alternating with nonhydrogen-bonded intercalated dA residues and not by AGAG tetrads. These results were corroborated by the finding of different UV melting profiles at 260 and 280 nm. These observations are consistent with findings of Murchie and Lilley (29), in their study of the hairpins and tetraplexes formed by the repetitive sequences $d(T-T-A-G-G-G)_4$ and $d(T-T-T-T-A-G-G-G)_4$ before and after replacement of particular dG residues with 7-deazaguanine residues. The lack of formation of an intramolecular hairpin helix after a single 7-deazaguanine residue substitution is completely consistent with formation of guanine base tetrads and excludes an AGAG tetrad model.

A structural requirement for a tetraplex formed by hydrogen bonding between two intramolecular hairpins is that adjacent antiparallel strands must have different glycosyl bond conformations. Consistent with this, it was determined by UVRR spectroscopy that dG residues in $d(A-G)_{10}$ are oriented both *anti* (mostly) and *syn* about the glycosyl bond (60). The central guanines were shown to be accessible to attack by S1 nuclease, whereas dG residues in the helical stems are not (Fig. 11) (13). Similar complexes stabilized by a continuum of G tetrads are formed by the sequence characteristic of telomeres from *Oxytricha* (61). Obviously, the two-hairpin dimers in the tetraplex can interact so as to form a number of electrophoretically distinct and structurally nonequivalent isomers (Fig. 12). For example, a two-hairpin tetraplex could be interlocked, with the turns at opposite ends. This type of complex has been observed using NMR spectroscopy in a structure containing G tetrads and T residues at the turns (62). Side-by-side association of single hairpins presents another way to form antiparallel tetraplexes (structure V in Fig. 1). Each hairpin with an overhanging dG residue can pair only with a limited subset of the other possible structures, which may explain the slow rate of hairpin dimer formation.

An intramolecular hairpin structure that can convert to a two-hairpin tetraplex is apparently more entropically favorable for such a short oligomer. At even quite high (millimolar) strand concentration, $d(A-G)_{10}$ is too short to nucleate a parallel duplex with $A \cdot A$ and $G \cdot G$ base pairs (or for that matter a linear tetraplex). On the other hand, the properties of complexes at neutral pH in 10 mM Mg^{2+} of the related shorter sequence, $d(G-A)_7G$, were interpreted in terms of linear parallel duplex formation involving $G \cdot G$ and $A \cdot A$ base pairs (5). Recognizing that poly(A) itself does not form a duplex with $A \cdot A$ base pairs at neutrality (16–20,23,24), it seems likely that the $G \cdot G$ base pairs help nucleate and contribute significantly to the stability of this parallel-stranded duplex, particularly by their strong tendency to stack. This may explain why parallel duplex formation prevails over intramolecular folding for the G-rich

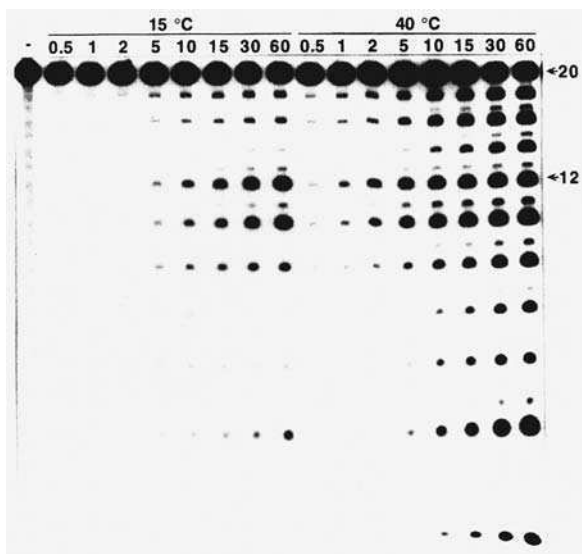


FIG. 11. S1-cleavage products of $d(A-G)_{10}$ at pH 7.4. PAGE (22% denaturing polyacrylamide) patterns obtained for cleavage at 15 and 40°C at the indicated times (min) of 0.64 mM $d(A-G)_{10}$ in 0.15 M Na^+ . Lane labeled (-) contains uncleaved oligonucleotide. The band corresponding to cleavage at residue G12 was identified by comigration with a band corresponding to uncleaved $d(A-G)_6$ (13).

oligomer $d(G-A)_7G$ despite being shorter than $d(A-G)_{10}$. Moreover, a two-hairpin conformer of $d(G-A)_7G$ would probably be less stable, given that it can form only three G tetrads.

There is some indication that a tetraplex may be formed by the longer repetitive $d(G-A)_{20}$ sequence at pH 8.3 (28). With increasing concentration, this oligonucleotide forms increasing amounts of a species containing two strands, that has an electrophoretic mobility intermediate between the mobilities of the unstructured single-strand and the intermolecular duplex. The equilibrium constant for formation of the latter complex is less by about an order of magnitude than that for the parallel-stranded duplex formed by $d(G-A)_{20}$ at pH 7.0. For all these reasons, a two-hairpin tetraplex appears to be the most reasonable structure, prevailing over linear parallel-stranded duplex at this elevated pH. This may be because a tetraplex better distributes the additional negative charges on some of the dG residues at the elevated pH, where they begin to ionize.

In almost all cases described, a smear is seen between the positions corresponding to hairpin and unfolded conformations, suggesting the

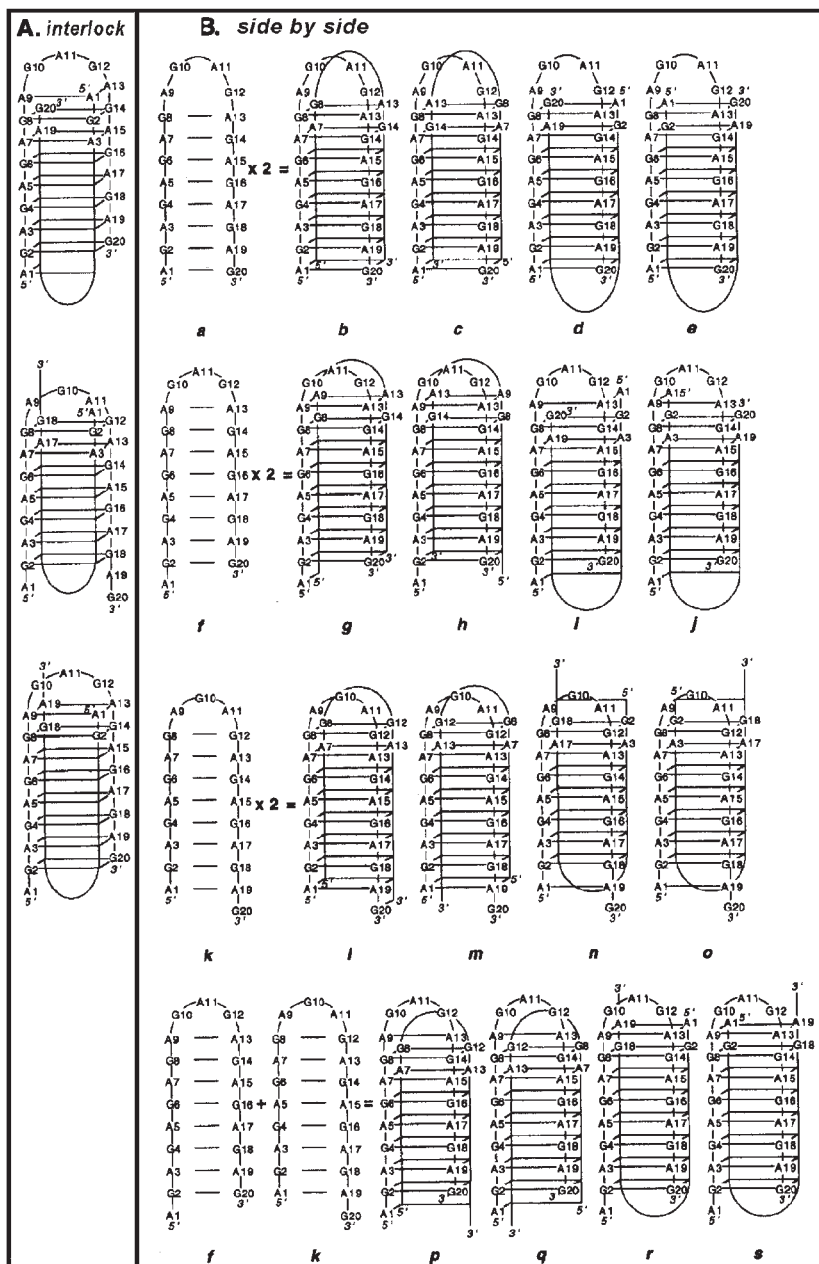


FIG. 12. Alternative interlocking and side-by-side conformers conceivable for the two-hairpin tetraplex(es) formed by d(A-G)₁₀ (13).

presence of multiple conformational species and/or a relatively slow interconversion between different forms. Analysis of the totality of the data indicates that the alternating sequences $d(A-G)_n$ or $d(G-A)_n$ differing in length of chain do not adopt any one stable conformation under defined conditions, but instead exist as an equilibrium between the intramolecular hairpin, bimolecular parallel-stranded duplex, and bimolecular tetraplex. These forms are in slow exchange on the NMR timescale. Conformational preferences must depend not only on the length of the A–G or G–A repeats, oligomer concentration, and solution conditions, but also on specific conformational nucleation, such as the presence of terminal residues that can form $C^+ \cdot C$ base pairs, flanking Watson–Crick base pairs, superstable turn sequences, etc.

V. Concluding Remarks

It is instructive to compare the two equilibria between intramolecular and intermolecular forms of $d(A-G)_n$: α -helix \rightleftharpoons acid-dependent duplex and single-hairpin duplex \rightleftharpoons two-hairpin tetraplex. Both equilibria are modulated primarily by ionic strength and oligonucleotide concentration, and the equilibrium between α -DNA and acid-dependent duplex is secondarily affected by pH and temperature. A general feature of both these equilibria is biphasic melting under boundary conditions, resulting in temperature-dependent conversion of the intermolecular structures, i.e., two-hairpin tetraplex or acid-dependent duplex to the more stable intramolecular hairpin and α -DNA helices, respectively, followed by second transitions to unstructured single strands (see, for example, Fig. 5).

For each conformation, then, there is a particular set of factors that defines its conformational space. Whether or not the dA residues are protonated and the ionic strength is low enough to permit those protonated residues to make ionic- and hydrogen-bond contacts intra- or intermolecularly, determines the presence or absence of stacked base pairs. The conformational space is further affected by the length of oligomer, since short length limits the probability for intermolecular helix nucleation, thereby allowing the potential for intramolecular conformations, the α -DNA helix and the hairpin duplex to be manifest, and from the latter, the intermolecular tetraplex.

We have described five discrete conformations undertaken by the sequence $d(A-G)_n$ between $n=6$ and 30, and the equilibria between the different conformers as conditions are varied. To present a phase diagram for the conformational possibilities for this seemingly simple repetitive sequence, a five-coordinate system is required that includes n , ionic strength, pH,

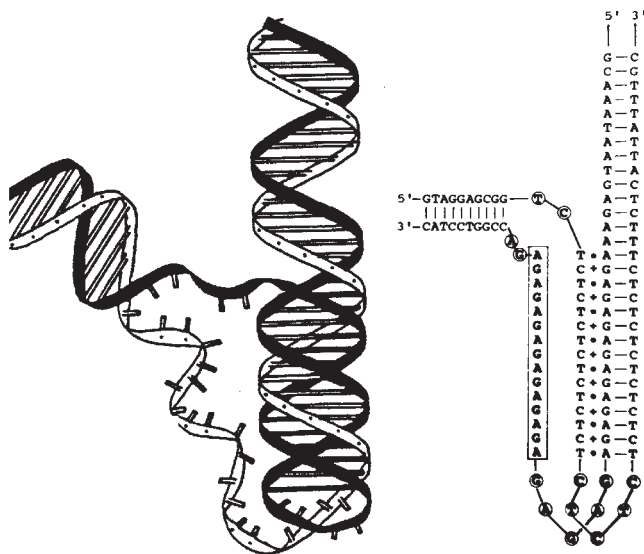


FIG. 13. Similarly oriented sequence (right) and analogous ribbon (left) diagrams of an H-DNA structural element with an intramolecular triplex and a “free” homopurine d(A-G)_n-strand segment capable of forming an α -DNA helix.

oligonucleotide concentration and temperature. The picture becomes even more complex when account is taken of the effect of specific nucleation sequences located on the termini of the oligonucleotides, such as C or Watson–Crick complementary residues.

That such sequences occur with greater than random expectation in regulatory elements of genomes is probably no mere coincidence. It will be interesting to see what roles the equilibria described here for the repetitive $d(A-G)_n$ sequence element may have in biology. One possibility is a role in the formation of H-DNA (intramolecular pyrimidine/parallel motif triplex) by affording a lower energy state for the homopurine-strand segment that becomes single stranded as a consequence of triplex formation (Fig. 13). In this connection, it may well be that the α -DNA helical structure is recognized and stabilized by A-G-A-G...-recognizing proteins (cf. Ref. (2)). Another is in affording single-strand secondary structural elements that make for the uneven DNA replication that must underlie sequence-extension mutations (63). Undoubtedly other effects will emerge.

Finally, we note that the knowledge gained from studies of the multiple conformations available to d(A-G)_n sequences such as those occurring in strand segments of one or the other complementary strands of DNA, along with the indications of their biological significance, should alert us to the

possibility of other repetitive strand elements whose conformational potential may have other functional consequences.

ACKNOWLEDGMENTS

For nearly a decade, the research described here from this laboratory was made possible by grants to J. R. F. from the National Institute of General Medical Sciences of the NIH. We are grateful to our colleagues Emory Braswell, John Fossella, Horst Klump, Laurence Lavelle, Ishita Mukerji, Mary Clare Shiber, Thomas Spiro, and Aylin Ulku, who each contributed to that research, and also to our colleague Olga Amosova for a critical reading of the present manuscript.

REFERENCES

1. R. D. Wells, D. A. Collier, J. C. Hanvey, M. Shimizu and F. Wohlrab, The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences, *FASEB J.* **2**, 2939–2949 (1988).
2. H. Manor, B. S. Rao and R. G. Martin, Abundance and degree of dispersion of genomic d(GA)_n.d(TC)_n sequences, *J. Mol. Evol.* **27**, 96–101 (1988).
3. R. Kolluri, T. A. Torrey and A. J. Kinniburgh, A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif, *Nucleic Acids Res.* **20**, 111–116 (1992).
4. A. Aharoni, N. Baran and H. Manor, Characterization of a multisubunit human protein which selectively binds single stranded d(GA)_n and d(GT)_n sequence repeats in DNA, *Nucleic Acids Res.* **21**, 5221–5228 (1993).
5. K. Rippe, V. Fritsch, E. Westhof and T. M. Jovin, Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex, *EMBO J.* **11**, 3777–3786 (1992).
6. D. Huertas, L. Bellolell, J. M. Casasnovas, M. Coll and F. Azorin, Alternating d(GA)_n DNA sequences form antiparallel stranded homoduplexes stabilized by the formation of G·A base pairs, *EMBO J.* **12**, 4029–4038 (1993).
7. M. Vorličková, I. Kejnovska, J. Kovanda and J. Kypr, Dimerization of the guanine-adenine repeat strands of DNA, *Nucleic Acids Res.* **27**, 581–586 (1999).
8. N. G. Dolinnaya and J. R. Fresco, Single-stranded nucleic acid helical secondary structure stabilized by ionic bonds: d(A⁺-G)₁₀, *Proc. Natl. Acad. Sci. USA* **89**, 9242–9246 (1992).
9. N. G. Dolinnaya, E. H. Braswell, J. A. Fossella, H. Klump and J. R. Fresco, Molecular and thermodynamic properties of d(A⁺-G)₁₀, a single-stranded nucleic acid helix without paired or stacked bases, *Biochemistry* **32**, 10263–10270 (1993).
10. M. C. Shiber, L. Lavelle, J. A. Fossella and J. R. Fresco, α -DNA, a single-stranded secondary structure stabilized by ionic and hydrogen bonds: d(A⁺-G)_n, *Biochemistry* **34**, 14293–14299 (1995).
11. I. Mukerji, M. C. Shiber, T. G. Spiro and J. R. Fresco, A UV resonance Raman study of d(A⁺-G)₁₀, a single-stranded helix without stacked or paired bases, *Biochemistry* **34**, 14300–14303 (1995).
12. N. G. Dolinnaya, A. Ulku and J. R. Fresco, Parallel-stranded linear homoduplexes of d(A⁺-G)_n>10 and d(A-G)_n>10 manifesting the contrasting ionic strength sensitivities of Poly(A⁺·A⁺) and DNA, *Nucleic Acids Res.* **25**, 1100–1107 (1997).

13. M. C. Shiber, E. H. Braswell, H. Klump and J. R. Fresco, Duplex-tetraplex equilibrium between a hairpin and two interacting hairpins of d(A-G)₁₀ at neutral pH, *Nucleic Acids Res.* **24**, 5004–5012 (1996).
14. J. G. Nadeau and P. T. Gilham, Anomalous hairpin formation in an oligodeoxyribonucleotide, *Nucleic Acids Res.* **13**, 8259–8274 (1985).
15. M. Vorličková, I. Kejnovská, J. Kovanda and J. Kypr, Conformational properties of DNA strands containing guanine-adenine and thymine-adenine repeats, *Nucleic Acids Res.* **26**, 1509–1514 (1998).
16. J. R. Fresco and P. Doty, Molecular properties and configurations of polyriboadenylic acid in solution, *J. Am. Chem. Soc.* **79**, 3928–3929 (1957).
17. J. R. Fresco, The X-ray diffraction patterns of solutions of the randomly coiled and helical forms of polyriboadenylic acid, *J. Mol. Biol.* **1**, 106–110 (1959).
18. A. Rich, D. R. Davies, F. H. C. Crick and J. D. Watson, The molecular structure of polyadenylic acid, *J. Mol. Biol.* **3**, 71–86 (1961).
19. R. D. Blake and J. R. Fresco, Thermodynamics of (A)_N·(U) from the dependence of T_M^N on oligomer length (see Appendix), *Biopolymers* **12**, 775–789 (1973).
20. J. Brahms, A. M. Michelson and K. E. van Holde, Adenylate oligomers in single- and double-strand conformation, *J. Mol. Biol.* **15**, 467–488 (1966).
21. H. Robinson and A.H.-J. Wang, 5'-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis, *Proc. Natl. Acad. Sci. USA* **90**, 5224–5228 (1993).
22. H. Robinson, J. H. van Boom and A.H.-J. Wang, 5'-CGA motif induces other sequences to form homo base-paired parallel-stranded DNA duplex: The structure of (G-A)_n derived from four DNA oligomers containing (G-A)₃ sequence, *J. Am. Chem. Soc.* **116**, 1565–1566 (1994).
23. W. M. Scovell, Structural and conformational studies of polyriboadenylic acid in neutral and acid solution, *Biopolymers* **17**, 969–984 (1978).
24. D. B. Lerner and D. R. Kearns, Proton and phosphorus NMR investigation of the conformational states of acid polyadenylic double helix, *Biopolymers* **20**, 803–816 (1981).
25. E. M. Evertsz, K. Rippe and T. M. Jovin, Parallel-stranded duplex DNA containing blocks of trans purine-purine and purine-pyrimidine base pairs, *Nucleic Acids Res.* **22**, 3293–3303 (1994).
26. S. B. Noonberg, J. C. Francois, T. Garestier and C. Hélène, Effect of competing self-structure on triplex formation with purine-rich oligodeoxynucleotides containing GA repeats, *Nucleic Acids Res.* **23**, 1956–1963 (1995).
27. B. Faucon, J.-L. Mergny and C. Hélène, Effect of third strand composition on triple helix formation: purine versus pyrimidine oligodeoxynucleotides, *Nucleic Acids Res.* **24**, 3181–3188 (1996).
28. J. M. Casasnovas, D. Huertas, M. Ortiz-Lombardía, J. Kypr and F. Azorín, Structural polymorphism of d(GA·TC)_n DNA sequences: Intramolecular and intermolecular association of the individual strands, *J. Mol. Biol.* **233**, 671–681 (1993).
29. A. I. Murchie and D. M. Lilley, Tetraplex folding of telomere sequences and the inclusion of adenine bases, *EMBO J.* **13**, 993–1001 (1994).
30. M. Ortiz-Lombardía, R. Eritja, F. Azorín, J. Kypr, I. Tejralova and M. Vorlickova, Divalent zinc cations induce the formation of two distinct homoduplexes of a d(GA)₂₀ DNA sequence, *Biochemistry* **34**, 14408–14415 (1995).
31. S.-H. Chou, L. Zhu, Z. Gao, J.-W. Cheng and B. R. Reid, Hairpin loops consisting of single adenine residues closed by sheared A·A and G·G pairs formed by the DNA triplets AAA and GAG: solution structure of the d(GTACAAAGTAC) hairpin, *J. Mol. Biol.* **264**, 981–1001 (1996).
32. S.-H. Chou, L. Zhu and B. R. Reid, Sheared purine-purine pairing in biology, *J. Mol. Biol.* **267**, 1055–1067 (1997).

33. M. Katahira, H. Sato, K. Mishima, S. Uesugi and S. Fujii, NMR studies of G·A mismatches in oligodeoxyribonucleotide duplexes modelled after ribozymes, *Nucleic Acids Res.* **21**, 5418–5424 (1993).
34. Y. Li, G. Zon and W. D. Wilson, Thermodynamics of DNA duplexes with adjacent G·A mismatches, *Biochemistry* **30**, 7566–7572 (1991).
35. J.-W. Cheng, S.-H. Chou and B. R. Reid, Base pairing geometry in GA mismatches depends entirely on the neighboring sequence, *J. Mol. Biol.* **228**, 1037–1041 (1992).
36. S. Ebel, A. N. Lane and T. Brown, Very stable mismatch duplexes: structural and thermodynamic studies on tandem G·A mismatches in DNA, *Biochemistry* **31**, 12083–12086 (1992).
37. H. A. Heus, S. S. Wijmenga, H. Hoppe and C. W. Hilbers, The detailed structure of tandem G·A mismatched base-pair motifs in RNA duplexes is context dependent, *J. Mol. Biol.* **271**, 147–158 (1997).
38. M. Wu, J. SantaLucia, Jr. and D. H. Turner, Solution structure of (rGGCAGGCC)₂ by two-dimensional NMR and the iterative relaxation matrix approach, *Biochemistry* **36**, 4449–4460 (1997).
39. S.-H. Ke and R. M. Wartell, The thermal stability of DNA fragments with tandem mismatches at a d(CXYG).d(CY'X'G) site, *Nucleic Acids Res.* **24**, 707–712 (1996).
40. L. S. Kan, S. Chandrasegaran, S. M. Pulford and P. S. Miller, Detection of a guanine X adenine base pair in a decadeoxyribonucleotide by proton magnetic resonance spectroscopy, *Proc. Natl. Acad. Sci. USA* **80**, 4263–4265 (1983).
41. G. G. Privé, U. Heinemann, S. Chandrasegaran, L. S. Kan, M. L. Kopka and R. E. Dickerson, Helix geometry, hydration, and G.A mismatch in a B-DNA decamer, *Science* **238**, 498–504 (1987).
42. E. P. Nikonowicz and D. G. Gorenstein, Two-dimensional proton and phosphorus-31 NMR spectra and restrained molecular dynamics structure of a mismatched GA decamer oligodeoxyribonucleotide duplex, *Biochemistry* **29**, 8845–8858 (1990).
43. T. Brown, W. N. Hunter, G. Kneale and O. Kennard, Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations, *Proc. Natl. Acad. Sci. USA* **83**, 2402–2406 (1986).
44. G. D. Webster, M. R. Sanderson, J. V. Skelly, S. Neidle, P. F. Swann, B. F. Li and I. J. Tickle, Crystal structure and sequence-dependent conformation of the A·G mispaired oligonucleotide d(CGCAAGCTGGCG), *Proc. Natl. Acad. Sci. USA* **87**, 6693–6697 (1990).
45. T. Brown, G. A. Leonard, E. D. Booth and J. Chambers, Crystal structure and stability of a DNA duplex containing A(anti).G(syn) base-pairs, *J. Mol. Biol.* **207**, 455–457 (1989).
46. X. Gao and D. J. Patel, G(syn)·A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-dependent conformational transition of G·A mispairs detected by proton NMR, *J. Am. Chem. Soc.* **110**, 5178–5182 (1988).
47. I. Hirao, Y. Nishimura, T. Naraoka, K. Watanabe, Y. Arata and K. Miura, Extraordinary stable structure of short single-stranded DNA fragments containing a specific base sequence: d(GCGAAAGC), *Nucleic Acids Res.* **17**, 2223–2231 (1989).
48. P. Sandusky, E. W. Wooten, A. V. Kurochkin, T. Kavanaugh, W. Mandecki and E. R. Zuidervog, Occurrence, solution structure and stability of DNA hairpins stabilized by a GA/CG helix unit, *Nucleic Acids Res.* **23**, 4717–4725 (1995).
49. I. Hirao, M. Ishida, K. Watanabe and K. Miura, Unique hairpin structures occurring at the replication origin of phage G4 DNA, *Biochim. Biophys. Acta* **1087**, 199–204 (1990).
50. M. Ortiz-Lombardía, A. Cortés, D. Huertas, R. Eritja and F. Azorín, Tandem 5'-GA:GA-3' mismatches account for the high stability of the fold-back structures formed by the centromeric *Drosophila* dodeca-satellite, *J. Mol. Biol.* **277**, 757–762 (1998).

51. E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco and E. H. Blackburn, Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs, *Cell* **51**, 899–908 (1987).
52. W. I. Sundquist and A. Klug, Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, *Nature* **342**, 825–829 (1989).
53. S. Yoshizawa, G. Kawai, K. Watanabe, K. Miura and I. Hirao, GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins, *Biochemistry* **36**, 4761–4767 (1997).
54. I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe and K. Miura, Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat, *Nucleic Acids Res.* **22**, 576–582 (1994).
55. D. Huertas and F. Azorín, Structural polymorphism of homopurine DNA sequences. d(GGA)_n and d(GCGA)_n repeats form intramolecular hairpins stabilized by different base-pairing interactions, *Biochemistry* **35**, 13125–13135 (1996).
56. J. R. Fresco, L. C. Klotz and E. G. Richards, A new spectroscopic approach to the determination of helical secondary structure in ribonucleic acids, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 83–90 (1963).
57. J. S. Lee, D. A. Johnson and A. R. Morgan, Complexes formed by (pyrimidine)_n-(purine)_n DNAs on lowering the pH are three-stranded, *Nucleic Acids Res.* **6**, 3073–3091 (1979).
58. J. S. Lee, D. H. Evans and A. R. Morgan, Polypurine DNAs and RNAs form secondary structures which may be tetra-stranded, *Nucleic Acids Res.* **8**, 4305–4320 (1980).
59. J. S. Lee, The stability of polypurine tetraplexes in the presence of mono- and divalent cations, *Nucleic Acids Res.* **18**, 6057–6060 (1990).
60. I. Mukerji, M. C. Shiber, J. R. Fresco and T. G. Spiro, A UV resonance Raman study of hairpin dimer helices of d(A-G)₁₀ at neutral pH containing intercalated dA residues and alternating dG-tetrads, *Nucleic Acids Res.* **24**, 5013–5020 (1996).
61. L. A. Klobutcher, M. T. Swanton, P. Donini and D. M. Prescott, All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus, *Proc. Natl. Acad. Sci. USA* **78**, 3015–3019 (1981).
62. F. W. Smith and J. Feigon, Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides, *Nature* **356**, 164–168 (1992).
63. G. R. Sutherland and R. I. Richards, Dynamic mutations: A newly discovered class of genetic mutations provides fresh insight into the curious inheritance patterns of some human diseases, *Am. Sci.* **82**, 157–163 (1994).

This Page Intentionally Left Blank

Index

A

ABBP-1, 31
 ABBP-2, 31–32
 ABC-R, 94
 ACF-CUGBP2 complex, 29
 ACF mRNA, 23
 Acid-dependent parallel duplex, 329–332
 Acid-dependent parallel
 duplex = pH-independent parallel
 duplex equilibrium, 332–4
 Activation peptide C4a, 275
 Adenocorticotrophic hormone (ACTH), 174,
 180, 182, 199
 Adipose tissue-specific glucocorticoid
 activation by 11-SD 1, 200
Alicycobacillus acidocaldarius, 298
 α -DNA, 325–329
 thermal stability, 327
 α -DNA = acid-dependent duplex equilibrium,
 327
 α -DNA helix = acid-dependent duplex
 equilibrium, 323–34
 α_2 macroglobulin protein family (A2M),
 267–270
 Amino acid sequences, 123–124
 Anaphylatoxins (ANAs), 240, 270
 Ancestral haplotypes (AHs), 232
 ApoB
 C to U RNA editing, 5–9
 cis-acting elements, 14–17
 trans-acting factors, 17–32
 ApoB RNA-editing core components, 17–26
APOBEC1 gene, 20–22
 Apobec-1, 17–22
 RNA-binding function of, 19
 Apobec-1-interaction, 30
 Apobec-1 complementation factor (ACF), 6,
 17, 22–26
 see also ACF; Human ACF
 Apparent mineralocorticoid excess (AME),
 182, 184–186
Arabidopsis thaliana, 82
 Arginine vasopressin (AVP), 180
 Arylsulfonamidothiazoles, 202

ATP, 89–90, 92–93, 146, 297–298
 ATP-binding cassette (ABC), 94–95
 ATPase, 146, 148, 158
 AU-rich elements (ARE), 15, 44, 64
 AU-rich sequence element (ARE), 19
 AUX240, 32

B

Bacillus stearothermophilus, 129
Bacillus subtilis, 82, 129, 131, 295, 298–301
 11 β -HSD 1, 174, 186–202
 adipose tissue-specific glucocorticoid
 activation by, 200
 characteristics, 184
 dehydrogenase or oxoreductase?, 186–187
 as dimeric enzyme, 188–189
 elevated levels in human adipose tissue,
 200–201
 gene organization, 191–192
 in glucocorticoid activation, 188–189
 glucocorticoid-induced, 197
 glycosylation status impact, 193–194
 hepatic, 196
 history of, 181–182
 identification of *cis*-acting elements,
 194–195
 inhibitors, 201–202
 involvement of *trans*-acting elements,
 195–196
 and metabolic syndrome, 199–201
 in obesity, 201
 in omental adipose tissues, 199
 (patho)physiological role, 182–183
 peptide, 192
 predominant *in vivo* reaction direction, 183
 as prereceptor regulative in intracellular GC
 signaling, 185
 regulation, 194
 regulation of reaction direction, 187–188
 schematic representation, 194
 structural determinants, 190–191
 11 β -HSD 1B
 structure, 192
 truncated, 192–193

- 11 β -HSD 2, 202
 - characteristics, 184
 - history of, 181–182
 - inhibition, 201
 - (patho)physiological role, 182–183
 - predominant *in vivo* reaction direction, 183
 - and syndrome of apparent mineralocorticoid excess, 184–186
- 11 β -hydroxysteroid dehydrogenase (11 β -HSD)
 - as prereceptor control, 180–181
- BHK cells, 72
- Biological membranes, 70

C

- C3, 219
 - evolution, 267–274
- C4, 217–292
 - determining 1, 2, 3, or 4 long or short loci, 223–226
 - divergence and polymorphisms of mouse and Slp proteins, 242–244
 - evolution, 267–274
 - genetic diversities, 221–232
 - genetic study, 221
 - historic background, 221–223
 - in MHC-associated diseases, 276
 - mobility polymorphism of α -chain and β -chain, 241
 - nucleotide and amino acid polymorphisms or mutations, 254
 - polygenic and gene size variations
 - correlation with serum protein concentrations and hemolytic activities, 256–257
 - polygenic and gene size variations in different races, 274–275
 - polygenic variations in Caucasians, 226–229
 - polymorphisms of mouse and Slp proteins, 235
 - post-translational processing, 275
 - presence on cell surfaces, 237–238
 - protein structures, 233
 - roles on tolerance and autoimmunity, 274
 - structural and functional basis of polymorphisms, 275
 - structural diversities of the human and mouse proteins, 232–245
 - structural variation created by incomplete processing, 236
 - three-dimensional structure, 275
 - transcripts present in multiple tissues, 255–256
 - tyrosine sulfation at carboxyl end of ain in, 232–234
 - variation in expression levels of mouse and Slp proteins, 259–260
- C4 gene
 - endogenous retrovirus mediating dichotomous size variation, 260–267
 - exon–intron structures of long and short, 246
 - genomic DNA sequence from first RCCX module, 247–253
- C4 gene expression
 - cytokines/reagents tested to affect, 256
 - regulatory motifs, 257–259
- C4 gene frequencies in Ohioan and Hungarian Caucasians, 228
- C4-deficient mice, phenotypes, 244–245
- C4A, 219–223, 225, 233–234, 270
 - activation peptide, 240–241
 - allotype heterogeneity, 241–242
 - deficiency and disease associations, 274
 - expression of transcripts and proteins, 255–260
 - isotypic residues, 237
 - local expression and polygenic variation, 275
 - protein sequence polymorphism and functional variation, 236–242
- C4A genes, 259, 274
 - nucleotide polymorphisms, 245–253
- C4A6, 239–240
- C4B, 219–223, 225, 233–234, 270
 - allotype heterogeneity, 241–242
 - expression of transcripts and proteins, 255–260
 - isotypic residues, 237
 - local expression and polygenic variation, 275
 - protein sequence polymorphism and functional variation, 236–242
- C4B genes, 259, 274
 - nucleotide polymorphisms, 245–253
- C4d
 - physiological relevance on erythrocytes and follicular dendritic cells, 276
 - X-ray crystal structure, 238–239
- C5, 219
 - binding site, 239–240
 - evolution, 267–274

- C. glutamicum*, 309
Caenorabditis elegans, 24
 Carbenoxolone derivative of glycyrrhetinic acid, 201
 CCCH tandem zinc finger proteins, 43–68
 CDP-choline, 84, 86
 CDP-ethanolamine, 73–74, 76, 82–84, 86
 CDP-ethanolamine:1,2-diacylglycerol ethanolaminephosphotransferase, 83
 C/EBP α , 195
 C/EBP β , 195–196
 Chido (Ch) blood-group antigens, 237–238
 Chinese hamster ovary cells *see* CHO cells
 CHO cells, 72
 CHO-K1 cells, 74–76, 84–86, 88–89, 98
 CHO-K1 mutants, 74
cis-acting elements, C to U ApoB RNA editing, 14–17
Clostridium perfringens, 82
 Complement *see* C3, C4, C5
 Congenital adrenal hyperplasia (CAH), 229
 RCCX modular variations in, 231
 Corticotrophic-releasing hormone (CRH), 180
 Cortisol-binding protein (CBP), 177
 Cortisol levels, 199
 Cortisol-secretion rate, 199
 C-terminal repeat domain (CTD), 28
 CTP:phosphocholine cytidyltransferase, 86–87
 CTP:phosphoethanolamine cytidyltransferase, 83
 CUGBP1, 30
 CUGBP2, 28–31
 mRNA splicing, stability, and translational control functions, 30
 Cushing's disease, 200
 Cushing's syndrome, 199–201
 CYP21 genes, 273
 Cytidine deaminase, 17–22
 Cytokines
 effects on 11 β -SD 1, 198
 glucocorticoid inhibition, 176
 proinflammatory, 178
 Cytokines/reagents tested to affect *C4* gene expression, 256
- ## D
- d(A-G)₁₀, 327–328
 S1-cleavage products, 340
 two-hairpin tetraplex(es) formed by, 341
 d(A-G)₂₀, native PAGE assays for complexes formed by, 333
 d(A-G)₃₀
 CD spectra, 328
 native PAGE assays for complexes formed by, 333
 stability, 331
 d(A-G)_n, 321–347
 conformational variants, 322
 homoduplex, 330
 secondary structures, 323
 structural form, 325
 Deoxyoligonucleotide, 338
 Diethylpyrocarbonate (DEPC), 330–336
 Dimethylsulfate (DMS), 331
 Dinucleotide polymorphism, 47
 Displacement loop formation, 161
 DMPK, 30
 DMS, 336
 DNA cleavage and ligation activities of initiator proteins, 124–126
 DNA damage, 155–159
 effect on HSV-1 DNA helicases, 157–158
 effect on HSV-1 DNA polymerase, 156–157
 DNA-protein complex, 119
 DNA-protein interactions, 113–137
 Double rolling-circle replication
 template, 159
 Double-strand origins of rolling-circle replicating plasmids, 116–119
Drosophila, 70, 82–83, 96–97, 270
Drosophila melanogaster, 24
 DYEnamic Direct cycle sequencing, 47
 DYEnamic ET primers, 47
- ## E
- Ehlers Danlos syndrome (EDS), 230
 Electrophoretic mobility shift assays (EMSA), 119
 Endoplasmic reticulum (ER), 79–81, 84, 86–91, 188
Enterococcus hirae, 299, 311–312
 Environmental Genome Project, 45
 Epstein-Barr virus (EBV), 140
Escherichia coli, 70, 81, 84–85, 131, 140, 154, 190, 192, 294, 297–300, 302, 304–306, 309–312
 Eukaryotic cells, 70, 72

Exon-intron structures of long and short *C4* genes, 246

F

Folding algorithms, 15
Free hormone hypothesis, 177

G

GC-GR complex, 176
Gel electrophoresis, 338
GGST motif, 85
Glucocorticoid activation, 11 β -HSD 1 in, 188–189
Glucocorticoid homeostasis, 179
Glucocorticoid metabolism, 173–215
Glucocorticoid receptor (GR), 174, 177–178
 ligands-activated, 175
Glucocorticoid resistance, 179–180
 familial, 180
 pathophysiology, 179–180
Glucocorticoid response elements (GREs), 175–176
Glucocorticoids
 effects on 11 β -HSD 1, 197–198
 and immune response, 176
 mode of action, 202
 molecular mechanism, 175–176
 negative action, 176
 nongenomic action, 178–179
 overview, 174–175
 physiological response, 179
 sensitivity, 179
Glycosylation of *C4*, structural variation
 created by difference in, 234–235
Glycyrrhetic acid, carbenoxolone
 derivative of, 201
Gram-negative bacteria, 295
Gram-positive bacteria, 295
Granulocyte-macrophage colony-stimulating
 factor (GM-CSF), 43
Growth hormone (GH), effect on 11 β -HSD 1, 198–199
GRY-RBP, 26–28
 mRNA splicing and stability functions, 27

H

Hairpin duplex structures, 334–342
Halobacterium halobium, 298

Halofeax volcanii, 298
H-DNA (intramolecular pyrimidine/parallel
 motif triplex), 343
Helicase-primase, 153
Herpes Simplex Virus (HSV)
 type-1, 139–171
 type-2, 140
Herpes viruses, classification, 140
Herpesviridae, 140
HERV-K(*C4*)
 family members in human genome,
 265–266
 in human and primate genomes, 263
 in intron 9 of long *C4* genes, 260–263
 LTRs of, 263–267
 solitary LTRs (solo-LTR) of, 268–269
hGR α , 178
hGR β , 178
High mobility group protein-1 (HMG-1), 149
Histidine decarboxylase, 84
hnRNP-C1, 32
HSV-1
 elongation, 151–154
 genome structure, 140–144
 homologous recombination, 159–163
 in vitro reconstitution of DNA replication,
 154
 initiation of replication, 144–151
 interaction of DNA replication machinery
 with DNA damage, 155–159
 life cycle, 140–144
 origin-dependent initiation of DNA
 replication, 152
 origins of replication, 144–145
 recombination, 159–160
 recombination-mediated DNA replication,
 162–164
 religation proteins, 156
 replication fork enzymes, 151–154
 replication genes, 140–144
 role of host proteins, 149–150
 strand exchange, 161–162
 strand invasion, 160–161
HSV-1 DNA helicases, effect of DNA damage
 on, 157–158
HSV-1 DNA polymerase, effect of DNA
 damage on, 156–157
HSV-1 DNA replication proteins, 144
Human ACF, 24
Human ACF gene, 25

Human adipose tissue, elevated 11β -HSD 1 levels in, 200–201
 Human CCCH proteins, genes encoding, 53
 Human gene *ZFP36*, 44
 Human Herpes Virus (HHV), 1–8, 140
 Human keratinocytes, 73
 Hybridoma cells, 73
 Hydrogen bonding between two intramolecular hairpins, 339
 Hydrogen-bonding scheme for sheared side-by-side A–G se pair, 337
 Hydroxysteroid dehydrogenases (HSDs), 181 physiological role, 181

I

ICP0, 142
 ICP22, 142
 ICP27, 142
 ICP47, 142
 ICP4, 142
 ICP8, 147–148, 150–151, 153–154, 156, 158, 160–163
 ICP8-UL9 complex, 147
 IMP, 271
 Initiator proteins, DNA cleavage and ligation activities, 124–126
 Insertion–deletion RNA editing, 2
 Insulin-like growth factor-I (IGF-I), effect on 11β -HSD 1, 198–199
 Inverted repeat (IR) elements, 116
 IRI, 116
 IRII, 116–121, 125–126, 128, 131
 IRIII, 116, 118

K

K^+
 channels, 303–304
 future directions, 312–313
 as intracellular activator, 309–311
 movement paths, 296–305
 regulation of internal pH by, 311–312
 roles in bacteria, 294
 transport in bacteria, 296
See also Potassium
 K^+ -efflux systems, 301–302
 K^+ -influx systems, 296–301
 K^+ -specific efflux systems, 306
 K^+ -transport systems, regulation, 305–309

K^+ -uptake systems, 300–301
 control of activity, 305
 control of expression, 306–309
 Kdp system, 299–300, 306–309
 Kephalin, 71–72
 Ktr system, 298–299

L

Lactobacillus lactis, 310
 Lesion bypass model, 158–159
 Leydig cells, steroidogenesis in, 187
 LGST motif, 85
 Liquorice
 medicinal properties, 201
 and mineralocorticoid side effects, 201
 Liver X receptors, 196
 Long-terminal repeats (LTRs), 259
 HERV-K(C4), 263–267
see also LTR
 Low-density lipoprotein receptor (LDLR), 7, 9
 Low-density lipoproteins (LDL), 9
 L–S junctions, 159
 3' LTR, 260–261
 5' LTR, 260–261
 LTR–LTR recombination, 262

M

Major histocompatibility complex *see* MHC
 Mammalian cell membranes, 71
 Mammalian cells, 72–73
 MBL/MASP complex, 270
 Membrane phospholipids, 70–71
 Metabolic syndrome and 11β -HSD 1, 199–201
Methanobacterium thermoautotrophicum, 303
Methanococcus jannaschi, 298
 Methylammonium, 302
 MHC, 217–292
 RCCX in, similarities and dissimilarities of human and mouse, 271–274
 RCCX modules in, 267–274, 276
 MHC-associated diseases, C4 in, 276
 MHC class III, 232
 gene organizations of human and mouse, 272
 MHC complement gene cluster (MCGC), 271
 MHC haplotype coding, 223–226
 Mineralocorticoid receptors (MRs), 184
 Mineralocorticoid side effects, 201

Mitochondria-associated membranes (MAM), 80–81, 87–91
 Mitogen-activated protein (MAP) kinases, 64
 Multidrug resistance proteins (MDR and MRP), 94
Mycobacterium tuberculosis, 304
 Myotonic dystrophy, 30

N

NAT1, C to U RNA editing, 12–14
 Netrin domain (NTR), 267
 NF1, C to U RNA editing, 10–12
 NS1-GRY-RBP interaction, 27
 NTR modules, 270
 Nuclear export signal (NES), 19
 Nucleotide polymorphisms in human *C4A* and *C4B* genes, 245–253
 Nucleotide sequence of plasmid pT181 origin of replication, 117

O

OpuA, 310–311
 Origin-binding factor (OF-1), 143, 145

P

P. pastoris, 192
 PABP, 27
 PAIB, 27
 PCNA, 153
 PCR-amplification conditions, tristetraprolin (TTP) family of CCH tandem zinc finger proteins, 46–47
 PCR products, 46–47
PDR1, 94
PDR3, 94
 PFGE, 226
 pH regulation by K^+ , 311–312
 pH-independent parallel duplex, 332–334
 pH-independent parallel duplex = single-hairpin duplex = two-hairpin tetraplex equilibria, 334–342
 Phosphatidylethanolamine (PE), 69–118
 biological functions, 96–97
 biosynthesis pathways, 71–74
 biosynthesis regulation, 86–87
 enzymes of biosynthesis, 82–85
 future directions, 97–98

 interorganelle transport in mammalian cells, 87–91
 interorganelle transport in yeast cells, 91–92
 intramembrane transport, 92–95
 Phosphatidylserine biosynthesis regulation, 85–86
 Phosphatidylserine (PS), 69–118
 biological functions, 95–96
 biosynthesis pathways, 71–74
 decarboxylation pathway, 84–85
 enzymes of biosynthesis, 74–82
 future directions, 97–98
 interorganelle transport in mammalian cells, 87–91
 interorganelle transport in yeast cells, 91–92
 intramembrane transport, 92–95
 Phosphatidylserine synthase-1 (PSS1), 72–73
 mammalian, 74–76
 metabolic difference between PS synthase-2 and, 77–79
 Phosphatidylserine synthase-1 (PSS1) cDNA, 75
 Phosphatidylserine synthase-2 (PSS2), 72–73
 mammalian, 76–77
 metabolic difference between PS synthase-1 and, 77–79
 Phosphatidylserine synthases, subcellular location, 79–81
Physarum polycephalum, 2
 Plasmids *see* pT181 rolling-circle replication of plasmids
 Polymorphic frozen blocks, 232
 Polypurine track (PPT), 260
 Potassium in bacteria, 293–320
 see also K^+
 PPAR γ , 196
 Primer-binding site (PBS), 260
 Prokaryotes, phosphatidylserine synthase in, 81–82
 Prokaryotic membranes, 71
 ProQ, 310
 Protein–protein interaction, 114, 233
 Proton-motive force (PMC), 297
 PSA-3 cells, 76, 86
PSD1, 84, 92
PSD2, 84, 92
 pT181 family
 future direction, 133
 initiator proteins, 122–126
 origins of replication of plasmids, 116–122

Rep proteins, mechanism of inactivation, 126–129
 replication
 initiation and termination, 119–122
 of lagging strand, 132
 rolling-circle replication of plasmids, 113–137
 pT181 famitl, replication enhancer, 122
 pT181 famiyl, RC replication, host proteins involved in, 129–132
 Purine–purine pairs, 338

R

Ras GTPase activating protein (GAP) activity, 10
 Rat mammary carcinoma cells, 73
 RCCX
 de novo recombinations or gene conversion-like events, 231–232
 genotypic and phenotypic analyses of structures, 225
 in MHC, similarities and dissimilarities of human and mouse, 271–274
 modular variations in Caucasians, 226–229
 RCCX length variants, 232
 detection of, 224
 heterozygosities, 229
 in Ohioan and Hungarian Caucasians, 228
 RCCX modules, 220–232, 258
 in Caucasian population, 227
 constituent genes, 222
 in MHC, 267–274
 located in MHC, 276
 structures in humans and in mice, 273
 variations in CAH, 231
 Recombinant ACF, 22
 Rep–PcrA helicase interactions, 129–131
 Rep proteins, 114–116, 119, 122–123
 biochemical activities, 123–126
 RepC protein, 122, 124
 RepC/RepC* protein, 126, 128
 RepD protein, 122, 124–125
 RepD/RepD* protein, 126, 128
 Replication–initiation complex, 119
 Replication terminator protein (RTP), 131
 Restriction fragment length polymorphism (RFLP), 47–48, 60, 223, 226
 Rev response element (RRE), 16
 RNA-binding function of apobec-1, 19
 RNA editing, 2–41

A to I, 4
 C to U, 4–14
 in mammals, 4
 of ApoB, 5–9
 cis-acting elements, 14–17
 core components, 17–26
 trans-acting factors, 17–32
 of NAT1, 12–14
 of NF1, 10–12
 physiological modulators, 26–31
 definition, 2
 insertion-deletion, 2
 in mammals, 3
 regulators, 31–32
 substitutional, 3–5
 U to C, 4–5
 Wilms' tumor susceptibility gene (WT1), 4–5
 RNA polymerase, 2
 Rodgers (Rg) blood-group antigens, 237–238
 Rolling-circle replication of plasmids of pT181
 see pT181 family *RP-C4-CYP21-TNX* (RCCX) modules, 217–292
 RRM, 24, 28
 RRM2, 25
 RRM3, 25

S

S. cerevisiae, 83
Salmonella typhimurium, 82, 294, 297, 309, 311
SEC14 gene, 92
 Sequence alignment of replication origins of pT181, pC221, and pS194 plasmids, 117
 Sequence-specific DNA-binding activity of initiator proteins, 123–124
 Sequence-specific DNA-binding domain, 123
 Sex hormone-binding globulin (SHBG), 177
 Short-chain dehydrogenase/reductase (SDR), 190
 Single-hairpin (antiparallel) duplex structure, 335–338
 Single-nucleotide polymorphisms (SNPs), 54–55, 60, 65, 245–246
 Single-strand DNA-binding protein (SSB), 115
 Single-stranded α -DNA, 326
 Solitary LTRs (solo-LTR) of HERV-K(C4), 268–269

Spectroscopic analysis, 294
Staphylococcus aureus, 129, 131, 300
 Steroid 21-hydroxylase deficiency,
 recombinations leading to, 229–230
 Sterol response element binding protein
 (SREBP), 96–97
Streptococcus pyogenes, 131
Streptomyces lividans, 303
 Structural variation created by difference in
 glycosylation of C4, 234–235
 Substitutional RNA editing, 3–5
 Superfamily I (SFI), 129

T

TEP1-TEP4, 270
 TFII-DAB complex, 176
 THF/THE ratio, 199
TNXA gene, 273
TNXA sequences, 258
 Topoisomerase, type II, 151
trans-acting factors, C to U ApoB RNA editing,
 17–32
 Transcription elongation complexes (TECs), 2
 TRAPS-like syndromes, 54–55
 Tristetraprolin (TTP) family of CCHH tandem
 zinc finger proteins, 43–68
 discussion of results, 63–66
 DNA sequencing, 47
 EST database searches, 48
 nomenclature for human gene coding, 45
 overview, 44–45
 PCR-amplification conditions, 46–47
 results of study, 52–63
 sequence analysis, 47
 sequencing strategy, 46
 subjects and DNA, 45–46
 Trk system, 296–298
 kinetic analyses, 305
 Troponin T pre-mRNA, 30
 TTP deficiency
 in man, 44
 in mice, 44
 Tumor necrosis factor α (TNF α), 43–44
 Turgor, 294–295
 Two-hairpin tetraplex

 formed by d(G-A)₁₀, 341
 structure, 338–342
 Type II topoisomerase, 151
 Tyrosine sulfation at carboxyl end of α -chain of
 C4, 232–234

U

UL30, 151, 153, 156
 UL42, 151, 153
 UL9, 153, 158, 160
 interacting proteins, 147–149
 model for initiation, 150–151
 origin-binding protein, 145–147
 Ultraviolet hypochromism, 327
 Ultraviolet resonance Raman (UVR) spectroscopy, 324, 326, 338–339

V

Varicella-Zoster Virus (VZV), 140
Vibrio alginolyticus, 298–299
 Wilms' tumor susceptibility gene (WT1), RNA
 editing, 4–5

X

Xenopus, 60

Y

Yeast, phosphatidylserine synthase in, 81–82
 Yeast cells
 interorganelle transport of PE in, 91–92
 interorganelle transport of PS in, 91–92

Z

ZFP36, 45, 52, 65–66
 polymorphisms, 49, 52–56
ZFP36L1, 44, 47, 52
 nonsplicing allele, 59
 polymorphisms, 50, 56–61, 63–64
ZFP36L2, 45, 48, 52
 polymorphisms, 51, 61–63