

PROGRESS IN
NUCLEIC ACID RESEARCH
AND MOLECULAR BIOLOGY

VOLUME 59

EDITED BY
KIVIE MOLDAVE



ACADEMIC PRESS

PROGRESS IN

Nucleic Acid Research
and Molecular Biology

Volume 59

This Page Intentionally Left Blank

PROGRESS IN

Nucleic Acid Research and Molecular Biology

edited by

KIVIE MOLDAVE

*Department of Molecular Biology and Biochemistry
University of California, Irvine
Irvine, California*

Volume 59



ACADEMIC PRESS

San Diego London Boston New York
Sydney Tokyo Toronto

This book is printed on acid-free paper. ∞

Copyright © 1998 by ACADEMIC PRESS

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-1998 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.
0079-6603/98 \$25.00

Academic Press

a division of Harcourt Brace & Company

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

<http://www.apnet.com>

Academic Press Limited

24-28 Oval Road, London NW1 7DX, UK

<http://www.hbuk.co.uk/ap/>

International Standard Book Number: 0-12-540059-4

PRINTED IN THE UNITED STATES OF AMERICA

97 98 99 00 01 02 QW 9 8 7 6 5 4 3 2 1

Contents

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

| | |
|--|---|
| Rhodopsin: A Prototypical G Protein-Coupled Receptor | 1 |
|--|---|

Thomas P. Sakmar

| | |
|---|----|
| I. Structure and Function of Rhodopsin: A Prototypical G Protein-Coupled Receptor | 2 |
| II. Spectral Tuning and the Mechanism of the Opsin Shift | 10 |
| III. Light-Induced Conformational Changes in Rhodopsin | 13 |
| IV. Molecular Switches and Determinants of the Active Receptor Conformation | 20 |
| V. Coupling of Light-Induced Conformational Changes to Transducin Activation | 22 |
| VI. Structural Modeling of Rhodopsin | 25 |
| VII. Rhodopsin Mutations as a Cause of Human Disease | 28 |
| VIII. Conclusions | 29 |
| References | 30 |

| | |
|--|----|
| Cell Membrane and Chromosome Replication in <i>Bacillus subtilis</i> | 35 |
|--|----|

Noboru Sueoka

| | |
|--|----|
| I. Introduction | 36 |
| II. Early Evidence of Membrane—Chromosome Association | 36 |
| III. Chromosome Initiation Mutants of <i>Bacillus subtilis</i> | 39 |
| IV. Preparation of Origin—Membrane and Terminus—Membrane Complexes | 39 |
| V. The <i>dnaB</i> Gene: Critical for Chromosome Initiation and Replication Origin Membrane Attachment | 40 |
| VI. Chromosomal Membrane Attachment Sites | 44 |
| VII. <i>In Vitro</i> Initiation of Chromosome Replication Using the Membrane Fraction | 46 |
| VIII. Membrane Attachment to the Terminus | 47 |
| IX. Differences in Replication Initiation in Two Systems | 47 |
| X. Unsolved Questions | 47 |
| References | 51 |

| | |
|---|-----|
| Stability and Structure of Model DNA Triplexes and Quadruplexes and Their Interactions with Small Ligands | 55 |
| Richard H. Shafer | |
| I. Triple-Helical Structures | 57 |
| II. Guanine Quadruplex Structures | 79 |
| III. Summary | 91 |
| References | 91 |
| | |
| On the Physiological Role of Casein Kinase II in <i>Saccharomyces cerevisiae</i> | 95 |
| Claiborne V. C. Glover III | |
| I. General Properties of CKII | 96 |
| II. <i>Saccharomyces cerevisiae</i> CKII | 100 |
| III. Potential Functions of CKII in <i>Saccharomyces cerevisiae</i> | 109 |
| IV. Substrates of CKII in <i>Saccharomyces cerevisiae</i> | 119 |
| V. The Physiological Role of CKII | 127 |
| References | 129 |
| | |
| The Heparan Sulfate—Fibroblast Growth Factor Family: Diversity of Structure and Function | 135 |
| Wallace L. McKeehan, Fen Wang, and Mikio Kan | |
| I. Diversity and Ubiquity of the Fibroblast Growth Factor Family | 136 |
| II. Diversity of Structure and Function | 142 |
| III. Structure, Assembly, and Control of the FGF Receptor Complex | 155 |
| IV. The FGF Family in Liver Growth and Function | 164 |
| V. The FGF Family in Prostate and Prostate Tumors | 168 |
| References | 173 |
| | |
| The Ribosomal Elongation Cycle and the Movement of tRNAs across the Ribosome | 177 |
| Knud H. Nierhaus, Heinrich B. Stuhmann, and Dmitri Svergun | |
| I. Introduction | 178 |
| II. Functional Aspects: Models of the Elongation Cycle | 180 |

| | |
|---|------------|
| III. Structural Aspects: The Shape of Ribosomes and the Localization of tRNAs | 188 |
| IV. Conclusions | 201 |
| References | 202 |
| | |
| Life on the Salvage Path: The Deoxynucleoside Kinases of <i>Lactobacillus acidophilus</i> R-26 | 205 |
| David H. Ives and Seiichiro Ikeda | |
| I. Historical Background—Nucleotide Metabolism in Lactobacilli | 207 |
| II. Purification of Deoxynucleoside Kinases from <i>Lactobacillus acidophilus</i> R-26 | 212 |
| III. Steady-State Kinetics | 224 |
| IV. Assignment of Subunit Functions | 230 |
| V. Cloning the Genes for dAK/dCK or dAK/dGK | 232 |
| VI. dCK and dGK Are Products of the Same Gene | 238 |
| VII. Probing the Active Site and Subunit Contacts | 243 |
| VIII. Summary | 250 |
| References | 252 |
| | |
| Molecular Analyses of Metallothionein Gene Regulation | 257 |
| Susan L.-A. Samson and Lashitew Gedamu | |
| I. Overview of Metallothioneins | 258 |
| II. Metallothionein Gene Regulation | 259 |
| III. Metallothionein Promotor Organization and Function | 261 |
| IV. MRE-Binding trans-Acting Factors | 274 |
| V. Conclusions and Suggestions for Further Research | 285 |
| References | 285 |
| | |
| Transcriptional Regulation of the Steroid Receptor Genes | 289 |
| M. Vijay Kumar and Donald J. Tindall | |
| I. Structure of a Steroid Receptor Gene | 290 |
| II. Molecular Mechanism of Transcription | 291 |
| III. Regulation of the Androgen Receptor Gene | 293 |
| IV. Regulation of the Glucocorticoid Receptor Gene | 298 |
| V. Regulation of the Progesterone Receptor Gene | 300 |

| | |
|--|-----|
| VI. Regulation of the Estrogen Receptor Gene: Characterization of the 5' Flanking Region | 301 |
| VII. Concluding Remarks | 303 |
| References | 304 |
| | |
| Molecular Evolution of Snake Toxins: Is the Functional Diversity of Snake Toxins Associated with a Mechanism of Accelerated Evolution? | 307 |
| M. Ohno, R. Ménez, T. Ogawa, J. M. Danse, Y. Shimohigashi, C. Fromen, F. Ducancel, S. Zinn-Justin, M. H. Le Du, J.-C. Boulain, T. Tamiya, and A. Ménez | |
| I. About Snake Toxins | 309 |
| II. Snake Toxins with a Phospholipase A ₂ -Type Fold | 311 |
| III. Snake Toxins with a Three-Fingered Fold | 339 |
| IV. General Conclusion on the Evolution of Snake Toxins | 356 |
| References | 357 |
| | |
| INDEX | 365 |

Some Articles Planned for Future Volumes

Structure and Transcription Regulation of Nuclear Genes for the Mouse
Mitochondrial Cytochrome c Oxidase

NARAYAN G. AVADHANI, ARUNA BASU, CARMEN SUCHAROV,
AND NIBEDITA LENKA

Tissue Transglutaminase—Retinoid Regulation and Gene Expression

PETER J. A. DAVIES AND SHAKID MIAN

Genetic Approaches to Structural Analysis of Membrane Transport Systems

WOLFGANG EPSTEIN

Intron-Encoded snRNAs

MAURILLE J. FOURNIER AND E. STUART MAXWELL

Mechanisms for the Selectivity of the Cell's Proteolytic Machinery

ALFRED GOLDBERG, MICHAEL SHERMAN, AND OLIVER COUX

Mechanisms of RNA Editing

STEPHEN L. HAJDUK AND SUSAN MADISON-ANTENUCCI

The Nature of DNA Replication Origins in Higher Eukaryotic Organisms

JOEL A. HUBERMAN AND WILLIAM C. BURTIANS

A Kaleidoscopic View of the Transcriptional Machinery in the Nucleolus

SAMSON T. JACOB

Function and Regulatory Properties of the MEK Kinase Family

GARY L. JOHNSON, P. GERWINS, C. A. LANGE-CARTER, A. GARDNER,
M. RUSSELL, AND R. R. VILLIANCOURT

Sphingomyelinases in Cytokine Signaling

MARTIN KRONKE

Mammalian DNA Polymerase Delta: Structure and Function

MARIETTA Y. W. T. LEE

DNA Helicases: Roles in DNA Metabolism

STEVEN W. MATSON AND DANIEL W. BEAM

Inosine Monophosphate Dehydrogenase: Role in Cell Division and Differentiation

BEVERLY S. MITCHELL

Specificity of Eukaryotic Type II Topoisomerase: Influence of Drugs, DNA Structure, and Local Sequence

MARK T. MULLER AND JEFFREY SPITZNER

Immunoanalysis of DNA Damage and Repair Using Monoclonal Antibodies

MANFRED F. RAJEWSKY

Positive and Negative Transcriptional Regulation by the Retinoblastoma Tumor Suppressor Protein

PAUL D. ROBBINS AND JOHN HOROWITZ

Organization and Expression of the Chicken α -Globin Genes

KLAUS SCHERRER AND FELIX R. TARGA

Mechanism of Regulatory GTPase in Protein Biosynthesis

MATHIAS SPRINZL AND ROLF HILGENFELD

Molecular Genetic Approaches to Understanding Drug Resistance in Protozoan Parasites

DYANN WIRTH, SARAH VOLKMAN, AND LARRY CHOW

Rhodopsin: A Prototypical G Protein-Coupled Receptor

THOMAS P. SAKMAR

*The Howard Hughes Medical Institute
Laboratory of Molecular Biology and
Biochemistry
Rockefeller University
New York, New York 10021*

| | |
|---|----|
| I. Structure and Function of Rhodopsin: A Prototypical G Protein-Coupled Receptor | 2 |
| A. The Extracellular Domain of Rhodopsin | 6 |
| B. The Membrane-Embedded Domain and the Chromophore-Binding Site | 7 |
| C. The Cytoplasmic Domain of Rhodopsin | 9 |
| II. Spectral Tuning and the Mechanism of the Opsin Shift | 10 |
| A. Spectral Properties of Recombinant Rhodopsins | 10 |
| B. Spectral Properties of Recombinant Cone Pigments | 11 |
| III. Light-Induced Conformational Changes in Rhodopsin | 13 |
| A. Molecular Changes in the Membrane-Embedded Domain of Rhodopsin | 13 |
| B. Functional Interactions between Transmembrane Helices 3 and 6 | 15 |
| C. Light-Induced Molecular Changes on the Cytoplasmic Surface | 18 |
| IV. Molecular Switches and Determinants of the Active Receptor Conformation | 20 |
| V. Coupling of Light-Induced Conformational Changes to Transducin Activation | 22 |
| VI. Structural Modeling of Rhodopsin | 25 |
| VII. Rhodopsin Mutations as a Cause of Human Disease | 28 |
| VIII. Conclusions | 29 |
| References | 30 |

A variety of spectroscopic and biochemical studies of recombinant site-directed mutants of rhodopsin and related visual pigments have been reported over the past 9 years. These studies have elucidated key structural elements common to visual pigments. In addition, systematic analysis of the chromophore-binding pocket in rhodopsin and cone pigments has led to an improved understanding of the mechanism of the opsin shift, and of particular molecular determinants underlying color vision in humans. Identification of the conformational changes that occur on rhodopsin photoactivation has been of particular recent concern. Assignments of light-dependent molecular alterations to specific regions of the chromophore have also been attempted by studying native opsins regenerated with

synthetic retinal analogs. Site-directed mutagenesis of rhodopsin has also provided useful information about the retinal-binding pocket and the molecular mechanism of rhodopsin photoactivation. Individual molecular groups have been identified to undergo structural alterations or environmental changes during photoactivation. Analysis of particular mutant pigments in which specific groups are locked into their respective “off” or “on” states has provided a framework to identify determinants of the active conformation, as well as the minimal number of intramolecular transitions required to switch between inactive and active conformations. A simple model for the active state of rhodopsin can be compared to structural models of its ground state to localize chromophore–protein interactions that may be important in the photoactivation mechanism. This review focuses on the recent functional characterization of site-directed mutants of bovine rhodopsin and some cone pigments. In addition, an attempt is made to reconcile previous key findings and existing structural models with information gained from the analysis of site-directed mutant pigments. © 1998 Academic Press

I. Structure and Function of Rhodopsin: A Prototypical G Protein-Coupled Receptor

Visual pigments comprise a branch of a large family of G protein¹-coupled receptors (I). Although they share many similarities with other G protein-coupled receptor types, there is significant specialization in visual pigments not found in other receptor families. In particular, pigments are made up of opsin apoprotein plus chromophore. The chromophore is a cofactor and not a ligand in the classical sense because it is linked covalently via a protonated Schiff base bond to a specific lysine residue in the membrane-embedded domain of the protein (Fig. 1). One of two retinoids serve as the chromophore for nearly all visual pigments. The chromophore in most vertebrate pigments is the aldehyde of vitamin A, 11-*cis*-retinal (Fig. 2). The chromophore in many invertebrate, fish, and amphibian pigments is the aldehyde of vitamin A₂, 11-*cis*-3-dehydroretinal, which contains an additional carbon–carbon double bond in the β -ionone ring. An important structural feature of the retinal chromophore in rhodopsin, in addition to its Schiff base linkage, is its extended polyene structure, which accounts for its visible absorption properties and allows for resonance structures (2).

Rhodopsin has a broad visible absorption maximum (λ_{\max}) at about 500 nm. On photoisomerization of the chromophore, the pigment is converted to metarhodopsin II (MII) with a λ_{\max} value of 380 nm (Fig. 2). The MII intermediate is characterized by a deprotonated Schiff base chromophore linkage. MII is the active form of the receptor (R*, light-activated rhodopsin), which catalyzes guanine nucleotide exchange by transducin.

¹ G protein is a heterotrimeric guanine nucleotide-binding regulatory protein.

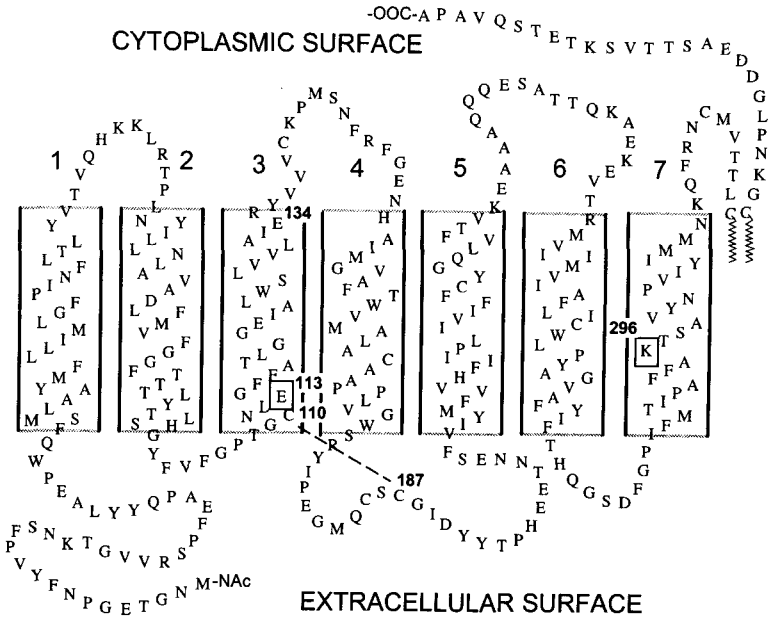


FIG. 1. Schematic representation of bovine rhodopsin secondary structure based on previous models. Seven putative TM helices, which are characteristic of G protein-coupled receptors, are shown. The carboxyl terminus and cytoplasmic surface are toward the top, and the amino terminus and extracellular (intradiskal) surface are toward the bottom of the figure. Rhodopsin resides in the specialized disk membrane of the outer segment of the rod cell of the retina and is responsible for dim-light (scotopic) vision. Lys²⁹⁶ is the site of the retinylidene Schiff base linkage. Glu¹¹³ is the negatively charged counterion to the positively charged protonated Schiff base. An essential disulfide bond on the intradiskal surface links Cys¹¹⁰ and Cys¹⁸⁷. Cys³²² and Cys³²³ in the carboxyl-terminal tail are palmitoylated, which results in a fourth cytoplasm loop.

Photoisomerization of the 11-*cis* to all-*trans* form of the retinylidene chromophore is the primary event in visual signal transduction, and it is the only light-dependent step (Fig. 2). Chromophore isomerization activates the pigment, allowing it to interact with a specific heterotrimeric G protein, transducin. In the case of the vertebrate visual system, transducin activation leads to the activation of a cyclic-GMP phosphodiesterase, and the closing of cyclic-GMP-gated cation channels in the plasma membrane of the rod cell. Light causes a graded hyperpolarization of the photoreceptor cell. The amplification, modulation, and regulation of the light response is of great physiological importance and has been discussed in detail elsewhere (3-5). However, it should be pointed out that despite the fact that the visual system functions over about a 10^6 -fold range of light intensity, the retinal rod cell

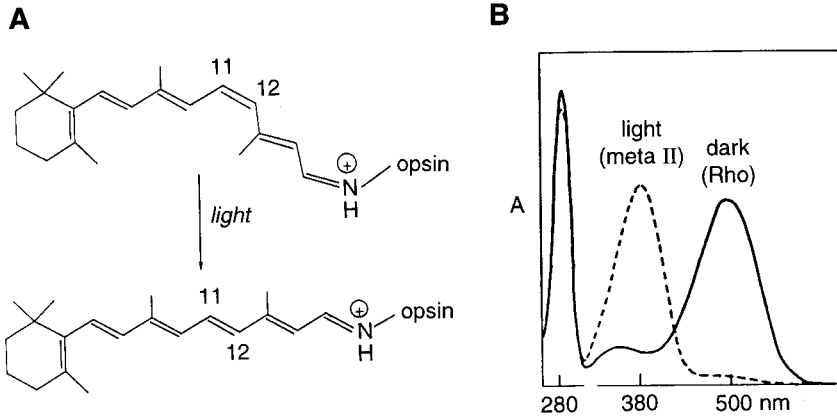


FIG. 2. (A) Photoisomerization of the 11-*cis*-retinylidene chromophore to the all-*trans* form is the only light-dependent event in vision. The 11-*cis*-retinal chromophore is covalently linked as a cofactor to Lys²⁹⁶ on TM helix 7 via a protonated Schiff base bond. Important photochemical properties of rhodopsin in the disk membrane include a very high quantum efficiency (~ 0.67 versus < 0.2 for 11-*cis*-retinal in solution) and an extremely low rate of thermal isomerization of the chromophore ($\sim 2.25 \times 10^{10}$ slower than in solution). (B) The UV-visible absorption spectrum of purified recombinant COS cell rhodopsin in detergent solution shows a characteristic broad visible absorbance with a λ_{\max} value of 500 nm. The 280-nm peak represents the protein component. After exposure to light, the pigment is converted to a species with a λ_{\max} value of 380 nm, characteristic of metarhodopsin II. This is the active form of the receptor that interacts with the G protein, transducin. Identical results can be obtained with rhodopsin from bovine retinas purified by concanavalin-A lectin affinity chromatography.

has single-photon detection capability due to extremely low levels of dark noise in rhodopsin and a significant degree of biochemical amplification. Thermal isomerization in a single rhodopsin molecule at physiologic temperature has been estimated to occur about once in 470 years (6). Activation of a single rhodopsin molecule by a single photon can prevent the entry of as many as 10^8 cations into the rod cell. The possibility of single-pheromone molecule detection by insect olfactory systems notwithstanding, the visual system is unique among sensory signal transduction systems in that it can detect single events.

The visual pigments of many species of vertebrates and invertebrates have been studied by absorption spectroscopy or microspectrophotometry of visual organs. Therefore, historically vertebrate visual pigments have been generally classified based on the photoreceptor cell type of the retina in which they are found (7). Rod cells, responsible for dim-light vision, contain rhodopsin ("red" opsin). Cone cells, responsible for bright-light and color vision, contain iodopsins ("violet" opsins), also known as cone pigments or col-

or vision pigments. The cloning of opsins from a variety of species has allowed more detailed comparisons and phylogenetic classifications based on structural, spectral, and biochemical properties of visual pigments (8). The homology in the opsin family of genes indicates that divergent evolution occurred from a single precursor retinal-binding protein to form long- and short-wavelength-absorbing prototypes. The long-wavelength prototype diverged to form red and green pigments. The short-wavelength prototype then diverged to form a blue pigment and the family of rhodopsins and rhodopsin-like green pigments.

Each visual pigment is tuned to a particular wavelength of maximal absorption, λ_{max} , although the visible absorbance peak tends to be quite broad. Even though a protonated retinylidene Schiff base chromophore is likely to be universal in visual pigments, the λ_{max} values of visual pigments span the visible spectrum (i.e., from near-ultraviolet at about 400 nm to far visible red at about 600 nm). Distinct chromophore-protein interactions are responsible, directly or indirectly, for spectral tuning in visual pigments. Thus, differences in primary structure result in differences in spectral properties.

Bovine rhodopsin is the most extensively studied G protein-coupled receptor. A large amount of pigment (0.5–1.0 mg) can be obtained from a single bovine retina by sucrose density gradient centrifugation preparation of the rod outer segment disk membranes. The pigment can be further purified by lectin affinity chromatography on concanavalin-A Sepharose resin. Rhodopsin is stable in solubilized form in a variety of detergents, including digitonin, dodecyl maltoside, and octyl glucoside. Bovine rhodopsin was the first G protein-coupled receptor to be sequenced by amino acid sequencing (9, 10) and the first to be cloned (11, 12). The cloning of the β_2 -adrenergic receptor (13) led to the identification of the structural homologies that now define the large family of G protein-coupled receptors (14, 15).

A number of structural features are shared by visual pigments (Fig. 1). Like all G protein-coupled receptors, they consist of seven putative transmembrane (TM) helices. A lysine residue that acts as the linkage site for the chromophore is conserved within the seventh TM segment in all pigments. In all vertebrate pigments, a carboxylic acid residue that acts as the counterion to the protonated, positively charged Schiff base is conserved within the third TM segment. The position of the Schiff base counterion is one helix turn away from the position of an aspartic acid residue conserved in biogenic amine receptors that serves as the counterion to the cationic amine ligands. A pair of highly conserved cysteine residues is found on the extracellular surface of the receptor and forms a disulfide bond. A Glu or Asp/Arg/Tyr or Phe tripeptide sequence is found at the cytoplasmic border of the third TM helix. This sequence is conserved in most G protein-coupled receptors and has been shown to be involved in G protein interaction (16, 17). Sites of light-

dependent phosphorylation (serine and threonine residues) are found at the carboxyl-terminal tail of most visual pigments. These sites are analogous to phosphorylation sites found on the carboxyl-terminal tails of other G protein-coupled receptors (18).

The topology of rhodopsin with respect to the disk membrane bilayer is defined by three structural domains: the extracellular (intradiskal) domain, the membrane-embedded domain, and the intracellular (cytoplasmic) domain.

A. The Extracellular Domain of Rhodopsin

The role of the extracellular domain in rhodopsin structure and function has been elucidated by a number of studies involving site-directed mutagenesis. The extracellular loops and amino-terminal tail of bovine rhodopsin have been shown in a deletion analysis to be important for proper folding of the receptor that allows cellular processing and chromophore binding (19). Insertional mutagenesis was also used in a related study to probe the topology of rhodopsin and to correlate the location of epitope insertion to stability and cell trafficking (20). Interestingly, several mutations that interfered with the formation of a correct tertiary structure on the intradiskal surface resulted in mutant opsins that appeared to be retained in the endoplasmic reticulum during heterologous expression, and were complexed with chaperonins (21).

The two conserved cysteine residues on the extracellular domains, Cys¹¹⁰ and Cys¹⁸⁷, are essential for proper folding of opsin expressed in COS cells (22). These two residues were shown to form a disulfide linkage in an elegant study in which the four cytoplasmic and three membrane-embedded cysteine residues were removed by site-directed mutagenesis to create a mutant receptor with only the three intradiskal cysteines remaining (23). In a related study, the double mutant C110A/C187A was shown to bind 11-*cis*-retinal to form a rhodopsin-like pigment (24). However, the MII-like photoproduct of the mutant pigment, which could activate transducin in response to light, was considerably less stable than native MII (24). In general, mutations on the intradiskal surface that might interfere with the formation of a disulfide bond between residues Cys¹¹⁰ and Cys¹⁸⁷ correlate to a loss of function phenotype. Namely, mutations affecting the ability of these two cysteine residues to juxtapose during translation or membrane translocation affect expression level, transport to the plasma membrane, and the ability of the mutant opsin to regenerate with 11-*cis*-retinal chromophore to form a pigment with normal stability.

Future studies of mutant opsins with defects in folding, membrane insertion, or cell trafficking will be facilitated by the recent report of a methodology to purify regenerated pigment from nonregenerated opsin. The paradigm was developed by again using the opsin mutant containing only the

three intradiskal cysteine residues (22). It was shown that the nonregenerated opsin, which could not bind 11-*cis*-retinal, was misfolded and was likely to have an incorrect disulfide bond pairing (25). This is additional evidence pointing to the role of the intradiskal domain, and in particular the early formation of a correct disulfide bond linkage, in the proper folding of the opsin apoprotein.

Rhodopsin is also known to be glycosylated on two asparagine residues (Asn² and Asn¹⁵) of its amino-terminal tail. A nonglycosylated rhodopsin, which was prepared in the presence of tunicamycin, was defective in light-dependent activation of transducin (26). Site-directed opsin mutants with replacements of these two residues or of neighboring consensus sequence residues were studied as well. It was concluded that glycosylation at Asn¹⁵ was required for full signal transduction activity, but apparently not for correct biosynthesis or folding (26).

B. The Membrane-Embedded Domain and the Chromophore-Binding Site

The rhodopsin chromophore lies within the seven-TM helical bundle of the membrane-embedded domain of the receptor. The exact position of the retinal has not been determined. However, many amino acid residues that seem to interact specifically with the chromophore have been identified through mutagenesis approaches. These residues might be divided into the following categories: (1) amino acid side chains that stabilize the 11-*cis*-isomeric structure of the ground state and inhibit thermal isomerization, (2) amino acid side chains involved in tuning the pigment to a particular wavelength of maximal absorption, (3) amino acid side chains that influence the efficiency and specificity of the *cis/trans* photoisomerization, and (4) amino acid side chains that convey the signal of retinal isomerization to the cytoplasmic surface of the receptor. These somewhat arbitrary classifications are not mutually exclusive and many amino acid residues may participate in more than one function.

A number of studies have been carried out on bovine rhodopsin to investigate retinal-opsin interactions in the membrane-embedded domain of bovine rhodopsin. Spectroscopic methods such as resonance Raman spectroscopy, Fourier-transform infrared (FTIR) difference spectroscopy, and nuclear magnetic resonance (NMR) spectroscopy have been employed (27, 28). Other approaches have included reconstitution of opsin apoprotein with synthetic retinal analogs (29) and photochemical crosslinking (30, 31). Early work on the structure and function of recombinant bovine rhodopsin focusing on the use of techniques of molecular biology has been reviewed (32, 33). The chromophore-opsin Schiff base linkage is discussed in this section, and recent work related specifically to protein-chromophore interactions within

the membrane-embedded domain, including molecular modeling, is the subject of later sections of this review.

Lys²⁹⁶ and Glu¹¹³ are two of the key amino acid residues that define the structure and function of the retinal chromophore in rhodopsin. The Schiff base linkage of the chromophore to Lys²⁹⁶ is a key feature of rhodopsin structure (34). Light-dependent Schiff base deprotonation is required for the formation of the active state of the receptor (35, 36). However, it has been shown that light can induce the receptor active state in the absence of a Schiff base chromophore linkage to the opsin (37). A model retinal Schiff base compound was prepared from 11-*cis*-retinal and *n*-butylamine and then regenerated with a recombinant mutant opsin K296G. A visible pigment was formed although no linkage was present between the model compound and the opsin. Photoisomerization of the model chromophore led to receptor activation as judged by transducin activation. The authors point out the similarities between the mutant receptor and its model chromophore ligand and other ligand-dependent G protein-coupled receptors, such as biogenic amine receptors. This theme was expanded in a concise review in which the retinal-binding domain of rhodopsin was compared and contrasted with the ligand-binding domains of other G protein-coupled receptor types (38).

Glu¹¹³ in bovine rhodopsin serves as the counterion to the positive charge of the retinylidene-protonated Schiff base (39–41). Glu¹¹³ is unprotonated and negatively charged in the ground state of rhodopsin (42). It becomes protonated on light-dependent formation of MII and is the net proton acceptor for the Schiff base proton (43). Glu¹¹³ is located near C-12 of the retinal polyene according to the two-photon spectroscopy and NMR spectroscopy of retinal analogs and semiempirical quantum mechanical orbital calculations (27, 44–46). This location is supported by resonance Raman (47) and FTIR difference spectroscopy studies (48) of recombinant mutant pigments lacking a carboxylic acid group at position 113. The Glu¹¹³-protonated Schiff base interaction serves to stabilize the Schiff base proton such that its acidity constant (pK_a) in rhodopsin is estimated to be >16 , compared to a value of ~ 7 in aqueous solution.

The stable interaction between Glu¹¹³ and the protonated Schiff base may also inhibit hydrolysis of the Schiff base linkage in darkness. For example, hydroxylamine does not react with the Schiff base of rhodopsin, but readily reacts with that of MII or with rhodopsin mutants in which Glu¹¹³ is replaced by a neutral amino acid residue by mutagenesis (39). This is an important consideration because the opsin alone, without the 11-*cis*-retinal chromophore, has a small but measurable activity (49–52). The high sensitivity of the rod cell depends on an extremely low intrinsic level of signaling in darkness. Dark noise can be generated by thermal isomerization events in rhodopsin, by the presence of opsin lacking the 11-*cis*-retinal chromophore,

which acts as an inhibitor of opsin activity, or by mutant opsins that display the property of constitutive activity, as discussed below.

The nature of the Glu¹¹³-protonated Schiff base interaction in the structure and function of rhodopsin has been elucidated indirectly by the discovery of constitutive activity among certain opsin mutants (51). Constitutive activity refers to the ability of an opsin to activate transducin in the absence of any chromophore. Generally, a mutation that disrupts a putative salt bridge between Glu¹¹³ and Lys²⁹⁶ in the opsin apoprotein leads to constitutive activity. Replacement of either Glu¹¹³ or Lys²⁹⁶ by a neutral amino acid results in a mutant opsin with constitutive activity. Other mutations such as G90D or A292E also result in constitutive activity, presumably because the introduction of the negatively charged residue into the membrane-embedded domain of the receptor affects the stability of the Glu¹¹³-Lys²⁹⁶ salt bridge (52, 53). The mechanism of constitutive activity of opsins and the potential relevance of constitutive activity to visual diseases such as congenital night blindness have been reviewed (54).

C. The Cytoplasmic Domain of Rhodopsin

The cytoplasmic domain of rhodopsin (Fig. 1) comprises three loops linking TM helices (loop 1–2, loop 3–4, loop 5–6) and the carboxyl-terminal tail. The carboxyl-terminal tail contains a fourth putative loop structure because palmitoylated cysteine residues of positions 322 and 323 between the cytoplasmic end of TM helix 7 and the carboxyl terminus form a membrane anchor. A number of cytoplasmic proteins are known to interact with R*. These include transducin and rhodopsin kinase. Rhodopsin kinase phosphorylates R* at specific serine residues (Ser³³⁴, Ser³³⁸, and Ser³⁴³) (55–57). Arrestin binds to phosphorylated R* to prevent it from activating transducin. The role of the distal carboxyl-terminal tail in the shut-off of the light signal in the rod cell was shown convincingly in a study using a transgenic mouse strain. The mice, whose rod cells expressed a rhodopsin transgene with a truncated carboxyl-terminal tail, were studied by electrophysiological techniques. Phosphorylation of the carboxyl-terminal tail was required for termination of light-induced signaling (58).

The biochemical mechanisms of background and light adaptation are known to involve light-dependent changes in intracellular concentrations Ca²⁺ as well as in cGMP. Calcium levels influence the rate of accumulation of R* (59). However, the precise effects of changes of calcium and cGMP concentrations on the rate of recovery during adaptation remain to be determined (60). The details of the regulation of the biochemical activities of rhodopsin kinase, phosducin, recoverin, transducin, and cGMP phosphodiesterase and their exact roles in signal termination and signaling system recovery and adaptation also remain to be elucidated in detail.

Perhaps the most extensively studied receptor–G protein interaction is that of bovine rhodopsin with transducin (61). Detailed biochemical and biophysical analysis of the R*–transducin interaction has been aided by mutagenesis of the cytoplasmic domain of bovine rhodopsin. Numerous rhodopsin mutants defective in the ability to activate transducin have been identified (17). Several of these mutant receptors were studied by flash photolysis (16), light scattering (62), or proton uptake assays (63). The salient result of these studies was that cytoplasmic loops 3–4 and 5–6 were involved in R*–transducin interaction. This finding was consistent with other approaches, including peptide competition studies (64). A class of interesting receptor mutants was also identified in which spectrally normal MII-like pigments were formed on illumination. The photolyzed mutants bound transducin in the presence of GDP, but failed to release the bound transducin in the presence of GTP (16, 62). These results were consistent with the idea that transducin binding and transducin activation were discrete steps mechanistically, which could be uncoupled by specific amino acid substitutions on the cytoplasmic surface of the receptor.

The role of the carboxyl-terminal tail of rhodopsin in the activation of transducin is less well defined. Although several biochemical studies have implicated the carboxyl-terminal tail in transducin activation (65, 66), alanine-scanning mutagenesis failed to confirm that the proximal portion of the tail was required (67, 68). Additional work is needed to elucidate any role, subtle or otherwise, of the carboxyl-terminal tail of rhodopsin in transducin activation. In addition, an important consideration will be to map the sites of specific transducin subunit interactions on the cytoplasmic surface of rhodopsin. This work may be facilitated by both the high-resolution crystal structure of a form of heterotrimeric transducin (69), and by the ability to produce, using a baculovirus expression system, purified recombinant heterotrimeric transducin and to reconstitute it with recombinant rhodopsin (70).

II. Spectral Tuning and the Mechanism of the Opsin Shift

A. Spectral Properties of Recombinant Rhodopsins

The molecular interactions between the opsin protein and the retinal chromophore define the ground-state spectral properties of a particular visual pigment. Rhodopsin has been used as a model pigment for a variety of chemical and spectroscopic studies to elucidate the mechanism of the opsin shift. Relevant to this question, one important result to arise from the study

of mutant bovine rhodopsin pigments was the identification of Glu¹¹³ as the retinylidene Schiff base counterion (39–41). This result, combined with the characterization of other membrane-embedded carboxylic acid residues by both UV–visible spectroscopy (39–41, 71) and microprobe resonance Raman spectroscopy (47), suggested that the retinal-binding pocket in bovine rhodopsin was electrostatically neutral (27). Additional studies, including the use of photoaffinity reagents (30, 31), retinal analogs regenerated with site-directed mutants (72), or site-directed mutant pigments (71, 73–75), have led to a more complete picture of the amino acid residues in the membrane-embedded domain of rhodopsin that interact with the retinal chromophore. Detailed analyses of particular mutant pigments in combination with molecular modeling algorithms have provided a basis for satisfactory models of the retinal-binding pocket. These models may focus attention on chromophore–opsin interactions relevant to understanding the opsin-shift mechanism.

B. Spectral Properties of Recombinant Cone Pigments

Human trichromatic color vision, at the level of the photoreceptor cell, requires the presence of three cone pigments with broad overlapping spectral absorption. Three genomic and cDNA clones encoding the opsin apoproteins of these pigments were cloned and characterized (76). The amino acid sequences of these opsins are about 41% identical to that of human rhodopsin. The green and red opsins are about 96% identical to each other and about 43% identical to the blue opsin. Previously, the spectral properties of human cone pigments had been studied by a variety of techniques ranging from psychophysical color matching to microspectrophotometry (7). Recently, however, the human cone pigment genes were expressed in tissue culture cells, reconstituted with 11-*cis*-retinal, and studied by UV–visible spectroscopy (77, 78). The λ_{max} values reported in the two studies were as follows: blue, 426 nm; green, 530 nm; red, 552 and 557 nm for polymorphic variants (78); and blue, 424 nm, green, 530 nm; and red, 560 nm (77). These studies confirmed the assignments based on genetic analysis of the cloned pigment genes.

Analysis of the arrangements of the cone opsin genes on the X chromosome has led to a detailed understanding of the molecular genetics of inherited variations in color vision (79). In males with normal color vision it was proposed that a single red opsin gene resides with one or more green opsin genes in a head-to-tail tandem array. Thus, in one type of color vision defect, anomalous trichromacy, unequal intragenic recombination can theoretically result in an opsin gene that is a hybrid between green and red opsin genes. It was proposed that these hybrids would have anomalous spectral properties (79, 80). This hypothesis was confirmed experimentally at the level of the

photoreceptor pigment by obtaining absorption spectra for heterologously expressed hybrid pigments responsible for anomalous trichromacy (81).

Quantitative polymerase chain reaction (PCR) methods have been applied to evaluate in more detail the numbers and ratios of middle- and long-wavelength-absorbing opsin genes in males with normal color vision. It was found that many subjects had more numerous opsin genes than previously suggested, and that in many cases more than one long-wavelength opsin gene was present on the X chromosome (82, 83). The molecular genetics of blue cone monochromacy has also been investigated (84). A genetic model to account for the absence of the green and red genes has been tested in transgenic mice (85). The results suggest that a conserved 5' region interacts with the green or red gene promoter to determine which gene is expressed in a given cone cell.

Sequence information has been obtained from PCR to allow predictions about the identity of specific amino acids involved in human red-green color vision (86, 87). New World monkeys, such as marmosets (*Callithrix jacchus*), tamarins (e.g., *Saguinus fuscicollis*), and squirrel monkeys (e.g., *Saimiri sciureus*), have dichromatic vision with a single long-wavelength pigment. However, there is a striking polymorphism in the pigment, such that females with two X-linked opsin genes are effectively trichromatic. Using pairwise comparisons of opsin gene sequences from a number of individual male monkeys with a range of pigment λ_{\max} values, a model was proposed in which three amino acid residues account for the spectral variation between human green and red cone pigments: positions 180, 277, and 285 (86). This model was tested by introducing mutations at these sites in bovine rhodopsin (88) and by expression of mutant and hybrid human pigment genes (81, 89). Thus, of the 15 amino acid differences between human green and red pigments, three hydroxyl-bearing amino acid residues were suggested to be predominantly responsible for the spectral shift: Ser¹⁸⁰, Tyr²⁷⁷, and Thr²⁸⁵ (81, 86, 88). A more complete analysis of the molecular determinants of red/green color vision was carried out using site-directed mutant pigments and chimeric pigments (90, 91). It was found that of the 15 differences between the 364-amino acid human green and red pigments, only seven amino acid changes were responsible for the observed 31-nm red shift in going from the green to the red pigment: Tyr¹¹⁶ to Ser, Ala¹⁸⁰ to Ser, Thr²³⁰ to Ile, Ser²³³ to Ala, Phe²⁷⁷ to Tyr, Ala²⁸⁵ to Thr, and Phe³⁰⁹ to Tyr.

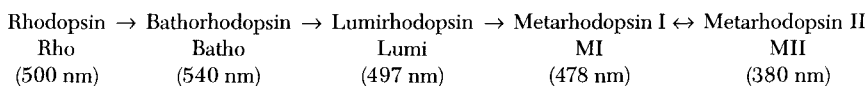
Red and green visual pigments have also been shown to bind chloride ions, resulting in a large red shift in their absorption maxima. An extensive mutagenesis study of 18 different positively charged amino acids identified His¹⁹⁷ and Lys²⁰⁰, in the extracellular loop linking TM helices 4 and 5, as the chloride ion binding site in these pigments (91). These residues are conserved in long-wavelength-absorbing opsins but not in short-wavelength-ab-

sorbing opsins or rhodopsins, suggesting that the ability to bind chloride was a key event in the evolution of color pigment genes.

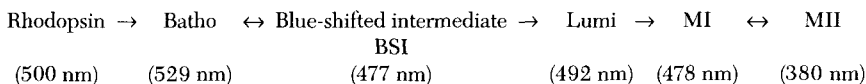
III. Light-Induced Conformational Changes in Rhodopsin

A. Molecular Changes in the Membrane-Embedded Domain of Rhodopsin

Photoisomerization of the 11-*cis*-retinal chromophore, which is covalently linked to Lys²⁹⁶ via a protonated Schiff base, induces the active receptor conformation R* (92–95). R* is spectroscopically indistinct from the rhodopsin photoproduct MII, which is characterized by an unprotonated Schiff base chromophore linkage and a λ_{max} value of 380 nm. At low temperature, a number of photointermediates of rhodopsin can be trapped and characterized by a variety of spectroscopic techniques. The classical photolysis pathway of rhodopsin at low temperature proceeds according to the following scheme:



Nanosecond time-resolved spectroscopy has identified a related, but different, pathway, which occurs at or near physiological temperature (96):



MII decays to metarhodopsin III (465 nm) and finally to opsin plus free all-*trans*-retinal.

Significant progress has been made in understanding structure–function relationships in the photointermediates of rhodopsin through optical, resonance Raman, FTIR, and NMR spectroscopic studies on a variety of native visual pigments and artificial pigments (native opsins reconstituted with synthetic retinal analogs) (45, 96–99). These types of studies provide specific detailed information about chromophore structures and about chromophore–opsin interactions. With recent advances in molecular biological techniques, UV–visible, resonance Raman, FTIR, and transducin activation experiments on recombinant mutant pigments have led to a better understanding of the photoactivation mechanism of rhodopsin (42, 47, 73, 100, 101). A critical evaluation of the experimentally observed molecular determinants of dark- and light-activated rhodopsin in conjunction with refine-

ment of the hypothetical structural model of rhodopsin should help to design additional mutant rhodopsins to address specific questions regarding the activation mechanism. Key elements taken from a more detailed discussion of the photoactivation mechanism in rhodopsin are presented below (100).

An inactive receptor conformation must be capable of changing to an active conformation, which catalyzes nucleotide exchange by a G protein. In rhodopsin, the 11-*cis*-retinal chromophore is in its “off” state, but switches to the “on”-state all-*trans* geometry by photoisomerization. Any amino acid side chain that is involved in such a conformational change must be able to exist in two different steric and/or electrostatic states, which can be designated as “on” or “off” depending on whether the particular state is observed in the active or inactive receptor, respectively.

Because receptor activation involves conformational changes at different topological locations, an individual amino acid or a structural domain in the chromophore-binding pocket may be in its “on” state without necessarily locking the transducin interaction domain into its active conformation, and vice versa. Thus, a concerted transition of individual groups may create the overall active conformation, but it may not be clear whether all of the measurable structural changes are actually essential. A minimal subset of molecular groups may be able to govern the transition to an active receptor conformation irrespective of the binary states (off/on) of all possible groups usually participating in receptor activation. The functional hierarchy of individual group transitions becomes obvious only in recombinant, or in otherwise modified receptors, in which the state of an individual amino acid side chain can be influenced, or even locked, into one of two states. Ideally, it should be possible to alter the receptor activation pathway such that certain transitions become uncoupled from other transitions. It has been argued that the *simplified binary model* of individual molecular transitions provides a framework to analyze functional data obtained from recombinant rhodopsins (100). Experimental results can be related to available structural predictions and may serve as a basis for the design of mutations specifically to address structure–function relations in rhodopsin.

According to this argument, only molecular changes occurring between dark rhodopsin and the R* state are relevant. Thus, FTIR difference spectroscopy has proved to be a well-suited technique for the study of light-induced conformational changes in retinal proteins (99, 102). FTIR difference spectroscopy measures frequency shifts only of those vibrational modes that are affected during photoproduct formation, irrespective of whether they are caused by the protein constituents or by the chromophore. In contrast, resonance Raman spectroscopy measures specifically chromophore vibrations and thereby helps to define the chromophore geometry and the Schiff base protonation state without the potential drawback of overlapping absorptions

of amino acid side chains. Both techniques have contributed significantly to knowledge of molecular changes occurring during MII formation. Application of these techniques to site-directed rhodopsin mutants have allowed the identification of a number of on/off states of particular amino acids during R* formation. Among the putative membrane-embedded carboxylic acid groups, light-induced changes of protonation states or hydrogen-bond strengths were deduced from characteristic frequency shifts of C=O stretching vibrations of protonated carboxylic acid groups in FTIR difference spectra. Their assignment to specific Asp or Glu residues was based on the disappearance of specific difference bands in site-directed mutants and revealed that Asp⁸³ (42, 103) and Glu¹²² (42) are protonated in both dark rhodopsin and MII, whereas Glu¹¹³ is ionized in the dark state and becomes protonated in MII (43).

At least five key individual molecular changes assigned to specific chemical groups in rhodopsin occur on MII formation. These light-induced changes occur in the retinal chromophore, at the Schiff base imine, at the counterion Glu¹¹³, and in the hydrogen-bonding environments of Asp⁸³ and of Glu¹²². For example, Glu¹¹³ is in its “off” state when it is ionized in dark rhodopsin, where it serves as the counterion to the protonated Schiff base (39–41). The “on” state of Glu¹¹³ is created by protonation of its side chain. Based on data derived from mutant pigments, two specific questions can be addressed: (1) which “on” states are crucial to maintain the active receptor conformation, and (2) which of the five known light-induced changes at specific sites in the receptor structure are essential for switching to the active receptor conformation?

B. Functional Interactions between Transmembrane Helices 3 and 6

Movement of α -helical domains is known to be involved in the signal transduction mechanisms of some TM receptor proteins, such as the bacterial chemoreceptors (104), and has been shown to occur during the proton pumping cycle following retinal isomerization in bacteriorhodopsin (BR), the seven-TM light-driven proton pump (105). Recent studies have suggested that steric and/or electrostatic changes in the ligand-binding pocket of rhodopsin may cause changes in the relative disposition of TM helices within the core of the receptor. These changes may be responsible for transmitting a “signal” from the membrane-embedded binding site to the cytoplasmic surface of the receptor. Tryptophan mutagenesis (73), mutagenesis of conserved amino acid residues on TM helices 3 and 6 (74, 75), and the introduction of pairs of histidine residues at the cytoplasmic borders of TM helices to create sites for metal chelation (106) have recently provided insights regarding the functional role of specific helix–helix interactions in rhodopsin.

The indole group of a tryptophan amino acid residue is often used as a noninvasive environment-sensitive probe of protein structure because of its unique absorption and fluorescence properties. Structural studies of bovine rhodopsin and its intermediates by fluorescence spectroscopy of tryptophan residues have been impractical because of the low fluorescence quantum yield (~ 0.022) apparently due to quenching by the retinal chromophore via energy transfer. In one study, however, this property was exploited and served as a basis for monitoring MII decay by measuring the fluorescence increase due to the release of free retinal out of the opsin after Schiff base hydrolysis (107). UV absorption spectroscopy has suggested that the local protein environment around tryptophan residues changes during the conversion of rhodopsin to MII. The difference spectrum obtained by subtracting the UV-visible absorption spectra of MII and rhodopsin shows distinct peaks at 278, 286, 294, and 302 nm. Because these wavelengths are close to the lowest $\pi-\pi^*$ optical transitions of a tryptophan, these UV difference bands were interpreted in terms of the environments of, at most, two tryptophan residues becoming more polar in MII (108). In addition, a linear dichroism study of these bands indicated a reorientation of an indole side chain during the MI \rightarrow MII conversion (109).

There are five tryptophan residues in the primary structure of bovine rhodopsin (positions 35, 126, 161, 175, and 265). Three of these (positions 126, 161, and 265) reside within the membrane-embedded domain of the pigment (Fig. 3). The tryptophan residues at positions 126, 161, and 175 are conserved in nearly all vertebrate rhodopsins and invertebrate visual pigments. The tryptophan residue at position 265 is conserved in all visual pigments except the family of blue cone pigments, where the position is occupied by a tyrosine residue. To determine if tryptophan residues were responsible for the absorbance changes previously reported, each of the five tryptophan residues in rhodopsin was replaced by either a phenylalanine or a tyrosine. The resulting pigments were examined by UV-visible photobleaching difference spectroscopy. Replacements of Trp¹²⁶ and Trp²⁶⁵ resulted in a decrease in the magnitudes of the UV difference peaks at 294 and 302 nm. These results were consistent with a change of Trp¹²⁶ and Trp²⁶⁵ to more polar environments during activation of the receptor (73). It was further suggested that the photoactivation of rhodopsin involved a change in the relative disposition of TM helices 3 and 6, which contain Trp¹²⁶ and Trp²⁶⁵ respectively, within the α -helical bundle of the receptor.

TM helix 3 of rhodopsin is known to be involved in chromophore-protein interactions, and contains the retinal Schiff base counterion, Glu¹¹³. Gly¹²¹, which is strictly conserved in all visual pigments, is located near the middle of TM helix 3. The α -carbon of Gly¹²¹ was predicted to point toward the 9-methyl group of retinal based on a model of the chromophore orienta-

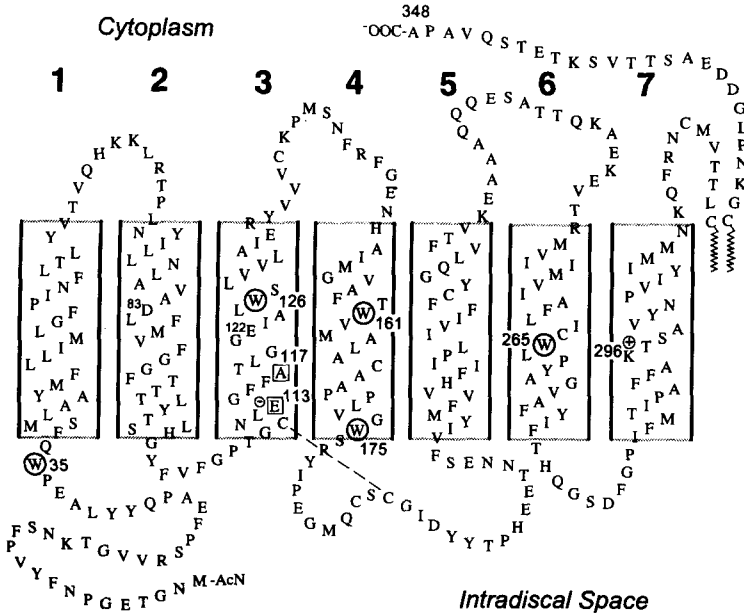


FIG. 3. Secondary structure model of bovine rhodopsin. The five tryptophan residues in the primary structure are circled. Trp¹²⁶ and Trp²⁶⁵ were found to experience more polar environments on formation of R*, consistent with a change in the disposition of TM helices 3 and 6 with respect to the α -helical receptor bundle (73). Reprinted with permission from S. W. Lin and T. P. Sakmar, *Biochemistry* **35**, 11149–11159 (1996). Copyright 1996 American Chemical Society.

tion in the protein derived from solid-state NMR constraints (46). Seven mutants were constructed with replacements at position 121 (G121A, S, T, V, L, I and W) (74). A progressive decrease in mutant pigment stability was reported with increasing residue size at position 121. A progressive blue shift of the mutant pigment λ_{\max} value, which was linear when plotted in wavenumbers against the volume of the side chain of the amino acid at position 121, was also observed. Replacement of Gly¹²¹ also resulted in a relative change of opsin selectivity for reconstitution with 11-*cis*-retinal over all-*trans*-retinal in membranes. All-*trans*-retinal could efficiently bind and activate the mutant opsin, even though none of them displayed constitutive activity. These results showed that TM helix 3, and Gly¹²¹ in particular, were important in determining the specificity of the chromophore-binding pocket in rhodopsin.

In a companion report (75), the G121L mutant was combined with second-site replacements at one of four different positions on TM helix 6:

Met²⁵⁷, Val²⁵⁸, Phe²⁶¹, or Trp²⁶⁵. The spectral and functional phenotypes of the G121L mutant were specifically rescued by replacement of Phe²⁶¹ by alanine, and not other larger substitutions at position 261 or by substitutions at the other three positions. The double mutant G121L/F261A had properties intermediate between those of the G121L mutant and native rhodopsin. These results led the authors to conclude that the retinal binding site may reside between Gly¹²¹ and Phe²⁶¹, and that functional interactions between TM helices 3 and 6 mediated by the chromophore were crucial for receptor photoactivation.

The functional interaction of TM helices 3 and 6 was further probed in a study in which metal ion binding sites were introduced between the cytoplasmic surfaces of TM helices with the aim of restraining specific activation-induced conformational changes (106). Pairs of histidine residues are capable of chelating metal ions such as Zn(II) if the distance and geometry between the residues are appropriate. Histidine residues substituted for the native amino acids at the cytoplasmic ends of TM helices 3 and 6, but not helices 3 and 7, created mutant proteins that were able to activate transducin in the absence, but not in the presence, of metal ions. It was concluded that specific metal-ion crosslinks between positions 138 and 251, or 141 and 251, on TM helices 3 and 6, respectively, prevented receptor activation. Taken together with previous findings described in this section, these results indicated a direct coupling of receptor activation to a change in the spatial disposition of TM helices 3 and 6. This could occur if movement of TM helices 3 and 6 was coupled to changes in the conformation of the connected intracellular loops, which are known to contribute to binding surfaces and tertiary contacts of rhodopsin with transducin (16, 17). Because the arrangement of the seven TM helices is likely to be evolutionarily conserved among the family of G protein-coupled receptors, the proposed motions of TM helices 3 and 6 may be a part of an activation mechanism shared among all receptor subtypes. In other receptor subtypes, ligand binding would have to be coupled to a change in the orientations of TM helices 3 and 6.

C. Light-Induced Molecular Changes on the Cytoplasmic Surface

Synthetic rhodopsin-derived peptides have been shown to compete with native rhodopsin for transducin binding (64). This has allowed the identification of the cytoplasmic loops 3–4 and 5–6, and a putative loop between the cytoplasmic termination of TM helix 7 and the palmitoylated Cys³²² and Cys³²³ residues, as transducin binding sites. Site-directed mutagenesis has further characterized groups of amino acids in these regions implicated in transducin binding and activation (16, 17). However, light-dependent alterations of their physical or chemical states are not well characterized even

where the importance of specific amino acids (e.g., the charged pair Glu¹³⁴/Arg¹³⁵) (110) for transducin activation is established. Infrared spectroscopic determination of protonation states or hydrogen bonding of specific amino acid side chains on the cytoplasmic surface of rhodopsin is to date less advanced than for the hydrophobic core of the receptor. This is in part due to the relatively small contribution of water-exposed amino acids to infrared absorption changes in typical FTIR samples (hydrated films) (111).

Other techniques have been more successful in monitoring light-dependent structural alterations occurring during MII formation. It has been shown by time-resolved (112) and static electron paramagnetic resonance (EPR) spectroscopy studies (113) on site-specific spin-labeled rhodopsin that the cytoplasmic terminations of TM helices 3 and 7 undergo structural rearrangements in the vicinities of Cys¹⁴⁰ and Cys³¹⁶, respectively. These changes have been specifically assigned to the MII conformation. Cys¹⁴⁰ is close to the highly conserved Glu or Asp/Arg/Tyr triad at the cytoplasmic border of TM helix 3 (positions 134–136 in bovine rhodopsin), which attracted attention in earlier studies because of its possible general importance for the function of G protein-coupled receptors.

It has been shown that replacement of Glu¹³⁴ by Gln renders the photoactivated pigment about eightfold more efficient in activating transducin at alkaline pH, compared to recombinant native rhodopsin (110). Therefore, it was suggested that Glu¹³⁴ is a good candidate for regulation of the transducin binding region and may undergo a light-induced transition from an ionized to a protonated state (110, 114). Recent measurements of light-induced pH changes in the bulk-water phase monitored simultaneously with R* formation of the mutants E134Q and E134D showed the involvement of Glu¹³⁴ in proton uptake reactions (63, 115). According to these results, it is likely that Glu¹³⁴ is a group that becomes protonated in MII. It has also been proposed that MII exists in two discrete conformations, termed MII_a and MII_b, which are indistinguishable spectroscopically (115). MII_b activates transducin and MII_a does not. It was proposed that one difference between these two forms of MII was the protonation state of Glu¹³⁴, whereby Glu¹³⁴ becomes protonated during the MII_a to MII_b transition.

The structural change detected by EPR spectroscopy may be directly related to protonation of Glu¹³⁴, which is expected to alter significantly the hydrogen-bonding properties of this amino acid. A rearrangement of neighboring hydrogen-bonding partners may then explain the conformational change. There is no straightforward approach to identify surface groups that interact with Glu¹³⁴, although the pH profile of transducin activation (110) as well as the abolishment of the uptake of two protons in E134Q (63) suggest the existence of other titratable groups influenced by Glu¹³⁴. Because Glu¹³⁴ is at the cytoplasmic border of TM helix 3 in a region demonstrated

to undergo structural alterations upon MII formation, it is not clear from the helical arrangement model which amino acids may interact with Glu¹³⁴, His⁶⁵ and to a lesser extent His¹⁵² have also been shown to affect the receptor conformation because their replacement favors MII formation as measured by 380-nm absorbance (116).

IV. Molecular Switches and Determinants of the Active Receptor Conformation

Except for photoisomerization of the chromophore, individual off/on transitions usually occurring in the hydrophobic core of rhodopsin are not required for switching from an inactive to an active conformation. This was shown by the UV photoactivation of the basic form of mutant E113Q (117), where neither the charge at residue 113 nor the Schiff base protonation state changed between the dark- and the light-activated states. Both groups were in their respective “on” states already in the dark pigment. Consequently, 11-*cis*-retinal efficiently abolishes any conformational change that the neutral states of the Schiff base and the Gln¹¹³ side chain may evoke in favor of an active receptor state. This feature has also been demonstrated by the inactivation of the constitutively active mutant opsin E113Q by the addition of 11-*cis*-retinal, as is also the case with other constitutively active mutants that can be regenerated with 11-*cis*-retinal (53, 114). The free energy of 11-*cis*-retinal binding obviously offsets the increase in the free energy of the mutant opsin due to incorporation of a neutral amino acid at position 113. Therefore, a sufficiently high energy barrier between the dark and photoactivated state of the 11-*cis*-retinal-regenerated rhodopsin is reestablished, which prevents measurable dark activity. It is mutant E113Q that also demonstrates, in a given 11-*cis*-retinal-regenerated opsin, that photoactivation may or may not involve a change of the Schiff base protonation state.

In mutant pigment E113Q, depending on pH, a protonated or an unprotonated 11-*cis*-retinal Schiff base photoisomerizes to the unprotonated all-*trans*-retinal chromophore and forms an active state R* (117). Because the opsin is the same in both cases, it is reasonable that the active states are identical. This mitigates against an essential involvement of the Schiff base protonation state in the switching mechanism, which is also in agreement with the existence of an active receptor conformation with a protonated Schiff base, as shown for mutant pigment E113A/A117E (Fig. 4). It is an astonishing result that a variety of electrostatic transitions occurring in the retinal-binding sites in various recombinant rhodopsins are compatible with the light-dependent formation of a cytoplasmic receptor surface that is recognized by transducin and catalyzes nucleotide exchange.

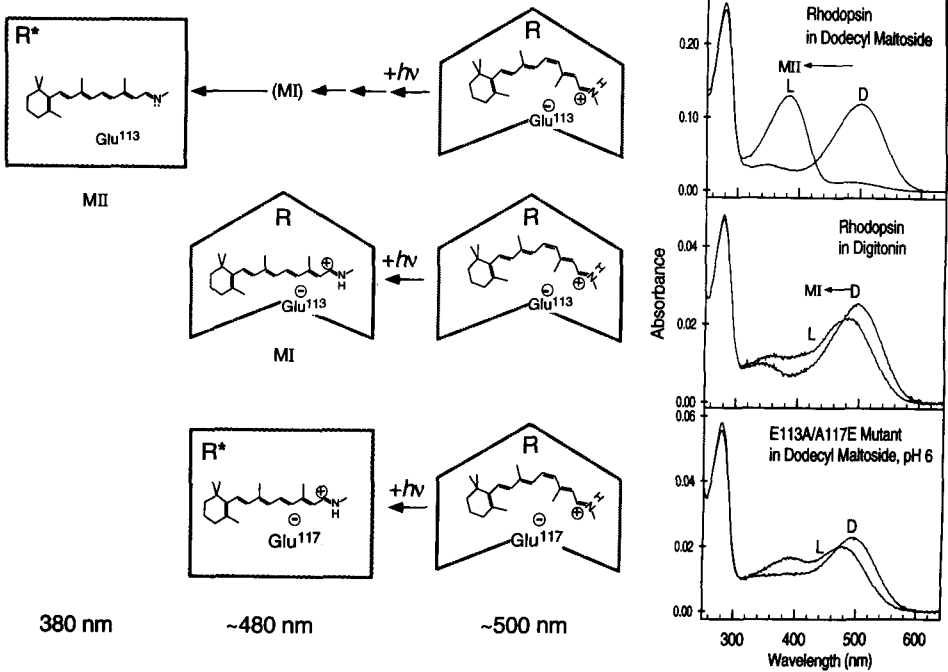


FIG. 4. Spectral forms of rhodopsin do not necessarily correlate to active states of the receptor. The three panels on the right show UV-visible spectra in darkness (D) and after illumination (L) for rhodopsin in dodecyl maltoside (DM) detergent, rhodopsin in digitonin detergent, and mutant pigment E113A/A117E in dodecyl maltoside detergent. Rhodopsin in DM readily forms metarhodopsin II (MII) on illumination. This spectral form of rhodopsin can activate transducin and is known as R^* . The characteristics of the retinal chromophore are shown on the left across the top row, where the active R^* form, shown in a rectangle, consists of a deprotonated (neutral) Schiff base, a protonated (neutral) Glu¹¹³, and a λ_{max} value of 380 nm. In contrast, the MI form of rhodopsin, which is stabilized in digitonin, is shown across the second row. MI will not activate transducin despite the presence of an all-trans chromophore. The protonated Schiff base is generally thought to be responsible for the inactivity of the MI photoproduct. However, in the third row, a mutant pigment in which the Schiff base counterion has been moved by one helix turn (E113A/A117E) is able to form an R^* species, which can activate transducin, even when the Schiff base remains protonated (119, 120). Thus, spectral forms sensitive to Schiff base protonation states do not generally correlate to the active state of the receptor. A similar situation wherein a photoproduct with a protonated Schiff base activates transducin occurs with the mutant G90D, which is responsible for congenital night blindness in humans (52, 121).

The different electrostatic chromophore environments obviously allow spectral tuning of the photoreceptor without interfering with the photoactivation mechanism. Most surprisingly, molecular substates typical of the inactive receptor can be stabilized by mutation in the hydrophobic core and

yet coexist with an active cytoplasmic surface conformation. This has not been anticipated on the basis of biochemical modification of native rhodopsin, which showed that Schiff base deprotonation was required for the formation of MII (36, 118). Similarly, a model that elegantly unifies the possible causes of constitutive activity in a large number of recombinant opsins suggests breakage of a salt bridge between Glu¹¹³ and Lys²⁹⁶ as a key event in receptor activation (51, 53, 114).

In several recombinant rhodopsins, the protonated Schiff base linkage appears to be stabilized by a mutation and yet the photoactivated pigment is still compatible with the R* state (Fig. 4). However, this has been explicitly shown only for mutants E113A/A117E and G90D (52, 119–121). The newly introduced electrostatic interaction between a protonated Schiff base and an “artificial” counterion in the photoproduct does not prevent the active receptor conformation if the charged pair is established between TM helices 3 and 7 (mutant E113A/A117E) and in another case between TM helices 2 and 7 (mutant G90D). Obviously, the active conformation of the receptor surface seems to be compatible with a variety of electrostatic Schiff base environments and stabilization of a particular protonation state is not required. Similarly, the hydrogen bonds of protonated Glu¹²² and Asp⁸³, although undergoing bonding energy changes highly characteristic of the R* formation (121, 122), cannot be essential to maintain the R* conformation, because replacement of either residue does not affect transducin activation (42). Both groups seem to be well-suited reporter groups for steric changes at TM helices 2 and 3 without being directly involved in transducin activation or proton transfer reactions.

V. Coupling of Light-Induced Conformational Changes to Transducin Activation

As discussed above, the domains of rhodopsin that interact with transducin have been studied by site-directed mutagenesis of bovine rhodopsin. Relatively large segments of cytoplasmic loops are required for proper rhodopsin–transducin interaction (17). However, single amino acid substitutions within these domains can have dramatic effects on transducin activation (39, 71). In addition, transducin binding and the activation of bound transducin were shown to be discrete steps involving different surface domains of the receptor (16). Mutant pigments with alterations of cytoplasmic loops 3–4 and 5–6 were characterized that formed spectrally normal MII-like photoproducts that bound transducin (16). However, the bound transducin was apparently defective in the release of GDP, which accounts for the block in pigment-catalyzed GTP uptake or GTPase activity by transducin observed when the mutant pigments were assayed (62).

Due to the free energy of transducin binding to MII, R* is stabilized by transducin binding. An interesting experiment was carried out by reisomerizing the all-trans chromophore in transducin-bound MII in order to create a mismatch between the isomeric “off” state 11-*cis*-retinal and a partially stabilized MII conformation (115). This approach revealed that transducin stays bound to the cytoplasmic surface of R* even if all-*trans*-retinal is replaced by 11-*cis*-retinal by photoconversion. Although the all-*trans* chromophore and a neutral residue at position 113 may be essential determinants of an active receptor conformation, this photoconversion experiment shows that there is no tight steric coupling between the transducin binding loops and the all-*trans* chromophore geometry once transducin has bound. This agrees well with the existence of a large number of constitutively active mutant opsins carrying amino acids of variable sizes at position 296 (114). It should be emphasized, however, that this result does not at all contradict the notion of photoisomerization being a necessary switch to reach the conformation of the free receptor that allows transducin binding.

The light-dependent alterations of the protonation or hydrogen-bonding states of specific molecular groups in the hydrophobic core and on the cytoplasmic surface of rhodopsin, when combined with known structural information, allow several conclusions regarding possible activation mechanisms in rhodopsin. Structural data include the proposed helix arrangement (123, 124) as well as FTIR results obtained with retinal analogs allowing identification of sterically important regions of the chromophore. A characteristic feature of the projection map of rhodopsin, discussed in detail in Section VI, is the smaller distance between the axes of TM helices 3 and 5 as compared to the distance in bacteriorhodopsin (BR), where these helices are effectively separated by TM helix 4. The rhodopsin structure accommodating the 11-*cis* isomer seems to have a more compact arrangement of TM helices 3, 4, and 5 than the more elongated projection structure of BR in which an all-*trans* isomer is incorporated. In a first approximation, photoisomerization in rhodopsin may be expected to cause chromophore–protein interactions that introduce sterical perturbations near the 3/4 and/or the 4/5 helical interfaces. Although this notion agrees with an earlier suggestion of relative movements among TM helices 3 to 5 (113) or 3 to 6 (125, 126), it does not necessarily imply global movements of helices. It may be that at these helical interfaces, amino acid side chains may be expected to experience changes in van der Waal contacts or hydrogen bond strengths on photoisomerization of the retinal chromophore.

A possible functional importance of the relative positions of TM helices 3 and 4 of rhodopsin is also suggested on the basis of a conserved disulfide bond between these helices on the extracellular receptor domain (19, 22, 23), which stabilizes the MII conformation (24). In terms of the chromophore, the

suggested localization of light-induced sterical changes would be expected mostly to affect locations distal to the Schiff base linkage, which is situated at the interfaces of TM helices 2, 3, and 7 (52, 123, 124).

Such a partitioning of crucial protein–chromophore interactions along the retinal is supported by FTIR studies on rhodopsin regenerated with retinal analogs showing that the sterical impacts of the two methyl groups of retinal are not equivalent. The more distal 9-methyl group is essential to induce those conformational changes that give rise to the typical MII absorption changes (127). A drastic reduction in transducin activation is observed in 9-desmethyl retinal rhodopsin. In contrast, the 13-methyl group, which is closer to the Schiff base linkage, does not play a crucial role to sustain the MII conformation (128).

Further evidence for the importance of steric interactions distal from the Schiff base comes from FTIR studies using ring-modified retinal analogs. Increased flexibility, as in 5,6-dihydro (129) or 7,8-dihydro analogs (97), reduces the usually observed torsions along the retinal chain in the batho intermediates at 80 K. In addition, the protein conformational changes observed at temperatures that stabilize the MI or MII intermediates differ from those observed in native rhodopsin. In an extreme case, illumination of a pigment regenerated with a retinal analog lacking the β -ionone ring fails to induce the complete set of infrared absorption changes typical of the MII conformation and results in reduced transducin activation (130). Therefore, the β -ionone ring must transmit important steric changes to the protein. According to the proposed model of light-induced steric perturbation at the TM helix 3/4 interface and the proposed anchoring of the β -ionone ring near TM helix 4, one would expect some effect of ring modifications on infrared absorption changes from amino acids at the TM helix 3/4 interface.

One model of G protein-coupled receptors locates Glu¹²² at the TM helix 3/4 interface (123, 124). The infrared absorption change assigned to the reduction of hydrogen bonding of Glu¹²² on MII formation was indeed not observed in rhodopsin regenerated with a retinal analog that lacks the β -ionone ring, whereas Asp⁸³ showed a normal absorption change. The proposed vicinity of the β -ionone ring and Glu¹²² is also in agreement with resonance Raman data on mutant pigment E122Q (47). Interestingly, infrared absorption changes of Glu¹²² occur prior (in MI) to those of Asp⁸³ (in MII). This is even more pronounced in mutant E122D, which exhibits hydrogen bond alterations of Asp¹²² in the batho intermediate at 80 K when most of the protein conformation is thermally fixed. This suggests even tighter steric coupling of Asp¹²² in E122D to the primary photochemical event than is the case for Glu¹²² in rhodopsin (131).

Taken together it seems likely that the primary steric effect of photoisomerization is mainly localized at the TM helix 3/4 interface where it is sensed

by Glu¹²². Ensuing thermal relaxation would transmit the steric alteration to the cytoplasmic surface and may thus explain why a structural change on MII formation is particularly observed in cytoplasmic loop 3–4 (111, 112).

VI. Structural Modeling of Rhodopsin

Bacteriorhodopsin, the light-driven proton pump of the halophilic bacterium *Halobacterium halobium*, contains seven TM helical segments and a retinylidene chromophore linked via a Schiff base to a specific lysine residue on TM helix 7. Although, these similarities to rhodopsin are intriguing, there is no apparent primary structural homology among BR and visual pigments (including rhodopsin), BR does not couple to G proteins, and BR has very short cytoplasmic loops compared to those of rhodopsin. Nevertheless, some of the methodologies developed for the study of BR have also been applied to the study of rhodopsin (32). In addition, some of the paradigms involving the photochemistry and biophysics of membrane proteins involved in energy transduction or vectorial transport can be applied to both systems (132).

BR forms an ordered two-dimensional lattice in the “purple membrane” of *H. halobium*, which has made it possible to undertake high-resolution structural studies. The three-dimensional structural model of BR (133) and its subsequent refinements have provided a widely accepted template for the superimposition of the seven TM segments of G protein-coupled receptors (134), including rhodopsin (Fig. 5). The lack of a clear evolutionary relationship between BR and rhodopsin makes such modeling approaches somewhat speculative at the outset. For example, it has been inferred that TM helical segments contiguous in the primary structure are juxtaposed in the tertiary arrangement of G protein-coupled receptors, as is the case for BR. This inference was supported by the knowledge that the amino and carboxyl termini of both proteins were known independently of the structural studies to be exposed to the extracellular and cytoplasmic sides of the cell, respectively.

The use of the BR structure as a template for rhodopsin and other G protein-coupled receptors has become perhaps less common since the report of a projection map of rhodopsin. Cryoelectron microscopy and image reconstruction was employed to determine the projection structure at 9 Å resolution of bovine rhodopsin reconstituted into phospholipid bilayers (123). A higher resolution map of rhodopsin was subsequently reported (135). Helix assignments and rotational orientations were proposed for the projection densities based on a comparison of a large number of G protein-coupled receptor amino acid sequences and on relevant biochemical and mutagenesis studies (123). The helix assignments for rhodopsin were consecutive, as was

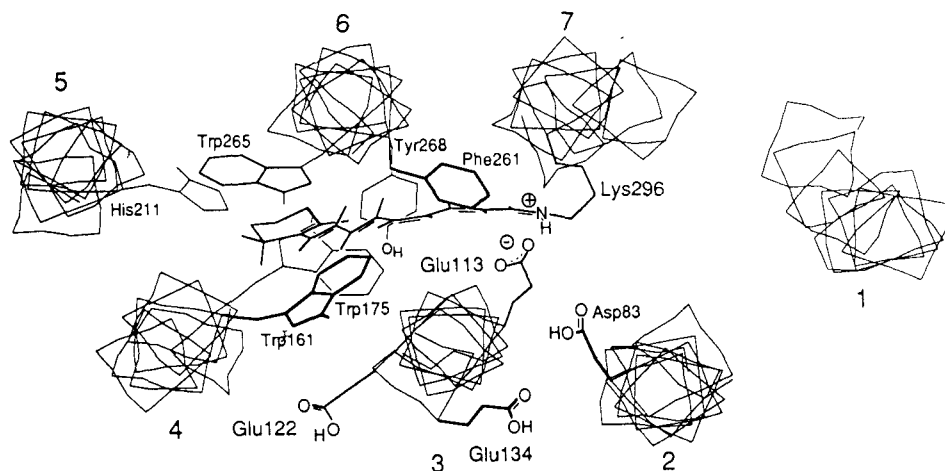


FIG. 5. A molecular graphics model of bovine rhodopsin (47). The helical arrangement in this model is clockwise, in analogy to the structure of bacteriorhodopsin (133). The view is perpendicular to the plane of the disk membrane from above the cytoplasmic surface. TM helix 3 contains the Schiff base counterion residue, Glu¹¹³. In this model, the carboxylic acid group of Glu¹¹³ is positioned between the Schiff base nitrogen and C-13 of the retinal polyene chain. More recent models (44, 75) place the Glu¹¹³ near C-12 of the retinal (see Fig. 6), which is also consistent with predictions based on the study of artificial pigments reconstituted with retinal analogs (98). Water molecules are thought to form a hydrogen-bond network in the vicinity of the Schiff base, but are not included in the model. The β -ionone ring of the retinal is oriented toward Trp²⁶⁵ on TM helix 6. Glu¹³⁴, which is located at the cytoplasmic border of TM helix 3, is highly conserved (as either Asp or Glu) in the family of G protein-coupled receptors. Reprinted with permission from S. W. Lin, T. P. Sakmar, R. R. Franke, H. G. Khorana, and R. A. Mathies, *Biochemistry* **31**, 5105–5111 (1992). Copyright 1992 American Chemical Society.

the case for BR, supporting the major assumption required of models based on the BR structure. However, according to the projection structures presented, significant differences between the proposed rhodopsin structure and the structure of BR were described (123). It was subsequently proposed that these differences might be explained largely by different tilt angles resulting from packing differences between protein arrays (136).

In any case, modeling based on the existing projection maps of rhodopsin has proved informative, and integrated approaches that consider a variety of information in addition to sequence alignments have been developed (137, 138). An example of one such model of rhodopsin is shown in Fig. 6. Another is available on the world wide web at <http://swift.embl-heidelberg.de/7tm/models/lin/index.html> (139). This web site also contains models of numerous other G protein-coupled receptors. Such models have been used extensively in ligand-binding site analysis with the aim of rational drug design. In addi-

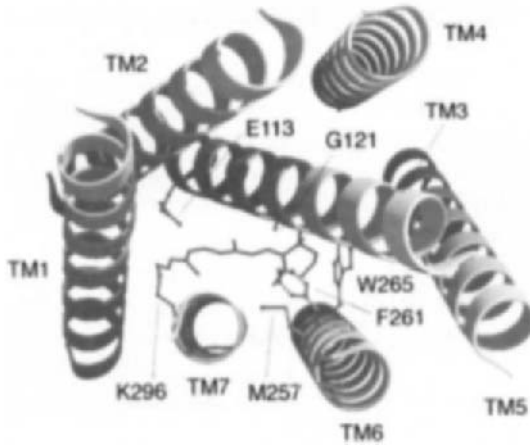


FIG. 6. Molecular graphics model of the retinal-binding pocket of bovine rhodopsin (75). The molecule is viewed down the axis of TM helix 7 from above the plane of the disk membrane. TM helix 7 is thought to be roughly perpendicular to the membrane plane. TM helix 3 is tilted and buried within the bundle of TM helices (123). Key residues that interact with the retinal (and are discussed in the text) include Glu¹¹³, Gly¹²¹, and Phe²⁶¹. At the intradiskal end of TM helix 3, Glu¹¹³ serves as the counterion to the protonated Schiff base localized near C-12 of the retinal (44). At the cytoplasmic end of TM helix 3 are three conserved residues (Glu¹³⁴, Arg¹³⁵, and Tyr¹³⁶), which are known to be involved in G protein binding. Replacement of Gly¹²¹ by a bulky group might move TM helix 3 in the plane of the membrane away from TM helices 6 and 7. An essential Cys¹¹⁰–Cys¹⁸⁷ disulfide bond at the intradiskal surface of helix 3 might act as the pivot point for this motion and restrict vertical translation of the helix relative to the other helices. In the double mutant G121I/F261A, increased steric interaction caused by the leucine residue at position 121 is compensated for by the smaller alanine residue at position 261 on TM helix 6. The results of this study indicate that the retinal lies roughly at the level of Gly¹²¹ and Phe²⁶¹. Gly¹²¹ on TM helix 3 of rhodopsin is conserved among all vertebrate and invertebrate visual pigments (152). Phe²⁶¹ on TM helix 6 is strictly conserved among all classes of G protein-coupled receptors (153). In long-wavelength cone pigments, it is replaced by a tyrosine, which is involved in spectral tuning, indicating its proximity to the retinal (88). The model is based on data from electron microscopy studies of reconstituted bovine and frog rhodopsin. The positions and assignments of the TM helices are according to the projection structure of the proteins, and a primary structure analysis of the G protein-coupled receptor family (123, 124). Gly¹²¹ is the only residue for which protons are displayed. The figure was produced by Dr. Steven O. Smith using the Program MOLSCRIPT (154). Reprinted from Ref. 75 with permission.

tion, when reasonable models are combined with experimental data on receptor activation, including studies of recombinant mutant pigments, a tentative assessment of intramolecular processes that may be of functional importance during activation of rhodopsin and of other G protein-coupled receptors might be attempted either empirically or by molecular dynamic simulations (124, 137, 138).

It should be noted that models based on the projection maps of rhodopsin generally only consider the membrane-embedded domain of the receptor. TM helices can be connected and secondary structures of the loops can be predicted using secondary structure prediction algorithms. However, the structure of the cytoplasmic surface of the receptor is not known. Recently, some three-dimensional structural information concerning the cytoplasmic surface of rhodopsin has been reported. The approach taken was to solve the solution structures of peptides corresponding to putative cytoplasmic domains. Multidimensional NMR (nuclear Overhauser effect) spectroscopy and circular dichroism were employed to study peptides corresponding to the third cytoplasmic loop (loop 5–6) and the carboxyl-terminal tail of bovine rhodopsin (140, 141). The loop 5–6 peptide seemed to fold to form a stable structure that could be docked onto the ends of the appropriate TM helices identified in the projection models.

VII. Rhodopsin Mutations as a Cause of Human Disease

Retinitis pigmentosa is a group of hereditary progressive blinding diseases with variable clinical presentations. One form of the disease, autosomal dominant retinitis pigmentosa (ADRP), was linked to a mutation in the gene for rhodopsin (142). This mutation at codon 23 would result in the change of a proline residue to a histidine residue. This key advance in the molecular genetics of human hereditary blinding diseases has led to considerable work involving recombinant rhodopsins. Over 60 other mutations have been reported in ADRP patients (142–144). Several studies have been carried out in which site-directed mutant opsin genes corresponding to ADRP genotypes were prepared (145–147). When expressed in tissue culture, the mutant opsins display a heterogeneity of spectral properties, biochemical properties, and cellular transport behavior. Some mutants are defective in chromophore binding, others in cellular transport and insertion into the plasma membrane. However, some mutations have no apparent effect *in vitro*.

Significant work in this area is continuing using transgenic mouse technology to produce mouse models of retinal degeneration. There appears to be no early developmental defect involving the retina in RP mouse models, and the work has potential relevance in understanding the molecular and cellular pathophysiology of some forms of ADRP (142–146). Transgenic mice expressing the P23H mutation on the intradiskal domain of rhodopsin have been characterized and results indicate that retinal degeneration may be due to improper mutant opsin expression and cellular transport (148, 149). Other interesting transgenic mouse models, such as those corresponding to mu-

tations of the conserved residues Val³⁴³ or Pro³⁴⁷, point to a defect in targeting of the defective gene product to the rod cell outer segments (150). Rhodopsin mutants with alterations or deletions of these residues near the carboxyl-terminal tail bind retinal to form a pigment. In response to light, they activate transducin and are phosphorylated normally. Further work is needed to clarify how the various rhodopsin mutations associated with ADRP lead to photoreceptor cell death and to degeneration of the many cell layers of the retina.

One particularly interesting mutation linked to retinal degeneration is a replacement of the Schiff base lysine by glutamic acid (A296E). This mutation should prevent chromophore Schiff base formation. Interestingly, the corresponding mutation in bovine rhodopsin was shown to cause constitutive activity of the mutant opsin without chromophore in *in vitro* transducin activation assays (51). Constitutive activity of mutant opsin G90D has also been implicated as a cause of congenital night blindness (52), a defect in scotopic vision in which the dark noise level of the rod cell is increased (151). The mutant pigment associated with congenital night blindness has been characterized in some detail (52, 121). A complete discussion of activating mutations of rhodopsin and other G protein-coupled receptors has been reported (54).

VIII. Conclusions

Over the last several years, a remarkable amount of information about structure–function relationships in G protein-coupled receptors has been obtained using techniques of molecular biology. In the study of visual pigments in particular, site-directed mutant pigments have been employed to elucidate key structural elements, the opsin-shift mechanism, and the mechanism of receptor photoactivation. One particular advantage of studying rhodopsin has been the opportunity to employ various spectroscopic methods in combination with site-directed mutagenesis. Optical spectroscopy and resonance Raman spectroscopy are possible because of the presence of the retinal chromophore, which is probed as a sensor of chromophore–protein interactions. Difference spectroscopy techniques, such as FTIR and UV–visible difference spectroscopy, make use of the chromophore as an optical switch. Overexpression of recombinant rhodopsin also allows a variety of biophysical methods to be used to address particular questions related to conformational changes and protein–protein interactions. Molecular models based on the projection structure of rhodopsin have also proved useful. In summary, it is clear that rhodopsin structure and the molecular mechanism of rhodopsin photoactivation are relevant to understanding the molecular mechanism of signal transduction by G protein-coupled receptors in general.

ACKNOWLEDGMENTS

I gratefully acknowledge the members of my laboratory, including students, postdoctoral fellows, and associates, who have worked on visual phototransduction over the past 5 years: Matthew Albert, Sophia Arnis, Mareike Beck, Kate Carroll, Theresa Chan, Aaron M. Cypess, Padmaja Deval, Oliver Ernst, Karim Fahmy, Stephen A. Gravina, May Han, Manija Kazmi, Melissa Lee, Richard Lee, Steven W. Lin, Ethan P. Marin, K. Christopher Min, Mariya Minkova, Dorothy Wang, Angie You, Wong Yu, Qing Zeng, and Tatyana A. Zvyaga. The assistance of Cliff Sonnenbrot and Carol Valli is also appreciated. In addition, collaborations with the laboratories of F. Siebert, K. P. Hofmann, R. A. Mathies, H. R. Bourne, S. G. Graber, H. Ostrer, S. O. Smith, and D. S. Kliger have been greatly appreciated. Support was provided by NIH Training Grants (GM 07739 and EY 07138) and by a Charles H. Revson Foundation Fellowship. T. P. S. is an Associate Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. T. P. Sakmar, in "Handbook of Receptors and Channels" (S. J. Peroutka, ed.), Vol. I, p. 257. CRC Press, Boca Raton, 1994.
2. R. R. Rando, *Chem. Biol.* **3**, 255 (1996).
3. G. Wald, *Nature (London)* **219**, 800 (1968).
4. M. Chabre, *Annu. Rev. Biophys. Chem.* **14**, 331 (1985).
5. L. Stryer, *J. Biol. Chem.* **266**, 10711 (1991).
6. D. A. Baylor, *Invest. Ophthalmol. Vis. Sci.* **28**, 34 (1987).
7. J. K. Bowmaker and H. J. A. Dartnall, *J. Physiol.* **298**, 501 (1980).
8. M. L. Applebury and P. A. Hargrave, *Vision Res.* **26**, 1881 (1986).
9. Y. A. Ovchinnikov, *FEBS Lett.* **148**, 179 (1982).
10. P. A. Hargrave, J. H. McDowell, D. R. Curtis, J. K. Wang, E. Juszczak, S. L. Fong, J. K. Mohanna Rao, and P. Argos, *Biophys. Struct. Mech.* **9**, 235 (1983).
11. J. Nathans and D. S. Hogness, *Cell* **34**, 807 (1983).
12. J. Nathans and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4851 (1984).
13. R. A. F. Dixon, B. K. Kobilka, D. J. Strader, J. L. Benovic, H. G. Kohlman, T. Frielle, M. A. Bolanowski, C. D. Bennett, E. Rands, R. E. Diehle, R. A. Mumford, E. E. Slater, I. S. Sigal, M. G. Caron, R. J. Lefkowitz, and C. D. Strader, *Nature (London)* **321**, 75 (1986).
14. C. M. Fraser, N. H. Lee, S. M. Pellegrino, and A. R. Kerlavage, *This Series* **49**, 113 (1994).
15. C. D. Strader, T. M. Fong, M. R. Tota, D. Underwood, and R. A. F. Dixon, *Annu. Rev. Biochem.* **63**, 101 (1994).
16. R. R. Franke, B. König, T. P. Sakmar, H. G. Khorana, and K. P. Hofmann, *Science* **250**, 123 (1990).
17. R. R. Franke, T. P. Sakmar, R. M. Graham, and H. G. Khorana, *J. Biol. Chem.* **267**, 14767 (1992).
18. J. L. Benovic, F. J. Mayor, R. L. Somers, M. G. Caron, and R. J. Lefkowitz, *Nature (London)* **321**, 869 (1986).
19. T. Doi, R. S. Molday, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4991 (1990).
20. J. Borjigin and J. Nathans, *J. Biol. Chem.* **269**, 14715 (1994).
21. A. Anukanth and H. G. Khorana, *J. Biol. Chem.* **269**, 19738 (1994).
22. S. S. Karmik, T. P. Sakmar, H.-B. Chen, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8459 (1988).
23. S. S. Karmik and H. G. Khorana, *J. Biol. Chem.* **265**, 17520 (1990).

24. F. F. Davidson, P. C. Loewen, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4029 (1994).
25. K. D. Ridge, Z. Lu, X. Liu, and H. G. Khorana, *Biochemistry* **34**, 3261 (1995).
26. S. Kaushal, K. D. Ridge, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4024 (1994).
27. R. R. Birge, *Biochim. Biophys. Acta* **1016**, 293 (1990).
28. F. Siebert, *Meth. Enzymol.* **189**, 123 (1990).
29. B. Honig, U. Dinur, K. Nakanishi, V. Balogh-Nair, M. A. Gawinowicz, M. Arnaboldi, and M. G. Motto, *J. Am. Chem. Soc.* **101**, 7084 (1980).
30. T. A. Nakayama and H. G. Khorana, *J. Biol. Chem.* **265**, 15762 (1990).
31. H. Zhang, K. A. Lerro, T. Yamamoto, T. H. Lien, L. Sastry, M. A. Gawinowicz, and K. Nakanishi, *J. Am. Chem. Soc.* **116**, 10165 (1994).
32. H. G. Khorana, *J. Biol. Chem.* **267**, 1 (1992).
33. J. Nathans, *Biochemistry* **31**, 4923 (1992).
34. P. A. Hargrave, D. Downs, and J. H. McDowell, *Meth. Enzymol.* **81**, 211 (1982).
35. J. Kibelbek, D. C. Mitchell, J. M. Beach, and B. L. Litman, *Biochemistry* **30**, 6761 (1991).
36. C. Longstaff, R. D. Calhoun, and R. R. Rando, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4209 (1986).
37. E. A. Zhukovsky, P. R. Robinson, and D. D. Oprian, *Science* **251**, 558 (1991).
38. D. D. Oprian, *J. Bioenerg. Biomembr.* **24**, 211 (1992).
39. T. P. Sakmar, R. R. Franke, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8309 (1989).
40. E. A. Zhukovsky and D. D. Oprian, *Science* **246**, 928 (1989).
41. J. Nathans, *Biochemistry* **29**, 937 (1990).
42. K. Fahmy, F. Jäger, M. Beck, T. A. Zvyaga, T. P. Sakmar, and F. Siebert, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10206 (1993).
43. F. Jäger, K. Fahmy, T. P. Sakmar, and F. Siebert, *Biochemistry* **33**, 10878 (1994).
44. M. Han, B. S. DeDecker, and S. O. Smith, *Biophys. J.* **65**, 899 (1993).
45. M. Han and S. O. Smith, *Biophys. Chem.* **56**, 23 (1995).
46. M. Han and S. O. Smith, *Biochemistry* **34**, 1425 (1995).
47. S. W. Lin, T. P. Sakmar, R. R. Franke, H. G. Khorana, and R. A. Mathies, *Biochemistry* **31**, 5105 (1992).
48. K. Fahmy, T. A. Zvyaga, T. P. Sakmar, and F. Siebert, *Biochemistry* **35**, 15065 (1996).
49. A. Surya, K. W. Foster, and B. E. Knox, *J. Biol. Chem.* **270**, 5024 (1995).
50. J. Buczylo, J. C. Saari, R. K. Crouch, and K. Palczewski, *J. Biol. Chem.* **271**, 20621 (1996).
51. P. R. Robinson, G. B. Cohen, E. A. Zhukovsky, and D. D. Oprian, *Neuron* **9**, 719 (1992).
52. V. R. Rao, G. B. Cohen, and D. D. Oprian, *Nature (London)* **367**, 639 (1994).
53. G. B. Cohen, D. D. Oprian, and P. R. Robinson, *Biochemistry* **31**, 12592 (1992).
54. V. R. Rao and D. D. Oprian, *Annu. Rev. Biophys. Biomol. Struct.* **25**, 287 (1996).
55. D. I. Papac, J. E. Oatis, R. K. Crouch, and D. R. Knapp, *Biochemistry* **32**, 5930 (1993).
56. H. Ohguro, K. Palczewski, L. H. Ericsson, K. A. Walsh, and R. S. Johnson, *Biochemistry* **32**, 4968 (1993).
57. J. H. McDowell, J. P. Nawrocki, and P. A. Hargrave, *Biochemistry* **32**, 4968 (1993).
58. J. Chen, C. L. Makino, N. S. Peachey, D. A. Baylor, and M. I. Simon, *Science* **267**, 374 (1995).
59. L. Lagnado and D. A. Baylor, *Nature (London)* **367**, 273 (1994).
60. M. P. Gray-Keller and P. B. Detwiler, *Neuron* **17**, 323 (1996).
61. K. P. Hofmann, S. Jäger, and O. P. Ernst, *Israel J. Chem.* **35**, 339 (1995).
62. O. Ernst, K. P. Hofmann, and T. P. Sakmar, *J. Biol. Chem.* **270**, 10580 (1995).
63. S. Arnis, K. Fahmy, K. P. Hofmann, and T. P. Sakmar, *J. Biol. Chem.* **269**, 23879 (1994).
64. B. König, A. Arendt, J. H. McDowell, M. Kahlert, P. A. Hargrave, and K. P. Hofmann, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6878 (1989).

65. D. J. Takemoto, D. Morrison, L. C. Davis, and L. J. Takemoto, *Biochem.* **235**, 309 (1986).
66. W. J. Phillips, S. C. Wong, and R. A. Cerione, *J. Biol. Chem.* **267**, 17040 (1992).
67. E. R. Weiss, S. Osawa, W. Shi, and C. D. Dickerson, *Biochemistry* **33**, 7587 (1994).
68. S. Osawa and E. R. Weiss, *Mol. Pharmacol.* **46**, 1036 (1994).
69. D. G. Lambright, J. Sondek, A. Bohm, N. P. Skiba, H. E. Hamm, and P. E. Sigler, *Nature (London)* **379**, 311 (1996).
70. K. C. Min, S. A. Gravina, and T. P. Sakmar, Unpublished data.
71. T. A. Nakayama and H. G. Khorana, *J. Biol. Chem.* **266**, 4269 (1991).
72. K. D. Ridge, S. Bhattacharya, T. A. Nakayama, and H. G. Khorana, *J. Biol. Chem.* **267**, 6770 (1992).
73. S. W. Lin and T. P. Sakmar, *Biochemistry* **35**, 11149 (1996).
74. M. Han, S. W. Lin, S. O. Smith, and T. P. Sakmar, *J. Biol. Chem.* **271**, 32330 (1996).
75. M. Han, S. W. Lin, M. Minkova, S. O. Smith, and T. P. Sakmar, *J. Biol. Chem.* **271**, 32337 (1996).
76. J. Nathans, D. Thomas, and D. S. Hogness, *Science* **232**, 193 (1986).
77. D. D. Oprian, A. B. Asenjo, and S. L. Pelletier, *Biochemistry* **30**, 11367 (1991).
78. S. L. Merbs and J. Nathans, *Nature (London)* **356**, 433 (1992).
79. J. Nathans, T. P. Piantanida, R. L. Eddy, T. B. Shows, and D. S. Hogness, *Science* **232**, 203 (1986).
80. J. Neitz, M. Neitz and G. H. Jacobs, *Nature (London)* **342**, 679 (1989).
81. S. L. Merbs and J. Nathans, *Science* **258**, 464 (1992).
82. M. Neitz and J. Neitz, *Science* **267**, 1013 (1995).
83. J. Neitz and M. Neitz, in "Molecular Genetics of Inherited Eye Disorders" (A. F. Wright and B. Jay, eds.), p. 217. Harwood Academic Publ., Chur, Switzerland, 1994.
84. J. Nathans, C. M. Davenport, I. H. Maumenee, R. A. Lewis, J. F. Hejtmancik, M. Litt, E. Lovrien, R. Weleber, B. Bachynski, F. Zwas, R. Klingaman, and G. Fishman, *Science* **245**, 831 (1989).
85. Y. Wang, J. P. Macke, S. L. Merbs, D. J. Zack, B. Klaunberg, J. Bennet, J. Gearhart, and J. Nathans, *Neuron* **9**, 429 (1992).
86. M. Neitz, J. Neitz, and G. H. Jacobs, *Science* **252**, 971 (1991).
87. A. J. Williams, D. M. Hunt, J. K. Bowmaker, and J. D. Mollon, *EMBO J.* **11**, 2039 (1992).
88. T. Chan, M. Lee and T. P. Sakmar, *J. Biol. Chem.* **267**, 9478 (1992).
89. S. L. Merbs and J. Nathans, *Photochem. Photobiol.* **58**, 705 (1993).
90. A. B. Asenjo, J. Rim, and D. D. Oprian, *Neuron* **12**, 1131 (1994).
91. Z. Wang, A. B. Asenjo, and D. D. Oprian, *Biochemistry* **32**, 2125 (1993).
92. T. Yoshizawa and G. Wald, *Nature (London)* **197**, 1279 (1963).
93. D. Bownds, *Nature (London)* **216**, 1178 (1967).
94. E. A. Dratz and P. A. Hargrave, *Trends Biochem. Sci.* **8**, 128 (1983).
95. A. R. Oseroff and R. H. Callender, *Biochemistry* **13**, 4243 (1974).
96. D. S. Kliger and J. W. Lewis, *Israel J. Chem.* **35**, 289 (1995).
97. I. Palings, J. A. Pardoën, E. van den Berg, C. Winkel, J. Lugtenburg, and R. A. Mathies, *Biochemistry* **26**, 2544 (1987).
98. R. R. Birge, L. P. Murray, B. M. Pierce, H. Akita, V. Balogh-Noir, L. A. Findsen, and K. Nakanishi, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4117 (1985).
99. F. Siebert, *Israel J. Chem.* **35**, 309 (1995).
100. K. Fahmy, F. Siebert, and T. P. Sakmar, *Biophys. Chem.* **56**, 171 (1995).
101. T. P. Sakmar and K. Fahmy, *Israel J. Chem.* **35**, 325 (1995).
102. K. J. Rothschild, *J. Bioenerg. Biomembr.* **24**, 147 (1992).
103. P. Rath, L. L. J. DeCaluwé, P. H. M. Bovee-Geurts, W. J. de Grip, and K. J. Rothschild, *Biochemistry* **32**, 10276 (1993).

104. M. V. Milburn, G. G. Privé, D. L. Milligan, W. G. Scott, J. Yeh, J. Jancarik, D. E. Koshland, and S.-H. Kim, *Science* **254**, 1342 (1991).
105. S. Subramaniam, M. Gerstein, D. Oesterhelt, and R. Henderson, *EMBO J.* **12**, 1 (1993).
106. S. Sheikh, T. A. Zvyaga, O. Lichtarge, T. P. Sakmar, and H. R. Bourne, *Nature (London)* **383**, 347 (1996).
107. D. L. Farrens and H. G. Khorana, *J. Biol. Chem.* **270**, 5073 (1995).
108. C. N. Rafferty, C. G. Mullenberg, and H. Shichi, *Biochemistry* **19**, 2145 (1980).
109. M. Chabre and J. Breton, *Photochem. Photobiol.* **30**, 295 (1979).
110. K. Fahmy and T. P. Sakmar, *Biochemistry* **32**, 7229 (1993).
111. U. M. Ganter, T. Charitopoulos, N. Virmaux, and F. Siebert, *Photochem. Photobiol.* **56**, 57 (1992).
112. Z. T. Farahbakhsh, K. Hideg, and W. L. Hubbell, *Science* **262**, 1416 (1993).
113. J. F. Resek, Z. T. Farahbakhsh, W. L. Hubbell, and H. G. Khorana, *Biochemistry* **32**, 12025 (1993).
114. G. B. Cohen, T. Yang, P. R. Robinson, and D. D. Oprian, *Biochemistry* **32**, 6111 (1993).
115. S. Arnis and K. P. Hofmann, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7849 (1993).
116. C. J. Weitz and J. Nathans, *Neuron* **8**, 465 (1992).
117. K. Fahmy and T. P. Sakmar, *Biochemistry* **32**, 9165 (1993).
118. U. M. Ganter, C. Longstaff, M. A. Pajares, R. R. Rando, and F. Siebert, *Biophys. J.* **59**, 640 (1991).
119. T. A. Zvyaga, K. Fahmy, and T. P. Sakmar, *Biochemistry* **33**, 9753 (1994).
120. K. Fahmy, F. Siebert, and T. P. Sakmar, *Biochemistry* **33**, 13700 (1994).
121. T. A. Zvyaga, K. Fahmy, F. Siebert, and T. P. Sakmar, *Biochemistry* **35**, 7536 (1996).
122. A. L. Klinger and M. S. Braiman, *Biophys. J.* **63**, 1244 (1992).
123. G. F. X. Schertler, C. Villa, and R. Henderson, *Nature (London)* **362**, 770 (1993).
124. J. M. Baldwin, *EMBO J.* **12**, 1693 (1993).
125. C. Pellicone, N. J. Cook, G. Nullans, and N. Virmaux, *FEBS Lett.* **181**, 184 (1985).
126. C. Pellicone, G. Nullans, N. J. Cook, and N. Virmaux, *Biochem. Biophys. Res. Commun.* **127**, 816 (1985).
127. U. M. Ganter, E. D. Schmid, D. Perez-Sala, R. R. Rando, and F. Siebert, *Biochemistry* **28**, 5954 (1989).
128. U. M. Ganter, W. Gärtner, and F. Siebert, *Eur. Biophys. J.* **18**, 295 (1990).
129. U. M. Ganter, T. Kashima, M. Sheves, and F. Siebert, *J. Am. Chem. Soc.* **113**, 4087 (1991).
130. F. Jäger, S. Jäger, O. Kräutle, N. Friedman, M. Sheves, K. P. Hofmann, and F. Siebert, *Biochemistry* **33**, 7389 (1994).
131. F. Jäger, T. P. Sakmar, and F. Siebert, in "Fifth International Conference on the Spectroscopy of Biological Molecules" (T. Theophanides, J. Anatassopoulou, and N. Fotopoulos, eds.), p. 223. Kluwer, Dordrecht, Netherlands, 1993.
132. J. K. Lanyi, *Nature (London)* **375**, 461 (1995).
133. R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, *J. Mol. Biol.* **213**, 899 (1990).
134. D. Röper, E. Jacoby, P. Krüger, M. Engels, J. Grötzinger, A. Wollmer, and W. Strassburger, *J. Recept. Res.* **14**, 167 (1994).
135. G. F. Schertler and P. A. Hargrave, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11578 (1995).
136. J. Hofflack, S. Trumpp-Kallmeyer, and M. Hibert, *Trends Pharmacol. Sci.* **15**, 7 (1994).
137. P. Herzyk and R. E. Hubbard, *Biophys. J.* **69**, 2183 (1995).
138. J. A. Ballesteros and H. Weinstein, *Meth. Neurosci.* **25**, 366 (1995).
139. S. W. Lin, R. A. Mathies, and T. P. Sakmar, Unpublished data.
140. P. L. Yeagle, J. L. Alderfer, and A. D. Albert, *Biochemistry* **34**, 14621 (1995).
141. P. L. Yeagle, J. L. Alderfer, and A. D. Albert, *Nature (Struct. Biol.)* **2**, 832–834 (1995).

142. T. P. Dryja, T. L. McGee, E. Reichel, L. B. Hahn, G. S. Cowley, D. W. Yandell, M. A. Sandberg, and E. L. Berson, *Nature (London)* **343**, 364 (1990).
143. T. P. Dryja, L. B. Hahn, G. S. Cowley, T. L. McGee, and E. L. Berson, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9370 (1991).
144. C.-H. Sung, C. M. Davenport, J. C. Hennessey, I. H. Maumenee, S. G. Jacobson, J. R. Heckenlively, R. Howakowski, G. Fishman, P. Gouras, and J. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6481 (1991).
145. S. Kaushal and H. C. Khorana, *Biochemistry* **33**, 6121 (1994).
146. C.-H. Sung, B. G. Schneider, N. Agarwal, D. S. Papermaster, and J. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8840 (1991).
147. K. C. Min, T. A. Zvyaga, A. M. Cypess, and T. P. Sakmar, *J. Biol. Chem.* **268**, 9400 (1993).
148. J. E. Olsson, J. W. Gordon, B. S. Pawlyk, D. Roof, A. Hayes, R. S. Molday, S. Mukai, G. S. Cowley, E. L. Berson, and T. P. Dryja, *Neuron* **9**, 815 (1992).
149. D. J. Roof, M. Adamian, and A. Hayes, *Invest Ophthalmol. Vis. Sci.* **35**, 4049 (1994).
150. C.-H. Sung, C. Makino, D. Baylor, and J. Nathans, *J. Neurosci.* **14**, 5818 (1994).
151. P. A. Sieving, J. E. Richards, F. Naarendorp, E. L. Bingham, K. Scott, and M. Alpern, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 880 (1995).
152. W. C. Probst, L. A. Snyder, D. I. Schuster, J. Brosius, and S. C. Sealfon, *DNA Cell Biol.* **11**, 1 (1992).
153. I. Alkorta and P. Du, *Protein Eng.* **7**, 1231 (1994).
154. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).

Cell Membrane and Chromosome Replication in *Bacillus subtilis*

NOBORU SUEOKA

*Department of Molecular, Cellular, and
Developmental Biology
University of Colorado
Boulder, Colorado 80309*

| | |
|---|----|
| I. Introduction | 36 |
| II. Early Evidence of Membrane–Chromosome Association | 36 |
| III. Chromosome Initiation Mutants of <i>Bacillus subtilis</i> | 39 |
| IV. Preparation of Origin–Membrane and Terminus–Membrane Complexes | 39 |
| V. The <i>dnaB</i> Gene: Critical for Chromosome Initiation and Replication Origin Membrane Attachment | 40 |
| VI. Chromosomal Membrane Attachment Sites | 44 |
| VII. <i>In Vitro</i> Initiation of Chromosome Replication Using the Membrane Fraction | 46 |
| VIII. Membrane Attachment to the Terminus | 47 |
| IX. Differences in Replication Initiation in Two Systems | 47 |
| X. Unsolved Questions | 47 |
| References | 51 |

This review covers studies of the structural and functional roles of the cell membrane on the replication of the *Bacillus subtilis* chromosome. A particular emphasis is placed on the essential roles of the membrane complex for the *in vivo* initiation and termination of the chromosome replication. A critical gene complex in *B. subtilis* for the role of membrane complex is the *dnaB* operon that most likely consists of four genes (*dnaB*, *dnaI*, *ORFZ/ORF213*, and *ORF ω /ORF281*). Detailed studies of these genes are currently available only for the *dnaB* and *dnaI* genes. The unique feature of the *dnaB* gene is that temperature-sensitive mutants of this gene simultaneously lose, at the nonpermissive temperature, chromosome attachment at *oriC* to the membrane as well as the new round of replication initiation at *oriC*. Further studies on the genes and their products of the *dnaB* operon are therefore essential for our understanding of the *in vivo* mechanism of the initiation of chromosome replication and its regulation.

The role of the membrane on the termination and segregation of the daughter chromosomes has not been discovered, but an important clue comes from the terminus area of the *B. subtilis* chromosome being bound to the membrane in a high-salt resistant and DnaB-independent fashion. © 1998 Academic Press

I. Introduction

Bacillus subtilis is a gram-positive, spore-forming soil bacterial species that has several features useful for the study of chromosomal replication: (1) the ability to grow in a minimal medium, (2) highly efficient DNA-mediated transformation, (3) a completely replicated chromosome with all loci occurring at equal frequency in the spore and in the stationary-phase cells of some strains (e.g., W23), and (4) a single lipid bilayer membrane. This review will cover mainly the association of the chromosome with the membrane and the role of the cell membrane proteins and their genes in the regulation of initiation and termination of chromosome replication. Details of earlier work on this subject have been reviewed by Winston and Sueoka (1), and studies on *in vitro* DNA replication using membrane fractions have been reviewed by Firshein (2). More general reviews have been made by Yoshikawa and Wake (3) on the various aspects of *B. subtilis* chromosome replication, and by Wake (4) on chromosomal partition and cell division. For other aspects of bacterial chromosome replication, the readers are referred to reviewed on *Escherichia coli* (5) and *B. subtilis* (6). Here the emphasis is on the points that this author has thought about over the years based on the observations generated mainly in his laboratory studies on *B. subtilis*, and the points that have not been treated explicitly by the previous reviewers. This reviewer also includes his own views of the subjects, some of which may not agree with the previous reviewers.

II. Early Evidence of Membrane–Chromosome Association

The possibility that the bacterial chromosome may be attached to the cellular membrane was first proposed by Jacob and co-workers in 1963 as a part of the replicon theory (7). The proposal was partly based on previously reported electron microscopic observations by Ryter and Jacob (8) that the nuclear body was closely associated with the membrane in *B. subtilis*. In 1967, Smith and Hanawalt (9) reported evidence suggesting that the replication fork of the *E. coli* chromosome was membrane associated. They pulse-labeled exponentially growing *E. coli* with [³H]thymidine and found that the ³H label was enriched in the membrane fraction separated from the soluble fraction by sucrose gradient centrifugation. This suggested that the replication fork may be associated with the membrane; similar findings were also reported in *B. subtilis* (10–12). These observations on the replication fork attachment to the membrane were made under low-salt conditions. Under a high-salt condition, membrane enrichment of pulse label in exponentially

growing cells was not observed, indicating that replication fork attachment is not high-salt resistant.

To identify other chromosomal sites of membrane attachment, it was essential to have available many genetic markers mapped along the chromosome and a method to quantitate the relative frequency of the markers in DNA preparations isolated from the membrane fraction and in those isolated from the soluble fraction. In 1963, we demonstrated that the *B. subtilis* chromosome replicates sequentially from a fixed origin to a fixed terminus (13, 14). For these studies, two methods were developed; one was marker frequency analysis and the other was synchronous density gradient centrifugation. The marker frequency analysis method was developed to show the relative frequencies of a number of chromosomal markers. Essentially, DNA was prepared from exponentially growing cells and compared by genetic transformation with standard DNA preparations (DNA from stationary cells of the W23 strain or DNA from spores) in which all markers are equally frequent. As expected, the ratio of the most frequent marker (*purB6*) to the least frequent marker (*metB5*) was approximately two in the DNA preparations isolated from the culture with a 40-min generation time. We also observed that the ratio of the most to the least frequent marker was four in DNA from the culture with a 20-min generation time (four origins and one terminus; dichotomous replication) (15, 16). The replication order of marker frequency analysis was also examined by synchronous density transfer experiments, which are simply density transfer experiments using synchronously growing cells. The results obtained by these two methods agreed remarkably well. We subsequently identified a marker, *purA16*, that was more closely located to the replication origin than *purB6* (17). By 1967, we determined the replication order of a number of markers and constructed a replication order map of the *B. subtilis* chromosome (17).

The marker frequency analysis was used to examine the possible membrane association of the chromosome by comparing the frequencies of the markers that were found in the membrane fractions and soluble fractions of exponentially growing cells (18). For this study, a marker located in the middle of the chromosome (*leu8*) was used as the standard. The results showed that genetic markers near the replication origin and terminus of the *B. subtilis* chromosome were specifically enriched in the DNA isolated from the membrane fraction (Fig. 1A). The results were subsequently confirmed by several reports (19–24) and were also expanded with additional markers (Fig. 1B) (19). When the chromosome is replicating dichotomously in a rich medium, the four copies of the origin area are also found associated with the membrane (25). The attachment of the replication origin area of the chromosome to the membrane was further supported by pulse-labeling of germinating spores with [³H]thymidine, chasing with nonradioactive thymidine, and fol-

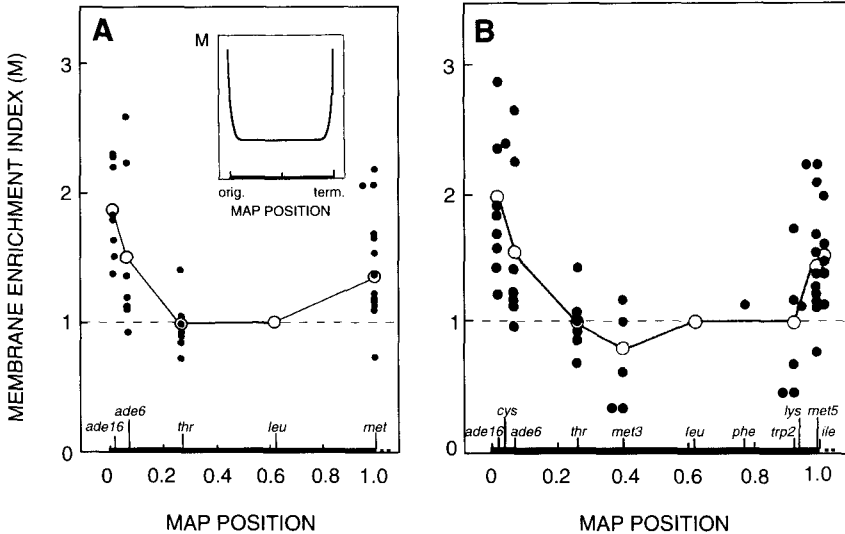


FIG. 1. Membrane enrichment of genetic markers of *B. subtilis*. (A) Initial evidence (18). The map position represents the replication order of markers; 0 corresponds to *oriC*, and 1 to the terminus; ●, the results of individual experiments; ○, the averages. (B) Additional evidence with more markers (19). All markers were isolated by using the UV-penicillin method (70). The membrane enrichment index (M) was calculated by the double ratio transformation method (18).

lowing the labeled DNA in the membrane and free fractions (26). The results showed that the most labeled DNA was found in the membrane fraction after the pulse labeling. Essentially all the label remained associated with the membrane and could not be chased out of the membrane fraction even after one and a half generations. This result suggested that the replication origin may be attached to the membrane throughout the cell cycle. Thus, these early findings strongly suggested that the *B. subtilis* chromosome is permanently associated with the membrane at the replication origin and also at the replication terminus.

Membrane attachment studies in *E. coli* have been intrinsically more difficult because of two reasons: (1) *E. coli* has two membranes, inner and outer, which may make results more variable than in *B. subtilis*, and (2) there is no simple transformation method with high efficiency available in *E. coli*. It has been reported in *E. coli* studies that the hemimethylated state of the *oriC* region is a necessary condition for membrane attachment (reviewed in Ref. 27). The situation in *B. subtilis* is quite different: there is no evidence of enrichment of methylation sites in the *oriC* area (3). Furthermore, *B. subtilis* requires an initiation gene, *dnaB*, for the membrane attachment of the origin

(28). In *E. coli*, there seems to be no operon analogous to the *B. subtilis dnaB* operon.

III. Chromosome Initiation Mutants of *Bacillus subtilis*

Temperature-sensitive mutants of DNA replication genes in bacteria were first isolated in *E. coli* by Kohiyama *et al.* (29) in 1963 and in *B. subtilis* by Mendelson and Gross (30) in 1967. Two types of DNA replication mutants were isolated; the first type was immediate-stop mutants, and the other was gradual-stop mutants. The mutants that gradually stopped DNA replication at a high temperature were particularly interesting because these mutants were possible candidates for initiation mutants in *E. coli* (31–33) and in *B. subtilis* (30, 34–38). Because we had developed unequivocal ways to examine the replication features of the *B. subtilis* chromosome by marker frequency analysis and the synchro-density-transfer technique, we decided to isolate more temperature-sensitive (ts) mutants, hoping to isolate indisputable initiation mutants. A number of ts mutants of the gradual-stop class were isolated, and one of the mutants (*dna-1*) was unequivocally shown as an initiation mutant (36). Subsequently, it was shown (39) that the *dna-1* mutant was located at a site proximal to a previously reported gene, *dnaB19* (34). These two mutants were shown to occupy different regions of the same gene (*dnaB*) (40). Although these two mutants have some similarities, they also have some different properties, for example, a *Staphylococcus*-derived plasmid, pUB110, cannot replicate in the *dna-1* mutant but replicates in the *dnaB19* mutant at the nonpermissive temperature (28).

IV. Preparation of Origin-Membrane and Terminus-Membrane Complexes

Around 1973, we devised the method of CsCl-sucrose double-gradient sedimentation analysis (high-salt condition) (20) to further fractionate the membrane fraction (M fraction) that had been defined previously by sucrose gradient centrifugation or by sucrose step centrifugation (low-salt condition). Using this double-gradient sedimentation technique, we found that the membrane-DNA complexes at *oriC* and at the terminus are high-salt resistant, and DNA and membrane do not dissociate even at 5 M CsCl. It was also found that this method separates the membrane fraction further into two sub-fractions (M1 and M2) (Fig. 2). The DNA/protein ratio was much higher in M2 than in M1 (about 20-fold). However, DNA in the two fractions was equally enriched for the near-origin marker (*purA16*) as well as the terminus

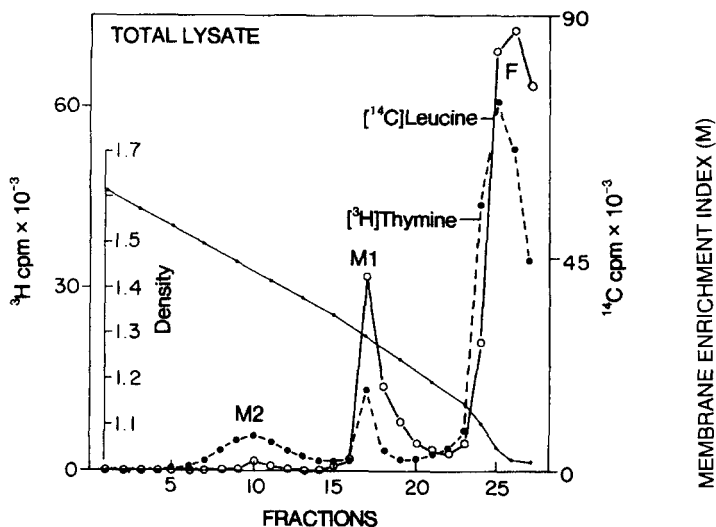


FIG. 2. Fractionation of cell lysate by sucrose–CsCl double-gradient sedimentation (20). See text for details.

marker (*met5*) relative to a midmarker (*leu8*) (20, 41). The M2 fraction was specifically enriched with four proteins of molecular masses 15, 22, 28, and 35 kDa (20). The M2 fractions were isolated from the *dna-1* mutant cultured at a permissive and nonpermissive temperatures, labeled with [^{14}C]leucine, and compared with the M2 fractions of the *dna-1*⁺ strain (168TT) labeled with [^3H]leucine at the respective temperatures. In the *dna-1* mutant, the M2 fraction did not contain the 35-kDa protein when it was labeled at a nonpermissive temperature, suggesting that the 35-kDa proteins may be coded by the *dnaB* gene (42).

V. The *dnaB* gene: Critical for Chromosome Initiation and Replication Origin Membrane Attachment

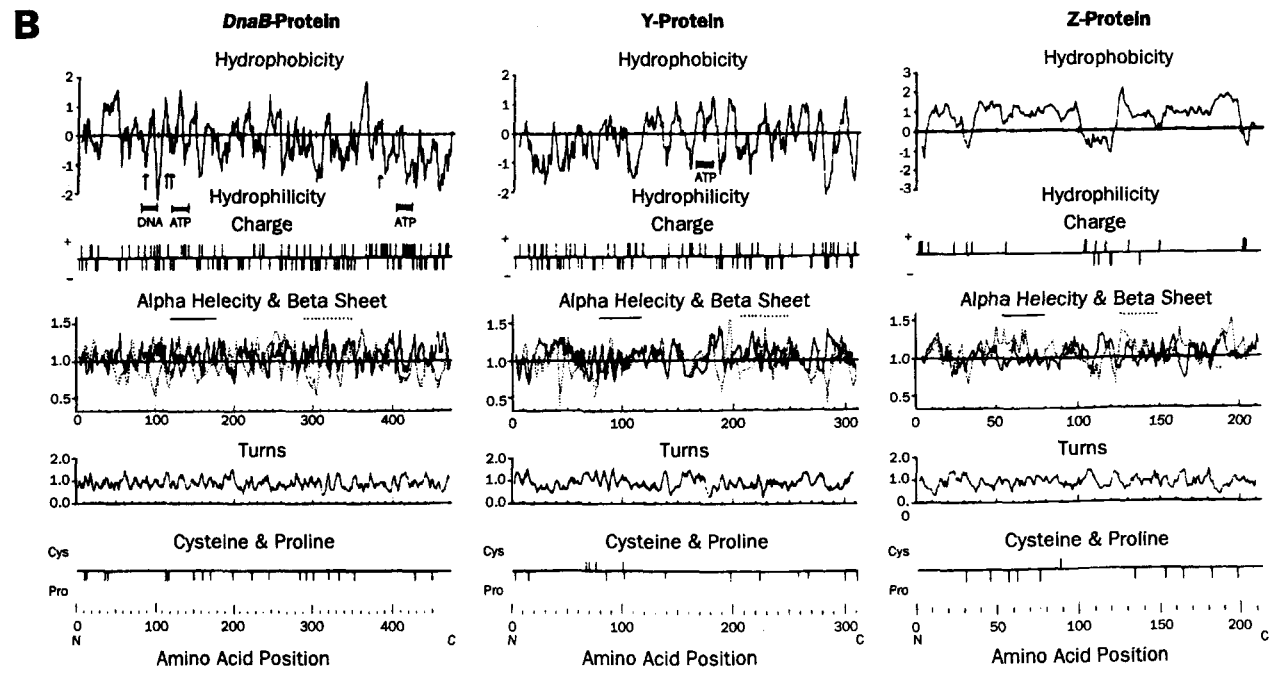
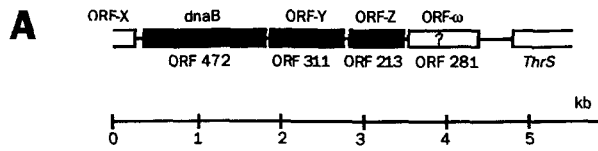
A major breakthrough on the membrane–chromosome work was the discovery that the function of the initiation gene, *dnaB*, is also essential for the origin–membrane binding. The *dnaB* gene, however, is not essential for the terminus–membrane association (28, 43). Therefore, it was concluded that membrane association of the chromosome at the origin may be essential for the initiation of chromosomal replication.

The *Staphylococcus aureus* plasmid pUB110 can replicate in *B. subtilis*

(44, 45). The plasmid is also bound to the membrane *in vivo* (at least ~20 copies per chromosome) and the association is high-salt resistant and its membrane attachment is DnaB protein dependent, as in the case of *oriC* of the chromosome (28). In the *dna-1* (*dnaB1*) mutant, pUB110 cannot replicate at a nonpermissive temperature. The plasmid can be specifically bound to the membrane *in vitro* (46), although the nature of the binding is quite distinct from that *in vivo*. Whereas *in vivo* binding is high-salt resistant and DnaB dependent (Type I binding), *in vitro* binding is salt sensitive and DnaB independent (Type II binding). The four regions that bound *in vitro* did not include the replication origin of the plasmid (47). These results suggest that there are two types of membrane binding in pUB110, Type I and Type II, and only one type (Type II) was reconstituted *in vitro*. Thus, membrane binding of the plasmid replication origin (Type I) requires the function of the *dnaB* gene, as the chromosome does. Type I binding (high-salt resistant and DnaB dependent) has not been achieved in *in vitro* experiments.

The *dnaB* locus was subsequently cloned in two laboratories (40, 48). Combining their results, it was found that the *dnaB* locus is the first gene of an operon (the *dnaB* operon, Fig. 3A) consisting of three or possibly four genes (*dnaB*/*ORF472*, *ORFY*/*ORF311*, *ORFZ*/*ORF213*, and *ORF ω* /*ORF281*; Fig. 3). The *dnaB* gene codes for a protein with 472 amino acid residues (calculated molecular mass, 55 kDa), lacks cysteine, and has both a hydrophobic region (amino acid residues 29 to 49) and a DNA binding motif near the N terminus (amino acid residues 79 to 98; Fig. 4). Subsequently, Watabe and Forough (49) reported that a 15-amino acid peptide near the C terminus (amino acid residues 444 to 458) of DnaB protein also has a DNA binding motif and showed by a filter binding assay that the corresponding synthetic peptide indeed binds to DNA. The *ORFY* was recently reported to be the same as the *dnaI* gene (50). The *dnaI* gene is known to be part of the *B. subtilis* primosome. The third *ORF* (*ORFZ*/*ORF281*) is extremely hydrophobic in almost all regions except proximal and midregions, suggesting that the product is a membrane-embedded protein. No mutant has so far been identified as *ORFZ*. Our laboratory was unable to create a knockout mutant of *ORFZ*, suggesting that it may be essential (J. Laffan and N. Sueoka, unpublished). These facts suggest that the products of these three or possibly four genes may be the major components of the membrane complex that are essential for the initiation of chromosome replication (initiation complex) (51).

Watabe and Forough (52) found that the plasmid pKW1, derived from a low-copy-number plasmid, pBS2, indigenous to *B. subtilis*, was not able to replicate in *dnaB19* at a nonpermissive temperature, whereas in *dnaB1* it replicated in dimeric form but not in monomeric form at a nonpermissive temperature. This implies that the *dnaB19* domain of the *dnaB* gene prod-



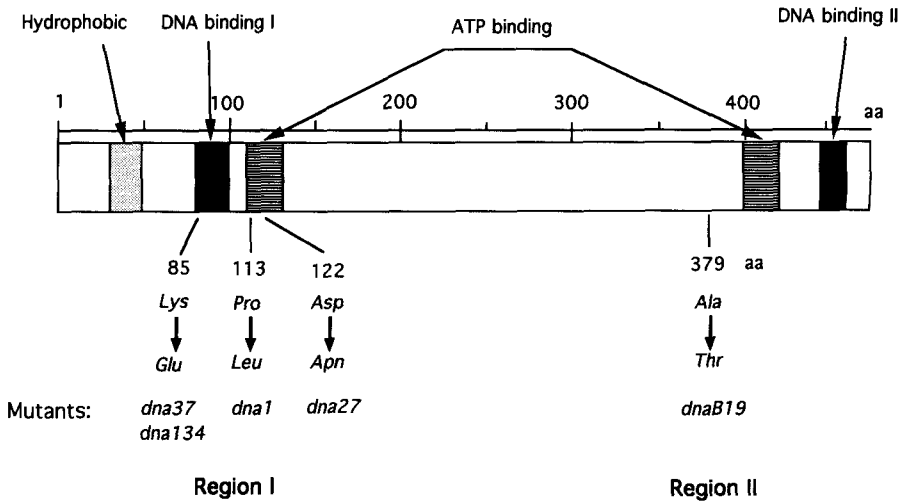
dnaB

FIG. 4. The *dnaB* gene. The DNA binding domain I and the two ATP binding domains are inferred from consensus sequences for DNA binding and for ATP binding (listed in Ref. 40). The 15-amino acid peptide corresponding to the DNA sequence of the DNA binding domain II has been reported to bind both double-stranded and single-stranded DNA by a filter binding assay (49). This C-terminal domain is likely to be removed from the 55-kDa protein to generate a 35-kDa protein in the membrane (J. Laffan and N. Sueoka, unpublished). In light of the fact that both 55- and 35-kDa proteins bind single-stranded DNA strongly, the result of the DNA binding II domain (40, 51) poses interesting functional possibilities about the before and after effects of the posttranslational cleavage of DnaB protein.

uct may also be involved in the resolution of the two nearly completed daughter chromosomes in the terminus region. They also constructed a single-stranded chimerical plasmid consisting of a single-stranded phage M13mp19 and pKW1 (M13mp19-pKW1). Using this single-stranded (ss) chimerical plasmid and double-stranded (ds) pKW1, they showed that *dnaB1* could release both ss and ds plasmids from the membrane but *dnaB19* could release

FIG. 3. The *dnaB* operon. (A) The genes of the *dnaB* operon (see text for details). The ORF notation is based on Refs. 40 and 51 (above the map) and Ref. 48 (below the map). (B) Physicochemical properties of the gene products of the *dnaB* operon deduced from the nucleotide sequence (based on Refs. 40, 48, and 51). The molecular masses of the DnaB, ORFY, and ORFZ proteins are 55, 36, and 23 kDa, respectively. Hydropathy (average values for four consecutive amino acids), α -helicity, and β -sheet potential (average values for four consecutive amino acids) were calculated by using the method of Chou and Fasman (71). Some mutations are shown by upward arrows.

only ds plasmids. They pointed out that the *dnaB* gene has at least two domains that influence the replication (initiation) of the chromosome and its membrane binding. These findings agree with our previous results on the *dnaB* gene (28, 40) and with our current result that DnaB protein binds tightly to ss DNA (J. Laffan and N. Sueoka, unpublished). As to the ds DNA binding to DnaB protein, no detailed study has been made.

Another interesting fact emerged from the study of the *dnaB* gene product. The putative *in vivo* product has a molecular mass (35 kDa) (42) that is different from the expected molecular mass (55 kDa) of the product derived from the DNA sequence of the *dnaB* gene. The 55-kDa protein was also identified by *in vitro* synthesis using the *dnaB* gene clone (40). The possibility that DnaB protein undergoes a posttranslational modification to become a transmembrane 35-kDa protein has some experimental support. Polyclonal antiserum against DnaB protein recognizes two proteins on Western blots, one at approximately 55 kDa and the other at about 35 kDa. This smaller DnaB protein is evident in the *dnaB19* mutant cultured at permissive temperatures but not when cultured at a restrictive temperature (J. Laffan and N. Sueoka, unpublished results).

The *dnaB* gene has been shown by Northern blots to be cotranscribed with the two other open reading frames (ORFs), i.e., *ORFY/dnaI* and *ORFZ*. Some associated proteins have been identified by immunoprecipitation and colocalization in isoelectric focusing gradients. DnaB protein is found in many subfractions—some with both the small and large forms, some with only the large form. A partially purified membrane fraction shows both forms in equal concentrations. The cytosolic fractions appear to have only the larger, unprocessed form (J. Laffan and N. Sueoka, unpublished results).

The DnaB protein is a strong, nonspecific single-stranded DNA binding protein. The binding may have some sequence preference for AT-rich sequences over GC-rich sequences. The binding is extremely salt resistant, as predicted from the Type I binding experiments (20, 28). There is preliminary evidence that the 55-kDa form may be phosphorylated under certain conditions (J. Laffan and N. Sueoka, unpublished results). Detailed biochemical studies of DnaB protein are an essential key to our understanding of the mechanism of initiation and termination and their regulation.

VI. Chromosomal Membrane Attachment Sites

So far, the search for membrane-bound regions of the *B. subtilis* chromosome has revealed three regions that bind to the membrane (Fig. 5). Region A most likely includes the origin (Type I binding; high-salt resistant and

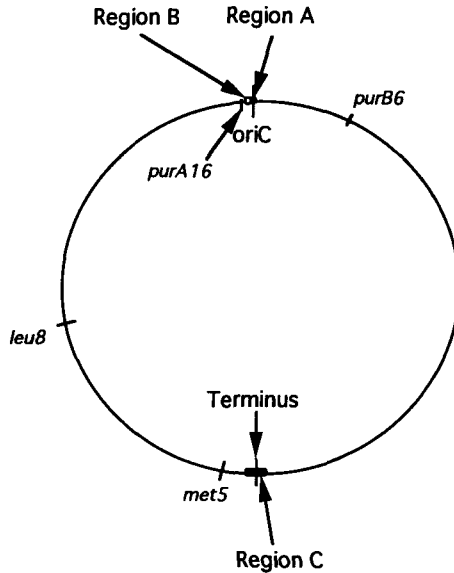


FIG. 5. Membrane binding regions of *B. subtilis* chromosome (see text for details). Chromosomal locations of the genetic markers are based on Anagnostopoulos *et al.* (72).

DnaB dependent) because the *oriC* region was lost from the membrane fraction at a nonpermissive temperature in the initiation mutant *dnaB* (28). Region B (Type II binding; salt sensitive and DnaB independent) was found close to the *purA16* locus (53), and a part of this region turned out to be the same as the membrane binding region (S complex) previously reported by Yamaguchi and Yoshikawa (21) and by Sargent and Bennet (54). The exact location of Region B was determined later to be 45 kb away, counterclockwise, from *oriC* (55). To our surprise, however, when Region B was deleted, the cells apparently grew normally (55). This may indicate either that an additional Type II site(s) exists elsewhere or that Region B may have some subtle effects on chromosome partition or cell division. These possibilities have not been examined. Region C (Type III binding; high-salt resistant and DnaB independent) was located in the terminus region, and the complex was found to be DnaB independent and high-salt resistant (51). Its exact location within the terminus region has not been determined. As described above, we have also shown that the *Staphylococcus* plasmid pUB110 shows both Type I and Type II membrane bindings and that multiple copies of the plasmid can occupy the membrane complex (43).

VII. *In Vitro* Initiation of Chromosome Replication Using the Membrane Fraction

Our attempt in late 1960s to develop an *in vitro* chromosome replication system using the membrane fraction plus soluble fraction was not successful. However, a semi-*in vitro* method using toluenized cells originally developed in *E. coli* by Moses and Richardson (56) also worked well in *B. subtilis* (57). Toluene treatment made cells permeable to nucleoside triphosphates. In this system, DNA replication of toluenized *B. subtilis* cells continued linearly for hours with ATP and the four deoxyriboside triphosphates. The mode of replication was semiconservative due to elongation of the preexisting replication forks, with a rate of approximately one-tenth the normal *in vivo* replication. However, the initiation of chromosome replication at *oriC* was missing from this system. These results are interesting because they suggest that the intact state of membrane may be essential for initiation. However, for elongation, the toluene treatment does not impair the semiconservative replication at the fork, although the replication proceeds at a slower rate.

In vitro replication of the replication origin using a membrane fraction was first observed by Firshein's laboratory (2). Although the reported data were somewhat obscured by high background due to repair DNA synthesis, the result was clearly positive in replication. The results, however, showed no evidence that the new initiation complex was formed *in vitro*. The possibility remains, therefore, that the observed initiation *in vitro* was due to activation or continuation of the initiation process of the preexisting initiation complex that had been formed *in vivo* prior to the preparation of membrane fraction. Moriya *et al.* reported in 1994 that they accomplished *in vitro* initiation using a *B. subtilis oriC*-carrying plasmid and a soluble fraction from *B. subtilis* (58). In their system, the plasmid was propagated in *E. coli* and purified in CsCl-ethidium bromide density gradient centrifugation. The system is clean of repair synthesis, replication after initiation covers the entire plasmid, the *in vitro* initiating plasmid has a structure similar to the cellular chromosome around *oriC* with two DnaA boxes, and the initiation requires DnaA and possibly DnaB proteins. The system, however, is short of the complete *in vivo* initiation in that (1) only 1.5% of the plasmid initiates, (2) only one round of initiation occurs, and (3) the role of DnaB protein was not convincing. Demonstration of the formation of the initiation complex and the biochemical details of the *in vitro* initiation processes with membrane complex have yet to be achieved, and particularly the roles of *dnaB* operon proteins have to be worked out. It is also interesting to note that the three- or four-gene *dnaB* operon of *B. subtilis* has a number of similarities to genes involved in eukaryotic initiation. It may be a short-sighted view to dismiss the role of the membrane components as "controversial" (e.g., Ref. 4) sim-

ply because study of the membrane component in *E. coli* does not give clear results.

VIII. Membrane Attachment to the Terminus

In spite of substantial progress made on the terminus of both *E. coli* (59; 60–63) and *B. subtilis* (3;59a; 64–66), the gene(s) responsible for the terminus–membrane association remains unknown. The structure of the terminus and the mechanism of termination in the two bacteria are similar with some minor differences. However, it is interesting to note that the *B. subtilis* replication terminator protein (RTP) shows some homology with the DnaB protein, particularly around the region of the *dna-1* mutation (67). It has also been reported that RTP protein is found in the temporary replication halt sites (Ster) in both sides of the *oriC* (within 200 kb) (68). Molecular biological studies in *B. subtilis* on this point seem quite feasible by combining partial deletions of regions around the terminus and membrane enrichment of a terminal marker by transformation using membrane-bound DNA.

IX. Differences in Replication Initiation in Two Systems

There are two important differences in the regulation of chromosomal replication and membrane attachment in *E. coli* and *B. subtilis*. First, the membrane attachment in *E. coli* requires a hemimethylated state of *oriC* (69). There is no evidence in *B. subtilis* that a hemimethylated state in the *oriC* region plays an important role for membrane attachment. Second, the *B. subtilis* *dnaB* gene and its three-gene operon (*dnaB–dnaI–ORFZ*) do not seem to exist in *E. coli* (40). In contrast, another major initiation gene, *dnaA*, functions principally in a similar fashion in both *E. coli* and *B. subtilis* (3). Further studies on the interaction of these two initiation genes should be extremely interesting from the viewpoint of the regulation of chromosome replication and its evolution.

X. Unsolved Questions

Further studies may address some of the following unanswered questions relevant to the subject of membrane–chromosome replication.

1. Are the initiation, replication, and termination complexes all associated with the membrane as one huge complex? This question is critical

for discriminating alternative models for the separation of daughter chromosomes, such as shown in Fig. 6.

- Does the initiation complex contain multiple copies of the *dnaB* operon proteins (*DnaB*, *ORFY/DnaI*, and *ORFZ*) and other proteins forming a large structure? Multiple copies of the *dnaB* operon proteins in

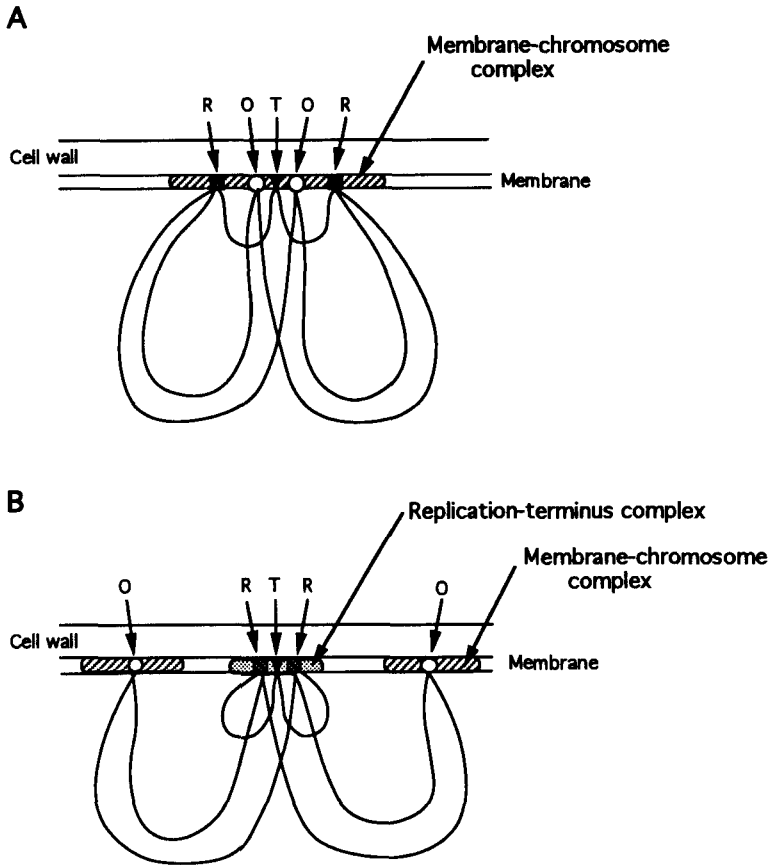


FIG. 6. Two of the possible configurations of chromosome binding to the membrane. (A) A large chromosome-membrane complex accommodates two *oriC*s (O), two replication forks (R), and one terminus (T) of the replicating chromosome. In this model, the complex cannot segregate to two daughter complexes until the replication completes by generating two termini. (B) Two *oriC*s separate further as the chromosome replicates. The two replication forks and the terminus form the replication-terminus complex that is separated from the two membrane-*oriC* complexes. In this model, separation of the chromosomes is gradual. In either model, the membrane complex has a capacity of accommodating about 20 molecules of pUB110 (28).

- the initiation complex are expected based on results of membrane attachment studies of pUB110 (multiple-copy membrane attachment).
3. Is the hydrophilic midregion of ORFZ protein exposed to the outside or the inside of the membrane? This is a critical question that should be answered to understand the role of ORFZ protein in the membrane–DNA complex (Fig. 6).
 4. Is DnaB protein instrumental in holding the replication origin in a certain configuration at the membrane (initiation complex) in order for replication initiation to proceed properly? A clue to the answer may be found in the fact that DnaB protein is a sequence-independent, single-stranded DNA binding protein (74) (Fig. 7).
 5. Do genes homologous to the genes of the *dnaB* operon exist in Archeobacteria or even in eukaryotes? If the functionally similar genes of the *dnaB* operon of a specific amino acid sequence identity exist also in other organisms of different lineages, such as Archeobacteria and eukaryotes, it is likely that the prototype genes of the *dnaB* operon may have already existed in progenotes. In such case, *E. coli* must have lost its *dnaB* operon proteins or modified them sufficiently so that we cannot recognize them today as cognates.
 6. Currently, models for the role of chromosome–membrane association, as shown in Fig. 8, are entertaining. How close is the model to the actual picture? A large part of the answer depends on our knowledge of the structural and functional roles of the *dnaB* operon proteins in *B. subtilis* and other organisms and their relationship with the products of the *dnaA* gene and some other genes.
 7. What is the role of the Type II binding of the *B. subtilis* chromosome? This question is currently difficult to answer because deletion of the

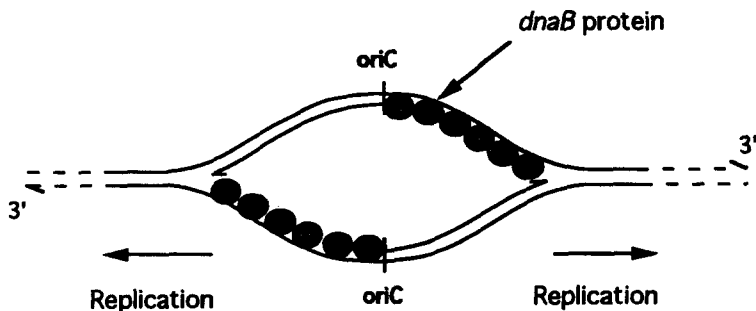


FIG. 7. The *dnaB*-encoded protein binds strongly with single-stranded DNA—a possible configuration of the initiation complex at *oriC*. (Based on unpublished data of L. Laffan and N. Sueoka.)

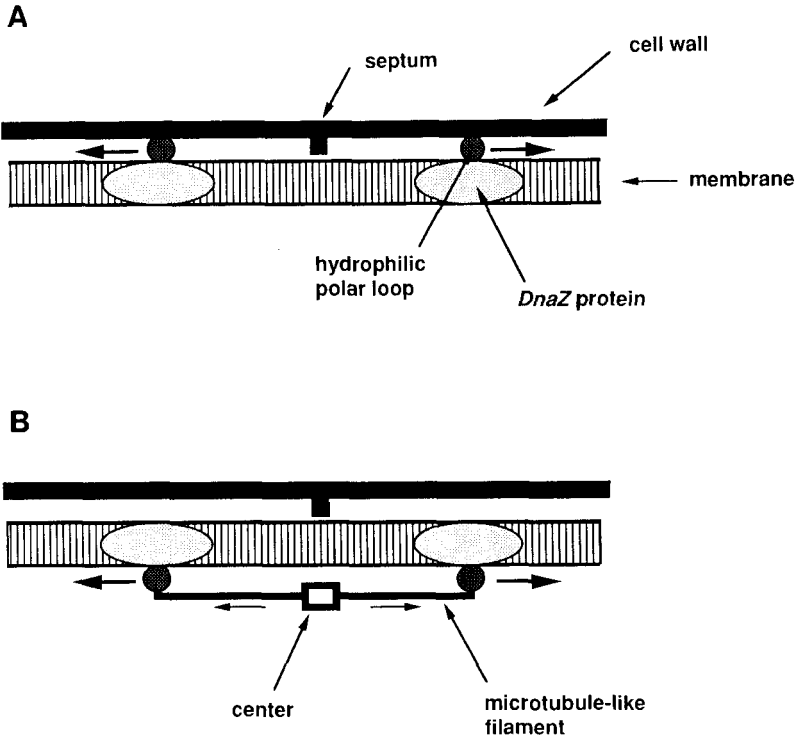


FIG. 8. Membrane sliding models of chromosome separation. In these models, the critical protein is ORFZ protein (Fig. 3), which is likely to be embedded in the cell membrane and may play a pivotal role in the formation of the initiation complex for DNA replication. DNA attachment to the membrane at *oriC* is mediated by DnaB protein, which may interact with ORFZ protein through the proximal hydrophobic region (51) (not shown in the diagrams for the sake of clarity). (A) If the hydrophilic midregion of the ORFZ protein is exposed to the outside of the membrane, separation of the two daughter *oriC*s may be effected by the growing cell wall, to which the ORFZ protein may be attached through the midhydrophilic region. This model is difficult because there is no evidence for the localization of a cell wall growth zone in the midsection of the cell. (B) If the hydrophilic midregion of the ORFZ protein is extruded into the cytoplasm, the separation of the initiation complex, including the *oriC*, may be effected by a hypothetical apparatus within the cytoplasm that is attached to the midregion of the ORFZ protein. Microtubule-like filaments encoded by the *ftsZ* gene have been reported in both *E. coli* and *B. subtilis*, but their involvement in chromosome segregation has not been demonstrated (73). In both Model A and Model B, the daughter initiation complexes are expected to slide in the membrane. (These models were presented at the 8th International Meeting on *Bacillus subtilis* in Paris, 1995.)

Type II binding area does not seem to affect chromosome replication and cell division (55). It is possible that an additional Type II binding region (4) may exist, or some subtle abnormality either in nucleoid segregation or in cell division (septum formation) might be present in the deletion mutant.

8. What is the significance of the salt-resistant terminus binding (Type III binding)? This question seems to be particularly important for the relationship between septum formation and chromosome separation in bacteria, although there is no evidence that the terminus region is essential for chromosome segregation or cell division in *E. coli*.

ACKNOWLEDGMENTS

I thank all the people who worked on *B. subtilis* in my laboratory and whose works are cited in this article. I value our scientific as well as personal interactions and above all our friendship. I thank D. John Laffan and Dr. Tamiko Kano-Sueoka for improving the manuscript. I am grateful to Dr. Kivie Moldave for his invitation to write this article, and for his patience.

REFERENCES

1. S. Winston and N. Sueoka, in "The Molecular Biology of the Bacilli" (D. A. Dubnau, ed.), p. 35. Academic Press, New York, 1982.
2. W. Firshein, *Annu. Rev. Microbiol.* **43**, 89 (1989).
3. H. Yoshikawa and R. G. Wake, in "Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics" (A. L. Sonnenshein, J. A. Hoch, and R. Losick, eds.), p. 507. American Society for Microbiology, Washington, D.C., 1993.
4. R. G. Wake, *Annu. Rev. Genet.* **29**, 41 (1995).
5. F. C. Neithard (ed.) "Escherichia coli and Salmonella: Cellular and Molecular Biology," Vol. I. American Society for Microbiology, Washington, D.C., 1996.
6. A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.), "Bacillus subtilis and Other Gram Positive Bacteria." American Society for Microbiology, Washington, D.C., 1993.
7. F. Jacob, S. Brenner, and F. Cuzin, *Cold Spring Harbor Quant. Biol.* **28**, 329 (1963).
8. A. Ryter and F. Jacob, *Compt. Rend. Acad. Sci.* **257**, 3060 (1963).
9. D. W. Smith and P. C. Hanawalt, *Biochim. Biophys. Acta* **149**, 519 (1967).
10. A. T. Ganesan and J. Lederberg, *Biochem. Biophys. Res. Commun.* **18**, 824 (1965).
11. R. D. Ivarie and J. J. Pene, *J. Bacteriol.* **104**, 839 (1970).
12. K. Yamaguchi, S. Murakami, and H. Yoshikawa, *Biochem. Biophys. Res. Commun.* **44**, 1559 (1971).
13. H. Yoshikawa and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **49**, 559 (1963).
14. H. Yoshikawa and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **49**, 806 (1963).
15. H. Yoshikawa, A. O'Sullivan, and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **52**, 973 (1964).
16. M. Oishi, H. Yoshikawa, and N. Sueoka, *Nature (London)* **204**, 1069 (1964).
17. A. O'Sullivan and N. Sueoka, *J. Mol. Biol.* **27**, 349 (1967).
18. N. Sueoka and W. G. Quinn, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 695 (1968).

19. N. Sueoka, R. J. Bishop, N. Harford, R. H. Kennett, W. G. Quinn, and M. A. O'Sullivan, in "International Symposium on the Genetics of Industrial Microorganisms," p. 73. Prague, 1970.
20. N. Sueoka and J. M. Hammers, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4787 (1974).
21. K. Yamaguchi and H. Yoshikawa, *J. Mol. Biol.* **110**, 219 (1977).
22. K. Yamaguchi and H. Yoshikawa, *J. Bacteriol.* **124**, 1030 (1975).
23. S. Winston and T. Matsushita, in "Microbiology—1976" (D. Schlessinger, ed.), p. 123. American Society for Microbiology, Washington, D.C., 1976.
24. J. Beeson and N. Sueoka, *J. Bacteriol.* **139**, 911 (1979).
25. M. A. O'Sullivan and N. Sueoka, *J. Mol. Biol.* **69**, 237 (1972).
26. W. G. Quinn and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 717 (1970).
27. W. Messer and C. Weigle, in "Escherichia coli and Salmonella: Cellular and Molecular Biology" (F. C. Neidhardt, ed.), Vol. 1, p. 1579. American Society for Microbiology, Washington, D.C., 1994.
28. S. Winston and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2834 (1980).
29. M. Kohiyama, H. Lanfrom, S. Brenner, and F. Jacob, *Compt. Rend. Acad. Sci.* **527**, 1979 (1963).
30. N. H. Mendelson and J. D. Gross, *J. Bacteriol.* **94**, 1603 (1967).
31. F. Bonhoeffer and H. Schaller, *Biochem. Biophys. Res. Comm.* **20**, 93 (1965).
32. M. Kohiyama, D. Cousin, A. Rytter, and F. Jacob, *Ann. Inst. Pasteur* **110**, 465 (1966).
33. W. L. Fangman and A. Novick, *Genetics* **60**, 1 (1968).
34. D. Karamata and J. Gross, *Mol. Gen. Genet.* **108**, 277 (1970).
35. S. J. Laurent, *J. Bacteriol.* **116**, 141 (1973).
36. K. White and N. Sueoka, *Genetics* **73**, 185 (1973).
37. P. Upercroft, H. J. Dyson, and R. G. Wake, *J. Bacteriol.* **121**, 121 (1975).
38. S. Murakami, N. Inuzuka, M. Yamaguchi, K. Yamaguchi, and H. Yoshikawa, *J. Mol. Biol.* **108**, 608 (1976).
39. S. Imada, L. E. Carroll, and N. Sueoka, *Genetics* **94**, 809 (1979).
40. T. Hoshino, T. McKenzie, S. Schmidt, T. Tanaka, and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 653 (1987).
41. S. Imada, L. E. Carroll, and N. Sueoka, in "ICN-UCLA Winter Conference in Molecular and Cellular Biology" (M. Goulian and P. Hanawalt, eds.), p. 187. University of California, Los Angeles, 1975.
42. S. Imada, L. E. Carroll, and N. Sueoka, in "Microbiology—1976" (D. Schlessinger, ed.), p. 116. American Society of Microbiology, Washington, D.C., 1976.
43. S. Winston and N. Sueoka, *J. Bacteriol.* **142**, 339 (1980a).
44. T. J. Gryczan, S. Contente, and D. Dubnau, *J. Bacteriol.* **134**, 318 (1978).
45. K. Keggins, P. Lovett, and E. J. Duvall, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1423 (1978).
46. R. Korn, S. Winston, T. Tanaka, and N. Sueoka, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 574 (1983).
47. T. Tanaka and N. Sueoka, *J. Bacteriol.* **154**, 574 (1983).
48. N. Ogasawara, S. Moriya, P. G. Massa, and H. Yoshikawa, *Nucleic Acids Res.* **14**, 9989 (1986).
49. K. Watabe and R. Forough, *Biochem. Biophys. Res. Comm.* **145**, 861 (1987).
50. C. Bruand and S. D. Ehrich, *Microbiology* **141**, 1199 (1995).
51. N. Sueoka, T. Hoshino, and T. McKenzie, in "Genetics and Biotechnology of Bacilli" (A. T. Ganesan and J. A. Hoch, eds.), p. 269. Academic Press, San Diego, 1988.
52. K. Watabe and R. Forough, *J. Bacteriol.* **169**, 4141 (1987).
53. Y. Sato, M. McCollum, T. McKenzie, J. Laffan, A. Zuberi, and N. Sueoka, *J. Bacteriol.* **173**, 7732 (1991).
54. M. G. Sargent and M. F. Bennett, *J. Bacteriol.* **166**, 38 (1986).
55. M. Itaya, J. J. Laffan, and N. Sueoka, *J. Bacteriol.* **174**, 5466 (1992).

56. R. E. Moses and C. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **67**, 674 (1970).
57. T. Matsushita, K. P. White, and N. Sueoka, *Nature (London)* **232**, 111 (1971).
58. S. Moriya, W. Firshein, H. Yoshikawa, and N. Ogasawara, *Mol. Microbiol.* **12**, 469 (1994).
59. T. M. Hill, in "Escherichia coli and Salmonella" (F. C. Neidhardt, ed.), Vol. 1, p. 1602. American Society of Microbiology, Washington, D.C., 1996.
- 59a. D. Bastia and B. K. Mohanty, in "DNA Replication in Eukaryotic Cells" (M. DePamphilis, ed.), p. 177. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996.
60. T. Horiuchi and Y. Fujimura, *J. Bacteriol.* **177**, 783 (1995).
61. P. Kuempel, A. Hogaard, M. Nielsen, O. Nagappan, and M. Tecklenburg, *Genes Dev.* **10**, 1162 (1996).
62. B. Sherma and M. T. Hill, *J. Bacteriol.* **18**, 54 (1995).
63. J. Louarn, F. Cornet, V. Francoir, J. Patte, and J. Louarn, *J. Bacteriol.* **176**, 7524 (1994).
64. D. E. Bussiere, D. Bastia, and S. W. White, *Cell* **80**, 651 (1995).
65. A. C. Manna, K. S. Pai, D. E. Bussiere, C. Davies, S. W. White, and D. Bastia, *Cell* **87**, 881 (1996).
66. A. H. Franks, A. A. Griffiths, and R. G. Wake, *Mol. Microbiol.* **17**, 13 (1995).
67. A. V. Kralicek, A. J. Day, R. G. Wake, and G. F. King, *Biochem. J.* **275**, 823 (1991).
68. A. Levine, S. Autret, and S. J. Seror, *Mol. Microbiol.* **15**, 287 (1994).
69. P. J. Leibowitz and M. Schaechter, *Int. Rev. Cytol.* **41**, 1 (1975).
70. B. Davis, *J. Am. Chem.* **70**, 4267 (1948).
71. P. Y. Chou and G. D. Fasman, *Adv. Enzymol.* **47**, 45 (1978).
72. C. Anagnostopoulos, P. J. Piggot, and J. A. Hoch, in "Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics" (A. L. Sonenshein, J. A. Hoch, and R. Losick, eds.), p. 507. American Society for Microbiology, Washington, D.C., 1993.
73. J. Lutkenhaus and A. Mukherjee, in "Escherichia coli and Salmonella: Cellular and Molecular Biology" (F. C. Neidhardt, ed.), Vol. 1, p. 1615. American Society for Microbiology, Washington, D.C., 1996.
74. J. Laffan and N. Sueoka, in preparation (1997).

This Page Intentionally Left Blank

Stability and Structure of Model DNA Triplexes and Quadruplexes and Their Interactions with Small Ligands

RICHARD H. SHAFER

*Department of Pharmaceutical Chemistry
School of Pharmacy
University of California, San Francisco
San Francisco, California 94143*

| | |
|--|----|
| I. Triple-Helical Structures | 57 |
| A. Background | 57 |
| B. Basic Triplex Properties | 57 |
| C. Spectroscopic Characterization of Triplexes | 61 |
| D. Triplex Thermodynamics | 69 |
| E. High-Resolution Structural Aspects of Triplexes | 75 |
| F. Ligand-Triplex Interactions | 77 |
| II. Guanine Quadruplex Structures | 79 |
| A. Background | 79 |
| B. Basic Properties of Quadruplexes | 81 |
| C. Spectroscopic Characterization of Quadruplexes | 82 |
| D. Quadruplex Thermodynamics | 85 |
| E. High-Resolution Structure of Quadruplexes | 86 |
| F. Quadruplex-Ligand Interactions | 89 |
| III. Summary | 91 |
| References | 91 |

This review focuses on the structural and thermodynamic characterization of model DNA triplex and quadruplex structures, taking into account effects of stoichiometry and sequence. Methods such as gel electrophoresis, UV melting, and scanning calorimetry, and the results thereof, are described for determination of the thermodynamic stability of such systems. Three classes of triplexes are considered based on the composition of the third strand, while quadruplex systems are limited to those based on the guanine quartet. X-ray crystallography and high resolution NMR studies are also described for these two classes of unusual structures. Ligand binding to triplexes and quadruplexes is also reviewed, with emphasis on specific molecular recognition. The availability of three-dimensional structures for triplex and quadruplex species sets the stage for structure-based development of ligands capable of binding to them specifically. To this end, we consider the application of DOCK, a program for the discovery of small molecules that can recognize macromolecular structures, to the problem of recognizing folded quadruplex structures. Such studies may ultimately lead to pharmaceutically active compounds. © 1998 Academic Press

For many years, structural diversity and complexity were characteristics often attributed to proteins but rarely to nucleic acids. However, it has become widely apparent that nucleic acids are capable of forming myriad structures composed of one, two, three, or four strands and involving a variety of different folding patterns. Interestingly, the first report of a three-stranded nucleic acid structure appeared only 4 years after Watson and Crick's historical announcement of the DNA double helix (1). Although a limited number of early studies appeared characterizing the triple helix, interest has been rekindled in the triplex conformation and other unusual conformations, fueled largely by intriguing hints of biological significance along with the ready availability of oligonucleotides of high purity. RNA structures have evolved an even greater richness in complexity, associated with a wider range of biological function.

Nucleic acids exist within the complex realm of the cellular milieu, and usually as segments within extremely long polymeric units, but the bulk of our detailed knowledge of their structure at the molecular and atomic level derives from studies on model systems composed of short oligonucleotides. For example, although structural models such as that of Watson and Crick were developed from fiber diffraction studies on polynucleotide systems, details of the Watson–Crick double helix at the atomic level became available only when short duplexes were crystallized, which required substantial amounts of highly purified material. With the advent of automated DNA, and now RNA, synthesis, large amounts of oligonucleotides are readily available, and this development has led to a vast increase in the number of structural and thermodynamic studies on model nucleic acid systems. Whereas very high-resolution three-dimensional structure analysis is possible in such model systems, results must be interpreted with care in terms of their applicability to biological systems. Nonetheless, from this approach, much can be learned about the structural possibilities of nucleic acids in their more complicated biological context.

The scope of this article is limited to selected studies on the structure and stability of short DNA triplexes and quadruplexes and their interactions with small molecules, rather than an exhaustive review of the current literature in this very rapidly changing field. More general reviews of triple-helical structures can be found in the recent monograph of Soyfer and Potaman (2) as well as in several articles (3–5). Similarly, several general reviews of quadruplex structures are available (6, 7). The thermodynamics of both triplexes and quadruplexes have been recently surveyed (8, 9). Developments in high-resolution nuclear magnetic resonance (NMR) have been utilized to determine solution structures of these systems and have been reviewed in a collection of articles (10). In what follows, we focus on a relatively small number of studies to illustrate the basic properties of these systems, covering a wide array

of experimental techniques. We have provided a brief explanation for some of the less well-known techniques before discussing their application to model triplex and quadruplex systems.

I. Triple-Helical Structures

A. Background

Felsenfeld *et al.* reported the first instance of triple-helical nucleic acid structures in a study on the RNA system composed of poly(A) and poly(U) (1). This system exhibits a complicated phase diagram involving single-, double-, and triple-stranded conformations, with temperature and salt concentrations being the principal variables governing changes in state. Subsequent reports appeared on the corresponding DNA system composed of poly(dA) and poly(dT) (11). The currently widespread interest in triplex structures was to a large extent influenced by investigations of single-strand-specific nuclease activity in chromatin and in closed circular plasmids containing homopurine•homopyrimidine inserts such as $d(GA)_n \cdot d(TC)_n$ (12–14). Experiments on plasmids were interpreted in terms of acid-induced formation of intramolecular triple helix structures as a mechanism to relieve superhelical stress. Thus, at low pH, the two strands containing the homopurine•homopyrimidine insert separate and half of the pyrimidine strand folds back onto an intact duplex region to form a triple helix, as illustrated in Fig. 1. Such structures were dubbed H-DNA (13, 14). More recently, evidence has been reported suggesting that such structures may play a role in DNA transcription and replication (15).

This work on supercoiled plasmids containing intramolecular H-DNA triplex forms stimulated interest in the triplex conformation. Shortly thereafter, work by Dervan (16) and Helene (17) demonstrated the feasibility of sequence-specific binding of oligonucleotides to double-helical DNA via triplex formation, which led to new opportunities for utilizing DNA as a gene-regulatory and chemotherapeutic agent. For completeness, we note that there is an additional form of triplex, induced by recombinase proteins, that is quite different from the triplexes described here (18, 19). Our discussion will not include these recombinase-assisted triplexes.

B. Basic Triplex Properties

Triple-helical structures are formed when oligonucleotides or polynucleotides bind to the major groove of a double helix. One of the most fundamental aspects of nucleic acid triple helices is the requirement for a homopurine•homopyrimidine motif in the duplex at the sequence targeted for

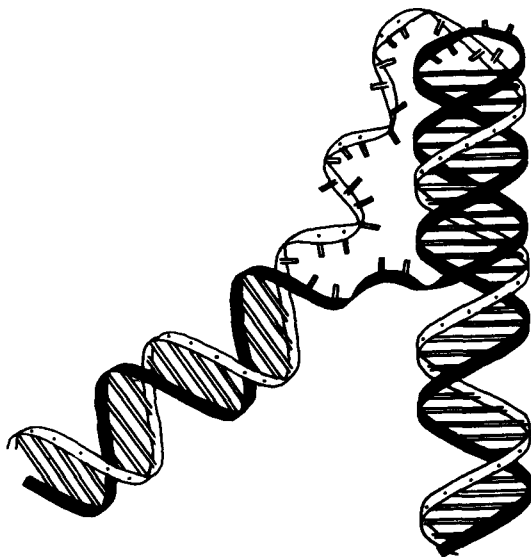


FIG. 1. Model of an intramolecular, H-form triplex helix. The central portion of this sequence undergoes strand separation and the pyrimidine strand (shaded) binds to the major groove of the adjacent duplex sequence to form a triple-helical region. Reprinted with permission (107).

oligonucleotide binding. This requirement arises from the fact that only the purine bases possess sufficient hydrogen bonding functionality such that they can interact with a third-strand base yet still maintain Watson–Crick hydrogen bonding in the underlying duplex. Thus the bases in the third strand must track the purine bases in the duplex in order effectively to hydrogen bond with them. If the purine bases alternate between the two duplex strands, it becomes sterically difficult or impossible for the third strand to achieve the conformation necessary for effective hydrogen bonding and stacking. However, when the purine bases are all on a single strand, efficient interaction of an additional oligonucleotide strand is indeed possible, leading to triplex formation.

In order for third-strand binding to occur, the ionic strength usually has to be increased to provide screening of the negative charge on all three strands. In contrast to Watson–Crick base pairing, which always places an adenine (A) opposite a thymine (T), and a guanine (G) opposite a cytosine (C), there is a certain degree of degeneracy in the hydrogen bonding interactions between the duplex purine and the third-strand bases, typically referred to as Hoogsteen or Hoogsteen-like hydrogen bonding. For example, an adenine

in the duplex purine strand can base pair with either a thymine or another adenine in the third strand. Similarly, a guanine in the duplex purine strand can base pair with either a cytosine or guanine in the third strand. This degeneracy leads to the possibility that several different oligonucleotide sequences can bind to the same duplex, giving rise to several different classes of triplexes. The principal classes include those in which the third strand is composed solely of pyrimidines (pyr*pur*pyr triplex), purines (pur*pur*pyr triplex), or a mixture of G and T (pur/pyr*pur*pyr triplex). The hydrogen bonding schemes characteristic of these three classes of triplexes are illustrated in Fig. 2. Note that pyr*pur*pyr triplexes containing cytosine in the third strand typically require acid pH in order to protonate the cytosine N-3 position, which is necessary for Hoogsteen hydrogen bond formation.

In the common B conformation (Watson-Crick) double helix, the two strands must be antiparallel to each other. Triplex formation involves somewhat more complicated polarity rules, due to the multiple classes of possible

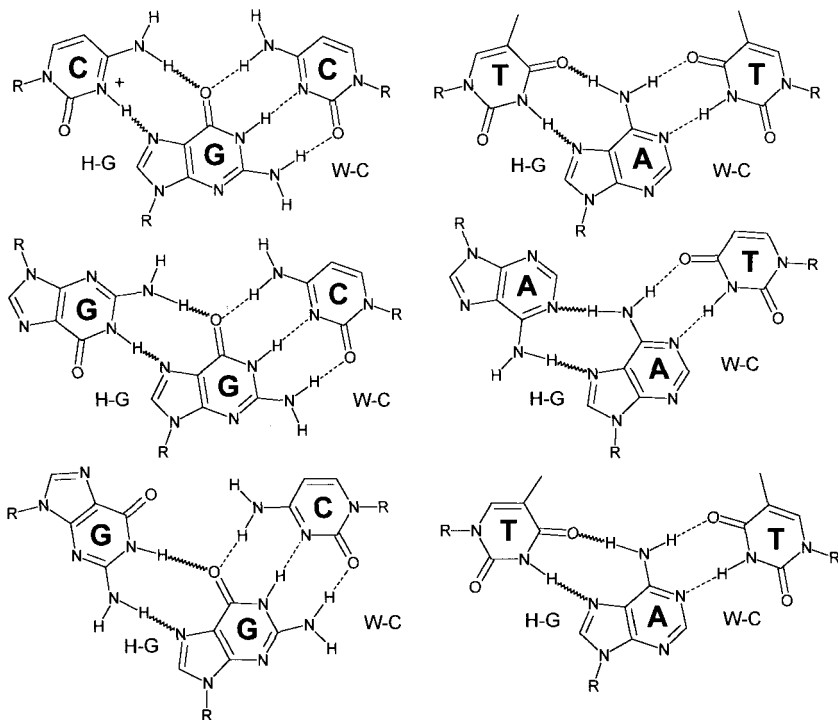


FIG. 2. Hydrogen bonding schemes for pyr*pur*pyr triplexes (top), pur*pur*pyr triplexes (middle), and pur/pyr*pur*pyr triplexes (bottom).

structures. In pyr•pur•pyr triplexes, the third strand is antiparallel to the equivalent pyrimidine strand in the duplex, whereas in pur•pur•pyr triplexes, the third strand is antiparallel to the cognate purine strand in the duplex. Sometimes these are referred to as parallel and antiparallel orientations, respectively, based on the duplex purine strand as a reference. The situation regarding strand orientation in pur/pyr•pur•pyr triplexes is more complicated. Here, it appears that the third-strand orientation depends on the third-strand sequence—when there are many TpG and GpT junctions, then the third strand prefers to be antiparallel to the purine strand of the duplex, whereas in the case of a small number of TpG or GpT sequences the parallel orientation is preferred (20).

Molecularity is another degree of freedom that can be manipulated in the design of triple helices. Just as a duplex can be formed from a single strand in a hairpin conformation, a triplex can be formed from a single strand containing three base-pairing sequences separated by two loop-forming sequences. Alternatively, a triplex can be composed of a single strand binding to a second, hairpin strand. Finally, three separate strands can form an intermolecular triplex. These possibilities are schematically illustrated in Fig. 3.

In general, then, there is considerably more freedom in designing a third strand for triple helix formation with a particular duplex than in designing an oligonucleotide for duplex formation. There are often three or more possible triplexes that can be formed on a given target duplex. For example, Giovannangeli *et al.* have reported a triplex in which the third strand is composed of T, C, and G (21), and triplexes containing a third strand composed of G, T, and A are also possible (P. V. Sacria and R. H. Shafer, unpublished results). Thus, in fact, there are many possible third-strand sequences capable of binding to a particular target duplex. One of the important practical questions concerns the relative stability of these different classes of triplexes. Another question of general importance regards the dependence of stability on sequence within each class of triplex. We will address these points later in terms of several triplex sequences.

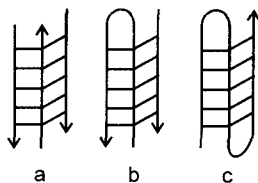


FIG. 3. Three possible configuration for triplexes: (a) three stranded, (b) two stranded, and (c) one stranded (unimolecular).

C. Spectroscopic Characterization of Triplexes

1. PYR*PUR*PYR TRIPLEXES

Following the earlier developments in the field of triple-helical nucleic acids, which commenced with studies on poly(A) and poly(U) or poly(dA) and poly(dT), several groups have examined short pyr*pur*pyr triplexes containing only A and T bases. Umemoto *et al.* have reported NMR spectra for the triplex $d(T)_6 \cdot d(A)_6 \cdot d(T)_6$ (22). In that study, triplex formation was apparent from the difference in the imino proton NMR spectra in the presence of either 100 mM NaCl or 20 mM $MgCl_2$, at a total strand concentration of 1.5 mM. Temperature-dependent imino proton spectra of the triplex sample and that of a closely related duplex, $d(AT)_3$. Triplex formation was apparent in the presence of either 100 mM NaCl or 20 mM $MgCl_2$, at a total strand concentration of 1.5 mM. Temperature-dependent imino proton spectra in 100 mM NaCl indicated that this triplex melted, in a monophasic process, directly to single strands.

Our work in the area of DNA triplexes started with investigations of the analogous triplex formed by the 10-mer oligonucleotides $d(A)_{10}$ and $d(T)_{10}$. Based on earlier polynucleotide work (11), we anticipated that, in the presence of sufficiently high ionic strength and a 2:1 mole ratio of $d(T)_{10}$ to $d(A)_{10}$, we should be able to form the $d(T)_{10} \cdot d(A)_{10} \cdot d(T)_{10}$ triplex. This was confirmed both by carrying out UV mixing curves, at 260 nm, as illustrated in Fig. 4, where it is readily apparent that the minimum occurs at a 1:1 strand stoichiometry in 100 mM NaCl but at a 2:1 $d(T)_{10}$ to $d(A)_{10}$ strand ratio in 50 mM $MgCl_2$. Similar experiments carried out at 10-fold lower oligonucleotide concentrations revealed only duplex formation, indicating the lower stability of the triplex in comparison to the duplex. Several features are worth noting about this mixing curve. First, these minima correspond to the expected strand ratios for the duplex and the pyr*pur*pyr triplex, respectively. Second, in the presence of $MgCl_2$, the only clearly indicated complex formed is the $d(T)_{10} \cdot d(A)_{10} \cdot d(T)_{10}$ triplex, although certainly the duplex must form as well at the 1:1 strand ratio. This implies that, at certain wavelengths, some complexes are detectable and others not. This is important in that the absence of a slope change in such mixing curves does not necessarily mean the absence of the complex. Third, there is no sign of the corresponding pur*pur*pyr triplex, $d(A)_{10} \cdot d(A)_{10} \cdot d(T)_{10}$.

Although UV mixing curves are of general utility in determining relative complex stoichiometry, the technique does suffer from some limitations, as mentioned above. A perhaps more direct measure of complex formation is afforded by polyacrylamide gel electrophoresis (PAGE), which permits direct visualization of all species, either separately or mixed together to form multistranded complexes. This is illustrated in Fig. 5. Here, it is evident that in lane 3, containing a 1:1 mixture of $d(T)_{10}$ and $d(A)_{10}$, a single band is observed with lower mobility than is seen for the bands in lanes 1 and 2. This

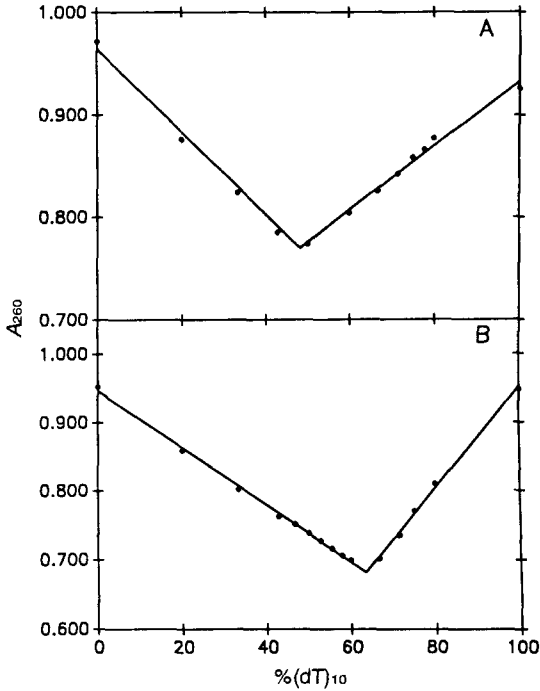


FIG. 4. UV mixing curves for the A_{10} - T_{10} system in either 100 mM NaCl (A) or 50 mM $MgCl_2$ (B). Reprinted with permission (23).

is expected for stoichiometric formation of a double helix. Similarly, a single, even more slowly migrating band is seen in lane 4 when these two strands are present in a 2:1 ratio, consistent with stoichiometric formation of a three-stranded structure. Note that although the UV mixing curve does not show duplex formation in the presence of $MgCl_2$, PAGE does. These two methods

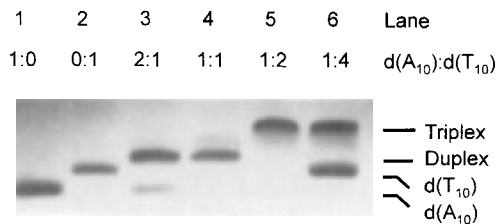


FIG. 5. Mixing experiments for the A_{10} - T_{10} system determined by PAGE.

have proved very useful for establishing the relative composition of higher order structures involving two or more species of non-self-complementary strands.

Although these types of experiments can provide information concerning the strand ratio in multistranded complexes, they provide little detail concerning the mode of interaction between the strands. In particular, the hydrogen bond pattern proposed for this pyr*pur*pyr triplex formed by $d(A)_{10}$ and $d(T)_{10}$ involves Hoogsteen-like hydrogen bonding between the third-strand thymine and the second-strand adenine. 1H NMR is a particularly powerful method for direct observation of protons shared between two heavier atoms in a hydrogen bond. If this bond is sufficiently strong, the rate of exchange of the proton with bulk water will be reduced enough to lead to an observable resonance if measured in H_2O . Observation of such resonances from Watson-Crick base pairs has long been routine.

Formation of Hoogsteen hydrogen bonding between the third strand and the duplex should lead to the appearance of a new set of imino protons, one for each triplet, arising from the T3NH proton that is hydrogen bonded to the A7N (see Fig. 2). This was investigated by measuring the low-field NMR spectra in H_2O of a 2:1 mixture of $d(T)_{10}$ and $d(A)_{10}$ as a function of increasing concentration of $MgCl_2$. At very low divalent salt concentrations, the system should consist of the $d(A)_{10} \cdot d(T)_{10}$ duplex and the unassociated single strand, $d(T)_{10}$. The latter is not expected to have any sharp imino proton resonances due to rapid exchange with water, whereas the former is expected to give rise to as many as 10 imino protons due to protection from solvent exchange as they participate in hydrogen bonding. Because of fraying at the ends, typically fewer than 10 resonances are seen for the duplex. As the concentration of $MgCl_2$ is increased, Fig. 6 shows the appearance of new imino proton resonances, as expected for the additional Hoogsteen base pairing resulting from triplex formation. Additional confirmation that these resonances arise from the proposed hydrogen bonding was obtained from nuclear Overhauser enhancement spectroscopy (NOESY) measurements (23).

One of the most common methods used to assess stability of nucleic acid structures is thermal denaturation, typically monitored by measuring UV absorbance as a function of temperature. Because nucleic acid systems usually undergo hypochromism (or a decrease in absorbance) on base pair formation, their UV absorbance increases with heating as base pairing is destroyed. When many base pairs are involved, the increase in UV absorbance occurs cooperatively, yielding a sigmoidal-shaped curve. When carried out reversibly, these "melting" curves can be analyzed in order to extract thermodynamic parameters, such as ΔH and ΔS , for the dissociation process (see below). This analysis is straightforward in the case of an all-or-none (two-state) transition, where the only possible structures are the fully formed complex

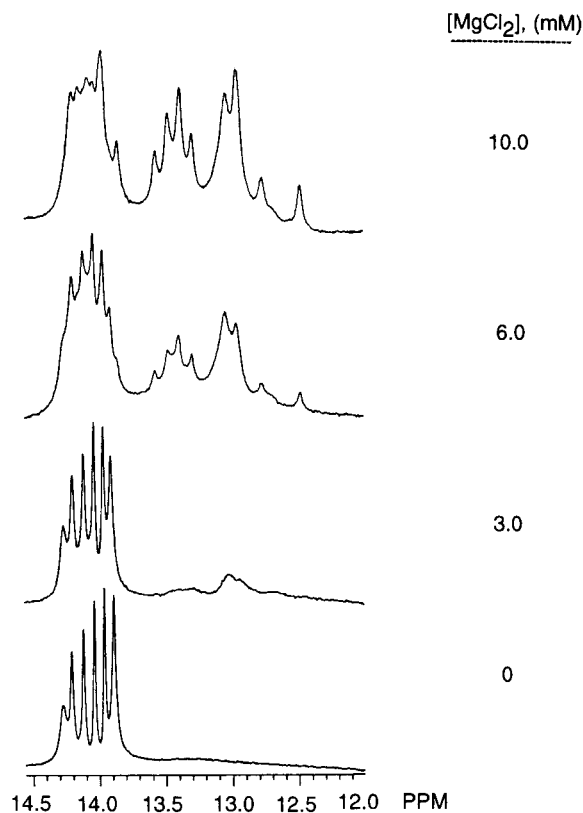


FIG. 6. Imino proton spectra of a $A_{10}:2T_{10}$ solution at the indicated $MgCl_2$ concentrations. Reprinted with permission (23).

and the completely dissociated single strands. Although duplexes show a melting profile consisting of a single transition, the presence of a third strand bound to a duplex in a triplex often gives rise to a biphasic melting profile. In this case, the low-temperature transition corresponds to loss of the third strand from the triplex, and the high-temperature transition involves dissociation of the remaining duplex. These two transitions can sometimes be completely resolved by measurements at different wavelengths such that only one transition is observable at one wavelength, the other being silent (23).

The presence of cytosine in $pyr \cdot pur \cdot pyr$ triplexes leads to the requirement for acid pH due to the necessity of protonating the N-3 of the third-strand cytosine in order to achieve Hoogsteen-like hydrogen bonding to the duplex guanine, as was shown in Fig. 2. This adds an additional environ-

mental variable in terms of conditions favoring triplex stability. The minimum pH required to form fully a particular cytosine-containing pyr*pur*pyr triple helix is a function of the oligonucleotide base composition and sequence along with the identity of monovalent, divalent, or higher valency salts. In general, however, it is significantly higher than the pK_a for cytosine residues in an unstructured single strand, which is 4.7 (24). This increase in effective pK_a suggests a close coupling between the protonation process and the structure-forming process. Furthermore, if cytosine is methylated at the 5-position, triplex formation occurs at considerably higher pH (12). This phenomenon is mainly due to the enhanced stability of the resulting triplex, because the pK_a of m^5dC is not much different from that of dC .

Some of the early physical chemical studies on pyr*pur*pyr triplexes containing cytosine involved NMR studies. Rajagapol and Feigon used proton NMR to investigate the triplex formed between $d(GA)_4$ and $d(TC)_4$ under acid conditions with the addition of $MgCl_2$ (25, 26). In the presence of excess pyrimidine strand, new imino protons could be observed, along with the expected NOE sequential connectivity. Due to the alternating sequence utilized, the underlying duplex was perfectly matched but binding of a second pyrimidine strand could occur in two ways, with either an overhang at the 5' end or at the 3' end of the purine strand. The latter appeared to be the major triplex conformer observed. Similarly, de los Santos *et al.* (27) reported NMR studies on both an 11-mer and a 7-mer intermolecular pyr*pur*pyr triplex containing cytosines designed to be maximally aligned in a major conformer, again demonstrating the presence of Hoogsteen hydrogen bonding between the purine strand of the duplex and the third strand. The presence of protonated cytosines in all of these studies could be definitively demonstrated through observation of their very low-field imino proton resonances (14.5–16 ppm).

Ts'o and co-workers examined a pair of somewhat longer oligonucleotides for their ability to form a pyr*pur*pyr triplex with a strictly alternating sequence, i.e., based on the $dA(GA)_7T$ oligonucleotide (28). Imino protons could be observed corresponding to Watson–Crick and Hoogsteen hydrogen bonding, although in this case resonances appeared in unresolved envelopes rather than resolved peaks for each proton, due to the longer sequence employed. Circular dichroism spectra were also useful in characterizing the triplex structure. Here again, the design of the sequence leads to the possibility of forming two triplexes with overhangs at either end of the purine strand. Interestingly, there was evidence for a conformational change from one form of triplex to another as a function of increasing temperature. Triplexes based on a hairpin duplex and separate third strand have also been investigated by spectrophotometric and gel electrophoretic techniques (24).

We initiated a series of studies of all three classes of triplexes based on

the duplex $d(G_3A_4C_3) \cdot d(C_3T_4C_3)$. This sequence was designed to be symmetric so that the 2:1 pyr*pur*pyr and pur*pur*pyr triplexes could be constructed from the two strands in the duplex with perfect alignment. This sequence turned out to be highly polymorphic, leading to a large number of different species, including, but not limited to, the three anticipated triplexes. Figure 7 illustrates the large number of structures that can form from these sequences under a variety of solution conditions.

Polyacrylamide gel electrophoresis (PAGE) experiments demonstrated the formation of the expected pyr*pur*pyr triplex at pH 5.5 in the presence of 50 mM $MgCl_2$ (29). This was confirmed with circular dichroism (CD), which is very sensitive to the geometry of nucleic acid conformations (30). Thus, in addition to measurements of UV absorbance spectra, UV CD spectra proved very helpful in discerning these new structures. This is a fairly general method for characterizing triplex structures, although there are reports of triplexes accompanied by very little change in CD as they form (31).

As in the case of the $d(T)_{10} \cdot d(A)_{10} \cdot d(T)_{10}$, formation of the $d(C_3^+T_4C_3^+) \cdot d(G_3A_4C_3) \cdot d(C_3T_4C_3)$ triplex should be accompanied by the appearance of additional imino protons protected from rapid exchange with bulk water due to their participation in Hoogsteen-like hydrogen bonds. These were indeed observed, including the very low-field peaks corresponding to protonated cy-

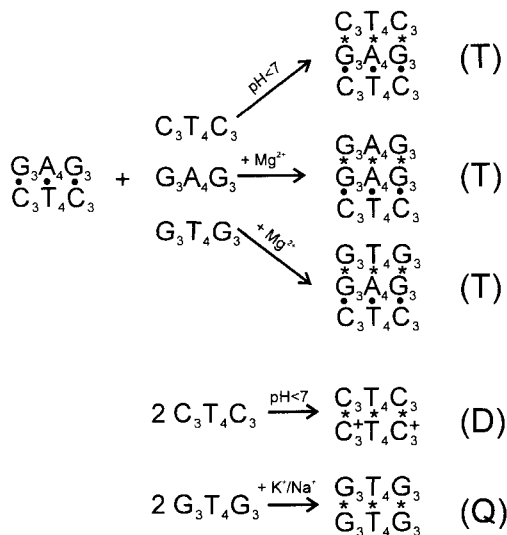


FIG. 7. Duplexes (D), triplexes (T), and quadruplexes (Q) based on the $d(C_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex. ●, Watson-Crick base pairs; *, Hoogsteen or other types of base pairs.

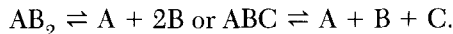
tosine. Again, as in the case of the $d(T)_{10} \cdot d(A)_{10} \cdot d(T)_{10}$ triplex, UV melting profiles were typically biphasic, with a low-temperature transition corresponding to dissociation of the third strand followed by a high-temperature transition relating to dissociation of the remaining duplex, although in the presence of 20 μM spermine, the third strand was sufficiently stabilized that a monophasic melting profile was observed (29).

There have been few systematic investigations of the sequence-dependent stability of $\text{pyr} \cdot \text{pur} \cdot \text{pyr}$ triplexes (32). Certain qualitative inferences can, however, be drawn. It appears that $\text{pyr} \cdot \text{pur} \cdot \text{pyr}$ triplexes containing arbitrary combinations of C and T in the third strand can be formed. When cytosine is present, pH becomes a major factor in determining stability and thus makes comparisons between different studies difficult. Furthermore, one can expect significant sequence effects on stability when protonated cytosines are present.

2. PUR*PUR*PYR TRIPLEXES

The second major class of triplexes involves those containing third strands composed of purines. This class of triplex does not involve base protonation and hence stable structures are formed at neutral pH. In contrast to $\text{pyr} \cdot \text{pur} \cdot \text{pyr}$ triplexes, which can often be formed in the presence of monovalent salts, $\text{pur} \cdot \text{pur} \cdot \text{pyr}$ triplexes usually require the presence of divalent or multivalent salts, such as Mg^{2+} , Mn^{2+} , Zn^{2+} , and/or spermine (33). In the case of a single-stranded, folded $\text{pur} \cdot \text{pur} \cdot \text{pyr}$ triplex, monovalent salts may suffice (34). One particularly unique property of $\text{pur} \cdot \text{pur} \cdot \text{pyr}$ triplexes is based on the observation of little or no UV hypochromism on binding of the third strand to the target duplex (34, 35). It is not known whether this behavior derives from properties of the purine strand alone, the base stacking and other interactions between the purine strand and the duplex, or both. This lack of hypochromism can make it difficult to detect triplex formation by UV mixing curves. In some cases, circular dichroism may be useful in demonstrating spectrophotometrically the binding of the purine third strand to the duplex.

Another interesting characteristic of $\text{pur} \cdot \text{pur} \cdot \text{pyr}$ triplexes studied to date is their propensity for melting in a single transition to separated strands, according to the equation



We have observed similar behavior with $\text{pyr} \cdot \text{pur} \cdot \text{pyr}$ triplexes under certain conditions, although under most conditions such triplexes exhibit biphasic melting behavior. In contrast, the $\text{pur} \cdot \text{pur} \cdot \text{pyr}$ triplexes we have investigated always denature in a monophasic fashion, typically at the same temperature as the underlying duplex. Malvy and co-workers have reported $\text{pur} \cdot \text{pur} \cdot \text{pyr}$

triplexes that melt at temperatures significantly higher than those corresponding to the underlying duplex (36, 37). Thus, in general, pur*pur*pyr triplexes possess thermal stability at least as great as the underlying duplex.

The $d(G_3A_4G_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex displays the above-mentioned characteristics of pur*pur*pyr triplexes (35). Both the duplex and the triplex melt at the same temperature. Mixing curves can be constructed to demonstrate formation of this triplex (29), and PAGE provides confirming evidence. CD also demonstrates the interaction of the second purine strand with the duplex. The anticipated Hoogsteen hydrogen bonding gives rise to additional imino proton resonances in the low-field region of the 1H NMR imino proton spectrum (35).

In contrast to the case of pyr*pur*pyr triplexes, there are substantial sequence effects on the stability of pur*pur*pyr triplexes. In particular, it is very difficult to form nucleic acid triplexes composed solely of A*A•T or A*A•U triplets, although several exceptions have been reported. Broitman *et al.* (38) have been able to form the RNA triplex poly(A)*poly(A)•poly(U), but only for poly(A) strands of a certain length. Pilch and Breslauer (39) have demonstrated formation of a hybrid poly(A)*poly(A)•poly(dT) triplex in the presence of berenil or 4',6-diamidino-2-phenylindole (DAPI). Howard *et al.* (40) have shown that, at high salt and moderate temperatures, one can form the barely stable poly(dA)*poly(dA)•poly(dT) triplex by adding the third strand to the duplex in an endothermic reaction. Another report suggests that certain porphyrins can induce the formation of intermolecular RNA triplexes based on A*A•U triplets alone (41) without constraints on polynucleotide length. The difficulty in forming such triplexes suggests that the A*A•T/U triplet is particularly unstable, for reasons unknown at this time. One obvious feature unique to the A*A Hoogsteen interaction is that it is based on amino protons alone, whereas all other known base pair hydrogen bonding schemes involve one imino proton.

3. PUR/PYR*PUR•PYR TRIPLEXES

Triplexes can also be formed with third strands composed of thymine and guanine. In one early study on a highly G-rich sequence, the corresponding G,T third-strand triplex was slightly more stable than the pur*pur*pyr (G,A third-strand) triplex. In general, however, pur/pyr*pur*pyr triplexes are less stable than the corresponding pur*pur*pyr triplex (42, 43). This most likely arises from the nonisomorphous nature of the T*A•T and G*G•C triplets, which leads to less favorable stacking interactions than in pyr*pur*pyr or pur*pur*pyr triplexes (20). Helene and co-workers have explored the question of strand orientation in pur/pyr*pur*pyr triplexes and concluded that the third strand is parallel to the duplex purine strand when it contains only a few GpT and TpG steps, whereas it is antiparallel in the presence of many

GpT and TpG steps (20). When the third strand is composed of G and T bases, there is an elevated probability it can form intra- or intermolecular complexes stabilized by guanine quartets (see below). Such structures can compete effectively with triplex formation and pose practical limitations to the design of this class of triplex.

D. Triplex Thermodynamics

I. UV MELTING

The stability of triplexes can be investigated by determining the thermodynamic parameters relevant to their formation, i.e., ΔH , ΔS , and ΔG . These functions can often be evaluated via analysis of UV melting curves. The thermodynamic analysis of single-transition melting curves characteristic of duplexes is well established (see Refs. 44 and 45 for reviews), but the biphasic nature of most triplex melting curves presents additional complications. There are two common measures of the melting temperature characterizing a transition induced by heat. The T_m is defined as the temperature at which the transition is 50% complete whereas T_{max} is the temperature corresponding to the maximum in the plot of the derivative

$$\partial\theta/\partial(1/T) = -(1/T^2) \partial\theta/\partial T$$

as a function of T (46). T_m involves conversion of the absorbance versus temperature melting curve to one in terms of θ versus temperature, where θ is the fraction of strands in the dissociated state (44, 45). This conversion process typically requires baseline data both at high and low temperatures. In contrast, T_{max} can be determined directly from the maximum in the absorbance data, assuming that the absorbance of the fully associated and dissociated states depends linearly on temperature.

UV melting curves for pyr*pur*pyr triplexes are frequently biphasic, with each transition maximum clearly defined, but only partially resolved in that baseline data in the region between the two transitions are incomplete. For this reason, we have applied the T_{max} method of analysis, treating each transition separately. Thus each transition is modeled by the reaction $AB \rightleftharpoons A + B$. For such a reaction, it can be readily demonstrated that T_{max} depends on oligonucleotide concentration according to the equation (30)

$$\frac{1}{T_{max}} = \frac{R}{\Delta H} \ln C_t + \frac{\Delta S + b}{\Delta H},$$

where b is a constant determined by the strand stoichiometry. This equation can be applied to each maximum in the derivative melting curve, corresponding to each of the two transitions, if they are sufficiently well resolved.

In addition to analysis based on the concentration dependence of T_m or

T_{\max} , thermodynamic parameters can be extracted from examination of the shape of the entire melting curve. We have found the approach implemented by Jovin and co-workers for duplexes particularly useful for monophasic transitions, because the same equation can be used to fit melting curves for self-complementary and non-self-complementary structures (47). Thus, for triplex structures corresponding to the formula ABC, AB₂, or A₃, where each letter represents a strand, the equilibrium constant for triplex dissociation can be represented by the expression

$$K = d \frac{f^3}{(1-f)} C_t^2,$$

where d is a statistical factor determined by the stoichiometry of the formation reaction relevant to each type of triplex, f is the fraction of strands in the dissociated, or random coil, state and C_t is the total strand concentration. By definition, $f = 1/2$ at $T = T_m$, which leads to $K(T_m) = (d/4)C_t^2$. Defining E_T by the equations

$$E_T = \frac{K(T_m)}{K(T)} = e^{\frac{\Delta H}{R} \left(\frac{1}{T} - \frac{1}{T_m} \right)}$$

where the second equation results from applying the van't Hoff equation for K , leads to the following equation for f as a function of T , applicable to all triplex forms:

$$f^3 + \left(\frac{1}{4E_T} \right) f - \frac{1}{4E_T} = 0.$$

The solution of this cubic equation for its one real, positive root can then be used to fit the observed dependence of UV absorbance from all species, A_t , on temperature according to the equation

$$A_t = A_h + (A_c - A_h)f,$$

where $A_h = b_h(1 + m_h T^C)$ is the absorbance of the helical form, assumed to be linear in temperature, expressed in Celsius units, and similarly for the coil form, $A_c = b_c(1 + m_c T^C)$. The constants b_h , b_c , m_h , and m_c are determined by baseline data; ΔH and T_m are determined by data in the transition region. Typically, melting curves should cover a wide range of oligonucleotide concentrations. We have applied this method to both the d[(GA)₅G]•d[(CT)₅C] duplex and the d[(GA)₅G]*d[(GA)₅G]•d[(CT)₅C] triplex (42).

Both methods of analyzing UV melting curves are based on the van't Hoff equation describing the temperature dependence of the equilibrium constant and rely on several assumptions. First is the assumption that the melting occurs reversibly, so that the triplex is in equilibrium with its separated com-

ponents at all temperatures throughout the transition. The best experimental verification of reversibility can be achieved by comparison of heating and cooling curves carried out at identical heating/cooling rates. Reversibility is manifested by superimposability of the heating and cooling curves. Triplexes exhibit slow kinetics that, under certain conditions, can lead to substantial irreversibility in UV melting experiments (48). The second assumption underlying analysis of UV melting curves for thermodynamic parameters is that the transitions studied exhibit all-or-none behavior. This means that there are only two states for the system, complete triplex structure or completely separated strands, implying a very high degree of cooperativity. States consisting of partially melted structures are not allowed according to this model. This assumption usually holds for short oligonucleotide structures. There are methods for dealing with systems that do not satisfy the all-or-none approximation, but they involve assumptions regarding the nature of the intermediate states that occur in the transition, and these assumptions are difficult to confirm (46, 47).

One method of examining the extent to which a given system exhibits all-or-none behavior is by comparing results for ΔH determined by the two methods described above. The effectiveness of this approach is based on the observation that fitting the shape of the entire melting curve can be more sensitive to the extent of cooperativity, compared to the method based on the variation of T_{\max} (or T_m) with concentration. Thus, the breadth of the transition may change with extent of cooperativity while T_{\max} remains relatively constant. Typically, when ΔH values measured by these two methods disagree, those determined by shape analysis are smaller, which can be interpreted as a manifestation of a significant population of intermediate states. To the extent that it is less sensitive to this effect, the method based on the concentration dependence of T_{\max} (or T_m) may be considered to provide more accurate results. The optimal method, however, for assessing the all-or-none aspect of oligonucleotide melting processes involves an independent, model-free approach to determining enthalpy changes.

2. CALORIMETRY

Differential scanning calorimetry (DSC) affords an alternative method for determining the thermodynamics of triplex dissociation with the significant advantage that it provides a direct measure of the heat required to disrupt the oligonucleotide complex and hence yields model-independent results. Thus accurate results can be obtained regardless of the applicability of the all-or-none approximation to the process being measured. The application of DSC to nucleic acid conformational transitions has been reviewed extensively (49), and will be described only briefly here. In a DSC experiment, both a sample and a reference cell are heated at a constant rate. In the reference

cell, containing buffer only, the amount of heat required to raise its temperature is determined by the heat capacity of the buffer, usually a very smooth, slowly changing function of temperature. In the sample cell, however, additional heat must be added to raise its temperature if its contents undergo a conformational transition accompanied by a change in enthalpy, ΔH . Thus the heat capacity of the sample cell includes, in addition to a term related to the heat capacity of the buffer, a term varying sharply through the transition, reflecting the heat required to induce the transition, or “melt” the structure. The heat capacity ($\partial H/\partial t$) is determined by multiplying the power ($\partial H/\partial t$), where t represents time, by the inverse of the heating rate (dt/dT). The difference between the power required to heat the sample and reference cells is then converted to the excess heat capacity, reflecting the additional heat associated with the transition under study. The associated model-independent enthalpy change, ΔH_{CAL} , then, is calculated via

$$\Delta H_{\text{CAL}} = \int C_{\text{p}}(\text{excess}) dT.$$

In this way, ΔH can be determined by numerically computing the area under the excess heat capacity versus temperature curve without making any assumptions concerning the nature of the measured transition. At the same time, the accompanying change in entropy can be determined by dividing the excess heat capacity by the (absolute) temperature and computing the area under the $C_{\text{p}}(\text{excess})/T$ versus T curve:

$$\Delta S = \int C_{\text{p}}(\text{excess})/T dT.$$

Because entropy involves measurement of the reversible heat, the heating process must be carried out reversibly for measuring ΔS .

Just as in the analysis of UV melting curves, one may model the transition measured by DSC and thereby determine a van't Hoff enthalpy change, $\Delta H_{\text{VH}(\text{cal})}$. This involves the same equilibrium analysis as presented above in the discussion of UV melting curves. For a single transition, we can write

$$H = H_0 + f\Delta H,$$

where f represents the fraction of strands dissociated, H_0 is the enthalpy of the reference, low-temperature state, and ΔH is the enthalpy change for the transition. This then, leads to the equation

$$C_{\text{p}}(\text{excess}) = \frac{\partial(H - H_0)}{\partial T} = \Delta H \frac{\partial f}{\partial T} + f\Delta C_{\text{p}}.$$

In many cases, the last term in the equation can be neglected because $\Delta C_{\text{p}} \approx 0$. Usually the van't Hoff enthalpy change, $\Delta H_{\text{VH}(\text{cal})}$, is used both in the equation above as well as in the equation for determining f . In this way,

$\Delta H_{\text{VH(cal)}}$ can be determined as that value giving the best fit to the experimental excess heat capacity curve, usually by nonlinear curve fitting methods. Alternative measures of $\Delta H_{\text{VH(cal)}}$ have also been discussed (49).

Manzini *et al.* (24) reported UV melting studies on a two-stranded pyr*pur•pyr triplex composed of a hairpin duplex and a single strand. They observed biphasic melting profiles (pH 5) and fit the entire melting curve to a model involving a two-step process: triplex \rightleftharpoons hairpin + third strand \rightleftharpoons coil + third strand. They obtained a mean ΔH per Hoogsteen base pair of 6.6 kcal/mol (discounting the estimated contribution from cytosine protonation). Subsequently, Xodo *et al.* (50) carried out similar studies on another two-stranded triplex, in which the hairpin strand was all pyrimidine. Here they found monophasic melting behavior (pH 5), and carried out both UV melting and calorimetric studies, obtaining fairly good agreement between the thermodynamic parameters determined by these two methods. After accounting for protonation of third-strand cytosines, they estimated ΔH as 5.8 kcal/mol Hoogsteen base pair.

Plum *et al.* (31) reported thermodynamic studies on an intermolecular triplex composed of a 15-mer pyrimidine strand binding to a 21-mer duplex. Calorimetric analysis revealed two well-separated transitions, and, interestingly, the calorimetrically determined ΔH (2 kcal/mol triplet at pH 6.5, 200 mM NaCl) was substantially smaller than that determined from van't Hoff analysis. Similar behavior was observed in thermodynamic analysis of an intramolecular pyr*pur•pyr triplex carried out at pH 6.7 but good agreement between the calorimetric and van't Hoff values for ΔH were obtained at pH 4.5 (51). Subsequently, Plum and Breslauer (52) investigated the thermodynamics of a different intramolecular pyr*pur•pyr triplex, again finding, at pH 6.5, van't Hoff ΔH values substantially larger than those determined directly by calorimetry. Thus, to date, the pyr*pur•pyr triplex systems appear to show larger ΔH_{VH} values than ΔH_{CAL} values when determined under conditions of marginal stability, i.e., pH close to neutral. It is clear that pH strongly affects stability in pyr*pur•pyr triplexes containing cytosine, and hence values for ΔH will depend on the pH at which it is measured. Wilson and co-workers, however, have shown that the situation is even more complicated in that buffer effects also contribute significantly to the measured thermodynamics (53). Breslauer and co-workers discuss these and other results in detail (9).

We have carried out calorimetric analysis of several triplex structures, targeted to the d(G₃A₄G₃)•d(C₃T₄C₃) and dG(AG)₅•dC(TC)₅ duplexes (42, 43). Examples of heat capacity curves of the two pur*pur•pyr triplexes are illustrated in Fig. 8, with results for the accompanying changes in thermodynamic parameters summarized in Table I. As can be seen for these two pur*pur•pyr triplexes, the agreement is excellent between the model-free thermodynamic parameters and those determined by van't Hoff analysis of

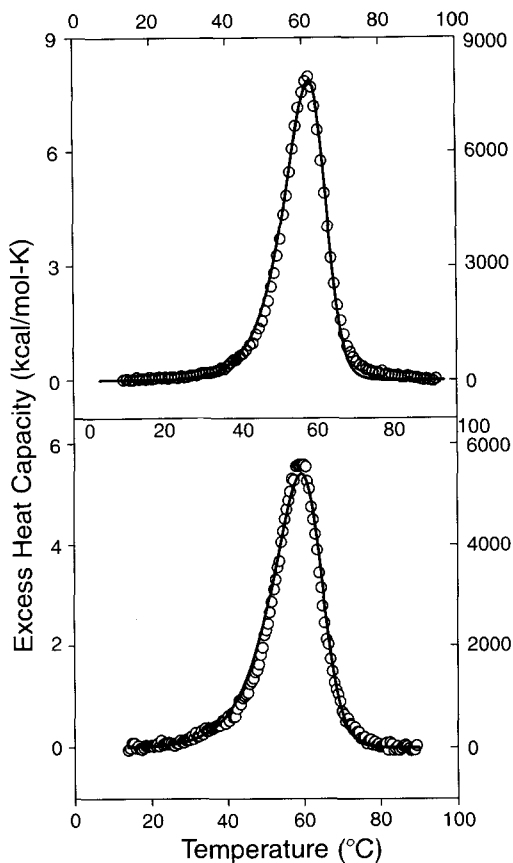


FIG. 8. Calorimetric heat capacity curves for the $d(G_3A_4G_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ (top) and $d[G(AG)_5] \cdot d[G(AG)_5] \cdot d[C(TC)_5]$ (bottom) pur*pur*pyr triplexes. Smooth lines represent the best fit from a van't Hoff analysis of the calorimetric data.

TABLE I
THERMODYNAMIC PARAMETERS FOR THE $d(G_3A_4G_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ (I)
AND $d[G(AG)_5] \cdot d[G(AG)_5] \cdot d[C(TC)_5]$ (II) TRIPLEXES

| Triplex | ΔH (kcal/mol) | | ΔS (cal/mol-K) | | ΔG (298) (kcal/mol) |
|-----------------|-----------------------|----------------------|------------------------|----------------------|--------------------------------|
| | ΔH_{cal} | $\Delta H_{VH(cal)}$ | ΔS_{cal} | $\Delta S_{VH(cal)}$ | |
| I ^a | 116 ± 1 | 115 ± 1 | 350 ± 4 | 309 ± 1 | 22.9 ± 0.3 |
| II ^b | 92.0 ± 2 | 93.9 ± 2 | 279 ± 6 | 246 ± 5 | 21.9 ± 0.6 |

^aSee Ref. 43 for details.

^bSee Ref. 42 for details.

the calorimetric data. Furthermore, when modeled as two transitions, the best fit to experimental data for these triplexes was obtained when the T_m values of the two transitions became identical. Thus we can be confident that dissociation of these triplexes takes place in a single, all-or-none process. Furthermore, the transition does not entail a significant change in heat capacity, ΔC_p , in accord with most other studies on short triplexes.

In terms of quantitative results, we observed a significant sequence dependence in the ΔH for triplex formation. For example, in the $d(G_3A_4G_3) \cdot d(C_3T_4C_3) \cdot d(C_3T_4C_3)$ triplex, we measured a ΔH of 11.6 kcal/mol triplet formed (43), whereas in the $d[G(AG)_5] \cdot d[G(AG)_5] \cdot d[C(TC)_5]$ triplex, $\Delta H = 8.4$ kcal/mol triplet (42). Thus despite its increased length, the overall enthalpy change is smaller for the alternating sequence than for the block sequence. A similar, but smaller difference was observed for the two underlying duplexes, implying that this sequence-dependent phenomenon is enhanced in the triplex relative to the duplex.

In our calorimetric studies on triplexes targeted to the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex, not only was the pur/pyr*pur*pyr triplex $d(G_3T_4G_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ the least enthalpically stable triplex, but also we measured a van't Hoff ΔH that was about 35% greater than the calorimetric ΔH . We postulated the possibility of end-to-end association as an explanation for this behavior. This interpretation was supported by the excellent fit to data obtained for a model that explicitly included such effects (43). The corresponding pur/pyr*pur*pyr triplex based on the $d[G(AG)_5] \cdot d[C(TC)_5]$ duplex was so marginally stable that DSC was not possible (42). In a qualitative sense, then, the sequence effect on stability described above for triplexes based on $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ is similar for triplexes targeted to the $d[G(AG)_5] \cdot d[C(TC)_5]$ duplex, in that the alternating pur/pyr*pur*pyr triplex is clearly less stable than the equivalent block sequence triplex. Although these results hint at important sequence-dependent effects in triplex formation, additional work is needed before a general picture of the thermodynamic behavior of triple helical structures emerges.

E. High-Resolution Structural Aspects of Triplexes

Early molecular models of the DNA triplex emerged from fiber diffraction studies on poly(dT)*poly(dA)*poly(dT) (54), and described the overall structure as A-like, with all sugars in the C-3'-endo pucker and a substantial displacement of the bases from the helical axis. Since then, the picture has changed considerably, based on NMR studies of triplexes in solution. X-Ray structure determination for a DNA triplex has not yet been achieved due to the difficulty in obtaining diffraction-quality crystals for such systems. A very restricted number of studies has appeared concerning NMR analysis of intermolecular (i.e., three-stranded) triplexes. The principal limitation in this

regard is unusually large NMR line widths, most likely due to the tendency for these systems to stack in an end-to-end fashion. Nonetheless, some useful information has appeared based on such studies. As mentioned earlier, Rajagopal and Feigon (25, 26) carried out NMR studies on the pyr*pur*pyr triplex formed from $d(GA)_4$ and $(CT)_4$ at pH 5.5. Assignments in the triplex were facilitated by finding conditions in which the system exhibited NOESY exchange cross-peaks between the triplex and the duplex and single strand. Thus, proton assignments in the triplex structure could be inferred from those known for the duplex without making assumptions concerning the DNA conformation in the former. In addition to mapping out the expected sequential connectivities for the Hoogsteen imino protons, they reported that the sugars in both pyrimidine strands adopted the N conformation (C-3'-endo) whereas the purine strand maintained sugars closer to the S (C-2'-endo) conformation.

A similar result was observed for the short, $d(T)_6 \cdot d(A)_6 \cdot d(T)_6$ triplex in that the sugars in both pyrimidine strands were determined to be in the C-3'-endo (N) conformation, based on a combination of NOESY and MINSY spectra triplex (22). Peak overlap prevented comparable measurements for the purine strand.

de los Santos *et al.* (27) carried out two-dimensional NMR studies on a short 11-mer pyr*pur*pyr triplex. Again, direct evidence for imino protons participating in Hoogsteen hydrogen bonding was reported. In contrast to the results described by Rajagopal and Feigon (25, 26), they described the purine strand as taking on an A-like conformation, based mainly on upfield chemical shifts for purine H-8 resonances. Similarly, differences in imino proton peaks arising from Watson-Crick base pairs in duplex and triplex samples were interpreted in terms of different ring current effects in the B and A form helices, respectively.

In contrast to these NMR results, Howard *et al.* (55) carried out Fourier transform infrared (FTIR) studies in D_2O on the $d(T)_n \cdot d(A)_n \cdot d(T)_n$ triplex and found that all nucleosides were in the C-2'-endo conformation. Modeling calculations established the feasibility that such a structure can still satisfy the helical pitch and number of residues per helical turn measured by fiber diffraction.

In order to circumvent problems arising from large line widths characteristic of intermolecular triplexes, several researchers have turned to unimolecular, folded triplexes (see Fig. 3). This construct leads to several advantages over intermolecular triplexes. First, less extreme solution conditions, such as the concentration of divalent salts, are needed to stabilize the triplex conformation fully. Second, the presence of loops extending beyond each end of the triplex stem should disrupt end-to-end aggregation. A significant number of high-resolution NMR studies of this nature has been carried out successfully, and summarized in a review (56).

Results from these studies establish the details concerning Hoogsteen hydrogen bonding, base stacking interactions, and sugar pucker. In general, these triplexes show little evidence for A-like conformations in that the sugar rings are mainly in the C-2'-endo configuration, characteristic of the B-DNA conformation. Results from studies on both pur•pur•pyr and pur•pur•pyr unimolecular triplexes indicate a surprising degree of similarity in overall structure. Analysis of triplexes containing G and T in the third strand provide direct evidence for the structural perturbations arising from the nonisomorphic nature of the G•G•C and T•A•T triplets. The reason for the apparent discrepancy in sugar conformations observed for the three-stranded and single-stranded triplexes has been ascribed to the difficulties in obtaining high-resolution, quantitative results on the former due to large line widths (57).

An alternative method for mitigating the tendency for intermolecular triplexes to stack end-to-end is based on designing such a triplex with one strand longer than the other. We took this “base overhang” approach in designing a pur/pyr•pur•pyr triplex and measured its imino proton NMR spectrum in comparison to that of the “blunt end” reference structure. The resulting decrease in line widths was dramatic (58). This strongly suggests that end-to-end stacking is indeed a major source of line broadening observed for short intermolecular triplexes. Constructing triplexes according to this design may lead to high-resolution structure determination for this type of triplex.

F. Ligand–Triplex Interactions

There are two major classes of small ligands that bind to double helical DNA—intercalators and minor groove binding agents. Because the third strand in triplexes lies in the major groove of the underlying duplex, both of these sites are potentially available for occupancy by ligands. In general, most, if not all, minor groove binding ligands that have been investigated interact less strongly with triplexes than with duplexes. For example, netropsin has been shown to bind to triplexes, but less strongly than to the underlying duplex, and this results in a lowering of the T_m for dissociation of the triplex third strand (59, 60). The closely related minor groove binder distamycin also lowers the thermal stability of pyr•pur•pyr triplexes (61). Similarly, the AT-specific minor groove binding agent Hoechst 33258 binds to, but destabilizes, the triplex (62). One exception involves the minor groove binding ligand berenil, which preferentially binds to triplexes in low-salt conditions but destabilizes the triplex form in high concentrations of salt (63). Another minor groove binding ligand, chromomycin, was shown to bind poorly to a folded triplex composed solely of G•G•C triplets (34). The rest of this discussion will focus in the interaction of intercalators with triplex structures.

One of the earliest reports of an intercalator binding to triple helices in-

volved the interaction of ethidium with the poly(dA)•2poly(dT) triplex (64). In this work, the enhanced fluorescence of ethidium in the presence of this DNA polynucleotide triplex was reported, although detailed studies on the mechanism of binding were not described. Subsequently, reports on the inability of ethidium to bind to triplexes such as poly(dAG)•2poly(dCT), involving protonated cytosines, appeared (12). We surmised that the apparently poor interaction of ethidium with pyr•pur•pyr triplexes containing cytosine in the third strand derived from the presence of a regular array of positive charges from the protonated cytosines. Thus we investigated in detail the interaction of ethidium with the poly(dA)•2poly(dT) triplex in terms of binding affinity and binding mechanism (65). We observed that, in contrast to the positive cooperativity characteristic of ethidium binding to the poly(dA)•poly(dT) duplex, the binding of ethidium to the poly(dA)•2poly(dT) triplex exhibited the more usual negative cooperativity typical of most intercalator–DNA complexes. More significantly, the intrinsic binding constant for the triplex was about 35 times greater than that for the duplex, thus demonstrating that ethidium not only binds to this triplex, but does so more strongly than to the underlying duplex. In agreement with this outcome, we observed a substantial increase in the thermal stability of the third strand, based on ethidium-induced increases in T_m for the well-resolved transition corresponding to the dissociation of the third strand from the underlying poly(dA)•poly(dT) duplex. This study represents the first demonstration of a ligand capable of preferential binding to triplexes relative to duplexes.

Not only can ethidium bind to polynucleotide triplexes, it can also bind to oligonucleotide triplexes. UV melting studies on the interaction of ethidium with the $d(G_3A_4G_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex, at neutral pH, showed a slightly greater increase in T_m than that observed for the underlying duplex, $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ (29). Due to the observation that such triplexes melt in a single transition, it is not possible to show preferential binding to the triplex form over the duplex form in the same UV melting experiment.

It is possible, however, to demonstrate preferential binding of ethidium to short pur/pyr•pur•pyr triplexes by UV melting experiments, because these triplexes typically display biphasic melting behavior. An example of binding to such a triplex is exemplified by the interaction of ethidium with the $d[G(TG)_5] \cdot d[G(AG)_5] \cdot d[C(TC)_5]$. In this case, we have been able to observe an increase in the thermal stability of the third strand as a function of increasing ethidium concentration. This behavior is consistent with our recently published results showing triplex stabilization by DNase I digestion studies (42). Thus, there is evidence indicating that ethidium binds to and stabilizes all three classes of triplexes, with the single exception of pyr•pur•pyr triplexes containing protonated cytosines.

In addition to ethidium, other intercalators have been examined. We have observed that daunomycin can also stabilize short triplexes, but not as well as ethidium (Y. He, P. V. Scaria, and R. H. Shafer, unpublished results). This may possibly be due to the unusual geometry of daunomycin intercalation into duplexes, involving the long axis of the chromophore lying almost perpendicular to the long axis of the base pairs. Coralyne, a planar antitumor antibiotic consisting of four fused aromatic rings, has been shown to stabilize poly(dA)•2poly(dT) preferentially as well as pyr•pur•pyr polynucleotide triplexes containing protonated cytosine (66). A report on 9-aminoacridine and a bis-acridine compound indicates that both can thermally stabilize poly(dA)•2poly(dT) (67). Interestingly, a report on propidium and a bis-ethidium shows that the former can thermally stabilize poly(dA)•2poly(dT), whereas the latter has no effect on the melting temperature corresponding to loss of the third strand (68). This latter result suggests that the favorable binding free energy for bis-ethidium is compensated by an unfavorable conformational strain.

Helene and co-workers have addressed the question of ligand binding to DNA triplexes by designing and synthesizing a benzo[e]pyridoindole derivative with optimized geometry for intercalation between base triplets (69–71). This compound indeed exhibits preferential binding to triplexes compared to duplexes. In addition, Wilson and co-workers have also synthesized a series of quinoline derivatives with alkylamine side chains and aryl substituents designed to bind to DNA triplexes preferentially (72, 73). These compounds bind as intercalators and some show substantial preferential triplex stability against thermal denaturation.

Although it is clear that certain intercalators can exhibit significant preferential binding to DNA triplexes in comparison to DNA duplexes, it is unlikely that this class of DNA binding agents will yield a ligand truly specific for triplex structures. That is, there are no known triplex intercalators that do not exhibit significant levels of duplex intercalation. The most likely avenue leading toward specific recognition of triplexes with little or no duplex binding is to target one of the two new grooves created by the third-strand binding in the duplex major groove.

II. Guanine Quadruplex Structures

A. Background

Guanine has long been known to aggregate in solution, in the nucleoside, oligonucleotide, and polynucleotide forms (6). The molecular basis for this behavior is rooted in the propensity of four guanine residues to associate in a square planar configuration called the guanine quartet (see Fig. 9). In such

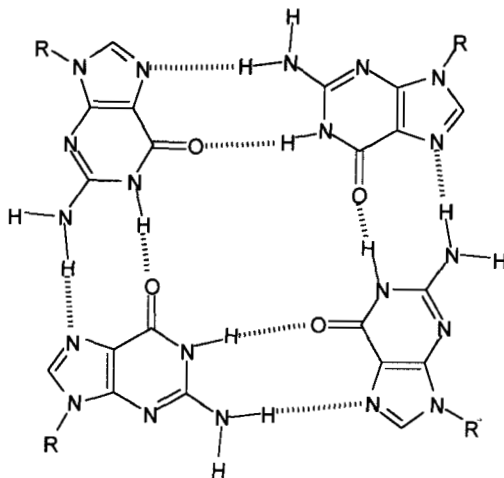


FIG. 9. Structure of the guanine quartet.

an arrangement, each guanine interacts with its two neighbors via Hoogsteen-like hydrogen bonding, resulting in a very tightly bound quartet. Seen as an extension of the notion of a base pair and a base triplet, it is obvious that this base quartet is a highly effective stacking unit with considerable surface area. Quadruplex structures are composed of stacks of guanine quartets, which can be realized by a variety of different strand configurations. Perhaps the simplest arises from four different strands, each containing a run of guanines. These strands can come together, each contributing one guanine to each quartet formed. Alternatively, two hairpin-shaped strands, each containing two runs of guanine separated by a loop consisting of several thymines, for example, can associate to form a hairpin dimer quadruplex. Finally, a single oligonucleotide containing four runs of guanine separated by three loop regions can fold into a unimolecular quadruplex. These different configurations are illustrated in Fig. 10.

Perhaps the most likely place to find guanine quadruplex structures in nature is the telomere, or terminal region of linear chromosomes. These are composed of many repeats of the type $G_n T_m$, sometimes containing an adenine, A, adjacent to guanine (74, 75). This sequence is ideally suited to form either hairpin dimer or unimolecular quadruplexes. The former may serve as a mechanism for chromosome association and the latter may provide a means for capping the ends of chromosomes with a nuclease-resistant structure. Other potential biological roles for this unusual DNA conformation have been discussed, such as facilitating homologous recombination (76). Proteins

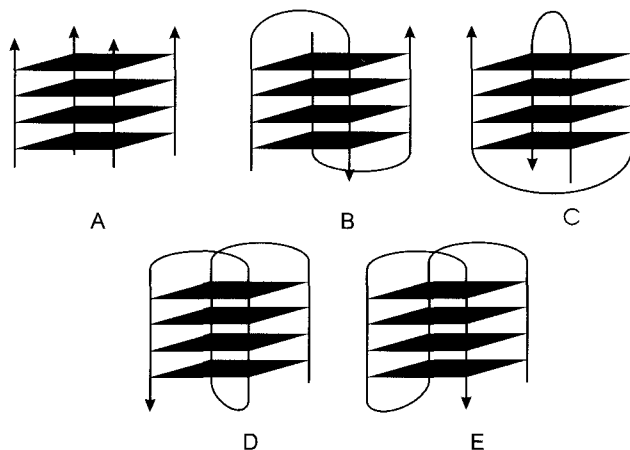


FIG. 10. Various forms of guanine quadruplexes composed of four guanine quartets (shaded rectangles). (A) Linear, four stranded; (B) edge-loop dimeric hairpin; (C) diagonal-loop dimeric hairpin; (D) unimolecular with diagonal central loop; (E) unimolecular with edge-type central loop.

exhibiting specific recognition of quadruplex structures have also been described (77, 78).

B. Basic Properties of Quadruplexes

Guanine quartet formation is strongly and specifically cation dependent. Sodium and potassium ions are among the most effective at stabilizing quadruplex structures. Because these cations are present in the cellular milieu, *in vivo* conditions are generally favorable for the existence of these structures. The ability of various cations to stabilize this conformation has been related to their ionic radius, based on a model of cation placement between adjacent guanine quartets and coordination to the guanine carbonyl oxygens (6). Some cations preferentially stabilize linear, four-stranded quadruplexes and others can preferentially stabilize folded forms in appropriate sequences. In the linear quadruplex configuration, all deoxyguanosines are usually in the anti conformation about the glycosidic bond. On the other hand, in dimeric hairpin and unimolecular quadruplexes, the deoxyguanosine nucleosides alternate between syn and anti conformations about the glycosidic bond, due to the requirements for quartet formation when all strand segments are not all parallel to each other. Furthermore, in the case of the folded quadruplexes, i.e., one- and two-stranded quadruplexes, there are several ways the central loop can bridge from one string of Gs to the other (shown in Fig. 10).

Stabilization of triplexes typically requires high concentrations of monovalent cations, due to the electrostatic cost of bringing a third negatively charged strand in close proximity to a duplex. In contrast, quadruplex structures, entailing four closely spaced strands, can be stabilized in the presence of much lower concentrations of potassium or sodium, e.g., 50–150 mM. This observation, along with that regarding the different effects of various cations, points to a specific interaction of cations with the quadruplex structure. Depending on the particular sequence, several different quadruplex forms can coexist in an equilibrium mixture, and there have been indications that different methods of preparation can lead to different populations of the various species (79). Several reports have described the very slow kinetics of quadruplex formation, accompanied by a negative energy of activation (80, 81). The slow kinetics can lead to practical difficulties in reaching true equilibrium conditions. One interesting aspect of quadruplex stability is revealed in the observation that the time scale for exchange of certain imino protons, within the guanine quartet, with bulk water can extend to a period of days or longer (82).

C. Spectroscopic Characterization of Quadruplexes

One of the early studies of structures formed by telomeric sequences demonstrated the presence of imino protons in a frequency range distinct from that observed for similar protons in duplexes or triplexes (83). More detailed studies using gel electrophoresis and circular dichroism demonstrated differential effects of sodium and potassium on the stability of intramolecular versus intermolecular quadruplex structures (79). Folded structures possess a positive CD peak near 295 nm whereas linear, four-stranded structures characteristically possess a positive CD maximum near 265 nm. These CD spectral differences have been utilized to show specific cation effects, e.g., Sr^{2+} preferentially stabilizes linear, four-stranded quadruplexes (84).

Although it was anticipated that half of the deoxyguanosines had to be in the syn conformation in the case of quadruplexes containing antiparallel strands, the study of Wang *et al.* (85) demonstrated that this occurs with sequential alternation, within each string of Gs, between the syn and anti conformations. Subsequently, Feigon and co-workers (82, 86) determined a high-resolution structure for the dimeric hairpin quadruplex constructed from two strands of $d(\text{G}_4\text{T}_4\text{G}_4)$. In addition to observing the syn–anti alternation, they showed that the thymine loop crossed diagonally over the terminal quartets, as illustrated in Fig. 10.

We first became interested in guanine quartet-based structures during our initial work on pur/pyr•pur•pyr triplexes, using $d(\text{C}_3\text{T}_4\text{C}_3)\cdot d(\text{G}_3\text{A}_4\text{G}_3)$ as the target duplex. As part of our studies to characterize the corresponding third strand, $d(\text{G}_3\text{T}_4\text{G}_3)$, we observed that it exhibited a relatively sharp con-

formational change on heating in sodium-containing buffer. In our triplex studies, this competing structure could be eliminated by avoiding buffers containing sodium or potassium, two cations that are strong inducers of the quadruplex structure. Tris was such a buffer, and so we examined the triplex structure by adding MgCl_2 to this buffer.

To investigate the self-structure possessed by the $d(\text{G}_3\text{T}_4\text{G}_3)$ strand, we simply added NaCl or KCl to the $d(\text{G}_3\text{T}_4\text{G}_3)$ oligonucleotide in Tris buffer. Under these buffer conditions, we could generate a cooperative UV melting transition. We assumed that there must be some hydrogen bonding underlying the structural transition and thus measured the imino proton spectrum of $d(\text{G}_3\text{T}_4\text{G}_3)$ in Tris alone and in Tris after addition of either NaCl or KCl. As expected, no peaks were observed in Tris alone, but in the presence of either added salt, 12 sharp imino protons were observed (87). Interestingly, the spectra for the two salts were significantly different in terms of chemical shifts. The frequencies of these chemical shifts were different from those characteristic of both duplexes and triplexes. Thus it was apparent that this was an unusual DNA structure.

From the NMR spectrum, it was not possible to ascertain the molecularity of the quadruplex structure, that is, the number of $d(\text{G}_3\text{T}_4\text{G}_3)$ strands complexed together. Although the sequence cannot form a unimolecular quadruplex, it can potentially form both a linear, four-stranded quadruplex and a hairpin dimer quadruplex. In order to resolve this question, we turned to equilibrium ultracentrifugation, as this is one of the few methods capable of determining absolute masses without the necessity of calibration against known standards. Briefly, in such an experiment, molecules sediment toward the bottom of the centrifuge tube, thereby generating a concentration gradient that, in turn, produces a diffusive flow back toward the top of the tube. Equilibrium is reached when these two opposing forces balance each other exactly, producing a time-independent concentration gradient throughout the centrifuge tube, determined by the buoyant mass of the molecule. For a single species, this concentration profile, $C(r)$, takes on the following form, as a function of r , the distance from the center of rotation:

$$C(r) = C(r_0) \exp \frac{M(1 - \bar{v}\rho)\omega^2(r^2 - r_0^2)}{RT},$$

where M is the solute molecular weight, \bar{v} is the solute partial specific volume, ω is the rotor speed, and ρ is the solution density.

Results from sedimentation analysis of $d(\text{G}_3\text{T}_4\text{G}_3)$ were quite conclusive regarding the number of strands complexed together (87). Samples were run in buffer alone and in the presence of KCl, yielding very different concentration profiles, with the steeper gradient occurring in the presence of KCl. Quantitative analysis of these curves, in terms of the above equation, pro-

duced a buoyant molecular mass, $M(1 - \bar{v}\rho)$, in the presence of KCl almost exactly twice that obtained in Tris alone. Assuming negligible differences in the buoyant density in the two buffer conditions, this implies a dimer or two-stranded structure for the quadruplex. Although this is one of the few applications of equilibrium sedimentation to the analysis of oligonucleotide complexes, its potential utility is clear.

The circular dichroism (CD) of quadruplexes can also provide a particularly helpful tool for identifying such structures, especially in the case of folded quadruplexes. Here, the CD spectrum is quite distinct in that it possesses a strong, positive peak centered near 290–295 nm. Significant peaks in this fairly long-wavelength region are not usually seen in duplexes or triplexes. However, structures containing the C⁺•C base pair also show a similar peak (88). Thus, at least in sequences lacking cytosine, such as d(G₃T₄G₃), the presence of a positive CD signal at 290 nm is very strong evidence for a folded quadruplex structure, in this case a dimeric hairpin quadruplex.

There were significant spectroscopic differences between d(G₃T₄G₃) in the presence of KCl and in the presence of NaCl, but another important difference involved the fact that, after incubation over a period of days or weeks in the presence of KCl, changes were observed in the CD spectrum. In particular, the peak at 290 nm decreased and a peak centered around 260 nm appeared. This was accompanied by a substantial broadening, but not loss, of the imino proton NMR resonances. We concluded that these changes were a manifestation of the presence of a four-stranded, linear quadruplex structure. CD spectra of samples incubated for 6 months in KCl showed both 260- and 290-nm peaks, representing an equilibrium mixture of linear and folded quadruplex forms. In contrast, no such indications of linear quadruplex structure were seen for d(G₃T₄G₃) in the presence of NaCl.

These cation differences point to the role played by metals in stabilizing this guanine quartet-based DNA structure. Early models proposed a structure in which the metal cation was sandwiched between two guanine quartets, octahedrally coordinated to the four guanine carbonyl oxygens above and four below. This picture provided a simple interpretation of quadruplex-inducing potential for a given cation in terms of its ionic radius. In the first X-ray structure of a quadruplex, composed of a stack of four guanine quartets in the dimer hairpin quadruplex [d(G₄T₄G₄)]₂, one K⁺ cation was observed in the region between the two central quartets (89). Subsequently, a short, linear quadruplex, [d(TGGGCT)]₄, was crystallized in the presence of NaCl, and there, too, cations were observed between the stacks of guanine quartets, although in some cases these cations were almost coplanar with some quartets (90). Thus, the observation that d(G₃T₄G₃) did not form linear, four-stranded quadruplexes was not due to the inherent inability of Na⁺ to stabilize such structures. Presumably, the free energy difference, in NaCl,

between the dimeric hairpin and linear forms is so large that the oligonucleotide is observed only in the folded form. In contrast, in the case of KCl, the free energy difference between these two forms is sufficiently small that both species are observed at equilibrium. In this case, equilibrium is reached over a long period of time due to the extremely slow kinetics of formation/dissociation typical of four-stranded quadruplexes.

D. Quadruplex Thermodynamics

There has been a relatively small number of thermodynamic studies characterizing quadruplexes. One of the earliest studies considered the quadruplex formed by the sequence d(GGTTTTTG) (91). Based on a comparison of ΔH values determined calorimetrically and by UV melting, it was concluded that this sequence formed a four-stranded or tetramolecular quadruplex. The resulting ΔH per quartet formed (in 1 M NaCl) was 28 kcal/mol, whereas the overall stability resulted in a ΔG per quartet of -3.8 kcal/mol. Subsequent NMR studies on the same sequence revealed the presence of both syn and anti deoxyguanosines (85), consistent with the positive maximum in the CD spectrum near 290 nm (91). This, then, appears to be an unusual case of a tetramolecular quadruplex with antiparallel strands. Breslauer and co-workers (80) subsequently carried out calorimetric studies on several short, tetramolecular quadruplexes, based on the two sequences d(TGGGT) and d(TGGGTTGGGT), carried out in the presence of 1 M KCl, with $\Delta H = -21$ kcal/mol (quartet) in each case.

Kallenbach and co-workers examined the effect of parallel versus antiparallel strand configurations by determining the thermodynamic stability of the $[d(G_4T_4G_4)]_2$ and $[d(G_4T_2-T_2G_4)]_2$ dimer hairpin quadruplexes, where the second sequence involves a 5'-p-5' linkage (92). Gel electrophoresis provided evidence that each quadruplex consists of two strands. Interestingly, the CD spectrum of $[d(G_4T_2-T_2G_4)]_2$ was characteristic of parallel-stranded, tetramolecular quadruplexes, with a positive maximum near 265 nm. This was interpreted as reflecting an all-parallel strand arrangement in the quadruplex formed by the modified oligonucleotide. Of course, for this to be the case, the loops would have to be at the same end of the quadruplex, in contrast to the more common arrangement, reported for $[d(G_4T_4G_4)]_2$ both in the solution and solid-state structures, involving thymine loops placed at opposite ends. Thermodynamic analysis of these two structures revealed that the parallel structure was considerably more stable in that its T_m was 17°C higher and its free energy was 5 kcal/mol more negative than the structure based on d(G₄T₄G₄). Thus quartets formed from stretches of guanine all parallel to each other appear to be more stable than those formed from antiparallel sequences, which may explain the observation that most linear quadruplexes involve parallel strands. This interpretation is subject to some

uncertainty due to the different loop configurations associated with the two bimolecular quadruplexes. In the all-parallel hairpin structure, there may be favorable loop-loop interactions not present in the antiparallel structure.

As part of our initial characterization of the dimeric hairpin quadruplex $[d(G_3T_4G_3)]_2$, we used thermal denaturation monitored by UV to determine the thermodynamics of this system, both in the presence of Na^+ and K^+ . We obtained ΔH values, per quartet, of 32 and 23 kcal/mol, respectively. Corresponding ΔG values were -4 and -2.6 kcal/mol. Thus the K^+ form is substantially more stable than the Na^+ form. These results are in qualitative agreement with those described above.

In carrying out thermodynamic measurements on quadruplex systems, especially linear, four-stranded structures, the question of slow kinetics must be addressed. Our experience, as well as that of others (80, 81), indicates that equilibrium can occur on a time scale of days. Thus, if UV melting curves are used for thermodynamic analysis, maintaining equilibrium during the heating process can pose a major challenge.

E. High-Resolution Structure of Quadruplexes

In contrast to triplexes, diffraction-quality crystals have been obtained for quadruplexes, the first example of which involved the dimeric hairpin structure $[d(G_4T_4G_4)]_2$ (89). In the presence of spermine, KCl, and $MgCl_2$, this structure also exhibited alternating syn-anti conformations for the guanine nucleosides but was found to have the two thymine loops crossing over the edge of the terminal quartets (see Fig. 10). As mentioned earlier, a single K^+ ion was located in the region between the two central quartets. In contrast to this solid-state structure, the hairpin dimer quadruplex formed by the same sequence in aqueous solution, determined in the presence of NaCl by two-dimensional NMR, possessed a diagonally looped configuration (82, 86). These results provide a vivid example of the possible differences between solution and solid-state structures.

Subsequently, a high-resolution NMR structure appeared for the linear, four-stranded quadruplex $[d(TGGGGT)]_4$ stabilized by NaCl (93); this structure possessed fourfold symmetry with all strands parallel to each other and all nucleosides in the anti conformation. The same sequence was also crystallized in a conformation qualitatively similar to that observed in solution (90). In the solid-state solution the terminal thymines projected out from the quadruplex core, and some of the sodium cations were found to lie almost coplanar with some of the guanine quartets. In the solution structure, the terminal thymines were similarly not well stacked over the quartet core. This result is in contrast to that reported for the closely related RNA quadruplex, $[d(UGGGU)]_4$, whose NMR structure revealed significant stacking of the terminal uracils, to the extent that, at low temperature, the 3'-terminal Us ap-

peared to form a quartet with an observable UN-3 imino proton protected from solvent exchange (94).

We pursued a detailed structural analysis of the sodium form of $d(G_3T_4G_3)$, which was readily amenable to high-resolution NMR analysis because the resonances were sharp and the system stable, with no added complications arising from the presence of a linear, four-stranded species (95, 96). That this dimer hairpin quadruplex was different from other hairpin dimer quadruplexes was first noted because the quadruplex stem was composed of an odd number of guanine quartets, in contrast to other systems, which contained two or four quartets (85, 82). In those systems, sequential deoxyguanosines were in strict alternation between syn and anti conformations. However, a different picture emerged from the nonexchangeable proton NOESY spectra obtained on $[d(G_3T_4G_3)]_2$. Due to the odd number of quartets, the complex is asymmetric, resulting in two separate sets of resonances, one for each strand in the complex. As expected, half of the guanine nucleosides were in the syn conformation, readily discernible from their large, intranucleoside H-8-H-1' cross-peaks. Unexpectedly, however, there was a sequential NOE cross-peak between two syn nucleosides, G1 and G2, indicating a lack of alternation between syn and anti nucleosides. This connectivity was confirmed by 1H - ^{31}P heteronuclear correlation experiments. In the final analysis, the two strands were determined to possess the following sequence, with regard to glycosidic bond conformation (*s* denotes syn and *a* denotes anti): G1*s*G2*s*G3*a*T4*a*T5*a*T6*a*T7*a*G8*s*G9*a*G10*a* and G11*s*G12*a*G13*a*T14*a*T15*a*T16*a*T17*a*G18*s*G19*s*G20*a*. The same result was also reported by Smith *et al.* (97). This was the first example of a nucleic acid system containing two adjacent syn guanine nucleosides.

In addition to determining the local, sequential variation in nucleoside conformation, it was necessary to determine the overall strand topology, or configuration. The two principal possibilities, illustrated in Fig. 10, with regard to the position of the T_4 loop segment, involve edge loop or diagonal loop orientations. The former was observed in the X-ray crystal structure determined for $[d(G_4T_4G_4)]_2$, whereas the latter was observed for the solution structure of $[d(G_4T_4G_4)]_2$ (see above). We were able to distinguish between these two possibilities in the solution structure of $[d(G_3T_4G_3)]_2$ by focusing on several NOE cross-peaks involving both nonexchangeable and exchangeable protons (96).

The observation of the unusual syn-syn sequence raises the question of the role of base stacking, the thymine loop, and the quadruplex stem-loop junction in determining the overall folding pattern in the quadruplex structure. We examined the base stacking component of the overall interactions from the viewpoint of unrestrained molecular dynamics calculations, comparing the energy differences between the three possible dinucleotide stacking arrange-

ments: 5'-*syn-anti-3'*, 5'-*anti-syn-3'*, and 5'-*syn-syn-3'/5'-anti-anti-3'*. The three possible quartet stacking patterns are indicated in Fig. 11. Note that in quartet stack a all four sequential base stacks are *syn-anti*, whereas in quartet stack b they are *anti-syn* and in quartet stack c they occur in pairs of *syn-syn* and *anti-anti*. In the structure determined for $[d(C_3T_4G_3)]_2$, stacks a and c are observed, but there is no corresponding stack b. Our calculations found that this last stack was the least favorable energetically. Thus the observed set of stacking interactions is consistent with the only configuration that does not include type b quartet stacks.

In the NMR spectra of $[d(C_3T_4G_3)]_2$ stabilized by NaCl, there is a large number of sharp cross-peaks involving protons in the T_4 loop. This indicates the presence of a well-defined structure for this region, rather than a floppy structure characteristic of many such loops. Most of the interactions between the thymine bases involve stacking, but there is some evidence of thymine–thymine hydrogen bonding within each loop.

The overall folding of dimeric hairpin quadruplexes depends on a variety of interactions. Thus, in addition to the separate contributions from the quartet stem and the loop regions, there are contributions from stacking and other forces between the stem and loops. All such interactions must be taken into account in order to achieve a full understanding of the principles involved in determining quadruplex structure.

Although no X-ray structure has been reported for a unimolecular quadruplex, several high-resolution NMR structures have appeared. The first such structure involved a thrombin-binding aptamer discovered by *in vitro* selection from a randomly generated library of sequences (98–100). This aptamer turned out to be structured as a single-stranded quadruplex stabilized by two quartets connected by an edge-looped central TGT sequence. Its tight thrombin binding properties have been assumed to derive from its three-dimensional structure complementing the thrombin binding site structure. A unimolecular quadruplex based on the human telomeric repeat sequence,

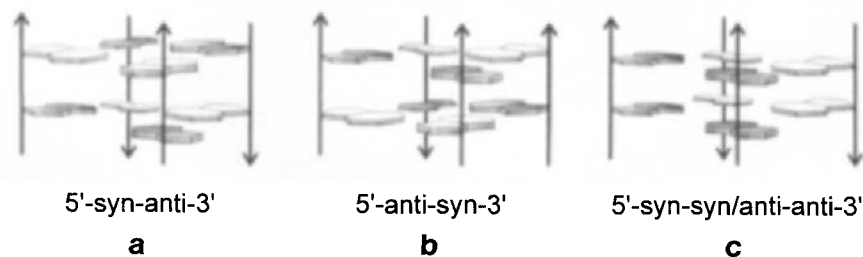


FIG. 11. Three types of possible guanine quartet stacks corresponding to dimeric hairpin quadruplex configuration displayed in Fig. 10C.

d(TTAGGG), has been analyzed (101). This structure, in contrast to that of the dimeric hairpin quadruplex $[d(G_3T_4G_3)]_2$ possesses strict syn-anti alternation. The central loop crosses the quadruplex stem diagonally and the other loops cross laterally. Most of the thymines were observed to be only loosely constrained.

Several unusual structures have been reported, including the dimeric hairpin quadruplex formed by the sequence d(G-C-G-G-T3-G-C-G-G) containing the fragile-X syndrome triplet repeat d(CGG) (102). Here the central two quartets are composed of the unusual combination of two guanine and two cytosine bases, all in the anti conformation, sandwiched by standard guanine quartets. In these central quartets, each hairpin maintains its G-C Watson-Crick hydrogen bonding and interacts with the opposing hairpin via major groove interactions. Each strand is antiparallel to its adjacent partner strand, in contrast to other dimeric hairpin quadruplex structures determined in solution. This strand topology, similar to that observed in the X-ray structure of $[d(G_4T_4G_4)]_2$ (89), leads to two narrow and two wide grooves.

Patel and co-workers have also reported a unimolecular quadruplex formed by the sequence $d(TTGGGG)_4$, containing four repeats of the tetrahymena telomeric sequence (103). In this structure, only three guanine quartets were formed, with a central TTG loop, another GTTC loop, and then an unusual TT string that involves a unique folding pattern. From these structures, it is clear that there are many degrees of freedom available to sequences composed of as few as two types of bases. This highly polymorphic character should result in many specific structures for targeting, ultimately, by small ligands.

F. Quadruplex-Ligand Interactions

As in the case of duplexes and triplexes, two major classes of quadruplex-binding ligands can be envisioned—groove binders and intercalators. In addition, as in the case of hairpin duplexes or folded triplexes, the possibility of loop-specific interactions with folded quadruplexes also exists. Intercalators are typically exemplified by ethidium, which has been shown to bind to a linear quadruplex with an affinity not very different from that for a linear duplex (104). Furthermore, evidence was presented supporting intercalation as the binding mode of ethidium to the linear quadruplex. We have observed an enhanced fluorescence from ethidium binding to the $[d(G_3T_4G_3)]_2$ quadruplex both by fluorescence spectroscopy and by gel staining (R. H. Shafer, unpublished results). Few of the other DNA intercalators have been investigated in terms of binding to quadruplexes.

One objective we have been pursuing is developing ligands capable of specific recognition of the quadruplex grooves. Our approach to this problem is based on the DOCK method, developed by Kuntz and co-workers, for

generating lead compounds targeted to a particular site (105). Briefly, this method involves several steps. First, a three-dimensional model of the binding site is constructed, either from an X-ray crystal or NMR solution structure, or from model building. Second, a "negative image" of the binding site is created by filling the site with a set of small spheres. Third, a computer database of small molecules is searched for compounds that match this negative image of the binding site. A differential search can be implemented in which candidate compounds are scored for high affinity to the targeted site and low affinity toward an undesired site.

In our initial application of the DOCK method, we started with the only high-resolution structure available at the time, that of the $[d(C_4T_4G_4)]_2$ quadruplex determined by X-ray crystal analysis (89). In that structure, there are four grooves, two narrow and two wide. We targeted one of the narrow grooves using the DOCK algorithm, at the same time scoring for ability to bind to the *minor groove of duplex DNA*. The best candidate ligands were those showing a good fit to the quadruplex groove and comparatively poor fit to the duplex minor groove.

Of the top 50 compounds generated by this search approach, we examined approximately 10 compounds experimentally and focused on one compound, diethyloxadycarbocyanine (DODC) (actually, the compound selected by the computerized searching program was the closely related structure diethyloxatricarbocyanine, but due to solubility limitations, we turned to DODC). Under our solution conditions, resembling physiological salt, we observed that DODC could bind nonspecifically to most nucleic acid forms, i.e., single stands and B-form duplexes. By nonspecifically, we mean that only limited changes in DODC spectrophotometric properties were observed on binding most DNA structures, i.e., there was a bathochromic shift in the ligand's visible absorbance, but no associated shift in wavelength. The fluorescence of the ligand remained substantially the same in the presence of these DNA structures, and no induced CD signal was detected in the visible region corresponding to the ligand absorbance.

In contrast to these results, binding of DODC to the $[d(C_4T_4G_4)]_2$ quadruplex led to a host of significant changes in its spectrophotometric behavior (106). A shoulder appeared on the short-wavelength side of the visible peak at 565 nm, a quenching of fluorescence was observed, and an induced CD peak in the visible region and protection from collisional quenching were also seen. None of these effects was observed for the complex of DODC with single- or double-stranded structures. One unexpected result was the detection of a very strong energy transfer effect near 254 nm, indicative of a stacking interaction between the ligand chromophore and a DNA base. Although this type of behavior is typically seen with intercalators, such as ethidium, intercalation (at least in the classical sense) is unlikely because DODC not only

does not bind specifically with duplexes, it also does not bind specifically with linear quadruplexes, such as $[d(\text{TGGGT})]_4$. This result leads one to consider the possibility of an interaction between the chromophore and a base in the loop of the $[d(\text{G}_4\text{T}_4\text{G}_4)]_2$ quadruplex. Because we have determined that DODC does not specifically recognize a hairpin duplex containing the same T_4 loop, $d(\text{CGCGTTTTCGCG})$, it is clear that the loop alone is not sufficient for recognition. Work is currently underway to determine the binding site geometry by high-resolution NMR analysis.

III. Summary

In this review, we have considered only a subset of the large number of unusual nucleic acid structures that have been described in the past decade. It is clear that both DNA and RNA are capable of forming structures of very high complexity and with a plasticity not imagined in the past. The biological significance of many of these structures remains to be determined, but it is highly likely that under *in vivo* conditions at least some of these conformations are sampled by certain regions of DNA and/or RNA. The ability to design small molecules capable of specific recognition of these unusual structures is a major challenge of paramount interest in the field of molecular recognition, and endeavors in this area may lead to pharmaceutically useful agents.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge many collaborators who have made significant contributions to much of the work described in this review: Q. Chen, Y. He, M. A. Keniry, I. D. Kuntz, C. Levenson, D. S. Pilch, P. V. Scaria, and G. D. Strahan. Financial support from NIH (GM-51650) and NSF (MCB 9218687) is also greatly appreciated.

REFERENCES

1. G. Felsenfeld, D. Davies, and A. Rich, *J. Am. Chem. Soc.* **79**, 2023 (1957).
2. V. N. Soyfer and V. N. Potaman, "Triple Helical Nucleic Acids," 1995.
3. J. S. Sun and C. Helene, *Curr. Opin. Struct. Biol.* **3**, 345 (1993).
4. N. T. Thuong and C. Helene, *Ang. Chem. (Int. ed. in English)* **32**, 666 (1993).
5. J. S. Sun, T. Garestier, and C. Helene, *Curr. Opin. Struct. Biol.* **6**, 327 (1996).
6. W. Guschlbauer, J. F. Chantot, and D. Thiele, *J. Biomol. Struct. Dynam.* **8**, 491 (1990).
7. J. R. Williamson, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703 (1994).
8. D. S. Pilch, G. E. Plum, and K. R. Breslauer, *Curr. Opin. Struct. Biol.* **5**, 334 (1995).

9. G. E. Plum, D. S. Pilch, S. F. Singleton, and K. J. Breslauer, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 319 (1995).
10. T. L. James, *Meth. Enzymol.* **261** (1995).
11. M. Riley, B. Maling, and M. J. Chamberlin, *J. Mol. Biol.* **20**, 359 (1966).
12. J. S. Lee, M. L. Woodsworth, L. J. Latimer, and A. R. Morgan, *Nucleic Acids Res.* **12**, 6603 (1984).
13. V. I. Lyamichev, S. M. Mirkin, and M. D. Frank-Kamenetskii, *J. Biomol. Struct. Dynam.* **3**, 667 (1986).
14. S. M. Mirkin, V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, and M. D. Frank-Kamenetskii, *Nature (London)* **330**, 495 (1987).
15. S. M. Mirkin and M. D. Frank-Kamenetskii, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 541 (1994).
16. H. E. Moser and P. B. Dervan, *Science* **238**, 645 (1987).
17. T. Le Doan, L. Perrouault, D. Praseuth, N. Habhoub, J. L. Decout, N. T. Thuong, J. Lhomme, and C. Helene, *Nucleic Acids Res.* **15**, 7749 (1987).
18. P. Hsieh, C. S. Camerini-Otero, and R. D. Camerini-Otero, *Genes Dev.* **4**, 1951 (1990).
19. V. B. Zhurkin, G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero, and R. L. Jernigan, *J. Mol. Biol.* **239**, 181–200 (1994).
20. J. S. Sun, T. De Bizemont, G. Duval-Valentin, T. Montenay-Garestier, and C. Helene, *C. R. Acad. Sci. III* **313**, 585 (1991).
21. C. Giovannangeli, M. Rougee, T. Garestier, N. T. Thuong, and C. Helene, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8631 (1992).
22. K. Umemoto, M. H. Sarma, G. Gupta, J. Luo, and R. H. Sarma, *J. Am. Chem. Soc.* **112**, 4539 (1990).
23. D. S. Pilch, C. Levenson, and R. H. Shafer, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1942 (1990).
24. G. Manzini, L. E. Xodo, D. Gasparotto, F. Quadrifoglio, G. A. van der Marel, and J. H. van Boom, *J. Mol. Biol.* **213**, 833 (1990).
25. P. Rajagopal and J. Feigon, *Biochemistry* **28**, 7859 (1989).
26. P. Rajagopal and J. Feigon, *Nature (London)* **339**, 637 (1989).
27. C. de los Santos, M. Rosen, and D. Patel, *Biochemistry* **28**, 7282 (1989).
28. L. S. Kan, D. E. Callahan, T. L. Trapane, P. S. Miller, P. O. Ts'o, and D. H. Huang, *J. Biomol. Struct. Dynam.* **8**, 911 (1991).
29. R. H. Shafer, D. S. Pilch, and P. V. Scaria, "Structure and Stability Studies on Short DNA Triple Helices" (R. H. Sarma and M. H. Sarma, eds.), p. 15. Adenine Press, Schenectady, New York, 1992.
30. D. S. Pilch, R. Brousseau, and R. H. Shafer, *Nucleic Acids Res.* **18**, 5743 (1990).
31. G. E. Plum, Y. W. Park, S. F. Singleton, P. B. Dervan, and K. J. Breslauer, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9436 (1990).
32. G. C. Best and P. B. Dervan, *J. Am. Chem. Soc.* **117**, 1187 (1995).
33. V. A. Malkov, O. N. Voloshin, V. N. Soyfer, and M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **21**, 585 (1993).
34. F. M. Chen, *Biochemistry* **30**, 4472 (1991).
35. D. S. Pilch, C. Levenson, and R. H. Shafer, *Biochemistry* **30**, 6081 (1991).
36. F. Svinarchuk, J. R. Bertrand, and C. Malvy, *Nucleic Acids Res.* **22**, 3742 (1994).
37. F. Svinarchuk, J. Paoletti, and C. Malvy, *J. Biol. Chem.* **270**, 14068 (1995).
38. S. L. Broitman, D. D. Im, and J. R. Fresco, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5120 (1987).
39. D. S. Pilch and K. J. Breslauer, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9332 (1994).
40. F. B. Howard, H. T. Miles, and P. D. Ross, *Biochemistry* **34**, 7135 (1995).
41. R. Lauceri, T. Campagna, A. Contino, and R. Purrello, *Ang. Chem. (Int. ed. in English)* **35**, 215 (1996).

42. Y. He, P. V. Scaria, and R. H. Shafer, *Biopolymers* **41**, 431 (1997).
43. P. V. Scaria and R. H. Shafer, *Biochemistry* **35**, 10985 (1996).
44. L. A. Marky and K. J. Breslauer, *Biopolymers* **26**, 1601 (1987).
45. J. D. Puglisi and I. Tinoco, Jr., *Meth. Enzymol* **180**, 304 (1989).
46. J. Gralla and D. M. Crothers, *J. Mol. Biol.* **78**, 301 (1973).
47. E. M. Evertsz, K. Rippe, and T. M. Jovin, *Nucleic Acids Res.* **22**, 3293 (1994).
48. M. Rougee, B. Faucon, J. L. Mergny, F. Barcelo, C. Giovannangeli, T. Garestier, and C. Helene, *Biochemistry* **31**, 9269 (1992).
49. K. J. Breslauer, E. Freire, and M. Straume, *Meth. Enzymol.* **211**, 533 (1992).
50. L. E. Xodo, G. Manzini, and F. Quadrioglio, *Nucleic Acids Res.* **18**, 3557 (1990).
51. J. Volker, D. P. Botes, G. G. Lindsey, and H. H. Klump, *J. Mol. Biol.* **230**, 1278 (1993).
52. G. E. Plum and K. J. Breslauer, *J. Mol. Biol.* **248**, 679 (1995).
53. W. D. Wilson, H. P. Hopkins, S. Mizan, D. D. Hamilton, and G. Zon, *J. Am. Chem. Soc.* **116**, 3607 (1994).
54. S. Arnott, P. J. Bond, E. Selsing, and P. J. Smith, *Nucleic Acids Res.* **3**, 2459 (1976).
55. F. B. Howard, H. T. Miles, K. Liu, J. Frazier, G. Raghunathan, and V. Sasisekharan, *Biochemistry* **31**, 10671 (1992).
56. I. Radhakrishnan and D. J. Patel, *Biochemistry* **33**, 11405 (1994).
57. R. Macaya, E. Wang, P. Schultze, V. Sklenar, and J. Feigon, *J. Mol. Biol.* **225**, 755 (1992).
58. P. V. Scaria, S. Will, C. Levenson, and R. H. Shafer, *J. Biol. Chem.* **270**, 7295 (1995).
59. M. Durand, N. T. Thuong, and J. C. Maurizot, *J. Biol. Chem.* **267**, 24394 (1992).
60. Y. W. Park and K. J. Breslauer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6653 (1992).
61. M. Durand and J. C. Maurizot, *Biochemistry* **35**, 9133 (1996).
62. M. Durand, N. T. Thuong, and J. C. Maurizot, *Biochimie* **76**, 181 (1994).
63. D. S. Pilch, M. A. Kirolos, and K. J. Breslauer, *Biochemistry* **34**, 16107 (1995).
64. A. R. Morgan, J. S. Lee, D. E. Pulleyblank, N. L. Murray, and D. H. Evans, *Nucleic Acids Res.* **7**, 547 (1979).
65. P. V. Scaria and R. H. Shafer, *J. Biol. Chem.* **266**, 5417 (1991).
66. J. S. Lee, L. J. Latimer, and K. J. Hampel, *Biochemistry* **32**, 5591 (1993).
67. H.-K. Kim, J.-M. Kim, S. K. Kim, A. Rodger, and B. Norden, *Biochemistry* **35**, 1187 (1996).
68. E. Tuite and B. Norden, *Bioorg. Med. Chem.* **3**, 701 (1995).
69. J. L. Mergny, G. Duvalvalentin, C. H. Nguyen, L. Perrouault, B. Faucon, M. Rougee, T. Montenaygarestier, E. Bisagni, and C. Helene, *Science* **256**, 1681 (1992).
70. D. S. Pilch, M. J. Waring, J. S. Sun, M. Rougee, C. H. Nguyen, E. Bisagni, T. Garestier, and C. Helene, *J. Mol. Biol.* **232**, 926 (1993).
71. G. Duval-Valentin, T. de Bizemont, M. Takasugi, J. L. Mergny, E. Bisagni, and C. Helene, *J. Mol. Biol.* **247**, 847 (1995).
72. W. D. Wilson, F. A. Tanious, S. Mizan, S. Yao, A. S. Kiselyov, G. Zon, and L. Strekowski, *Biochemistry* **32**, 10614 (1993).
73. S. P. Chandler, L. Strekowski, W. D. Wilson, and K. R. Fox, *Biochemistry* **34**, 7234 (1995).
74. E. H. Blackburn, *Nature (London)* **350**, 569 (1991).
75. E. H. Blackburn and C. W. Greider (eds.), "Telomeres." Cold Spring Harbor Laboratory, Plainview, New York, 1995.
76. D. Sen and W. Gilbert, *Nature (London)* **334**, 364 (1988).
77. K. Walsh and A. Gualberto, *J. Biol. Chem.* **267**, 13714 (1992).
78. P. Weisman-Shomer and M. Fry, *J. Biol. Chem.* **268**, 3306 (1993).
79. C. C. Hardin, E. Henderson, T. Watson, and J. K. Prosser, *Biochemistry* **30**, 4460 (1991).
80. R. Jin, B. L. Gaffney, C. Wang, R. A. Jones, and K. J. Breslauer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8832 (1992).
81. J. R. Wyatt, P. W. Davis, and S. M. Freier, *Biochemistry* **35**, 8002 (1996).

82. F. W. Smith and J. Feigon, *Nature (London)* **356**, 164 (1992).
83. E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, Jr., and E. H. Blackburn, *Cell* **51**, 899 (1987).
84. F. M. Chen, *Biochemistry* **31**, 3769 (1992).
85. Y. Wang, R. Jin, B. Gaffney, R. A. Jones, and K. J. Breslauer, *Nucleic Acids Res.* **19**, 4619 (1991).
86. P. Schultze, F. W. Smith, and J. Feigon, *Structure* **2**, 221 (1994).
87. P. V. Scaria, S. J. Shire, and R. H. Shafer, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10336 (1992).
88. V. P. Antao and D. M. Gray, *J. Biomol. Struct. Dynam.* **10**, 819 (1993).
89. C. Kang, X. Zhang, R. Ratliff, R. Moyzis, and A. Rich, *Nature (London)* **356**, 126 (1992).
90. G. Laughlan, A. I. Murchie, D. G. Norman, M. H. Moore, P. C. Moody, D. M. Lilley, and B. Luisi, *Science* **265**, 520 (1994).
91. R. Z. Jin, K. J. Breslauer, R. A. Jones, and B. L. Gaffney, *Science* **250**, 543 (1990).
92. M. Lu, Q. Guo, and N. R. Kallenbach, *Biochemistry* **32**, 598 (1993).
93. F. Aboul-ela, A. I. Murchie, D. G. Norman, and D. M. Lilley, *J. Mol. Biol.* **243**, 458 (1994).
94. C. Cheong and P. B. Moore, *Biochemistry* **31**, 8406 (1992).
95. G. D. Strahan, R. H. Shafer, and M. A. Keniry, *Nucleic Acids Res.* **22**, 5447 (1994).
96. M. A. Keniry, G. D. Strahan, E. A. Owen, and R. H. Shafer, *Eur. J. Biochem.* **233**, 631 (1995).
97. F. W. Smith, F. W. Lau, and J. Feigon, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10546 (1994).
98. R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe, and J. Feigon, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3745 (1993).
99. K. Y. Wang, S. H. Krawczyk, N. Bischofberger, S. Swaminathan, and P. H. Bolton, *Biochemistry* **32**, 11285 (1993).
100. K. Y. Wang, S. McCurdy, R. G. Shea, S. Swaminathan, and P. H. Bolton, *Biochemistry* **32**, 1899 (1993).
101. Y. Wang and D. J. Patel, *Structure* **1**, 263 (1993).
102. A. Kettani, R. A. Kumar, and D. J. Patel, *J. Mol. Biol.* **254**, 638 (1995).
103. Y. Wang and D. J. Patel, *Structure* **2**, 1141 (1994).
104. Q. Guo, M. Lu, L. A. Marky, and N. R. Kallenbach, *Biochemistry* **31**, 2451 (1992).
105. I. D. Kuntz, E. C. Meng, and B. K. Shoichet, *Acc. Chem. Res.* **27**, 117 (1994).
106. Q. Chen, I. D. Kuntz, and R. H. Shafer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2635 (1996).
107. R. D. Wells, D. A. Collier, J. C. Hanvey, M. Shimizu, and F. Wohlrab, *Faseb J.* **2**, 2939 (1988).

On the Physiological Role of Casein Kinase II in *Saccharomyces cerevisiae*

CLAIBORNE V. C. GLOVER III

*Department of Biochemistry and
Molecular Biology
Life Sciences Building
The University of Georgia
Athens, Georgia 30602*

| | |
|---|-----|
| I. General Properties of CKII | 96 |
| II. <i>Saccharomyces cerevisiae</i> CKII | 100 |
| A. Biochemistry | 100 |
| B. Molecular Biology | 100 |
| C. Genetics: Null and Conditional Alleles | 102 |
| D. Genetics: Multicopy Suppressors | 104 |
| III. Potential Functions of CKII in <i>Saccharomyces cerevisiae</i> | 109 |
| A. Flocculation/Gene Expression/Chromatin Structure | 109 |
| B. Cell Cycle Progression | 111 |
| C. Cell Polarity | 113 |
| D. Ion Homeostasis | 117 |
| IV. Substrates of CKII in <i>Saccharomyces cerevisiae</i> | 119 |
| A. Known Substrates | 119 |
| B. Predicted Substrates | 121 |
| V. The Physiological Role of CKII | 127 |
| References | 129 |

Casein kinase II (CKII) is a highly conserved serine/threonine protein kinase that is ubiquitous in eukaryotic organisms. This review summarizes available data on CKII of the budding yeast *Saccharomyces cerevisiae*, with a view toward defining the possible physiological role of the enzyme. *Saccharomyces cerevisiae* CKII is composed of two catalytic and two regulatory subunits encoded by the *CKA1*, *CKA2*, *CKB1*, and *CKB2* genes, respectively. Analysis of null and conditional alleles of these genes identifies a requirement for CKII in at least four biological processes: flocculation (which may reflect an effect on gene expression), cell cycle progression, cell polarity, and ion homeostasis. Consistent with this, isolation of multicopy suppressors of conditional *cka* mutations has identified three genes that have a known or potential role in either the cell cycle or cell polarity: *CDC37*, which is required for cell cycle progression in both G1 and G2/M; *ZDS1* and 2, which appear to have a function in cell polarity; and *SUN2*, which encodes a protein of the regulatory component of the 26S protease. The identity and properties of known CKII substrates in *S. cerevisiae* are also reviewed, and advantage is taken of the complete genomic sequence to predict globally the substrates of CKII

in this organism. Although the combined data do not yield a definitive picture of the physiological role of CKII, it is proposed that CKII serves a signal transduction function in sensing and/or communicating information about the ionic status of the cell to the cell cycle machinery. © 1998 Academic Press

Casein kinase II (CKII) is a highly conserved Ser/Thr protein kinase that is ubiquitously distributed among eukaryotic organisms. The enzyme was one of the first protein kinases discovered and has been studied for over 40 years (1). Nevertheless, both the mechanism of regulation of CKII and its physiological role remain poorly understood. Clues regarding the function of CKII *in vivo* have been obtained via the identification of probable physiological substrates and by molecular genetic studies in genetically tractable organisms. Over 100 well-characterized substrates of CKII are currently known (2). The enzyme phosphorylates a broad spectrum of targets, including both nuclear and cytoplasmic and both enzymatic and structural proteins. The multiplicity of substrates, many playing important roles in transcription, translation, signal transduction, cell cycle regulation, etc., has led to the view that CKII functions in global regulation of cellular metabolism. Consistent with the latter conclusion, molecular genetic studies have revealed that CKII is essential for viability in *Saccharomyces cerevisiae* (3) and probably also in *Schizosaccharomyces pombe* (4) and *Dictyostelium discoideum* (5). Other phenotypes observed in various genetic studies have provided intriguing glimpses into the possible physiological function(s) of CKII (3, 4, 6–11). However, the diversity of systems affected, ranging from the cell cycle to cell growth to cell polarity to ion homeostasis, render a coherent model of CKII function *in vivo* still beyond reach. The purpose of this review is to summarize the available information bearing on the physiological role of CKII in *S. cerevisiae*.

General features of CKII will be only briefly reviewed here; several broader reviews are available (12–15). Some aspects of the structure and biology of *S. cerevisiae* CKII have also been reviewed (16). Here the focus is on genetic studies, including the isolation of extragenic suppressors, which provide insights into the probable physiological function(s) of the enzyme. Advantage is also taken of the recent availability of the complete *S. cerevisiae* genomic sequence to predict in a global fashion the likely targets of CKII in this organism. The results of the latter analysis are compared with the genetic results.

I. General Properties of CKII

CKII from most sources is composed of two polypeptide subunits, α (35–44 kDa) and β (24–28 kDa), which combine to form a native $\alpha_2\beta_2$

tetramer (12–15). In many organisms two related α polypeptides are present, α and α' , and these are encoded by different genes. The presence of distinct β subunit isoforms appears to be rare but has been reported in *S. cerevisiae* (17) and *Arabidopsis thaliana* (18). A novel β subunit isoform is also present in *Drosophila melanogaster* (18a). The α polypeptide is catalytic and exhibits the expected homology to the catalytic domain of other protein kinases (19). The β subunit is subject to autophosphorylation (catalyzed by α) and is believed to serve a complex regulatory function (14, 15). Comparisons of the expressed free α subunit with native or reconstituted holoenzyme indicate that the β subunit stabilizes the α subunit against denaturation and stimulates catalytic activity approximately fivefold against most substrates. However, β also regulates α subunit specificity, and phosphorylation of some substrates, such as calmodulin, is completely blocked by the β subunit. The β subunit also mediates the activation of the catalytic subunit by polybasic effectors (see below). Recent studies indicate that the holoenzyme consists of a central regulatory subunit dimer flanked by catalytic subunit monomers (20, 21). CKII holoenzyme self-associates to form oligomers and polymers under physiological conditions *in vitro* (22–24), but the existence and significance of polymeric CKII *in vivo* remain to be demonstrated.

How CKII activity is regulated *in vivo* is unknown and represents a major deficiency in our understanding of the enzyme (1, 12–15). Because CKII is active as isolated from cells, the central problem is whether there exist mechanisms to restrain the activity *in vivo* and, if so, how these mechanisms are reversed to allow catalysis to occur. *In vitro*, CKII is inhibited by polyanions such as polyglutamate, polyaspartate, and heparin, modestly stimulated by polyamines, and strongly activated against specific substrates by polybasic compounds such as polylysine and protamine. Polyamines also block polymerization of CKII into linear filaments (24) and, surprisingly, prevent activation by polybasic effectors (25). Unfortunately, the physiological significance of these *in vitro* observations has not been established. Several laboratories have reported that CKII is activated in response to peptide growth factors, but these effects have been difficult to reproduce (26). Moreover, the design of most of these experiments would require that CKII activity be modulated by covalent modification. Although in some cases changes in autophosphorylation or phosphorylation by other kinases have been reported, a clear role for phosphorylation in regulation of CKII activity has not been established and appears unlikely (26). CKII also does not respond to known second messengers, though the existence of a novel messenger cannot be ruled out. These negative results leave open the possibility that CKII activity is constitutive, with regulation occurring via changes in enzyme localization, substrate effects, regulation of phosphatase activity, or other indirect

means. There is presently insufficient information to decide this important issue.

CKII is probably expressed in all tissues of metazoan organisms, but higher levels of activity are generally found in dividing cells, both normal and transformed (14). For example, CKII expression is elevated in proliferating cells of the early mouse embryo, in transformed human cell lines, and in solid human tumors. These observations suggest a positive role for CKII in cell proliferation. The intracellular distribution of CKII has been controversial but also appears to be broad (14). Cell fractionation studies in various laboratories have indicated that the enzyme is primarily nuclear, primarily cytoplasmic, or present in both compartments; and immunolocalization studies have yielded similarly diverse results. Critical analysis of the available data suggests that CKII is present in both nucleus and cytoplasm, with the preponderance of the enzyme being nuclear (27, 28). Pronounced nucleolar staining has also been observed in some studies, suggesting a potential role of CKII in this organelle (29, 30). Differences in localization obtained in different studies may reflect translocation of CKII from one cellular compartment to another. Translocation into the nucleus in response to growth factor stimulation has been reported (26), and nucleolar localization is lost as cells reach confluence (30). Translocation of the enzyme may represent an important mechanism of CKII regulation.

Over 100 probable physiological substrates of CKII are now known, and this number is likely to increase (1, 2). A surprising proportion of these proteins is involved in gene expression, growth control, signal transduction, and cell cycle regulation, consistent with a role for CKII in cell proliferation (15, 31, 32). Interestingly, several of the best substrates of CKII are nucleolar, including nucleolin (30) and the RNA polymerase I transcription factor, UBF (33), among others. Although the functional consequences of CKII phosphorylation are in most cases not known, effects on *in vitro* or *in vivo* activities have been identified in some cases (15, 31, 32). Phosphorylation by CKII regulates the activity of certain enzymes, including DNA ligase I and topoisomerase II; modulates the sequence-specific DNA binding of several transcription factors, including serum response factor, c-Myb, Max, and p53; and increases the rate of nuclear import of SV40 large T antigen. The latter may or may not reflect a general role of CKII in nuclear import (34). A requirement for CKII in rDNA transcription appears likely (30, 35) and correlates well with a role for the enzyme in cell proliferation.

Molecular genetic approaches have been used to explore the physiological role of CKII in *S. cerevisiae* (3, 6, 7), *S. pombe* (4, 8), and the mouse (11). The function of the enzyme in mammalian cells in culture has also been analyzed by overexpression of the enzyme subunits (36) and by antisense (9, 37)

or antibody (10, 38) inhibition of enzyme activity. Results obtained in *S. cerevisiae* are discussed in the body of this review. Results in the other systems are briefly summarized here.

Schizosaccharomyces pombe CKII appears to be encoded by only two genes, *cka1*⁺ and *ckb1*⁺, which encode the catalytic and regulatory subunits, respectively (8). The *cka1*⁺ gene has been identified as a temperature-sensitive recessive lethal mutation (*orb5*) that confers a spherical morphology on this normally filamentous yeast (4). Genetic and cytological studies indicate that this phenotype is due to a failure to reinitiate polarized growth after cytokinesis. Deletion of the *ckb1*⁺ gene is not lethal but results in slow growth, cold sensitivity, and an abnormal, rounded morphology reminiscent of *orb5* (8). Cells lacking *ckb1*⁺ also exhibit cell-cell aggregation and intense irregular staining with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) and Calcofluor and are resistant to digestion with glucanases, all suggestive of an altered composition or organization of the cell wall. Overexpression of *cka1*⁺ is without effect in *S. pombe*, but overexpression of *ckb1*⁺ results in severe growth inhibition and the production of multiseptated cells, the latter implying a defect in cytokinesis (8). Collectively, these results suggest a role for CKII in cell polarity and possibly in cell wall structure. A role for CKII in cell polarity is corroborated by studies in neuroblastoma cells, where antisense inhibition of CKII activity has been shown to inhibit neurogenesis (9).

Expression of the catalytic subunit of CKII in lymphocytes of transgenic mice leads to the stochastic production of lymphomas when expressed alone and to the production of acute lymphocytic leukemia when coexpressed with a *c-myc* transgene (11). Lymphocytes obtained from mice harboring the CKII transgene proliferate abnormally in culture, and this proliferation is sensitive to antisense inhibition of transgene expression, indicating that continued expression of the activity is required to maintain the proliferative state. These results imply a positive role for CKII in cell proliferation. Consistent with this conclusion, antisense or antibody inhibition of CKII activity in quiescent cells prevents mitogen-induced cell proliferation (37, 38). Moreover, this effect appears to reflect a requirement for CKII activity for cell cycle progression at several points in G₁ (10).

The above observations suggest that CKII is involved in transcriptional regulation, growth control, cell cycle regulation, and cell polarity, among other functions. Given the apparent complexity of CKII action, a combined biochemical, molecular, and genetic analysis of the enzyme is likely to be essential to success in determining its physiological role. Studies in genetically tractable systems, such as *S. cerevisiae*, are thus of great importance. The recent sequencing of the *S. cerevisiae* genome further enhances the power of this system.

II. *Saccharomyces cerevisiae* CKII

A. Biochemistry

Saccharomyces cerevisiae CKII has been purified to homogeneity and characterized both structurally and functionally (17, 39; for review, see 16). The purified enzyme is composed of two distinct catalytic subunits, α and α' , and two distinct regulatory subunits, β and β' , all of which are encoded by different genes (Fig. 1). The purified enzyme has been thoroughly characterized biochemically with regard to physical properties, substrate specificity, kinetics, autophosphorylation, etc. As expected, the enzyme is inhibited by polyanions such as heparin, stimulated by polycations such as spermine or polylysine, and exhibits autophosphorylation, which results in the phosphorylation of the β and β' subunits.

As is true of CKII from other organisms, the native holoenzyme is tetrameric. Indeed, available data are consistent with the proposition that *S. cerevisiae* CKII is an obligatory heterotetramer of α , α' , β , and β' . For example, deletion of either catalytic subunit gene alters the chromatography of all remaining CKII activity, implying the absence of tetramers containing only one type of catalytic subunit. The linear form of the holoenzyme shown in Fig. 1 is consistent with data bearing on the structure of CKII in higher systems (20, 21), and the order of subunits within the tetramer is consistent with synthetic phenotypes observed in strains lacking one catalytic and one regulatory subunit (40).

The free catalytic subunit does not appear to occur naturally in *S. cerevisiae*, but overexpression of active, monomeric *Drosophila* α subunit in *S. cerevisiae* produces no overt phenotype (41).

B. Molecular Biology

The *CKA1*, *CKA2*, *CKB1*, and *CKB2* genes encoding the α , α' , β , and β' subunits of *S. cerevisiae* CKII, respectively, have been cloned and sequenced

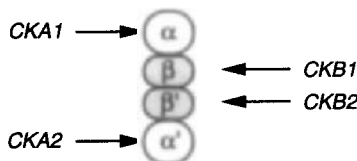


FIG. 1. *Saccharomyces cerevisiae* CKII. The four subunits of *S. cerevisiae* CKII and the genes encoding them are shown. Depiction of the holoenzyme as a heterotetramer of all four subunits, the linear form of the holoenzyme, and the specific order of subunits within the tetramer are consistent with available data (see text).

(3, 6, 40, 42). Computer searches of the recently completed *S. cerevisiae* genome confirm that no other closely related sequences are present. The predicted sequences of the yeast subunits are clearly homologous to the CKII α and β subunits of other organisms but are the most divergent sequences yet reported. This high divergence allows a better definition of the essential features of CKII.

The CKII catalytic subunit consists of a typical protein kinase domain preceded by a short N-terminal extension containing the conserved motifs, SEARVY and EYWDYE, and a variable C-terminal tail. The N-terminal extension bears some resemblance to the N-terminal extension of the catalytic subunit of cyclic AMP-dependent protein kinase (PKA) and has been proposed to play a role in stabilizing the hinge region between the large and small lobes of the catalytic domain (43). Distinct C-terminal tails are present on the α and α' subunits of mammalian CKII, and the difference in length between these tails is largely responsible for the greater molecular weight of the α subunit. Neither Cka1 nor Cka2 possesses a significant C-terminal extension, and the larger size of the Cka1 polypeptide is due to a unique insertion between subdomains III and IV of the catalytic domain (19).

The regulatory subunit contains an N-terminal autophosphorylation site, an internal acidic region, and a conserved CPX₃CX₂₂CPXC motif that may represent a metal-binding site. Mutation or deletion of the autophosphorylation site of the human β subunit has no significant effect on either holoenzyme assembly or activation of the α subunit *in vitro* (44). In contrast, mutations in the acidic region simultaneously inhibit autophosphorylation and activate the enzyme against exogenous substrates, suggesting a role in regulation (44–46). Neither type of mutation has yet been examined *in vivo*. Computer searches indicate that the cysteine-rich motif is most closely related to the zinc-binding site present in the regulatory subunit of aspartate transcarbamoylase (C. Glover, unpublished observations). In the latter protein, the motif constitutes a separately folded domain that contacts the catalytic subunit. Unusual features of the yeast regulatory subunits include the presence of a substantial insertion in the β subunit and the presence of sizable N-terminal extensions preceding the putative autophosphorylation site of both β and β' . Such N-terminal extensions are also present in *S. pombe* and plant β subunits.

Cka1 and Cka2 are as divergent from one another as they are from the catalytic subunits of other organisms, and the same is true of Ckb1 and Ckb2. This high degree of sequence divergence is consistent with the proposal that the holoenzyme is an obligatory heterotetramer of all four subunits. The high divergence of Cka1 and Cka2 is presumably also relevant to the distinct phenotypes displayed by *cka1^{ts}* and *cka2^{ts}* mutants (see below).

C. Genetics: Null and Conditional Alleles

Site-directed and/or random mutagenesis techniques have been used to construct null alleles of the *CKA1*, *CKA2*, *CKB1*, and *CKB2* genes as well as conditional (*GAL*-regulated or temperature-sensitive) alleles of *CKA1* and *CKA2*. The structure and properties of these alleles are summarized in Table I.

TABLE I
MUTANT ALLELES OF CKII IN *Saccharomyces cerevisiae*

| Allele | Mutation ^a | Phenotype | Ref. |
|------------------------------|-----------------------|-----------------------------------|-------|
| Null alleles | | | |
| <i>CKA1</i> | | | |
| <i>cka1-Δ1::HIS3</i> | Δ27–241 | Synthetic lethal with <i>cka2</i> | (42) |
| <i>cka1-Δ3</i> | Δ7–27 | Synthetic lethal with <i>cka2</i> | (49a) |
| <i>CKA2</i> | | | |
| <i>cka2-Δ1::TRP1</i> | Δ42–321 | Synthetic lethal with <i>cka1</i> | (3) |
| <i>CKB1</i> | | | |
| <i>ckb1-Δ1::URA3</i> | Δ15–208 | Salt sensitive | (6) |
| <i>ckb1-Δ1::HIS3</i> | Δ15–208 | Salt sensitive | (6) |
| <i>CKB2</i> | | | |
| <i>ckb2-Δ1::URA3</i> | Δ40–204 | Salt sensitive | (40) |
| <i>ckb2-Δ1::LEU2</i> | Δ40–204 | Salt sensitive | (40) |
| Conditional alleles | | | |
| <i>CKA1</i> | | | |
| <i>GAL1-CKA1^b</i> | None | Galactose dependent | (3) |
| <i>ckal-Δ2</i> | Δ7–12 | Temperature sensitive | (49a) |
| <i>ckal-13^c</i> | D253N | Temperature sensitive | (49a) |
| <i>CKA2</i> | | | |
| <i>cka2-7</i> | A190T, T336M | Temperature sensitive | (7) |
| <i>cka2-8</i> | E51K, G102D | Temperature sensitive | (7) |
| <i>cka2-11</i> | D225N, E299K | Temperature sensitive | (7) |
| <i>cka2-12</i> | A190V, H294Y | Temperature sensitive | (7) |
| <i>cka2-13^c</i> | D225N | Temperature sensitive | (7) |

^aNumbers refer to amino acid positions. The wild-type Cka1, Cka2, Ckb1, and Ckb2 proteins contain 372, 339, 278, and 258 amino acid residues, respectively. The *cka1-Δ2* and *cka1-Δ3* alleles are in-frame deletions. Point mutations are indicated in the one letter code.

^bThis allele is a transcriptional fusion of the *GAL1* promoter to a point 61 bp upstream of the *CKA1* protein-coding region. Strains that carry *GAL1-CKA1* as the only source of catalytic subunit (e.g., JC35-1b) exhibit essentially wild-type behavior in galactose medium but pronounced slow growth and flocculation in glucose medium (3). The survival of such strains on glucose is presumably due to basal expression from the *GAL1* promoter and/or residual function of the truncated *CKA1* promoter region.

^c*ckal-13* and *cka2-13* contain the identical replacement at the same site. The affected residue, which corresponds to D220 of PKA, is invariant in all protein kinases (19).

1. NULL ALLELES

Null alleles of the *CKA1*, *CKA2*, *CKB1*, and *CKB2* genes have been constructed by gene replacement (3, 6, 40, 42). In addition, all possible pairwise combinations of two null alleles (*cka1 cka2*; *ckb1 ckb2*; and *cka1 ckb1*, *cka1 ckb2*, *cka2 ckb1*, *cka2 ckb2*) as well as two of the four possible triple mutant combinations (*cka1 ckb1 ckb2* and *cka2 ckb1 ckb2*) have been constructed (3, 6, 40; J. Reed, A. Bidwai, and C. Glover, unpublished).

Deletion of *CKA1* or *CKA2* alone has no overt phenotypic effect on normal media, but deletion of both is lethal, demonstrating that CKII is essential for viability in *S. cerevisiae* (3). The inviability of *cka1 cka2* strains can be rescued by expression of the *Drosophila* α subunit, implying that the biochemical function of the enzyme has been conserved over large evolutionary distances. The inability of active site mutants of the *Drosophila* catalytic subunit to rescue demonstrates that the phosphotransferase activity of the enzyme is essential (47). Cells arrested by depletion of CKII continue to increase in size, suggesting that metabolism continues after cell division is halted. The arrested population consists of both unbudded and budded cells, with a significant fraction of the latter displaying an aberrant, elongated bud. CKII-depleted cells also exhibit flocculation, a Ca^{2+} -dependent lectin-mediated cell-cell aggregation (48). The significance of these phenotypes is discussed further below.

Deletion of either or both regulatory subunits has no detectable phenotypic consequences on normal media but results in specific sensitivity to high concentrations of Na^+ and Li^+ (6). Such strains do not display generalized osmosensitivity, and the Na^+ and Li^+ sensitivity is suppressible by high concentrations of K^+ . This constellation of phenotypes is consistent with a defect in Na^+ homeostasis (Section III,D). The fact that deletion of either gene results in a level of sensitivity identical to that displayed by the double mutant indicates that *ckb1* and *ckb2* are not functionally redundant with regard to this phenotype. One possible interpretation of this is that β and β' form an obligatory heterodimer whose function is lost on deletion of either gene. However, the sensitivity of the *ckb1* mutant is partially suppressed by multicopy *CKB2* (the converse is not true) and the sensitivity of both *ckb1* and *ckb2* strains is weakly suppressed by overexpression of *Drosophila* β , so there may be some ability for alternative regulatory dimers to form under these circumstances. *A priori*, the salt sensitivity of *ckb1* and/or *ckb2* strains could be due either to qualitative changes in enzyme targeting or specificity or simply to decreased kinase activity. We have recently found that *cka1* and *cka2* mutants are also salt sensitive, suggesting that the latter may be the case (J. Reed, A. Bidwai, and C. Glover, unpublished).

All four *cka ckb* double mutants are viable. In addition to their expected

salt sensitivity, all four strains exhibit synthetic slow growth and flocculation. The two triple mutants, one retaining only the α subunit and the other only α' , are also viable. In fact, such strains do not display any significant phenotypic deficit beyond that of the four *cka ckb* double mutants. The latter result is consistent with the identical phenotype displayed by the *ckb1 ckb2* double mutant relative to *ckb1* or *ckb2* mutants (6).

In general, CKII mutants grow on both fermentable and nonfermentable carbon sources, mate and sporulate normally, and exhibit normal resistance to nitrogen starvation and heat shock in stationary phase (3, 6, 40, 42).

2. CONDITIONAL ALLELES

Five independent temperature-sensitive alleles of the *cka2* gene have been isolated by a plasmid shuffling technique (7). Four of the five alleles (*cka2-7*, *-8*, *-11*, and *-12*) contain two amino acid replacements and the fifth (*cka2-13*) contains a single Asp-to-Asn replacement at a position corresponding to D220 of PKA. The latter residue is invariant in all members of the protein kinase family and is thought to play a structural role in stabilizing the catalytic loop (49). All five alleles are recessive to wild type. Analysis of strains harboring one of these alleles as the only source of CKII catalytic activity has revealed that CKII is required for cell cycle progression in both G_1 and G_2/M (7). These studies are discussed in detail in Section III,B.

Two temperature-sensitive alleles of the *CKA1* gene have been constructed by site-directed mutagenesis (49a). The first of these (*cka1- Δ 2*) was obtained serendipitously in the course of analyzing the function of the N-terminal conserved sequences of the catalytic subunit, and the other (*cka1-13*) was constructed by analogy to the *cka2-13* allele (Table I). Remarkably, these mutants do not exhibit the cell cycle arrest characteristic of *cka2^{ts}* alleles but instead exhibit a defect in cell polarity (Section III,C).

D. Genetics: Multicopy Suppressors

Multicopy suppression screens have been carried out using either the *cka2-13* allele (49b) or the *cka1- Δ 2* allele (A. Rethinaswamy and C. Glover, unpublished). In addition to the known interactions among the enzyme subunits, these screens have identified a genetic interaction of CKII with three novel functions: *CDC37*, identified in both screens; *ZDS1*, identified in the *cka2-13* screen; and *SUN2*, identified in the *cka1- Δ 2* screen. Each of these genes has subsequently been tested with other and in some cases all of the available temperature-sensitive alleles shown in Table I. In addition, there exists a homolog of *ZDS1* in the yeast genome; an explicit test of this gene, termed *ZDS2*, reveals behavior similar to that of *ZDS1* (multicopy plasmid kindly provided by D. Stillman). Interactions identified to date are summarized in Fig. 2.

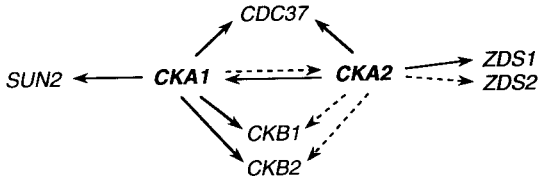


FIG. 2. Multicopy suppressors of *cka1* and *cka2* temperature-sensitive alleles. Solid lines indicate suppressors identified in a random screen; dashed lines indicate suppression events identified only by direct test. Not illustrated is the fact that *cka1* and *cka2* are also of course suppressed by *CKA1* and *CKA2*, respectively.

As expected, the wild-type *CKA1* and *CKA2* genes completely suppress the temperature sensitivity of either *cka1*^{ts} or *cka2*^{ts} strains, and in fact bypass the requirement for the mutant gene. In contrast, all of the other suppressors exhibit only partial suppression, and none of them bypasses the requirement for the mutant catalytic subunit. The *CKB1* and *CKB2* genes display varying degrees of allele specificity, consistent with their physical association with the catalytic subunit. To the extent we have tested them, *SUN2* and *ZDS1,2* exhibit gene but not allele specificity (though very weak suppression of *cka1-Δ2* by *ZDS* has occasionally been observed), whereas *CDC37* successfully suppresses all five *cka2*^{ts} alleles and both *cka1*^{ts} alleles. The fact that all four CKII subunit genes were identified in these screens confirms the power of multicopy suppression to detect relevant targets in this system. The *cka2-8* allele has in fact been used successfully to isolate a CKII β subunit from *Arabidopsis thaliana* (18). Additional information on the three “new” genes identified in these screens, *CDC37*, *ZDS1,2*, and *SUN2*, is presented below.

1. *CDC37*

The *CDC37* gene was originally identified by Reed (50) in a screen for temperature-sensitive mutants that arrest at Start in the G₁ phase of the cell cycle. The gene has been cloned and sequenced (51), but the predicted Cdc37 polypeptide exhibits no homology to known sequence motifs. *CDC37* interacts genetically with a remarkable diversity of genes encoding the catalytic subunit of protein kinases. In addition to *CKA1* and *CKA2*, these include *CDC28* (51), *KIN28* (52), *MPS1* (A. Schutz and M. Winey, personal communication), *BCK1* and *MPK1* (K. Irie and K. Matsumoto, personal communication), and mammalian *v-src* (expressed in yeast) (53). Similarly, the homolog of *CDC37* in *Drosophila* interacts genetically with the receptor tyrosine kinase encoded by the *sevenless* locus as well as with *Dmcdc2* (54). In all of these cases the interaction has the same sense: that is, overexpres-

sion of *CDC37* suppresses a deficiency in a kinase, and reduction of *CDC37* function enhances such a deficiency. These results suggest that *Cdc37* functions either as a general kinase activator or chaperone or as an inhibitor of a phosphatase.

Recent biochemical and molecular studies in mammalian cells in fact indicate that *Cdc37* is a subunit of a protein kinase-specific chaperone. Heat-shock protein 90 (*Hsp90*) functions as a molecular chaperone, involved in both the folding and transport of numerous proteins within the cell (55), and distinct *Hsp90* complexes are believed to direct this activity to distinct substrates (56). A complex containing *Hsp90* and a 50-kDa subunit termed p50 associates with several different protein kinases, including the tyrosine kinase v-src and the Ser/Thr kinase Raf1, implying that this form of *Hsp90* functions as a protein kinase-specific form of the chaperone. Consistent with the genetic studies described above, mammalian *Cdc37* has recently been cloned and shown to encode p50 (57). Moreover, a complex containing *Hsp90*, *Cdc37*, and the cyclin-dependent kinase *Cdk4* is detectable in several cell lines (57, 58), and purified *Cdc37* interacts with *Cdk4* directly *in vitro* (57). The latter suggests that *Cdc37* is the kinase-targeting component of the complex. The exact substrate specificity of the *Cdc37/Hsp90* chaperone remains to be defined, but it appears unlikely that the complex recognizes all protein kinases, because *Cdc37* does not interact with most cyclin-dependent kinases tested (57, 58).

The identification of *Cdc37* as a kinase-specific chaperone provides a plausible mechanism for the isolation of *CDC37* as a multicopy suppressor of various protein kinase mutants, including CKII mutants, namely, that an increase in the expression of the *Cdc37/Hsp90* chaperone increases the concentration of properly folded mutant enzyme. Although the observed lack of allele (or even gene) specificity is somewhat surprising in this context (because this is generally taken to mean that the mechanism of suppression does not involve the physical association of the proteins involved), the apparent paradox is resolved if *Cdc37* is designed, as it presumably must be, to detect some generic feature of the protein kinase domain and to ignore even quite dramatic changes in primary sequence.

2. *ZDS1,2*

At the time we originally isolated it, the *ZDS1* gene had not previously been described and displayed no homology to any known sequence. However, the sequence of this gene as well as that of the related gene, *ZDS2*, was subsequently deposited in GenBank as part of the global *S. cerevisiae* sequencing project. Shortly thereafter, *ZDS1* (as well as the homolog) was identified as a negative regulator of cell polarity, specifically as a gene that, in high copy, would reverse the ability of multicopy *CDC42* to suppress the cell po-

larity defect of a *cdc24* mutation (59). Within the last year, *ZDS1* and/or *ZDS2* have been isolated in a remarkable diversity of genetic screens. In addition to the interaction with *CKA1* and *CKA2*, *ZDS1* and/or *ZDS2* interact genetically with *CDC28* (M. Blondel and C. Mann, personal communication); *CEG1*, which encodes the guanylyltransferase responsible for mRNA capping (60); *CDC20*, which encodes a WD40 repeat-containing protein required for proper spindle function (D. Burke, personal communication); *SIN4*, which plays a role in chromatin structure (D. Stillman, personal communication); as well as several other genes (for review, see 59). The diverse nature of these interactions is responsible for the naming of this gene pair as *ZDS1* and *ZDS2* (zillion different screens 1 and 2) (59).

In spite of the diversity of these interactions, several observations suggest that *ZDS1* and *ZDS2* have a primary function in cell polarity. First, as noted above, *ZDS1* and *ZDS2* behave as negative regulators of cell polarity in their interactions with *CDC24* and *CDC42*. Second, although deletion of *ZDS1* or *ZDS2* alone results in no overt phenotype, deletion of both results in a viable cell that grows with an elongated bud (59). The latter phenotype is indicative of hyperpolarization (61) and is thus what one would predict from inactivation of a negative regulator of cell polarity. Third, *in situ* immunofluorescence of an epitope-tagged derivative of Zds1 localizes the protein to presumptive bud sites and the apex of small buds, a localization that closely mimics that of Cdc42 (59). The apparent involvement of *ZDS1,2* in other functions could reflect the increasingly diverse functions of Cdc42 (62, 63). Nevertheless, it remains possible that *ZDS1* and *ZDS2* have a primary function in some more global process and that the cell polarity and other effects observed are secondary.

Given the uncertainties in the function of *ZDS1* and *ZDS2*, a detailed mechanistic explanation for the genetic interaction of these genes with CKII cannot be given, though the absence of allele specificity suggests that physical association with the kinase is not involved.

3. *SUN2*

At the time we originally isolated and sequenced it, *SUN2* had not previously been identified in *S. cerevisiae* (the sequence was subsequently deposited in GenBank as part of the global *S. cerevisiae* sequencing effort and latter as *SUN2*; see below). *SUN2* is unique in the yeast genome but is homologous to a gene previously described in *Caenorhabditis elegans*, *Drosophila*, mice, cows, and humans. The gene was first identified in the mouse as the locus responsible for the P91A transplantation antigen (64), but no information on the normal function of the protein was obtained. The *C. elegans* gene was identified as part of the global sequencing effort in that organism, and again no functional information is available. The *Drosophila* ho-

molog was originally thought to encode Dox-A2, one of three diphenol oxidases detectable in this organism (65), but subsequent cloning of the gene (designated *Dox-A2*) failed to reveal the expected homology to the multicopper oxidase family (66). The effect on diphenol oxidase activity is thus presumably indirect. Consistent with the latter conclusion, the bovine and human homologs have recently been identified as a subunit of the regulatory complex of the 26S protease (67, 68). The 26S protease is composed of the catalytic 20S proteasome combined with a regulatory complex and functions in both ubiquitin- and non-ubiquitin-dependent proteolysis of intracellular proteins (for a review of 26S protease structure and function, see Refs. 68–70). Interestingly, the point mutation (R193H) responsible for the P91A transplantation antigen (64) results in presentation of a peptide carrying the mutation bound to class I major histocompatibility proteins, a function in which the 26S protease has been implicated (69, 70).

The *S. cerevisiae* homolog (*SUN2*, suppressor of *nin1-1*) has been identified in a screen for multicopy suppressors of a temperature-sensitive allele of *NIN1* (nuclear integrity) (71). The latter gene is required for cell cycle progression at both the G_1/S and G_2/M transitions, and *nin1-1* cells fail to activate Cdc28 kinase activity at either point (72). An important clue to the mechanism of action of Nin1 is provided by the discovery that it, too, is a subunit of the regulatory complex of the 26S protease (72). As discussed further in Section III,B, ubiquitin-mediated proteolysis of cyclin B and other proteins is required for cell cycle progression (73–75). Nin1 appears to be a positive regulator of protease function because *nin1-1* cells accumulate polyubiquitinated proteins at the restrictive temperature (72). The identity of these polyubiquitinated proteins has not been established, but Cdk inhibitors (e.g., Sic1) (74) would be logically consistent with the observed failure to activate Cdc28 kinase activity.

The ability of multicopy vectors carrying *SUN2* to suppress *nin1-1* may be explained either by a shared function (though the proteins exhibit no sequence identity) or possibly by a physical interaction of the two proteins within the protease. Construction and analysis of null alleles indicates that *SUN2* is essential for viability in *S. cerevisiae* (71; A. Rethinaswamy and C. Glover, unpublished), and this inviability can be rescued by expression of a cDNA encoding *Drosophila* Dox-A2 (71). Interestingly, strains rescued by Dox-A2 are temperature sensitive, and the mutant cells arrest with a large-budded phenotype with the nucleus lodged in the neck between mother and daughter. The latter phenotype resembles the terminal phenotype of *nin1-1* cells.

As discussed above, *SUN2* was isolated as a high-copy suppressor of the *ckal- $\Delta 2^{ts}$* allele. The identification of Sun2 as a subunit of the 26S protease offers several possibilities for the mechanism of interaction with CKII. First, overexpression of Sun2 might compensate for reduced protease function in

a *cka1^{ts}* mutant. CKII in fact copurifies with the 20S proteasome of human erythrocytes and phosphorylates a 30-kDa proteasome subunit (76), and the Sun2 protein contains a C-terminal CKII recognition motif. Second, if Sun2 were to function as a negative regulator of protease function, its overexpression could prevent or reduce degradation of CKII. This possibility appears unlikely given that *SUN2* behaves as a positive activator of protease function in its interaction with *nin1-1*, but it is possible that Sun2 alters the specificity of the protease, promoting degradation of some proteins and sparing others. Third, overexpression of Sun2 might increase CKII activity indirectly, by altering the degradation of some other protein that interacts with it.

III. Potential Functions of CKII in *Saccharomyces cerevisiae*

The genetic and other results outlined above implicate CKII in at least four important biological functions: flocculation (which, as discussed below, is probably a manifestation of altered gene expression), cell cycle regulation, cell polarity, and ion homeostasis. Although many hypomorphic CKII mutants also exhibit a slow growth phenotype, in general CKII does not appear to be limiting for metabolism and growth in this organism. *cka2^{ts}* strains do exhibit a twofold reduction in the rate of total RNA synthesis at the nonpermissive temperature, but total protein and DNA synthesis appear to be unaffected and cell mass continues to increase after division is halted (3, 7). The slow-growth phenotype of some CKII mutant strains is thus presumably due to effects on other systems, such as cell cycle progression and ion homeostasis, rather than to an effect on growth per se.

A. Flocculation/Gene Expression/Chromatin Structure

Flocculation in *S. cerevisiae* is thought to be mediated by the binding of specific extracellular proteins, termed flocculins, to the high-mannose side chains of cell wall-associated glycoproteins (77). Flocculins are encoded by a small gene family consisting of at least four functional genes, *FLO1*, *FLO5*, *FLO9*, and *FLO10*, as well as three pseudogenes, all of which have a subtelomeric location analogous to that of the *SUC*, *MAL*, and *MEL* gene families (78). Recent work indicates that *FLO* gene expression is negatively regulated by the Tup1/Ssn6 complex, a general transcriptional repressor involved in repression of mating type-specific genes, glucose-repressible genes, and oxygen-sensitive genes, among others (79). Ssn6 serves as an adapter to link the complex to targeting factors specific for each class of genes, whereas Tup1 is responsible for repression per se. Repression is believed to involve a change in chromatin structure because it is accompanied by the induction

of strong nucleosome phasing for long distances away from the site of initial recruitment of the complex. Consistent with this notion, the repression domain of Tup1 interacts directly with the N-terminal arms of histones H3 and H4 (80). Not surprisingly, loss-of-function mutations in either *TUP1* or *SSN6* result in derepression of genes regulated by the Tup1/Ssn6 complex. Such mutations also cause gross flocculation of *S. cerevisiae* (81, 82), and recent data confirm that flocculation results from derepression of *FLO* gene expression (83). The relevant cis-acting element(s) and targeting factor(s) in this case have not been identified.

Although *FLO* mRNA levels have not yet been examined, the flocculation phenotype of CKII mutants provides indirect evidence of a role for CKII in expression of the *FLO* genes. This could involve a specific effect directed at the *FLO* gene promoter, an effect on the Tup1/Ssn6 repressor complex, or perhaps a more generic effect on RNA polymerase II transcription and/or the establishment or maintenance of chromatin structure. Interestingly, *S. pombe* strains carrying a null allele of the *ckb1⁺* gene encoding the CKII regulatory subunit also exhibit cell-cell aggregation (8), so the phenomenon may be conserved between these two distantly related species.

In mammalian systems CKII phosphorylates a broad spectrum of proteins involved in transcriptional regulation, including subunits of the polymerases and numerous transcription factors (15, 31, 32). Although some of the data are indirect, as in the case of the genetic data just discussed, biochemical and/or genetic studies suggest that CKII affects transcription by all three classes of RNA polymerase in *S. cerevisiae*.

Studies in mammalian systems indicate a significant role for CKII in transcription by RNA polymerase I. CKII can be highly enriched in the nucleolus, particularly in proliferating cells (29, 30); several of the most efficient CKII substrates are nucleolar, including nucleolin and the polymerase I transcription factor UBF (30, 33), and polymerase I transcription is stimulated *in vitro* by phosphorylation by CKII (30, 35). CKII also physically associates with nucleolin (84). Although no direct data relating CKII to RNA polymerase I transcription are available in *S. cerevisiae*, *cka2^{ts}* strains incubated at the restrictive temperature exhibit a reduction in the rate of total RNA synthesis, which may reflect decreased synthesis of rRNA (7).

In addition to the probable involvement of CKII in *FLO* gene expression, we have identified one other RNA polymerase II-transcribed gene whose transcription is affected by CKII. As discussed in Section III,D, transcription of the *ENA1* gene is not induced in response to salt in *ckb* mutant strains (K. Tenney and C. Glover, unpublished).

Direct support for a requirement for CKII in transcription by RNA polymerase III has been obtained in *S. cerevisiae* (85). Transcription of 5S and tRNA genes by RNA polymerase III is reduced both *in vivo* and *in vitro* in a

cka2^{ts} strain, and the *in vitro* defect can be corrected by addition of purified *S. cerevisiae* CKII.

B. Cell Cycle Progression

Cell cycle progression in *S. cerevisiae* is mediated by Cdc28 (Cdk1), a member of the cyclin-dependent protein kinase (Cdk) family (86). As is the case for Cdks generally, Cdc28 is positively regulated by phosphorylation by Cdk-activating kinase (CAK) and by association with cyclins, and negatively regulated by association with Cdk inhibitors (CDIs or CKIs) as well as by inhibitory phosphorylation within the ATP binding site (for a review of Cdk structure and regulation, see Ref. 87).

The CAK responsible for activating phosphorylation of Cdc28 at Thr-169 has recently been identified and shown to be essential for viability (88, 89), but the regulatory significance of this phosphorylation event remains to be defined. As in other systems, entry into various cell cycle stages is regulated by the periodic synthesis of cyclins, and Cdc28 associates with at least nine different species: Cln1, 2, and 3, required for entry into the cell cycle (Start); Clb3, 4, 5, and 6, which promote S phase entry and spindle assembly; and Clb1 and 2, required for entry into mitosis (90). Entry into S phase also critically depends on Sic1, a CDI that specifically inhibits Clb/Cdc28 activity (74). In addition to the regulated synthesis of cyclins, selective ubiquitin-mediated proteolysis of these and other proteins is also crucial to cell cycle progression (69, 70). A ubiquitin-protein ligase complex containing Cdc53 and the ubiquitin-conjugating enzyme Cdc34 selectively degrades Clns and probably also Sic1 (91). Degradation of Sic1 is required for entry into S phase (74, 92). A second such complex containing Cdc16, Cdc23, and Cdc27 selectively degrades several mitotic proteins, including Clb cyclins; degradation of an as yet unidentified target is required for sister chromatid separation (metaphase-to-anaphase transition), and degradation of Clb cyclins is required for exit from mitosis (75, 93).

Inhibitory phosphorylation of Tyr-19 is catalyzed by Swe1 protein kinase (94) and removed by the Mih1 phosphatase (95). Functions operating upstream of Swe1 have recently been described (96). In contrast to its role in other organisms (97), this phosphorylation event in *S. cerevisiae* functions as part of a checkpoint that monitors cell morphogenesis (98) (see Section III,C).

All five temperature-sensitive alleles of *CKA2* display defects in cell cycle progression in both G_1 and G_2/M (7). Arrest occurs within a single cell doubling following a shift to nonpermissive temperature, and the arrested population consists of approximately equal numbers of budded and unbudded cells. The unbudded half of the population contains a single nucleus and a single focus of microtubule staining. The budded half of the population is heterogeneous: the majority of the cells contain segregated chromatin and

an extended spindle, indicating arrest in anaphase, but the remainder contain an undivided nucleus with a short, thick intranuclear spindle, indicating arrest in G_2 and/or metaphase. Flow cytometry of synchronized cells confirms that CKII is required in G_1 , at a point which must lie at or beyond Start but prior to DNA synthesis, and again in G_2 and/or M. The data do not reveal a requirement for CKII in DNA replication, though the design of the experiments does not exclude a requirement for CKII in the initiation of DNA synthesis. Cells arrested at nonpermissive temperature rapidly become inviable, apparently as a result of the G_2 /M block, which is irreversible. This, combined with the fact that the majority of G_2 /M arrested cells have proceeded into anaphase, suggests that CKII is required downstream of mitotic Cdc28 activity and that CKII functions in execution rather than regulation of cell cycle events in this phase of the cycle. In contrast, the CKII G_1 block is partially reversible: cells bud, enter and complete S phase, but then die in G_2 /M. The reversibility of G_1 arrest suggests that CKII may serve a regulatory function in G_1 .

The isolation of *CDC37* as a multicopy suppressor of CKII mutations is intriguing in light of the above results (49b). Although the discovery that Cdc37 is part of a kinase-specific chaperone offers a simple model to explain its genetic interaction with CKII and other protein kinases, the cell cycle arrest of *cdc37* mutations still requires explanation. Indeed, analysis of additional *cdc37* alleles has revealed that Cdc37 is required not only for Start, but for the G_2 /M transition as well (99). Because most of the kinases that are known to interact with Cdc37 have no clear relationship to the cell cycle, those that do, notably Cdc28 and CKII, are likely to be involved in the G_1 and/or G_2 /M arrest of *cdc37* strains. The obvious candidate is Cdc28, and loss of Cdc37 function has in fact been shown to reduce the association of Cdc28 with both G_1 and mitotic cyclins *in vivo* (100). However, the post-S phase arrest of *cdc37* mutants differs significantly from that of *cdc28* mutants: *cdc28* cells arrest as large-budded cells with a single nucleus and a short, thick intranuclear spindle (101, 102), whereas a substantial proportion of *cdc37* cells arrest with well-separated nuclei and an extended spindle (99). This suggests a requirement for Cdc37 function downstream of Cdc28 activation at the G_2 /M boundary. In addition, *cdc37* strains that exhibit G_2 /M arrest rapidly lose viability at the restrictive temperature (99), whereas comparable *cdc28* strains do not (101). Thus, failure to activate Cdc28 appears insufficient to explain all aspects of the *cdc37* phenotype. Of the remaining kinases that interact genetically with Cdc37 in *S. cerevisiae*, CKII has the clearest cell cycle connection. In fact, because the Cdc37 arrest phenotype closely mimics that of *cka2^{ts}* strains (7), failure to activate CKII could in theory explain all aspects of the phenotype of *cdc37* mutants. Additional stud-

ies will be required to determine the relative importance of Cdc28, CKII, and other protein kinases as downstream targets of Cdc37.

The interaction of CKII with Cdc37 has two unique features that may be significant. First, mammalian CKII physically associates with Hsp90 both *in vivo* and *in vitro* to form an 8S complex (103, 104). Formation of the complex *in vitro* does not require additional proteins, demonstrating that the interaction is direct. Association with Hsp90 antagonizes CKII polymerization and stimulates kinase activity against several substrates, including Hsp90. Hsp90 is a well-characterized substrate of CKII in mammals (105), though the sites are not conserved in the *S. cerevisiae* homolog, Hsp82. Second, Cdc37 contains a well-conserved CKII recognition site near the N-terminus of the protein, raising the possibility that Cdc37 is a substrate of CKII *in vivo*. Consistent with this prediction, phosphorylation of Cdc37 is temperature sensitive in a *cka2-13^{ts}* strain, and mutation of the CKII recognition site eliminates the majority of Cdc37 phosphorylation *in vivo* (R. McCann and C. Glover, unpublished). Moreover, the mutant derivative fails to suppress the temperature sensitivity of a *cdc37-1* strain. These observations suggest that CKII may regulate Cdc37, in addition to serving as one of its targets.

Combined with the genetic data, the above observations raise the possibility that CKII and Cdc37 function as a molecular switch in which CKII activates Cdc37, which then activates CKII as well as a variety of other kinases, including Cdc28. Such positive-feedback loops constitute an important part of cell cycle regulation (106).

C. Cell Polarity

Rearrangements of the actin cytoskeleton and secretory apparatus are believed to drive the dramatic morphological changes that occur during cell division and mating in *S. cerevisiae* (107, 108). During bud initiation, a ring of cortical actin patches assembles at the presumptive bud site, and actin cables terminating on the cortical patches appear to direct secretory vesicles to the site of bud growth (109, 110). Both the cortical actin patches and secretion are directed to the apex of the bud early in the budding cycle but become randomly distributed later, the shape of the bud being determined by the balance between these apical and isotropic modes of growth. These cytoskeletal and secretory rearrangements are dependent on proper establishment and maintenance of cell polarity.

Establishment of cell polarity in *S. cerevisiae* is mediated by the guanine nucleotide exchange factor encoded by *CDC24* and the Rho-family GTPase encoded by *CDC42* (111, 112). Temperature-sensitive *cdc24* or *cdc42* mutants arrest as unbudded spherical cells that exhibit an apolar actin cytoskeleton and elevated and isotropic deposition of cell wall chitin. Arrested

cells increase in size, indicating that metabolism and growth are not impaired, and a significant fraction become multinucleate, indicating that the nuclear cycle continues. This classical apolar phenotype is also conferred by mutations in several other genes, including *CDC43*, *BEM1*, and *BEM2* (112–114). The mechanism by which Cdc42 activation is coupled to downstream events is poorly understood, but physical interaction with Ste20-like kinases may play a part (62, 63). Genes apparently required for the maintenance of cell polarity have also been described, notably *RHO3* and its homolog *RHO4* (115). Temperature-sensitive *rho3* mutants exhibit a phenotype very similar to that of *cdc24* or *cdc42* mutants except that budded as well unbudded cells are observed in the arrested population (116).

As mentioned in Section II,C,2, both currently available *cka1^{ts}* alleles confer a terminal phenotype that differs dramatically from that of all five *cka2^{ts}* alleles (49a). In spite of their slow growth at permissive temperature, *cka1^{ts}* strains continue to divide for at least three cell divisions following a shift to restrictive temperature, and then arrest as a mixed population of budded and unbudded cells with a spherical morphology. Arrested cells continue to increase in size, indicating that metabolism continues, and a significant fraction become bi- or multinucleate, indicating that the nuclear cycle often continues. Chitin is elevated and randomly distributed in unbudded cells and mother cells, and the actin cytoskeleton is completely delocalized. This phenotype strongly resembles that of other cell polarity mutants, but the presence of budded cells in the population suggests that *cka1^{ts}* strains, like *rho3^{ts}* strains, are defective in maintenance rather than in initiation of cell polarity.

The phenotype of *cka1^{ts}* strains shares many features in common with that of *orb5* mutants of *S. pombe* (4). However, the underlying mechanism appears to be fundamentally different in the two cases. Genetic and cytological data indicate that *orb5* strains are not defective in polarization per se but in the reinitiation of polarized growth following cytokinesis. Because of the filamentous nature of *S. pombe*, continued division in the absence of growth ultimately results in the production of spherical cells that are too small to divide. In contrast, *S. cerevisiae cka1* strains retain the ability to grow but lose cell polarity. The consequent isotropic growth ultimately produces the spherical morphology. Indeed, the mechanism proposed to explain the *S. pombe orb5* mutants cannot in principle explain the development of spherical morphology in budding yeast, where growth is a prerequisite to cell division. How similar the roles of CKII may be in the two organisms thus remains uncertain.

That five different alleles of *CKA2* exhibit a cell cycle phenotype whereas two different alleles of *CKA1* (one of which is identical to one of the *cka2* alleles) exhibit a cell polarity defect clearly indicates that these are gene-

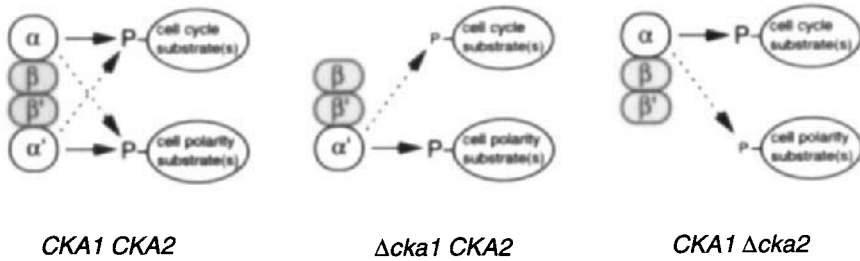


FIG. 3. A model to explain the distinct terminal phenotypes of *cka1^{ts}* and *cka2^{ts}* alleles. The *S. cerevisiae* CKII holoenzyme is drawn as in Fig. 1. Cka1 is assumed to phosphorylate preferentially one or more substrates required for cell cycle progression (solid arrow) but to phosphorylate only weakly one or more substrates required for cell polarity (dashed arrow); the opposite is assumed to be true for Cka2. The level of phosphorylation on these substrates in various strains is indicated by the size of the letter P. $\Delta cka1$ *CKA2* and *CKA1* $\Delta cka2$ strains are viable because all crucial substrates are phosphorylated sufficiently to achieve essentially normal function, but distinct proteins are hypophosphorylated in the two cases. Loss of the remaining catalytic subunit (e.g., by inactivation of a temperature-sensitive allele) leads to eventual dephosphorylation of all CKII substrates, but the phosphorylation of cell cycle targets becomes limiting before that of cell polarity targets in the former case, whereas the opposite is true in the latter.

rather than allele-specific phenotypes. This result is surprising in view of the functional redundancy of *CKA1* and *CKA2* (3, 42). One possible way to resolve the apparent paradox is illustrated in Fig. 3. In this model we assume that Cka1 and Cka2 are both able to phosphorylate all essential CKII substrates, but that the two isoforms have either different kinetic constants for these substrates or differing access to them. Either subunit would thus be capable of supporting viability in the absence of the other, but *cka1* and *cka2* strains would contain different levels of phosphorylation on individual substrates. Given this scenario, the phosphorylation of different substrates would initially become limiting when the remaining catalytic subunit is inactivated. Based on the phenotypes observed, we propose that *CKA1* preferentially phosphorylates essential cell cycle targets (so that these are hypophosphorylated in a $\Delta cka1$ *cka2^{ts}* strain prior to a shift to the restrictive temperature), whereas *CKA2* preferentially phosphorylates targets involved in cell polarity. Regardless of the explanation, the distinct behavior of temperature-sensitive *cka1* and *cka2* strains provides evidence for functional specialization of *CKA1* and *CKA2*.

The occurrence of both cell cycle and cell polarity phenotypes in CKII mutants may reflect the existence of intimate connections between the cell cycle and cell morphogenesis in budding yeast (108). Studies of *S. cerevisiae* strains lacking or overexpressing different cyclins indicate that specific cy-

clin/Cdc28 complexes orchestrate the changes in cell polarity required to drive the morphological changes occurring during the cell cycle (61). Specifically, Cln/Cdc28 complexes promote apical growth and Clb/Cdc28 complexes promote isotropic growth. In addition, a checkpoint exists that retards cell cycle progression when morphogenesis is incomplete (98). Because of these underlying connections, the distinct phenotypes of different CKII mutants may be mediated by a primary role of CKII in only one of the two processes.

The apolar phenotype of *cka1^{ts}* strains is perhaps most difficult to reconcile with the apparent hyperpolarization of *cka2^{ts}* strains (recall that the latter produce a significant fraction of elongated buds at restrictive temperature). Although seemingly contradictory, such behavior is not unprecedented and has been observed even in the context of a single allele of a single gene (*cdc24-4*) (111). A possible explanation for this ambivalent behavior may reside in the complexity of the controls regulating cell polarity. In addition to Cdc42, at least three other Rho-family GTPases are involved in cell polarity in *S. cerevisiae* (107, 114, 116) as well as numerous protein-protein interactions mediated by pleckstrin homology or SH3 domains (117). Mutations that affect this system may lead to abnormal interactions among the components, with unpredictable results.

The isolation of *ZDS1,2* as a multicopy suppressor of *cka2^{ts}* alleles provides additional evidence of a role for CKII in cell polarity. Although *ZDS1,2* exhibits pleiotropic interactions with diverse genes, several lines of evidence, including the null phenotype, suggest that this gene pair functions as a negative regulator of cell polarity (Section II,D,2). Consistent with such an interpretation, *ZDS1,2* behaves as a strong multicopy suppressor of *cka2^{ts}* mutants, which exhibit a tendency to hyperpolarize, but not of apolar *cka1^{ts}* mutants (one might predict that *ZDS1,2* would behave as a multicopy enhancer of *cka1^{ts}* alleles, but this is not observed). Further speculation regarding the function of *ZDS* and the mechanism of its genetic interaction with CKII must await additional studies. Nevertheless, it is tempting to speculate that the pleiotropy of *ZDS* interactions is indicative of a crucial interaction with CKII, itself one of the most pleiotropic of enzymes.

Assuming that CKII activity is completely eliminated at restrictive temperature, the behavior of *cka1^{ts}* strains has one very important implication for the possible function of CKII in cell cycle progression. Specifically, the ability of *cka1^{ts}* cells to proceed efficiently through several cell divisions at restrictive temperature prior to arrest implies that no essential cell cycle target of CKII is dephosphorylated in an obligatory way during each cell cycle. This in turn implies that CKII is not an integral part of the cell cycle machinery per se. Rather, it is more likely to function in a signal transduction capacity, feeding some information on the status of the cell or its environ-

ment to the cell cycle engine. A caveat of course is that the activity is completely inactivated at the nonpermissive temperature. However, the slow growth of both *cka1-Δ2* and *cka1-13* strains at permissive temperature, the very different nature of these two alleles, and the similar behavior of *S. pombe orb5* strains all argue that this is the case. The nature of the hypothetical signal transduced by CKII is unknown, but information of the ionic status of the cell is one possibility (Section III,D).

D. Ion Homeostasis

Na⁺ homeostasis in *S. cerevisiae* involves the coordinate regulation of Na⁺ influx and efflux across the plasma membrane (118, 119). Influx of Na⁺ and other alkali metals occurs primarily via a K⁺ uptake system encoded by the *TRK1* and *TRK2* genes, whereas Na⁺ efflux is mediated by a putative P-type ATPase encoded by the *ENA1-4* gene cluster. *TRK1* and *TRK2* encode homologous membrane proteins with a topology reminiscent of other transporters. However, because these proteins lack sequence similarity to known transporters, it is uncertain whether these proteins directly mediate K⁺ uptake or, alternatively, serve to regulate a more conventional transporter. Trk1 is the more highly expressed and appears to be responsible for the majority of high-affinity K⁺ uptake in normal cells (120, 121). The *ENA1-4* genes are believed to encode Na⁺ pumps, though direct biochemical proof of this is currently lacking. *Ena1*, encoded by the first gene of the tandem array, is the most highly expressed and appears responsible for most of the salt tolerance of yeast (122, 123). Both systems respond to Na⁺ stress: the Trk1 K⁺ uptake system is converted to a higher affinity form that discriminates more efficiently against Na⁺ and other alkali metals (124), whereas *ENA1* gene expression is induced (123). Loss of these responses presumably underlies the Na⁺ and Li⁺ sensitivity of *trk1* or *ena1* null mutants.

A number of genes involved in the regulation of Na⁺ homeostasis in *S. cerevisiae* have been identified, most by virtue of a specific effect on Na⁺ and Li⁺ sensitivity. A remarkable number of these encode protein kinases or phosphatases, suggesting that a complex network of kinases and phosphatases regulates ion homeostasis in *S. cerevisiae*. The latter genes include *YCK1* and *YCK2*, which encode redundant isoforms of casein kinase I (CKI) (125); *YCR101*, which encodes a putative but uncharacterized protein kinase (126); *BCY1*, which encodes the regulatory subunit of PKA (127); *CNAL2* and *CNBL*, which encode the catalytic and regulatory subunits of protein phosphatase 2B (calcineurin) (118, 128, 129); and *PPZ1* and *PPZ2*, which encode a novel Ser/Thr protein phosphatase related to protein phosphatase I (130). Interestingly, loss-of-function mutations in CKI, Ycr101, Bcy1, or calcineurin result in salt sensitivity, whereas loss-of-function mutations in Ppz1 or Ppz2 result in salt tolerance. For Bcy1, calcineurin, and Ppz the salt-sensitive

(or -resistant) phenotype has been shown to involve an effect on expression of *ENAI* and/or Trk1 function.

Additional genes involved in Na⁺ homeostasis have been identified by screening for genes whose overexpression confers supranormal resistance to Na⁺ or Li⁺ (halotolerance, or *HAL* genes). *HAL1* (131) and *HAL3* (119) influence intracellular K⁺ and Na⁺ concentrations and hence appear to serve a regulatory role. *HAL3* is required for optimal expression of *ENAI* and probably also affects conversion of Trk1 to the high-affinity form (119). Neither Hal1 nor Hal3 contains known protein motifs that give a clue to their biochemical function.

As outlined in Section II,C,1, loss-of-function mutations of yeast CKII, notably those involving the *CKB1* and/or *CKB2* genes, exhibit specific sensitivity to Na⁺ and Li⁺, and this sensitivity is K⁺ suppressible (6). This behavior is diagnostic of a defect in Na⁺ homeostasis and implies that CKII affects either Na⁺ influx, efflux, or both. Consistent with the latter prediction, we have recently found that *ckb1*, *ckb2*, and *ckb1 ckb2* strains are defective in the ability to induce *ENAI* gene expression in response to high external Na⁺ (K. Tenney and C. Glover, unpublished observations). Whether an effect on the Trk system is also involved has not yet been determined. The simplest model to explain these results is that CKII phosphorylates and activates one or more proteins required for induction of *ENAI* expression in response to salt stress (alternatively, CKII might phosphorylate and inactivate an inhibitor of *ENAI* expression). Furthermore, because *ppz* null mutants exhibit salt tolerance, Ppz1 and Ppz2 could be the phosphatase responsible for dephosphorylating these CKII substrates.

Considering the phenotypes observed, all of the kinases and phosphatases known to be involved in Na⁺ homeostasis can be organized into two parallel pathways (Fig. 4). Three of the four protein kinases, CKI, CKII, and Ycr101, behave as positive regulators of salt tolerance, and the action of all three is potentially reversible by Ppz1 and Ppz2 (130). In contrast, PKA behaves as a negative regulator of salt tolerance because constitutive activation of the kinase by deletion of *BCY1* results in salt sensitivity (127). Because loss-of-function mutations in calcineurin cause salt sensitivity, dephosphorylation of PKA substrates relevant to Na⁺ homeostasis could be catalyzed by this phosphatase. In none of these cases have the relevant substrates been identified. One or more transcription factors required for activation of the *ENAI* gene are an obvious possibility, but the cis- and trans-acting elements regulating expression of this gene have not been well defined.

CKII mutants are not osmosensitive, and there are currently no data implicating CKII in systems that sense or regulate osmotic pressure, the other principal component of salt stress (132).

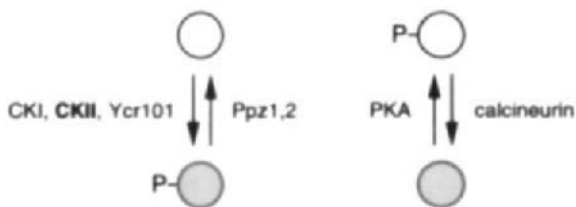


FIG. 4. Possible functional relationships among protein kinases and phosphatases involved in ion homeostasis in *S. cerevisiae*. As discussed in the text, CKI, CKII, Ycr101, and calcineurin promote salt tolerance, whereas Ppz1,2 and PKA have the opposite effect. Given that CKI, CKII, Ycr101, and PKA are protein kinases and Ppz1,2 and calcineurin are protein phosphatases, these observations can be explained by assuming that CKI, CKII, and Ycr101 phosphorylate substrates (or sites) that promote salt tolerance in their phosphorylated form, whereas PKA phosphorylates substrates (or sites) that promote salt tolerance in their dephosphorylated form. Ppz1,2 may reverse the activating phosphorylation event in the former case, and calcineurin may reverse the deactivating phosphorylation event in the latter. CKI, CKII, and Ycr101 are presumed to act on different sites, but these may be present on the same or different substrates. Activating and deactivating sites may also be present on the same or different substrates. Available data are consistent with this model, but are not sufficient to establish it. Circles represent substrates, the letter P represents phosphorylation, and shading represents an activated form that promotes salt tolerance.

IV. Substrates of CKII in *Saccharomyces cerevisiae*

Identification and analysis of CKII targets in *S. cerevisiae* provides a complementary approach to defining the physiological role of CKII in this organism. Known substrates are discussed first, and then a systematic search for potential CKII substrates is described.

A. Known Substrates

A compelling case that a protein is a physiological target of a given protein kinase generally requires three types of information. First, the protein must be a substrate of the kinase *in vitro*. Ideally, this analysis should be done with purified enzyme and substrate, and the kinetics of the reaction should be ascertained. Second, the sites of *in vitro* phosphorylation must correspond to sites that are phosphorylated *in vivo*. Third, modulation of kinase activity *in vivo* must result in an appropriate alteration in the phosphorylation of the putative target. Modulation of the activity may be achieved either by treatment with an appropriate agonist or inhibitor or by mutation or overexpression of the kinase. Unfortunately, false positive (or negative) results may be obtained at any point in the analysis (a given *in vitro* site may be phospho-

rylated by a different kinase *in vivo*, changes in phosphorylation in response to an agonist may be indirect, etc.). Nevertheless, substrates that meet all three criteria are generally taken to be physiological. A separate issue is the physiological role of the phosphorylation event, which can be pursued either biochemically or, in amenable organisms, genetically.

Because there is no known mechanism for regulation of CKII activity *in vivo*, modulation of the activity *in vivo* is restricted to the use of mutants or overexpression. Furthermore, because the enzyme is essential, only hypomorphic or conditional mutants can be used. A conditional mutant potentially eliminates activity completely but introduces a number of complications. These include the effects of the shift to restrictive conditions, the unknown rates at which different substrates are dephosphorylated, and possible indirect effects of the accompanying cell cycle arrest, loss of viability, etc. A hypomorphic allele avoids these complications but does not completely eliminate the activity. Both conditional and hypomorphic alleles are available in *S. cerevisiae* (Table I), and both have been used successfully to define *in vivo* substrates in this organism (133, 134). Strains that overexpress CKII activity are also available but have not yet been used in the identification of any endogenous substrate. Even quite dramatic increases in activity have no observable phenotype in *S. cerevisiae* (41), so it is possible that CKII activity is not normally limiting, but this problem deserves further investigation.

Only a handful of CKII substrates has been identified in *S. cerevisiae*. Two of the best characterized are topoisomerase II (Top2) and eIF-2 α (Sui2), both of which have been defined by all three of the criteria above. Topoisomerase II is phosphorylated *in vitro* by purified CKII at multiple sites, most of which are located in the C-terminal tail; these sites correlate with sites modified *in vivo*, and phosphorylation of topoisomerase II is virtually eliminated in a *cka2-8^{ts}* strain (133). Phosphorylation of topoisomerase II is required for enzyme activity *in vitro* (135) and promotes binding of the enzyme to DNA (136). Because topoisomerase II mutants die in the process of undergoing a lethal mitosis (137), failure to phosphorylate this protein could play a role in the G₂/M arrest of *cka2^{ts}* strains (7). Counter to this conclusion, deletion of the C-terminal tail has only modest effects *in vivo*, but it is possible that the tail has an inhibitory function that is antagonized by phosphorylation (135). The translation initiation factor eIF-2 α is phosphorylated by purified CKII at three sites in the C-terminal tail of the molecule, and again these sites are phosphorylated *in vivo* and hypophosphorylated in CKII mutant strains (134). Replacement of all three serines with alanine does not yield an observable phenotype in wild-type yeast but does elicit a slow-growth phenotype in *gcn2*, *gcn3*, and *gcd7* mutant backgrounds, suggesting a subtle role in translational regulation.

Additional well-characterized substrates of CKII in *S. cerevisiae* include two nucleolar proteins: Srp40, a homolog of mammalian Nopp140 (138), and Fpr3, a nucleolar-specific immunophilin (139, 139a). The function of these proteins is not well understood, but Nopp140 has been proposed to function as a chaperone in ribosome assembly and/or in nuclear transport of nucleolar components (138). If either protein is directly or indirectly required for rRNA transcription, failure to phosphorylate one or both substrates could contribute to the decrease in total RNA synthesis observed in *cka2^{ts}* mutants (7). As noted in Section III,B, Cdc37 is also a probable target of CKII in *S. cerevisiae*, and mutagenesis of the relevant site impairs the ability of the protein to suppress a temperature-sensitive *cdc37-1* allele *in vivo*.

An unusual feature of CKII is the ability to form a stable complex with a number of its substrates (84, 104). In *S. cerevisiae*, this is true of topoisomerase II (140) and Srp40 (138). The *in vivo* significance of this behavior remains to be established.

B. Predicted Substrates

The availability of the complete genomic sequence of *S. cerevisiae* provides an unprecedented opportunity to examine eukaryotic cellular function from a global perspective. Perhaps the most difficult yet exciting challenge is to understand the complex network of regulatory interactions that underlie the coordinate control and integration of cellular activity. Reversible protein phosphorylation plays a central role in this regulatory network, and protein kinases and phosphatases in fact constitute two of the largest gene families in eukaryotic organisms, including *S. cerevisiae*. A global analysis of protein phosphorylation in *S. cerevisiae* should thus be particularly informative. Although the recognition sites of protein kinases are uncomfortably short, a computer search for potential substrates offers one potential avenue toward such a global analysis.

CKII appears to be an excellent candidate for computer-based identification of substrates. Of prime importance, the CKII recognition site has been exhaustively defined through studies of both natural targets and synthetic peptides (2). This site consists of a Ser or Thr residue located in an acidic context extending from approximately -2 to +5 (relative to the modified residue) and often farther in either direction. An acidic residue at +3 is most important but is not crucial in an otherwise acidic context. Basic residues are poorly tolerated anywhere in the core -2 to +5 region, and proline is not tolerated at +1. In addition to Glu and Asp, phosphoserine and phosphothreonine can also serve as acidic determinants (141), leading to the possibility of sequential phosphorylation of complex sites containing multiple Ser and Thr residues. A key feature of the CKII recognition site is that it is highly hydrophilic. Because such a hydrophilic stretch is not readily buried in the in-

terior of a protein, potential CKII sites are expected to be solvent exposed and hence accessible for modification. Given the soluble nature and broad intracellular distribution of CKII, this fact makes it likely that any predicted target will in fact be a substrate *in vivo*. Exceptions to the latter deduction will include targets that are cotranslationally exported or sequestered within a membrane-bound organelle, but these can generally be identified by examination of the sequence and/or the literature.

To identify potential CKII substrates in the *S. cerevisiae* genome, a series of ProSite searches of the SwissProt database and BLAST searches of the *Saccharomyces* Genome Database have been conducted, both via the world wide web. The former search requires a perfect match to the query sequence but the latter does not. In spite of the tedium of searching with multiple queries, the ProSite algorithm has the advantage of providing an unambiguous identification of all targets meeting a given criterion. Table II shows pooled results obtained using the ProSite algorithm with nine queries (see footnote *a*) designed to retrieve extremely good matches to the CKII consensus. Together, these highly stringent searches retrieve a total of 107 proteins from the yeast genome. Of these, 58 are known proteins, and 49 are hypothetical open reading frames. Of the five *S. cerevisiae* proteins previously identified as CKII targets on the basis of biochemical and/or genetic studies, three (Srp40, Fpr3, and Sui2) are included in Table II, confirming the ability of the screen to detect relevant targets. The absence of the other two (Top2 and Cdc37) in fact suggests that the search criteria are too stringent. Top2 would be identified in a screen that relaxed the acidic constraint at two of the seven acidic positions, Cdc37 in a screen that relaxed this constraint at three positions. At some point, further reductions in stringency will retrieve a large number of false positives, because many mildly acidic sites may not be solvent accessible or otherwise functional. Biochemical/genetic identification of additional substrates will be required to determine where this point lies. Nevertheless, given the arguments in the preceding paragraph, it seems probable that most, if not all, of the proteins listed in Table II are substrates of CKII *in vivo* and that they are among the best, if not the best, CKII substrates in yeast. Additional support for this contention is provided by the occurrence of potential substrates that are homologous to well-characterized CKII substrates of other organisms, for example, Hhg1 (HMG1,2) and Nsr1 (nucleolin).

The list in Table II is remarkable both for what it contains and for what it does not. More than half of the 58 known proteins are involved in some aspect of gene expression, from RNA synthesis through protein degradation. Another large group is involved with the cell cycle. The remainder affect diverse functions including DNA replication/repair, cell polarity/morphogenesis, and ion transport, among others. Conspicuous by their relative absence

TABLE II
A SYSTEMATIC SEARCH FOR CKII SUBSTRATES IN *Saccharomyces cerevisiae*^a

| Protein | Function | CKII motif(s) ^b | Location ^c |
|--------------------------------------|---|----------------------------------|-----------------------|
| Chromatin structure | | | |
| Hgh1 | HMG1,2 homolog | EEEEEESEEEEDDDEEDE | 0.96 |
| DNA replication/repair | | | |
| Tel2 | Telomere length regulation protein | DCSDSDDEDENDE | 0.61 |
| Rad2 | Single-stranded endonuclease, excision repair | DFSEDEED | 0.69 |
| Srs2 | 3' → 5' DNA helicase, UV repair | DESEDEEEDQE | 0.58 |
| Transcription | | | |
| Rpa1 | RNA pol I 190-kDa subunit | EDTEDEMGE | 0.09 |
| Rpb2 | RNA pol II 140-kDa subunit | ESEDDSE | 0.13 |
| Tfc4 | TFIIIC 131-kDa subunit (TPR motifs) | EISEDDEED | 0.08 |
| Srp40 | Sup of AC40 subunit of RNA pol I and III | SSSSSSSDSSSESESEDE | 0.26 |
| Rrn3 | RNA pol I transcription initiation factor | EDDDDLDDSGDDDDDE | 0.44 |
| Spt5 | Spt4/5/6 transcription regulator | EVSDDEDEDEDEED | 0.18 |
| Spt6 | Spt4/5/6 transcription regulator | FDSSSEDEDIDEDEDE | 0.03 |
| | | EDDEFSDDEDDE | 0.14 |
| Ssl1 | C ₂ H ₂ zinc finger, sup of stem-loop protein | SESEDED | 0.02 |
| Rim1 | C ₂ H ₂ transcription factor (meiosis) | EEVDDESEIEDD | 0.89 |
| Asf1 | Regulation of silent loci | ENSDGDEEEGEEE | 0.83 |
| RNA processing/transport/degradation | | | |
| Rrp3 | Pre-rRNA processing | EESENEEESEEEEEEDDED ₈ | 0.11 |
| | | DCSDSDSDSETSSDDENDID | 0.14 |
| Nsr1 | Nucleolin homolog | DSSSSSDSSSDEEEEE | 0.21 |
| Fpr3 | Immunophilin, nucleolar | EDEIDESSEEEEE | 0.19 |
| | | AESESEDDEEDDDEDE | 0.25 |
| Nmd2 | Nonsense-mediated mRNA decay | DESEDEDEDED ₁₂ | 0.81 |
| Pan1 | Poly(A) tail exonuclease | QESDDEDEDDEE | 0.77 |

(continues)

TABLE II (Continued)

| Protein | Function | CKII motif(s) ^b | Location ^c |
|---------------------|---|----------------------------|-----------------------|
| Protein synthesis | | | |
| Sui2 | eIF-2 α | SDSEDDDEDESDDDE-cooh | 0.97 |
| Tif5 | eIF-5 | AESDDDEEDDE-cooh | 0.98 |
| Rit1 | Initiator tRNA phosphoribosyl transferase | EDTDDEEE | 0.75 |
| Rpl7 | Ribosomal protein L30 | EETEEEEEDD | 0.37 |
| Ard1 | N-Terminal acetylation (with Nat1) | EDSDEQDEND | 0.24 |
| Protein targeting | | | |
| Npl1 | Preprotein translocation | EYSTDDDETESDDDES | 0.95 |
| Pse1 | Protein secretion enhancer | IESDDTDDEEE | 0.31 |
| Uso1 | Protein transport, ER to golgi | EISSDEEDDEEDDEEEDEEE | 0.99 |
| Sec7 | Traffic from golgi | DESGSEEEEEEEEEEEEE | 0.08 |
| Adb6 | Clathrin assembly protein (β -adaptin) | EESSEDEDESEESSDDDE | 0.96 |
| Srp1 | Nuclear import (importin/karyopherin α) | ADSDEEDE | 0.12 |
| Nip1 | Nuclear import | DDSDYSSDEESDEEDG | 0.12 |
| Mft1 | Mitochondrial protein import | DEDYSDEEDE | 0.68 |
| Protein degradation | | | |
| Ubc8 | E2 ubiquitin-conjugating enzyme (24 kDa) | EEDSDSDEDMD | 0.81 |
| Cdc34 | E2 ubiquitin-conjugating enzyme (34 kDa) | EDESEDVEDVE | 0.96 |
| Ubr1 | E3 ubiquitin-protein ligase (N-terminus recognition) | DEDDSDNDSDSDE | 0.99 |
| Pim1 ^d | LON protease (mitochondrial matrix) | EEEESEELDD | 0.31 |
| Cell cycle | | | |
| Cdc68 | Required for cyclin gene expression at Start | DESEEGEDWDELE | 0.98 |
| Sic1 | CDK inhibitor | QESEDEED | 0.71 |
| Sap2 | Cell cycle-dependent association with Sit4 | DELSDDSDDEEYDNCEDD | 0.95 |
| Sap3 | Cell cycle-dependent association with Sit4 | DNEISSDEEDSEDEDEEND | 0.97 |
| Clb3 | Cyclin (G ₂ /M) | EDDGESEEEDEEDQE | 0.27 |
| Mif2 | Centromere protein (CENP-C) | DESEDEEE | 0.59 |
| Cbf2 | Centromere binding factor (chromosome segregation and movement) | EETEDLDEEEEE | 0.11 |

| | | | |
|-----------------------------|--|----------------------------|------|
| Mad1 | Spindle assembly checkpoint | NDSDDDD | 0.45 |
| Pds1 | Metaphase–anaphase transition | EESDDDEGNEDSE | 0.50 |
| Mak16 | Nuclear events of cell cycle | EENSQDEED | 0.65 |
| | | SESDSESDSDSDEE | 0.83 |
| Cell polarity/morphogenesis | | | |
| Spa2 | Mating-induced morphogenesis (shmoo) | DWSSEEEEE | 0.12 |
| | | EEEESEDDDEEEEDDFD | 0.88 |
| Metabolism | | | |
| Gac1 | Regulates glycogen synthase, PP1 regulator | DLYDSEDEDDIDE | 0.20 |
| Tps3 | Trehalose phosphate synthase/phosphatase | EDTDESDDIDSLETD | 0.23 |
| Prp1 | PRFP synthase | VDSEDEEE | 0.69 |
| Channels/pumps | | | |
| Pma2 ^e | Plasma membrane proton pump | DDSDSDEDID | 0.07 |
| Trk1 ^e | High-affinity potassium transporter | EDTDTEDDGNEDDDDEENE | 0.82 |
| Other | | | |
| Pph22 | Protein phosphatase 2A subunit (essential) | DEDTDEELED | 0.11 |
| Pas8 | Peroxisome biogenesis (Sec 18 family, ATP) | EENDESDDDEDE | 0.36 |
| Son1 | Nuclear integrity | EFDSDDDDISDDEEIEE | 0.41 |
| Prp20 | RCC1 homolog | DADDSSDDEDGD | 0.28 |
| Leo1 | Weak similarity to Npl6 | SDSDSDEDDEGE | 0.18 |
| Tsm1 | Essential | DEEEEEGESEEEEEGEEE | 0.23 |

^a*Saccharomyces cerevisiae* proteins containing at least one close match to a canonical CKII phosphorylation motif are shown. A canonical site is defined as a serine or threonine residue that is preceded by two and followed by five acidic residues. Searches were carried out using the ProSite search routine available via SwissProt on the world wide web. Specific queries were designed to return (1) a perfect match to the canonical site, (2) a single mismatch (other than lysine or arginine) at one of the acidic positions, or (3) 15 consecutive residues of glutamate, aspartate, and/or serine. All proteins with a known function retrieved in searches 1 and 2 are shown. Search 3 was culled by hand for likely CKII substrates (Asf1, Mak16, Nsr1, Sec7, Srp40, and Uso1 were identified). In addition to the 58 known proteins shown in the table, these searches also yielded 49 proteins currently classified as hypothetical. Proteins shown in bold are known CKII targets in *S. cerevisiae*.

^bFlanking regions have been extended in both the N- and C-terminal directions to include the full extent of the acidic region associated with each site. Serine and threonine residues likely to be modified by CKII are shown in bold.

^cThe location of the CKII site is expressed as a decimal fraction of protein length. The N terminus corresponds to 0 and the C terminus to 1.

^dThe localization of Pim1 in the mitochondrial matrix makes it unlikely that this protein is phosphorylated by CKII *in vivo*.

^eThe predicted membrane topology of both Pma1 and Trk1 would place the CKII recognition motif in the cytoplasm.

are the hundreds of proteins involved in intermediary metabolism. The few such enzymes that are present, such as trehalose phosphate synthase/phosphatase or PRPP synthase, are important regulatory enzymes of the pathways of which they are a part (142). Collectively, the results appear to argue against the notion that CKII functions merely to complete the covalent structure of its substrates, because it is difficult to see why so many regulatory proteins should require this modification while the majority of other proteins do not. The general picture that emerges is that CKII is involved in global regulation of higher cell functions such as gene expression and cell cycle progression.

Genetic and biochemical studies in *S. cerevisiae* identify a requirement for CKII in at least four processes: gene expression, cell cycle progression, cell polarity, and ion homeostasis (Section III). In each case, one or more predicted substrates were identified that could provide a mechanism for the observed connection to CKII. Predicted substrates relevant to gene expression include proteins involved in transcription initiation by all three classes of RNA polymerase as well as in RNA transport, processing, and degradation. Another large collection of proteins is involved in protein synthesis, targeting, and degradation. Predicted substrates relevant to cell cycle progression include the cell division cycle proteins Cdc34 and 68, the CDK inhibitor Sic1, cyclin Clb3, and a variety of other proteins. Some of these proteins are potentially relevant to the G₁ phase arrest of *cka2^{ts}* mutants (e.g., Sic1 and Cdc34), others to the G₂/M arrest of these mutants. One predicted substrate relevant to cell polarity was identified. This protein, Spa2, localizes to presumptive bud sites and the apex of developing buds (143) and is part of a complex of proteins that exhibit similar behavior (107, 108). Additional proteins of this complex include Cdc42, Bem1, and Rho1. The Trk1 potassium transporter is of obvious potential relevance to the salt-sensitive phenotype of CKII mutants. Although the rate of Na⁺ or Li⁺ import has not yet been examined in these mutants, it is tempting to speculate that phosphorylation of Trk1 by CKII is required for the conversion of this transporter to the high-affinity form that discriminates more efficiently against Na⁺ (124).

The presence of two ubiquitin-conjugating enzymes (Ubc8 and Cdc34) and one ubiquitin-protein ligase (Ubr1) among the predicted substrates deserves additional comment. Because of the role of ubiquitin-mediated proteolysis in cell cycle progression in both G₁ and G₂/M, phosphorylation of one or more of these proteins could be relevant to the cell cycle defects of *cka2^{ts}* mutants. For example, failure to phosphorylate Cdc34 could prevent ubiquitination and subsequent degradation of Sic1, thereby preventing entrance into S phase. Alternatively, ubiquitin-dependent proteolysis may be relevant to CKII expression itself. Several groups have noted that the CKII β subunit is unstable *in vivo* in the absence of the catalytic subunit (41, 144), and it has been proposed that CKII β contains a destruction box similar to

that found in cyclins (15). Degradation of free β subunit could provide a means to regulate the stoichiometry of α and β subunits or could have some deeper significance. The association of CKII with mammalian proteasomes (76) and the isolation of *SUN2* as a multicopy suppressor of *cka1^{ts}* alleles provide additional support for a possible role of CKII in ubiquitin-mediated proteolysis. Additional studies will be required to determine the functional interactions among CKII, ubiquitin-protein ligase complexes, and the 26S protease.

The availability of a global list serves to identify possible substrates mediating a known CKII response (e.g., Trk1 and ion homeostasis) as well as new systems in which CKII might play a role. Consistent with the latter, some predicted targets are relevant to functions proposed for CKII in other systems but not currently documented in *S. cerevisiae*, such as nuclear import (145).

V. The Physiological Role of CKII

Both the broad substrate specificity of CKII and the diverse spectrum of phenotypes associated with CKII mutations imply that CKII is a complex, pleiotropic activity that affects numerous cellular functions. Nevertheless, reason dictates that the enzyme must possess some coherent purpose that underlies its numerous effects. One might refer to such a coherent function as *the* physiological role of the enzyme. What might be the nature of this role? Given the phenotypes and other information currently available, one attractive possibility is that CKII serves a signal transduction function that is linked to the cell cycle.

Elucidation of signal transduction pathways in *S. cerevisiae* has proceeded rapidly in recent years (146), and many of these pathways impinge on the cell cycle machinery (147). In general, cells respond to two different types of signals, and both may affect the cell cycle: external signals provide information about the extracellular environment, including the presence or absence of essential nutrients, stressful conditions, a potential mating partner, etc., and internal signals provide information on the cell's own status. Many of the latter function as checkpoints that ensure the orderly progression of cell cycle events (148, 149). A requirement for CKII in cell cycle progression in *S. cerevisiae* has been established by studies of *cka2^{ts}* alleles (7). Assuming that CKII serves a signal transduction function linked to the cycle, what might be the nature of the putative signal to which CKII responds? Consideration of the behavior of *cka1^{ts}* alleles suggests that CKII is more likely to monitor an external environmental variable than to function as part of an internal checkpoint mechanism (Section III,C). Given the salt sensitivity of *ckb* mutants,

one possibility is that CKII has a fundamental role in ion homeostasis and that it feeds information on the ionic status of the cell to the cell cycle machinery.

Unfortunately, virtually nothing is known about possible signaling pathways linking ion homeostasis with the cell cycle. However, a potential connection between these two functions has recently been provided by the discovery that the halotolerance gene *HAL3* is allelic to *SIS2*, a multicopy suppressor of *sit4* (119). *SIT4* encodes a type 2A protein phosphatase that is essential for expression of genes encoding G_1 cyclins and other proteins during late G_1 (150), and reduced expression of these genes in *sit4* mutants is suppressed by overexpression of *SIS2* (151). Thus, overexpression of the same gene, *SIS2/HAL3*, affects both cyclin gene expression in G_1 and salt tolerance. Remarkably, a potential connection between CKII and Sit4 is also available, in the form of *SAP2* and *SAP3*. These proteins physically associate with Sit4 in a cell cycle-dependent manner (152), and both proteins also contain dramatic CKII motifs (Table II). Additional studies will be required to address the significance of these coincidences, but they may reflect an important connection between CKII (a protein kinase) and Sit4 (a protein phosphatase) in linking G_1 cell cycle progression to the ionic status of the cell.

A role for CKII in monitoring the ionic milieu would be consistent with the fact that CKII function is strongly ionic strength dependent. The self-association of CKII to form oligomers and polymers is salt dependent (22–24), as is the K_m for at least some protein substrates (22). Either of these facts could provide a direct mechanism for sensing intracellular ionic strength. Moreover, CKII activity and polymerization are regulated by polyamines, which are both growth associated and important for the interaction of polycations (basic proteins) and polyanions (nucleic acids) in transcription, translation, and other processes (153). CKII is inhibited by polyanions and activated by polybasic compounds, and the latter effect is modulated by polyamines (25). The enzyme thus appears ideally suited to monitoring and/or regulating the intracellular ionic environment.

Other phenotypes exhibited by *S. cerevisiae* CKII mutants are not readily explained by this model but might be secondary to effects on the cell cycle. In particular, the altered cell polarity of temperature-sensitive *cka1* and *cka2* mutants may reflect the intimate connections that exist between cell cycle progression and morphogenesis in budding yeast (108). Interestingly, calcineurin, in addition to its role in ion homeostasis, also affects cell polarity and cell morphology in *S. cerevisiae* (129).

It must be emphasized that the above scenario is highly speculative. An alternative view, equally consistent with the available facts, is that CKII affects some global process, such as gene expression or chromatin structure. In this model, the diverse phenotypes associated with CKII mutants are ex-

plained indirectly, via alterations in the expression of genes involved in different and possibly unrelated processes.

Obviously much remains to be done. A major disappointment of the genetic studies to date is that they have shed no light on the crucial issue of CKII regulation. Is the enzyme constitutively active *in vivo*, or is it inhibited in some manner? Is there a physiological equivalent of polybasic activator? Obviously, multicopy suppression of a temperature-sensitive allele is unlikely to identify a CKII inhibitor. Moreover, because CKII almost certainly has multiple essential targets, this approach is also unlikely to identify downstream functions. Synthetic lethal or other genetic screens, as well as two-hybrid screens for proteins that physically interact with the catalytic or regulatory subunits, should circumvent some of these limitations and may yield novel interactions.

A final issue concerns the relevance of the results obtained in *S. cerevisiae* to other systems; in other words, is the physiological role of CKII identical in different organisms? Based on existing precedents, such as Cdk1, one can anticipate that the fundamental role will be the same, but that the details may vary considerably. Certainly this appears to be the case for the role of CKII in cell polarity in *S. cerevisiae* and *S. pombe* (Section III,C). Parallel studies of the physiological role of CKII in different systems are thus essential.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to past and present members of my laboratory who have contributed to the work described here: Ramesh Padmanabha, Joan Chen, Amit Saxena, David Hanna, Rick McCann, Mark Birnbaum, Ashok Bidwai, Craig Reed, Asokan Rethinaswamy, Sricharan Bandhakavi, Wenfan Zhao, and Kirsten Tenney. Research from my laboratory was supported by National Institutes of Health Grant GM33237 and American Cancer Society Grant VM-19.

REFERENCES

1. L. A. Pinna, *Cell. Mol. Biol. Res.* **40**, 383 (1994).
2. F. Meggio, O. Marin, and L. A. Pinna, *Cell. Mol. Biol. Res.* **40**, 401 (1994).
3. R. Padmanabha, J. L. P. Chen-Wu, D. E. Hanna, and C. V. C. Glover, *Mol. Cell. Biol.* **10**, 4089 (1990).
4. V. Snell and P. Nurse, *EMBO J.* **13**, 2066 (1994).
5. U. Kikkawa, S. K. O. Mann, R. A. Firtel, and T. Hunter, *Mol. Cell. Biol.* **12**, 5711 (1992).
6. A. P. Bidwai, J. C. Reed, and C. V. C. Glover, *J. Biol. Chem.* **270**, 10395 (1995).
7. D. E. Hanna, A. Rethinaswamy, and C. V. C. Glover, *J. Biol. Chem.* **270**, 25905 (1995).
8. I. Roussou and G. Draetta, *Mol. Cell. Biol.* **14**, 576 (1994).

9. L. Ulloa, J. Diaz-Nido, and J. Avila, *EMBO J.* **12**, 1633 (1993).
10. R. Pepperkok, P. Lorenz, W. Ansorge, and W. Pyerin, *J. Biol. Chem.* **269**, 6986 (1994).
11. D. C. Seldin and P. Leder, *Science* **267**, 894 (1995).
12. L. A. Pinna, *Biochim. Biophys. Acta* **1054**, 267 (1990).
13. P. T. Tuazon and J. A. Traugh, in "Advances in Second Messenger and Phosphoprotein Research" (P. Greengard and G. A. Robinson, eds.), p. 123. Raven Press, New York, 1991.
14. O.-G. Issinger, *Pharmacol. Theor.* **59**, 1 (1993).
15. J. E. Allende and C. C. Allende, *FASEB J.* **9**, 313 (1995).
16. C. V. C. Glover, A. P. Bidwai, and J. C. Reed, *Cell. Mol. Biol. Res.* **40**, 481 (1994).
17. R. Padmanabha and C. V. C. Glover, *J. Biol. Chem.* **262**, 1829 (1987).
18. M. A. Collinge and J. C. Walker, *Plant Mol. Biol.* **25**, 649 (1994).
- 18a. A. Bidwai, W. Zhao, and C. Glover, in preparation (1997).
19. S. K. Hanks and A. M. Quinn, *Meth. Enzymol.* **200**, 38 (1991).
20. R. D. Gietz, K. C. Graham, and D. W. Litchfield, *J. Biol. Chem.* **270**, 13017 (1995).
21. B. Boldyreff, U. Mietens, and O.-G. Issinger, *FEBS Lett.* **379**, 153 (1996).
22. C. V. C. Glover, *J. Biol. Chem.* **261**, 14349 (1986).
23. M. D. Mamrack, *Mol. Cell. Biochem.* **85**, 147 (1989).
24. E. Valero, S. De Bonis, O. Filhol, R. H. Wade, J. Langowski, E. M. Chambaz, and C. Cochet, *J. Biol. Chem.* **270**, 8345 (1995).
25. S. Sarno, O. Marin, F. Meggio, and L. A. Pinna, *Biochem. Biophys. Res. Commun.* **194**, 83 (1993).
26. D. W. Litchfield, G. Dobrowolska, and E. G. Krebs, *Cell. Mol. Biol. Res.* **40**, 373 (1994).
27. W. Krek, G. Maridor, and E. Nigg, *J. Cell. Biol.* **116**, 43 (1991).
28. N. Chester, I. J. Yu, and D. R. Marshak, *J. Biol. Chem.* **270**, 7501 (1995).
29. M. Pffaf and F. A. Anderer, *Biochim. Biophys. Acta* **969**, 100 (1988).
30. P. Belenguer, V. Baldin, C. Mathieu, H. Prats, M. Bensaid, G. Bouche, and F. Amalric, *Nucleic Acids Res.* **17**, 6625 (1989).
31. H. Meisner and M. P. Czech, *Curr. Opin. Cell Biol.* **3**, 474 (1991).
32. D. W. Litchfield and B. Luscher, *Mol. Cell. Biochem.* **127/128**, 187 (1993).
33. R. Voit, A. Schnapp, A. Kuhn, H. Rosenbauer, P. Hirschmann, H. G. Stunnenberg, and I. Grummt, *EMBO J.* **11**, 2211 (1992).
34. D. A. Jans, *Biochem. J.* **311**, 705 (1995).
35. R. Voit, A. Kuhn, E. E. Sander, and I. Grummt, *Nucleic Acids Res.* **23**, 2593 (1995).
36. R. A. Heller-Harrison and M. P. Czech, *J. Biol. Chem.* **266**, 14435 (1991).
37. R. Pepperkok, P. Lorenz, A. Jakobi, W. Ansorge, and W. Pyerin, *Exp. Cell Res.* **197**, 245 (1991).
38. P. Lorenz, R. Pepperkok, W. Ansorge, and W. Pyerin, *J. Biol. Chem.* **268**, 2733 (1993).
39. A. P. Bidwai, J. C. Reed, and C. V. C. Glover, *Arch. Biochem. Biophys.* **309**, 348 (1994).
40. J. C. Reed, A. P. Bidwai, and C. V. C. Glover, *J. Biol. Chem.* **269**, 18192 (1994).
41. A. P. Bidwai, D. E. Hanna, and C. V. C. Glover, *J. Biol. Chem.* **267**, 18790 (1992).
42. J. L.-P. Chen-Wu, R. Padmanabha, and C. V. C. Glover, *Mol. Cell. Biol.* **8**, 4981 (1988).
43. M. Veron, E. Radzio-Andzelm, I. Tsigelny, L. F. Ten Eyck, and S. S. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10618 (1993).
44. B. Boldyreff, F. Meggio, L. A. Pinna, and O.-G. Issinger, *Biochemistry* **32**, 12672 (1993).
45. B. Boldyreff, F. Meggio, L. A. Pinna, and O.-G. Issinger, *J. Biol. Chem.* **269**, 4827 (1994).
46. F. Meggio, B. Boldyreff, O.-G. Issinger, and L. A. Pinna, *Biochemistry* **33**, 4336 (1994).
47. M. J. Birnbaum and C. V. C. Glover, *Biochem. Biophys. Res. Commun.* **181**, 524 (1991).
48. G. B. Calleja, in "The Yeasts" (A. H. Rose and J. S. Harrison, eds.), p. 165. Academic Press, London, 1987.

49. D. R. Knighton, J. Zheng, L. F. Ten-Eyck, V. A. Ashford, N. H. Xuong, S. S. Taylor, and J. M. Sowadski, *Science* **253**, 407 (1991).
- 49a. A. Rethinaswamy, M. Birnbaum, and C. Glover, submitted (1997).
- 49b. R. McCann, D. Hanna, S. Bandhakari, and C. Glover, submitted (1997).
50. S. I. Reed, *Genetics* **95**, 561 (1980).
51. J. Ferguson, J.-Y. Ho, T. A. Peterson, and S. I. Reed, *Nucleic Acids Res.* **14**, 6681 (1986).
52. J.-G. Valay, M. Simon, M.-F. Dubois, O. Bensaude, C. Facca, and G. Faye, *J. Mol. Biol.* **249**, 535 (1995).
53. F. Boschelli, *Mol. Cell. Biol.* **13**, 5112 (1993).
54. T. Cutforth and C. M. Rubin, *Cell* **77**, 1027 (1994).
55. W. B. Pratt, *J. Biol. Chem.* **271**, 21455 (1993).
56. J. K. Owens-Grillo, M. J. Czar, K. A. Hutchison, K. Hoffmann, G. H. Perdew, and W. B. Pratt, *J. Biol. Chem.* **271**, 13468 (1996).
57. L. Stepanova, X. Leng, S. B. Parker, and J. W. Harper, *Genes Dev.* **10**, 1491 (1996).
58. K. Dai, R. Kobayashi, and D. Beach, *J. Biol. Chem.* **271**, 22030 (1996).
59. E. Bi and J. R. Pringle, *Mol. Cell. Biol.* **16**, 5264 (1996).
60. B. Schwer and S. Shuman, *Gene Express.* **5**, 331 (1996).
61. D. J. Lew and S. I. Reed, *J. Cell Biol.* **120**, 1305 (1993).
62. F. Cvrckova, C. De Virgilio, E. Manser, J. R. Pringle, and K. Nasmyth, *Genes Dev.* **9**, 1817 (1995).
63. M.-N. Simon, C. De Virgilio, B. Souza, J. R. Pringle, A. Abo, and S. I. Reed, *Nature (London)* **376**, 702 (1995).
64. C. Lurquin, A. Van Pel, B. Mariame, E. De Plaen, J.-P. Szikora, C. Janssens, M. J. Reddehase, J. Lejeune, and T. Boon, *Cell* **58**, 293 (1989).
65. E. S. Pentz, B. C. Black, and T. R. F. Wright, *Genetics* **112**, 823 (1986).
66. E. S. Pentz and T. R. F. Wright, *Gene* **103**, 239 (1991).
67. G. N. DeMartino, C. R. Moomaw, O. P. Zagnitko, R. J. Proske, M. Chu-Ping, S. J. Afendis, J. C. Swaffield, and C. A. Slaughter, *J. Biol. Chem.* **269**, 20878 (1994).
68. W. Dubiel, K. Ferrell, and M. Rechsteiner, *Mol. Biol. Rep.* **21**, 27 (1995).
69. A. Ciechanover, *Cell* **79**, 13 (1994).
70. M. Hochstrasser, *Curr. Opin. Cell Biol.* **7**, 215 (1995).
71. M. Kawamura, K.-I. Kominami, J. Takeuchi, and A. Toh-E, *Mol. Gen. Genet.* **251**, 146 (1996).
72. K.-I. Kominami, G. N. Demartino, C. R. Moomaw, C. A. Slaughter, N. Shimbara, M. Fijumuro, H. Yokosawa, H. Hisamatsu, N. Tanahashi, Y. Shimizu, K. Tanaka, and A. Toh-E, *EMBO J.* **14**, 3105 (1995).
73. M. Ghislain, A. Udvardy, and C. Mann, *Nature (London)* **366**, 358 (1993).
74. E. Schwob, T. Bohm, M. D. Mendenhall, and K. Nasmyth, *Cell* **79**, 233 (1994).
75. S. Irniger, S. Piatti, C. Michaelis, and K. Nasmyth, *Cell* **81**, 269 (1995).
76. R. L. Ludemann, K. M. Lerea, and J. D. Etlinger, *J. Biol. Chem.* **268**, 17413 (1993).
77. A. W. R. H. Teunissen, and H. Y. Steensma, *Yeast* **11**, 1001 (1995).
78. E. J. Louis, *Yeast* **11**, 1553 (1995).
79. S. Y. Roth, *Curr. Opin. Genet. Dev.* **5**, 168 (1995).
80. D. G. Edmondson, M. M. Smith, and S. Y. Roth, *Genes Dev.* **10**, 1247 (1996).
81. P. N. Lipke and C. Hull-Pillsbury, *J. Bacteriol.* **159**, 797 (1984).
82. J. Schultz and M. Carlson, *Mol. Cell. Biol.* **7**, 3637 (1987).
83. A. W. R. H. Teunissen, J. A. Van Den Berg, and H. Yde-Steensma, *Yeast* **11**, 435 (1995).
84. D. Li, G. Dobrowolska, and E. G. Krebs, *J. Biol. Chem.* **271**, 15662 (1996).
85. D. J. Hockman and M. C. Schultz, *Mol. Cell. Biol.* **16**, 892 (1996).

86. K. Nasmyth, *Curr. Opin. Cell Biol.* **5**, 166 (1993).
87. D. O. Morgan, *Nature (London)* **374**, 131 (1995).
88. P. Kaldis, A. Sutton, and M. J. Solomon, *Cell* **86**, 553 (1996).
89. J.-Y. Thuret, J.-G. Valay, G. Faye, and C. Mann, *Cell* **86**, 565 (1996).
90. E. Schwob and K. Nasmyth, *Genes Dev.* **7**, 1160 (1993).
91. A. R. Willems, S. Lankester, E. E. Patton, K. L. Craig, T. F. Nason, N. Mathias, R. Kobayashi, C. Wittenberg, and M. Tyers, *Cell* **86**, 453 (1996).
92. B. L. Schneider, Q.-H. Yang, and A. B. Futcher, *Science* **272**, 560 (1996).
93. J. R. Lamb, W. A. Michaud, R. S. Sikorski, and P. A. Hieter, *EMBO J.* **13**, 4321 (1994).
94. R. N. Booher, R. J. Deshaies, and M. W. Kirschner, *EMBO J.* **12**, 3417 (1993).
95. P. Russell, S. Moreno, and S. I. Reed, *Cell* **57**, 295 (1989).
96. X.-J. Ma, Q. Lu, and M. Grunstein, *Genes Dev.* **10**, 1327 (1996).
97. W. G. Dunphy, *Trends Cell Biol.* **4**, 202 (1994).
98. D. J. Lew and S. I. Reed, *J. Cell Biol.* **129**, 739 (1995).
99. B. Dey, J. J. Lightbody, and F. Boschelli, *Mol. Biol. Cell* **7**, 1405 (1996).
100. M. R. Gerber, A. Farrell, R. J. Deshaies, I. Herskowitz, and D. O. Morgan, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4651 (1995).
101. S. I. Reed and C. Wittenberg, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5697 (1990).
102. U. Surana, H. Robitsch, C. Price, T. Schuster, I. Fitch, A. B. Futcher, and K. Nasmyth, *Cell* **65**, 145 (1991).
103. Y. Miyata and I. Yahara, *J. Biol. Chem.* **267**, 7042 (1992).
104. Y. Miyata and I. Yahara, *Biochemistry* **34**, 8123 (1995).
105. S. P. Lees-Miller and C. W. Anderson, *J. Biol. Chem.* **264**, 2431 (1989).
106. A. Murray and T. Hunt, "The Cell Cycle." W. H. Freeman, New York, 1993.
107. J. Chant, *Trends Genet.* **10**, 328 (1994).
108. D. J. Lew and S. I. Reed, *Curr. Opin. Genet. Dev.* **5**, 17 (1995).
109. A. Bretscher, B. Drees, E. Harsay, D. Schott, and T. Wang, *J. Cell Biol.* **126**, 821 (1994).
110. M. D. Welch, D. A. Holtzman, and D. G. Drubin, *Curr. Opin. Cell Biol.* **6**, 110 (1994).
111. B. F. Sloat, A. E. M. Adams, and J. R. Pringle, *J. Cell Biol.* **89**, 395 (1981).
112. A. E. M. Adams, D. I. Johnson, R. M. Longnecker, B. F. Sloat, and J. R. Pringle, *J. Cell Biol.* **111**, 131 (1990).
113. A. Bender and J. R. Pringle, *Mol. Cell. Biol.* **11**, 1295 (1991).
114. J. Peterson, Y. Zheng, L. Bender, A. Myers, R. Cerione, and A. Bender, *J. Cell Biol.* **127**, 1395 (1994).
115. Y. Matsui and A. Toh-E, *Mol. Cell. Biol.* **12**, 5690 (1992).
116. J. Imai, A. Toh-E, and Y. Matsui, *Genetics* **142**, 359 (1996).
117. L. Bender, H. S. Lo, H. Lee, V. Kokojan, J. Peterson, and A. Bender, *J. Cell Biol.* **133**, 879 (1996).
118. I. Mendoza, F. Rubio, A. Rodriguez-Navarro, and J. M. Pardo, *J. Biol. Chem.* **269**, 8792 (1994).
119. A. Ferrando, S. J. Kron, G. Rios, G. A. Fink, and R. Serrano, *Mol. Cell. Biol.* **15**, 5470 (1995).
120. C. H. Ko and R. F. Gaber, *Mol. Cell. Biol.* **11**, 4266 (1991).
121. J. Ramos, R. Alijo, R. Haro, and A. Rodriguez-Navarro, *J. Bacteriol.* **176**, 249 (1994).
122. R. Haro, B. Garciadeblas, and A. Rodriguez-Navarro, *FEBS Lett.* **291**, 189 (1991).
123. B. Garciadeblas, F. Rubio, F. J. Quintero, M. A. Banuelos, R. Haro, and A. Rodriguez-Navarro, *Mol. Gen. Genet.* **236**, 363 (1993).
124. R. Haro, M. A. Banuelos, F. J. Quintero, F. Rubio, and A. Rodriguez-Navarro, *Physiol. Plant.* **89**, 868 (1993).
125. L. C. Robinson, E. J. A. Hubbard, P. R. Graves, A. DePaoli-Roach, P. J. Roach, C. Kung,

- D. W. Haas, C. H. Hagedorn, M. Goebel, M. R. Culbertson, and M. Carlson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 28 (1994).
126. J. Skala, B. Purnelle, M. Crouzet, M. Aigle, and A. Goffeau, *Yeast* **7**, 651 (1991).
127. D. Hirata, S. Harada, H. Namba, and T. Miyakawa, *Mol. Gen. Genet.* **249**, 257 (1995).
128. T. Nakamura, Y. Liu, D. Hirata, H. Namba, S. Harada, T. Hirokawa, and T. Miyakawa, *EMBO J.* **12**, 4063 (1993).
129. I. Mendoza, F. J. Quintero, R. A. Bressan, P. M. Hasegawa, and J. M. Pardo, *J. Biol. Chem.* **271**, 23061 (1996).
130. F. Posas, M. Camps, and J. Arino, *J. Biol. Chem.* **270**, 13036 (1995).
131. R. Gaxiola, I. F. Larrinoa, J. M. Villalba, and R. Serrano, *EMBO J.* **11**, 3157 (1992).
132. R. Serrano, *Int. Rev. Cytol.* **165**, 1 (1996).
133. M. Cardenas, Q. Dang, C. V. C. Clover, and S. M. Gasser, *EMBO J.* **11**, 1785 (1992).
134. L. Feng, H. Yoon, and T. F. Donahue, *Mol. Cell. Biol.* **14**, 5139 (1994).
135. M. E. Cardenas and S. M. Gasser, *J. Cell Sci.* **104**, 219 (1993).
136. Q. Dang, G.-C. Alghisi, and S. M. Gasser, *J. Mol. Biol.* **243**, 10 (1994).
137. C. Holm, T. Goto, J. C. Wang, and D. Botstein, *Cell* **41**, 553 (1985).
138. U. T. Meier, *J. Biol. Chem.* **271**, 19376 (1996).
139. L. K. Wilson, B. M. Benton, S. Zhou, J. Thorner, and G. S. Martin, *J. Biol. Chem.* **270**, 25185 (1995).
- 139a. L. K. Wilson, N. Dhillon, J. Thorner, and G. S. Martin, *J. Biol. Chem.* **272**, 12961 (1997).
140. M. E. Cardenas, R. Walter, D. Hanna, and S. M. Gasser, *J. Cell Sci.* **104**, 533 (1993).
141. F. Meggio, J. W. Perich, R. B. Johns, and L. A. Pinna, *FEBS Lett.* **237**, 225 (1988).
142. J. M. Thevelein and Hohmann, *Trends Biosci.* **20**, 3 (1995).
143. M. Snyder, S. Gehrung, and B. D. Page, *J. Cell Biol.* **114**, 515 (1991).
144. B. Luscher and D. W. Litchfield, *Eur. J. Biochem.* **220**, 521 (1994).
145. D. A. Jans and P. Jans, *Oncogene* **9**, 2961 (1994).
146. J. M. Thevelein, *Yeast* **10**, 1753 (1994).
147. C. Wittenberg and S. I. Reed, *Curr. Opin. Cell Biol.* **8**, 223 (1996).
148. L. H. Hartwell and T. A. Weinert, *Science* **246**, 629 (1989).
149. A. W. Murray, *Curr. Opin. Genet. Dev.* **5**, 5 (1995).
150. M. J. Fernandez-Sarabia, A. Sulton, T. Zhong, and K. T. Arndt, *Genes Dev.* **6**, 2417 (1992).
151. C. J. Di Como, R. Bose, and K. T. Arndt, *Genetics* **139**, 95 (1995).
152. M. M. Luke, F. D. Setta, C. J. Di Como, H. Sugimoto, R. Kobayashi, and K. T. Arndt, *Mol. Cell. Biol.* **16**, 2744 (1996).
153. C. W. Tabor and T. H., *Annu. Rev. Biochem.* **53**, 749 (1984).

This Page Intentionally Left Blank

The Heparan Sulfate–Fibroblast Growth Factor Family: Diversity of Structure and Function

WALLACE L. MCKEEHAN,
FEN WANG, AND MIKIO KAN

*Center for Cancer Biology and Nutrition
Albert B. Alkek Institute of Biosciences
and Technology
Department of Biochemistry and
Biophysics
Texas A & M University
Houston, Texas 77030*

| | |
|---|-----|
| I. Diversity and Ubiquity of the Fibroblast Growth Factor Family | 136 |
| A. Three Types of Subunits Compose the FGF Receptor Signal Transduction Complex | 136 |
| B. Intrinsic Regulator in Embryonic Development and Tissue Homeostasis in the Adult | 136 |
| II. Diversity of Structure and Function | 142 |
| A. The FGF Polypeptides | 142 |
| B. The FGF Receptor Tyrosine Kinase | 143 |
| C. The FGF Receptor Heparan Sulfate | 151 |
| III. Structure, Assembly, and Control of the FGF Receptor Complex | 155 |
| A. Models of the FGF Receptor Complex | 155 |
| IV. The FGF Family in Liver Growth and Function | 164 |
| A. Analysis of FGF in Hepatocyte-like Hepatoma Cells | 164 |
| B. Role of the FGF Signal Transduction System in Normal Hepatocytes and Resting and Regenerating Liver | 166 |
| V. The FGF Family in Prostate and Prostate Tumors | 168 |
| A. Andromedic Stromal-to-Epithelial Activities of the FGF Family | 168 |
| B. Switch from Homeostatic Paracrine Regulator to Autocrine Autonomy during Prostate Tumor Progression | 173 |
| References | 173 |

The fibroblast growth factor (FGF) receptor complex is a ubiquitous regulator of development and adult tissue homeostasis that bridges the peri-cellular matrix and the intracellular environment. Diverse members of the FGF polypeptide family, the FGF receptor tyrosine kinase (FGFR TK) family and the FGF receptor heparan sulfate proteoglycan (FGFRHS) family combine to result in active and specific FGFR signal transduction complexes. Regulated alternate splicing and combination of variant subdomains give rise to diversity of FGFR TK monomers. Divalent cations cooperate with the FGFRHS to conformationally restrict FGFR TK

trans-phosphorylation, which causes depression of kinase activity and facilitates appropriate activation of the FGFR complex by FGF. Diffusional and conformational molecular models of the oligomeric FGFR complex are presented to explain how different point mutations in the FGFR TK commonly cause craniofacial and skeletal abnormalities of graded severity by graded increases in FGF-independent activity of total FGFR complexes. The role of the FGF family in liver growth and function and in prostate tumor progression is discussed. © 1998 Academic Press

I. Diversity and Ubiquity of the Fibroblast Growth Factor Family

A. Three Types of Subunits Compose the FGF Receptor Signal Transduction Complex

The fibroblast growth factor (FGF) family consists of three branches; these branches comprise the three components that participate in the FGF receptor (FGFR) signal transduction complex. These are regulatory polypeptides (FGF), transmembrane tyrosine kinases (FGFR TK), and heparan sulfate proteoglycans (FGFR HS). There are 10 genetically distinct genes currently known to encode FGF polypeptides. These polypeptides have been demonstrated to participate in assembly of FGFR complexes (1, 2), and have by consensus been named FGF-1 to FGF-10, although other acronyms are still used for individual members of the family (Figs. 1 and 2). Recently, four more members of the polypeptide family, comprising a subfamily based on homology and unique exon-intron structure, have been reported, but their bioactivity has not been evaluated (3-7) (Fig. 3). (*Note:* All amino acid and nucleic acid sequences discussed herein were obtained from GenBank.)

The FGFR TK family currently consists of the products of four genes, some of which give rise to an extremely large number of products by combinatorial alternate splicing at transcription, and possibly by posttranslational modification (Fig. 4). The FGFR HS branch of the family, which is mimicked by heparin, is the least characterized in terms of its specificity and diversity.

As described below, the FGFR HS component is expected to exhibit specificity and diversity equal to or exceeding that of the FGF polypeptide and FGFR TK families, and integrates the FGFR complex with the pericellular tissue matrix.

B. Intrinsic Regulator in Embryonic Development and Tissue Homeostasis in the Adult

Studies to date show that the FGF signal transduction system is ubiquitous in tissues of both the developing and the adult organism. In contrast to

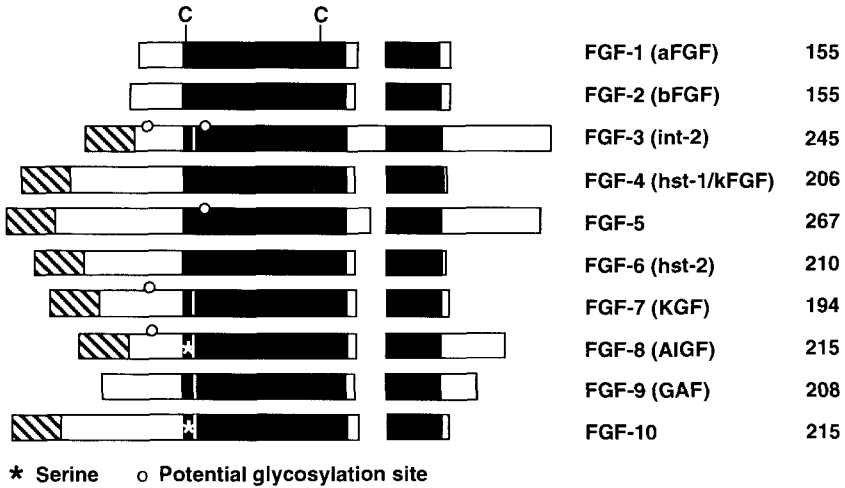


FIG. 1. Schematic of homologous and unique sequence domains of 10 members of the FGF polypeptide family. Other acronyms are in parentheses. The number of amino acid residues in the full-length product is indicated on the right.

signal transduction systems in which a component involved in the transduction complex arrives external to tissues, components of the FGF regulatory system originate and often preexist within tissues. It is unlikely that any of the three components of the FGF system are involved in significant long-range diffusion and, therefore, in conventional equilibrium reactions, except when complexed or chaperoned by binding factors in the tissue environment. The involvement of pericellular HS proteoglycan as an integral component of the FGFR system and the local origin of FGF and FGFRTK in tissues suggest that the FGFR signal transduction complex is an integrated part of the tissue or extracellular matrix, with regard to its external domain, and a part of the intracellular environment of cells, with respect to its intracellular domain. The fact that components of the FGFR complex often preexist in tissues suggests that restriction of their activity and access to each other are major features of their regulation, in addition to overall synthesis and metabolism.

The FGF system mediates communication among cells to maintain homeostasis within tissues. A wide variety of bioactive components in the external environment impinge on the FGF system and, therefore, the system serves as a monitor both of the external environment of tissues at the level of the organism and of the local extracellular environment at the tissue and cell level. As a consequence, there are likely few external environmental condi-

| | | | | | | |
|-------|--|---------|-------|-------------------------|---------------------|----------|
| FGF-1 | TPNEECLFLEERLEENHYNTYISKKHAEKN-----W | FVGLKKN | GSC | KRG-- | PRTHYGOKAIL-FLPLPVS | aFGF |
| FGF-2 | CVTDECFEERLESNNYNTYRSRKYTS-----W | YVALKRT | GQY | KLK-- | SKTGPQOKAIL-FLPMSAK | bFGF |
| FGF-3 | HYSAECEFFVERIHELGYNTYASRLYRTVSSTPGARRQPSAERLWVYVSVNGKGRPRRGFKTR-RT-QKSSL-FLPRVLD | | | | | INT-2 |
| FGF-4 | FFTDECTFKEILLPNNYNAYESYKYPG-----M | FIALSK | NGKTK | KKG-- | NRVSPTMKVTH-FLPRL | HST/kFGF |
| FGF-5 | KFTDDCKFRERFOENSYNTYASAIHRTEKTGRE-----W | YVALNKR | GKAKR | GCSPRVK | QPHISTH-FLPRFKQ | FGF-5 |
| FGF-6 | SFQEECKFRETLLPNNYNAYESDLYQG-----T | YIALSK | YGRV | KRG-- | SKVSPIMTVTH-FLPRI | HST-2 |
| FGF-7 | ECNEDCNFKELILENHYNTYASAKWTHNGGE-----M | FVALNQ | KGIP | VRGK | TK-KE-QKTAH-FLPMAIT | KGF |
| FGF-8 | GKGKDCVFTEIVLENNYNTALQNAKYEG-----W | YMAFTR | KGRPR | KSKTR | QHQREVHFMKRLPRGHH | AIGF |
| FGF-9 | KLTOECVFREQFEENWYNTYSSNLYKHVDGTGRR-----Y | VVALNK | DGT | PREG-- | TRTKRHOKFTH-FLPRPPV | GAF |
| FGF10 | EFNNDCKLKERIEENGYNTYASFNWQHNGRQ-----M | YVALNG | KGAPR | RGOKTR-RK-NTSAH-FLPMVVH | | FGF10 |
| FGF-1 | SD | | | | 155 | aFGF |
| FGF-2 | S | | | | 155 | bFGF |
| FGF-3 | HRDHMVRQLQSGLPRPPGKGVQPRRRRQKQSPDNLEPSHVQASRLGSQLEASAH | | | | 239 | INT-2 |
| FGF-4 | | | | | 206 | HST/kFGF |
| FGF-5 | SEQPELSFTVTVPEKKNPPSPIKSKIPLSAPRKNNTNSVKYRLKFRFG | | | | 268 | FGF-5 |
| FGF-6 | | | | | 198 | HST-2 |
| FGF-7 | | | | | 194 | KGF |
| FGF-8 | TTEQLSRFEFLNYPPFTRSLRGSQRTWAPEPR | | | | 215 | AIGF |
| FGF-9 | DPDKVPELYKDILSQS | | | | 208 | GAF |
| FGF10 | S | | | | 215 | FGF10 |

FIG. 2. Amino acid sequence of FGF-1 through FGF-10. Other acronyms are indicated on the right and the number of total residues is given in the bottom panel. Underscored residues are those that are conserved with respect to property (e.g., charge, hydrophobicity, serine/threonine, and tyrosine/phenylalanine) in at least six FGFs. Coded bars over the sequence indicate the β -strand assignments from different studies; see Ref. 3 (Eriksson *et al.*), Ref. 4 (Zhang *et al.*), Ref. 5 (Moy *et al.*), and Ref. 6 (Zhu *et al.*).

| | | | |
|-------|---|----------------------|-------|
| FGF-1 | | | |
| FGF-2 | | MAAGSITTLPALPEDGGSGA | FGF-2 |
| FHF-1 | MAAAIASSLIRQKRQARESNSDRVSASKRRSSPSKDGRLSLCERHVLGVFSKVRFCSGRKRVPVRR | | FHF-1 |
| FHF-2 | MAAAIASSLIRQKRQAREREKSNACKCVSSPSKGKTSCKDNKLVNFSRVKLVFGSKKRRRR | | FHF-2 |
| FHF-3 | MAALASSLIRQKREVREPGGSRPVSAQRRVCPRGTKSLCQQLLLSKVRLCGGRPARPDR | | FHF-3 |
| FHF-4 | MAAAIASGLIRQKRQAREQHWDRPSASRRRSSPSKNRGLCNGNLVDIFSKVRIFGKKRRLRR | | FHF-4 |
| FGF-7 | MHKWILTWLPTLLYRSCFHIIICLVGTISLACNDMTPEQMATNVNCSSEPERHTRSYPDQ | | FGF-7 |
| FGF-1 | <u>LP-PGNYK-KPKLLYCSNGGHFLRILPDGTVDGTDRDRSDQHIQLQLSAESVG--EVYIKSTETGQYLAMDTDGLLYGSQ</u> | | FGF-1 |
| FGF-2 | <u>FP-PGHFK-DPKRLYCKNGGFFLRIHDPGRVDGVREKSDPHIKLQLQAERG--VVSIKGVCANRYLAMKEDGRLLASK</u> | | FGF-2 |
| FHF-1 | <u>RPEPQ-LKGIVTRLFSQQG-YFLQMHPDGTIDGTDKENDSYTLFNLIP--VGLRVVAIQGVKASLYVAMNGEGYLYSSD</u> | | FHF-1 |
| FHF-2 | <u>RPEPQ-LKGIVTKLYSRQG-YHLQLQADGTIDGTDKEDSTYTLFNLIP--VGLRVVAIQGVQTKLYLAMNSEGYLYTSE</u> | | FHF-2 |
| FHF-3 | <u>GPEPQ-LKGIVTKLFCRQG-FYLQANPDGSIQGTPEDTSSFTHFNLIP--VGLRVVTIQSAKLGHYAMNAEGLLYSSP</u> | | FHF-3 |
| FHF-4 | <u>-QDPQ-LKGIVTRLYCRQG-YLQMHDPDGAIDGTDKDDSTNSTLFNLIP--VGLRVVAIQGVKTGLYIAMNGEGYLYPSE</u> | | FHF-4 |
| FGF-7 | <u>YMEGGDIR--VRRLF CRTQ-WYLRIDKRGKVKGTQEMKNYINIMEIRTVAVGI--VAIKGVESEFYLAMNKEGKLYAKK</u> | | FGF-7 |
| FGF-1 | <u>TPNEECLFLERLEENHYNTYISKKAHEKN---WFVGLKKNCSCKRGPRTHYGQK-AILFLPLPVS</u> | | FGF-1 |
| FGF-2 | <u>CVTDECFEERLESNNYNTYRSRKYTS-----WYVALKRTGOYKLGSKTGPGQK-AILFLPMSAK</u> | | FGF-2 |
| FHF-1 | <u>VFTPECKFKESVFENYVVIYSSLYRQOESGRAWFLGLNKEGQIMKGNRVKKT-KPSSHFPKPIE</u> | | FHF-1 |
| FHF-2 | <u>LFTPECKFKESVFENYVVTYSSMLYRQQSGRGWYGLNKEGEMKGNHVKKK-KPAAHFLPKPLK</u> | | FHF-2 |
| FHF-3 | <u>HFTAECRFKECVFENYVVLASALYRQRRSGRAWYGLDKEGQVMKGNRVKKT-KAAAHFLPKLLE</u> | | FHF-3 |
| FHF-4 | <u>LFTPECKFKESVFENYVVIYSSMLYRQOESGRAWFLGLNKEGQAMKGNRVKKT-KPAAHFLPKPLE</u> | | FHF-4 |
| FGF-7 | <u>ECNEDCNFKELILENHYNTYASAKWTHN--GGEMFVALNQKGI PVRGKKTKEQK-TAHFLPMAIT</u> | | FGF-7 |
| FGF-1 | SD | | FGF-1 |
| FGF-2 | S | | FGF-2 |
| FHF-1 | VCMYREPSLHEIGEKQG--RSRKSSGTPT---MNGGKVVN-QDST | | FHF-1 |
| FHF-2 | VEPSLHDLTEFSRSGSGTPTKRSVSGVL----NGGKSMHNEST | | FHF-2 |
| FHF-3 | VAMYQEPSLHSVPEASPSPPAP | | FHF-3 |
| FHF-4 | VAMYREPSLHDVGETVPKPGVTPSKSTASAIMNGGKPVNKSKT | | FHF-4 |
| FGF-7 | | | FGF-7 |

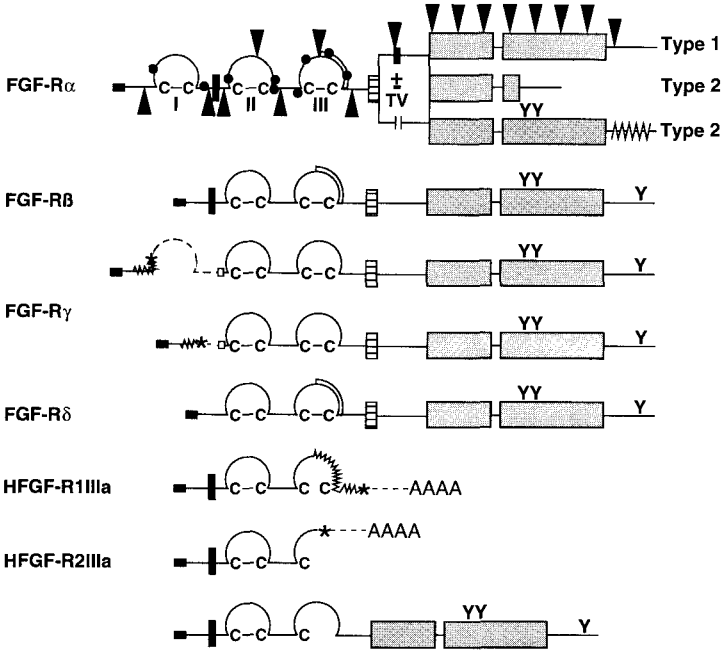


FIG. 4. Schematic of FGFRTK isoforms. Immunoglobulin-like modules are depicted with disulfide bonds. Triangles indicate known exon/intron junctions (18). Boxes indicate confirmed or putative translational initiation sites; solid boxes indicate a site followed by a secretory signal. The vertical solid rectangle indicates the acidic box sequence, and circles indicate putative glycosylation sites. The double line in loop III indicates alternate exons in the FGFRI, 2, and 3 genes. The transmembrane domain is indicated by the hatched rectangle; the wavy line at the COOH terminus indicates a reading frame shift in coding sequence. The three significant phosphotyrosine sites are indicated (Y). The asterisk indicates a stop codon and dotted lines indicate noncoding sequences. H, Human.

tions associated with health and disease that do not have an impact on the FGF system within tissues. Because of its ubiquity, its role in monitoring the external environment, mediating communication among cells in diverse tissues, and its position as an integral part of the extracellular matrix that communicates with the interior of cells, the FGF signal transduction system has a regulatory impact on most general, as well as tissue- and cell-specific, functions. Few studies to date have shown that components of the FGF system

FIG. 3. Four FGF homologous factors (FHF) associated with the nervous system. Homologous residues, when present in three of the four FGF homologs FHF-1 to FHF-4, are underlined and compared to FGF-1 and FGF-2 (top) and FGF-7 (bottom). For details, see Ref. 7.

are not present or that the system is not involved either directly or indirectly in regulatory processes. A major challenge lies in dissection of the combinatorial specificity and control of function of the three components of the FGF family in diverse tissues and regulatory processes in homeostasis and disease. This review concentrates on work from the author's laboratory on biochemical diversity of the FGFR family members related to biochemical function, with an emphasis on function of the family in liver and prostate. It is not intended as a comprehensive review of the field and literature, and the reader is referred to several reviews that cover the general FGF field (8–10) and specifically liver growth and function (11).

II. Diversity of Structure and Function

A. The FGF Polypeptides

The 10 cloned genes coding for members of the FGF polypeptide family produce products varying from 155 to 267 amino acid residues; these exhibit a core sequence homology split by one or two areas of lesser homology (Fig. 1). The core sequence is flanked at the NH₂ and COOH termini by unique sequences. FGF-1 and FGF-2, also called acidic FGF (aFGF) and basic FGF (bFGF), which exhibit the widest expression among diverse cells and tissues, were first to be cloned and characterized, and consequently have been most widely studied. Neither exhibits an apparent NH₂-terminal secretory signal sequence and each appears in intracellular compartments in addition to the pericellular matrix. This has raised questions concerning the mode of exit of the two FGFs from cells. Except for FGF-9 and the four most recent homologs that appear to be within the family (7), the other FGF polypeptides exhibit a conventional secretory signal and are thought to exit cells by conventional secretory mechanisms.

The structure of both FGF-1 and FGF-2 has been determined by X-ray crystallography (3–6) and, most recently, FGF-2 has been complexed with a heparin hexasaccharide (12). All FGFs exhibit a specific affinity for heparin and heparan sulfate that aids in purification and characterization, both from native and recombinant sources. Structural studies revealed that FGF-1 and FGF-2 have a similar backbone structure, with the single most significant difference in the inter- β strand 9/10, which was predictable from sequence differences in the otherwise homologous core of FGFs (Figs. 1–3). Structural and functional analyses indicate that the heparin-binding domain of FGF is composed of amino acid residues from both the NH₂ and COOH termini of the core homology domain among FGFs on each side of the unique inter- β strand 9/10 sequences (12–14). Functional analysis by site-directed mutage-

nesis of FGF receptor binding and bioactivity studies indicate that the most inviolate segment of FGFs lies within sequences between β -strands 10 through 12 in the COOH terminus of the core homology domain (15), a segment that also contributes to the heparin-binding domain (12, 13) (Fig. 2).

The FGFs exhibit multiple cysteines, two of which (except in FGF-8 and FGF-10) are conserved (Figs. 1 and 2). In the crystal structures of FGF-1 and FGF-2, the conserved cysteines are neither in position to form intrachain disulfide bonds, nor accessible to solvent (3, 4). However, some FGFs aggregate and form disulfide bonds between surface cysteines in solution, particularly in the absence of heparin and at high concentrations (16, 17). Aggregation via disulfides generally correlates with loss of FGF activity and has been proposed to play a role in storage and trafficking of FGFs that are not secreted by conventional pathways (18). These results suggest that active FGF polypeptides may exist only in the reduced monomeric forms in association with heparin/heparan sulfate.

B. The FGF Receptor Tyrosine Kinase

The FGFRTK family consists of the multiple products of four genes that exhibit the structural modules illustrated in Fig. 4. The longest product of the four genes consists of an extracellular domain containing three immunoglobulin (Ig)-like disulfide motifs split by intervening sequences flanked by a secretory signal sequence at the NH₂ terminus and a juxtamembrane sequence at the COOH terminus. The extracellular domain is followed by a transmembrane domain and the intracellular domain. The latter consists of the two tyrosine kinase subdomains (ATP binding site and catalytic domains) split by a short intervening sequence (the interkinase domain) and flanked by a juxtamembrane sequence at the NH₂ terminus and a COOH-terminal sequence outside the catalytic domain of the kinase. Alternate splicing at transcription gives rise to variant products of the FGFRTK in the extracellular and intracellular domains and in all subdomains except Ig loop II, the connecting sequence between loops II and III (interloop II/III domain), the ATP binding domain, and the interkinase domain. To date, over 20 single-site splice variations have been described, most of which occur in the FGFR1 and FGFR2 gene products. The variations occur in subdomains that affect practically every known function or property of a transmembrane tyrosine kinase receptor. These include ectodomain subdomains, which affect ligand (FGF and heparan sulfate) affinity and specificity; the kinase catalytic domain, which affects kinase activity; phosphothreonine and phosphotyrosine sites, which affect interaction with substrates and metabolism of the receptor; ectodomain alterations, which affect trafficking among cellular compartments; and variations in the COOH-terminal domain of unknown function. The combinatorial possibilities of the single-site splice variations

suggest a vast potential for diversity of function of monomeric products from a single FGFR TK gene.

I. THE EXTRACELLULAR DOMAIN

a. NH₂-Terminal Loop I and the FGFR γ Isoform. The FGFR TK ectodomain consists of distinct structural subdomains encoded in one or two exons, some of which are subject to alternate splicing and others which are invariant (Fig. 4). The first invariant exon encodes the NH₂-terminal domain headed by a secretory signal sequence. A single exon encodes an NH₂-terminal Ig disulfide loop (loop I) that is subject to alternate splicing by "exon skipping or excision," and its presence or absence defines the FGFR α and FGFR β isoforms, respectively (20, 21). The alternate splicing of loop I has not been reported in the FGFR3 or FGFR4 genes, therefore, to date they are known to express only the FGFR α isoforms. Except for the transmembrane domain, loop I is least conserved with respect to amino acid sequence of the structural subdomains of the FGFR TK family (Fig. 5). Ig loop I is not required for the binding of either FGF or heparin/heparan sulfate to FGFR TK, but affects the affinity of both when all other conditions are constant (22, 23).

Splicing at an alternate acceptor site in the FGFR1 gene results in a five-nucleotide deletion in coding sequence for loop I and a shift in reading frame resulting in a downstream stop codon (24). This event gives rise to the FGFR γ mRNA (Fig. 4). The consequence is predicted to be a short, secreted peptide and possibly initiation of translation at an alternative site, resulting in an intracellular product (consisting of loops II and III and downstream domains) (Fig. 4) that is not targeted to the cell membrane. A functional product of this mRNA has not been demonstrated despite its significant and variable expression among cell types and the ability of its cDNA to transform fibroblasts on transfection (24). An alternative role of this splice event may be the short, secreted NH₂-terminal polypeptide and an otherwise nonfunctional mRNA. The mRNA may simply be a consequence of the control of the

```

R1 ARPSPTLPEQAQ-----PWGAPVEVESFLVH-PGDLLQLR- CRLRDDVQSIN
R2 :::FS:V:DTTLEPEEP:TKYQISQPEVY:AA::ES:EV:- :M:K:AA-V:S
R3 SSES LGTEQRVVGRAAEV:GPE:GQQ:Q-::FGS::AVE:S- :PPPGGGMPGP
R4 SLEAEEV:LE-----:CL::SLEQQEQELTVALGQPV:L :CG:AERGGH-

R1 --WLRDGVQLAESNRTRITGEEVEVQDSVPADSLGLYA CVTSSPSGSDTTYFSVNVS
R2 --:TK:::H:GPN::VLI::YLQIKGAT:R::: :AAARTVD:E:L::M:::T
R3 TV:VK:::TG:VP:E:VLVGPORLQ:LNASHE:::A:S :RQLTQRVLCH-:::R:T
R4 --:YKE:SR::PAG:V:GWRGRL:IASFL:E:A:R:L :LARGSMIVLQNLTLITG

```

FIG. 5. Sequence comparison of immunoglobulin loop I. Sequences from human FGFR2, FGFR3, and FGFR4 are compared to FGFR1, with conserved sequences indicated by colons. Dashed lines depict gaps introduced to achieve maximum alignments among all four FGFRs.

no direct functional role of specific acidic box sequences has been demonstrated. The interloop I/II sequence has been implicated indirectly as a structural element that cooperates with loop I in affecting affinity of FGFRTK for both heparin/heparan sulfate and FGF (21–23).

c. Invariant Loop II Composes the Primary Heparan Sulfate and FGF Binding Site. Ig loop II is encoded by two exons that are invariant in the four FGFRTKs. At the NH₂ terminus of loop II is a sequence of about 20 residues that comprise a heparin-binding domain (22, 25, 26); this is highly conserved among the four mammalian FGFRTKs, but not the FGFRTK homologs in *Drosophila melanogaster*, sea urchin, and *Caenorhabditis elegans* (Fig. 7). Although highly conserved, each of the four human FGFRTKs exhibit one unique residue (boldface letters, Fig. 7) that may alter the structure of the domain. The invariant nature of the two exons encoding loop II in both normal and abnormal tissues is consistent with the essential role of the module in formation of the base, minimal binding site for all FGF polypeptides, in addition to the presence of the heparin-binding domain, which plays multiple and critical roles in formation and regulation of activity of the FGFR complex (22, 25).

d. Interloop II/III and Loop III Comprise a Homeo-Interaction Domain and Secondary FGF-Specific Binding Sites. Invariant sequence domains that comprise the interloop II/III and NH₂-terminal half of loop III and the variant alternately spliced COOH terminus of loop III cooperate in formation of secondary interaction domains of FGFRTK with FGF; these secondary do-

| | | |
|--------|-------------|--|
| HFGFR1 | KME | KKLHAVPAA K TVKFKC |
| HFGFR2 | KME | KRLHAVPAANTV K FRC |
| HFGFR3 | RMD | KKL L AVPAANTVRFRC |
| HFGFR4 | RME | KKLHAVPAGNTV K FRC |
| XFGFR1 | KME | KKLHAVPAA K TVKFKC |
| XFGFR2 | KME | KKLHAV S AANTV K LRC |
| DFGFR | RKE | L KRL Q SLSGNTV N LAC |
| SFGFR | RME | - P E K PLPS N TKV R LE C |
| EFGFR | V LF | N ETHALPAG R T L KL N C |

FIG. 7. Comparison of the heparin/heparan sulfate binding domain among FGFRTKs. Nineteen residues within Ig loop II, containing the major heparin-binding domain, are indicated for human (H) FGFRTK1 through FGFRTK4, the *Xenopus* (X) receptors 1 and 2, and the *Drosophila melanogaster* (D), sea urchin (S), and *Caenorhabditis elegans* (E) homologs. The basic residues that generally flank the heparin-binding domain within proteins (26) are italicized. Unique residues in each FGFRTK are indicated in bold.

| | Loop II | Inter-loop | Loop III |
|--------|---------|------------|-----------|
| HFGFR1 | LDVV | ERSPHRPI | LQAGLPANK |
| HFGFR2 | LDVV | ERSPHRPI | LQAGLPANA |
| HFGFR3 | LDVL | ERSPHRPI | LQAGLPANQ |
| HFGFR4 | LDVL | ERSPHRPI | LQAGLPANT |
| XFGFR1 | LDVV | ERSPHRPI | LQAGLPANT |
| XFGFR2 | LDVI | ERSSHRPI | LQAGLPANT |
| DFGFR | VQIK | HRTRSAPI | I--VVPQNV |
| SFGFR | VKIR | ERLVPVKPI | --MSPMKNV |
| EGFR | VIIV | NRMRRPPI | IVPNILANQ |

FIG. 8. Comparison of the homeo-interaction domain among FGFRTKs. Four residues at the COOH terminus of Ig loop II, the interloop II/III, and nine residues of the NH₂ terminus of loop III are shown. The three domains were separated and assigned by homology modeling to characterized Ig modules (21). The interaction interface lies within the interloop and loop III sequence (25a) H, Human; X, *Xenopus*; D, *D. melanogaster*; S, sea urchin; E, *C. elegans*.

mains determine specificity of the FGFR complex for FGF (22). The interloop II/III sequence and NH₂ terminus of loop III are encoded in an invariant exon, whereas the COOH-terminal half is encoded in another, which is alternately spliced. The interloop sequence is invariant in all four mammalian FGFRTKs, but is not conserved in the FGFR homologs in lower organisms (Fig. 8). The interloop sequence is thought to be important in the structural relationship between loops II and III. An interaction interface between FGFRTK ectodomains has been identified within the sequence composed of the interloop II/III sequence and the NH₂ terminus of Ig loop III (25a).

The FGFR1, 2, and 3 genes exhibit two adjacent exons, IIIb and IIIc, in the genome that codes for variants in the COOH-terminal half of loop III (19). In FGFR2, the two exons are alternately spliced by a mutually exclusive mechanism; e.g., isoforms with variants in the COOH-terminal half of loop III (Fig. 9) are not expressed in the same cell type at the same time (27, 28). Expression of one or the other defines a cell phenotype (see Section V). Although the second half of Ig loop III encoded within exons IIIb and IIIc is not required for the FGF and heparin/heparan sulfate association with FGFRTK, the domain has dramatic impact on specificity of the receptor gene for different FGF polypeptides (27, 28). Similar to the NH₂-terminal loop I, the COOH-terminal half of loop III in the FGFR4 gene is invariant and alternative exons coding for the domain are not present in the genome (29).

An additional variant potentially arises in loop III in the FGFR1 and FGFR2 gene products by alternate chain termination and poly(A) addition at a site in the otherwise intronic sequence downstream of the first exon coding for loop III (19). The event results in a truncated and secreted product, the length of product dependent on receptor gene and species. The function

| f12 | E | b13 | F | f13 | G |
|------|-----------|--------------------|------------------|-------|-------------------|
| TSGI | NSSD--AEV | murine R1IIIb | | | |
| HSGI | NSSD--AEV | LTLFNVTEAQS | GEYVCKVSN | YIGEA | NQSAWL TVT R1IIIb |
| HSGI | NSSN--AEV | LALFNVTEADA | GEYICKVSN | YIGQA | NQSAWL TVL R2IIIb |
| SWIS | ENVE-ADAR | LRLANVSEKDG | GEYLCRATN | FIGVA | EKAFWL RVH R3IIIb |
| TAGV | NTTDKEMEV | LHLRNVSFEDA | GEYTCLAGN | SIGLS | HHSAWL TVL R1IIIc |
| AAGV | NTTDKEIEV | LYIRNVTFEDA | GEYTCLAGN | SIGIS | FHSAWL TVL R2IIIc |
| TAGA | NTTDKELEV | LSLHNVTFEDA | GEYTCLAGN | SIGFS | HHSAWL VVL R3IIIc |
| TADI | NSSE--VEV | LYLRNVSAEDA | GEYTCLAGN | SIGLS | YQSAWL TVL R4III |
| TTDI | NISE--VQV | murine R4III | | | |

FIG. 9. Variations in the COOH-terminal half of Ig loop III caused by alternate splicing. Invariant residues are indicated in bold (Leu/Ile and Thr/Ser were considered similar). Sequences are grouped into Ig loop subdomains as previously described (21). All sequences are human except the indicated murine sequences.

of this FGFRTK variant is currently unclear. The potential expression products exhibit the heparin-binding domain and the minimal FGF binding domain. It has been proposed that the secreted products may sequester FGF away from the transmembrane isoforms of FGFRTK or play a dominant negative role by dimerization between ectodomains (19, 30). In the one natural product, secreted FGFR1, that has been examined, the affinity for FGF is drastically reduced relative to the transmembrane isoforms (30). Recent results suggest that secreted fragments of the FGFRTK ectodomain are relatively unstable and rapidly lose affinity for ligands (31). This suggests that the interaction of secreted fragments of the FGFRTK ectodomain with ligands may not be competitive with that of the full-length transmembrane receptor complex. An alternative role of mRNA chain termination and poly(A) addition may be simply to control the level of full-length transmembrane FGFRTK, as discussed earlier for the γ isoforms of FGFR1.

Large numbers of point mutations in the coding sequence of the FGFR1, 2, and 3 genes beginning downstream of the COOH terminus of loop II cause single amino acid changes associated with a variety of craniofacial and skeletal syndromes in humans, and are thought commonly to cause a gain of function through constitutive dimerization of the FGFRTK (32) (see Section III). One silent mutation, with respect to a codon for alanine in exon IIIc of the FGFR2 gene, introduces a new splice donor site in the coding sequence for exon IIIc, which results in an in-frame deletion of the coding sequence for the 17 COOH-terminal residues of loop III (33). A similar deletion of the 13 COOH-terminal residues in exon IIIc in the FGFR1 gene has been reported in mouse L cells, and it likely occurs by the same type of mutation (31). Although impact of these deletions on FGF binding has not been tested, the mutations shorten the distance between the FGFR ectodomain and

the cell membrane and likely change the general relationships among structural subdomains in the FGFR ectodomain.

2. THE TRANSMEMBRANE AND JUXTAMEMBRANE DOMAINS

The extracellular juxtamembrane, the transmembrane, and most of the intracellular juxtamembrane domain are encoded in one exon whose amino acid sequence is relatively divergent among the four FGFRTKs. The role of each of these structural domains has not been clearly elucidated; however, by analogy to other receptors, the three domains are predicted to be important in oligomerization of FGFRTK in cell membranes, either by participation in a direct interaction or by indirectly affecting structure of other interaction domains. Point mutations in the transmembrane domain of FGFR3 are associated with achondroplastic dwarfism through a gain of function elicited by constitutive dimerization (35, 36). A mRNA variant of the FGFR gene that results from skipping/excision of the two exons coding for both the COOH terminus of Ig loop III and the juxtamembrane/transmembrane domain has been reported (37). The predicted product is a secreted receptor tyrosine kinase because of lack of a membrane stop-transfer signal; however, the product appears to accumulate in the particulate fraction of cells (37).

Other than an indirect role as a spacer between the extracellular domain of FGFRTK and the kinase domain, the intracellular juxtamembrane domain has not been implicated in a direct function of FGFRTK. Splicing at an alternate 5' donor site results in the presence and absence of a threonine/valine dipeptide in the intracellular juxtamembrane domain that is a substrate for serine/threonine kinases and may be involved in metabolism of FGFRTK (20, 38). The dipeptide and a similar sequence context are not present in the FGFR4 gene, although a serine is in the homologous position as the threonine. The threonine/valine variation has been demonstrated only in the FGFR1 and FGFR2 gene products.

3. THE INTRACELLULAR KINASE DOMAIN AND COOH-TERMINAL TAIL

The intracellular kinase and COOH-terminal domains of FGFRTK are encoded in at least eight exons (19). The kinase domain consists of the ATP binding site and catalytic domains split by a 14-residue interkinase domain. The sequence of the latter is relatively divergent among the four FGFRTK, whereas the domains on each side are the most highly conserved within the FGFRTK family. Three functionally important tyrosine autophosphorylation sites have been identified in the FGFRTK (23, 39, 40). Two of them (Tyr-653 and Tyr-654 by FGFR1 numbering) comprise a tyrosine/tyrosine dipeptide within an autorepressor structural domain within the catalytic domain of the

| | | | |
|--------|----------------------|-------------------------|-----------------|
| HFGFR1 | ADFLGARDIHHI | DYYKKTT | NGRLPVK |
| HFGFR2 | ADFLGARDINNI | DYYKKTT | NGRLPVK |
| HFGFR3 | ADFLGARDVHNL | DYYKKTT | NGRLPVK |
| HFGFR4 | ADFLGARGVHHI | DYYKSTS | NGRLPVK |
| XFGFR1 | ADFLGARDIHHI | DYYKKTT | NGRLPVK |
| XFGFR2 | R DFLGARDVNNI | DYYKSTS | NGRLPVK |
| DFGFR | ADFLGARDIHHI | DYYKKTT | NGRLPVK |
| SFGFR | C DFLGARDIHYI | D FYRKTT | D GRLPVK |
| EFGFR | S DFGLSRDVHCN | D YYRKR G | NGRLPIK |
| TFGFR3 | ADFLGARDVHNL | DYYK E TT | NGRLPVK |

FIG. 10. Comparison of the kinase repressor loop domain among FGFRTKs. The repressor loop sequence domains of FGFRTKs from the four human (H) genes, *Xenopus* (X), *D. melanogaster* (D), sea urchin (S), and *C. elegans* (C) are compared. TFGFR3 is a mutant occurring in thanatophoric dysplasia type II (50). Residues unique to each are in bold.

kinase (41) (Fig. 10). Conformational change in the domain resulting from trans-autophosphorylation of one or both of the tyrosines within adjacent dimers or higher order oligomers of FGFRTK is thought to relieve restriction of the access of external substrates to the catalytic site and to be the fundamental basis for the requirement of oligomerization of FGFRTK in signal transduction. The third functional tyrosine (Tyr-766 in FGFR1) subject to autophosphorylation is outside the kinase domain in the COOH terminus and is required for stable interaction of *src* homology 2 (SH2) domains of phospholipase C γ (PLC γ) with FGFRTK (23, 42). Despite the ubiquity of the FGFR signal transduction complex in tissues, its broad effect on gene expression and cellular functions and the effect of its manipulation on most intracellular signaling pathways, PLC γ is the only substrate so far identified that directly interacts with the intracellular domain of FGFRTK with an affinity significant enough to dissect and study the features of the enzyme-substrate complex (23). Site-directed mutagenesis experiments suggest that the interaction with and phosphorylation of PLC γ by FGFRTK can be dissociated from the growth-stimulatory activity of the FGF signal transduction system (43, 44). The broad spectrum of intracellular signal transduction pathways activated by the FGF receptor system, the increasing evidence that the primary role of autophosphorylation may be simply to derepress the FGFR kinase, and the paucity of FGFRTK binding proteins raise major questions about how intracellular substrates are kept from, or appropriately recruited to, the activated FGFR complex.

Multiple variants in the intracellular domain of FGFRTK have been reported. An alternate donor site splice in the FGFR1 gene in human hepatoma cells results in a shift in reading frame that truncates the catalytic domain of the kinase to 24 amino acid residues followed by 44 unique COOH-terminal residues (20, 23). The event obliterated the three functional tyrosine phos-

phorylation sites. The variant called the FGFR1 type 2 isoform (FGFR1-2) is thought to underlie the FGF concentration-dependent biphasic growth response of cells by a dominant-negative heterodimerization with the intact kinase isoforms (23) (see Section IV). An alternate exon skipping/excision event in the FGFR2 gene in rat prostate results in an isoform exhibiting most of the kinase domain, with a unique COOH terminus that lacks the tyrosine and that, when phosphorylated, interacts with PLC γ (45). The isoform called FGFR2 type 2 (FGFR2-2) has subsequently been shown to exhibit reduced or no kinase activity measured by autophosphorylation, although it can be trans-phosphorylated when associated with an intact FGFRTK (M. Kan, F. Wang, and W. L. McKeehan, unpublished results). Additional variants in the distal COOH-terminal domain outside the kinase domain of the FGFR2 gene (46–48), and more recently the FGFR1 gene (49), have been described, some of which arise by alternate splicing mechanisms consistent with current knowledge of genomic structure, and others that have yet to be explained.

The COOH-terminal tail sequence and NH₂-terminal flanking domains were removed from the FGFRTK to achieve resolution of the core kinase structure by crystallography (41). Because of lack of understanding of the structural relationship of the COOH tail in relation to the core kinase domain, the mechanism of phosphorylation of the PLC γ -binding phosphotyrosine (Tyr-766 in FGFR1) and the functional impact of variants in the COOH terminus of the FGFRTK monomer are unclear. It is noteworthy that heterozygous mutations in FGFR resulting in abnormal extensions of the COOH terminus appear to cause the severe craniofacial and skeletal abnormalities associated with thanatophoric dysplasia (neonatal lethality) (50). These series of mutations appear commonly to cause gain of function in the FGFR signal transduction complex, which suggests that the COOH-terminal domain may also normally play a repressor role in FGFRTK activity and function (see Section III).

C. The FGF Receptor Heparan Sulfate

The independent interaction of FGF polypeptides with heparin and heparan sulfates has been extensively studied. Numerous electrolytes with properties similar to heparin have been reported to interact with FGFs. These range from unsulfated di- and trisaccharides to molecules with ringed structures bearing various charge densities, to various fragments from heparin, to intact glycosaminoglycans from native sources. Molecular modeling (31, 51), site-directed mutagenesis (51), calorimetry (51), and cocrystallization with various heparin-like agents, including tetrameric and hexameric fragments of heparin (12), suggest that the heparin-binding domain of FGF, which spans the width of an FGF, accommodates a pentameric polysaccharide and

is composed primarily of amino acid residues from the first and last third of the FGF sequence. The many heparin-like agents that bind to FGF impact the bioactivities of FGF to different extents, dependent on the specific assay and conditions employed. Generally, the range of heparin-like agents that affect FGF activity is much more restricted than the range of agents that bind to the isolated factor. Heparin-like agents, from small molecules to preparations of heparin containing chains exceeding 20,000 kDa, have little effect on the structure of FGF, as detected by current analytical methods (12); however, fragments of heparin exceeding four to five monosaccharides protect FGF against heat denaturation and proteolytic digestion (14).

FGFs bind to pericellular matrix sites with properties of heparan sulfate chains of proteoglycans, distinguishable from FGFR complexes by their extractability with heparin or high salt concentrations and their failure to associate covalently with FGF in the presence of bifunctional protein-protein crosslinking agents (17, 53). Unless precisely controlled with respect to concentration, time, and temperature, extraction with high salt also removes FGF bound to FGFRTK complexes (53; M. Kan and W. L. McKeehan, unpublished results). FGF bound to FGFRTK complexes can be distinguished from FGF bound to pericellular matrix heparan sulfate sites in two ways: (1) by the requirement for extraction with detergent after removal of the heparan sulfate sites by heparin and (2) the ability to form covalently crosslinked complexes with FGF (17, 53). Based on these observations, the independent interaction of FGF with pericellular matrix heparan sulfate binding sites has been proposed to reflect a role of matrix heparan sulfate proteoglycans in maintenance of the stability and the lifetime of FGF and in sequestering FGF either away from or in proximity to signal-generating FGFRTK complexes (Fig. 11). In view of the wide range of heparan sulfate-like molecules that interact with FGFs and their abundance in the tissue matrix, it is unlikely that FGFs exist in the unbound state in tissues. Whether there is specificity within the repertoire of pericellular matrix binding sites for FGF and how heparan sulfate-bound FGF and FGFRTK come together to form active FGFR signal transduction complexes is a major problem in understanding FGF signal transduction.

As described in Section II,B,1,c, the NH₂ terminus of Ig loop II of FGFRTK exhibits a major heparin-binding domain (25) (Fig. 7). The FGFRTK domain is within a linear sequence of 20 amino acid residues that conforms to the general structure and composition of heparin-binding domains in a wide variety of proteins (26, 53). The linear sequence is in contrast to the discontinuous heparin-binding domain within FGF. The heparin-binding property and structural domain of FGFRTK were identified by analysis of purified recombinant FGFRTK ectodomain away from cellular products. When washed free of residual cell-derived heparin-like factors, the

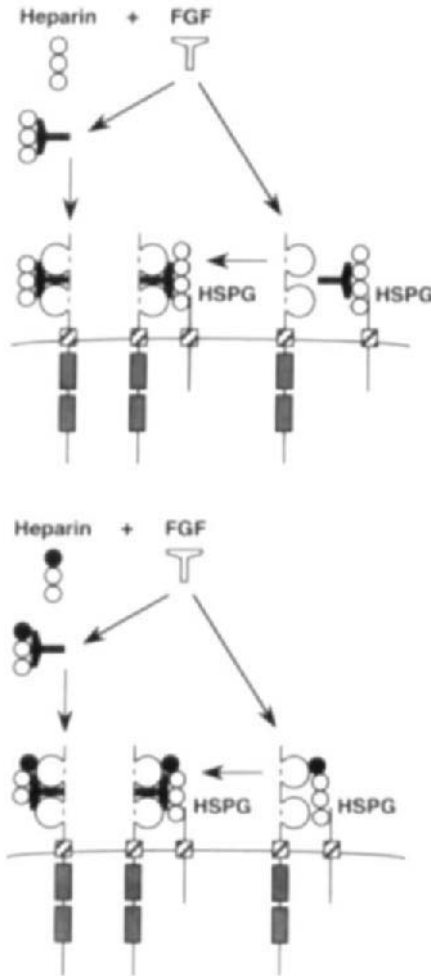


FIG. 11. Two views of the role of heparin/heparan sulfate in binding of FGF to FGFR TK. The upper view describes a two-step process in which FGF bound to pericellular matrix heparan sulfate chains of proteoglycans (HSPG) or soluble heparin fragments binds to the FGFR TK ectodomain independent of an interaction of the oligosaccharide with FGFR TK. The lower view describes the formation of a ternary complex of heparin/heparan sulfate, FGF, and FGFR TK, whereby FGF binds to a preformed duplex of oligosaccharide and FGFR TK.

isolated FGFR TK failed to bind FGF in the absence of heparin or heparan sulfate, but bound to immobilized heparin-agarose, which also restored FGF binding (25). A relatively specific independent interaction of heparin with the ectodomain of FGFR TK was indicated by the ability of heparin to protect the

active FGF-binding core of FGFR1 from proteolytic degradation. A combined analysis using tryptic fragmentation, epitope mapping with specific monoclonal antibodies (21), and site-directed mutagenesis was used to localize the heparin-binding domain within Ig loop II (25) (Fig. 7). Mutations in the heparin-binding sequence, which affected heparin binding, affected FGF binding proportionately. A synthetic peptide K18K with the sequence of the heparin-binding domain of FGFR1 bound heparin and inhibited FGF binding to FGFR1 (25, 31). These results suggest that the FGF receptor may be a duplex of FGFRTK and the heparan sulfate portion of a pericellular matrix proteoglycan (FGFRHS) (Fig. 11). In contrast to the interaction with FGF polypeptides, a high-affinity interaction of heparin with FGFRTK ($K_d = 10$ nM) requires divalent cations (Ca^{2+} or Mg^{2+}) at extracellular physiological levels (31). In the absence of divalent cations, heparin interacts with FGFRTK with very low affinity ($K_d > 10$ μM), which probably reflects nonspecific electrolyte interactions. These results suggest that the interaction of heparin and heparan sulfate with FGF polypeptides and FGFRTK is quite different. The conditions required for interaction of heparan sulfate with FGFRTK are more stringent than those for interaction with FGF, and therefore may be more specific. The difference suggests that heterotypic sequences and proper spacing of them in a single heparan sulfate chain may be required if a single chain is to span both FGFRTK and FGF to form a ternary complex.

Not only are divalent cations required for the high-affinity interaction of heparin or heparan sulfate with FGFRTK, they appear to stabilize the FGFRTK ectodomain against time- and temperature-dependent loss of affinity for FGF (31), as well as against proteolytic degradation (M. Kan and W. L. McKeehan, unpublished results). Although the affinity for FGF increases ($K_d = 200$ pM) in the presence of divalent cations and heparin, FGFRTK binds FGF with relatively high affinity ($K_d = 600$ pM) in the absence of divalent cations and heparin (31). Notably, divalent cations suppress FGF binding in the absence of heparin to levels insufficient to quantitate affinity coincident with their stabilization of FGFRTK. The nature of the heparin-independent binding of FGF in the absence of divalent cations, relative to that in the presence of the cations and heparin, remains to be established. Because FGFRTK is unlikely to exist in a divalent cation- and heparan sulfate-deficient environment in normal cell membranes, the heparin-independent binding of FGF to FGFRTK is probably not of physiological significance. An analysis of autophosphorylation activity of FGFRTK revealed that, in the absence of divalent cations and heparin, autophosphorylation of FGFRTK occurs independent of FGF (23, 31). Divalent cations and heparin suppress FGF-independent autophosphorylation to near background levels. In sum, divalent cations maintain the FGFRTK ectodomain in a heparan sulfate-dependent conformation with respect to FGF binding, divalent cations are re-

quired for interaction of the FGFRHS with the FGFRTK, and the two cooperate to maintain activity of the FGFRTK in an FGF-dependent state. A major challenge is to determine the nature and specificity of the FGFRHS component of the FGFR signal transduction complex in various cell types and its character in homeostasis and in pathologies of various tissues.

III. Structure, Assembly, and Control of the FGF Receptor Complex

A. Models of the FGF Receptor Complex

Three general models of the oligomeric FGF complex have been proposed (Fig. 12). The simplest is that proposed by Ornitz and colleagues (54) and adopted by Schlessinger and co-workers (55) in which heparan sulfate chains with tandemly bound FGFs result in oligomerization of two or more FGFRTKs. Neither an interaction of heparan sulfate with FGFRTK nor an interactive interface between FGFRTK plays a central role in regulation of the complex in this model. A second model follows the growth hormone paradigm, which involves adjacent FGFRTKs linked together by a single FGF with high and low affinity for each of the two FGFRTKs (15, 51). FGF binds to FGFRTK first through the high-affinity interaction and then recruits a second FGFRTK through the low-affinity interaction. Heparan sulfate is proposed to play a role in the low-affinity binding step of assembly of the complex. An interaction between FGFRTK occurs via an interface in loop III. In both models, the oligomerization of FGFRTK is assumed to be limited by concentration and diffusional mobility of FGFRTK and heparan sulfate proteoglycan, whose chains bear FGF, diffusible FGF–heparan sulfate complexes, or diffusible free FGF, which must come together to form the FGFR complex. We have proposed a third model in which activity of the FGFR complex is limited by conformation and is not necessarily subject to the random and probabilistic processes of diffusion (Fig. 12). This model accommodates all current structure–function analyses in the FGF field to date.

I. MODEL OF A TERNARY COMPLEX OF FGFRTK, FGFRHS, AND FGF

Although the backbone of FGF-1 and FGF-2 has been determined from crystal structures, there is neither information on the structure of an FGF when bound to FGFRTK or an FGFRTK–HS complex nor a static structural template of the ectodomain of FGFRTK. We have used well-characterized Ig modules as templates to construct homology models of the ectodomain of FGFR α and FGFR β (21, 56). When the sequences of Ig loops II and III are

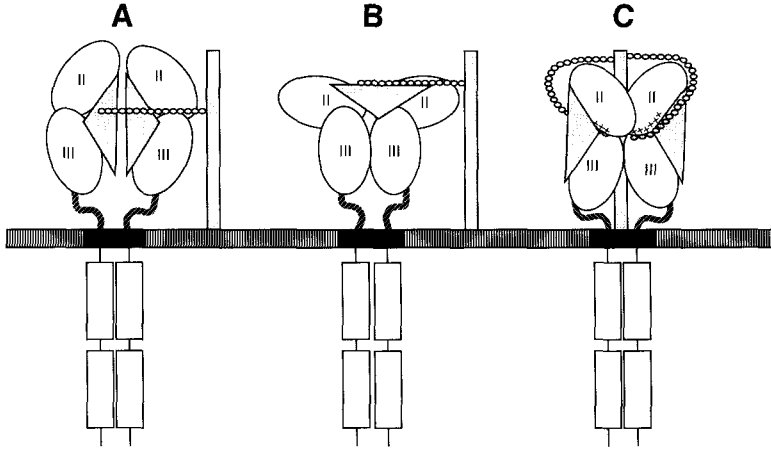


FIG. 12. Three models of assembly of FGFR dimeric complexes. Model A (54, 55) suggests that a heparan sulfate chain bridges two FGF polypeptides, recruiting two FGFRs (two FGFs per two FGFRs) and thereby resulting in holding the FGFR intracellular domains in proximity for trans-phosphorylation. The growth hormone receptor complex paradigm (one FGF per two FGFRs) illustrated in model B (15, 51) involves high-affinity binding of FGF to an FGFR (left) and then recruitment of a second FGFR through formation of a ternary complex of FGF, FGFR, and a heparan sulfate chain (right) through a low-affinity interaction of a secondary binding domain on FGF. The recruitment brings the two intracellular domains in proximity for trans-phosphorylation. Model C (31) suggests that there is a preexisting duplex of an FGFR and heparan sulfate chain. To achieve molecular symmetry, different antiparallel heparan sulfate chains from the same proteoglycan core run in front and in back of back-to-back dimers of FGFR (see Fig. 13). A requirement for separate chains places additional restrictions on spacing of the serine glycosylation sites and composition of the heparan sulfate proteoglycan, which can serve as coreceptor.

assigned β -strand and β -strand connector domains using the C2 Ig module as template, the clusters of highest sequence conservation throughout the family of FGFR are at the back face of loop II and the front face of loop III (21). When the two modules are placed in tandem connected by the inter-loop sequence, which is 100% conserved among FGFR variants, well-conserved subdomains from the back face of loop II and the front face of loop III come together to form a potential FGF binding pocket of FGFR β (21). When the resolved atomic structure of FGF-1 or FGF-2 is oriented with the binding site by alignment of heparin-binding domains in both FGFR and FGF, atomic models reveal a continuous S-shaped channel, formed by the heparin-binding domains of each, that accommodates a decameric heparin or heparan sulfate oligosaccharide chain (14, 31). As discussed earlier, interaction of the heparan sulfate chain with FGFR requires divalent

cations, whereas interaction with FGF does not. In the model ternary complex, a single heparan sulfate chain must be capable of binding to both FGFR TK and FGF and must possess the proper length and character to assume the indicated conformation.

2. FGF-SPECIFIC SUBDOMAINS IN THE ECTODOMAIN OF FGFR TK

In the context of cell membranes, where divalent cations and heparan sulfate proteoglycans coexist with FGFR TK and are conceivably synthesized and metabolized coordinately, we have proposed that the unoccupied FGFR is a duplex of FGFR TK rigidly anchored through divalent cation coordination sites to a heparan sulfate chain from FGFRHS (Figs. 12C and 13). The primary interaction of FGF-1 with the FGFR TK-HS complex is supported by the interaction of FGFRHS and Ig loop II (25, 57), whereas the interac-

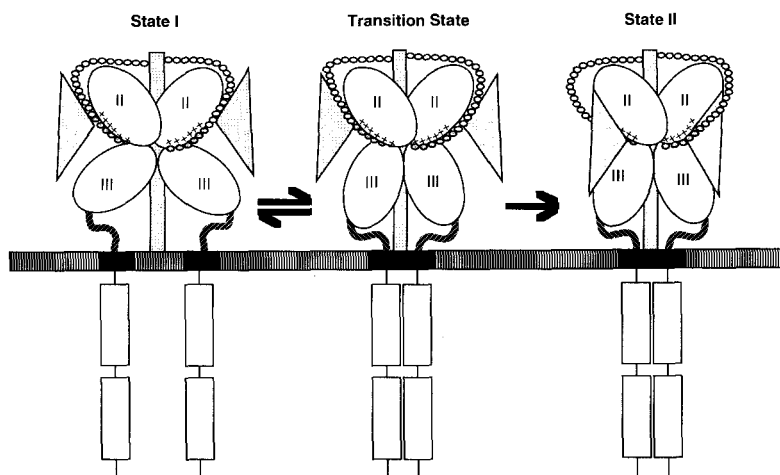


FIG. 13. Three hypothetical conformational states of the FGFR complex within the pericellular matrix. State I is the predominant resting conformation in which the angle between Ig loops II and III is rigidly fixed by anchorage of loop II to a heparan sulfate chain of the FGFRHS subunit through divalent cations (+++). The FGFRHS protein core is shown behind and between two FGFR TKs anchored to antiparallel chains from the FGFRHS core protein. The orientation of loop III is such that in state I the intracellular kinases are kept apart through cooperation of the extracellular juxtamembrane domain, the transmembrane domain, and potentially the intracellular juxtamembrane domain. Flexibility of orientation of loop III relative to rigidly anchored loop II results in the reversible, transition conformation in which the intracellular kinases transiently assume an enzyme-substrate relationship. Docking of FGF into the transition conformation results in essentially irreversible state II complexes in which adjacent kinases maintain an enzyme-substrate relationship.

tion with FGF-2 and FGF-7 is supported by loop II and 30 additional residues that comprise the interloop sequence, the A β -strand, b11, and part of the B β -strand of Ig loop III (57). It is noteworthy that the additional sequence required for high-affinity binding of FGF-2 and FGF-7 also contains the interactive interface between FGFRTK ectodomains (Fig. 8). Therefore, it cannot be distinguished whether the additional structural domains required for high-affinity binding of FGF-2 and FGF-7 relative to FGF-1 are formed within the extra sequence within a monomer or by the dimeric interaction between two of the additional sequence domains. Sequences downstream of the A-b11-B domain of loop III have a secondary or indirect impact on FGF binding to the base FGFRTK-HS complex by interaction with a secondary binding domain on FGF polypeptides that is distinct from the major heparin-binding domain and loop II binding domains (57). A solvent-exposed loop composed of the inter-9/10 β -strand domains, which is an obvious difference between the backbone structures of FGF-1 and FGF-2 (Fig. 2) and which differs significantly from the other eight members of the family, has been proposed to be involved in the secondary interaction of FGFs with loop III sequence domains (57, 58). The secondary interaction of FGF with loop III plays a major role in the specificity of the interaction of FGFRTK-HS with different FGF polypeptides. Loop III structural domains involved in the secondary interaction are a composite of the invariant relatively conserved NH₂-terminal half (downstream of the A-b11-B domains) and the variant COOH-terminal half of loop III (57).

3. FGF- AND HEPARAN SULFATE-INDEPENDENT KINASE ACTIVITY AND ASSOCIATION OF THE FGFRTK ECTODOMAIN

As discussed earlier, the *trans*-phosphorylation between FGFRTK is thought to be essential in derepression of the active site of the kinase and possibly in creation of phosphotyrosine sites involved in binding or release of signal transduction substrates. Normally this process is dependent on FGF and the FGFRTK ectodomain containing the FGF binding site *in vivo*. When removed from its ectodomain by recombinant methods, a construct of the intracellular kinase, including the COOH-terminal tail of FGFRTK in solution, exhibits an intrinsic activity that results in autophosphorylation of the COOH-terminal tail tyrosine (Tyr-766) (42). When the COOH-terminal tail sequence was removed, multiple other phosphotyrosine peptides from the FGFR1 kinase could be detected (40), including repressor loop Tyr-653, first described by our laboratory (39), and Tyr-654 (40). When removed from cell membranes and the pericellular matrix, full-length FGFRTK with an intact ectodomain also exhibits an intrinsic kinase activity resulting in phosphorylation *in trans* of Tyr-653 (and possibly Tyr-654) in FGFR1 in the absence of heparan sulfate, FGF, and other cofactors (23, 31). This suggests that *trans*-

phosphorylation between FGFRTK kinases resulting in derepression and activation of the kinase simply requires sufficient proximity of two FGFRTKs to establish and maintain an enzyme–substrate relationship. The question arises as to what normally limits the proximity and enzyme–substrate relationship between FGFRTKs and how the appearance or activation of FGF overcomes the limitations.

4. RESTRICTION AND MAINTENANCE OF FGF DEPENDENCE OF THE ENZYME–SUBSTRATE RELATIONSHIP BETWEEN FGFRTKS BY HEPARIN/HEPARAN SULFATE AND DIVALENT CATIONS

a. Subversion by Mutation of Restrictions on FGF-Dependent Trans-Activation between FGFRTKs. The question of what restricts the proximity and enzyme–substrate relationship between autoactivating FGFRTKs has become more important with the findings that nearly 30 different autosomal dominant mutations in FGFRTK result in a variety of abnormal craniofacial and skeletal conditions of graded severity dependent on the individual mutation (32) (Fig. 14). Although several mutations occur in the intracellular domain, the majority occur beginning from the COOH terminus of loop II through the transmembrane domain. Among the mutations whose impact on function has been analyzed, the common feature appears to be a graded gain of FGF-independent function rather than a loss of function, due to loss of ability to interact with either FGF or heparan sulfate, or to loss of kinase activity (31, 36, 59–61). This is consistent with the fact that the diverse mutations show no common effect on FGF binding activity; some cause loss and some have no effect (59–62) (F. Wang, M. Kan, and W. L. McKeehan, unpublished results). Coincident with the gain of function caused by several mutations in the ecto- and transmembrane domains is the increase in dimerization between FGFRTKs; this is promoted by acquisition of an unpaired cysteine residue (60) or a charged residue in the core of the transmembrane domain, which has been shown to cause constitutive dimerization in the *erbB2* receptor kinase system (36).

How can such diverse mutations commonly result in constitutive dimerization and consequently increased trans-phosphorylation between kinases? Based on results discussed earlier, we have proposed that pericellular matrix FGFRTK heparan sulfate chains anchor FGFRTK loop II through divalent cations bridges so that the relationship of loop III to a fixed loop II exists in at least two conformational states (31) (Fig. 13). In state I, which is predominant when FGF is not docked into the active site of the complex, the relationship is such that the intracellular kinases cannot sustain an enzyme–substrate relationship. State II, in the absence of active FGF, is a minority transitional and reversible conformation, but is stabilized and rendered relatively irreversible when FGF is able to dock into the active site. Some acti-

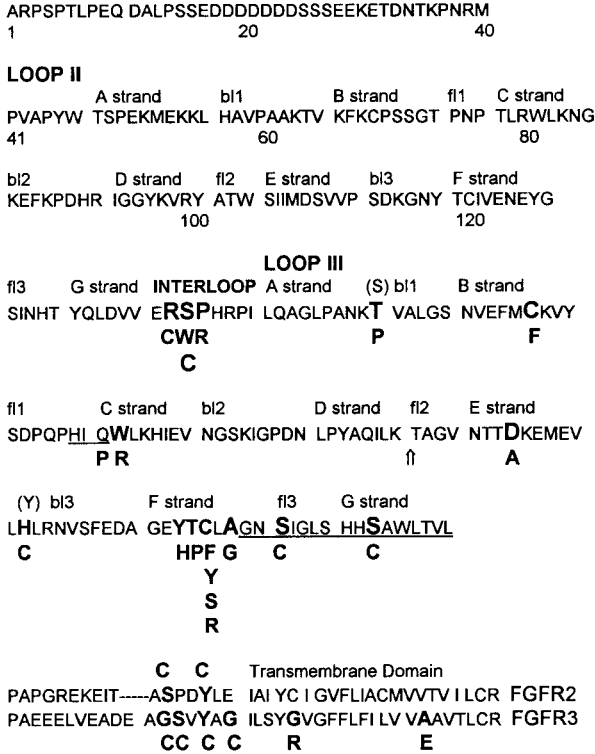


FIG. 14. Single-site heterozygous mutations in the homeo-interaction domain, loop III, the extracellular juxtamembrane, and transmembrane domains of FGFR1K; these may cause graded levels of FGF-independent activation of FGFR signal transduction complexes, causing skeletal and cranial abnormalities in humans. The sequence of FGFR1 β through loop III is indicated and is divided into putative structural subdomains characteristic of tandem immunoglobulin motifs (see Ref. 21). Residues are numbered beginning at the mature NH₂ terminus of FGFR1. The NH₂ terminus containing the acidic box is at the top; the start point of loop II and loop III and the interloop II/III sequence is indicated in bold. Residues and mutations are indicated in bold. Residues in parentheses are the actual mutated residues in the FGFR2 sequence. Underlined residues represent in-frame deletions caused by either a nucleotide deletion (3-base pair multiple) or a splice site mutation (arrow).

vating trans-phosphorylation may occur in the transient state, but it is likely too minor to be of consequence and may even be removed by phosphatase activity. In stable state II complexes, the kinases are able to assume the prolonged relationship required for sustained trans-phosphorylation, activation of the kinases, and delivery of signal. Dependent on location, mutations in

the ectodomain downstream of the COOH terminus of loop II destabilize the predominant state I conformation to different degrees and contribute to stability and increase the number of complexes that reside in state II in the absence of FGF.

Coincident with the intrinsic kinase activity of the isolated FGFRTK, described above, is the ability of the isolated FGFRTK ectodomain intrinsically to self-associate, independent of the intracellular kinase domain and cellular factors, by interaction of the highly conserved sequence that connects loop II to loop III (23). In its strategic location between loops II and III, the interaction domain plays a key role in the orientation of loop III relative to loop II and, in turn, the structure of the interaction domain will be affected by alteration of the position of loop III relative to loop II. Therefore, the domain is predicted to be pivotal in determining conformation that either restricts or allows the enzyme substrate relationship between intracellular kinase domains.

b. Mutation Site-Dependent Levels of FGF-Independent Kinase Activity Cause Phenotypes of Graded Severity. A striking feature associated with the diverse mutations in FGFRTK is their diverse phenotypic consequences, particularly on craniofacial and skeletal features, without obvious effect on the function of adult tissues in which the same FGFRTK isotype is present and believed to be functional. Moreover, the phenotypic consequences, when obvious, appear to be graded in severity, ranging from short limbs and unusual facial features in otherwise normal achondroplastic dwarfs to profound dwarfism, which is accompanied by multiple other abnormalities associated with neonatal lethality (thanatophoric dysplasia). The model in Fig. 13 provides an explanation for the graded phenotypes and suggests even more detail about potential control and mechanism of action of the FGFR signal transduction complex. In heterozygous cells expressing a particular FGFRTK isotype, of which 50% are mutant and 50% are wild type, 25% of FGFR complexes will be composed of mutant–mutant, 50% of mutant–wild type, and 25% of wild type–wild type (Fig. 15). The conformational model requires that the orientation of loop III relative to immobile loop II in both partners in FGFR complexes be fixed in a favorable conformation for the sustained enzyme–substrate relationship. Therefore, wild-type FGFRTKs, whether in partnership with wild-type or mutant monomers, require occupancy by FGF to achieve a stable state II conformation (Fig. 13). Therefore, for mutations that disrupt the equilibrium between state I and state II complexes, 75% of complexes remain FGF dependent, including the 50% mutant–wild-type complexes (Fig. 15, top panel). Maximum FGF-independent signal cannot exceed 25% of the total possible, even if all mutant–mutant complexes are

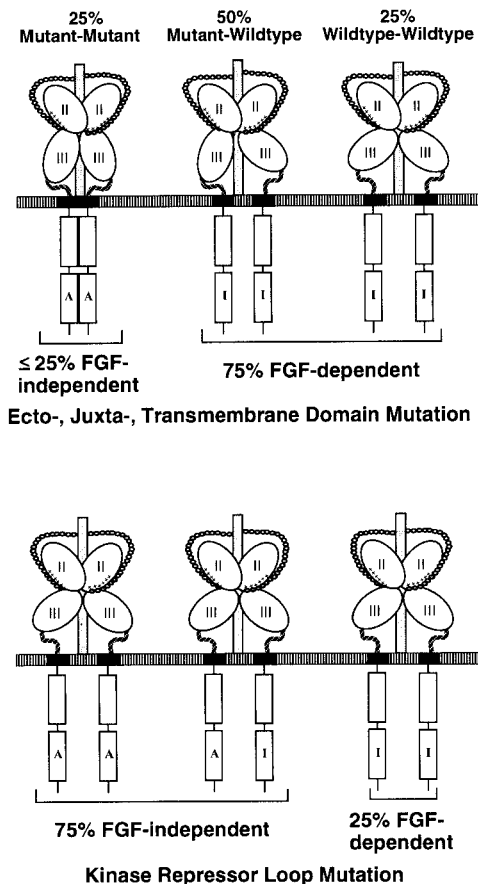


FIG. 15. Mutation site-dependent levels of FGF-independent signal transduction. The top panel indicates the consequence of mutations in the ectodomain, extracellular juxtamembrane domain, or transmembrane domain. Wild-type kinase domains require sustained mutual association either as substrate or enzyme for trans-phosphorylation, which is required for sustained derepression and activation (A) of kinase activity. The bottom panel illustrates the impact of an FGFRTK with an activating mutation in the kinase domain. Association and trans-phosphorylation are not required for activation of constitutively derepressed and active (A) mutant kinases. The percentages below each panel refer to percentage of total signal transduction complexes.

able to achieve a stable enzyme-substrate conformation. The severity of the specific developmental abnormality is proportional to the number of mutant-mutant complexes in the FGF-independent stable state II conformation up to the maximum 25% that can result from the heterozygous condition. Mutations promoting the 25% maximum stable state II complexes or some

critical fractional threshold thereof cause neonatal lethality. The 25% maximum FGF-independent signal or fraction of it is insufficient to impact functions of the same FGFRTK isotype whose activity in adult tissue homeostasis remains FGF dependent. This is consistent with the observation that a heterozygous single amino acid change in the FGFR3 gene causes the skeletal features of achondroplastic dwarfism without apparent impact on normal function of the central nervous system, whereby FGFR3 is expressed in addition to the cartilaginous growth plates of bones (63).

In conditions known as thanatophoric dysplasia (TDI and TDII), several mutations in the FGFR3 gene result in neonatal lethality accompanied by profound skeletal abnormalities. The most severe condition, TDII, is accompanied and probably caused by a lysine-650 to glutamate-650 mutation in the cis-repressor loop of FGFR3 (50, 61) (Fig. 10) that results in a constitutively activated tyrosine kinase in the monomeric form. In contrast to other mutant FGFRTKs, this mutant presumably requires no trans-phosphorylation for derepression and activation of the kinase, and is consequently independent of the interaction with other FGFRTK molecules. Our model predicts that the full 50% of FGFRTK bearing the lysine to glutamate mutation in the kinase domain will be constitutively active independent of whether they are complexed with another mutant or wild-type FGFRTK (Fig. 15, lower panel). This results in 75% of total FGFR complexes that are active independent of FGF. This level of constitutive activity is lethal and results in more profound skeletal and craniofacial phenotypes relative to other mutants whose activity, although still lethal, is capped at 25% of maximum FGF-independent signal. This conclusion is also supported by the fact that gain-of-function mutants with mutations in the ectodomain through the transmembrane domain can be suppressed by coexpression with kinase-defective FGFR, but the kinase-defective FGFR cannot suppress the constitutive activity of the FGFR mutant with the lysine to glutamate change in the repressor domain of the kinase (59).

In the FGFR3 gene, heterozygous mutations that result in sequence extension of the COOH-terminal tail sequence also are neonatally lethal (35), which suggests that the normal COOH-terminal tail also plays a role as a repressor of kinase activity in addition to or in cooperation with the cis-repressor loop within the kinase domain. Heterozygous mutations to cysteines within two subdomains of the FGFR3 extracellular domain are also neonatally lethal (35). Three of these occur in a seven-residue sequence just NH₂ terminal to or within the transmembrane domain in the case of FGFR3, and two occur in the NH₂ terminus of the homeo-interaction domain between Ig loops II and III (Fig. 14). The severity of these mutations relative to others resulting in new cysteine residues at other sites within loop III and noncysteine mutations at the same sites in the homeo-interaction domain are con-

sistent with our model and suggest that the two structural domains are contact domains as shown. Because of their position in a contact domain, new cysteines in the two domains efficiently form stable disulfides that maintain a large percentage of the 25% mutant–mutant complexes in the stable state II conformation, independent of FGF.

Last, despite the fact that the interloop II/III contact domain is 100% conserved among isoforms, the model in Fig. 13 predicts that the heterodimerization among coexpressed FGFRTK ectodomain isoforms may be restricted if the heparan sulfate chain of the FGFHRHS subunit is specific for the FGFRTK subunit of the FGFR complex, and a single heparan sulfate chain or two identical chains from the same proteoglycan core maintain preexisting dimeric FGFRTKs in the state I conformation.

IV. The FGF Family in Liver Growth and Function

A. Analysis of FGF in Hepatocyte-like Hepatoma Cells

The FGF family was first implicated in liver cell growth and function when it was demonstrated that well-differentiated human hepatoma cells (HepG2), which reflect many qualities of liver parenchymal cells, display high-affinity FGFR complexes (53). HepG2 cells exhibit both high- and low-affinity classes of FGFRTK sites in higher total number than most cell types; these can be distinguished from the even more abundant pericellular matrix heparan sulfate sites. In further contrast to most cell types, the predominant activity of FGF-1 was inhibition of HepG2 cell growth rather than the more common stimulation of growth (53). Closer examination of the response of HepG2 cells revealed that FGF-1 elicited stimulation of growth at low concentrations and then inhibited growth as concentration increased. Growth inhibition was accompanied by an increase in secretory activities. Scatchard analysis revealed that the dose-dependent biphasic response with respect to growth correlated with the presence of high- and low-affinity FGFR complexes containing FGFRTK. About 15,000 receptors per cell exhibited an apparent K_d of 9 pM and about 180,000 sites exhibited a K_d of about 2 nM (53). This was in contrast to other hepatoma cell lines, which exhibited a positive mitogenic response to low levels of FGF-1 and exhibited only the high-affinity class of FGFR (53). Direct cloning and mRNA analysis revealed that HepG2 cells expressed high total levels of the FGFR1 gene and numerous splice variants coding for diverse monomeric products with variant ectodomain, intracellular juxtamembrane domain, and intracellular kinase and COOH-terminal domains (20). Further analysis revealed that combinatorial splicing of variants in the ectodomain with variants in the intracellular kinase domain is restricted, e.g., all combinations are not expressed random-

ly. HepG2 cells express the FGFR1 α 1, FGFR1 β 1, and FGFR1 α 2 isoforms, but not the FGFR1 α 2 combination (23). As discussed earlier, the three-Ig loop FGFR α ectodomain is a low-affinity receptor for FGF and the two-loop FGFR β is the high-affinity species of receptor (22). The dose-dependent biphasic response of HepG2 cells to FGF-1 can be explained by the scheme shown in Fig. 16. At low levels of FGF-1, homodimeric complexes of high-

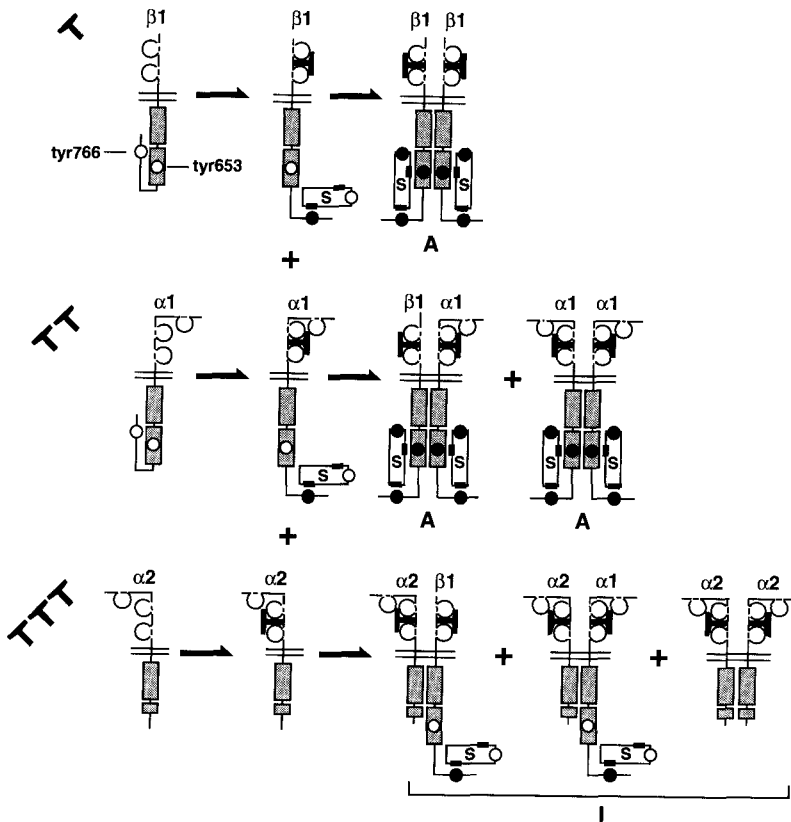


FIG. 16. Dose-dependent control of FGFR signal transduction by regulated combinatorial alternate splicing of extracellular and intracellular domains of FGFR TK. At low levels of FGF (top), only high-affinity FGFR β 1 isoforms are occupied in liver cells, which results in active (A) homodimeric kinase domains inside cells. As FGF levels rise (second row), low-affinity FGFR α 1 isoforms participate in formation of active complexes. As FGF rises to maximum levels (bottom row), the abundant low-affinity FGFR α 1 isoforms begin to compete for type 1 isoforms and eventually dominate by formation of inactive (I) heterodimers with respect to the intracellular domains. A putative substrate (S), PLC γ , which binds to phosphotyrosine-766 through SH $_2$ domains (rectangles), is indicated. Reprinted with permission from Ref. 23, American Society for Microbiology.

affinity FGFR1 β 1 are activated and, as FGF rises, the low-affinity FGFR1 α 1 isoforms begin to participate in formation of active homodimeric complexes with respect to the intact kinase. As FGF continues to rise and occupy the low-affinity kinase-inactive FGFR1 α 2 isoform, which appears to be the predominant FGFR1 isoform after prolonged exposure to FGF-1 (23, 53), inactive heterodimers of FGFR1 α 1 or FGFR1 β with FGFR1 α 2 begin to dominate and dampen proliferation.

After incubation of HepG2 cells with FGF-1 under conditions in which FGFR complexes are internalized, radiolabeled FGF-1 is long-lived inside cells and persists for up to 12 hr (53, 64). A labeled FGF-1 complex of about 40 kDa appears and is resistant to detergents, acid, heat, reducing agents, and other denaturing agents. The complex consists of a 24-kDa cellular component in addition to labeled FGF-1 (64). Immunochemical analysis with FGFRTK- and epitope-specific antibodies reveals that the complex contains a fragment of Ig loop I of the FGFR α isoform of both FGFR1 and FGFR4, which is associated with the particulate fraction of cells (65). Immunocytochemical analysis suggests that the loop I epitope is localized at or in the nucleus of cells. Separate experiments indicate that specifically the full-length FGFR1 α 1 relative to the FGFR1 β 1 kinase migrates to the perinuclear region of cells after FGF-induced internalization (65–67). The functions of the long life and the specific intracellular migration of both FGF polypeptides and FGFRTK remain to be established.

B. Role of the FGF Signal Transduction System in Normal Hepatocytes and Resting and Regenerating Liver

Because of the results with hepatocyte-like HepG2 cells, the mitogenic response to FGF-1, FGF-2, and FGF-7 in primary cultures of normal hepatocytes was examined. FGF-1 stimulated hepatocyte DNA synthesis three- to fivefold with a half-maximum at 60 pM, with 30 to 80% of individual hepatocytes participating in the response (68). FGF-1 was additive with suboptimal levels of EGF; however, in the presence of high levels of EGF, FGF-1 exhibited no stimulatory effect, but inhibited the EGF stimulus. In contrast to HepG2 cells, normal hepatocytes exhibit no response to FGF-2 (68) and exhibit a variable and weak response to FGF-7 relative to FGF-1 (M. Kan and W. L. McKeehan, unpublished results). Analysis of FGFRTK in primary hepatocyte cultures revealed that, in contrast to HepG2 cells, normal parenchymal cells do not express the FGFR1 gene, the expression of which is limited to the nonparenchymal cells (69). Thus the expression of the FGFR1 isoforms in HepG2 cells either indicates an acquisition that occurred during tumorigenesis or the expression pattern of a special hepatocyte-like cell that

gave rise to HepG2 cells. Subsequent mRNA and immunochemical analyses revealed that the HepG2 cells express predominantly the FGFR1 and FGFR4 genes, and less, but significant, levels of the FGFR2IIIc and FGFR3IIIc isoforms (M. Kan and W. L. McKeehan, unpublished results). Primary hepatocytes, when properly prepared free of nonparenchymal cells, display predominantly the FGFR4 gene; traces of the FGFR1 gene are detectable by the polymerase chain reaction (PCR), although all four FGFR genes are expressed in intact liver tissue (M. Kan and W. L. McKeehan, unpublished results). Only the FGFR4 product has been detected by covalent affinity crosslinking to radiolabeled FGF-1 and subsequent immunochemical analysis. Expression of the FGFR2 and FGFR3 genes is probably associated with nonparenchymal epithelial cells or other specialized cell types in liver.

The resting liver and cultured hepatocytes exhibit abundant levels of FGF-1 and FGF-2 proteins, whereas mRNA levels are very low (68). Partial hepatectomy or chemical damage to liver results in an increase in expression of mRNA that peaks at 24 hr, and is at 80% of the peak at 4 hr (68). Because the liver stores of FGF-1 are high prior to the mRNA rise, it is thought that the level of transcription of mRNA for FGF-1 is not a rate-limiting factor for induction of the immediate compensatory response to damage, but rather functions to replenish the abundant stores of FGF-1 that may be depleted as a result of the damage event and the response to it. Despite extensive study, none of eight other FGF polypeptides has been demonstrated to have a significant presence in liver or isolated hepatocytes (M. Kan and W. L. McKeehan, unpublished results). Therefore, FGF-1 and the FGFR4 tyrosine kinase are the current default candidates for mediating the activity of the FGF family in compensatory growth and differentiation of hepatocytes, whereas both FGF-1 and FGF-2 may play a role in the nonparenchymal cell response via FGFR2 and FGFR3 isoforms.

Transforming growth factor beta (TGF β) is a potent inhibitor of EGF-stimulated DNA synthesis in cultured hepatocytes, is up-regulated after liver injury, and depresses the regenerative response (68). TGF β is a much less effective inhibitor of FGF-1-stimulated DNA synthesis in hepatocytes than it is of EGF-stimulated synthesis (68). Although TGF β will inhibit the FGF-1 response at sufficiently high levels, FGF-1 can overcome the inhibition of DNA synthesis at low levels of TGF β . TGF β 1 originates in the nonparenchymal compartment of liver and therefore is a paracrine regulator acting on the parenchymal cell fraction (68).

From our collective results, we have proposed that the compensatory response of liver to loss of mass may unfold in a stepwise fashion, whereby a population of cells is initially recruited into DNA synthesis by FGF-1 (or a yet uncharacterized homolog) and then further expanded by response to

TGF α (the actual EGF homolog present in liver) (11). The former step is mediated by rapid activation of preexisting high-affinity complexes of FGFR4 (or a yet unidentified homolog), FGFRHS, and FGF-1 ligand, a process that is temporarily immune to the inhibitory action of TGF β . TGF α , concurrent with its expansion of the pool of proliferating hepatocytes, induces low-affinity forms of the FGFR complex, which, with rising levels of FGF-1 arising by activation of preexisting FGFR and induction of FGF-1 gene expression, then limits the FGF-dependent pool of proliferating cells. Although the first-stage FGF-stimulated pool of proliferating cells escapes inhibition by TGF β , the second-stage pool expanded by TGF α is extremely sensitive to limitation by the presence of TGF β , and subsequently expansion of the first-stage pool becomes limited as TGF β rises to high levels. We have proposed that it is potentially the exogenous EGF/TGF α (which is added to primary liver cell cultures to achieve maximal DNA synthesis) together with the TGF α produced by cells in the culture that indirectly limit the long-term proliferation and selection of proliferative cell types in primary cultures. The precise relationship of signal transduction by the FGF family and the insulin-like growth factor (IGF) family, and variants of the hepatocyte growth factor (HGF) ligands and the met kinase receptor, both of which also play roles in compensatory growth of liver, has not been established.

V. The FGF Family in Prostate and Prostate Tumors

A. Andromedic Stromal-to-Epithelial Activities of the FGF Family

Unlike the liver, which has a nonparenchymal compartment composed of bile duct epithelial cells, sinusoidal endothelial cells, and a variety of minor cell types, the prostate gland consists of parenchymal and nonparenchymal compartments composed of epithelium and stroma separated by a basement membrane (Fig. 17). Although microheterogeneous in detail, the prostate epithelium can be broadly divided into highly differentiated luminal secretory cells and less differentiated basal epithelial cells; the stromal cells consist largely of smooth muscle cells and fibroblasts. *In vivo*, the prostate is an androgen-dependent organ with respect to maintenance and growth; however, androgen is not a direct-acting mitogen for isolated epithelial cells from the prostate (70). In the absence of androgen, members of several growth factor families (insulin-like growth factors, epidermal growth factor, and possibly others) directly support epithelial cell growth, and it is apparent that one or more of these factors may mediate the action of androgen *in vivo* and may be regulated by androgen (andromedins). The FGF polypeptide

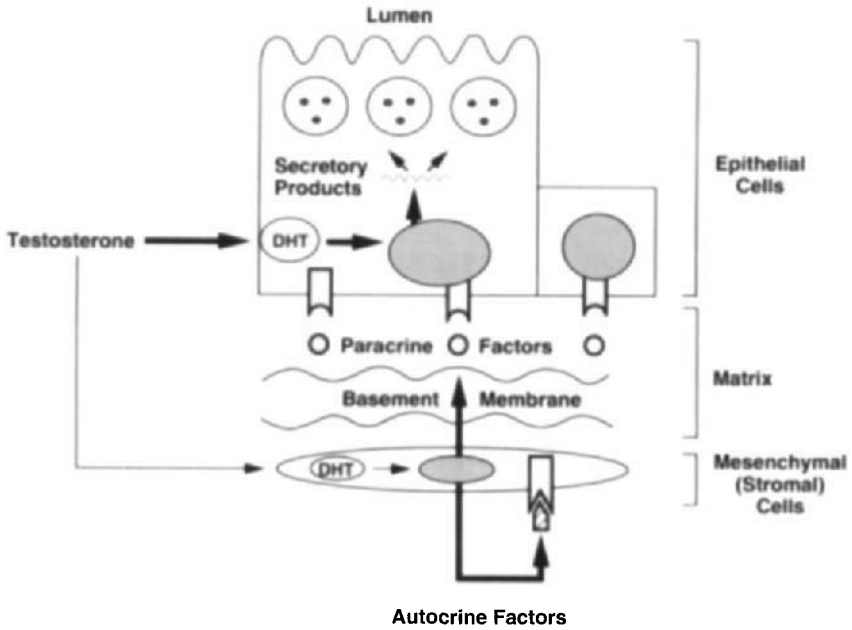


FIG. 17. Schematic of dual, indirect control of stromal-epithelium communications by androgen in the prostate. Testosterone, converted to dihydrotestosterone (DHT) within cells, acts to control directionally specific paracrine factors in the nucleus and autocrine factors in the stroma. Androgen, which exercises well-characterized control of functional, secretory products in the parenchymal epithelial cells in the prostate, possibly also controls reception mechanisms for paracrine stromal factors in the epithelium.

family was first implicated in regulation of prostate cell growth when it was found by direct isolation and protein microsequencing that a major activity in neural or pituitary extracts that supported the growth of isolated rat prostate epithelial cells in primary culture was FGF-1 (prostatropin) (71). In contrast to FGF-2, the first chemically characterized member of the FGF polypeptides that had been isolated earlier, FGF-1 in large part substituted for the crude neural extracts in epithelial cell cultures. Both FGF-1 and FGF-2 exhibit a potent mitogenic effect on prostate stromal cells in culture (72). Although both FGF-1 and FGF-2 are barely detectable in normal adult prostate tissue (FGF-1 > FGF-2) (28), expression of FGF-1 is elevated at progressively higher levels in androgen-responsive (not androgen dependent) Dunning R3327PAP and completely androgen-independent R3327AT3 tumors (73). The former tumor consists of both well-differentiated epithelial and stromal cell compartments, similar to normal tissue, is slow growing, and is relatively benign. The latter tumor, which arises after castration of hosts

bearing the former tumor, consists of undifferentiated epithelial-like cells with no clear stromal compartment, is fast growing, and is malignant (28). FGF-2 expression is markedly increased in the AT3 tumors and cells (73). The elevated FGF-1 expression in the PAP tumors and derived cells is predominantly in the stromal cells (73). Analysis of receptors from normal prostate and the two types of tumors reveals that normal epithelial cells exhibit both low- and high-affinity FGFR complexes for FGF-1, but no receptor sites for FGF-2, whereas normal stromal cells exhibit both classes of sites for both FGF-1 and FGF-2 (72, 73). Attempts to demonstrate that androgen modified FGF-1 or FGF-2 synthesis or activity failed. However, the presence of an androgen-stimulated growth response of epithelial cells when cocultured with stromal cells and the ability of stromal cell products to support directly epithelial cell growth suggested that other unidentified members of the FGF polypeptide family from stromal cells may play an andromedic role in the prostate (72). Subsequent experiments indicated that FGF-7 has the properties of a stromal-derived andromedic regulator of epithelial cell growth and function (28, 72). FGF-7 expression is limited strictly to the stromal cell compartment of both normal prostate and PAP tumors and binds specifically to the FGFR2IIIb isoform, which is expressed strictly in the epithelial cells (28, 72). FGF-7 expression is barely detectable in normal prostate stromal cells in primary culture in the absence of androgen, but is supported in its presence (28, 72). PAP tumor stromal cells in culture exhibit a higher baseline of androgen-independent synthesis of FGF-7 mRNA, which increases constitutively with time in culture, but the level is still responsive to androgen at early times (28, 72). Subsequently, it was demonstrated that both normal prostate epithelial cells and epithelial cells from the PAP tumor express exclusively the FGFR2IIIb splice variant (28), which is not present in the stromal cells (65, 68). Therefore, the androgen-regulated FGF-7 expression in stromal cells that have no receptor for it and the exclusive expression of FGFR2IIIb in epithelial cells that express no FGF to activate it constitute a directionally specific paracrine signaling system between the two cell compartments. In contrast to the normal and PAP tumor prostate epithelial cells, AT3 tumor-derived cells express only the IIIc isoform of the FGFR2 gene (28). In addition, the AT3 cells activate the normally stromal cell FGFR1 gene in addition to FGF-2 (described above) and the FGF-3 and FGF-5 genes. Thus the progression to the androgen-independent, fully malignant state represented by AT3 tumors that arise from castrated hosts bearing the differentiated, and androgen-responsive PAP-like tumors appears to occur by disruption of the directional communication between stromal and epithelial cells via FGF-7 and FGFR2IIIb by splice switching to the FGFR2IIIc isoform (28). This is accompanied by activation of new FGF polypeptides in the epithelial cells capable of acting on the FGFR2IIIc iso-

form, as well as the *FGFR1* gene, whose activation also accompanies the progression. These redundant autocrine loops within the tumor epithelial cells likely underlie their autonomy with respect to growth and environment.

To determine whether these events can occur from a single PAP tumor epithelial cell, PAP tumor epithelial cells expressing exclusively *FGFR2IIIb* and undetectable levels of *FGFR1* and FGF polypeptides were cloned and injected into syngeneic rat hosts alone and in the presence of PAP tumor stromal cells (28). PAP tumor epithelial cells alone progressed over several months into tumors and cells with properties similar to the AT3 tumor cells, e.g., the cells switched from exclusive expression of *FGFR2IIIb* to the *FGFR2IIIc* variant, activated *FGFR1* and FGF-2, FGF-3, and FGF-5. In contrast, when mixed with stromal cells, the differentiated slow-growing tumors that resulted were composed of both epithelial and stromal compartments (28). *FGFR2IIIb* remained the predominant isoform of the *FGFR2* gene expressed in the tumors, although a low level of *FGFR2IIIc* could be detected associated with appearance of small foci of undifferentiated cells. No activation of *FGFR1* or FGF-2 was apparent. These results suggest that the switch from exclusively the *FGFR2IIIb* isoform in nonmalignant epithelial cells is delayed by the presence of stromal cells and may be an early clonal and irreversible event that relieves epithelial cells from the control of their stroma. This event alone is insufficient to confer autonomy; however, relieved of the pressure from stromal signals that control proliferation and promote differentiation, epithelial cells may survive to acquire additional changes that confer autonomy as activation of FGF-2 to complete an autocrine support system (Fig. 18).

A notable difference between normal prostate epithelial cells and those from the R3327PAP tumor was the loss of a low-affinity class of FGFR complexes (73). As discussed earlier (22, 23), alternately spliced NH₂-terminal loop I of FGFRTK confers a 5- to 10-fold decrease in affinity for FGF, and in liver cells appears to be associated with inhibition of cell growth. Comparative analysis of the *FGFR2* gene products in normal prostate and PAP tumor tissue revealed that the *FGFR2* α , *FGFR2* β , and *FGFR2* $\beta(\Delta AE)$ (*FGFR* δ ; see Fig. 4) ectodomain variants, which exhibit low, high, and high affinity, respectively, are present in normal prostate (22). However, the single variant *FGFR* δ missing loop I and the acidic box is expressed in the PAP epithelial cells and persists in the derived malignant AT3 cells (22). Currently, it is unclear whether the three *FGFR2* ectodomain isoforms are expressed in the same or different epithelial cells in normal prostate. Clearly the loss of the *FGFR* α splice variant, which represents the low-affinity class of FGFRTK and is specifically targeted to the perinuclear/nuclear compartment (65–67), occurs in prostate tumor epithelial cells and correlates with malignancy of pancreatic and brain tumor cells as well (74, 75). In the an-

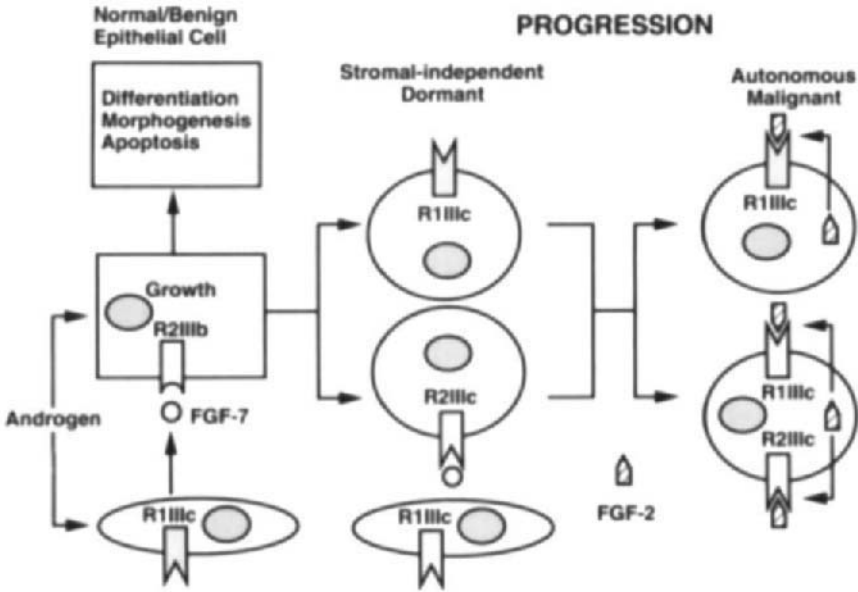


FIG. 18. Schematic of progression of prostate tumors from paracrine control by stroma, to the stromal-independent state, to malignancy. Dual indirect control of paracrine communication from stroma to epithelium is mediated by the FGF-7 signal from stroma and the FGFR2IIIb reception system in epithelium (see Fig. 17). FGF-7 drives epithelial cell growth and associated differentiation, morphogenesis, and apoptosis to maintain homeostasis in normal tissue and benign nonmalignant androgen-responsive tumors. Clones of epithelial cells that arise switch at splicing from FGR2IIIb to FGFR2IIIc or activate the FGFR1 gene. Independent of the pressure to terminally differentiate stromal signals, the cells lie dormant until FGF-2 or other FGFs are activated to complete an autocrine support system, resulting in selection for the malignant phenotype.

drogen-independent, malignant prostate tumors that have activated the normal stromal cell FGFR1 gene, the FGFR1 α isoform is suppressed by splicing of the loop I exon at an alternate 3' acceptor site (24). The alternate splice event results in a deletion of five base pairs, a reading frame shift, and a stop codon downstream of the splice junction, resulting in either an intracellular product or an unproductive mRNA with respect to a transmembrane product. It remains to be established whether combinatorial splicing of the FGFR2 ectodomain with variants in the intracellular kinase domain is regulated as described earlier for the FGFR1 gene in liver cells. In the normal rat prostate, the intact type 1 FGFR2 intracellular domain and a splice variant FGFR2-2 (Fig. 4), which results from skipping of exon 16, are expressed in about equal proportions (45). The FGFR2-2 isoform exhibits most of the ki-

nase domain and associated tyrosine phosphorylation sites, but does not exhibit the counterpart of FGFR1 tyrosine-766, which, when phosphorylated, is required for association with and phosphorylation of PLC γ by FGFRTK (23, 42).

B. Switch from Homeostatic Paracrine Regulator to Autocrine Autonomy during Prostate Tumor Progression

From combined results from our laboratory using rat prostate and the Dunning R3327 tumors and derived cells, we propose that members of the FGF family play key roles in prostate homeostasis and that a series of step-wise changes in the FGF family may contribute to the progression of prostate tumors to malignancy (Fig. 18). The directionally specific paracrine communication from stromal to epithelial cells mediated by stromal cell-derived FGF-7 (or homologs) and epithelial cell-derived FGFR2IIIb contributes to normal prostate homeostasis, impacting cell growth, morphogenesis, differentiation, and apoptosis. The indirect control of these processes by androgen is partitioned in the stromal compartment through control of signal, e.g., control of FGF-7 (or homologs) activity, and in the epithelium by control of reception, e.g., activity and impact of FGFR complexes. Although less stringently controlled by androgen and labile with respect to quantitative aspects, the homeostatic FGF-7/FGFR2IIIb paracrine system remains intact to some extent in well-differentiated, slow-growing, nonmalignant tumors. Contributing to the tumor phenotype is the change in alternate splicing mechanisms and selection for epithelial cells no longer expressing the growth-controlling FGFR2 α isoform. Long-term survival of an epithelial cell that no longer exclusively expresses the FGFR2IIIb isoform—e.g., irreversible exon switching to FGFR2IIIc, loss of the FGFR2 gene, or activation of the FGFR1 gene—results in a new epithelial cell phenotype, less dependent on the control of stromal cells, that is mediated by FGF-7 (or homologs) and androgen. This single change alone is insufficient to confer autonomy and the accompanying malignancy, but allows undifferentiated cells to survive for long periods in a dormant state, independent on stromal influence. Activation of FGF-2 or other FGF polypeptides that activate the new FGFR phenotype confers autonomy and enhances evolution of the malignant cell phenotype.

REFERENCES

1. D. M. Ornitz, J. Xu, J. S. Colvin, D. G. McEwen, C. A. McArthur, F. Coulier, F. Gao, and M. Goldfarb, *J. Biol. Chem.* **271**, 15292 (1996).
2. M. Yamasaki, A. Miyaki, S. Tagashira, and N. Itoh, *J. Biol. Chem.* **271**, 15918 (1996).

3. A. S. Eriksson, L. S. Cousens, L. H. Weaver, and B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3441 (1991).
4. J. Zhang, L. S. Cousens, P. J. Barr, and S. R. Sprang, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3446 (1991).
5. F. J. Moy, A. P. Seddon, E. B. Campbell, P. Bohlen, and R. Powers, *J. Biomol. NMR* **6**, 10965 (1995).
6. X. Zhu, H. Komiya, A. Chirino, S. Faham, G. M. Fox, T. Arakawa, B. T. Hsu, and D. C. Rees, *Science* **251**, 90 (1991).
7. P. M. Smallwood, I. Munoz-Sanjuan, P. Tong, J. P. Macke, S. H. C. Hendry, D. J. Gilbert, N. G. Copeland, N. A. Jenkins, and J. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9850 (1996).
8. W. H. Burgess and T. Maciag, *Annu. Rev. Biochem.* **58**, 575 (1989).
9. M. Jaye, J. Schlessinger, and C. A. Dionne, *Biochim. Biophys. Acta* **1135**, 185 (1992).
10. D. E. Johnson and L. T. Williams, *Adv. Cancer Res.* **60**, 1 (1993).
11. M. Kan and W. L. McKeehan, "Liver Growth and Repair. D. Growth Factors and their Receptors. ii. FGF (A. Strain and A. M. Diehl, eds.). Chapman & Hall, London, 1997. (In press.)
12. S. Faham, R. E. Hileman, J. R. Fromm, R. J. Linhardt, and D. C. Rees, *Science* **271**, 1116 (1996).
13. L. D. Thompson, M. W. Pantoliano, and B. A. Springer, *Biochemistry* **33**, 3831 (1994).
14. Y. Luo, J. L. Gabriel, F. Wang, X. Zhan, T. Maciag, M. Kan, and W. L. McKeehan, *J. Biol. Chem.* **271**, 26876 (1996).
15. B. A. Springer, M. W. Pantoliano, F. A. Barbera, P. L. Gunyuzlu, L. D. Thompson, W. F. Herblin, S. A. Rosenfeld, and G. W. Book, *J. Biol. Chem.* **269**, 26879 (1994).
16. W. L. McKeehan and J. W. Crabb, *Analyt. Biochem.* **164**, 563 (1987).
17. M. Kan, E. Shi, and W. L. McKeehan, *Meth. Enzymol.* **198**, 158 (1991).
18. A. Jackson, F. Tarantini, S. Gamble, S. Friedman, and T. Maciag, *J. Biol. Chem.* **270**, 33 (1995).
19. D. E. Johnson, J. Lu, H. Chen, S. Werner, and L. T. Williams, *Mol. Cell. Biol.* **11**, 4627 (1991).
20. J. Hou, M. Kan, K. McKeehan, G. McBride, P. Adams, and W. L. McKeehan, *Science* **251**, 665 (1991).
21. J. Xu, M. Nakahara, J. W. Crabb, E. Shi, Y. Matuo, M. Fraser, M. Kan, J. Hou, and W. L. McKeehan, *J. Biol. Chem.* **267**, 17792 (1992).
22. F. Wang, M. Kan, G. Yan, J. Xu, and W. L. McKeehan, *J. Biol. Chem.* **270**, 10231 (1995).
23. E. Shi, M. Kan, J. Xu, J. Hou, G. McBride, and W. L. McKeehan, *Mol. Cell. Biol.* **13**, 3907 (1993).
24. G. Yan, F. Wang, Y. Fukabori, D. J. Sussman, J. Hou, and W. L. McKeehan, *Biochem. Biophys. Res. Commun.* **183**, 423 (1992).
25. M. Kan, F. Wang, J. Xu, E. Shi, J. W. Crabb, J. Hou, and W. L. McKeehan, *Science* **259**, 1918 (1993).
- 25a. F. Wang, M. Kan, K. McKeehan, S. Feng, and W. L. McKeehan, *J. Biol. Chem.* **272** (in press) (1997).
26. H. Margalit, N. Fischer, and S. A. Ben-Sasson, *J. Biol. Chem.* **268**, 19228 (1993).
27. G. Yan, Y. Fukabori, S. Nikolaropoulos, F. Wang, and W. L. McKeehan, *Mol. Endocrinol.* **6**, 2123 (1992).
28. G. Yan, Y. Fukabori, G. McBride, S. Nikolaropoulos, and W. L. McKeehan, *Mol. Cell. Biol.* **13**, 4513 (1993).
29. S. Vainikka, J. Partanen, P. Bellosta, F. Coulier, C. Basilico, M. Jaye, and K. Alitalo, *EMBO J.* **11**, 4273 (1992).
30. D. S. Duan, S. Werner, and L. T. Williams, *J. Biol. Chem.* **267**, 16076 (1992).
31. M. Kan, F. Wang, M. Kan, B. To, J. Gabriel, and W. L. McKeehan, *J. Biol. Chem.* **27**, 26143 (1996).

32. M. Muenke and U. Schell, *Trends Genet.* **11**, 308 (1995).
33. X. Li, W. J. Park, R. E. Pyeritz, and E. W. Jabs, *Nature (Genet.)* **9**, 232 (1995).
34. N. J. Fasel, M. Bernard, N. Delgon, M. Rousseaux, R. J. Eisenberg, C. Bron, and G. H. Cohen, *Biochem. Biophys. Res. Commun.* **178**, 8 (1991).
35. J. Bonaventure, F. Rousseau, L. Legeai-Mallet, M. Le Merrer, A. Munnich, and P. Maroteaux, *Am. J. Med. Genet.* **63**, 148 (1996).
36. M. K. Webster and D. J. Donoghue, *EMBO J.* **15**, 520 (1996).
37. C. L. Johnston, H. C. Cox, J. J. Gomm, and R. C. Coombes, *J. Biol. Chem.* **270**, 30463 (1995).
38. L. L. Gillispie, G. Chen, and G. D. Paterno, *J. Biol. Chem.* **270**, 22758 (1995).
39. J. Hou, K. McKeehan, M. Kan, S. A. Carr, M. J. Huddleston, J. W. Crabb, and W. L. McKeehan, *Protein Sci.* **2**, 86 (1993).
40. M. Mohammadi, I. Dikic, A. Sorokin, W. H. Burgess, M. Jaye, and J. Schlessinger, *Mol. Cell. Biol.* **16**, 977 (1996).
41. M. Mohammadi, J. Schlessinger, and S. R. Hubbard, *Cell* **86**, 577 (1996).
42. M. Mohammadi, A. M. Honegger, D. Rotin, R. Fischer, F. Bellot, W. Li, C. A. Dionne, M. Jaye, M. Rubinstein, and J. Schlessinger, *Mol. Cell. Biol.* **11**, 5068 (1991).
43. M. Mohammadi, C. A. Dionne, W. Li, T. Spivak, A. M. Honegger, M. Jaye, and J. Schlessinger, *Nature (London)* **358**, 681 (1992).
44. K. G. Peters, J. Marie, E. Wilson, H. E. Ives, J. Escobedo, M. Del Rosario, D. Mirda, and L. T. Williams, *Nature (London)* **358**, 678 (1992).
45. G. Yan, G. McBride, and W. L. McKeehan, *Biochem. Biophys. Res. Commun.* **194**, 512 (1993).
46. P. Champion-Arnaud, C. Ronsin, E. Gilbert, M. C. Gesnel, E. Houssaint, and R. Breathnach, *Oncogene* **6**, 979 (1991).
47. Y. Hattori, H. Odagiri, H. Y. Nakatani, K. Miyagawa, K. Naito, H. Sakamoto, O. Katoh, T. Yoshida, T. Sugimura, and M. Terada, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5983 (1990).
48. M. Seno, R. Sasada, T. Watanabe, K. Ishimaru, and K. Igarishi, *Biochim. Biophys. Acta* **1089**, 244 (1991).
49. L. Y. Wang, S. P. Edenson, Y. L. Lu, L. Senderowicz, and C. W. Turck, *Biochemistry* **35**, 10134 (1996).
50. P. L. Tavormina, R. Shiang, L. M. Thompson, Y. Z. Zhu, D. J. Wilkin, R. S. Lachman, W. R. Wilcox, D. L. Rimoim, D. H. Cohn, and J. J. Wasmuth, *Nature (Genet.)* **9**, 321 (1995).
51. M. W. Pantoliano, R. A. Horlick, B. A. Springer, D. E. Van Dyk, T. Tobery, D. R. Wetmore, J. D. Lear, A. T. Nahapetian, J. D. Bradley, and W. P. Sisk, *Biochemistry* **33**, 10229 (1994).
52. F. Y. Zhou, M. Kan, R. T. Owens, W. L. McKeehan, J. A. Thompson, R. J. Linhardt, and M. Hook, *Eur. J. Cell Biol.* **73**, 71 (1997).
53. M. Kan, D. DiSorbo, J. Hou, H. Hoshi, P. E. Mansson, and W. L. McKeehan, *J. Biol. Chem.* **263**, 11306 (1988).
54. A. Yayon, M. Klagsbrun, J. D. Esko, P. Leder, and D. M. Ornitz, *Cell* **64**, 841 (1991).
55. T. Spivak-Kroizman, M. A. Lemmon, I. Dikic, J. E. Ladbury, D. Pinchasi, J. Huang, M. Jaye, G. Crumley, and J. Schlessinger, *Cell* **79**, 1015 (1994).
56. W. L. McKeehan and M. Kan, *Mol. Reproduct. Dev.* **39**, 69 (1994).
57. F. Wang, M. Kan, J. Xu, G. Yan, and W. L. McKeehan, *J. Biol. Chem.* **270**, 10222 (1995).
58. A. P. Seddon, D. Aviezer, L. Lu-Yuan, P. Bohlen, and A. Yayon, *Biochemistry* **34**, 731 (1995).
59. K. M. Neilson and R. E. Friesel, *J. Biol. Chem.* **271**, 25049 (1996).
60. M. C. Naski, Q. Wang, J. Xu, and D. M. Ornitz, *Nature (Genet.)* **13**, 233 (1996).
61. M. K. Webster, P. Y. D'Avis, S. C. Robertson, and D. J. Donoghue, *Mol. Cell. Biol.* **16**, 4081 (1996).
62. J. Hou, M. Kan, J. Xu, M. Nakahara, G. McBride, K. McKeehan, and W. L. McKeehan, *J. Biol. Chem.* **267**, 17804 (1992).

63. K. Peters, D. Ornitz, S. Werner, and L. Williams, *Dev. Biol.* **155**, 423 (1993).
64. E. Shi, M. Kan, J. Xu, and W. L. McKeehan, *J. Biol. Chem.* **266**, 5774 (1991).
65. S. Feng, J. Xu, F. Wang, M. Kan, and W. L. McKeehan, *Biochim. Biophys. Acta* **1310**, 67 (1996).
66. I. Prudovsky, N. Savion, X. Zhan, R. Friesel, J. Xu, J. Hou, W. L. McKeehan, and T. Maciag, *J. Biol. Chem.* **269**, 31720 (1994).
67. I. A. Prudovsky, N. Savion, T. M. LaVallee, and T. Maciag, *J. Biol. Chem.* **271**, 14198 (1996).
68. M. Kan, J. Huang, P.-E. Mansson, H. Yasumitsu, B. Carr, and W. L. McKeehan, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7432 (1989).
69. M. Kan, G. Yan, J. Xu, M. Nakahara, and J. Hou, *In Vitro Cell. Dev. Biol.* **28A**, 515 (1992).
70. W. L. McKeehan, P. S. Adams, and M. P. Rosser, *Cancer Res.* **44**, 1998 (1984).
71. J. W. Crabb, L. G. Armes, S. A. Carr, C. M. Johnson, G. D. Roberts, R. S. Bordoli, and W. L. McKeehan, *Biochemistry* **25**, 4988 (1986).
72. G. Yan, Y. Fukabori, S. Nikolaropoulos, F. Wang, and W. L. McKeehan, *Mol. Endocrinol.* **6**, 2123 (1992).
73. P. E. Mansson, P. Adams, M. Kan, and W. L. McKeehan, *Cancer Res.* **49**, 2485 (1989).
74. F. Yamaguchi, H. Saya, J. M. Bruner, and R. S. Morrison, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 484 (1994).
75. M. S. Kobrin, Y. Yamanaka, H. Friess, M. E. Lopez, and M. Korc, *Cancer Res.* **53**, 4741 (1993).

The Ribosomal Elongation Cycle and the Movement of tRNAs across the Ribosome

KNUD H. NIERHAUS,* HEINRICH B. STUHRMANN,[†] AND DMITRI SVERGUN^{‡,1}

**Max-Planck-Institut für Molekulare Genetik*

D-14195 Berlin, Germany

[†]*GKSS Forschungszentrum*

Makromolekulare Strukturforschung

D-21502 Geesthacht, Germany

[‡]*European Molecular Biology Laboratory*

D-22603 Hamburg, Germany

| | |
|---|-----|
| I. Introduction | 178 |
| II. Functional Aspects: Models of the Elongation Cycle | 180 |
| A. Allosteric Three-Site Model | 180 |
| B. Hybrid-Site Model | 182 |
| C. α - ϵ Model | 184 |
| III. Structural Aspects: The Shape of Ribosomes and the Localization of tRNAs | 188 |
| A. The Shape of Ribosomes | 188 |
| B. The Arrangements of tRNAs before and after Translocation | 199 |
| IV. Conclusions | 201 |
| References | 202 |

Ribosome research has reached an exciting state, where two lines of experimental research have considerably improved our understanding of the ribosomal functions. On one hand, functional analysis has elucidated principles of both the decoding process and the tRNA movement on the ribosome during translocation. Experimental data leading to current competing models of the ribosomal elongation cycle can be reconciled by a new model, the α - ϵ model, according to which both tRNAs are tightly bound to a movable ribosomal domain. This α - ϵ domain carries the tRNA₂ mRNA complex from the A and P sites to the P and E sites in the course of translocation maintaining the binding of both tRNAs. On the other hand, the location of tRNAs within the elongating ribosome can be directly determined for the first time by neutron scattering and electron microscopy. Both lines of evidence complement each other and define a frame for the first experimentally sound functional model of the elongating ribosome. © 1998 Academic Press

¹ On leave from Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia.

I. Introduction

Three phases can be distinguished in protein synthesis: initiation, elongation, and termination. The phases are steered and driven by specific proteins, the so-called factors. Accordingly, we distinguish initiation, elongation, and termination factors. The central elongation phase is also termed a "cycle," because a cycle of three basic reactions is used to elongate the nascent peptide chain by one amino acid (see Fig. 1).

The first reaction of the elongation cycle is the decoding step, i.e., the selection of the cognate ternary complex aminoacyl-tRNA·EF·Tu·GTP according to the codon exposed at the decoding center on the small subunit of the ribosome. The second reaction is the peptidyl transfer of the nascent peptide chain to the amino acid of the newly bound aminoacyl-tRNA, with the corresponding active center being located on the large ribosomal subunit. The third reaction is translocation, which is the movement of the ribosome, relative to the mRNA, one codon length.

During the past few years our knowledge of each of these steps has progressed considerably. The selection of the cognate ternary complex proceeds

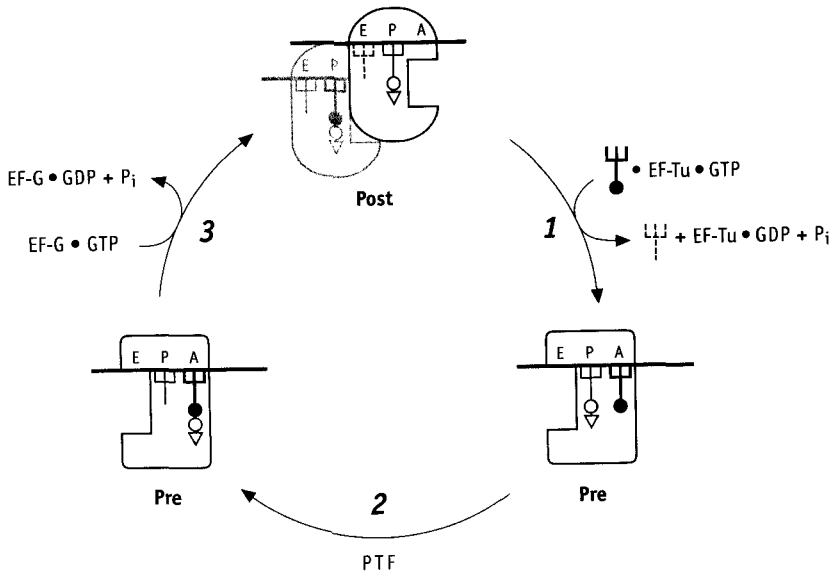


FIG. 1. The allosteric three-site model (see Ref. 1 for review): 1, decoding step during occupation of the A site; 2, peptide-bond formation (the nascent peptide chain is linked to the newly bound aminoacyl-tRNA); 3, translocation reaction. PTF, Peptidyltransferase. For details, see text.

in at least two subreactions. The first one probably allows only codon-anticodon contacts, excluding contacts between EF-Tu and the ribosome as well as between tRNA outside the anticodon stem-loop structure and the ribosome (1). A consequence of this feature is that about 35 noncognate ternary complexes in *Escherichia coli* (i.e., complexes containing aminoacyl-tRNAs with anticodons dissimilar to the cognate one) do not interfere with the selection process, a fact that has been previously demonstrated experimentally. Therefore the selection process has to deal with only five to seven ternary complexes containing the cognate ternary complex (correct anticodon with respect to the codon at the decoding center) and the near-cognate ternary complexes (anticodons similar to the cognate one). The precision of the selection during this subreaction probably does not require a proofreading mechanism [a proofreading mechanism is defined as a process that exploits more than once the accuracy-defining parameter of codon-anticodon interactions (2)], but rather the initial binding reaction provides sufficient discrimination energy to account for the precision of the selection process (3).

After the decoding process is finished the ternary complex is bound tightly to the ribosome in the second subreaction, during which the EF-Tu-dependent GTPase center is activated by an as yet unknown mechanism resulting in a release of EF-Tu-GDP (4). The newly bound aminoacyl-tRNA is now ready to accept the nascent peptide chain via a peptide bond.

Peptide bond formation is the second basic reaction of the elongation cycle. A minimal set of candidates for participation in this reaction comprises the ribosomal proteins L2, L3, and L4 and the 23S rRNA (5). Especially the role of 23S rRNA has been studied (6), and the participation of single nucleotides of the 23S rRNA of *E. coli* ribosomes at the peptidyl transferase center has been demonstrated. The universally conserved bases G2252 (7) and G2581 (8) are probably involved in fixation of the third and the second to last nucleotides respectively, of the universally conserved CCA-3' end of all tRNAs at the ribosomal P site. A current view of the third basic reaction, the translocation reaction, will be considered in Section II,C.

In this chapter we deal exclusively with a general aspect of the elongation phase. As a first step toward an understanding of protein synthesis at the molecular level we have to know where the tRNAs are located within the ribosome and how they are positioned and mutually arranged. The second step is the analysis of the topography of the tRNAs, i.e., the identification of the neighboring components and molecular groups such as nucleotides and aminoacyl residues. We have already accumulated a remarkable amount of knowledge concerning the second step (9, 10). In contrast, only recently have methods become available that allow a direct measurement of the position of tRNAs (11, 12), the results of which will be reviewed here. Through such

measurements, we have gathered new insights concerning localization and arrangement of the tRNAs on the ribosome in the course of protein synthesis. The reason for the slow progress is that the localization of tRNAs is a formidable task of structural analysis due to the sheer size of the ribosomes: at least two tRNAs are on the elongating ribosome, and these tRNAs, with a total molecular mass of 50 kDa, represent only about 2% of the molecular mass of a ribosome (about 2500 kDa in prokaryotes).

We begin in Section II with an overview of current models of the elongation cycle and close this section with a description of a new model (the α - ϵ model) that reconciles the various observations on which the previous models were based. The new model makes some predictions concerning the mutual arrangements of the tRNAs on the ribosome during the elongation phase. In Section III we survey the new results concerning the tRNA locations in the ribosome and compare these results with the predictions made by the functional α - ϵ model.

II. Functional Aspects: Models of the Elongation Cycle

Three tRNA binding sites have been found on ribosomes derived from organisms of all evolutionary domains, i.e. on ribosomes of bacteria (*E. coli*) (13, 14), archaea (*Halobacterium halobium*) (15), and eukarya (yeast) (16). In contrast, four tRNA binding sites were described on ribosomes of mammalia (80S ribosomes from rabbit liver) (17). However, this finding could not be confirmed in a recent analysis, wherein only three tRNA binding sites could be demonstrated on ribosomes from the rabbit liver (18a). Therefore, the existence of three tRNA binding sites seem to be a universal feature of ribosomes.

The way in which the three sites cooperate in the course of protein synthesis is described differently by current models of the elongation cycle. Essentially two models for the elongation cycle have been proposed: the allosteric three-site model and the hybrid-site model. Both models are based on different experimental observations and are therefore characterized by different features. To begin with, the two models will be outlined, and this outline will be followed by a description of the lines of evidence that have led to a new model, the α - ϵ model.

A. Allosteric Three-Site Model

Functional characterizations of the three tRNA binding sites have led to the allosteric three-site model. This model has two essential features:

1. The first and third sites, the A and E sites, respectively, are coupled in a reciprocal fashion (1, 16, 18, 19). This means that occupation of the

E site induces a low affinity of the A site and, vice versa, occupation of the A site drastically reduces the affinity of the E site, with the consequence that a deacylated tRNA present at this site is expelled. This feature can also be demonstrated with native polysomes (20). Consequently, the elongating ribosome contains two tightly bound tRNAs, at the A and P sites before translocation and at the P and E sites after translocation (see Fig. 1). A second consequence is that the ribosome oscillates between two main states in the course of elongation, the state before and that after translocation. These two main states are separated by high activation-energy barriers (more than 80 kJ/mol), and the transition from one state to the other is catalyzed by the elongation factors (4).

2. Both tRNAs present on the elongating ribosome are simultaneously interacting with the corresponding codons of the mRNA via base pairing between anticodon and codon (1, 16, 19).

The two features probably explain essential properties of the ribosomes. The first feature seems to be related to the remarkable selection power of the ribosome: in *E. coli*, 41 virtually identical ternary complexes aminoacyl-tRNA·EF-Tu·GTP (21) compete for the A site, and discrimination among them rests on codon–anticodon interactions, the correctness of which has to be recognized by the ribosomal decoding center. The decoding center is an integral part of the A site and resides on the small ribosomal subunit. Occupation of the E site induces the low-affinity state of the A site, which obviously means that interactions of the large ternary complex are reduced to those at the decoding center, thus avoiding those interactions with the A site that do not participate in the discrimination process. For example, EF-Tu within the ternary complex increases the affinity of the aminoacyl-tRNA to the A site by several orders of magnitude (22), but because this factor (contributing two-thirds of the mass of the ternary complex) is identical in all 41 complexes, the interaction of this factor with the ribosomal A site would disturb rather than contribute to the discrimination process. Indeed, it can be shown that a ribosome with a free E site incorporates even noncognate aminoacyl-tRNAs with an anticodon not at all complementary to the codon at the A site (codon UUU, incorporation of Asp-tRNA, corresponding codon GAC). Such a disastrous incorporation was not observed when the E site was occupied with a cognate deacylated tRNA (23).

It has been speculated (1) that the second feature (two adjacent codon–anticodon interactions) might explain the astonishing capacity of the ribosome for keeping the frame, with a frameshift frequency of less than 1 per 6000 decoding events (24). The idea is that the ribosome provides a “window” that allows just six base pairs between two adjacent codons and the an-

ticodons of the corresponding tRNAs, and that any disturbance of this window resulting in a reduction or increase of the allowed number of six base pairs might induce frameshifts. Stereochemical arguments support this view (24a). However, this interesting hypothesis still lacks experimental proof.

Although the model explains some important traits of the protein-synthesizing machinery, it does not shed light on what is probably the most mysterious reaction of translation, namely the mechanism of translocation. How do molecules as large as tRNAs and mRNA move across the surface of the ribosome between both subunits? We will see in the Section II,B how this problem is tackled by the hybrid-site model.

B. Hybrid-Site Model

This model was proposed by Harry Noller and his group (25). It is based on experimental data that were obtained by protection experiments. Reagents that specifically modify one of the four bases of an RNA were used. Prerequisites of modification of a base are that the base is unpaired and that access of the modifying reagent is not hindered. The modified base can be identified by reverse transcription starting from a defined probe complementary to a distinct sequence of the RNA. The reverse transcriptase cannot read the modified base, which therefore blocks the transcription at this point. The corresponding fragment can be detected on a sequence gel. When a tRNA is bound to a defined site of the ribosome, an analysis of the modification pattern allows identification of unpaired bases that are protected by the tRNA. The protected base is probably located at or near the binding region of the tRNA, although indirect (allosteric) effects cannot be excluded without further experiments.

The intriguing observation was that the protection pattern of the 23S rRNA did not change on translocation, in contrast to that of the 16S rRNA. Therefore, the pattern of the 16S rRNA of the small subunit followed the expected behavior, i.e., it changed on translocation, thus reflecting the movement of the tRNA from the A to the P site; protected bases of the 16S rRNA were not found when the deacylated tRNA was at the E site. In striking contrast, an AcPhe-tRNA at the A site protected the same set of bases of the 23S rRNA after translocation to the P site. The same was true for the deacylated tRNA when translocated from the P site to the E site. Because a Phe-tRNA at the A site showed a somewhat different pattern on the 23S rRNA compared to an AcPhe-tRNA, an analog of a peptidyl-tRNA, the conclusion was that the tRNAs on the ribosome move from A to P and from P to E on the large subunit on peptide formation, whereas the corresponding movement on the small subunit occurs at the subsequent translocation step (25). Accordingly, it was inferred that an aminoacyl-tRNA that is at the A site before peptide-bond formation occupies a "hybrid-site" A/P after peptide-bond for-

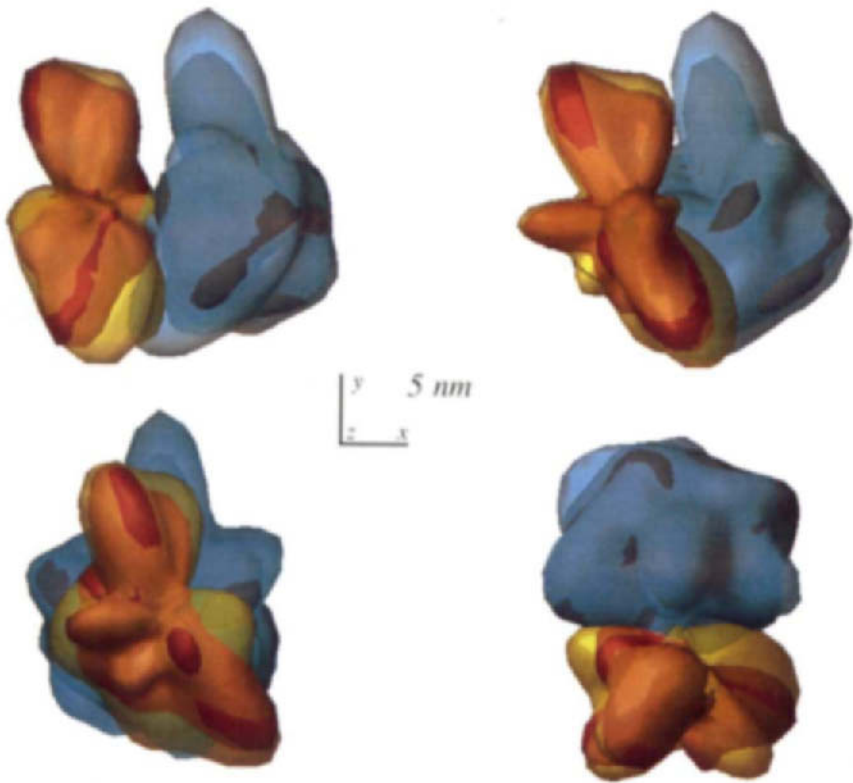


FIG. 7 Low-resolution four-phase model of the 70S ribosome obtained from contrast variation solution scattering data. Several orientations are presented, with the 30S subunit displayed in yellow and the 50S in blue. The semitransparent envelopes of the subunits allow the inner envelopes of the corresponding RNA moieties to be seen. The regions where RNA protrudes to the surface of the subunits are displayed in purple (30S) and dark blue (50S). Reproduced from Ref. 65.

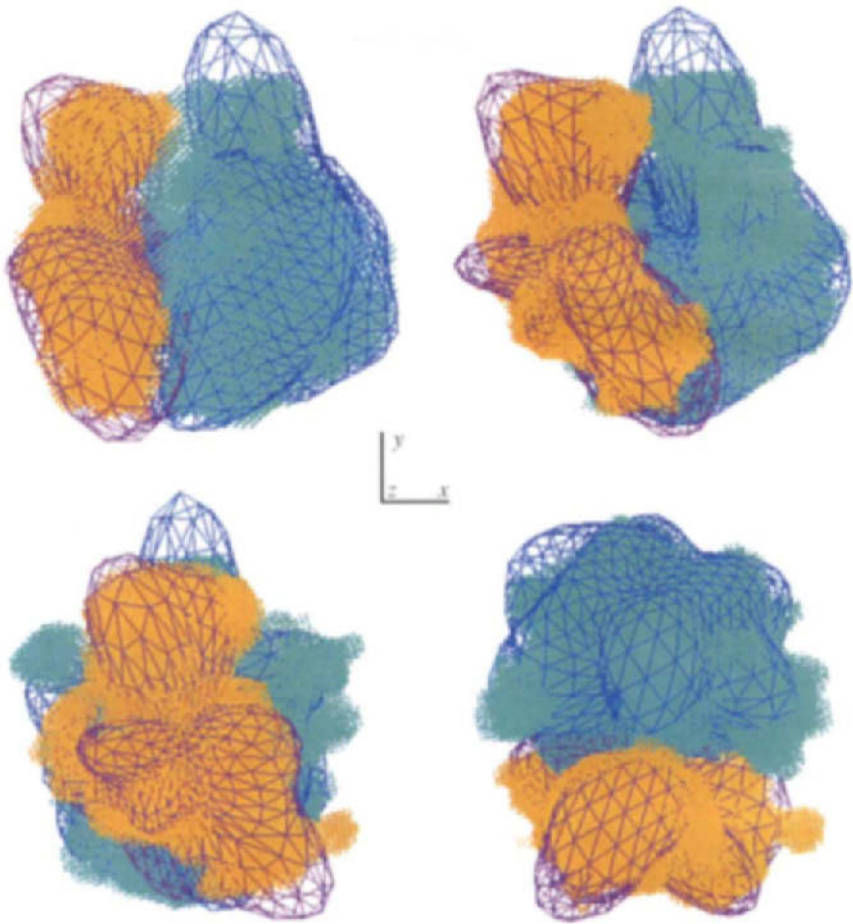


FIG. 9 Comparison of the envelopes of the 30S and 50S subunits (represented by wire frames) in the four-phase model and the Frank model (pixels; 30S is displayed in yellow, 50S in blue). The orientations are the same as in Fig. 7. Reproduced from Ref. 65.

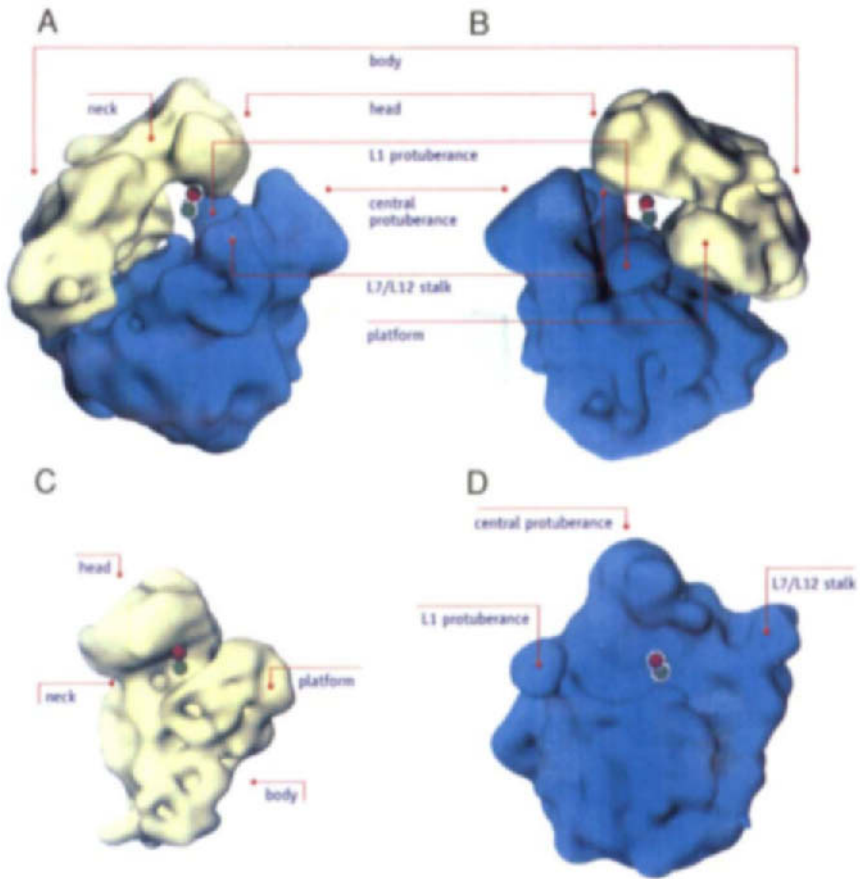


FIG. 10 Location of the mass centers of gravity of the mRNA-tRNA₂ complex. The 70S ribosome is the Frank model from 1995 (42), with the 50S subunit in blue and the 30S subunit in yellow. The green sphere indicates the mass center in the pretranslocational state, red in the posttranslocational state. (A) View from the L7/L12 side. (B) View from the L1 side. (C) View to the interface of the 30S subunit (50S is made invisible). (D) View into the interface of the 50S subunit (30S invisible). Reproduced from Ref. 12.

This Page Intentionally Left Blank

mation, i.e., it is present at the A-site region of the small subunit but already at the P-site region of the large subunit (Fig. 2). After translocation the peptidyl-tRNA is at the P site (P/P) and the deacylated tRNA is at the E site.

This model has the following attractive consequence: the tRNAs seem to move independently along the two subunits at different steps of the elongation cycle. The advantage is similar to that of an artist on the trapeze when he holds on to a swing with at least one limb (hand or leg) while progressing

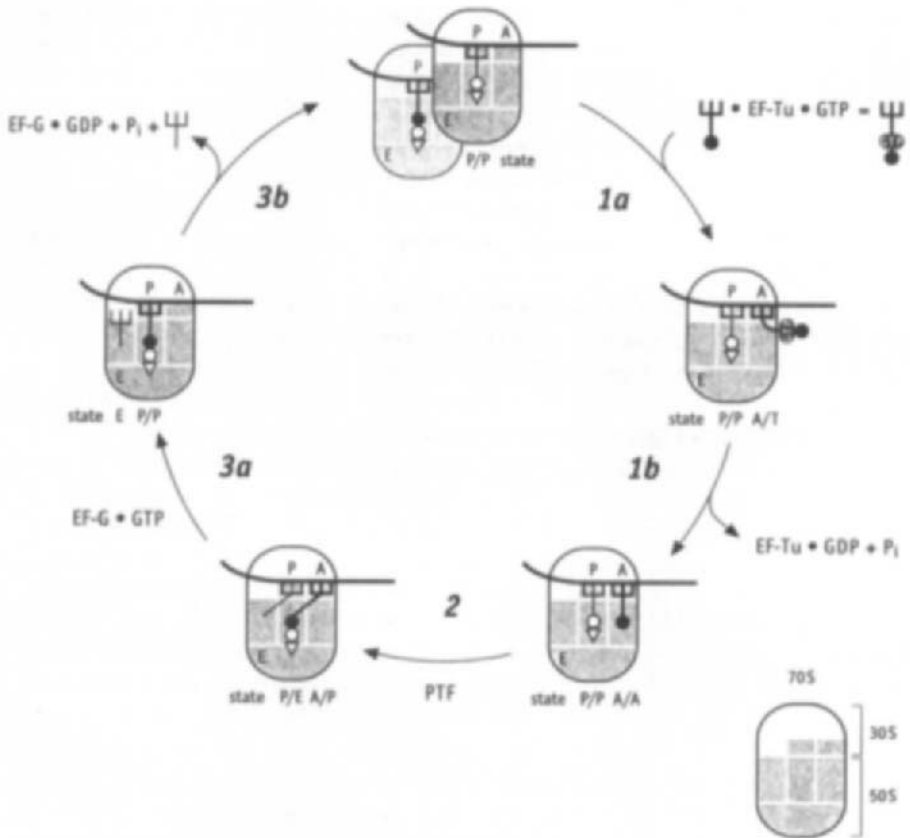


FIG. 2. The hybrid-site model (25). An aminoacyl-tRNA is at the A site. After linking the already synthesized peptide chain to the aminoacyl residue via a peptide bond, the peptidyl-tRNA elongated by one aminoacyl residue moves to the P site moiety on the large subunit but remains with the anticodon region at the A site moiety on the small subunit (A/P hybrid site). Likewise, the deacylated tRNA moves to the P/E hybrid site. After translocation, the peptidyl-tRNA is at the P site and the deacylated tRNA is at the E site, which is thought to be present exclusively on the large ribosomal subunit. PTF, Peptidyltransferase.

to the next position, over a situation in which he is forced to let go of one swing before connecting with the second swing. The pivoting movement imposed by attachment at one end restrains the movement at the other end. With a large molecule, detachment at both ends simultaneously might be fatal for protein synthesis. Similarly, the audience holds its collective breath only when the artist is disconnected and floats in midair, because of the potential fatal consequences.

However, the problem of how molecules as large as tRNAs are translocated is not solved by dividing the movement from one site to the other into two steps. Furthermore, this model abandons the operational definition of the P site, according to which an acyl-tRNA occupies the P-site region of the large subunit when the acyl residue can react with puromycin, whereas location at the A-site region is characterized by a negative puromycin reaction (26). According to this classic definition, a peptidyl-tRNA occupies the A-site region of the large subunit after peptide-bond formation, whereas the hybrid-site model specifies a location of the peptidyl residue at the P-site region of the large subunit, although this peptidyl residue cannot react with puromycin.

Two arguments can be raised against the conclusion drawn from the protection data: the hybrid-site model rests on the protection pattern of the 23S rRNA, whereby 16 of the 17 protected nucleotides depend on the presence of the CCA-end of the tRNA (27). However, the flexible CCA end of a tRNA may not be a reliable reporter structure for the site location of the relatively rigid tRNA. The second argument addresses the wide range of the Mg^{2+} concentrations (5 to 25 mM) applied in these protection studies (25, 27). It is known that the properties of the various sites are drastically and differently changed in response to such dramatic Mg^{2+} shifts, making a comparison of results obtained at various Mg^{2+} concentrations prohibitively difficult (28).

In the Section II,C we describe a new model of the elongation cycle, the α - ϵ model, which reconciles the various data sets of the previous models and, in addition, gives a simple explanation for the mechanism of translocation.

C. α - ϵ Model

It does not appear likely that a model of the ribosome at atomic resolution will be available in the immediate future, even though significant progress has been achieved in crystallizing ribosomes (29). Therefore, current models at a resolution of 20 to 35 Å are particularly valuable, because they allow tRNA to be localized within the ribosome (see Section III). The same is true for methods that yield a high-resolution picture of special aspects of the ribosome. One example is zero-length crosslinking experiments (10). Another example is the phosphorothioate method (30), which can record the contact pattern of all the phosphates of an RNA within a complex.

The contact pattern determined for a tRNA in a complex with its cognate aminoacyl-tRNA synthetase was found to correlate well with X-ray analysis data (31).

The principle of the method is as follows: the RNA under observation is transcribed *in vitro*, and phosphorothioates are randomly incorporated at about 20% of either the A, G, C, or U positions during transcription. The result is that a nonbridging oxygen is replaced by a sulfur atom at the phosphate residues of the corresponding nucleotides. After it is confirmed that the modified RNA has not lost its biological activity, the RNA is incorporated into the complex. The addition of the small I_2 molecule cleaves the sugar-phosphate backbone at the position of phosphorothioated nucleotides with a low efficiency (about 5%). If the RNA is ^{32}P labeled at, for example, the 5' end, then the contact pattern can be assessed on a sequence gel.

We performed such an experiment with tRNA bound to ribosomes. Although the *in vitro* transcribed tRNA lacked the modifications found on native tRNA^{Phe}, the transcribed tRNAs were fully active in protein synthesis regardless whether they contained thioated nucleotides. The tRNAs at the A or P sites showed a strikingly different contact pattern (32). The startling observation was, however, that neither tRNA changed its contact patterns during translocation, during which the A-site-bound AcPhe-tRNA moves to the P site and the P-site-bound tRNA^{Phe} to the E site (33, 33a). About 65 out of 76 phosphate positions could be judged, covering the tertiary structure of the tRNAs from the anticodon loop to the amino acid stem. The conclusion was that the microtopography of both tRNAs remained unchanged during translocation. A consequence of this conclusion is that both tRNAs obviously are tightly bound to a ribosomal domain, and this domain moves both tRNAs (as well as the mRNA) from the A-P sites to the P-E sites in the course of translocation. The domain region carrying the AcPhe-tRNA was called α , and that carrying the deacylated tRNA^{Phe}, ϵ . At the A site, only α appears, and at the E site, only ϵ , hence the name (see Fig. 3). Because ribosomes can be saturated with three deacylated tRNAs (although the third one binds only at high excess of tRNA over ribosomes) (13), and because the posttranslocational ribosome with occupied P and E sites starts the selection of the ternary complex aminoacyl-tRNA·EF-Tu·GTP according to the codon displayed at the A site, we postulate a decoding center δ that is stably positioned at the A site (Fig. 3).

This new α - ϵ model for the elongating ribosome has a number of important and testable consequences:

1. Reciprocal linkage between A and E sites: The allosteric three-sites model has interpreted the reciprocal connection between the first and the third site as an allosteric effect in the sense of negative coopera-

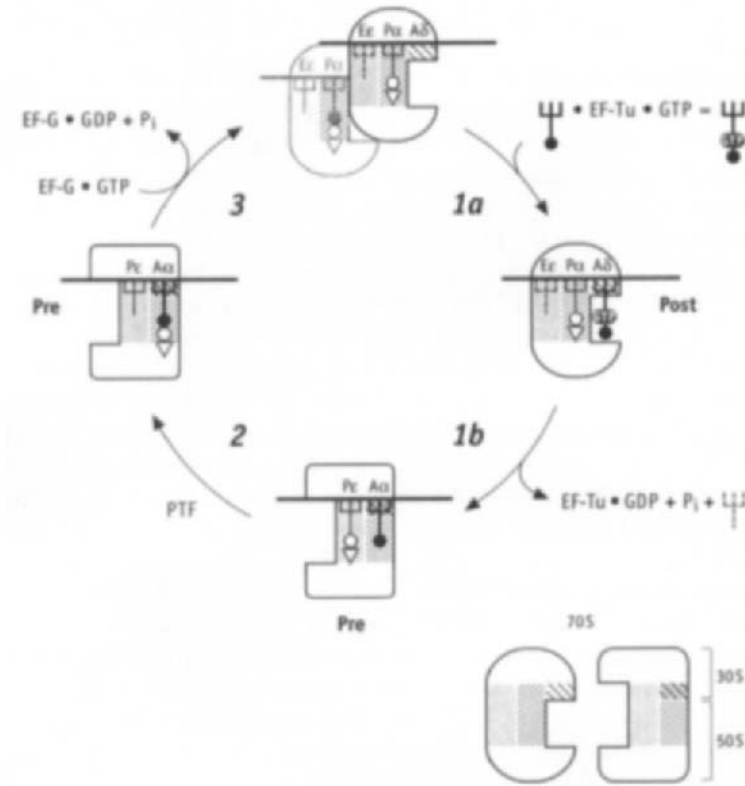


FIG. 3. The α - ϵ model (33). Both tRNAs bound to the elongating ribosome are tightly fixed on a movable ribosomal domain, or tRNA carrier, which moves both tRNAs from the A-P site region of the ribosome to the P-E site region in the course of translocation. The part of the movable domain that binds an acylated tRNA and exposes it at the A site of pretranslocational ribosomes is called α ; the one that binds deacylated tRNA and exposes it at the E site of posttranslocational ribosomes is called ϵ . The decoding center δ allows more or less exclusively codon-anticodon interaction and cannot move but is fixed at the ribosomal A site. For details, see text.

tivity (1): when the E site was occupied during translocation, the A-site affinity dropped due to this allosteric linkage. The interpretation of the α - ϵ model differs in that the α - ϵ domain moves from the A-P sites to the P-E sites. The binding domain has left the A site, leaving behind the decoding center δ , which represents the low-affinity state of the A site in the posttranslocational ribosome.

2. Number of tRNA binding sites: The binding region α overlaps with

the decoding center δ in the pretranslocational state (Fig. 3); the binding region ϵ is at the P site, whereas the E site lacks any binding region. It follows that the pretranslocational ribosome has two binding sites and the posttranslocational ribosome has three binding sites: the two high-affinity regions, α and ϵ , at P and E sites, respectively, and the low-affinity δ at the A site.

3. Translocation: The α - ϵ model provides a simple explanation for the translocation mechanism: The α - ϵ domain binds tightly two tRNAs and precisely moves them from A and P sites to the P and E sites. A recent contact pattern analysis of the mRNA supports this view (34). The accessibility of the phosphate groups of an mRNA outside the decoding codons was practically identical to that of the same mRNA in solution with two exceptions: the phosphate two positions upstream from the decoding region was protected and the phosphate immediately downstream of the decoding region was more reactive. This pattern also did not change during translocation, indicating that the microtopography of the decoding region remained the same during translocation, in agreement with the α - ϵ model. Movable domains are well known from allosterically regulated enzymes. For example, domain 1 of the prokaryotic elongation factor EF-Tu moves away from domains 2 and 3 on hydrolysis of the bound GTP to GDP (35).
4. According to the α - ϵ model, the critical step of the elongation cycle is the A-site occupation (to be precise, the second part of this reaction). The A-site occupation is divided into two parts. Reaction 1a (Fig. 3) represents the selection step of the aminoacyl-tRNA cognate to the codon displayed at the A site. This step probably is restricted to codon-anticodon interaction recognized by the decoding center δ , i.e., EF-Tu does not participate in the selection process (3, 23). An important consequence is that proofreading processes do not exist on the ribosome; in other words, EF-Tu-dependent GTP cleavage does not contribute to the accuracy of protein synthesis on the ribosome. This view is supported by the fact that EF-Tu binds the aminoacyl-tRNA at the short arm of the tRNA about 50 Å away from the anticodon (36), and thus hardly has a chance to interfere with the mechanism that selects the correct codon-anticodon interaction. Step 1b is, as mentioned already, the critical step of protein synthesis, because at this step the α - ϵ domain detaches from the tRNAs at P and E sites and switches back to the A and P sites. The result is that the peptidyl-tRNA is now fixed by the ϵ region of the movable domain and the aminoacyl-tRNA is tightly bound by the α region at the A site. The fact that the occupation of the A site is the rate-limiting step of the elongation cycle (22, 37) is consistent with this explanation.

An important prediction of the α - ϵ model is that the mutual orientation of both tRNAs should not be altered in the course of translocation, because the microtopography of the $(\text{tRNA})_2 \cdot \text{mRNA}$ complex seems to be invariant during translocation. We will compare this prediction with the first results concerning the direct localization of tRNAs within the ribosome (see Section III).

III. Structural Aspects: The Shape of Ribosomes and the Localization of tRNAs

In the past decade, electron microscopic analysis of biological specimens has made great strides in three specific areas: (1) imagery of biological samples embedded in a frozen, vitreous water/buffer layer, representing a breakthrough (38); (2) computer-aided data collection, leading to reduction of the dose per projection to extremely low levels (39, 40); and (3) development of methods of image processing that exploited projection images obtained in all three dimensions, and collected at very dose levels to reconstruct a three-dimensional model of the biological specimen (41). These improvements have greatly enhanced our understanding in particular with regard to ribosome structure, and reconstructions of the 70S ribosome from *E. coli* have been obtained by two groups (42, 43). Furthermore, tRNAs within the ribosome have become visible for the first time (11).

Both the Frank model (42), with a more compact structure, and the van Heel/Brimacombe model (43), which is more open, with a network of channels and caves, were evaluated by neutron and X-ray scattering. The Frank model was used as a starting point for modifying the model in order to optimize the fit with the measured scattering curves. Furthermore, a new method for neutron scattering, the proton-spin contrast variation method (44), was applied, improving the signal-to-noise ratio by factor of 20, thus enabling for the first time assessment of the positions and mutual arrangement of the tRNAs within the elongating ribosome. The results of the scattering analyses are discussed in Section III,A.

A. The Shape of Ribosomes

Scattering techniques can be used for the determination of structure parameters of components within biological complexes in solution or in vitreous ice, provided that the partial structure to be studied within the complex scatters sufficiently different from the surrounding matrix or buffer system. The relative difference in scattering efficiency is termed "contrast": the larger the contrast the more precise the extractable data.

Two sets of structural parameters can be determined with high precision by means of scattering techniques:

1. The distance between two scattering centers within a complex can be deduced from the interference pattern of the collected spectra, provided that the centers have a fixed distance. If each scattering center is a larger structure or molecule within the complex, the deduced distance gives the distance between the corresponding mass center of scattering, which is usually almost identical to the mass center of gravity. In this way many protein distances within both ribosomal subunits have been previously measured (45, 46); the distances were used to construct the spatial distribution of the protein positions in the ribosome applying a technique called "triangulation" (47).
2. If various models of a distinct complex are on the market and a monodisperse suspension of the complex is available (this is a solution that contains practically only monomers, rather than dimers or aggregates of the complex), then the corresponding intensity curves of the complex in solution can be compared with the calculated "scattering curves" of the models. That model with the best-fitting calculated scattering curve is the best description of the structure of the naturally occurring complex. In other words, the scattering analysis can judge and rate competing models. This analysis was done for the two new ribosome models (see Section III,A,2.)

There is a third important application that does not give an absolute value, in contrast to the two aforementioned applications, but rather a probable description of the structure of the complex. For example, if one has the scattering curve of a complex and a model that is already a good structural description, one can modify the model so as to optimize the fit of its calculated scattering curve with respect to the measured intensities. If the fit can be improved the modified model is probably a better description of the complex structure than the starting model. Examples of this procedure are presented in the following sections. In principle, such a fit-optimization routine can be used to derive the structure of a complex *ab initio*, as has been done for the 50S subunit of the *E. coli* ribosome (48), but a starting with the best available model will improve both the correctness and resolution of the derived structure.

X-Ray or neutron scattering from an isolated atom is characterized by the so-called scattering length, which is a measure of the scattering efficiency of an atom (we consider elastic scattering—that is, scattering without loss of energy). Photons are scattered by electrons, and the X-ray scattering length of an atom increases with its atomic number; i.e., the higher the charge of the

atomic nucleus, the stronger the scattering. Neutrons are scattered by the potential of the atomic nuclei, and the scattering lengths of the isotopes of the same element can be remarkably different. The most prominent example is hydrogen (H) and its heavy isotope deuterium (D): The scattering of deuterium is similar to that of the other most abundant isotopes of elements found in biological samples (mainly C, N, O, P, and S), whereas the scattering of hydrogen differs in two important aspects. The first aspect is related to the sign of the scattering length, which is positive for D and most other elements, but negative for H. This means that the phase of the neutron wave scattered by an H atom is shifted by 180° as compared to those scattered by deuterium and the other elements. The scattering of a volume element of a sample can be calculated from the sum of scattering intensities of all atoms in that volume element, and because hydrogen is the most abundant element in biological samples (about 30% in nucleic acids and 50% in proteins), it is clear that the hydrogen content of a molecule determines its scattering intensity. This fact is highly relevant for structure research. For example, if a deuterated molecule is incorporated in an otherwise protonated complex, the molecule will be "stained" for the neutron beam and can therefore be measured specifically *in situ*. Apart from it, as the scattering length densities of H_2O and D_2O are significantly different, valuable additional structural information is obtained by performing measurements in solvents with various D_2O concentrations.

Elastic interaction of neutrons with the nuclear potential produces so-called coherent scattering (spin-independent scattering). Because the atomic positions within the particle are fixed, stable interference patterns are formed, leading to an angle-dependent scattering from which structural information can be deduced.

Neutrons are, however, also scattered in a way that is dependent on the nuclear spin states (proton spin contrast variation) (49). A striking example is the scattering of polarized neutrons by polarized nuclei of the hydrogen isotope ^1H . If the polarization of the proton spin, with respect to an external magnetic field, is pointing in the same direction as the polarization of the incident neutron beam, then the scattering process is coherent. No incoherent scattering is observed in this case. The scattering length of a neutron-proton interaction is comparable to that of neutrons with other nuclei, e.g., C, N, and O, and also with that of its heavier isotope, deuterium. This case is not very interesting from the point of view of contrast. The situation is quite different when the polarization of the incident neutron spin (or that of the proteins) is reversed, i.e., when neutrons and protons with spins of antiparallel orientation interact. There is a strong incoherent scattering, even stronger than in the unpolarized state, and there is an unusually negative length of coherent scattering, which is even more negative than for unpolarized spins. This is an

ideal case for high contrast. Because the price is an increased incoherent scattering, the advantage of the large contrast should be used in a way that leads to a favorable high coherent scattering due to the large contrast combined with a minimum of incoherent scattering. In practice this means that only those structural regions of interest in a macromolecule will be left protonated, whereas the residual hydrogens, both of the particle and of the solvent, will be deuterated. The gain in contrast due to proton-spin polarization is impressive. It is a factor of about 2.5 with respect to the unpolarized sample, both for proteins and for RNA. A further increase of contrast is due to the solvent, which is a mixture of D_2O and deuterated glycerol (9/11) instead of D_2O . It is an additional factor of almost two for RNA and a somewhat smaller factor for proteins (44). Hence, the total increase in contrast due to proton-spin polarization and solvent composition is a factor of 5 for RNA and about 4 for proteins. Because the intensities are proportional to the square of the corresponding contrasts, the observable neutron scattering intensities of the protonated labels exceed those of the unpolarized sample by a factor of 25 in the case of RNA and of 16 for proteins.

Below, we describe the use of these peculiarities of neutron scattering from hydrogen atoms to deduce a low-resolution shape of the ribosome in solution and the positions of the tRNA molecules before and after translocation.

1. METHODS

Biological macromolecules in dilute solutions are randomly oriented. The solution scattering intensity $I(s)$, with $s = 4\pi \sin \theta/\lambda$ (here 2θ is the scattering angle and λ is the wavelength of the radiation), is therefore an isotropic function obtained as the spherically averaged intensity scattered by a single particle (50). The spherical average leads to a significant loss of structural information, and only a few (typically 10 to 20) independent structural parameters characterizing the particle can be extracted from an experimental solution scattering curve (51, 52). Restoration of the particle shape from solution scattering data is therefore only possible at low resolution, where the shape can be adequately described by a few parameters.

A convenient and economic description of the particle shape at low resolution is given by introducing an angular envelope function $r = F(\omega)$, where $(r, \omega) = (r, \theta, \varphi)$ are spherical coordinates. As illustrated in Fig. 4, the value $r = F(\omega)$ gives, for each angular direction ω , the distance from the origin to the particle's border. Any envelope function can be parameterized using the following representation (53):

$$F(\omega) \approx \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega), \quad (1)$$

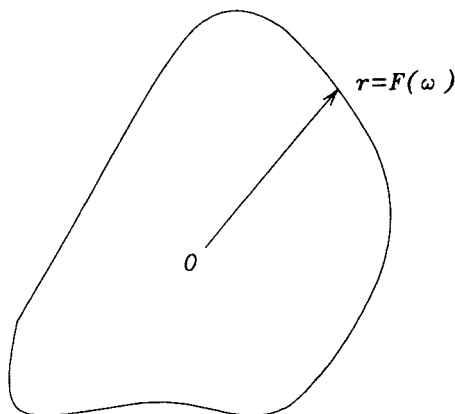


FIG. 4. Representation of a particle shape using the angular envelope function. (From Ref. 65).

where f_{lm} are complex numbers and $Y_{lm}(\omega)$ are spherical harmonics. The particle shape is thus represented by a set of numbers (values of coefficients f_{lm}) that can be determined by minimizing the discrepancy between the experimental and calculated scattering curves. Here, the truncation value L describes the resolution: the larger L , the more detailed is the description of the particle shape. Methods were developed (54–56) in order to evaluate rapidly the scattering intensity $I(s)$ for a given shape and to perform the nonlinear minimization to restore the f_{lm} coefficients from the scattering curve. Furthermore, the necessary conditions for an unambiguous low-resolution shape restoration from solution scattering data were formulated (57).

The direct shape restoration proved to be a useful tool to investigate quaternary structures of proteins in solution (58–60). Let us consider how this approach can be applied to ribosome studies. Given the restriction in the number of parameters, direct shape determination of the entire 70S ribosome would yield a resolution of about 10 nm, a value that is not able to compete with the resolution of modern electron microscopic studies (42, 43). The most important advantage of the scattering experiments is, however, their ability to distinguish among the different structural components of the ribosome. Proteins and RNA have, both for X-rays and neutrons, different scattering densities, and their scattering contributions can be separated by performing experiments in solutions with varying solvent scattering densities (contrast variation) (61). Moreover, important additional information can be obtained in neutron scattering studies using selectively deuterated samples. Thus, in a hybrid 70S ribosome reassociated from one protonated and one deuterated subunit, the scattering densities of RNA (and proteins) belonging

to different subunits will be significantly different. Therefore, contrast variation studies on hybrid ribosomes enable distinction not only between RNA and proteins, but also between the two ribosomal subunits. Considering the 70S ribosome as a four-phase system consisting of RNA and protein moieties in both the 30S and 50S subunits, it can be shown that contrast variation of hybrid ribosomes yields 10 times more information than is contained in a single scattering curve.

RNA molecules in both subunits are known to occupy preferentially the central parts, whereas the proteins are predominantly located at the periphery (62–64). A plausible low-resolution model of the ribosome can therefore be built using four envelope functions, as illustrated in Fig. 5. Here, the envelopes F_1 and F_3 describe the shapes of the entire 30S and 50S subunits, whereas the functions F_2 and F_4 describe the shapes of their RNA moieties, respectively. By expressing these envelope functions in Eq. (1), the four-phase model is represented via four sets of multipole coefficients. In the following data fitting procedure (65) the multipole resolution $L = 7$ is used, requiring $64 \times 4 = 256$ parameters and corresponding to a spatial resolution of about 3.5 nm.

Hybrid ribosome samples were prepared in the Max-Planck-Institut für Molekulare Genetik, Berlin, and the neutron scattering experiments with the 70S particles and with free subunits were performed at the Risø National Laboratory, Roskilde, in Denmark, and the GKSS Research Center, Geesthacht, in Germany. Seven different types of hybrid 70S particles included: (1) fully protonated (H30 + H50) and fully deuterated (D30 + D50) particles; (2) one subunit fully deuterated, the other fully protonated (D30 + H50 and H30 + D50); (3) the total protein (TP) moiety of one subunit deuter-

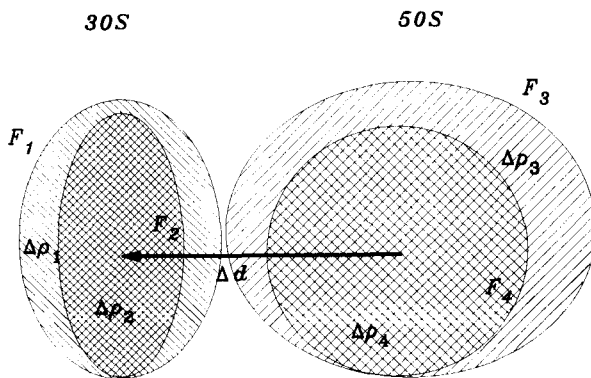


FIG. 5. Schematic representation of the four-phase model of the 70S ribosome in terms of four envelope functions. The center of the 50S subunit is placed at the origin, and that of the 30S subunit is shifted along the z axis by Δd . (From Ref. 65).

TABLE I
MEASURED SAMPLES OF HYBRID 70S AND FREE SUBUNITS

| | | |
|--|--------------------------------------|-----------|
| Contrast variation by solvent exchange (neutrons) | | |
| H30 + H50 | 0, 35, 50, 75, 100% D ₂ O | 5 curves |
| H30 + D50 | 0, 35, 50, 75% D ₂ O | 4 curves |
| D30 + H50 | 0, 35, 50, 75, 100% D ₂ O | 5 curves |
| D30 + D50 | 0, 35, 50, 75, 100% D ₂ O | 5 curves |
| D-TP30 + H50 | 0, 40, 60, 100% D ₂ O | 4 curves |
| H30 + D-TP50 | 0, 40, 60, 100% D ₂ O | 4 curves |
| H30 | 0, 100% D ₂ O | 2 curves |
| D30 | 0% D ₂ O | 1 curve |
| H50 | 0, 100% D ₂ O | 2 curves |
| D50 | 0% D ₂ O | 1 curve |
| Spin-dependent contrast variation (neutrons) | | |
| H30 + D50 | Polarization = 0 and 1 | 2 curves |
| D30 + H50 | Polarization = 0 and 1 | 2 curves |
| D-TP + H-RNA | Polarization = 0 and 1 | 2 curves |
| Synchrotron radiation X-ray scattering | | |
| All hybrid 70S samples in all solvents | | 1 curve |
| All 50S samples in all solvents | | 1 curve |
| All 30S samples in all solvents | | 1 curve |
| Total | | 42 curves |

ated, the rest of the ribosome protonated (D-TP30 + H50 and H30 + D-TP50); and (4) total proteins of the entire 70S deuterated and the total ribosomal RNA protonated (D-TP + H-RNA). All hybrid particles were also measured with X-rays at the EMBL beamline (HASYLAB, DESY, Hamburg). The X-ray measurements were of special importance because they allowed comparison of hybrid samples: hydrogen and deuterium both have only one electron and deuteration therefore does not change the X-ray scattering pattern of the particle. The constancy of the measured X-ray curves along with the preservation of biological activity indicated that the structure of the hybrid 70S particles was not noticeably affected by the various preparation procedures. All measured samples (total of 42) are listed in Table I.

Small-angle scattering intensities from a four-phase system are expressed in terms of the scattering amplitudes of the components:

$$I_{\text{theor}}(S) = \left\langle \left[\sum_{n=1}^4 \Delta\rho_n \exp(i\mathbf{s}\mathbf{r}_n) \cdot A_n(\mathbf{s}) \right]^2 \right\rangle_{\Omega} \quad (2)$$

where $\langle \rangle$ represent the spherical average over the solid angle Ω in the reciprocal space, the scattering vector is $\mathbf{s} = (s, \Omega)$, $A_n(\mathbf{s})$ is the normalized scat-

tering amplitude from the n th phase, and $\Delta\rho_n$ and \mathbf{r}_n are the contrast and position of the center of mass of the n th phase, respectively. The contrasts of the phases for each scattering curve are known from the chemical compositions and the extent of deuteration of both the ribosome components and the solvent. The scattering intensity [Eq. (2)] for each of the samples listed in Table I is therefore a function of the shapes of the components and positions of their centers. The distance between the centers of the subunits could be estimated in advance from the invariance of the scattering curves and was found to be $\Delta d = 10.5$ nm, whereas the centers of the RNA and protein moieties in both subunits coincided approximately. To speed up the calculations, the two phases describing the 50S subunit were centered at the origin, whereas the 30S subunit was moved by Δd along the z axis. This configuration allowed the intensity [(Eq. (2))] to be evaluated rapidly for each combination of contrasts using the methods developed by Svergun (55, 56), thus enabling an optimization procedure to find the four sets of shape coefficients fitting the entire set of the experimental data.

Before fitting the experimental data, model calculations were performed on simulated contrast variation data sets. It was found that unequivocal restoration of the structure of the four-phase particle requires a good initial approximation for at least two shape functions. The initial approximation for the envelopes F_1 and F_3 of the entire 30S and 50S subunits can be taken from the electron microscopic data. For this purpose, two recent electron microscopic models (42, 43) were available. To achieve a model-independent assessment of consistency of these models with our experimental data, scattering curves from the two models were evaluated and compared with the measured neutron scattering from a fully deuterated 70S ribosome (particle D30 + D50) in fully protonated solvent. Because the scattering length densities of deuterated proteins and deuterated RNA are very close to each other and both are very different from water density, the latter curve, representing a "shape scattering" from the 70S ribosome, is ideal for such a direct comparison. As seen in Fig. 6, both models fail to fit the experimental curve, with the systematic deviations for the model of van Heel/Brimacombe being more significant at the very beginning of the scattering curve (that is, at very low resolution). This is caused in particular by the "Swiss cheese"-like structure of the model, which contains a variety of channels; the fit is somewhat improved if the channels are filled by matter. The more compact Frank model provides better agreement with the experimental data and has therefore been used for the initial approximation. After visual inspection, the model was divided into subunits and their envelope functions were expressed by spherical harmonics according to Eq. (1) to represent the envelopes F_1 and F_3 of the initial approximation. For the unknown RNA envelopes F_2 and F_4 , spheres were taken with volumes equal to approximately 60% of the volumes of the

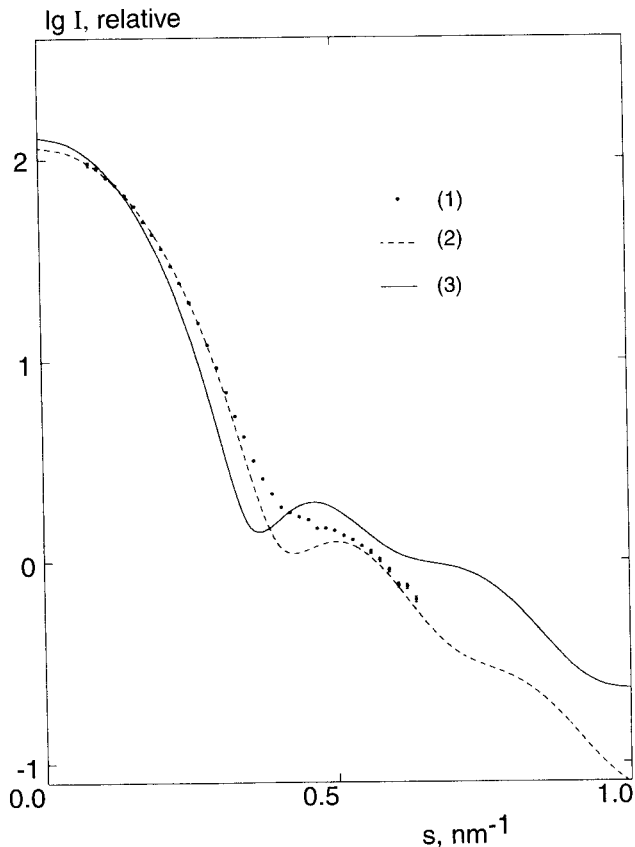


FIG. 6. Comparison of the experimental neutron scattering with that calculated from the electron microscopic models. (1) Neutron scattering from fully deuterated 70S (D30 + D50) in H_2O ; (2) and (3), evaluated scattering curves from the Frank and the van Heel/Brimacombe models, respectively. (From Ref. 65.)

whole subunits [the volume fraction of hydrated RNA in ribosomes is known from the chemical composition and solution scattering studies (62, 63, 66)].

The four sets of the multipole coefficients thus obtained were used as a starting point in the nonlinear refinement procedure to minimize the overall discrepancy:

$$\chi^2 = \frac{1}{M} \sum_{k=1}^M \frac{\sum_j \{ [I_{\text{exp}}^{(k)}(s_j) - I_{\text{theor}}^{(k)}(s_j)] / \sigma^{(k)}(s_j) \}^2}{\sum_j [I_{\text{exp}}^{(k)}(s_j) / \sigma^{(k)}(s_j)]^2}, \quad (3)$$

where M is the total number of curves and $I_{\text{exp}}^{(k)}(s_j)$ and $\sigma^{(k)}(s_j)$ are the measured scattering intensity and its standard deviation, respectively, in the j th experimental point of the k th curve. $I_{\text{theor}}^{(k)}(s_j)$ is the corresponding scattering from the four-phase model evaluated using Eq. (2) for the contrasts of the components corresponding to the k th curve. In order to take into account the instrumental effects, theoretical neutron scattering curves were appropriately smeared as previously described (66). Minimization was performed using the nonlinear least-squares package NL2SOL (67).

2. RESULTS

The optimized four-phase model of the *E. coli* 70S ribosome presented in Fig. 7 (see color plate) has an overall discrepancy of 0.067 [Eq. (3)], and a typical fit to the neutron scattering data is illustrated in Fig. 8. The Frank model with the RNA moieties of the two subunits selected to occupy 60% of the entire ribosome yielded the overall discrepancy 0.13. Optimization of the structure therefore led to a significantly better fit of the neutron scattering data than that provided by electron microscopy, i.e., the fit of the optimized model is better by factor of 2 in terms of the overall discrepancy.

How different is the optimized model of the 70S ribosome as compared to the initial electron microscopic one in terms of the overall shape and the internal structure? Comparisons of the initial envelopes of the 30S and 50S subunits in the Frank model and those of the optimized model are presented in Fig. 9 (see color plate). Qualitatively, the heads of both the 30S and the 50S subunits become more elongated and the two subunits appear to be somewhat more separated; otherwise, the two models look remarkably similar. A quantitative estimate of the agreement between two shape functions, say, F_1 and F_2 , is given by the average least-square deviation

$$R^2_{\omega} = \frac{\int [F_1(\omega) - F_2(\omega)]^2 d\omega}{\int F_1(\omega) F_2(\omega) d\omega}. \quad (4)$$

These values are 0.12 and 0.10 for the 30S and 50S subunits, respectively; that is, the differences in the overall shape of the subunits are not dramatic. The improvement of the fit is therefore largely due to the internal structure of the subunits: The RNA distributions obtained differ significantly from the tentative RNA shapes inferred from the electron microscopic data. The latter are practically concentric to the envelopes of the whole subunits (42), whereas in the four-phase model (Fig. 7) the RNA moieties do not show this distribution. In the small subunit the RNA is more irregular than the subunit itself, whereas the RNA in the large subunit forms a very compact structure with protuberances, as previously observed (66, 68). The RNA

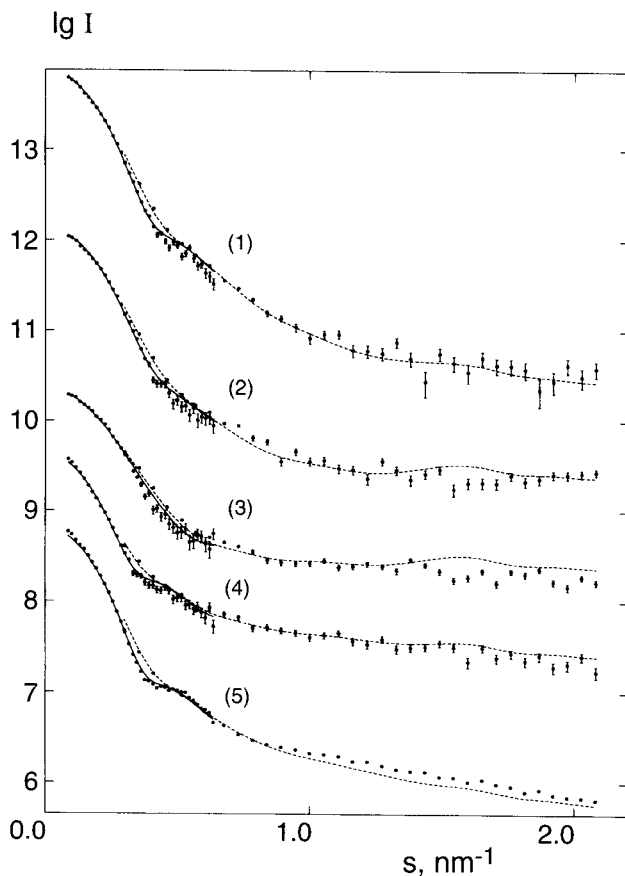


FIG. 8. Example of the experimental data fitting by the four-phase model in Fig. 7. The experimental and calculated scattering curves from fully protonated 70S ribosome (H30 + H50) in solvents with different D_2O concentrations are displayed (curves 1–5 correspond to 0, 35, 50, 75, and 100% D_2O). Two settings with different sample-detector distances overlapping in the range $0.3 < s < 0.7 \text{ nm}^{-1}$ are fitted for each D_2O concentration. For better visualization, successive curves are shifted one logarithmic unit down the abscissa axis. (From Ref. 65.)

molecules in both subunits occupy parts of the surface (those areas are depicted by darker colors in Fig. 7).

Considering the four-phase model obtained, one should be aware of the limitations of the model. The restricted resolution in the representation of the envelope functions does not allow fine details of the surface to be reproduced (e.g., in the 50S subunit the L1 protuberance and the L7/L12 stalk are not resolved, and the shape of the head is smeared).

The current model represents the intersubunit space as well as the distribution of rRNA and ribosomal proteins within both subunits, but beyond these features the model, which uses a solid-body representation of each phase, does not contain holes or channels. The latter point has an important consequence: keeping in mind that the model with a network of holes and caves (43) yielded a poor fit in a model-independent comparison with the experimental data (Fig. 6), it is not very likely that the ribosome contains such a distinct system of holes and channels.

The four-phase model presented here accounts—as already mentioned—for the internal structure of the ribosome in terms of the distribution of rRNA and ribosomal proteins. It fits a large body of experimental neutron and X-ray scattering data and is in good agreement with the results of electron microscopy. The model allows further refinement and positioning of smaller structural ribosomal elements (single proteins, rRNA fragments, tRNA) using selective labeling with the method of spin-dependent neutron scattering described in the next section.

B. The Arrangements of tRNAs before and after Translocation

Three-dimensional reconstructions of ribosomes derived from electron microscopic investigations have achieved a resolution of 20 to 30 Å (42, 43). Because the L-shaped tRNAs have an overall size of about 75 Å (69), it is feasible to subtract the structure of the empty ribosome from the structure of a programmed ribosome carrying tRNAs, yielding a “difference mass.” The difference mass thus represents mRNA and tRNAs but also shifted regions of the ribosomal matrix caused by conformational changes of the ribosome. In particular, the latter aspect, the possibility of conformational changes, complicates a straightforward interpretation. Nevertheless, tRNAs have been fitted successfully into a difference mass (11). The precise localization of tRNAs within the two main states of the elongating ribosome, i.e., the pre- and the posttranslocational states, is subject of ongoing research by the groups of J. Frank (Albany)/K. H. Nierhaus (Berlin) and M. van Heel/R. Brimacombe (Berlin). We will concentrate here on those observations that are supported by different methods and thus seem to be well documented.

In addition to electron microscopic studies, neutron scattering analysis also allows the localization of tRNAs. Conventional techniques of neutron scattering only allow the determination of protein positions within the ribosome (45, 46), whereas the determination of positions of small RNAs has not been possible. This insensitivity of conventional techniques is due to the lower proton content in RNA (33%) compared to that in proteins (51%), which results in an unfavorable contrast between the protonated small RNAs and the deuterated ribosomal matrix. It presents a particularly severe drawback

because the most important ligands of the ribosome during protein synthesis are small RNAs, such as tRNAs and an mRNA sequence of about 40 nucleotides covered by the ribosome (70).

The new neutron scattering technique of the proton spin contrast variation developed by the Stuhrmann group (49) provided a way out of this dilemma. The principles of this technique were described in Section III,A,1. The ability to align the spins of the neutrons by using a polarized neutron beam and to align the spins of the reflecting protons inside the sample leads to an improvement of the signal-to-noise ratio by a factor of more than 20, which for the first time allowed the determination of the tRNA positions (12).

In a first application of this new technique to the problem of tRNA localization, a pretranslocational complex was prepared with the help of a 46-nucleotide heteropolymeric mRNA that contained two unique codons, AUG-UUC, in the middle. Accordingly, tRNA^{fMet} (codon AUG) was bound to the P site and AcPhe-tRNA (codon UUC) was bound to the A site. A sister aliquot of this preparation was translocated to a posttranslocational complex with the help of translocation factor EF-G and GTP. Both tRNAs as well as the mRNA were protonated and the ribosome was fully deuterated (better than 98%). The technique of proton spin contrast variation yielded a clear signal from which the location of the mass center of gravity of the protonated mass (mRNA and two tRNAs) inside the ribosome could be extracted unequivocally (Fig. 10; see color plate). The centers of gravity were found in the interface cavity separating the ribosomal subunits near the neck of the 30S subunit. The center moves during translocation by $12 \pm 3 \text{ \AA}$ toward the head of the 30S subunit and the L1 protuberance of the 50S subunit.

The fact that the mass center of the tRNA·mRNA complex moves by a codon length indicates that the tRNAs do not simply make a turn to pull the mRNA for a codon length through the ribosome, but rather that the whole tRNA·mRNA complex moves. This means in the framework of the α - ϵ model (Section II,C) that the ribosomal α - ϵ domain, with both tightly bound tRNAs, performs a corresponding movement.

The α - ϵ model predicts that the mutual arrangement of the two tRNAs does not change in the course of translocation, i.e., it is the same before and after translocation. The signal of the neutron scattering analysis also allowed this prediction to be tested. The results provided the following data:

1. The E site of the posttranslocational complexes, which had been purified via centrifugation through a sucrose cushion, was fully occupied. This demonstrates that deacylated tRNA is quantitatively and tightly bound to the E site, a point that has been subject to controversy in the past (19).

2. The mutual arrangement of the two tRNAs was practically identical before and after translocation.
3. The angle between the tRNAs was about 110° before and after translocation (70a). The latest Frank model (42) was not used at an early processing stage, thus leading to a different stable angle of 50 to 55° between the tRNAs in the course of translocation (33). This indicates the sensitive dependence of the results on the model that is used as input for the processing procedure.

A recent attempt to obtain an arrangement of tRNAs and mRNA that (1) conforms with codon–anticodon interactions and (2) fits the difference mass reported by Agrawal *et al.* (11) also revealed an angle of 110° between the tRNAs (J. Frank and S. Harvey, personal communication).

A recent analysis of equivalent pre- and posttranslocational states by means of electron microscopy (70b) already supported the main conclusions drawn from the neutron scattering analysis. The difference masses before and after translocation were almost identical; that of the posttranslocational state was shifted toward the L1 protuberance by about 10 \AA . Although the fitting of tRNAs into the difference mass of the ribosomal elongating states has not yet been completed, the following conclusions can be already drawn; (1) A tRNA is not lost during translocation, again indicating a quantitative occupation of the E site and supporting the results derived from the neutron scattering analysis (see above); (2) the mutual arrangement of both tRNAs after translocation is very similar to that before translocation; and (3) the mRNA·tRNA₂ complex moves as a block in the course of translocation by about a codon length.

IV. Conclusions

Recent biochemical and structural analyses of functional complexes of the ribosome yield a satisfyingly consistent picture of the elongating ribosome. The new α – ϵ model for the elongation cycle (Section II,C) offers a simple explanation for the translocation reaction. Its critical step is the tight binding of the aminoacyl-tRNA to the A site, when the α – ϵ domain flips back from the E–P region to the P–A region (reaction 1b, Fig. 3). This step takes place after the decoding has been completed and, thus, does not jeopardize the information flow from the mRNA to the protein structure, in contrast to all previous models of the elongation phase, whereby the critical step was the translocation reaction. According to the previous models, translocation was understood as the movement of tRNAs on the surface of the ribosome from

one ribosomal region to another, i.e., a precise movement of both tRNAs essential for keeping the reading frame was difficult to imagine.

An essential prediction of the α - ϵ model, namely, that the mutual arrangement of both tRNAs before and after translocation does not change, is independently confirmed by two different structural methods—neutron scattering and electron microscopy. The basis of the α - ϵ model was that a movable ribosomal domain is carrying both tRNAs, together with the mRNA, side by side during translocation. Accepting this explanation, it is now easy to imagine why codon-anticodon interaction is observed at the E site after translocation (16, 19, 71) when it occurs at the P site before translocation, as is generally accepted. There is no discrepancy among the various attempts to place the two tRNAs inside the elongating ribosome in terms of the general position. Rather, the current discussion centers around the precise orientation of both tRNAs. We expect that this problem will be solved in the near future.

ACKNOWLEDGMENTS

I thank J. Beckmann, N. Burkhardt, and C. M. T. Spahn (Berlin) for helpful discussions and V. I. Lim (Moscow) for making available to me unpublished material. I am particularly grateful to J. Frank (Albany) for many valuable comments.

REFERENCES

1. K. H. Nierhaus, *Mol. Microbiol.* **9**, 661 (1993).
2. C. G. Kurland, *Trends Biochem. Sci.* **12**, 126 (1987).
3. A. P. Potapov, F. J. Triana-Alonso, and K. H. Nierhaus, *J. Biol. Chem.* **270**, 17680 (1995).
4. K. H. Nierhaus, *Nature (London)* **379**, 491 (1996).
5. F. J. Franceschi, and K. H. Nierhaus, *J. Biol. Chem.* **265**, 16676 (1990).
6. H. F. Noller, *J. Bacteriol.* **175**, 5297 (1993).
7. R. R. Samaha, R. Green, and H. F. Noller, *Nature (London)* **377**, 309 (1995).
8. C. M. T. Spahn, M. A. Schäfer, A. A. Kravetsky, and K. H. Nierhaus, *J. Biol. Chem.* **271**, 32857 (1996).
9. H. F. Noller, *Annu. Rev. Biol.* **60**, 191 (1991).
10. R. Brimacombe, *Eur. J. Biochem.* **230**, 365 (1995).
11. R. K. Agrawal, P. Penczek, T. A. Grassucci, Y. Li, A. Leith, K. H. Nierhaus, and J. Frank, *Science* **271**, 1000 (1996).
12. J. Wadzack, N. Burkhardt, R. Jünemann, G. Diedrich, K. H. Nierhaus, J. Frank, P. Penczek, W. Meerwink, M. Schmitt, R. Willumeit, and H. B. Stuhmann, *J. Mol. Biol.* **266**, 343 (1997).
13. H.-J. Rheinberger, H. Sternbach, and K. H. Nierhaus, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5310 (1981).
14. R. A. Grayevskaya, Y. V. Ivanov, and E. M. Saminsky, *Eur. J. Biochem.* **128**, 47 (1982).
15. H. Saruyama and K. H. Nierhaus, *Mol. Gen. Genet.* **204**, 221 (1986).
16. F. J. Triana-Alonso, K. Chakraborty, and K. H. Nierhaus, *J. Biol. Chem.* **270**, 20473 (1995).

17. M. V. Rodnina and W. Wintermeyer, *J. Mol. Biol.* **228**, 450 (1992).
18. H.-J. Rheinberger and K. H. Nierhaus, *J. Biol. Chem.* **261**, 9133 (1986).
- 18a. A. V. El'skaya, G. V. Ovcharenko, S. S. Palchevskii, Z. M. Petrushenko, F. J. Triana-Alonso, and K. H. Nierhaus, *Biochemistry* (1997). In press.
19. A. Gnirke, U. Geigenmüller, H.-J. Rheinberger, and K. H. Nierhaus, *J. Biol. Chem.* **264**, 7291 (1989).
20. J. Remme, T. Margus, R. Vilems, and K. H. Nierhaus, *Eur. J. Biochem.* **183**, 281 (1989).
21. Y. Komine, T. Adachi, H. Inokuchi, and H. Ozeki, *J. Mol. Biol.* **212**, 579 (1990).
22. S. Schilling-Bartetzko, F. Franceschi, H. Stembach, and K. H. Nierhaus, *J. Biol. Chem.* **267**, 4693 (1992).
23. U. Geigenmüller and K. H. Nierhaus, *EMBO J.* **9**, 4527 (1990).
24. F. Jørgensen and C. G. Kurland, *J. Mol. Biol.* **215**, 511 (1990).
- 24a. V. I. Lim, *J. Mol. Biol.* **266**, 877 (1997).
25. D. Moazed and H. F. Noller, *Nature (London)* **342**, 142 (1989).
26. R. E. Monro, J. Cerná, and K. A. Marcker, *Proc. Natl. Acad. Sci. U.S.A.* **61**, 1042 (1968).
27. D. Moazed and H. F. Noller, *Cell* **57**, 585 (1989).
28. H.-J. Rheinberger and K. H. Nierhaus, *J. Biomol. Struct. Dynam.* **5**, 435 (1987).
29. F. Schlünzen, H. A. S. Hansen, J. Thygesen, W. S. Bennett, N. Volkmann, I. Levin, J. Harms, H. Bartels, A. Zaytzev-Bashan, Z. Berkovitch-Yellin, I. Sagi, F. Franceschi, S. Krumbholz, M. Geva, S. Weinstein, I. Agmon, N. Bøddeker, S. Morlang, R. Sharon, A. Dribin, E. Maltz, M. Peretz, V. Weinrich, and A. Yonath, *Biochem. Cell Biol.* **73**, 739 (1995).
30. D. Schatz, R. Leberman and F. Eckstein, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6132 (1991).
31. J. Rudinger, J. D. Puglisi, J. Pütz, D. Schatz, F. Eckstein, C. Florentz, and R. Giege, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5882 (1992).
32. M. Dabrowski, C. M. T. Spahn, and K. H. Nierhaus, *EMBO J.* **14**, 4872 (1995).
33. K. H. Nierhaus, D. Beyer, M. Dabrowski, M. A. Schäfer, C. M. R. Spahn, J. Wadzack, K.-U. Bittner, N. Burkhardt, G. Diedrich, R. Jünemann, D. Kamp, H. Voss, and H. B. Stuhmann, *Biochem. Cell Biol.* **73**, 1011 (1995).
- 33a. M. Dabrowski, C. M. T. Spahn, M. A. Schafer, and K. H. Nierhaus, manuscript in preparation (1997).
34. E. V. Alexeeva, O. V. Shpanchenko, O. A. Dontsova, A. A. Bogdanov, and K. H. Nierhaus, *Nucleic Acids Res.* **24**, 2228 (1996).
35. H. Burchtold, L. Reshetnikova, C. O. A. Reiser, N. K. Schirmer, M. Sprinzl, and R. Hilgenfeld, *Nature (London)* **365**, 126 (1993).
36. P. Nissen, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F. C. Clark, and J. Nyborg, *Science* **270**, 1464 (1995).
37. N. Bilgin, L. A. Kirsebom, M. Ehrenberg, and C. E. Kurland, *Biochemie* **70**, 611 (1988).
38. J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowell, and P. Schultz, *Rev. Biophys.* **21**, 129 (1988).
39. A. J. Koster, W. J. de Ruijter, A. van Den Bos, and K. D. Mast, *Ultramicroscopy* **27**, 251 (1989).
40. K. Dierksen, J. Typke, R. Hegerl, and W. Baumeister, *Ultramicroscopy* **49**, 109, (1993).
41. J. Frank, "Three-Dimensional Electron Microscopy of Macromolecular Assemblies." Academic Press, New York, 1996.
42. F. Frank, J. Zhu, P. Penzcek, Y. Li, S. Srivastava, A. Verschoor, M. Rademacher, R. Grassucci, R. K. Lata, and R. K. Agrawal, *Nature (London)* **376**, 441 (1995).
43. H. Stark, F. Mueller, E. V. Orlova, M. Schatz, P. Dube, T. Erdemir, F. Zemlin, R. Brimacombe, and M. van Heel, *Structure* **3**, 815 (1995).
44. H. B. Stuhmann, N. Burkhardt, G. Diedrich, R. Jünemann, W. Meerwinck, M. Schmitt, J. Wadzack, R. Willumeit, J. Zhao, and K. H. Nierhaus, *Nucl. Instr. Meth. Phys. Res.* **A356**, 124 (1995).

45. M. S. Capel, D. M. Engelman, B. R. Freeborn, M. Kjeldgaard, J. A. Langer, V. Ramakrishnan, D. G. Schindler, D. K. Schneider, B. P. Schoenborn, I.-Y. Yabuki, and P. B. Moore, *Science* **238**, 1403 (1987).
46. R. P. May, V. Nowotny, P. Nowotny, H. Voss, and K. H. Nierhaus, *EMBO J.* **11**, 373 (1992).
47. R. P. May, in "Neutron, X-ray and Light Scattering: Introduction to an Investigative Tool for Colloidal and Polymeric Systems" (P. Lindner and T. Zemb, eds.), p. 119. Elsevier, Amsterdam, 1991.
48. H. B. Stuhmann, J. Haas, K. Ibel, B. De Wolf, M. H. J. Koch, R. Parfait, and R. R. Crichton, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2379 (1976).
49. W. Knop, M. Hirai, H.-J. Schink, H. B. Stuhmann, R. Wagner, J. Zhao, O. Schärpf, R. R. Crichton, M. Krumpolc, K. H. Nierhaus, A. Rijllart, and T. O. Niikoski, *J. Appl. Crystallogr.* **25**, 155 (1992).
50. L. A. Feigin and D. I. Svergun, "Structure Analysis by Small-Angle X-ray and Neutron Scattering." Plenum, New York, 1987.
51. C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication." Univ. of Illinois Press, Urbana, 1949.
52. P. B. Moore, *J. Appl. Crystallogr.* **13**, 168 (1980).
53. H. B. Stuhmann, *Z. Phys. Chem.* **72**, 177 (1970).
54. D. I. Svergun and H. B. Stuhmann, *Acta Crystallogr.* **A47**, 736 (1991).
55. D. I. Svergun, *J. Appl. Crystallogr.* **24**, 485 (1991).
56. D. I. Svergun, *Acta Crystallogr.* **A50**, 391 (1994).
57. D. I. Svergun, V. V. Volkov, M. B. Kozin, and H. B. Stuhmann, *Acta Crystallogr.* **A52**, 419 (1996).
58. S. Koenig, D. I. Svergun, M. H. J. Koch, G. Huebner, and A. Schellenberger, *Eur. Biophys. J.* **22**, 185 (1993).
59. B. Schmidt, S. Koenig, D. I. Svergun, V. Volkov, G. Fischer, and M. H. J. Koch, *FEBS Lett.* **372**, 169 (1995).
60. D. I. Svergun, *J. Appl. Crystallogr.* (submitted).
61. H. B. Stuhmann and R. G. Kirste, *Z. Phys. Chem.* **46**, 247 (1965).
62. M. H. J. Koch and H. B. Stuhmann, *Meth. Enzymol.* **59**, 670 (1979).
63. I. N. Serdyuk, *Meth. Enzymol.* **59**, (1979).
64. W. Kuehlbrandt and P. N. T. Unwin, *J. Mol. Biol.* **156**, 431 (1982).
65. D. I. Svergun, N. Burkhardt, M. H. J. Koch, J. Skov Pedersen, V. V. Volkov, M. B. Kozin, H. B. Stuhmann, W. Meerwink, G. Diedrich, and K. H. Nierhaus, manuscript in preparation.
66. D. I. Svergun, J. Skov Pedersen, I. N. Serduyk, and M. H. J. Koch, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11826 (1994).
67. J. Dennis, D. Gay, and R. Welsch, *ACM Trans. Math. Soft.* **7**, 348 (1981).
68. A. P. Korn, P. Spitnik-Elson, D. Elson, and F. P. Ottensmeyer, *Eur. J. Cell Biol.* **31**, 334 (1983).
69. S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, J. L. Sussmann, A. H. J. Wang, N. C. Seeman, and A. Rich, *Science* **185**, 435 (1974).
70. D. Beyer, E. Skripkin, J. Wadzack, and K. H. Nierhaus, *J. Biol. Chem.* **269**, 30713, 1994.
- 70a. K. H. Nierhaus, J. Wadzack, N. Burkhardt, R. Jüneman, J. Frank, P. Penczek, R. Willumeit, and H. B. Stuhmann, manuscript in preparation (1997).
- 70b. R. K. Agrawal, S. Srivastava, R. A. Grassucci, J. Frank, N. Burkhardt, R. Jünemann, and K. H. Nierhaus, in progress (1997).
71. H.-J. Rheinberger, H. Sternbach and K. H. Nierhaus, *J. Biol. Chem.* **261**, 9140 (1986).

Life on the Salvage Path: The
Deoxynucleoside Kinases of
Lactobacillus acidophilus
R-26

DAVID H. IVES AND
SEIICHIRO IKEDA

*Department of Biochemistry
The Ohio State University
Columbus, Ohio 43210*

| | |
|--|-----|
| I. Historical Background—Nucleotide Metabolism in <i>Lactobacilli</i> | 207 |
| A. Unusual Deficiency Reveals New Functions | 207 |
| B. Deoxynucleotide Metabolism in <i>Lactobacillus acidophilus</i> R-26; Contrasts with Other Bacteria | 209 |
| II. Purification of Deoxynucleoside Kinases from <i>Lactobacillus</i> <i>acidophilus</i> R-26 | 212 |
| A. Assay of Kinase Activities | 212 |
| B. Resolution of Activities—Affinity Chromatography Based on the Phosphate-Donor Site | 214 |
| C. Affinity Chromatography Based on Deoxynucleoside Derivatives .. | 217 |
| D. Affinity Chromatography Based on the Combined Deoxynucleoside and Phosphate-Donor Sites | 219 |
| III. Steady-State Kinetics | 224 |
| A. pH Dependency | 224 |
| B. Different Kinetic Mechanisms for dAK and dCK or dGK | 224 |
| C. Heterotropic Activation of dAK | 226 |
| D. Isosteric Patterns of Inhibition by dNTP End Products | 228 |
| IV. Assignment of Subunit Functions | 230 |
| V. Cloning the Genes for dAK/dCK or dAK/dGK | 232 |
| A. Use of Hybridization Probes | 232 |
| B. Colony Screening by Polymerase Chain Reaction | 234 |
| VI. dCK and dGK Are Products of the Same Gene | 238 |
| A. Evidence for a <i>dck</i> Gene Is Entirely Negative | 238 |
| B. dCK Engineered from dGK Is Identical to Native dCK | 241 |
| VII. Probing the Active Site and Subunit Contacts | 243 |
| A. Limited Proteolysis | 243 |
| B. Affinity Labeling | 244 |
| C. Site-Directed Mutagenesis | 248 |
| VIII. Summary | 250 |
| References | 252 |

In *Lactobacillus acidophilus* R-26, the synthesis of DNA precursor deoxynucleotides occurs exclusively by salvage of deoxynucleosides, beginning with phosphorylation by four deoxynucleoside kinases. Subunits bearing three of these activities are uniquely organized into two heterodimers, deoxyadenosine/deoxycytidine kinase (dAK/dCK) and deoxyadenosine/deoxyguanosine kinase (dAK/dGK), which, along with a distinct deoxythymidine kinase (TK), catalyze the parallel first committed steps of dNTP biosynthesis. Whereas TK is common to most prokaryotes (and eukaryotes), the other three activities that are the emphasis of this review are quite unusual in bacteria. Each activity is regulated in *cis* by its homologous end-product (dNTP) which is understood to act as a multisubstrate inhibitor capable of binding to both nucleoside and phosphate subsites. Conversely, the inactive dAK subunit is progressively activated by 1) association with a dGK or dCK subunit and 2) the conformationally driven heterotropic affect of dGuo or dCyd bound to the opposing subunit. Limited proteolysis has proven to be a powerful probe of conformational states. Further indication of conformational or structural differences between dAK and dGK (or dCK) is that the former follows an ordered kinetic path, while dGK or dCK exhibits rapid-equilibrium random kinetics. The multi-substrate behavior of end-product binding provides a convenient new diagnostic tool for distinguishing kinetic mechanisms.

Tandem *dak-dgk* genes have been cloned from *Lactobacillus* DNA and expressed in *Escherichia coli* as dAK/dGK, utilizing the associated promoter. Sequence alignments reveal 65% identity in their DNA and 61% in their derived amino acid sequences. Encoded N-terminal sequences are identical for the first 18 residues, and both subunits share conserved sequences in common with adenylate kinase and viral TK. A more unusual conserved element, which appears to play a role in the activation of dAK, resembles the G2 loop of p21ras. Remarkably, no homologous gene(s) for the dAK/dCK pair could be found. Comparisons of amino acid sequences, isoelectric pHS and subunit masses strongly indicated that native dCK and dGK are identical in sequence, except at their extreme N-termini (M-IVL for dCK and -TVIVL for dGK), suggesting that processing of a common precursor occurs in *Lactobacillus*. Accordingly, deletion of codons 2 and 3 from *dgk* resulted in the expression of dAK/dCK in the *E. coli* host; its kinetic properties are indistinguishable from those of native dAK/dCK.

Subcloning the *dgk* or engineered *dck* gene resulted in expression of active dGK or dCK homodimers, each with a virtually unchanged K_m toward its primary deoxynucleoside. However, in common with human dCK, dCK (or dGK) homodimer exhibits secondary activities with much larger K_m s towards dAdo and dGuo (or dCyd). dCTP (or dGTP) is the best inhibitor of all three activities of the respective homodimer. Fully active heterodimers can be reconstituted simply by mixing a homodimer with independently expressed (inactive) dAK. © 1998 Academic Press

The deoxynucleoside kinases from *Lactobacillus acidophilus* R-26 (ATCC 11506) represent a sort of alternative universe with regard to DNA precursor biosynthesis. In the absence of a functional ribonucleotide reductase and *de novo* deoxyribose synthesis found in most other growing cells (1, 2), including other lactobacilli (3–6), they are endowed with a most unusual—especially for bacteria—biosynthetic network that achieves a balanced al-

location of salvaged deoxyribose to the four deoxynucleoside triphosphates. The comprehensive review of O'Donovan and Neuhard (7) still provides a good overall picture of pyrimidine metabolism in bacteria, so we limit our discussion to the unusual occurrences in *L. acidophilus* R-26, contrasting these with other bacteria. In particular, we present our current understanding of the catalytic and regulatory behavior of *L. acidophilus* R-26 deoxynucleoside kinases. It is rare that the responsibility for uncovering the properties of an interesting group of enzymes will have resided almost entirely with only one laboratory for a quarter-century! Undoubtedly more would have been known much sooner had other workers taken up the study of these enzymes, but at least we have the privilege of perspective as we share our interim understanding of how these deoxynucleoside kinases function and are regulated. We find it to be an irresistible temptation to present this understanding somewhat chronologically, in hopes that the false leads and incremental gains we have experienced will be instructive. The hurried reader may wish to turn to the end for a summation, however.

I. Historical Background—Nucleotide Metabolism in Lactobacilli

A. Unusual Deficiency Reveals New Functions

Thermobacterium acidophilus, Orla-Jensen strain R-26 (more recently classified as *Lactobacillus acidophilus* or *Lactobacillus* sp. *johnsonii*, ATCC 11506) was first applied by Hoff-Jørgensen to the microbiological assay of DNA or deoxyribonucleosides (8), which are among its numerous nutritional requirements (9, 10), whereas deoxynucleoside polyphosphates added to the growth medium apparently are not taken up by the cells (11). The organism also requires a pyrimidine and folic acid; both of these can be partially replaced by thymine, but at reduced growth rates (12–14). However, there is no specific requirement for 5-methylpyrimidine, i.e., thymine base (14). In contrast with other lactobacilli, and including the neotype strain of *L. acidophilus*, the requirement for deoxynucleoside in the R-26 strain cannot be replaced by vitamin B₁₂ (D. H. Ives, unpublished results), a cofactor of ribonucleotide reductase in lactobacilli. This deoxynucleoside requirement is understood now to be due to the functional absence of ribonucleotide reductase activity from this strain. The observation that a single deoxynucleoside can supply all of the deoxyribose needed for DNA synthesis in the R-26 strain has raised some interesting issues regarding the means of deoxyribose distribution to the four deoxynucleotides.

MacNutt first documented the existence of “*trans-N*-glycosidases” (now

nucleoside 2-deoxyribosyl transferase) capable of generating the four deoxynucleosides by transferring deoxyribosyl units from base to base (15). Enzyme partially purified from *Lactobacillus helveticus* (16, 17) and *T. acidophilus* (*L. acidophilus*) R-26 (18) revealed a relatively thermostable protein having a slightly acidic pH optimum, and for which all of the endogenous bases were potential recipients of deoxyribose. Transfer was strongly inhibited by Tris or azahypoxanthine (16), or by azathymidine (18). An alternative acceptor is water, particularly in the absence of an acceptor base, but transfer to purine or pyrimidine nitrogen occurs at substantially higher rates (16). A useful practical application of this hydrolase function has been found in the nondestructive analysis of DNA nucleosides, because it efficiently releases intact deoxyribose from the four bases (19). The crystal structure of the hexameric *Lactobacillus leichmannii* enzyme has been resolved to 2.5 Å (20), and residues participating in the active site are being identified (21).

It occurred to us that further utilization of the deoxynucleosides produced in this manner would require either four deoxynucleoside kinase activities or else nonspecific phosphotransferases to phosphorylate them. On assay, all four deoxynucleoside kinase activities were found in the deoxynucleoside-requiring *L. acidophilus* R-26, but also in the less fastidious commercial Hansen's strain (22). They are also found in the *L. acidophilus* type strain, ATCC 4356, which will grow either when supplied with vitamin B₁₂ or with deoxynucleoside in a defined growth medium (D. H. Ives, unpublished results). Although not all species have been surveyed, it appears likely that these deoxynucleoside kinases are characteristic throughout the genus. In addition to all four activities being reported in *L. leichmannii* (23), we have found deoxycytidine kinase (dCK) activity in *L. helveticus* (ATCC 15009), and it was present but less active in *Lactobacillus plantarum* (ATCC 8014); the other activities were not surveyed in this unpublished study. The deoxynucleoside kinases appear to be constitutive, even in *L. leichmannii*, in contrast with ribonucleotide reductase, which is derepressed in response to reduced levels of cyanocobalamin and deoxyribosides (24). Partial purification from the R-26 strain readily separated thymidine kinase, whereas the other three activities remained unresolved by conventional chromatographic steps (22). Separate active sites were implicated for these three, because each was inhibited by its own end-product deoxynucleoside triphosphate, whereas dCyd¹ or dGuo stimulated dAdo phosphorylation.

¹ dCyd, Deoxycytidine; dGuo, deoxyguanosine; dAdo, deoxyadenosine; dThd, deoxythymidine.

B. Deoxynucleotide Metabolism in *Lactobacillus acidophilus* R-26; Contrasts with Other Bacteria

The lactobacilli possess a number of metabolic distinctions that set them apart from other enteric bacteria, including the presence of all four deoxynucleoside kinases. The ability to incorporate dCyd, dAdo, and dGuo appears to be lacking in *Escherichia coli* B (25), confirming our own assays of *E. coli* extracts, nor was dCyd incorporation, per se, or dCyd phosphorylation found in *Salmonella typhimurium* (26, 27). The only other genus in which all of the deoxynucleoside kinase activities have been found is *Bacillus*. Wachsman and Morgan reported the phosphorylation of dAdo, dGuo, and dThd, but not dCyd, by crude extracts of *Bacillus megaterium* KM, a nonenteric bacterium (28). Later, these authors found that dCyd phosphorylation also could be demonstrated, with GTP as the preferred phosphate donor, or at about one-fourth the rate, with dCTP (29). Substantial levels of deoxycytidine kinase were found in *Bacillus subtilis* strains lacking cytidine deaminase (30). On 950-fold purification of the *B. subtilis* enzyme (31), a deoxyadenosine/deoxycytidine kinase was resolved but had very little activity toward dGuo or any other nucleoside. Again, GTP was the preferred P donor. Mutual inhibition by substrate deoxynucleosides (dAdo and dCyd) and end-product triphosphates (dATP and dCTP) indicates a common site of phosphorylation for these nucleosides.

The lactobacilli also differ from other enteric bacteria in that most of the nucleoside degradative enzymes are absent (32). Surveying *L. helveticus* extracts, Roush and Betz (16) reported the absence of cytidine deaminase, adenosine deaminase, and nucleoside phosphorylase, as well as of adenase, guanase, and cytosine deaminase and xanthine oxidase at the level of purine or pyrimidine bases. A possible explanation for the unusual absence of nucleoside phosphorylases is that their function is replaced by nucleoside 2-deoxyribosyl transferase (20). The phosphate-dependent degradation of thymidine reported to occur in *L. acidophilus* R-26 (32) may be attributed to the hydrolytic capability of nucleoside-2-deoxyribosyl transferase, rather than to thymidine phosphorylase, inasmuch as phosphate relieves the strong inhibition of the former by Tris. On the other hand, thymidine phosphorylase has been isolated from mutant *Lactobacillus casei* cells partially deficient in thymidylate kinase. The consequently reduced level of dTTP, a regulatory end-product of the phosphorylase, presumably results in its derepression (33). It is not known whether a repressible thymidine phosphorylase might exist in *L. acidophilus* R-26, nor, given the presence of an active nucleoside deoxyribosyl transferase, what benefit it would provide to the organism.

The highly active cytidine/deoxycytidine deaminase of *E. coli* (34-36) di-

verts deoxycytidine into the thymidylate pathway in which thymidine kinase and thymidylate synthetase participate. The absence of this enzyme from *L. acidophilus* R-26 suggests a different pathway for deoxycytidine utilization. The connection between the deoxycytidine and thymidine nucleotide pathways actually occurs at the level of dCMP. Siedler and Holtz tentatively identified a deoxycytidylate deaminase in crude extracts of *L. acidophilus* R-26 (37). Although there was some evidence for inhibition by dTMP, the effects of dCTP or dTTP were not examined, probably because the first of the reports of coordinate regulation of dCMP deamination by these end-products in another organism had not yet appeared (38). However, when the *L. acidophilus* enzyme was subsequently purified sufficiently to remove dCMP kinase and phosphatase activities, it was found to be inhibited strongly by dTTP (39). The inhibition could be competitively overcome by addition of dCTP, which appears to act as an allosteric activator affecting the binding of dCMP, inasmuch as sigmoid saturation plots became hyperbolic in the presence of 0.7 mM dCTP. The presence of a dCMP deaminase, normally characteristic of eukaryotes, is certainly another distinctive trait of this bacterium, further distinguishing its DNA anabolism from other enteric bacteria. This enzyme has, however, also been found in *Staphylococcus aureus* (39). In the absence of a regulated ribonucleotide reductase, maintenance of balanced deoxynucleotide pools must involve a combination of negative and positive regulatory mechanism acting on both dCMP deamination and deoxynucleoside phosphorylation.

To examine the question of utilization of deoxycytidine for dCTP and dTTP biosynthetic pathways, uniformly labeled deoxycytidine was introduced as the sole source of deoxyribose for growing *L. acidophilus* R-26 cells (19). We found very similar sugar/base labeling ratios in the dCMP and dTMP derived from the DNA, whereas label from the sugar only appeared significantly in the purine deoxynucleotides (Table I). Both of the bases were diluted somewhat by deoxyribosyl transfer to the unlabeled pyrimidine of the medium, but because the dilution was slight, the phosphorylation of the deoxycytidine must have taken place more rapidly than deoxyribosyl transfer. Our experiments clearly revealed the utilization of deoxycytidine as an unbroken unit after phosphorylation, by both the deoxycytidylate and thymidylate pathways. Moreover, these results suggest the necessary participation of thymidylate synthase in methylating the deamination product. The presence of functioning thymidylate synthase would also be predicted on the basis of the requirement for either folate or thymine by this strain. Thus, we may assume a fully developed pyrimidine deoxynucleotide network exists that is based on deoxyribosyl transfer, deoxynucleoside phosphorylation, pyrimidine interconversion at the nucleotide level, and, we assume, deoxynucleotide kinase(s) and nucleoside diphosphokinase. The principal pathways are sketched in Fig. 1.

TABLE I
RADIOACTIVITY OF NUCLEOTIDES ISOLATED FROM DNA OF *Lactobacillus acidophilus* R-26^a

| Determination | Nucleoside or nucleotide (dpm) | Sugar (dpm) | Base (dpm) | Sugar/base ratio |
|--------------------------|--------------------------------|-------------|------------|------------------|
| [G- ¹⁴ C]dCyd | 57,600 | 31,700 | 26,000 | 1.22 |
| | 60,000 | 32,800 | 27,200 | 1.21 |
| dCMP | 94,900 | 72,200 | 22,700 | 3.18 |
| | 91,300 | 67,500 | 23,800 | 2.84 |
| dTMP | 169,500 | 128,000 | 41,600 | 3.07 |
| | 158,200 | 121,000 | 37,300 | 3.24 |
| dAMP | 126,000 | 124,000 | ≤528 | — |
| | 124,600 | 121,900 | ≤821 | — |
| dGMP | 61,000 | 60,700 | 241 | — |
| | 54,200 | 53,000 | 1135 | — |

^aResults were determined after 190 min of growth in B₁₂ assay medium containing [G-¹⁴C]dCyd (19). Nucleotides isolated from DNA digested with DNase and venom diesterase were cleaved with crude nucleoside hydrolase (nucleoside deoxyribosyl transferase?) from *L. acidophilus* R-26. Sugar and base were separated by paper chromatography and counted by liquid scintillation. The entire analysis was performed in duplicate.

Although thymidylate synthase does not seem to have been specifically identified in *L. acidophilus* R-26, it has been isolated and crystallized from *L. casei*, in which this homodimeric protein is elevated in the presence of aminopterin (40). The *L. casei* enzyme has been the subject of very extensive study in several laboratories. It is one of very few enzymes to have been completely sequenced at the polypeptide level (41). Substrate binding follows an ordered Bi Bi kinetic mechanism, with dUMP binding first to the free en-

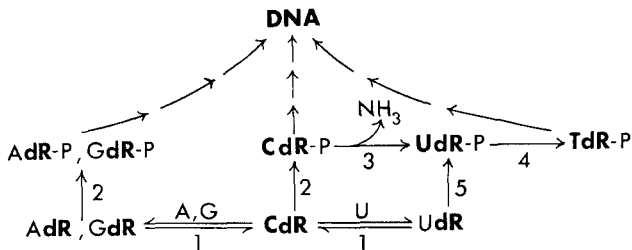


FIG. 1. Predominant pathways of deoxynucleoside metabolism in *L. acidophilus*. Enzymes involved are (1) nucleoside deoxyribosyl transferase, (2) deoxynucleoside kinases, (3) deoxycytidylate aminohydrolase, (4) thymidylate synthase, and (5) deoxythymidine kinase. AdR-P, Deoxyadenosine monophosphate; GdR-P, deoxyguanosine monophosphate; A, adenine; G, guanine; Cdr-P, deoxycytidine monophosphate; Udr-P, deoxyuridine monophosphate; Tdr-P, deoxythymidine monophosphate. (From Ref. 19.)

zyme, followed by 5,10-methylene-5,6,7,8-tetrahydrofolate (42). As does its mammalian counterpart, the enzyme forms an abortive covalent bond with the mechanism-based inhibitor, 5-fluoro-dUMP (43), but surprisingly also with 5-fluoro-dCMP (44). Molecular cloning (45) and crystallographic solutions (46) have opened the way to more detailed study of structure–function relationships of thymidylate synthase (47–49).

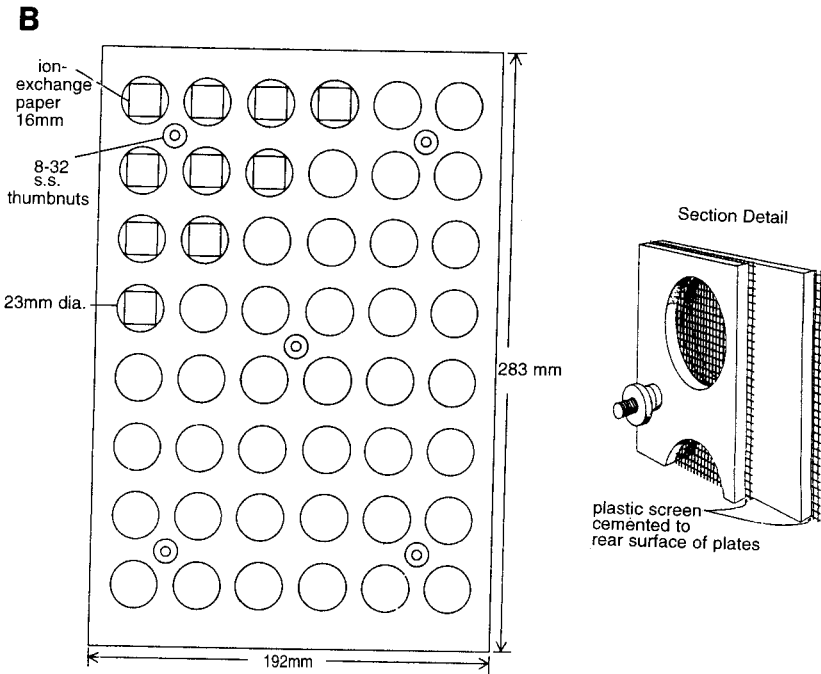
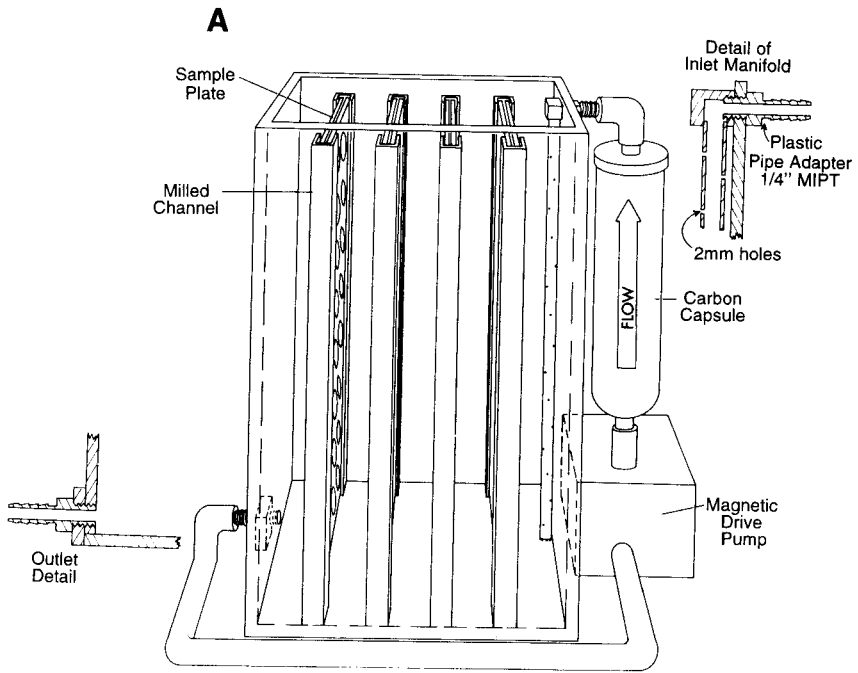
II. Purification of Deoxynucleoside Kinases from *Lactobacillus acidophilus* R-26

A. Assay of Kinase Activities

The modest levels of activity in crude *Lactobacillus* extracts are assayed most conveniently with the radiometric method developed for assay of mammalian deoxynucleoside kinases (50). In a fixed-time assay (usually 30 min), deoxynucleoside labeled with ^{14}C or ^3H was phosphorylated with the γ -phosphate of ATP. The resulting dNMP was retained on disks or squares of ion-exchange paper, and after washing away the unreacted nucleoside, the product was counted by liquid scintillation. To minimize quenching due to energy absorption by the paper, the nucleotide was eluted with HCl and KCl (the latter to ensure the release of dNTPs) and taken up into the continuous phase of a water-miscible cocktail. Initially the strong-base resin-impregnated anion-exchange paper SB-2 (Reeve-Angel) was used, but its discontinuation by the manufacturer required adapting the assay conditions to the use of DEAE-cellulose paper. This product has poor wet strength and tends to break apart during the washing operation. Therefore, a washing device (Fig. 2) providing a protective niche for each sample disk was developed (51). Additional advantages are convenient ordered organization of disks for sample application, washing, and recovery, and removal of the radioactive nucleoside from the wash liquid by means of a charcoal cartridge through which the liquid is pumped.

It was surprising that an assay temperature of 20°C is required to avoid loss of activity, given the ability of parental *Lactobacillus* to grow at 37–43°C.

FIG. 2. (A) Perspective of the complete system. The tank and sample plates are constructed of clear Lucite sheet stock. The inlet manifold creates a prevailing cross-flow of wash solution through the screened compartments holding the sample disks. The radioactive solution is pumped out at the bottom and through a charcoal cartridge, which efficiently removes radioactivity from the solution reentering the manifold. (B) View of the sample plate, showing the compartment for paper sample disks. Each plate has a plastic screen cemented to one side. Two plates arranged front-to-back, as shown in the section detail, create a series of small individual cages, each enclosing a sample disk. The five thumbnuts hold the plates together tightly enough to prevent the disks from slipping between plates. (From Ref. 51.)



We thought that this might have been due to the effects of protease in early work with cruder preparations, so the thermostability of the enzyme (dAK/dGK) has been reinvestigated with pure protein (S. Guo, personal communication). Whereas both activities are fully stable at 20–22°C, complete inactivation of dAK occurs within 5 min at 37°C, presumably with concomitant dissociation. The residual dGK activity (now monomeric or homodimeric), however, is stable for several hours at 37°C. (The residual dGK activity is only 25–50% of its original activity, because, as will be shown, it requires association with the dAK subunit for full activity.)

This assay has been applied successfully to many steady-state kinetics experiments in our laboratory. However, one peculiarity should be noted here: when the nucleoside substrate is varied, while maintaining a fixed level of radio-labeled tracer, the total count drops off as the substrate concentration (and velocity) is increased. This is distressingly counterintuitive to the inexperienced investigator, in contrast with spectrally based assays in which the precision of the assay increases in proportion to the velocity. As larger quantities of enzyme have become available, it has been possible to assay the enzyme by the classical technique of coupling ATP regeneration by pyruvate kinase with lactate dehydrogenase, following the disappearance of NADH spectroscopically (52). The latter method is ideal for the determination of K_m values falling in the millimolar concentration range.

B. Resolution of Activities—Affinity Chromatography Based on the Phosphate-Donor Site

1. BLUE SEPHAROSE PSEUDO-AFFINITY CHROMATOGRAPHY

Comigration of all three activities had been observed when applying various physical methods of isolation, including anion-exchange chromatography, preparative electrophoresis, hydroxylapatite or calcium phosphate adsorption, and gel permeation chromatography. Moreover, these activities all exhibited indistinguishable isoelectric pH values. Clearly these enzymes have very similar physicochemical properties. However, it was apparent from the time of their discovery that the three deoxynucleoside kinase activities derive from separate active sites, given the lack of competition between the substrates, and their differential response to triphosphate end-products (22). Whether these sites all reside on a single monomeric or oligomeric protein was unclear. Although the activity peaks observed on gel filtration through Bio-Gel P100 were essentially superimposable, two unexplained aberrations were to have considerable impact on future work with these enzymes. First, an activity peak was always found in the column void volume whenever Blue Dextran had been used previously for calibration. On concentration by ultrafiltration, a distinctly blue color could be detected in the protein sample.

This did not occur when Sephadex G-150 was used, however. Close scrutiny of the Bio-Gel P100 column (acrylamide beads) revealed some dye that was adsorbed near the top of the column and that became dislodged by the protein sample. Evidently the retained dye component was bound more strongly to the enzyme than to the gel matrix. This clue was to lead eventually to the resolution of the closely related deoxynucleoside kinases.

A cautionary tale can be drawn from the second unusual observation in early gel permeation experiments: two peaks of enzyme activity were seen, ~50 kDa and ~33 kDa (R. Reznik and D. H. Ives, unpublished observations). Whereas dCK and dGK were congruent in both, dAK was associated mainly with the larger peak. Similar results were seen in sucrose gradient sedimentation and preparative electrophoresis experiments. Considerable experimental effort was expended on the possibility of ligand-induced monomer-dimer interconversions, without being able to demonstrate reversibility. All of these experiments were done with extract derived from a single, large batch fermentation. Later small preparations yielded only monotonic peaks, thus we were forced to conclude that proteolytic degradation of a portion of the enzyme during its somewhat prolonged extraction with a Mantin-Gaulin mill had produced a truncated active molecule. The lactobacilli are richly endowed with proteases and peptidases, which, given the opportunity, can progressively degrade an enzyme preparation. On one occasion, gel permeation was performed with a sample of extract taken from a freezer deprived of power for a weekend, and a very broad band of kinase activity, ranging from 50 to less than 20 kDa, was observed. As a preventive against the serine proteases, at least, phenylmethylsulfonyl fluoride is now added to the cell extraction buffer. Perhaps for the same reason, the crude enzyme could not be assayed at 37°C or even room temperature, without loss of linearity over the 30-min radiometric assay. Therefore, 20°C was adopted for the "standard" assay (53, 54).

The interaction with Blue Dextran observed during Bio-Gel P100 chromatography led to attempts to use this substance for enzyme preparation by sedimentation of the complex through sucrose gradients. The strongly adsorbed enzyme was still assayable in the complex, and sedimentation profiles of the three activities suggested some degree of resolution. However, at that time it was not immediately apparent how the activities could be released from the polymer, high concentrations of KCl having little or no effect. Thompson and co-workers (55) had constructed a column by coupling the Blue Dextran to Sepharose, noting that it had affinity for enzymes possessing the "dinucleotide fold" characteristic of enzymes utilizing NAD⁺, but also for a number of other nucleotide-binding enzymes. For most of these enzymes, NADH, NAD⁺, ATP, or other nucleotides were effective eluants, strongly suggesting that the polyaromatic dye was performing as a pseudosubstrate type

of affinity ligand (56). Dr. R. L. Easterday of Pharmacia kindly provided us a sample of the free dye, Cibacron Blue F3G-A, which has a chlorotriazine group that readily reacts with polysaccharides (57). With this and Sepharose CL-6B, Blue Sepharose was prepared, substituted to about $4.6 \mu\text{mol dye/ml}$. Soluble sucrose-complexed dye strongly competed with MgATP for dCyd phosphorylation [$K_i(\text{app}) = 10 \mu\text{M}$]. Accordingly, the kinases were retained by Blue Sepharose columns equilibrated with 15% glycerol in 15 mM K-phosphate, pH 6.6, and eluted with MgATP gradients (53), as shown in Fig. 3. Substantial separation of dCK from dGK was achieved, whereas dAK was associated with both peaks. The tailing of the peaks is indicative of suboptimal elution conditions, but steeper MgATP gradients merely eluted the activities together. Somewhat improved resolution was achieved by isocratic elution of dAK/dCK at 1mM MgATP, but a total of three passages through columns of various sizes was required for complete resolution and 1200-fold increase in specific activities (54). In one of these, it was necessary to wash with NADH to remove lactate dehydrogenase, a persistent impurity. Later, baseline separation of the two enzymes was achieved by eluting the enzymes with bisubstrate mixtures: 0.5 mM dCyd plus 1 mM MgATP eluted the dAK/dCK activity, followed by 1 mM dGuo plus 5mM MgATP to elute dAK/dGK (58). Thus, for reasons that are not yet clear, it appears that the two pairs of kinases bind to Blue Sepharose through the dCK or dGK active sites,

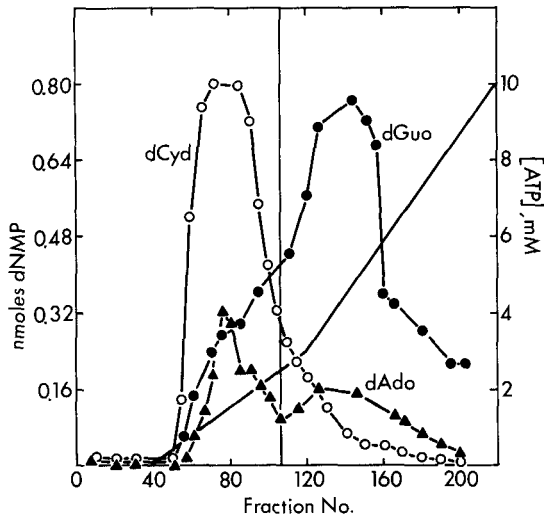


FIG. 3. Blue Sepharose CL-6B affinity chromatography of deoxynucleoside kinase activities; elution by MgATP gradients: 0–3 mM and 3–10 mM, in 15 mM K-phosphate, pH 6.6, and 15% glycerol. (From Ref. 53.)

rather than that of the associated dAK. The latter procedure by itself yielded less than 50-fold purification, so additional chromatographic steps were required for purification of the desired dAK/dGK, including UDP-Sepharose and high-pressure liquid chromatography (HPLC)-anion exchange (58). Consequently, only about 0.1 μg of apparently pure protein (2% yield of dGK) was obtained. A stainable band of protein comigrated with dGK and dAK activities in nondenaturing gel electrophoresis at several concentrations and at pH 7.5 and 8.8. Because the quaternary structure of the protein was unknown, a small amount was also subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). On silver staining, a band of about 56 kDa was seen, leading us to conclude tentatively that the protein was a monomeric multifunctional enzyme. We now know that dAK/dGK is expressed as a heterodimer, thus the band observed must have been the result of disproportionate staining of an impurity, for which silver staining is notorious.

The broad specificity of Blue Sepharose results in retention of many nucleotide-binding proteins. Nevertheless, it is a powerful tool for preliminary purification of these kinases and offers several advantages over the much more specific affinity media developed later; it is far cheaper for use on a large scale and has a much greater capacity.

2. UDP-SEPHAROSE

As with mammalian dCK (59, 60), UDP strongly inhibits the *Lactobacillus* deoxynucleoside kinases, particularly dGK (22). UTP is a fairly good phosphate donor; although it yields lower velocities than does ATP, it partially overcomes the effect of end product inhibition by tighter binding to the triphosphate site. Although competition experiments have not been carried out with the bacterial enzymes, it seems likely that UDP, as a reaction product, may compete for the phosphate-donor site, perhaps forming a dead-end complex with the enzyme.

A commercial preparation of UDP-Sepharose (P-L Biochemicals, Milwaukee, WI), in which UDP was linked through the ribose hydroxyls, was found to retain a Blue Sepharose fraction of dAK/dGK quantitative in the presence of 25 mM MnCl_2 (61). After washing with salt, the enzyme was quantitatively eluted with 5 mM EDTA, but with only about sixfold additional purification. Further fine tuning of this affinity system appears warranted, however.

C. Affinity Chromatography Based on Deoxynucleoside Derivatives

1. MODIFICATION OF DEOXYRIBOSE

The arduous path to pure enzyme by more conventional procedures—along with correspondingly poor yields—convinced us that a biospecific affin-

ity chromatography medium would be essential in obtaining protein of sufficient quality and quantity for sequence determinations and other structural studies. There had been but little effort reported on development of media for dCK, dAK, or dGK, but the successful purification of thymidine kinase by easily prepared media based on linkage of the thymidine molecule to a polymer matrix (62–65) was encouraging and suggested that, by analogy, similar results might be possible with the *Lactobacillus* deoxynucleoside kinases. However, reasoning by analogy is often a hazardous undertaking for biochemists!

Our very first attempt at developing an affinity chromatography system was to synthesize 5'-amino-2',5'-dideoxycytidine (66). The dCyd analog was coupled directly to cyanogen bromide (CNBr)-activated Sepharose, or to Affigel-10 (BioRad) by means of an *N*-hydroxysuccinimide linker. These media proved to be rather nonspecific or of low capacity. Perhaps a more careful reading of the analogous paper by Rohde and Lezius (62) would have dissuaded us from trying this approach, because the TK had been eluted by salt rather than substrate, suggesting an interaction that is mainly ionic.

The other position available for derivatization of the sugar is the 3'-OH group. Again, we turned to a technique that had been quite successful in the purification of TK (65), using a phosphodiester linkage between the hydroxyl and an aminophenol linker. Our job was made more difficult by the necessity of making the starting compounds, 3'-(4-nitrophenyl phosphate)-deoxynucleosides, whereas the thymidine derivative was commercially available at that time. Moreover, unlike thymidine, the reactive amino groups on the bases had to be protected during synthesis. On catalytic reduction of the nitro group on each compound and coupling to Sepharose by a linker, the 3'-derivatized deoxynucleosides were tested with a variety of enzymes (67). Although each of the columns had been used successfully for the purification of at least one enzyme—mitochondrial dCK (67), mitochondrial dGK (68), and barley nucleoside phosphotransferase (69) on 3'-derivatized dCyd-, dGuo-, and dAdo-Sepharose, respectively—none of the three *Lactobacillus* kinases was retained by any of the columns. The high K_i values of the soluble derivatives toward the *Lactobacillus* kinases (≥ 1 mM) indicate that a bulky group cannot be tolerated within the active site at the 3'-hydroxyl position.

2. MODIFICATIONS AT THE BASE

Noting that Lindberg and co-workers (70) had been able to derivatize the adenine amino of NADH by carboxymethylation, we then decided to apply similar chemistry to dAdo to provide a handle by which this nucleoside could be attached to agarose. N^6 -Carboxymethyl-2'-deoxyadenosine was successfully prepared and coupled to aminohexyl-Sepharose by means of water-soluble carbodiimide (N^6 -dAdo-Sepharose) (71). Another dAdo derivative tar-

getting dAK, 8-(6-aminohexylamino)-deoxyadenosine, was made by replacing the bromine of 8-bromodeoxyadenosine with 1,6-diaminohexane and coupling to CNBr-activated Sepharose (C^8 -dAdo-Sepharose). In the same manner, dGuo was derivatized to make C^8 -dGuo-Sepharose. A putative affinity medium for dCK was prepared (72) from N^4 -carboxymethyl-2'-deoxycytidine, which had been synthesized by the facile transamination of the deoxycytidine-bisulfite adduct and then coupled to aminoethyl-Sepharose (N^4 -dCyd-Sepharose). Among these potential new affinity media, however, only the N^6 -dAdo-Sepharose retained either *Lactobacillus* or mammalian kinases, but these interactions proved to be due largely to electrostatic forces.

As reflected by the K_1 values of corresponding soluble deoxynucleoside derivatives, interactions with these enzymes were weakened about two orders of magnitude by the attachment of linkers to these positions on the bases (66, 71; S. Ikeda, unpublished). Clearly the active site clefts of the *Lactobacillus* kinases must fit around their respective deoxynucleoside substrates with very tight tolerances indeed, whereas other enzymes, including thymidine kinase, appear to be structurally more forgiving.

D. Affinity Chromatography Based on the Combined Deoxynucleoside and Phosphate-Donor Sites

I. BISUBSTRATE ANALOG (dNp₄A)-BOUND SEPHAROSE AFFINITY CHROMATOGRAPHY

The preceding results with many deoxynucleoside derivatives convinced us that further work in this direction was not likely to be fruitful with these highly specific bacterial kinases. The binding determinants about the deoxynucleoside sites appear to allow little structural expansion for attachment of a linker. On the other hand, the nucleotide (phosphate donor) site is much less specific, but is therefore unlikely to yield large purification factors. The mechanism of phosphorylation with these enzymes begins with a ternary complex of enzyme, donor, and acceptor molecules, and theoretically proceeds through an in-line transfer of the phosphoryl group, with an intermediate pentavalent phosphorous stretched between the oxygens of the donor and acceptor molecule. Tightest binding, then, would be to a transition-state analog, as has been exploited in the case of adenosine deaminase and cytidine/deoxycytidine deaminase (73–75). Unfortunately, no stable transition-state analog of these kinases could readily be envisioned. Nevertheless, a linker extension at the 5'-hydroxyl of the deoxynucleoside ought to be tolerated, notwithstanding the previous results with a 5'-nitrogen. As we were considering how best to accomplish this, a timely suggestion from Professor H. B. F. Dixon led us to try a multisubstrate approach. Such compounds, when they fit perfectly, are bound to the active site by virtue of the combined binding

determinants of both substrates, with a dissociation constant that is the product of those of the individual substrates (76). In other words, the binding of a kinase having a K_d of 1 mM for ATP and 1 μ M for the nucleoside, might approach a K_d value of 1 nM with a good multisubstrate analog.

We proceeded to synthesize a family of dNp₄A (deoxynucleoside 5'-adenosine 5'''-P¹,P⁴-tetraphosphate analogs, with provision of a linker group on the nonspecific ATP end of the molecules (77). In exact analogy with adenylate kinase (78), an extra phosphate is required for optimum inhibition by the free analog, possibly to span the distance between the substrate sites, but, as is much more likely, to provide a fourth negative charge needed for electrostatic interaction with the enzyme (58). On the other hand, there was little difference in the degree of inhibition by tetraphosphate and pentaphosphate analogs. The four tetraphosphate dNp₄A derivatives were conveniently linked to adipic acid dihydrazide-Sepharose following periodate oxidation of the ribose ring to the dialdehyde. Unfortunately, these media exhibited very limited capacities for the *Lactobacillus* kinases. However, they proved to be very biospecific in terms of the activities bound, and served to establish the copurification of dAK with either dCK or dGK on dCp₄A-Sepharose or dGp₄A-Sepharose, respectively, providing further proof of the existence of two distinct sites on each protein.

2. dNTP-SEPHAROSE IS A MORE EFFICIENT MULTISUBSTRATE AFFINITY CHROMATOGRAPHY MEDIUM

As will be discussed in greater detail in a later section, the kinetic behaviors of the end-product inhibitors are qualitatively identical to those of the formal bisubstrate analogs, dNp₄A, but the dNTPs bind even more tightly ($K_i = 0.4\text{--}3\ \mu\text{M}$) than the bisubstrate compounds ($K_i = 1.4\text{--}9.2\ \mu\text{M}$). The regulatory model most consistent with all evidence is that the deoxynucleoside moiety of dNTP fits optimally at the deoxynucleoside binding site of the enzyme and thereby provides the basis for the inhibitor's specificity. The dNTP phosphates interact with the positively charged portion of the ATP-binding site, reinforcing the binding initiated by the deoxynucleoside portion, and yielding K_i values that are smaller than the K_s of the nucleoside alone. In other words, the use of multiple binding determinants makes a dNTP a more potent inhibitor. This mechanism is depicted in Fig. 4. It seems that the bulky purine base of adenine does not fit optimally at the triphosphate binding site, perhaps even interfering with proper alignment of the phosphates with their binding determinants within the site. It is interesting that the pyrimidine nucleotide UTP has been shown to produce higher catalytic efficiencies as a phosphate donor for human dCK (79, 80), and also tends to overcome end-product inhibition of either human or bacterial enzyme, presumably by binding more strongly than ATP at the triphosphate site, thereby displacing dCTP.

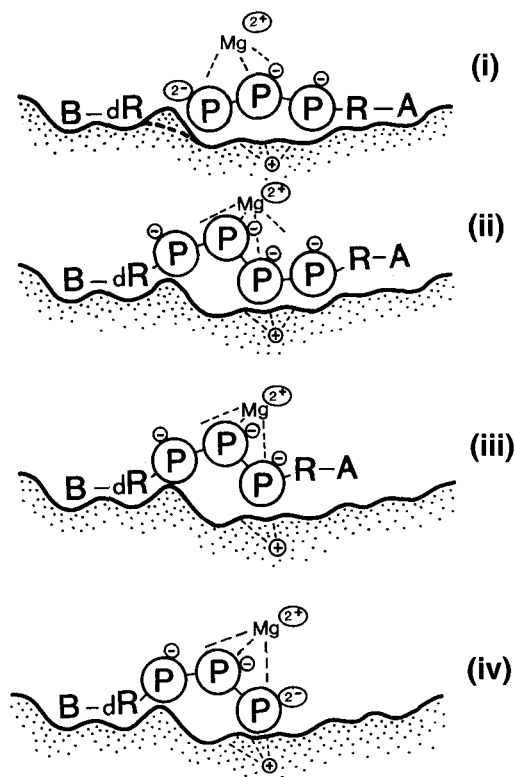


FIG. 4. Putative modes of binding of substrates and multisubstrate analogs at the active site of deoxynucleoside kinases. A, Adenine; B, any base; R, ribose; dR, deoxyribose; P, phosphate group; (i) ternary complex of enzyme and substrates (deoxynucleoside and MgATP); (ii) dNp_4A ; (iii) dNp_3A ; (iv) dNTP on the active site. (From Ref. 58.)

There is no evidence whatever for an independent regulatory site on either the human or bacterial dCKs, but direct proof of this model of multisubstrate inhibition has yet to be obtained.

On the basis of these observations, we decided to construct another affinity medium in which the elements of dNTP are linked to Sepharose through its terminal phosphate via 6-aminohexyl phosphate (81). Analogous to dNp_4A , an additional phosphoryl group was "inserted" to replace the negative charge lost on covalent linkage of the dNTP γ -phosphate. The synthesized affinity ligand for dCK, P^3 -[(6-aminohexyl)phosphoryl]-dCTP, was found to inhibit dCK versus dCyd [$K_i(\text{app}) = 61 \mu M$, at 1 mM ATP]. This ligand was coupled to CNBr-activated Sepharose ($\sim 1.2 \mu\text{mol}$ bound ligand/ml

Sepharose). An 18-ml column of this medium retained all of the dCK and about half of the dAK of an ammonium sulfate fraction of *Lactobacillus* extract. After washing with 0.1 M KCl, dAK and dCK activities were concomitantly eluted with 0.1 mM dCTP, again attesting that these two activities reside on one protein. This product proved to be about 20 μ g of homogeneous dAK/dCK, representing a 19,000-fold increase in the specific activity of dCK. At last, we had the breakthrough needed to enable us to proceed to structural studies! SDS-PAGE revealed a single polypeptide of \sim 26,000 daltons (81), which for the first time indicated that the enzyme is composed of two subunits of similar size, inasmuch as the holoenzyme was estimated by gel permeation to be \sim 50,000 daltons (53, 54). The homogeneous enzyme remains active for years at -20°C when stored in the eluting dCTP (an effective protective agent).

A second affinity column targeting dAK was prepared in the same manner (82). The runoff from the dCTP column, which contained no residual dCK, was applied to this dATP-Sepharose column. After washing with 0.3 M KCl in equilibration buffer (EB), dAK and dGK activities together were eluted with 0.1 mM dATP in EB. This dAK/dGK preparation, too, was homogeneous, and it also yielded a single 26.0 kDa peptide band by SDS-PAGE, just as with dAK/dCK. It then became clear that our earlier tentative conclusion, that the dAK/dGK is a larger monomeric protein (61), cannot be true.

We now offer some observations based on extensive experience with these two affinity media. Notwithstanding the very valuable ability to obtain homogeneous protein in a single pass, these columns have certain drawbacks. First, as may be obvious, they require expensive starting materials and are time-consuming to prepare. Overall yields of ligand amounted to only about 15% of the dNTP starting material, in part because of the rapid disproportionation of polyphosphates in organic solvents. Isolation of the desired product by ion-exchange chromatography is necessary. A second limitation is that only a tiny fraction of the bound ligand actually has the ability to retain enzyme strongly, so column capacity is quite limited. We believe it is necessary, in this case, to balance two nonspecific interactions (ionic and hydrophobic) that supplement the biospecific interaction between enzyme and affinity ligand. For example, the optimal salt concentration needed to assist in the biospecific elution of dAK/dGK from dATP-Sepharose was found to be about 0.3 M KCl, whereas almost no enzyme was eluted with dATP in either 0 or 0.5 M KCl. On repetitive use of either column, the biospecific recovery of kinases progressively deteriorates from $>60\%$ initially, to less than 10%, even though the retention of kinase activity is still nearly quantitative. We have observed that reasonable recovery could be regained by reducing, then eliminating, salt washing. At the latter stage, it was necessary to have Triton X-100 (final concentration, 0.02%) present in the column throughout the entire op-

eration. Apparently, the affinity columns gradually take on a more hydrophobic, less ionic, character, possibly due to deterioration of affinity ligands. However, by adjusting the washing and elution conditions in the manner described, the dCTP- and dATP-Sepharose columns have been used successfully 20–30 times over a 2-year period. It might be supposed that the columns are gradually filling up with tightly bound enzyme, but that seems unlikely because the columns are stripped with guanidine-HCl between runs.

3. PURIFICATION OF RECOMBINANT DEOXYNUCLEOSIDE KINASES

As will be presented below, the genes for the *Lactobacillus* kinases have been cloned and expressed in *E. coli*, and a number of mutants have been constructed. The relatively large fraction of expressed recombinant enzyme protein in crude *E. coli* extracts greatly simplifies purification to a high state of purity by conventional procedures, whereas the highly specific dNTP-Sepharose affinity columns lack the capacity for purifying the enzymes on a milligram scale or higher. For kinetic studies, purification of recombinant enzymes has been found to be even more essential than for native enzyme from the original host. The chief problem affecting activity measurements in crude *Lactobacillus* extracts is the tendency for the highly active proteolytic enzymes present to degrade the kinases. On the other hand, as has been noted, degradation of enzyme substrates does *not* tend to occur. This cannot be said for the recombinants. The abundant cytidine/deoxycytidine deaminase and purine nucleoside phosphorylase activities of *E. coli* compete for the dCyd and dGuo substrates; if not removed, erroneous determinations of activities or kinetic parameters will result. Degradation of dAdo, on the other hand, is not a major problem.

The simplest purification procedure for the recombinant kinases, which is adequate for many purposes, is passage through a 5-foot column of Sephacryl S-200-HR, resulting in enzyme of about 80% purity, as estimated by SDS-PAGE. As may be necessary, this protein can be purified further by Blue Sepharose chromatography, eluting with ATP. Alternatively, the dCTP- or dATP-Sepharose columns retain the homologous recombinant enzymes and yield microgram quantities of homogeneous enzymes having properties virtually identical to native enzymes (52). Of course, mutations affecting K_m or K_i values can be expected to weaken the interaction with these affinity columns, or with Blue Sepharose.

As an example of how the specificity of the affinity columns can be used to characterize, as well as to purify, mutant enzymes, we can cite our recent use of dCTP-Sepharose to distinguish an “engineered dAK/dCK” from the recombinant dAK/dGK from which it was genetically derived (52). Exactly as with the native enzyme, the engineered dAK/dCK was quantitatively re-

tained and eluted by dCTP as a homogeneous protein; there was no characteristic 26 kDa band in the runthrough fraction. Given the fact that recombinant dAK/dGK is not retained by this column, we could conclude that (1) the site-directed mutation truly did switch the specificity to protein indistinguishable from native dAK/dCK and (2) the mutant protein is fully active; inactive or misfolded enzyme with altered affinity for substrate was not present.

III. Steady-State Kinetics

A. pH Dependency

Within the limits of their thermostability, these enzymes are active and stable over a relatively broad range pH values, ranging from pH 6.5 to 9.5. As shown in Fig. 5, dCK exhibits an optimum maximum velocity at about pH 7.5 (66). This experiment was carried out at constant ionic strength, because dCK activity was found to increase in proportion to increasing concentrations of non-denaturing anions. In reciprocal fashion, the apparent K_m for dCyd is at a minimum at about the same pH value.

The associated dAK exhibits quite different pH dependency, however. In the absence of an allosteric activator, its specific activity gradually rises with increasing pH, reaching a maximum at pH 9.5. But, when the activator, dCyd, is present, a second pH optimum is apparent at about 7.8. The dAK activity is also affected strongly by increasing concentrations of anions, but in the opposite direction from dCK; its activity tends to fall off in a variety of 50–150 mM salts (all compared with a common potassium counterion).

B. Different Kinetic Mechanisms for dAK and dCK or dGK

Prior to larger scale isolation of enzymes, one means of gaining some understanding of their behavior was by application of steady-state kinetics. For example, kinetics has enabled us to address such several questions: Do the enzymes transfer phosphoryl groups by way of a phosphoenzyme intermediate, or is the transfer accomplished on a ternary complex of enzyme, ATP, and deoxynucleoside? If the latter occurs, are substrates bound in ordered or random fashion? These questions were answered by systematically varying the concentrations of MgATP and deoxynucleoside and by observing the patterns of competition exerted by products and multisubstrate analogs (58, 83).

For dCK, plots of reciprocal velocity versus reciprocal dCyd concentration at various concentrations of MgATP yielded a converging pattern of lines characteristic of a ternary complex (84), as shown in Fig. 6. Because the point of convergence does not occur on the horizontal axis, the apparent K_m of

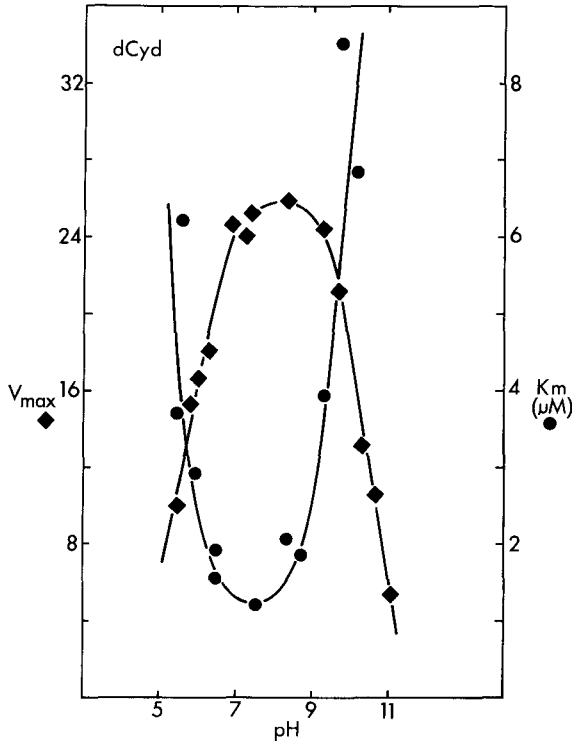


FIG. 5. The effects of variation in pH on the V_{max} and K_m of deoxycytidine phosphorylation at 20°C. Fraction V-A protein (purified on Blue Sepharose) was used and all buffers were prepared to an approximately equal ionic strength of 0.1. MgATP was near saturation at 10 mM. Buffers included K-phosphate, imidazole-HCl, Tris-HCl, and glycine-KOH. (From Ref. 66.)

dCyd becomes a function of the particular concentration of MgATP used for a particular plot line. When the data were plotted as $1/v$ versus $1/MgATP$, the point of convergence was also in the upper left-hand quadrant. True K_m values must then be obtained from slope and intercept replots. A very similar converging pattern was obtained with dGK (72), plotting $1/v$ versus $1/dGuo$. However, plots of $1/v$ versus $1/MgATP$ converge on the abscissa, indicating that dGuo concentrations have no effect on MgATP binding. To reveal whether the ternary complex is formed by random or ordered binding of substrates, product inhibition was studied. For dCK and dGK activities in their respective heterodimers, both products compete with both substrates, and we infer that the kinetic pathways for both enzymes are of the rapid-equilibrium random-order type (72, 83).

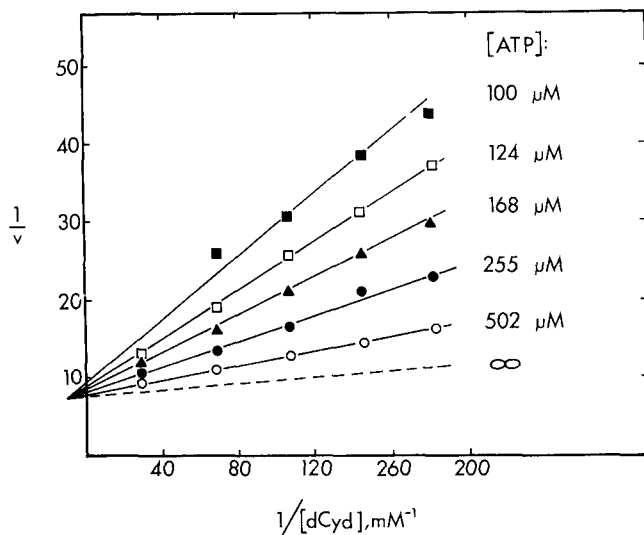


FIG. 6. Effect of varying dCyd concentrations, at fixed, subsaturating concentrations of MgATP, on the velocity of phosphorylation. (From Ref. 83.)

Surprisingly, a different pattern is observed for the dAK component of each heterodimer, compared with dCK or dGK. First, the primary plots converge on the abscissa, regardless of whether the data are fitted with dAdo or ATP as the varied substrate (72) (S. Ikeda, unpublished), indicating there is no interaction between substrates binding at the two subsites and simplifying routine K_m determinations from one or two plot lines. Second, dAMP and ATP represent the only competitive pair, and we conclude that an ordered path to the ternary complex is followed, with ATP and dAMP as the leading substrates in the forward and reverse direction, respectively (58). This is only one of a number of properties distinguishing the behavior of dAK from that of its partners! Others will be presented in other portions of this review.

C. Heterotropic Activation of dAK

With dAK/dCK or dAK/dGK isolated in the absence of substrates, the dAK exhibits only about one-seventh the activity of either of its partners. However, when the active site of dCK or dGK is saturated with dCyd or dGuo, respectively, the dAK activity increases four- to sixfold. These results (83), depicting the activity of dAK (of dAK/dCK) with or without added dCyd, are shown in Fig. 7. At dAdo concentrations ranging from 5 to 70 μM this amounts to a purely V_{max} effect, the K_m for dAdo being unchanged. As such,

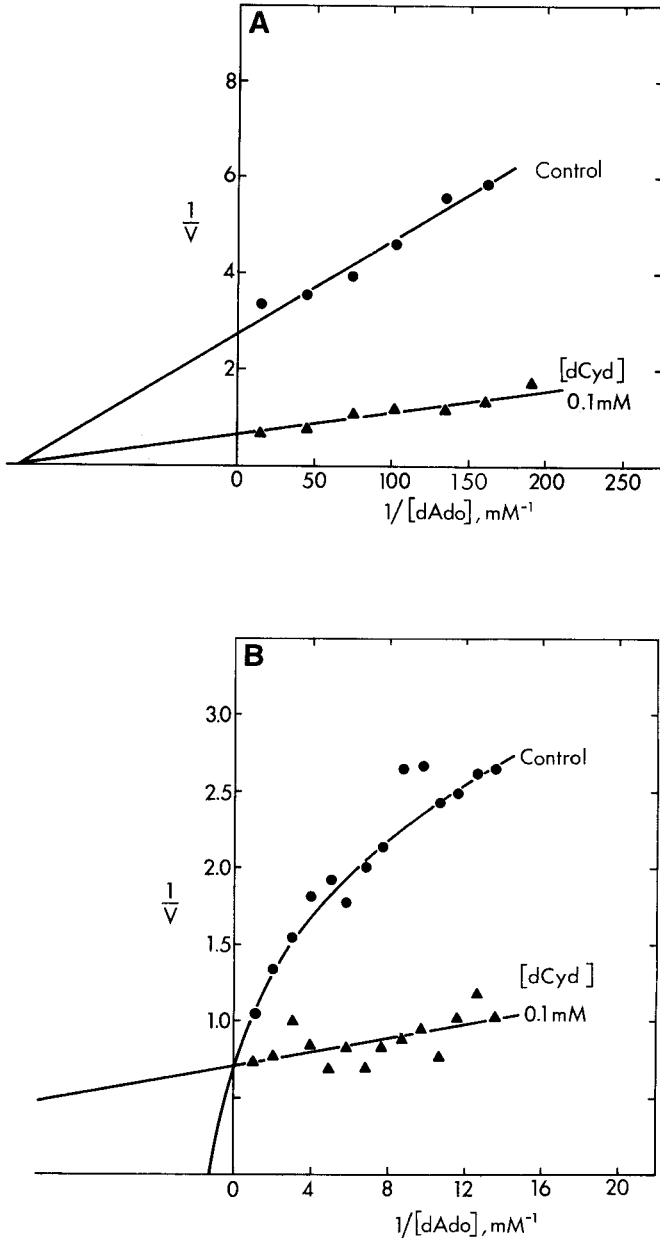


FIG. 7. Effect of varying dAdo concentrations on the rate of phosphorylation, with MgATP at 10 mM; stimulation by an effector, dCyd (0.1 mM). (A) 5 to 50 μM dAdo; (B) 60 to 1000 μM dAdo. (From Ref. 83.)

this represents a form of positive heterotropic or allosteric control of the very unusual V type. However, when the concentration of dAdo is further increased into the millimolar range, the same maximum velocity is achieved without dCyd. This is understood to mean that dAdo binds to the dCK and serves as its own heterotropic activator ligand. (Later, we will provide direct evidence for the binding and turnover of dAdo on dCK, when supplied at these higher concentrations.) An identical pattern of stimulation was observed with dAK/dGK, with dGuo serving as the heterotropic effector (61). Consistent with the understanding that they bind at the respective dCyd or dGuo sites, dCTP or dGTP also produce identical increases in the V_{\max} of dAK, also without affecting its K_m . It is clear that intersubunit communication is involved in this heterotropic effect, and, as will be demonstrated below, a conformational change is induced in dAK, analogous to the classical taut-to-relaxed transition of many other allosteric enzymes. ATP, on the other hand, produces no such effect, although it is, of course, a substrate for dAK as well as for dCK or dGK.

D. Isosteric Patterns of Inhibition by dNTP End Products

Each of the three kinase activities is strongly and specifically inhibited by its homologous end product (22, 58, 83). It would be tempting to assume that this regulatory effect, too, is exerted by means of an allosteric site on each subunit, but we believe we can rule out this possibility on a number of grounds. On the basis of kinetics, each dNTP end product behaves exactly like the corresponding bisubstrate analog, but binds even more tightly (58). Thus, for dCK or dGK, dCTP or dGTP competes with both deoxynucleoside and MgATP, as expected for a multisubstrate analog of an enzyme having a random pathway to the ternary complex. But for dAK, which follows an ordered path for substrate binding, dATP competes only with the leading substrate, ATP, exactly like the bisubstrate compound dAp₄A.

The multisubstrate analogs—or the multisubstrate behavior of the dNTPs—provide us with a very convenient means of diagnosing or confirming the kinetic pathways of these enzymes and their mutants. A novel method of plotting has been developed by which the K_i of the inhibitor can be determined, extrapolating to zero substrate concentration, as shown in Fig. 8 (58). In situations where the dNTP is in competition with both substrates, as when substrates bind in a random order, these plots of $K_i(\text{app})$ versus either [MgATP] or [deoxynucleoside] will have a positive slope, as is the case with dCK or dGK. However, where competition is with only one substrate, only that substrate will produce a slope in the $K_i(\text{app})$ versus [substrate] plot. Plotted versus the noncompeting substrate, these plots are characteristically hor-

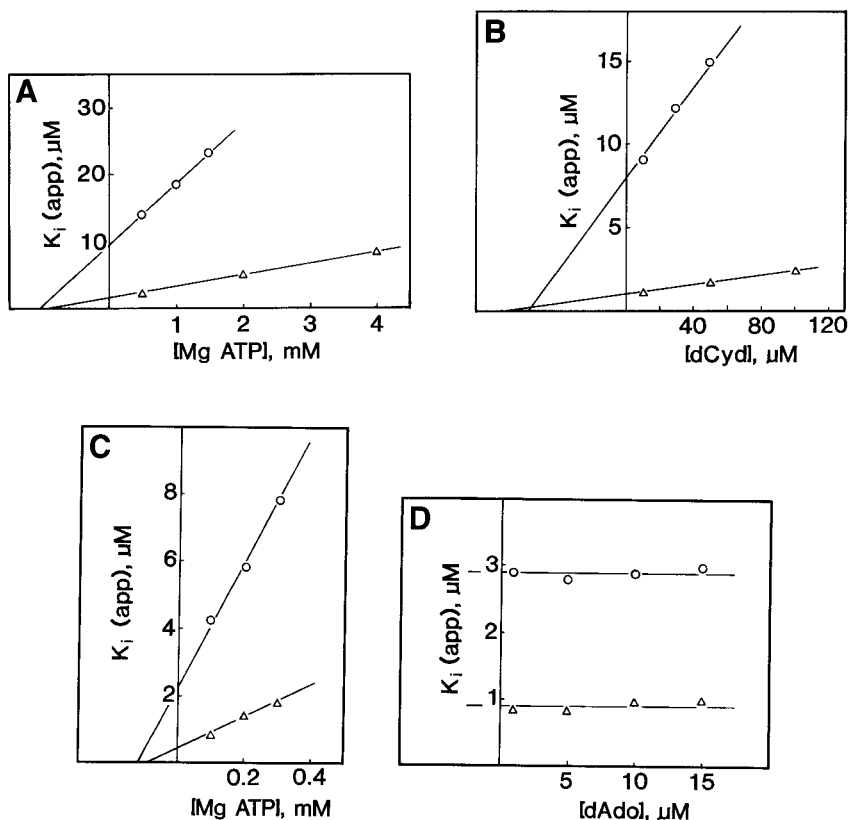


FIG. 8. Effect of several fixed substrate concentrations on the $K_i(\text{app})$ values of multisubstrate inhibitors, dNp_4A or dNTP , as they inhibit dCK or dAK . (A) Values obtained varying $[\text{dCyd}]$ at several fixed concentrations of MgATP : \circ , dCp_4A ; \triangle , dCTP . (B) Values obtained by varying $[\text{MgATP}]$ at several concentrations of dCyd : \circ , dCp_4A ; \triangle , dCTP . (C) Values obtained varying $[\text{dAdo}]$ at several fixed concentrations of MgATP : \circ , dAp_4A ; \triangle , dATP . (D) Values obtained by varying $[\text{MgATP}]$ at several concentrations of dAdo : \circ , dAp_4A ; \triangle , dATP . (From Ref. 58.)

izantal. Thus, dAp_4A or dATP competes with ATP , the leading substrate in the ordered dAK kinetic path, but not with dAdo . By using this procedure, any change in kinetic mechanism produced by a mutation (such as from random to ordered) should be readily detectable. In practice, this type of experiment is considerably easier to perform than classical product inhibition kinetics. (Plots of ADP inhibition are found to be particularly difficult to obtain, owing to large K_i values and to the marked tendency for ADP to disproportionate into ATP and AMP while in the freezer.)

The kinetic properties of the native heterodimeric deoxynucleoside kinases are summarized in Table II. Each of the subunits is highly specific for its primary deoxynucleoside substrate, with limiting K_m values of 2–8 μM . Approximating its large intracellular concentrations, MgATP exhibits K_m values that are two or three orders larger. However, conformational differences among subunits seem to have a much more profound effect on the triphosphate portion of the active site, with dGK having a much larger K_m value than the very similar dCK. Moreover, dAK exhibits an obligatory order, binding MgATP first, in contrast to the random kinetic path of dCK and dGK. The limiting dissociation constants for the dNTPs are of the order of 1 μM or less. This represents tighter binding than occurs at either substrate site, and is consistent with the multisubstrate model of dNTP inhibition.

IV. Assignment of Subunit Functions

With the availability of the dCTP- and dATP-Sepharose affinity columns, we were able, at last, to begin studying the structures of these enzymes and to clone their genes. Although it was clear that they have a dimeric quaternary structure, we did not know if the two active sites are on separate subunits, or, if one of the subunits might contain both catalytic centers, with the other serving a regulatory function. Batches of ~ 25 μg of pure dAK/dCK were isolated and submitted for preliminary Edman gas-phase sequencing, which yielded the first 28 amino acid residues (81):

(M-I-V-L-S-G-P-I-G-A-G-K-S-S-L-T-G-L/I-L-A-G/K-Y-L-G-T-N/Q-P/A-F)

This short sequence provided several pieces of valuable information. First, sufficient information for the preparation of a DNA probe was provided. Second, within only six residues of the N terminus, the sequence contained the conserved P-loop element, Gly-X-X-Gly-X-Gly-Lys, common to many ATP- or GTP-binding proteins, including adenylate and thymidine kinases (85, 86). Finally, although only one sequence was in evidence overall, deviations occurred at residues 18, 21, 26, and 27, with similar yields of two different residues for each. Our interpretation, which later was verified, was that the two subunits have very similar but nonidentical N-terminal sequences.

A very useful stroke of serendipity eased the task of assigning subunit functions. Starting with affinity-purified dAK/dCK and dAK/dGK, we planned to obtain additional sequence information with electroblotted protein (87). Because the high pH ordinarily used for SDS-PAGE is thought to be detrimental to protein sequenceability, we used the pH 6.6 MZE 3328.IV buffer system recommended by Moos *et al.* (88). A doublet with two distinct bands was observed, in contrast to the single band found at pH 9.5 (89). The

TABLE II
KINETIC PROPERTIES OF *Lactobacillus* dAK (I)/dCK AND dAK (II)/dGK HETERODIMERS

| Kinase | Substrate | K_m (μM) | V_{max} (nmol/min per mg) | V_{max}/K_m | K_m for ATP (μM) | Mechanism | K_i (μM) | Activation factor |
|----------------------------|-----------|---------------------------|-----------------------------------|----------------|---------------------------------|------------------------------|--|----------------------|
| Native dAK (I)/ dCK | dCyd | 2.2 | 2540 | 1150 | 26 | Random Bi Bi | dCTP, 1.2 (C vs. dCyd, C vs. ATP) | 1.1 by dAdo |
| | dAdo | 6.9 | 320 | 46 | 90 | Ordered Bi Bi (ATP first) | dATP, 0.7 (NC vs. dAdo, C vs. ATP) | 6.1 by dCyd |
| | | Complex, $\geq 1000^a$ | 1430 ^a | — ^b | — | — | — | — |
| Native dAK (II)/ dGK | dGuo | 170 | 1070 | 6.3 | — | — | — | — |
| | dGuo | 11.0 | 2150 | 200 | 1700 | Random Bi Bi | dGTP, 0.6 (C vs. dGuo, C vs. ATP) | 1.1 by dAdo |
| | dAdo | 8.3 | 280 | 34 | 110 | Ordered Bi Bi (ATP first) | dATP, 0.6 (NC vs. dAdo, C vs. ATP) | 5.1 by dGuo |
| | | Complex, $\geq 1000^a$ | 1820 ^a | — | — | — | — | — |
| | dCyd | — | — | — | — | — | — | — |

^aThe combined effects of secondary dAdo turnover at the dCK or dGK site, plus self-activation of dAK due to the binding of dAdo at the dCK or dGK site.

^bNot determined.

sharpness and separation of these bands were improved considerably by introducing a 12–13.2% concentration step halfway down the gel slab. The subunit peptides were then electroblotted and their N termini sequenced, with the results shown in Fig. 9. With dAK/dCK, the individual components showed the same deviations from being identical that were seen previously with the combined subunits of the heterodimeric protein over residues 1–28. With the dAK-dGK there was less protein available for sequencing, so reliable comparison ended with residue 17. However, these two peptides can be seen to differ in the first two residues, with the slower moving component being identical to the first 16 residues of both dCK and its associated dAK.

Limited tryptic proteolysis rapidly inactivated both activities of each heterodimer, by a single cleavage of each polypeptide into 11- and 16-kDa fragments (89). However, when a dNTP end product was added to the proteolytic mixture, the homologous activity was stabilized, and only its partner's activity was lost rapidly. The opposite occurred when the end product of the partner kinase was added instead. By observing the correlation between the loss of an activity and the disappearance of the faster or slower moving polypeptide, we were able to assign an activity to each of the subunits (89). Thus, we learned that N terminus of dGK lost its initiating methionine, whereas the native dCK and both of the slower moving dAK subunits were entirely identical over the first 17 residues. Moreover, the selective protection of the homologous active site by dNTP is further evidence for our proposed regulatory model in which the deoxynucleoside portion of the dNTP binds at the deoxynucleoside subsite, with its polyphosphates extending into the phosphate-binding portion of the ATP subsite.

We also should note here that the pH 6.6 electrophoretic system has produced a remarkable anomaly. The faster moving peptides are actually larger than the slower moving components by roughly 1600 Da. When we compared these assignments with the deduced amino acid sequence of dAK/dGK from its cloned genes, it was seen that dGK has nine additional residues at its C terminus (90) [as does dCK, by actual sequencing of its C-terminal tryptic peptide (52)]. Apparently, this anomaly is the result of differences in the hydrophobicity of the two subunits affecting the relative binding of SDS at pH 6.6 versus pH 9.5.

V. Cloning the Genes for dAK/dCK or dAK/dGK

A. Use of Hybridization Probes

The first sequence information available to us—that of the first 28 residues of the combined dAK and dCK subunits—presented two problems relative to

**Kinase Complex I,
dAK(I)/dCK:**

| | 5 | 10 | 15 | 20 | 25 | 30 | 35 | <u>Subunit Assignment</u> | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------|---|----|----|----|----|----|----|---------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----------------|
| S-Component: | M | I | V | L | S | G | P | I | G | A | G | K | S | S | L | T | S | L | L | A | E | ? | L | G | T | Q | A | F | Y | E | G | V | D | N | dAK(I) Subunit |
| F-Component: | M | I | V | L | S | G | P | I | G | A | G | K | S | S | L | T | L | | | | | | L | | L | G | T | | F | | | | | | dCK Subunit |

**Kinase Complex II,
dAK(II)/dGK:**

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--|--|--|--|--|--|--|-----------------|-------------|
| S-Component: | M | I | V | L | S | G | P | I | G | A | G | K | S | S | L | T | S | L | L | A | E | ? | ? | G | ? | Q | A | | | | | | | | dAK(II) Subunit | |
| F-Component: | T | V | I | V | L | S | G | P | I | G | A | G | K | S | S | L | T | | | | | | | | | | | | | | | | | | | dGK Subunit |

FIG. 9. Comparison of N-terminal amino acid sequences of the subunits of heterodimeric dAK(I)/dCK and dAK(II)/dGK from *L. acidophilus* R-26, and their functional assignments. A question mark in a sequence represents a residue position where it was not possible to identify any amino acid with certainty. Identical sequences from each subunit are enclosed in the box. Reprinted with permission from S. Ikeda, G. T. Ma, and D. H. Ives, *Biochemistry* **33**, 5328 (1994). Copyright 1994 American Chemical Society.

designing a DNA probe: it contained many highly degenerate codons and included the highly conserved P-loop sequence common to many nucleotide-binding proteins. Consequently, it was difficult to prepare a probe of reasonable size that, at the same time, could provide adequate concentrations of a sequence unique to dCK or dAK. Many partial genomic libraries were prepared from size-selected DNA that was found, by Southern blotting, to hybridize with the probes; these libraries were then screened by colony hybridization techniques. Although this approach yielded numerous positive colonies, none were found to express dCK activity. Only later did we discover that dAK and dGK activities should also have been checked routinely, because, as it turns out, dCK activity is never expressed in *E. coli* from unmodified *Lactobacillus* genes (see below). Consequently, many months of effort by several people yielded naught.

B. Colony Screening by Polymerase Chain Reaction

Clearly, a more specific method of colony selection was needed. Efforts were made to obtain sequence data from internal tryptic peptides, but because of the small amounts of material available from electroblotted fragments and the often poor sequenceability of peptide manipulated in this manner, combined with repeated equipment failures, these experiments yielded little useful information. We then decided to apply polymerase chain reaction (PCR) methodology, which was then still in its infancy, to the preparation of a nondegenerate hybridization probe, utilizing the entire 28 known codons of dAK/dCK. Initially, short (14-mers) primers with modest degeneracy (256–384) were used without success. Longer primers, of 20–21 coding nucleotides and higher degeneracy (1025–1536), each extended by added restriction sites, yielded the desired product consisting of the 77-bp target sequence within the 93-bp probe. An *Xba*I partial library constructed in pBlue-script with strongly hybridizing genomic fragments (3.5–4.7 kb) was probed by colony hybridization. Two positively hybridizing clones (from ~1000 recombinants) were obtained and found to contain an open reading frame (ORF) corresponding to dAK in tandem with a partial ORF containing the N-terminal sequence of either dCK or dGK. None of the three activities could be detected, however. On the basis of restriction-site mapping, it was deduced that a *Kpn*I fragment should enclose the entire genes for both dAK and dGK, and so a partial genomic library was constructed from 2- to 4-kb *Kpn*I fragments.

Because of the large number of false positives obtained by the colony hybridization procedure, a more specific method based on PCR has been developed for secondary screening. Twenty colonies grown on replica plates are suspended together and lysed to provide mixed templates for the PCR. Ten

such reactions are run simultaneously, employing the same primers used to generate the hybridization probe. Thus, about 200 colonies can be screened at one time. DNA from individual colonies of any pool that yields the characteristic PCR product is then run again until the positive colony is identified. In this manner, a *KpnI* clone was isolated, and it contained two open reading frames (90). Surprisingly, the activities expressed were dAK and dGK, not the expected dCK.

The DNA sequence and the deduced amino acid sequence for the 1.72-kb clone are shown in Fig. 10. Tandem genes for dAK (*dak*) and dGK (*dgk*) are separated by a 21-bp spacer that includes the Shine–Dalgarno ribosome-binding sequence for the second gene. The 5'-untranslated region contains weak -10 and -35 promoter elements, but also several conserved elements characteristic of gram-positive promoters, including the “-45 A cluster” beginning at -72 (91). The upstream S/D sequence differs by one base from that found in the connector—a difference that later proved to have important consequences in protein expression. Northern blotting experiments revealed that the transcribed mRNA is bicistronic in size, so it may be presumed that it is transcribed by means of a single endogenous promoter (52). A well-defined transcription terminator loop follows the *dgk* structural gene. Expressed in *E. coli* XL-1 Blue host cells, the dAK/dGK protein amounts to ~3% of soluble cellular protein (90).

Alignment of the deduced amino acid sequences reveals a number of interesting features (Fig. 11). A relatively high degree of amino acid identity (61%) is seen between the perfectly aligned main portions of the dAK and dGK polypeptides. However, dGK, at ~26 kDa, is eight residues longer than dAK, ~25 kDa. (As will be detailed later, native dAK and dGK peptides also differ from each other and from the sequence encoded in the first three codons of their respective genes, presumably due to differences in translational processing.) Both polypeptides have the P loop (Site i) and arginine-rich motif (Site iii) found in the phosphate-binding and -transfer domains of other kinases (92, 93). Site ii, the DRS motif, is thought to be analogous to a conserved DRH motif, which, in herpesvirus TKs, is implicated as contributing to the deoxynucleoside site (92, 94–97). Site iv, which resembles the G-2 region of p21^{ras} (98, 99), is a most unusual occurrence in a prokaryotic protein, but we shall present evidence for its involvement in intersubunit communication. Identical placement of glycines within the two polypeptides suggests that helix turns occur at the same locations in both and that the corresponding domains are of similar size. Finally, it is noteworthy that neither peptide contains cysteine, a residue apparently essential for the activity of mammalian deoxynucleoside kinases (59, 68, 100).

ggtaccagcat -271
 ctatcttacaagatcaccagatgtaacagttatTTgtgatgaagttgctgccgcaaaacttgatccaaaatatagaaactaatcctcat -181
 taattagaccgTTTacttgagcgaattaatctTTTTtagattagttcgTTTTctTTTTtgcgattTTTTtgcgTgttcgcttgactTTT -91
 TTTCTTTatagcaaaATAAAAAaccactTgTTTTcataccagatattgaattTACTtattaaagcaaaactatatatgtagaAGAacgTg -1
 A cluster -35 -10 S/D
 atgacagttattgtattgagcgggccattggagccgaaaatccagTTtaaccagTcttcttggccgaacatttaggcactcaagcctTT 90
 M T V I V L S G P I G A G K S S L T S L L A E H L G T Q A F
 tatgagggtgtagataacaatccaatTTTgccactTTattataaagatagTgctcattataactTTTctTTTaaatacttatcttctaaac 180
 Y E G V D N N P I L P L Y Y K D M A H Y T F L L N T Y L L N
 caccgTTtagctcaaattaaccaagctattcgcgatcataacagTgtgctgatcgtTcaatTTtatgaagatgccttattTTTcaaaatg 270
 H R L A Q I N Q A I R D H N S V S D R S I Y E D A L F F K M
 aatgTTgatagTggtattgctgatcctacagaattcaagatttatgatagTTTacttgagaatagTatggaacaagcacctgTaatcca 360
 N V D S G I A D P T E F K I Y D S L L E N M M E Q A P G N P
 agtaagaaaccagatctTTTaatTTtatattcatgTTTcttTgatactatgctTcaccgaattcaaaaacgTggtcgtaagTTTgaacaa 450
 S K K P D L L I Y I H V S L D T M L H R I Q K R G R K F E Q
 ttatcaaccgatccaagTTTaaagattactatgctcgcTTTTatcatattacgagcctTggtacgaaaagtataatgcatcccctaag 540
 L S T D P S L K D Y Y A R L L S Y Y E P W Y E K Y N A S P K
 atgatgattgTggtgataaatacgactTTgTTgccaatgaagacgcaagaaggaaagTTtattaacgcaattgatcaaaaacttattgat 630
 M M I D G D K Y D F V A N E D A R R K V I N A I D Q K L I D

s/D

```

atagggaaatstaactagDttaacgaatagaaggaacgtgatgacagttattgtattaagcgggccattggagccgaaaatccagtcta 720
I G N L N . M T V I V L S G P I G A G K S S L

acaggtatccttatctaaatatttgggtactaatcccttttatgaaagtgtagatgacaatcctgsttttgccattattctatgaaaacct 810
T G I L S K Y L G T N P F Y E S V D D N P V L P L F Y E N P

aaaaagtatgcctttttactgcaagtttatttcttaaatactcgsttttcggagtattaagtcagccttaactgatgataataatgtactt 900
K K Y A F L L Q V Y F L N T R F R S I K S A L T D D N N V L

gaccgttctatctacgaagatgctcttttcttccaaatgaatgcagatattggcgctgctactccagaagaagtcgatacttactatgag 990
D R S I Y E D A L F F Q M N A D I G R A T P E E V D T Y Y E

ctcttgcacaatatgatgagtgaactagatcggatgcctaagaagaatcctgatctcctggttcatatcgatgtctcatatgatacaaatg 1080
L L H N M M S E L D R M P K K N P D L L V H I D V S Y D T M

ctcaagagaattcaaaaacgtggtcgtaactatgaacaattgagttatgatccgactctagaagattactacaagcgtctacttctggttat 1170
L K R I Q K R G R N Y E Q L S Y D P T L E D Y Y K R L L R Y

tacaaccttgggatgcaagatgactattcaccaaaaatgactattgatgggtgataaacttgatttcatggcaagtgagaagatcgt 1260
Y K P W Y A K Y D Y S P K M T I D G D K L D F M A S E E D R

caagaagtcctaaatcaaattgtggctaagctcaaagaatgggtaaacttgaagacgactggaacctaatttagttaataaagcaaa 1350
Q E V L N Q I V A K L K E M G K L E D D W K P N L V K . ---

actgscggtccaatattggatgcgagtttttatttttttaatcctaacaatcaatgtcttDgatgcaccacttacagcatcagctgcttcacct 1440
-----> <-----
gttaactgatttt 1453

```

FIG. 10. Sequences of the dAdo kinase (dAK) and dGuo kinase (dGK) genes from *L. acidophilus* R-26. Lowercase letters indicate the nucleotide sequence; uppercase letters indicate the amino acid sequence. Promoter elements (-35, -10, etc.) and ribosome binding sites (S/D) are underlined and indicated as marked. The transcription terminator is marked by opposing arrows. (From Ref. 90.)

| | | (Site i) | | | | | |
|------------|--|---|------|------|-----------|------|------|
| | | 10v | 20v | 30v | 40v | 50v | 60v |
| dAK | | MTVIVLS GP IGAGKSSLTSLLAEHLGTAQAFYEGVDNPNILPLYYKDMAHYTFLLNTYLLN | | | | | |
| | | MTVIVLS GP IGAGKSSLT::L::LGT::FYE:VD:NP:LPL:Y::Y:FLL::Y:LN | | | | | |
| dGK | | MTVIVLS GP IGAGKSSLTGILSKYLGTNPFYESVDDNPNVLPFYENPKKYAFLLQVYFLN | | | | | |
| | | 10^ | 20^ | 30^ | 40^ | 50^ | 60^ |
| | | (Site ii) | | | | | |
| | | 70v | 80v | 90v | 100v | 110v | 120v |
| dAK | | HRLAQINQAIRDHNSVSD DR SIYEDALFFKMNVDSGIADPTEFKIYDLSLENMMEQAPGNP | | | | | |
| | | R: I: A: D:N:V DR SIYEDALFF:MN.D G A.P.E .Y .LL:NMM.: P | | | | | |
| dGK | | TRFRSIKSALTDNNVLD DR SIYEDALFFQMNADIGRATPEEVDTYYEILHNMMSELDRMP | | | | | |
| | | 70^ | 80^ | 90^ | 100^ | 110^ | 120^ |
| | | (Site iii) | | | (Site iv) | | |
| | | 130v | 140v | 150v | 160v | 170v | 180v |
| dAK | | SKKPDLLIYIHVSLDTMLH RIQKRGR KFEQLST DP SLK DY YARLLSYYEPWYEKYNASPK | | | | | |
| | | .K:PDLL:.I:VS DTML. RIQKRGR ::EQLS DP :L. DY Y RLL.YY.PWY.KY: SPK | | | | | |
| dGK | | KKNPDLLVHIDVSYDTML KRIQKRGR NYEQLS YDPT LE DY YKRLRLRYKPYAKYDYSFK | | | | | |
| | | 130^ | 140^ | 150^ | 160^ | 170^ | 180^ |
| | | 190v | 200v | 210v | | | |
| dAK | | MMIDGDKYDFVANEDARRKVINAI DQ KLIDIGNLN X | | | | | |
| | | M IDGDK DF:A:E::R::V:N.I .KL ::G:L:: | | | | | |
| dGK | | MTIDGDKLDFMASEEDRQEVLNQIVAKLKEMGKLEDDWKPNLVK | | | | | |
| | | 190^ | 200^ | 210^ | 220^ | | |

FIG. 11. Homology of derived amino acid sequences of dAdo kinase (dAK) and dGuo kinase (dGK). There is 61% identity in the 215-amino acid overlap (DNA* AANW). A colon denotes amino acids that are positively related using the "probability of acceptable mutation" matrix; a blank denotes negatively related amino acids and a period indicates a neutral relationship. Consensus regions, sites i-iv, described in the text, are indicated in boldface. (From Ref. 90.)

VI. dCK and dGK Are Products of the Same Gene

A. Evidence for a *dck* Gene Is Entirely Negative

Exhaustive efforts dedicated to isolation of a putative gene for dCK (*dck*) have involved the construction and probing of many, many genomic libraries (52). The application of several variations of PCR should have resulted in amplification of stretches of genomic DNA between *dak*-*dgk* and the putative *dck* [because dCK has an N-terminal sequence identical to dGK (and dAK) between residues 3 and 17]. In all of these experiments, DNA with the *dak*-*dgk* nucleotide sequence was recovered repeatedly, but only dAK and dGK activities were expressed in *E. coli* by any of these clones. Moreover, restriction mapping with many different endonucleases invariably revealed on Southern blots only the one or two bands that were attributable, in each case, to hybridization of various probes with *dak* and/or *dgk*.

In need of a DNA probe unique to dCK, we committed our resources to peptide mapping and sequencing of native *Lactobacillus* dCK subunit. A new dCTP-Sepharose affinity column was synthesized and cycled over a dozen times to accumulate sufficient enzyme protein for this analysis. The subunits were resolved by HPLC, on a C-4 reversed-phase column. The dCK polypeptide was digested exhaustively with trypsin, the peptide mixture was resolved by reversed-phase HPLC on a C-18 column, and 10 peptides were sequenced by the gas-phase Edman procedure. As shown in Fig. 12, each of the peptide sequences was identical to a corresponding region of dGK, and the peptides chosen were distributed from the N terminus to the C terminus of dGK and were equivalent to 63% of all the dGK residues (52). The single deviation from the deduced dGK sequence occurs at the extreme N terminus, where the second and third deduced residues are "deleted."

Differences between dGK and dCK, could, of course, exist in the peptides not yet sequenced. However, rather than pursue the diminishing returns of

| | |
|------------|--|
| dCK | M--IVLSGPIGAGKSSLTGILSKYLGTNPF |
| <i>dgk</i> | MTVIVLSGPIGAGKSSLTGILSKYLGTNPFYESVDDNPVLPFLFYENPKKY |
| <i>dak</i> | MTVIVLSGPIGAGKSSLTSLLAEHLGTQAFYEGVDNPNILPLYYKDMAHY |
| dCK | DDNNVLDRSIYEXA QMNAXIGRATPE |
| <i>dgk</i> | AFLQVYFLNTRFRS IKSALTDNNVLDRSIYEDALFFQMNADIGRATPE |
| <i>dak</i> | TFLNNTYLLNHRLAQINQAIRDHNSVSDRSIYEDALFFKMNVDSGIADPT |
| dCK | EVDTYEYELLHNMSELDR NYEQ |
| <i>dgk</i> | EVDTYEYELLHNMSELDRMPKKNPDLVHIDVSYDTMLKRIQKRGRNYEQ |
| <i>dak</i> | EFKIYDSLLENMMEQAPGNPSKKPDLIYIHVSLDTMLHRIQKRGRKFEQ |
| dCK | LSYDPTLEDY YKRYKPYAKYDYS PKMTIDGDKLDFMASEEDRQEV |
| <i>dgk</i> | LSYDPTLEDYKRLRLRYKPYAKYDYS PKMTIDGDKLDFMASEEDRQEV |
| <i>dak</i> | LSTDPSLKDYYARLLSYEYPWYEKYNASPKMMIDGDKYDFVANEDARRKV |
| dCK | LNQI LKEMGKLEDDWKPNLVK |
| <i>dgk</i> | LNQIVAKLKEMGKLEDDWKPNLVK |
| <i>dak</i> | INAIDQKLIDIGNLNX |

FIG. 12. Edman gas-phase sequencing of tryptic peptides from native deoxycytidine kinase subunits reveals only sequences identical to deoxyguanosine kinase. Sequences obtained from tryptic peptides of the dCK subunit are shown in boldface. They are aligned with the complete deduced amino acid sequences from the *dgk* and *dak* DNA sequences (8), which are shown in italics. (See Table III, column 2, for actual processed N-terminal sequences of dGK and dAK peptides.) Single residues that could not be assigned without ambiguity are denoted by X, but in each case one of the two possible assignments is the residue found in dCK. (From Ref. 12.)

further sequencing, we elected to compare the subunits of dAK/dGK and dAK/dCK on the basis of their molecular mass, using both matrix-assisted laser desorption/ionization (MALDI) and the more precise nanoelectrospray equipment (Table III). Also shown in the table are the actual N-terminal sequences of each native and recombinant subunit, as determined by Edman gas-phase sequencing. Native dGK (and its recombinant) has lost its initiating methionine, consistent with the second-codon effect of threonine (101), whereas dCK and both native dAKs apparently have lost residues two and three (Thr and Val), while retaining the Met. (Recombinant dAK, on the other hand, is processed in *E. coli* exactly like dGK, but without apparent effects on its activity and specificity.) Adjusting the predicted masses accordingly, we concluded that the mass of dCK must be identical to that of dGK, within experimental error, when the N-terminal difference is factored in. It seemed very likely, therefore, that mRNA editing or processing at the translational level must somehow be generating two molecular species having different functional specificities, i.e., dCK and dGK, and that similar processing alters the N-terminal sequence of dAK, but without changing its specificity. Moreover, the practically identical masses of the dAKs from the two native heterodimers strongly suggest that they are from a common subunit pool and are encoded by a common gene.

TABLE III
MOLECULAR MASS MEASUREMENTS OF DEOXYNUCLEOSIDE KINASE SUBUNITS BY MALDI
AND NANO-ELECTROSPRAY MASS SPECTROMETRY^a

| Subunit | Sequenced N terminus | Calculated MH ⁺ (residues) ^b | MALDI MH ⁺ | Δ | Nanospray MH ⁺ ($\pm\sigma$) | Δ |
|---------------------------------|----------------------|--|-----------------------|----------|---|----------|
| Native dCK ^c | MIVLSC... | 26, 116 (222) ^g | 26,138 | +22 | 26,122 (± 4.8) | +6 |
| Native dAK (I) ^c | MIVLSC... | 24, 499 (213) ^h | 24,504 | +5 | 24,500 (± 2.3) | +1 |
| Native dGK ^d | TVIVLSC... | 26, 185 (223) | 26,196 | +11 | n.d. ⁱ | - |
| Native dAK (II) ^d | MIVLSC... | 24, 499 (213) | 24,499 | 0 | 24,501 (± 1.7) | +2 |
| Recombinant dGK ^e | TVIVLSC... | 26, 185 (223) | 26,194 | +9 | 26,187 (± 2.8) | +2 |
| Recombinant dAK(I) ^e | TVIVLSC... | 24, 568 (214) | 24,561 | -7 | 24,568 (± 3.8) | 0 |
| Mutant dGK ^f | MIVLSC... | 26, 116 (222) | n.d. ⁱ | - | 26,125 (± 7.5) | +9 |

^aFrom Ref. 52.

^bAssumes sequence deduced from gene is modified at N terminus as determined by Edman sequencing (see previous column).

^cFrom dAK/dCK(I) heterodimer of *Lactobacillus acidophilus* R-26.

^dFrom dAK/dGK(II).

^eFrom recombinant dAK/dGK.

^fFrom mutant dGK/dGK homodimer.

^gAssumes sequence for dCK is same as for dGK, except at the N terminus (column 2).

^hAssumes sequence for dAK(I) is the same as dAK(II).

ⁱNot determined.

B. dCK Engineered from dGK Is Identical to Native dCK

1. COMPARISON OF HETERODIMERS

The only difference that could be discerned in primary structures of native dCK and dGK is at the N terminus, and because of the different processing of dAK in the parental and host organisms, we decided to test the effect of deleting codons two and three from the *dgk* gene by site-direct mutagenesis (52). The result, shown in Table IV, is a recombinant heterodimer that is indistinguishable from native dAK/dCK in terms of specificity, substrate K_m values, and specific activity of the affinity-purified protein. Not shown in the table is the fact that the engineered dCK subunit (of dAK/dCK) responds just as its native counterpart to homotropic inhibition by dCTP, not dGTP. Moreover, it was found that its associated dAK is now heterotropically activated by dCyd, not dGuo.

These results provide further, direct, evidence for the overall sequence identity of dCK and dGK, consistent with a common genetic origin of dCK and dGK subunits. If this is the case, the *Lactobacilli* must have a mechanism or combination of mechanisms not present in *E. coli* by which, in effect, they excise residues two and three from the peptide destined to be dCK subunits, but retain or replace Met. The same mechanism seems to apply to dAK subunits in *Lactobacilli*. Therein lies a problem in distribution: all of the

TABLE IV
KINETIC PROPERTIES OF NATIVE, RECOMBINANT, AND MUTANT HETERODIMERS^a

| Kinase | Substrate | K_m (app) (μM) ^b | V_{max} (nmol/min per mg) | V_{max}/K_m | dAK activation factor | |
|---|-----------|---|-----------------------------------|---------------|--------------------------|-----------------------|
| | | | | | dGuo (50 μM) | dCyd (50 μM) |
| Recombinant dAK/dCK | dGuo | 7.8 | 1660 | 210 | — | — |
| | dAdo | 8.4 | 120 | 14 | 5.1 | 1.1 |
| | dCyd | 1680 | 600 | 0.4 | — | — |
| Recombinant mutant dAK/dGK (dAK/dCK activity) | dGuo | 250 | 1000 | 4.0 | — | — |
| | dAdo | 8.2 | 140 | 17 | 1.2 | 8.6 |
| | dCyd | 4.5 | 2190 | 490 | — | — |
| Native dAK/dCK | dGuo | 170 | 1070 | 6.3 | — | — |
| | dAdo | 6.9 | 320 | 46 | 1.1 | 6.1 |
| | dCyd | 2.2 | 2540 | 1150 | — | — |

^aFrom Ref. 52.

^b[MgATP] = 10 mM.

native dAK is processed, but only half of the *dgk* gene product can be processed in this manner if balanced amounts of the two heterodimers are to be produced. The mystery deepens when we recognize that *dak* and *dgk* are identical over the first 18 codons, so other portions of the structure would have to dictate this differentiation.

From analysis of the *Lactobacillus* mRNA, we believe that mRNA editing is probably not involved (52). Using primers specific for the *dak-dgk* DNA, *Lactobacillus* RNA was copied by reverse transcriptase, and regions encoding the N termini were amplified by PCR. The predicted PCR product(s) should be 124 bp if no editing occurred, but should include an 118-bp product if two codons are edited out. Of the three PCR products (~120–140 bp) obtained, no 118-bp product was found. On cloning and sequencing these products, two were found to be artifacts, but the third was the predicted 124-bp product with the Thr and Val codons retained. Although it is conceivable that repeated experiments had failed to amplify an edited mRNA, it seems much more likely that the differentiation of dCK from dGK occurs at the translational level. However, such a mechanism would be most unusual, and has yet to be identified. Were there an enzyme capable of removing half of the threonine and valine residues after excision of the initiating methionine (the latter also occurring in *E. coli*), replacement of the methionine might be accomplished by a methionine aminoacyl-tRNA-protein transferase analogous to those that mark proteins destined for degradation (102–104).

2. EXPRESSION OF SINGLE GENES YIELDS TWO ACTIVE HOMODIMERS AND ONE INACTIVE SUBUNIT

The subcloning and expression of *dgk* in its native and mutant forms yields active forms of dGK and dCK, respectively (52), whereas dAK, expressed by itself, is completely inactive (105). The heterodimers are readily reconstituted by simply mixing the dAK subunits with either dCK or the engineered dCK subunits. On reconstitution, dAK activity is fully restored and dCK or dGK specific activity is doubled. Clearly, subunit interaction must play an important role in establishing the proper conformations necessary for normal activity and in facilitating the regulation of the activities. To the extent that dGK and dCK subunits interact as weakly associated homodimers, their association is strongly overridden by heterodimer formation, with a K_d estimated by “titrating” one subunit with another, to be of the order of 10^{-8} M.

The steady-state kinetics of the dCK and dGK homodimers reveal several important features (Table V). First, the K_m values for the primary substrates, dGuo or dCyd, respectively, are the same as in the heterodimers, whereas the specific activities are reduced substantially (even to zero in the

TABLE V
KINETIC PROPERTIES OF HOMODIMERIC RECOMBINANT AND MUTANT KINASES^a

| Kinase | Substrate | K_m (app) (μ M) ^b | V_{max} (nmol/min per mg) | V_{max}/K_m | Inhibition (% activity remaining) ^c | | |
|--|-----------|--|-----------------------------------|---------------|---|------|------|
| | | | | | dGTP | dATP | dCTP |
| dGK homodimer | dGuo | 5.8 | 1280 | 220 | 1.3 | 99 | 71 |
| | dAdo | 240 | 200 | 0.8 | 0.5 | 94 | 69 |
| | dCyd | 1800 | 1190 | 0.7 | 0.1 | 80 | 72 |
| Mutated dGK homodimer (dCK activity) | dGuo | 110 | 290 | 2.6 | 52 | 81 | 4.8 |
| | dAdo | 1800 | 4200 | 2.3 | 58 | 78 | 3.0 |
| | dCyd | 2.8 | 640 | 230 | 75 | 91 | 11 |

^aFrom Ref. 52.

^b[MgATP] = 10 mM.

^cFor inhibition experiment, [MgATP] = 2 mM, [dN] = 0.02 mM, [dNTP] = 0.1 mM.

case of dAK). Thus, the stimulation that occurs on reconstitution is purely an effect on turnover, just as are the heterotropic effects of dGuo or dCyd (61, 83). The homodimer kinetics also reveal that each active site is capable of phosphorylating secondary substrates, but with K_m values that are one or two orders larger than those of the primary substrates and with efficiency ratios that are only 1%, or less, in comparison with the preferred substrate. In this respect, the dCK homodimer strongly resembles the human and other mammalian dCKs (106). Moreover, each active site is preferentially inhibited by the dNTP corresponding to the preferred nucleoside substrate; e.g., the several activities of homodimeric dCK are preferentially inhibited by dCTP, not dATP or dGTP. This can best be understood in terms of binding competition for the deoxynucleoside site. Again, this behavior precisely mimics that of human dCK, so we have, for the first time, a reasonable bacterial model of mammalian dCK. It will be of interest, therefore, to explore its behavior with some of the various chemotherapeutic deoxynucleoside analogs that depend on the human enzyme for their metabolic activation.

VII. Probing the Active Site and Subunit Contacts

A. Limited Proteolysis

As mentioned earlier in reference to subunit specificity assignments, trypsin cleaves either heterodimeric native polypeptide at a single site when not protected by the homologous dNTP (89). This technique has turned out to be very useful also for analyzing the conformational states of the subunits

and for probing their conformationally mediated interaction within the heterodimer (82). It has been revealed that (1) the proteolytic inactivation of the dAK subunit by trypsin is significantly slower than that of its dGK (or dCK) counterpart, but its rate of inactivation is elevated by dGTP (or dCTP) up to the rate observed in the opposite subunit, and (2) the activating effect of dGuo (or dCyd) on dAK was abolished in the course of differential proteolytic inactivation of dGK (or dCK) in the presence of dATP. These results, together with evidence from previous steady-state kinetics experiments and the results of affinity labeling (discussed below), lead us to the following model for the conformational states of the subunits, and the role of conformational change on the heterotropic activation of dAK:

1. There is an intrinsic difference in the conformation of the two subunits, with dAK being in a constrained (closed) state and the counterpart dGK (or dCK) subunit being in a relatively relaxed (open) state. This is reflected in their relative V_{\max} values and in the different degrees of mutual heterotropic activation; the extreme conformational state may be seen in dAK, which, when expressed without dGK or dCK, is completely inactive.
2. The conformational change induced by the binding of dGuo or dGTP (or dCyd or dCTP) at the active site of the dGK (or dCK) subunit is transmitted to dAK, therefore causing its activation through subunit-subunit interaction.

A schematic representation of this model is given in Fig. 13.

B. Affinity Labeling

At earlier stages of this work, attempts were made to identify mechanistically important amino acid residues of the three deoxynucleoside kinase activities by chemical modification, using partially purified dAK/dCK (53) or dAK/dGK (61). Now, with the availability of entire amino acid sequences of each subunit, we are in a position to reevaluate those studies. The amino acid composition of each subunit, based on the sequence deduced from the DNA sequence of each gene, but modified according to the actual N terminus of each polypeptide, is given in Table VI. Both dAK and dCK were completely inactivated by the arginine-specific reagents cyclohexanedione and butanedione (53). ATP, but not dCyd, partially blocked the inactivation by these reagents, implicating an arginine at the ATP-binding site. The arginine-rich motif (R-X-X-X-R-X-R) common among kinases is found in each of the deduced *Lactobacillus* deoxynucleoside kinase sequences (Fig. 11, Site iii), so it is likely that an arginine from this motif, which generally contributes to the phosphate transfer mechanism, was modified. Little or no inactivation was

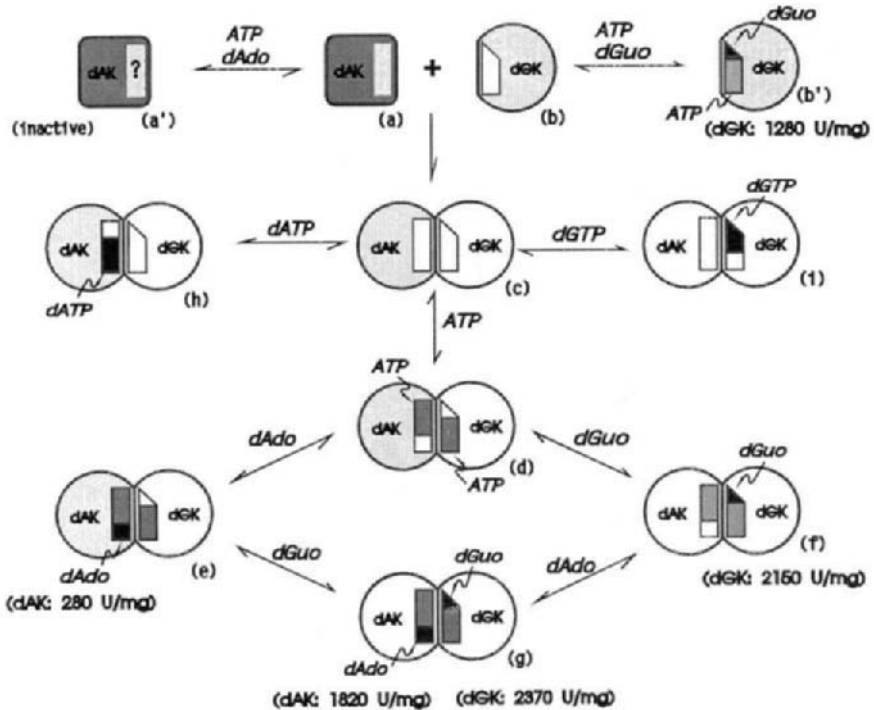


FIG. 13. Schematic representation of a model for the conformational change in the subunits of heterodimeric dAK/dGK from *Lactobacillus* on subunit association and substrate binding. The round (white), round (shaded), and rectangular (shaded) subunits represent the fully active (or open), partially active (or constrained), and inactive states of enzyme conformation, respectively. The shape, size, and location of the active site on each subunit carry no real dimension; rather they represent the relationship of substrate binding sites to the sites for dNTP binding (89). Because dAK and dGK exhibit very similar structural and functional features, the active sites for two kinase subunits are tentatively placed in a symmetrical orientation. Also, based on the results of crosslinking experiments (105), the two active sites are placed near the subunit interface, although the dimension of the interface is tentative. An exactly parallel model can be applied to the other *Lactobacillus* kinase heterodimer, dAK/dCK.

observed with the sulfhydryl reagent, diisopropylfluorophosphate. This was very perplexing at the time the experiment was performed because of the observed sulfhydryl sensitivity of mammalian dCK; we now understand the absence of a sulfhydryl effect to be due to the total absence of cysteine from any of the *Lactobacillus* kinase subunits.

The results of other chemical modification experiments are more difficult to explain in the light of present sequence information, and perhaps they illustrate the hazards inherent in this methodology. For example, the lack of

TABLE VI
 AMINO ACID COMPOSITION OF DEOXYNUCLEOSIDE
 KINASE SUBUNITS OF *Lactobacillus acidophilus* R-26

| Amino acid | Kinase subunit | | |
|----------------|-----------------|--------|--------|
| | dAK | dGK | dCK |
| A | 14 ^a | 9 | 9 |
| C | 0 | 0 | 0 |
| D | 18 | 21 | 21 |
| E | 10 | 14 | 14 |
| F | 7 | 8 | 8 |
| G | 11 | 9 | 9 |
| H | 6 | 2 | 2 |
| I | 16 | 10 | 10 |
| K | 14 | 19 | 19 |
| L | 25 | 27 | 27 |
| M | 8 | 8 | 9 |
| N | 14 | 12 | 12 |
| P | 10 | 12 | 12 |
| Q | 6 | 6 | 6 |
| R | 9 | 11 | 11 |
| S | 15 | 13 | 13 |
| T | 7 | 10 | 9 |
| V | 7 | 12 | 11 |
| W | 1 | 2 | 2 |
| Y | 15 | 18 | 18 |
| Total residues | 213 | 223 | 222 |
| MW | 24,498 | 26,184 | 26,115 |

^aThese numbers are based on the amino acid sequences deduced from DNA sequences, adjusted according to N-terminal residues observed by actual protein sequencing.

inactivation by methyl acetimidate, which should attack the ϵ -amino group of lysines, is surprising given the presence of the P-loop motif (G-X-X-G-X-G-K-S/T) at the N terminus of each peptide. The lysine within this motif is believed to participate directly in binding the γ -phosphate of ATP, and has been shown to be essential for the activity of various other kinases and ATPases (107–109). Therefore, the failure of methyl acetimidate to inactivate the *Lactobacillus* kinases suggests that the presumably essential lysine is not sufficiently exposed for the reagent to attack it. Consistent with these results, another reagent specific for the primary amino group of the lysine residue, adenosine 5'-di(or tri, or tetra)phosphopyridoxal, also failed to inactivate these kinases, even though this reagent has been used successfully in labeling the active-site lysines of other ATP-binding proteins (110–112).

Two other reagents that were able to inactivate dAK/dCK completely were *N*-acetylimidazole and rose bengal, which induced photooxidation; by default, these are assumed to attack histidine and/or possibly tyrosine. The rate of inactivation of dCK by *N*-acetylimidazole was reduced in the presence of 10 mM ATP and 2 mM dCyd, suggesting that the group attacked may be at, or near, the active site (66). Photooxidation attacked dCK and dGK much more rapidly than it did dAK, additional early evidence suggesting separate domains (or subunits) for these activities (53). So far, there is no independent evidence implicating either histidine or tyrosine in substrate binding or catalysis, so it is possible that modification of surface residues might induce global conformational changes.

A clear distinction between active sites was also provided on labeling with 5'-[*p*-(fluorosulfonyl)benzoyl]adenosine, which selectively inactivated the dGK of dAK/dGK and coordinately eliminated the ability of dGuo to stimulate dAK, suggesting that the dAK and dGK active sites are in allosteric communication (61).

Attempts at labeling the active site and/or effector site of each subunit with a wide range of deoxynucleoside and deoxynucleotide analogs have generally yielded quite limited labeling and inactivation or else rather nonspecific modifications. The list of reagents and experiments includes: (1) photoaffinity labeling with 5-iodo-dCyd, 8-azido-dAdo, 8-azido-dGuo, 8-azido-ATP, 8-azido-dGTP, or γ -(4-azidophenylphospho)-dATP; (2) direct photoaffinity labeling with dCyd, dAdo, dGuo, dCTP, dATP, or dGTP; (3) laser photolabeling with dCyd or dGuo; and (4) chemical labeling with 5'-[*p*-fluorosulfonylbenzoyl]-dGuo or diazotized 3'-(4-aminophenylphospho)-dCyd. All of the 8-azidopurine derivatives tested so far proved to be ineffective in labeling these kinases, except 8-azido-Ade. Apparently the deoxynucleoside sites of these kinases do not readily accommodate 8-azidopurine nucleosides, which are forced into the syn configuration about the glycosidic axis, rather than the apparently preferred anti configuration (113).

8-Azido-Ade, being free of rotational steric hindrance, was found to inactivate heterodimeric dGK almost completely within 3 min (82). This inactivation was prevented by the presence of 0.1 mM dGuo, indicating that the labeling takes place at the dGK active site. dAK, on the other hand, was labeled at a much lower rate than dGuo. Moreover, when 0.1 mM dAdo was added to the labeling mixture to protect the dAK site, dAK actually acquired an elevated activity concurrent with the complete inactivation of dGK. Thus, photoaffinity labeling of the dGK active site with 8-azido-Ade produces a new species of the dAK/dGK heterodimer in which the dAK is permanently activated, i.e., conformational changes induced at the dGK active site are transmitted to dAK, permanently switching it into its relaxed conformation.

C. Site-Directed Mutagenesis

I. TENTATIVE IDENTIFICATION OF THE DEOXYNUCLEOSIDE SITE

Apart from the highly conserved ATP-binding elements, alignment with the primary sequences of other deoxynucleoside kinases yielded very little in the way of additional conserved motifs. However, the conservation of the -DRH- (or FDRH) motif within the viral TKs, and the data implicating its participation in the thymidine site, led us to focus on a similar motif, DRS-, at the beginning of an 11-residue stretch in which dGK and dAK sequences are identical (90). Moreover, it aligns with a very similar stretches in human and mammalian dCKs, which begin with ERS- (114-117). Site-directed replacement of the aspartate, D78, of both dAK/dGK subunits with Glu, Asn, or Ala reduced their activities to 0.2%, or less, of the native enzyme activities. The corresponding residue from herpes simplex virus type-1 TK, D162, is also required for its activity (95, 118). Recent 2.8- to 3-Å preliminary crystallographic solutions place the F¹⁶¹DRH within the active site (96, 97). These models assign the role of Mg²⁺ coordination to D162.

The Arg-163 of HSV-1 TK is pointed toward the deoxyribose, hydrogen bonding to the 5'-O. The equivalent Arg-79 in each of the *Lactobacillus* dAK/dGK polypeptides also appears to be intimately involved in substrate binding and/or catalysis. The charge-conservative R79K replacement was carried out either in both subunits together, or in dGK only (119). In the tandem mutant the K_m values for both deoxynucleosides were increased 3- to 10-fold, whereas for MgATP, the K_m doubled for dGK but was decreased by two-thirds for dAK, another indication of the differing active conformations of the two subunits. The most dramatic contrast, however, occurs in the turnover of dAK, when dGK alone bears the R79K mutation: a 28-fold increase in V_{max} was observed, which can be interpreted as a conformational change induced in dGK locks dAK into a more relaxed configuration. As with all such effects transmitted to dAK from dGK, it is purely a V-type effect; no significant change in the K_m for either dAdo or MgATP was seen for dAK. It must be mentioned here that the K_m s listed for recombinant dGKs in this study (82) are actually K_s values. The actual limiting K_m s for U.M.C.E. are 5.5 μ M and 500 μ M for dGuo and MgATP, respectively, while for the R79K:dGK mutant they are 780 μ M and 80 μ M for dGuo and MgATP, respectively. The values given for dAK are correct, since $K_m = K_s$ for this enzyme.

We have noted before that, beginning with D⁷⁸RS, there is a stretch of 11 identical residues in dAK and dGK, suggesting their importance for substrate turnover, but presumably not for substrate specificity. Within this stretch, D84 also appears to perform a critical function, inasmuch as its replacement

in the dAK/dGK heterodimer by Asn, Glu, or Ala reduces activity to less than 1% (I. Park, unpublished data).

2. PROBING THE HETERODIMER INTERFACE

The preferred method for identifying the interfacial domains of the two subunits is by chemical crosslinking. Among the various reagents tried, 1,5-difluoro-2,4-dinitrobenzene, with its short-range 3-Å arm length, gave the most complete inactivation, with more than 80% loss of dGK but only 50% inactivation of dAK (105). These differential reactivities thus parallel the different rates of inactivation of the two subunits by rose bengal photooxidation and 5'-[*p*-fluorosulfonylbenzoyl]-Ado (53, 61). However, bifunctional crosslinking amounted to only about 20%, whereas monofunctional labeling predominated, greatly complicating the isolation of crosslinked peptide following proteolysis. Nevertheless, a crosslinked band of intact heterodimeric dAK/dGK was identified by SDS-PAGE, and quantitated by laser scanning, enabling us to assess the possible blocking effects of ligands on the crosslinking reaction. dAdo, dATP, dGTP, and ATP all inhibited crosslinking substantially, but, surprisingly, dGuo did not. Because dGuo and dGTP produce identical effects on dAK activation, and because of the kinetic evidence supporting a common binding site for the two ligands, it seems most likely that the nucleoside sites of both dAK and dGK, and their ATP sites as well, are very near the dimer interface. The ability of dGTP but not dGuo to inhibit crosslinking therefore suggests the inhibition is due to direct steric shielding of reactive sites, rather than being due to indirect positional switching resulting from a conformational change.

The intriguing presence in these prokaryotic enzymes of a conserved sequence, normally characteristic of the eukaryotic Ras-like proteins, raises the question of its function here. In p21^{ras}, the G2 loop, which the Site iv (see Fig. 11) sequences resemble, is believed to participate in the binding of Mg-GTP and in enabling the GTPase activating protein (GAP) to bind to the complex. Like the dAK subunit, the Ras GTPase is inactive until complexed with GAP. With this analogy in mind, mutagenesis of two dAK Site iv residues was carried, looking for changes in effects due to subunit interaction, i.e., in the inherent activity of dAK with or without dGK and with or without heterotropic activators (105). The dAK(S156T)/dGK replacement was performed because of the highly conserved nature of the threonine within the Ras superfamily (98, 120, 121), but its replacement produced no discernible effect. Replacement of the proline [dAK(P155A)/dGK], however, increased the dAK activity twofold, but reduced the extent of its heterotropic activation by half. In other words, the mutation appears to have changed the conformation of dAK halfway to its fully activated state, but its maximum potential turnover was not changed. Therefore, as might be expected, the remaining

potential activation, mediated by ligands of dGK, was reduced accordingly. Interestingly, the mutant dAK remains totally inactive until associated with dGK (or dCK) in a heterodimeric protein. These results also indicate that the Ras-like motif of dAK participates in its heterotropic activation. Replacement of proline by alanine would be expected to confer greater flexibility on the G2-like loop, perhaps enabling it to mimic the movement detected in the equivalent loop during Ras-GAP interaction.

The activation of dAK thus occurs in two stages, as illustrated in Fig. 13: on association with dGK or dCK, its turnover increases from zero to about one-seventh of its partner's turnover; then, on saturation of the partner with nucleoside substrate or end-product dNTP, the resulting conformational changes are transmitted to dAK, bringing its turnover to its full potential. The configuration of dGK (or dCK), on the other hand, is closer to that of the fully activated state. Association with the dAK subunit approximately doubles its turnover, but dAdo (or dATP) produces only an additional 20% of heterotropic stimulation. We do not know if self-association enables dGK or dCK to be partially active in the absence of dAK, nor do we know yet what region(s) or motifs within the primary structures of dGK (or dCK) and dAK might spell the difference between active and inactive conformational states of these highly homologous polypeptides.

VIII. Summary

The deoxynucleoside kinases of *L. acidophilus* R-26 (*Lactobacillus* sp. *johnsonii*) catalyze the salvage pathways obligatory for DNA precursor biosynthesis in this strain. In addition to thymidine kinase (TK), common to nearly all bacteria, lactobacilli contain deoxyadenosine (dAK), deoxycytidine (dCK), and deoxyguanosine (dGK) kinase-activities that are rarely found in other bacteria. Each of these three activities resides on its individual subunit within two heterodimeric proteins, dAK/dCK or dAK/dGK.

Current evidence strongly indicates that all three activities are encoded by two tandem genes, *dak-dgk*. On development of an affinity chromatography system capable of purifying nonabundant dAK/dCK to homogeneity, the N-terminal sequence of the combined subunits was determined over the first 28 residues. Aided by PCR amplification of the corresponding genomic DNA sequence, we succeeded in cloning these genes. Their expression in *E. coli*, utilizing their endogenous promoter, is mediated by a bicistronic mRNA transcript, which is then translated into fully active heterodimeric dAK/dGK. The two genes are separated by a 21-bp spacer that contains the Shine-Dalgarno ribosome-binding sequence for the second open reading frame. The two genes are identical over the first 18 codons and have identi-

cally situated conserved elements, including glycine- and arginine-rich ATP-binding elements, a putative nucleoside site that begins with -DRS-, and sequences closely resembling loop G2 of p21^{ras}. The C terminus of *dgk* extends eight codons further than that of *dak*, so the predicted masses of the two subunits are 26 and 25 kDa, respectively. However, determination of the N-terminal sequences of each native polypeptide has revealed that a type of translational processing must occur in the parental *Lactobacillus*, but of which *E. coli* appear incapable. Somehow, this processing serves, in effect, to remove the second and third encoded residues of dAK and dCK, while retaining or replacing the initiating methionine. dGK, on the other hand, has lost the initial Met, but retains residues two and three, Thr and Val. Remarkably, except for the first three residues, the sequences of dGK and dCK appear to be identical, and all evidence indicates they are the products of the same *dgk* gene. Accordingly, removal of codons two and three from the *dgk* of the tandem genes by site-directed mutagenesis results in the expression in *E. coli* of dAK/dCK, and not dAK/dGK.

The *Lactobacillus* deoxynucleoside kinase activities are tightly regulated, as is to be expected for the "first committed steps" of parallel pathways leading the deoxynucleoside triphosphate precursors of DNA. Several distinct, but integrated, regulatory mechanisms affect these activities. First, each activity is negatively controlled by its dNTP end product. Kinetic, genetic, and structural evidence all support the notion that these products bind to the active site as bisubstrate inhibitors; the deoxynucleoside moiety binds very specifically to the deoxynucleoside site, and the phosphate groups interact with the adjacent ATP-phosphate subsite, the combined binding resulting in dissociation constants that are smaller than that of either substrate alone. Moreover, as expected of multisubstrate analogs, dCTP (or dGTP) competes with both ATP and dCyd (or dGuo) in the random kinetic mechanism of dCK (or dGK), but in the ordered kinetic path of dAK, dATP competes only with the leading substrate, ATP. A second level of regulation occurs on the association of the completely inactive (*taut*) dAK subunit, with a partially active dCK or dGK subunit, possibly involving the participation of a conserved sequence resembling the G2 loop of p21^{ras}. The dCK or dGK activity increases approximately twofold, whereas the dAK activity is increased from zero to about one-seventh of the dCK or dGK activity. The final stage of activation then occurs on saturation of the dCK or dGK active site with either the deoxynucleoside substrate or the dNTP end product, and the dAK activity is increased an additional five- to sixfold. This is an example of the unusual V-type control; turnover is affected, but substrate binding (K_m) is not. The active sites (and perhaps the Ras-like motif of dAK) lie very near the dimer interface and transmit the conformational effects of ligand binding to the opposite subunit. This is an asymmetric effect, with much larger changes occur-

ring within dAK than in dCK (or dGK). The kinetic properties of each of these heterodimeric kinases were summarized in Table II, and a proposed model for their regulatory mechanisms was presented in Fig. 13.

Both unmodified dAK/dGK and "engineered" dAK/dCK are readily expressed in the *E. coli* host, with soluble protein amounting to 3–6% of the total extract. In all of their properties, these heterologously expressed enzymes appear to be identical to the native proteins, and are readily purified nearly to homogeneity by gel permeation chromatography. Subcloning and expression of the individual subunits have provided additional valuable insights. Whereas dGK and dCK retain most of their activities, dAK is completely inactive. However, reconstitution of active heterodimers is accomplished simply by mixing the dAK and dCK (or dGK) subunits. (It is now clear that the dAK subunits of both native heterodimers are identical.) Because the K_m values for dCyd or dGuo are the same in the homo- and heterodimeric states, it may be presumed that subunit association is also a V-type process, affecting turnover, but not substrate binding to dCK or dGK. (The inactive dAK subunit does not stick to a dATP–Sephrose affinity column, but we do not know yet if ATP binds to the protein in the absence of an opposing subunit.) Having active dCK and dGK homodimers free of the complicating presence of dAK has enabled us to further examine the specificities of their active sites. Besides their efficient primary activity, both were found to have secondary activities toward the other two deoxynucleoside substrates, but with much larger K_m values. Taking dCK/dCK as an example, we are struck with its similarity to human dCK in structure, specificity, and regulation. Both phosphorylate dGuo and dAdo as secondary substrates, but each is most strongly inhibited by dCTP. The dCK homodimer derived from *Lactobacillus* may thus be a useful model to further probe the structural elements controlling substrate specificity toward analogs of dCyd, dAdo, or dGuo. It is easily expressed at higher levels than the heterodimers, and should be a good candidate for crystallographic solutions.

REFERENCES

1. P. Reichard and L. Rutberg, *Biochim. Biophys. Acta* **37**, 554 (1960).
2. P. Reichard, *Annu. Rev. Biochem.* **57**, 349 (1988).
3. M. Downing and B. S. Schweigert, *J. Biol. Chem.* **220**, 521 (1956).
4. R. L. Blakley and H. A. Barker, *Biochem. Biophys. Res. Commun.* **16**, 391 (1964).
5. M. Goulian and W. W. Beck, *J. Biol. Chem.* **241**, 4233 (1966).
6. D. Panagou, M. D. Orr, J. R. Dunstone, and R. L. Blakley, *Biochemistry* **11**, 2378 (1972).
7. C. A. O'Donovan and J. Neuhard, *Bacteriologic. Rev.* **34**, 278 (1966).
8. E. Hoff-Jorgensen, *Biochem. J.* **50**, 400 (1952).
9. J. Soska, *J. Bacteriol.* **91**, 1840 (1966).

10. T. Morishita, Y. Deguchi, M. Yajima, T. Sakurai, and T. Yura, *J. Bacteriol.* **148**, 64 (1981).
11. W. C. Schneider and R. L. Potter, *Proc. Soc. Exp. Biol. Med.* **94**, 798 (1957).
12. T. Okazaki and R. Okazaki, *Biochim. Biophys. Acta* **35**, 434 (1996).
13. S. Lovtrup and D. Shugar, *J. Bacteriol.* **82**, 623 (1961).
14. R. V. Sawula, S. Zamenhof, and P. J. Zamenhof, *Can. J. Microbiol.* **21**, 501 (1975).
15. W. S. MacNutt, *Biochem. J.* **50**, 384 (1952).
16. A. H. Roush and R. F. Betz, *J. Biol. Chem.* **233**, 261 (1958).
17. W. S. Beck and M. Levin, *J. Biol. Chem.* **238**, 702 (1963).
18. J. C. Marsh and M. E. King, *Biochem. Pharmacol.* **2**, 146 (1959).
19. M. B. Davis and D. H. Ives, *J. Bacteriol.* **126**, 1136 (1976).
20. S. R. Armstrong, W. J. Cook, S. A. Short, and S. E. Ealick, *Structure* **4**, 97 (1996).
21. S. A. Short, S. R. Armstrong, S. E. Ealick, and D. J. T. Porter, *J. Biol. Chem.* **271**, 4978 (1996).
22. J. P. Durham and D. H. Ives, *Biochim. Biophys. Acta* **228**, 9 (1971).
23. J. W. Powell and J. T. Wachsman, *Appl. Microbiol.* **25**, 869 (1973).
24. R. K. Chambeer and R. L. Blakley, *J. Biol. Chem.* **241**, 4710 (1966).
25. H. O. Karlström, *Eur. J. Biochem.* **17**, 68 (1970).
26. J. Neuhard, *J. Bacteriol.* **96**, 1519 (1968).
27. C. Janion, *Mol. Gen. Genet.* **153**, 179 (1977).
28. J. T. Wachsman and D. D. Morgan, *J. Bacteriol.* **105**, 787 (1971).
29. J. T. Wachsman and D. D. Morgan, *Appl. Microbiol.* **25**, 506 (1973).
30. B. K. Rima and I. Takahashi, *J. Bacteriol.* **129**, 574 (1977).
31. H. Möllgaard, *J. Biol. Chem.* **255**, 8216 (1980).
32. R. V. Sawula, S. Zamenhof, and P. J. Zamenhof, *J. Bacteriol.* **117**, 1358 (1974).
33. Y. Avraham, N. Grossowicz, and J. Yashphe, *FEMS Microbiol. Lett.* **82**, 287 (1991).
34. T. P. Wang, H. Z. Sable, and J. O. Lampen, *J. Biol. Chem.* **184**, 17 (1950).
35. R. M. Cohen and R. Wolfenden, *J. Biol. Chem.* **246**, 7561 (1971).
36. G. W. Ashley and P. A. Bartlett, *J. Biol. Chem.* **259**, 13615 (1984).
37. A. J. Siedler and M. T. Holtz, *J. Biol. Chem.* **238**, 697 (1963).
38. G. F. Maley and F. Maley, *J. Biol. Chem.* **237**, PC3311 (1962).
39. R. C. Sergott, L. J. Debeer, and M. J. Bessman, *J. Biol. Chem.* **246**, 7755 (1971).
40. R. B. Dunlap, N. G. L. Harding, and F. M. Huennekens, *Biochemistry* **10**, 88 (1971).
41. G. F. Maley, R. L. Bellisario, D. U. Guarino, and F. Maley, *J. Biol. Chem.* **254**, 1301 (1979).
42. H. H. Daron and J. L. Aull, *J. Biol. Chem.* **253**, 940 (1978).
43. R. J. Cisneros, J. W. Zapf, and R. B. Dunlap, *J. Biol. Chem.* **268**, 10102 (1993).
44. L. Liu and D. V. Santi, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1209**, 89 (1994).
45. K. Pinter, V. J. Davisson, and D. V. Santi, *DNA* **7**, 235 (1988).
46. J. Finer-Moore, E. B. Fauman, P. G. Foster, K. M. Perry, D. V. Santi, and R. M. Stroud, *J. Mol. Biol.* **232**, 1101 (1993).
47. S. C. Climie, C. W. Carreras, and D. V. Santi, *Biochemistry* **31**, 6032 (1992).
48. L. Liu and D. V. Santi, *Biochemistry* **31**, 5100 (1992).
49. U. Schellenberger, P. Balaram, V. S. N. K. Francis, B. K. Shoichet, and D. V. Santi, *Biochemistry* **33**, 5623 (1994).
50. D. H. Ives, J. P. Durham, and V. S. Tucker, *Anal. Biochem.* **28**, 192 (1969).
51. D. H. Ives, *Anal. Biochem.* **136**, 416 (1984).
52. N. Ma, S. Ikeda, S. Guo, A. Fieno, I. Park, S. Grimme, T. Ikeda, and D. H. Ives, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14385 (1996).
53. M. R. Deibel, Jr., and D. H. Ives, *J. Biol. Chem.* **252**, 8235 (1977).
54. M. R. Deibel, Jr., and D. H. Ives, *Meth. Enzymol.* **51**, 346 (1978).
55. S. T. Thompson, K. H. Cass, and E. Stellwagen, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 669 (1975).

56. R. L. Easterday and I. M. Easterday, "Immobilized Biochemicals and Affinity Chromatography" (R. B. Dunlap, ed.), p. 123. Plenum, New York, 1974.
57. H. J. Böhme, G. Kopperschläger, J. Schulz, and E. Hofmann, *J. Chromatogr.* **69**, 209 (1972).
58. S. Ikeda, R. Chakravarty, and D. H. Ives, *J. Biol. Chem.* **261**, 15836 (1986).
59. J. P. Durham, and D. H. Ives, *J. Biol. Chem.* **245**, 2276 (1970).
60. J. C. Sarup, M. A. Johnson, V. Verhoef, and A. Fridland, *Biochem. Pharmacol.* **38**, 2601 (1989).
61. R. Chakravarty, S. Ikeda, and D. H. Ives, *Biochemistry* **23**, 6235 (1984).
62. W. Rohde and A. G. Lezius, *Hoppe-Seyler's Z. Physiol. Chem.* **352**, 1507 (1971).
63. E. P. Kowal and G. Markus, *Prep. Biochem.* **6**, 369 (1976).
64. P. Gröbner and P. Loidl, *J. Biol. Chem.* **259**, 8012 (1984).
65. N. Tamiya, T. Yusa, Y. Yamaguchi, R. Tsukifuji, N. Kuroiwa, Y. Moriyama, and S. Fujimura, *Biochim. Biophys. Acta* **995**, 28 (1989).
66. M. R. Deibel, Jr., Ph.D dissertation. The Ohio State University, Columbus, Ohio, 1977.
67. S. Ikeda, I. Park, P. Gardner, and D. H. Ives, *Biochemistry* **23**, 1914 (1984).
68. I. Park and D. H. Ives, *Arch. Biochem. Biophys.* **266**, 51 (1988).
69. D. C. Prasher, M. C. Carr, D. H. Ives, T.-C. Tsai, and P. A. Frey, *J. Biol. Chem.* **257**, 4931 (1982).
70. M. Lindberg, P. O. Larsson, and K. Mosbach, *Eur. J. Biochem.* **40**, 187 (1973).
71. S.-M. Wang, Ph.D dissertation. The Ohio State University, Columbus, Ohio, 1980.
72. R. Chakravarty, Ph.D dissertation. The Ohio State University, Columbus, Ohio, 1985.
73. R. Wolfenden, *Pharmacol. Ther.* **60**, 235 (1993).
74. L. Betts, S. Xiang, S. A. Short, R. Wolfenden, and C. W. Carter, Jr. *J. Mol. Biol.* **235**, 635 (1994).
75. A. Radzicka and R. Wolfenden, *Meth. Enzymol.* **249**, 284 (1995).
76. L. D. Byers, *J. Theor. Biol.* **74**, 501 (1978).
77. S. Ikeda and D. H. Ives, *J. Biol. Chem.* **260**, 12659 (1985).
78. C. E. Lienhard and I. I. Secemski, *J. Biol. Chem.* **248**, 1121 (1973).
79. J. C. White and R. L. Capizzi, *Cancer Res.* **51**, 2559 (1991).
80. D. S. Shewach, K. K. Reynolds, and L. Hertel, *Mol. Pharmacol.* **42**, 518 (1992).
81. S. Ikeda, R. P. Swenson, and D. H. Ives, *Biochemistry* **27**, 8648 (1988).
82. S. Ikeda and D. H. Ives, *Biochemistry* **33**, 13373 (1994).
83. M. R. Deibel, Jr., R. B. Reznik, and D. H. Ives, *J. Biol. Chem.* **252**, 8240 (1977).
84. I. H. Segel, "Enzyme Kinetics. Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems," p. 273. Wiley, New York, 1975.
85. H. Otsuka and S. Kit, *Virology* **135**, 316 (1984).
86. D. C. Fry, S. A. Kuby, and A. S. Mildvan, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 907 (1986).
87. P. Matsudaira, *J. Biol. Chem.* **262**, 10035 (1987).
88. M. Moos, Jr., M. Y. Nguyen, and T.-Y. Liu, *J. Biol. Chem.* **263**, 6005 (1988).
89. S. Ikeda, G. T. Ma, and D. H. Ives, *Biochemistry* **33**, 5328 (1994).
90. G. T. Ma, Y. S. Hong, and D. H. Ives, *J. Biol. Chem.* **270**, 6595 (1995).
91. M. C. Graves and J. C. Rabinowitz, *J. Biol. Chem.* **261**, 11409 (1986).
92. N. K. Balasubramaniam, V. Veerisetty, and G. A. Gentry, *J. Gen. Virol.* **71**, 2979 (1990).
93. G. E. Schulz, E. Schiltz, A. G. Tomasselli, R. Frank, M. Brune, A. Wittinghofer, and R. H. Schirmer, *Eur. J. Biochem.* **161**, 127 (1986).
94. D. K. Dube, J. D. Parker, D. C. French, D. S. Cahill, S. Dube, M. S. Z. Horwitz, K. M. Munir, and L. A. Loeb, *Biochemistry* **30**, 11760 (1991).
95. M. E. Black and L. A. Loeb, *Biochemistry* **32**, 11618 (1993).
96. K. Wild, T. Bohner, A. Aubry, G. Folkers, and G. E. Schulz, *FEBS Lett.* **368**, 289 (1995).

97. D. G. Brown, R. Visse, G. Sandhu, A. Davies, P. J. Rizkallah, C. Melitz, W. C. Summers, and M. R. Sanderson, *Nature (Struct. Biol.)* **2**, 876 (1995).
98. H. R. Bourne, D. A. Sanders, and F. McCormick, *Nature (London)* **349**, 117 (1991).
99. D. Markby, R. Onrust, and H. R. Bourne, *Science* **262**, 1895 (1993).
100. C. Bohman and S. Eriksson, *Biochem. (Life Sci. Adv.)* **9**, 11 (1990).
101. P.-H. Hirel, J.-M. Schmitter, P. Dessen, G. Gayat, and S. Blanquet, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8247 (1989).
102. J. L. Sohl, "Transfer RNA: Biological Aspects" (D. Söll, J. N. Abelson, and P. Schimmel, eds.), p. 493. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1980.
103. A. Varshavsky, *Cell* **69**, 725 (1992).
104. G. Abramochkin and T. E. Schrader, *J. Biol. Chem.* **270**, 20621 (1995).
105. S. Guo, N. Ma, and D. H. Ives, *J. Biol. Chem.* **272**, 6890 (1997).
106. M.-Y. Kim, S. Ikeda, and D. H. Ives, *Biochem. Biophys. Res. Commun.* **156**, 92 (1988).
107. D. Parsonage, M. K. Al-Shawi, and A. E. Senior, *J. Biol. Chem.* **263**, 4740 (1988).
108. J. Reinstein, M. Brune, and A. Wittinghofer, *Biochemistry* **27**, 4712 (1988).
109. G. Tian, H. Yan, R.-T. Jiang, F. Kishi, A. Nakazawa, and M.-D. Tsai, *Biochemistry* **29**, 4296 (1990).
110. Q. Dong and H. J. Fromm, *J. Biol. Chem.* **265**, 6235 (1990).
111. J. K. Tamura and M. Gellert, *J. Biol. Chem.* **265**, 21342 (1990).
112. M. Tagaya and T. Fukui, *Biochemistry* **25**, 2958 (1986).
113. J. J. Czarnecki, *Biochim. Biophys. Acta* **800**, 41 (1984).
114. E. G. Chottiner, D. S. Shewach, N. S. Datta, E. Ashcraft, D. Gribbin, D. Ginsburg, I. H. Fox, and B. S. Mitchell, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1531 (1991).
115. A. Karlsson, M. Johansson, and S. Eriksson, *J. Biol. Chem.* **269**, 24374 (1994).
116. L. Y. Wang, U. Hellman, and S. Eriksson, *FEBS Lett.* **390**, 39 (1996).
117. M. Johansson and A. Karlsson, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7258 (1996).
118. J. Fetzer, G. Folkers, I. Müller, and G. M. Keil, *Virus Genes* **7**, 205 (1993).
119. Y. S. Hong, G. T. Ma, and D. H. Ives, *J. Biol. Chem.* **270**, 6602 (1995).
120. E. F. Pai, U. Krengel, G. A. Petsko, R. S. Good, W. Kabsch, and A. Wittinghofer, *EMBO J.* **9**, 2351 (1990).
121. M. V. Milburn, L. Tong, A. M. DeVos, A. Brüner, Z. Yamaizumi, S. Nishimura, and S.-H. Kim, *Science* **247**, 939 (1990).

This Page Intentionally Left Blank

Molecular Analyses of Metallothionein Gene Regulation

SUSAN L.-A. SAMSON AND
LASHITEW GEDAMU

*Department of Biological Sciences
The University of Calgary
Calgary, Alberta T2N 1N4*

| | |
|--|-----|
| I. Overview of Metallothioneins | 258 |
| A. Short Review of Biochemical Characteristics and Putative Functions | 258 |
| B. Conservation of MT Gene Structure in Humans and Rainbow Trout | 258 |
| II. Metallothionein Gene Regulation | 259 |
| A. Short Review of Inducers | 259 |
| B. Metals | 259 |
| III. Metallothionein Promoter Organization and Function | 261 |
| A. General Cis-Acting Elements | 261 |
| B. Metal-Responsive Elements | 261 |
| C. The Trout MT-B Promoter: Response to Oxidative Stress | 270 |
| IV. MRE-Binding trans-Acting Factors | 274 |
| A. General Background | 274 |
| B. <i>In Vitro</i> and <i>In Vivo</i> Activation of Human and Mouse MTF | 276 |
| C. Cotransfection Studies on Human and Mouse MTF cDNAs | 280 |
| V. Conclusions and Suggestions for Further Research | 285 |
| References | 285 |

Metallothionein (MT) genes encode small proteins that chelate metal ions through metal-thiolate bonds with cysteine residues. MTs may have a role in cellular zinc homeostasis and metal detoxification. Congruent with these putative functions, MT gene transcription is induced by metals via multiple metal-responsive elements (MREs) present in the MT gene 5'-regulatory regions. This chapter mainly is focused on studies of the functional and physical interactions of MRE binding proteins with MT promoters from human and rainbow trout. In addition to mediating zinc induction, MREs may make important contributions to non-metal induced promoter activity. In part, differential basal activity of MREs appears to be determined by sequence and position in the promoter. During zinc induction, increased functional MRE activity correlates with increased activity of mammalian MRE binding proteins by zinc treatment *in vivo* or *in vitro*, as detected by electrophoretic mobility shift assays. Interestingly, the addition of cadmium *in vitro* or *in vivo* has no detectable effect even though it strongly induces MT gene expression in the same time course. This raises questions about how the ef-

fects of cadmium are mediated by MREs. The molecular masses and MRE complex migration of the zinc-responsive factors we detect are consistent with mouse and human metal-responsive transcription factor (MTF) and expression of the MTF cDNAs increases co-transfected MT promoter activity in both mammalian and trout cell lines underlining the conservation of MRE binding factor function among diverse species. © 1998 Academic Press

I. Overview of Metallothioneins

A. Short Review of Biochemical Characteristics and Putative Functions

Metallothioneins (MTs) are small molecular mass (6 to 7 kDa) nonenzyme proteins that act as biological chelators of heavy metals through formation of metal–thiolate bonds with their numerous cysteine residues (1). Although the precise cellular roles of MTs have not been defined, their metal-binding capacity suggests involvement in trace metal homeostasis, such as for zinc and copper, as well as detoxification of more poisonous metals such as cadmium and mercury, which have no known cellular function (2). Metallothionein genes are organized as multigene families (3) that display developmentally regulated expression patterns (4, 5) as well as cell-type-specific regulation (6, 7). In addition, the MT primary sequences are highly conserved among vertebrates (1). Collectively, these observations suggest that MTs have an essential cellular role.

In an attempt to better define the function of MTs, researchers have inactivated the MT-I and MT-II genes in mice (8, 9). Although there was no observable effect on development with MT gene disruption, the mice were more susceptible to cadmium poisoning, which is not surprising considering that amplification of MT genes in cell lines increases cadmium resistance (8, 10). In other experiments, cells cultured from null MT-I/II mice were more sensitive to oxidants (9), suggesting that MTs provide protection from oxidative stress. Further support for an antioxidative role for MT has been accumulating from a variety of laboratories. Pretreatment *in vivo* with MT-inducing agents provides protection from chemicals that generate oxidative stress (11). Likewise, yeast or monkey MT expression confers protection from oxidants in yeast strains lacking superoxide dismutase (12) and, *in vitro*, rabbit liver cadmium or zinc MT scavenges hydroxyl radicals (13).

B. Conservation of MT Gene Structure in Humans and Rainbow Trout

Despite the ubiquity of MTs in the animal kingdom, there are wide differences in the number and complexity of MT genes observed in different or-

ganisms. The human (h)MTs are encoded by a multigene family of at least 12 members (14), with numerous nonfunctional (15) and processed pseudogenes (16) and a number of functional isoforms. So far, these include one MT-III isoform gene (17), one MT-II isoform gene, MT-IIA (14), and five known functional MT-I isoform genes, MT-IA (18), MT-IB (19), MT-IE (15), MT-IF (15, 20), and MT-IG (31), clustered on chromosome 16 (21, 22). The trout (t)MT system is less complex, consisting of two known genes and possibly three or four genes in total (23). Our laboratory has cloned both the cDNA and genomic sequences of two trout isoforms, tMT-A and tMT-B (28, S. Schieffman, unpublished data).

Both the human and trout MT genes encode 61 or 62 amino acid polypeptides, and the primary amino acid sequences of the trout MTs are comparable to those in humans with greater than 60% homology (28). More importantly, the functional conservation of trout and human MTs is apparent in that the positions of 19 out of 20 cysteine residues are conserved in trout with the expected Cys-X-Cys or Cys-Cys arrangements, which allow the coordination of metals (1). Such conservation strengthens a widely accepted view of common ancestral origins for teleost and mammalian MTs.

Although the introns of trout and human MT genes differ in length, the tripartite gene structure—three exons interrupted by two introns at amino acids 9 and 31 (or 32 for tMT-A)—is conserved among all higher eukaryotic MTs. Intron 2 occurs at the junction of the α and β domains of the polypeptides in mammals and probably trout, so that the α domain contains 11 cysteines and the β domain has 9 (28, 23).

II. Metallothionein Gene Regulation

A. Short Review of Inducers

In considering the regulation of MT gene expression, the appropriate question is “What doesn’t induce MT synthesis?” Kagi (1991) (13a) has provided a comprehensive list of inducers, in addition to metal ions, that increase MT synthesis in cell culture and whole organisms. Growth factors, oncogene products, tumor promoters, and second messengers such as cAMP and diacylglycerol have all been documented to increase MT message and protein. Also, MT is generated in response to vitamins and hormones, inflammatory agents, and cytotoxic chemicals (and their downstream mediators, cytokines) as part of the acute-phase response. In fact, MT synthesis is associated with almost any chemical or physical stress.

B. Metals

Although a variety of agents have been shown to induce MTs, metals are still the common and most potent inducers of all MT isoforms. Considering

the possible physiological importance of trace metal metabolism, fine control of MT transcription may be required in order for MTs to accommodate the changing metal environment or requirements of the cell. Whether this means greater induction of more effectual isoforms or a precise combination of isoforms has not been demonstrated. Even so, MT regulation has been shown to be very complex. Differential patterns of expression are observed as variable transcription activity in different tissues and cell types, diverse levels of transcription among MT isoforms, and diverse responses to metal inducers.

The trout MT genes and selected members of the human MT genes have been shown to be regulated in a cell-type-specific manner in response to metal ions. In each case, DNA methylation appears to play a role in the observed cell-type-specific patterns of MT gene expression (6, 7, 19).

In humans, differential MT gene expression is observed as varied transcription responses to different metals. This involves both the rate and the maximal level of transcript accumulation. The differential responses of MT-IIA, MT-IG, and MT-IF genes to Zn, Cd, and Cu in HepG2 cells have been quantified using probe excess titration followed by RNase protection analysis (24). Using primer extension analysis of transcripts of all known hMT isoforms in HepG2 cells, the different genes were shown to respond to Cd, in the following order of transcription activity: MT-IE \geq MT-IIA > MT-IG > MT-IF > MT-AI (25). The trout MT-A and MT-B genes also exhibit differential transcription regulation in response to metals (26, 27). The kinetics and the level of MT mRNA accumulation differed in various tissues in response to Zn, Cd, and Cu (27). The transcriptional activity of MT genes is related to the region of the 5' flanking DNA that exhibits promoter activity, although there is evidence of metal-specific posttranscriptional regulation (29). For example, the differential transcriptional regulation of the human MT-IF and MT-IG genes has been localized to the proximal 150 bp of the promoter regions (30). In fact, these genes share greater than 80% sequence homology in this region. One important variation between the two promoters was that MT-IG has a consensus TATAA box whereas MT-IF has a TATCA sequence (20, 31). Using site-directed mutagenesis and functional analysis, the TATA box sequence was shown to make an important contribution to the differential regulation of these genes (25).

The differential responses to metals are not surprising considering the varied physiological roles of the metals that bind MTs. These can range from very toxic with no biological function, such as Cd, to metals that are involved in numerous biological processes and are not normally cytotoxic, as with Zn (2, 32). In fact, the optimal dose of Zn, Cu, or Cd that is required for maximum transcription induction in both human and trout cell lines differs for each metal (6, 7), which may be a function of differential metal uptake and

efflux (34). Also, MTs display differential binding affinities for metals and it is possible that metal-binding transcription factors might also exhibit differential affinities for the metals, which in turn would affect their capacity to activate transcription. Alternatively, each metal might bind different transcription factors that show differential specificity for metal-responsive elements (MREs).

III. Metallothionein Promoter Organization and Function

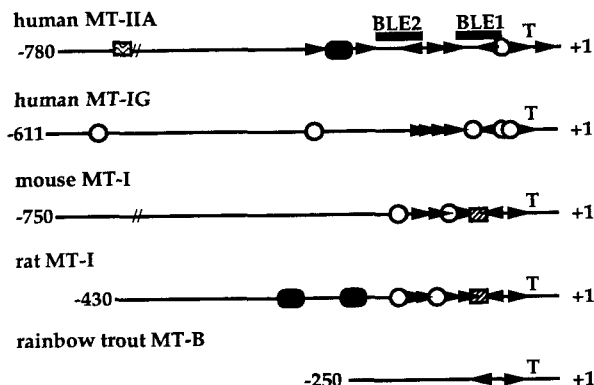
A. General Cis-Acting Elements

Because MTs are mainly regulated at the transcriptional level (35), the end effector molecules of MT induction are trans-acting factors that work at the level of the MT promoter. In many cases, the cis-acting elements that mediate induction have been characterized. In warm-blooded vertebrates, a common motif is the GC box, which is a potential binding site for the Sp1 transcription factor (36). The human MT-IIA regulatory region is one of the more complex MT promoters containing basal level enhancers (BLEs), which bind factors, such as AP1 and AP2, to mediate transcription activation by serum growth factors, tumor-promoting phorbol esters, and other activators of protein kinase A and protein kinase C pathways (37, 38). As well, an interferon-responsive element (39) has been identified at a distal locale. Glucocorticoid-responsive elements are also present in the human MT-IIA-(37) and rat MT-1 promoters (40). More recently, a number of MT promoters from divergent species have been recognized to contain antioxidant-responsive elements (41), which mediate induction by hydrogen peroxide and redox-active quinones (42). This lends further support to the possible role of MT as a cellular antioxidant.

B. Metal-Responsive Elements

1. SYNTHETIC SEQUENCES DEFINE CONSENSUS FOR METAL REGULATION

Although many MT promoter cis-acting elements are unique to one gene or species, the single common motif from invertebrates to vertebrates is the metal-responsive element, which is always present in multiple copies (Fig. 1) and mediates induction by metals. The MRE consensus was originally identified by comparisons of MT promoter sequences, and subsequent functional analysis of MT promoter deletions and synthetic MRE sequences narrowed the consensus (44–46). The MRE consists of a conserved 7-bp core



LEGEND:








- BLE1** = Basal Level Enhancer 1 containing Ap1, Ap2 and Ap4 transcription factor binding sites
BLE2 = Basal Level Enhancer 2 containing Ap1 and Ap2 transcription factor binding sites
 = Ap1 binding site/Antioxidant Responsive Element (ARE)
 = Glucocorticoid Responsive Element (GRE)
 = GC box, putative Sp1 transcription factor binding site
 = Interferon Responsive Element (IRE)
 = Metal Responsive Element (MRE) in forward orientation
 = Metal Responsive Element (MRE) in reverse orientation
T = TATA box Element
 = 50 b.p.

FIG. 1. Schematic representations of metallothionein 5'-flanking regions. The cis-acting elements contained in MT gene regulatory regions are diagrammed based on the following sources: human MT-IIA gene (39, 107), human MT-IG gene (31), mouse MT-I gene (44, 50), rat MT-I gene (40), and the rainbow trout MT-B gene (23).

sequence (underlined) surrounded by semiconserved flanking sequences: CTNTGC(G/A)CNCGGCCC.

2. ROLE OF MREs WITHIN NATIVE PROMOTER SEQUENCES

Our focus was on determining the contributions of cis-acting elements, namely MREs, to metallothionein promote activity. However, where previous studies have employed synthetic metal-responsive promoters, we have examined MRE contributions to transcription regulation in the context of native MT promoter organization and sequences. Metallothionein promoters

from rainbow trout and humans were convenient for investigating this problem from the perspective of a conserved metallothionein gene regulatory mechanism.

a. Synergistic Interaction of MREs in Response to Zinc. Using transfection analysis, we have determined that the two MRE sequences present in the rainbow trout hepatoma cell line tMT-B promoter cooperate for a significant response to zinc. This strong synergism allows metal induction even when both MREs are point mutated at the first nucleotide T of the MRE core (43). This mutation has previously been shown to destroy synthetic MRE dimer activity even with the first MRE forward and the second reversed, as is found in the tMT-B promoter (45, 46). Because of the remaining inducibility of the tMRE point mutants, substitutions of the entire tMREa and tMREb sequences were required to reveal the role of each MRE in tMT-B promoter activity (Fig. 2). tMREa appears to be able to function in an independent manner because substitution of tMREb still allows basal activity above the TATA box alone and elicits a significant response to metals. However, substitution of MREa devastates basal promoter activity and allows only nominal metal induction, indicating that MREb is not able not function efficiently independent of MREa and requires synergism with MREa (43). Both MREs must cooperate for a response to metals because the sum of the fold inductions generated by each MRE when the other was substituted does not equal that of the native promoter.

b. MREs Have a Basal Role in Addition to Metal Induction. *i. BASAL ROLE OF HUMAN MT-IG MREa.* Previous studies have shown that MRE activity is modulated by distance from the TATA box (45, 47). However, we have observed that, in transfected HepG2 cells, the TATA box of MT promoters could be dependent on the activity of the MREs, which appear to compensate for a TATAA to TATCA mutation during zinc induction even though basal activity is abolished. A biologically relevant example is the hMT-IF promoter, which has lower absolute levels of activity due to a TATCA sequence but displays greater metal fold induction than hMT-IG (25). We have observed a qualitative difference in the physical interaction of the TATA box binding protein (TBP) with TATCA. To extrapolate from this, newly recruited MRE binding factors could facilitate TBP binding to this variant during metal induction (48). This could occur through direct physical interaction or, indirectly, through effects on the DNA conformation in the proximal promoter to increase the affinity of TBP for the TATCA sequence (49), compensating for the mutation.

Although the devastating effect of the TATCA mutation could suggest

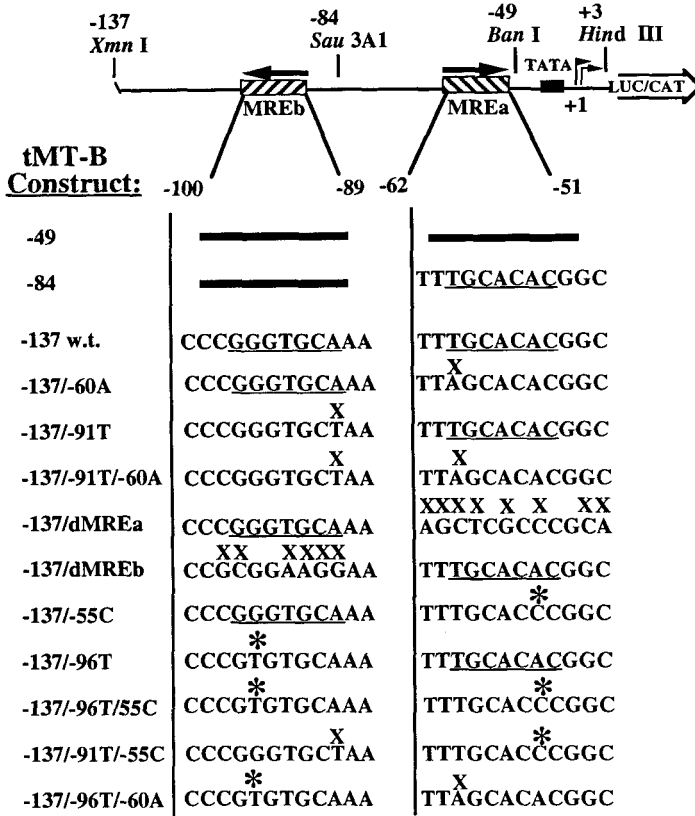


FIG. 2. The rainbow trout MT-B promoter and constructed metal-responsive element mutations. A schematic diagram of the tMT-B promoter is shown with the restriction site used to create the 5' deletions -49, -84, and -137. As well, the sequences of the wild-type (underlined) and mutated MREa and MREb elements are shown. Mutations intended to inactivate MRE function are marked with an X and mutations in flexible MRE consensus nucleotides are marked with an asterisk. The wild-type and MRE mutant promoter fragments were fused to CAT and LUC reporter genes in the vectors pMEV1R and pMEV35R, respectively.

that the TATA box is solely responsible for MT-IG basal promoter activity in HepG2 cells, the core mutation of MREa (Fig. 3) also substantially decreases general promoter activity (48). This also is consistent with a requirement for MRE effects on TATA box activity, with these two elements working in concert to promote transcription initiation. In the past, basal contributions by MREs have not been discussed because *in vivo* footprinting of rat and mouse MT-I promoters generally supports metal-dependent occupation of

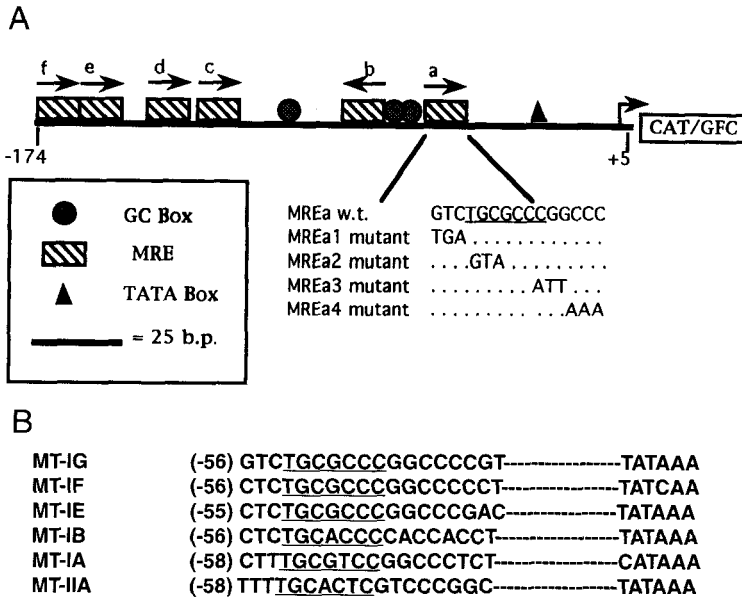


FIG. 3. The MT-IG gene regulatory region and comparison of MT-I promoter MREa sequences. (A) A schematic representation of the MT-IG (-174/+5) is shown with the positions of the TATA box, GC boxes, and MREs (arrow direction conveys MRE orientation). The sequences of MREa and the available trinucleotide mutants are given (108). (B) The MREa and TATA box sequences of the following human MT promoters are compared: MT-IG (25), MT-IF (20), MT-IE (15), MT-IB (19), MT-IA (18), and MT-IIA (14). The MREa heptanucleotide core sequence is underlined.

MREs (40, 50). However, a lower strand G residue of the rat MT-I MREb was constitutively protected in a cadmium-resistant Fao cell line containing amplified MT genes (40). In the work of Mueller *et al.* (50), it was found that amplified mouse MT-I promoters displayed basal occupancy of MREd, and the pattern of protection was increased but consistent in metal-treated cells, but inconsistent with Sp1 occupation of the juxtaposed GC box sequence. In addition, examination of the data also reveals that other mouse MT-I MREs were somewhat protected in non-metal-treated cells (50), suggesting that there is some constitutive occupancy of MREs.

Because of the basal contributions of MT-IG MREa, mutation in its core actually increases fold induction, which is inconsistent with a role as a metal-responsive element. However, MREa is still essential to metal induction because it is necessary for efficient expression in spite of the presence of five other MREs. These additional MREs appear to be active because addition of the distal promoter fragment (MREs, c, d, and e) to the proximal region

(MREs a and b) increases metal induction of MT-IG in HepG2 cells (30). Together, these observations suggest that the efficient function of distal MREs *in vivo* is dependent on the proximal promoter region, and specifically MREa. Perhaps MT promoter elements are organized such that they require occupancy of an MRE (MREa) with the TATA box for constitutive expression levels as well as for communication with remaining MREs during metal induction. This could explain why gene disruption of the MRE binding factor MTF-1 in embryonic mouse cells abolishes both basal and metal-induced MT-I expression (51), in spite of the presence of other non-MRE basal elements (37). If MRE binding factors are regulated directly by cellular metal ion concentrations, even the normal level of zinc ions in the cell could promote partial activation of the metal regulatory factor pool for maintenance of basal MT gene expression.

ii. **BASAL AND METAL-INDUCIBLE ROLES OF TROUT MT-B MREs.** In the trout MT-B promoter (Fig. 2), we also have noted that tMREa is zinc responsive without cooperative interactions with a second MRE. This unusual characteristic has been observed only with the mouse MT-I MREd sequence (46). Because distance from the TATA box has been shown to affect MRE function in synthetic promoters (45, 47), the basal as well as independent metal-induced activity of tMREa may, in part, be a consequence of close proximity to the TATA box. This would explain why the devastating effect of the MREa mutant (-60A) on basal promoter activity cannot be reversed fully if the tMREb sequence is altered to that of tMREa (-137/-96T/-60A). The activity of -137/dMREa is also not rescued by the tMREb/-96T (43).

Basal MT levels are significant in rainbow trout tissues and primary cell culture (27, 52), so that the activity of the transfected MT-B promoter without metal induction can be considered biologically relevant. From our results, other than tMREa in the proximal tMT-B 5' flanking sequence, there are no additional elements that drive promoter activity in the absence of metal treatment. Further, contributions to basal activity by distal elements are also dependent on tMREa.

In past discussions, MREs have been considered only as inducible elements because mammalian MT promoters contain numerous other potential basal elements (40, 50, 53) and MREs generally are not protected *in vivo* prior to metal induction (40, 50). Nevertheless, tMREa also contributes to tMT-B promoter activity in mammalian cell lines, so the basal activity of MRE binding factors is conserved from trout to mammals (43).

Multiple MRE-specific proteins that have been detected within one cell line or species could have distinct functions for basal and metal-induced MT expression. In our hands, altering conditions of denaturation and renaturation during Southwestern analysis allows detection of multiple MRE binding

factors (S. Samson and L. Gedamu, unpublished data). It is also possible that one MRE binding factor is responsible for both basal and metal-induced levels of MT expression. Here, we have designated the term "basal" to indicate the level of promoter activity without metal treatment. However, zinc obviously is present at significant levels in the cell as an essential structural component of enzymes and transcription factors (54). If metal regulatory factors, namely, MRE binding factors, are sensitive to zinc concentrations in the cell, then some of these factors could be active even without metal treatment, and these could be responsible for MRE activity under basal conditions.

Previous studies with synthetic MRE sequences have shown that synergism between MREs is required for metal induction, similar to our observations with the tMT-B MREb. The mouse MT-I gene MREd is the only sequence that has been shown to have independent activity assayed by transfection, with basal activity partially justified by its overlap with an Sp1 transcription factor binding site (45, 46, 50). In this sense, the observed basal activity and independent metal inducibility of the tMT-B MREa are novel.

We considered that the observed differential function of MREa and MREb for basal and metal-induced tMT-B activity could be due to position relative to the TATA box, orientation or subtle sequence differences between the MREs. MREa and MREb differ by only one nucleotide (underlined), **TTTGCACNCGG**, which is -55A in MREa and -96G (C in forward orientation) in MREb. This position is the least conserved nucleotide of the MRE consensus; all four bases have been found to occupy this position in natural MREs (37). To determine if this nucleotide was responsible for differential basal and metal-induced MRE activity, reciprocal point mutants were made to alter MREa to the MREb core sequence, -137/-55C, and MREb to the MREa core sequence, -137/-96T (Fig. 2). If this nucleotide difference was responsible for the observed basal activity of MREa, then -55C should reduce the basal activity of MREa, and therefore the promoter, whereas -96T should increase the basal contributions of MREb. These constructs served to show that the N position of the MRE core sequence is an important determinant of MRE contributions to basal activity. When both MREs have the sequence of MREa, the basal activity of the promoter is increased, whereas the opposite is true when both MREs have the MREb sequence. This is consistent with the previous results of the MRE inactivating mutations, which suggested that MREa is an important contributor to basal level expression. However, the mutation of MREb to an MREa sequence (Fig. 2) could not confer independent activity on MREb, because the -96T mutation did not completely rescue the basal activity of the -60A mutant (43).

Therefore, the basal character of tMREa is partially due to a nucleotide difference of A instead of C at the variable position of the core sequence,

TGC(G/A)CNC. Searle *et al.* (45) also have observed a substantial increase in basal activity of a synthetic MRE dimer when A is substituted for C at this position. An A rather than G nucleotide at the fourth position of the TGC(G/A)CNC core, as in tMREa, also raises basal activity of a synthetic MRE sequence (45) as well as increasing binding of a zinc-dependent factor from rat liver nuclear extracts (55). Therefore, the functional effects of an A at these two variable positions of the MRE core may be a reflection of increased affinity of MRE binding factor(s), and the tMT-B MREa could be an "ideal" core sequence for factor binding. If only a low concentration of the factor is functional in the cell, binding would be limited to high-affinity MREs, such as the tMREa sequence, for basal activity. In the presence of metal, higher concentrations of MRE binding factor could be available to bind lower affinity sites through cooperative interactions with the constitutively bound MRE(s), eliciting a metal response that is significantly higher than basal expression. There is sufficient variability among MRE sequences in one MT promoter, both within the core as well as flanking sequences, to allow two levels of MRE-dependent transcription (37). The tMREa sequence is the usual MRE variant of the piscine MT genes from stone loach, pike, and rainbow trout (23, 56, 57). As well, the consensus MREs of the *Drosophila* MT promoter (58) and sea urchin (*Strongylocentrus purpuratus*) MT-A promoter have the tMREa sequence (59). However, the tMREa sequence is rare in mammalian MT promoters, and of the numerous MRE sequences we have examined, the tMREa sequence is present as mouse MT-I MREe, mouse MT-II MREc, rat MT-I MREf, sheep MT-I MREf, and human MT-IF MREc (37). In most of these cases, the MRE is located a great distance from the TATA box and/or the 3' flanking G nucleotide is mutated, which decreases MRE activity (45). The one apparent exception is the hMT-IIA promoter, which displays high basal activity and has three such consensus tMREa sequences (MREd, f, and h) (37). Interestingly, the activity of the proximal basal level enhancer of the human MT-IIA is abolished without TGCACAC sequence in opposite orientation at its 3' boundary (60), suggesting that this MRE sequence is required for hMT-IIA basal activity. This same tMREa core sequence is also found in the distal basal level enhancer (60). Other human MT promoters with less significant basal level activity compared to hMT-IIA may have less high-affinity sites because their numerous MREs, in cooperation with other non-MRE elements, would result in high basal activity and a loss of metal regulation.

c. Functional Interactions of Distal MREs with Proximal MREs in the Trout MT-B Promoter. In addition to the two proximal MREs of the tMT-B promoter, two MRE sequences are present at a distance of greater than 500

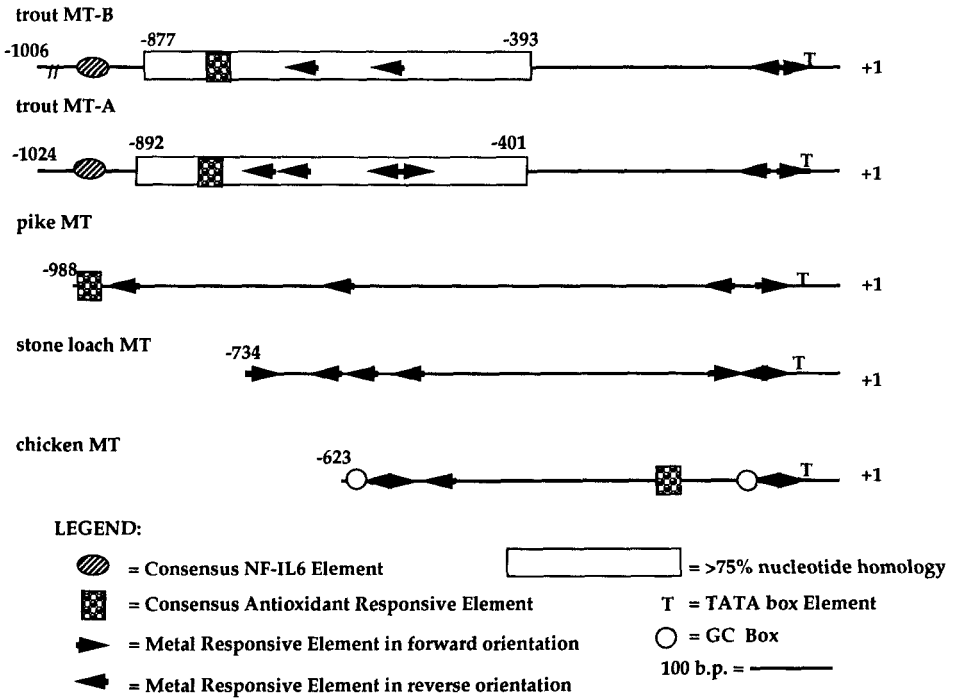


FIG. 4. Schematic representations of metallothionein 5' regulatory regions exhibiting bipartite structure. The rainbow trout MT-B, rainbow trout MT-A (57), pike MT, stone loach MT (56), and chicken MT (61) 5' regulatory regions are diagrammed with putative cis-acting elements identified.

bp from the TATA box (Fig. 4): MREc (-571 to -577) and MREd (-676 to 682). Both MREc and MREd are in reverse orientation and contain the same core sequence as MREa (TGCACAC). MREc is positioned in the center of a jumble of four "near" MREs, all of which differ from the MRE core consensus by only one nucleotide. An additional near MRE is present at -726 to -731.

Based on mutagenesis, the proximal promoter region of the tMT-B gene was dependent on the activity of MREa for efficient basal and metal-induced activity. If the 5' boundary of tested tMT-B regulatory sequence is extended to include these additional MREs, the general promoter activity, but not metal fold induction, is increased, indicating that elements in the distal region contribute to overall promoter activity. If the proximal MRE sequences are substituted, there is little activity above background (60a). These results

demonstrate that the contributions of distal elements are dependent on a functional interaction with the proximal MREs. Because the addition of the distal region did not increase metal fold induction compared to -137 , and no metal induction was observed without the proximal MREs, it was possible that the distal MRE sequences were nonfunctional and other cis-acting elements were responsible for the increased activity of the -738 fragment. However, the distal MREs are active on their own for basal and metal-induced activity when fused in close proximity to the TATA box.

The distal region from -393 to -877 displays greater than 75% nucleotide sequence homology with the comparable region of tMT-A, including conservation of the MT-B MRE_c and MRE_d sequences, whereas sequences between the proximal promoters and conserved distal regions cannot be aligned (S. Schieman *et al.*, unpublished data). Therefore, the conservation between the two promoters is limited to the regions containing MREs, arguing that these sequences are of particular importance to the regulation of the trout MT genes. Multiple distal (>500 bp) consensus MREs are also found in the pike and stone loach promoters (56, 57) as well as the chicken MT promoter (61). This bipartite organization appears to parallel the presence of simple proximal MT promoter with only two MREs (three with stone loach MT), whereas the mammalian MT promoters, with numerous proximal MREs, do not generally contain distal MRE sequences (compare Fig. 1 with Fig. 4) (37). It is not difficult to imagine that the distal MREs are brought into close proximity to the promoter through physical interactions between distal and proximal bound MRE binding proteins. This would require DNA looping, as has been proposed numerous times for the action of distal enhancers on proximal promoters (62–64). Basal MT levels display extensive variation in trout tissues (27) and although cell-type-specific levels of MRE binding activity could be responsible for these differences, it is also tempting to speculate that distal MRE contributions to basal MT activity could be modulated in a tissue-specific manner by events removed from the proximal promoter, such as chromatin structure or DNA methylation.

C. The Trout MT-B Promoter: Response to Oxidative Stress

In addition to the distal MRE sequences, the 5' flanking region of the tMT-B gene contains a consensus antioxidant-responsive element, or ARE (Fig. 4) (65). AREs, also called electrophile-responsive elements, mediate transcription induction by hydrogen peroxide as well as redox-active quinones and so called phenolic antioxidants (65, 66). A recent examination of MT promoters from vertebrates and invertebrates reveals that the ARE consensus, TGACNNNGC, is almost as common among MT promoters as

MREs, often overlapping previously defined AP1 binding sites or, in mouse and hamster MT-I, major late transcription factor sites (41). The common presence of AREs in MT promoters, combined with the observations that MT provides protection from oxidative stress (12, 67, 68) and can quench hydroxyl radicals (13), suggests that these elements are conserved for a reason.

The distal tMT-B regulatory region also contains a consensus ARE, which is emerging as a common element among MT promoters at both proximal and distal locales (see Figs. 1 and 4) (41). However, in transfection analysis, a tMT-B regulatory fragment containing this element (-3000/+8) had no more significant oxidant induction than the -738 5' deletion, suggesting that it is not required for the response of the MT-B regulatory region in transfected cells. In preliminary experiments using hydrogen peroxide, the -738/+8 region responded but -137/+8 did not (60a). This led us to hypothesize that MREs were not mediating the response to oxidative stress agents. Instead, the possible candidates were the four ARE half-sites (TGACNNN) present in the tMT-B promoter distal region. These mutant AREs are prevalent in the regulatory regions of fish MT promoters: tMT-A has 10 (57) and pike and stone loach have 7 each (56). One possibility is that the multiplicity of ARE-like sequences overcomes the lack of full consensus, although synthetic AREs appear to work additively rather than cooperatively (69). Another possibility is that MREs and the "near" AREs functionally interact. Dalton and co-workers (41) have identified an ARE consensus present in the mouse MT-I promoter 70 bp downstream of the potent MREd. Deletion of both elements, not just the ARE alone, is required to abrogate hydrogen peroxide responsiveness (41). Concatemers of MREd are also induced by hydrogen peroxide, suggesting that MREs may also respond to oxidants. In experiments that we have performed, a small but consistent response was observed for the proximal tMT-B promoter to *tert*-butyl hydroperoxide, so MREa and MREb could be involved in the response to oxidative stress. However, induction is increased substantially with the addition of tMT-B sequences from -137 to -738, suggesting that if the proximal MREs mediate oxidant induction, then the distal MREs (or near AREs) must be required to amplify the response (60a).

The possibility that MREs are involved in a positive response to oxidative stress is perplexing in that most MRE binding proteins appear to require metal in some capacity for DNA binding activity (see Table I). The trout, mouse, and human MRE binding proteins require zinc as assessed by chelation (43, 48) and mouse and human factor-MRE interactions are increased by *in vivo* zinc treatment (69a). Moreover, the primary sequences of mouse and human MTF MRE binding proteins both contain zinc finger motifs (70, 71). In transcription factors containing zinc, the common zinc-protein interaction is through the formation of metal-thiolate bonds with cysteine residues (54).

TABLE I
SUMMARY OF REPORTED MRE BINDING FACTORS

| Factor | Source | Detected by ^a | Molecular mass | MRE binding ^b | <i>In vitro</i> metal response | <i>In vivo</i> metal response | Ref. |
|---------|---------------------------------------|--------------------------|----------------------------------|--------------------------|--|---|---------|
| MEP-1 | Mouse L cell nuclei (CD) ^c | SW; ExoIII; DP | 108–110 kDa | MRE variants | Constitutive; EDTA sensitive; Zn ²⁺ restores but not Cd ²⁺ , Cu ²⁺ , Mn ²⁺ , or Hg ²⁺ | Cultured in 50 μM CdCl ₂ and 80 μM ZnCl ₂ ; or 5 μM CdCl ₂ and 50 μM ZnCl ₂ | (86–88) |
| MBF-1 | Mouse L cell nuclei | DP; UV | 74 kDa | mMREd/et, MREa/b | Zn ²⁺ for <i>in vitro</i> transcription | N.D. | (102) |
| | Mouse L cell nuclei | ExoIII; DP; EMSA | N.D. ^d | mMREd | Cd ²⁺ , not Zn ²⁺ or Cu ²⁺ | N.D. | (109) |
| | Rat Fao cell nuclei | EMSA; UV; not DP | 39 kDa | MRE dimer | Cd ²⁺ | Cd ²⁺ | (93) |
| ZAP | Rat liver nuclei | EMSA | N.D. | MRE variants | Zn ²⁺ not Cd ²⁺ | N.D. | (55) |
| MRE-BF1 | Human cell nuclei | EMSA; UV; SW | 86 kDa | hMT-I MREa | No response to Zn ²⁺ or Cd ²⁺ | Inactivation by Zn ²⁺ , Cd ²⁺ , Cu ²⁺ | (110) |
| MRE-BF2 | See MRE-BF1 | See MRE-BF1 | Two: 28 kDa (SW); 29/32 kDa (UV) | See MRE-BF1 | No response to Zn ²⁺ or Cd ²⁺ | Activated by Zn ²⁺ , Cd ²⁺ , and Cu ²⁺ | (110) |

| | | | | | | | |
|--------------------|--|------------------|--------------------------------------|--------------------------|---|--|----------|
| ZRF | HeLa cell nuclei | EMSA; DP; UV; SW | 116 kDa | 28 bp of hMT-IIA MREa | Zn ²⁺ but not Cd ²⁺ , Cu ²⁺ , or Hg ²⁺ | No response to Cd ²⁺ or Zn ²⁺ | (80, 84) |
| M96 cDNA | Mouse T cell cDNA library | EMSA | 45 kDa | hMT-IIA MRE 3/4 fragment | During bacterial expression: Zn ²⁺ , not Cd ²⁺ , Hg ²⁺ | N.D. | (111) |
| MRE-BP | HeLa cell nuclei | EMSA | 103 kDa | 14 bp of hMT-IIA MREa | Constitutive; inhibited at high Zn ²⁺ and Cd ²⁺ | N.D. | (84) |
| | HeLa and Cd ²⁺ HeLa cell nuclei | EMSA; UV | 118, 93.5, 85, 67, and 50.5 kDa (UV) | hMT-IIA MRE 3/4 | Constitutive; EDTA sensitive; restored by Zn ²⁺ , not Cd ²⁺ or Cu ²⁺ | Constitutive (EMSA); Cd ²⁺ increased 67/50.5 kDa (UV) | (112) |
| Human MTF-1 (cDNA) | HeLa cell nuclei; cDNA library | EMSA; MI | 753 a.a. (cDNA); 120 kDa (SW) | mMREd | Zn ²⁺ , not Cd ²⁺ | N.D. | (71, 85) |
| Mouse MTF-1 (cDNA) | cDNA expression library | EMSA; SW | 675 a.a. (cDNA); 100 kDa (SW) | mMREd; modified mMREd | Constitutive; EDTA sensitive | Zn ²⁺ | (70) |

^aAbbreviations: EMSA, electrophoretic mobility shift assays; DP, DNase I protection; SW, Southwestern analysis; UV, UV crosslinking; MI, methylation interference; ExoIII, exonuclease III protection.

^bMRE promoter sources: h, human, m, mouse MT-I; t, rainbow trout MT-B.

^cCd²⁺ indicates cell line selected for cadmium resistance.

^dN.D., Not determined.

There is some evidence that such sulfhydryl-dependent DNA binding domains are inactivated by oxidation *in vitro* and *in vivo* (72). However, it is also plausible that the oxidative release of zinc from cellular proteins (73, 74) could result in an elevated concentration of free zinc in the cell, which could increase zinc-dependent MRE binding activity. In any case, further analysis of the tMT-B 5'-regulatory region is merited in order to narrow the identity of sequences that are contributing to oxidant induction.

IV. MRE-Binding trans-Acting Factors

A. General Background

With the MRE consensus defined, the next logical step toward understanding metal regulation is the identification and characterization of the factors that bind to MREs. Studies employing nuclear runoff assays have demonstrated that MT gene regulation occurs mainly at the level of transcription (75). The rapid induction of MT transcription by metals occurs even if protein synthesis is inhibited by cyclohexamide, indicating that the factors that mediate the response already exist in the cell (75, 76). Competition experiments in transiently transfected cells have established that the increase in MT transcription in response to metals is mediated by a titratable, positively acting factor that interacts with MREs (30, 78, 79). In confirmation, *in vivo* footprinting of the rat MT-I (40) and mouse MT-I (50) promoters has demonstrated that metal treatment of the cells causes the occupation of MRE elements. Thus, the search for the MRE-binding metalloregulatory factor was on. Unfortunately, with numerous investigators attempting to identify, purify, and clone MRE binding proteins, the result has been a confusing array of factors with apparently different MRE sequence requirements and metal responses *in vitro* or *in vivo*. In considering all of these factors, proposed regulatory mechanisms by different laboratories are rife with contradictions. In addition, methodological oversights, such as including the chelator EDTA during cell extract preparations, cause difficulties in interpreting the level of MRE binding factor activity in the cell under basal or metal-induced conditions. A comprehensive list of reported MRE binding factors is given in Table I; a review, which includes some of these factors, has been presented elsewhere (81).

By far, the most significant recent advance in understanding MRE-directed transcription has been the isolation of cDNAs encoding mouse and human MTF-1 (70, 71). The mouse cDNA was isolated from an expression library using a modified mMREd sequence as a probe (70). Subsequently,

hybridization with the mouse MTF cDNA enabled the isolation of fragments of the human MTF-1 cDNA, which was gap-filled with reverse transcriptase polymerase chain reaction (PCR) of human hepatocarcinoma RNA (71). The primary sequences of the mouse and human homologs share 93% amino acid identity (71). However, recombinant mouse MTF activates MRE-directed transcription when cotransfected in mouse cells without requiring zinc treatment (70), whereas the human MTF homolog apparently is zinc dependent (71). Although the results of coexpression experiments are not conclusive as to the role of mouse MTF in metal regulation, allelic disruption of the MTF gene in mouse embryonic cells completely abolishes all basal and metal-induced transcription of the endogenous MT genes (51). Further, transcription directed by a transfected mouse MT-I promoter in this knock-out cell line is dependent on coexpression with recombinant mouse MTF (51). Stable transfected antisense MTF also prevents MRE-directed transcription activation by a series of metals (77), confirming the importance of MTF in MRE activation by all metals.

In comparing the characteristics of MTF and previously reported MRE binding factors, mouse L MEP-1 appears to be equivalent to mouse MTF, whereas ZAP may be the rat MTF homolog. Also, ZRF is now known to be human MTF, as determined by amino acid sequencing (80).

In mouse 3T6 cells, endogenous MTF MRE binding activity is detected without the addition of zinc (70), although whether it is activated further by *in vitro* zinc addition is unclear (51, 70). However, *in vivo* zinc treatment consistently increases the amount of MRE binding factor activity that is detected (70), even though the level of MTF transcript accumulation is not affected by *in vivo* Zn²⁺ (70) or by Cd²⁺, Cu²⁺, Ni²⁺, or Pb²⁺ (51). Unfortunately, the effects of metals other than zinc for MTF MRE binding activation *in vitro* or *in vivo* have not been reported.

Because the coexpressed mouse MTF is apparently zinc independent, and endogenous MTF activity is increased by *in vivo* zinc, one explanation is that MTF is not the metal sensor, but it is regulated by a regulatory inhibitor that releases MTF in the presence of metal (70). In apparent support of this hypothesis, Palmiter determined that a cell variant with an abnormally high level of MRE-directed expression could be reverted by fusion to a normal cell type. The interpretation of these results was that the regulatory inhibitor was inactive in the mutant cell line, causing abnormally high amounts of MTF activity (77). Interestingly, when the "deregulated" zinc-sensitive cell type was used to screen genetically for rat cDNAs that could rescue the phenotype, the majority of isolated sequences encoded a transmembrane protein that appears to regulate zinc efflux from the cell (82). Thus, the MTF inhibitor remains elusive.

B. *In Vitro* and *In Vivo* Activation of Human and Mouse MTF

We have assayed the functional contributions of the MRE sequences to MT promoter activity and have observed that increased MT expression with *in vivo* zinc treatment is paralleled with increased MRE-specific binding activity in nuclear extracts. It was important to establish if MRE binding activity was accessible to zinc *in vitro* and how other metal ions, which are known to activate MRE activity, affect factor binding.

Using Southwestern analysis, we have consistently identified a 125-kDa nuclear protein from HepG2 cells that bound to the human MRE α and trout MRE α sequences (43). Both of these sequences also bind to an *in vivo* zinc-induced factor in electrophoretic mobility shift assays (EMSAs). Throughout the course of many assays, the tMRE α sequence was found to detect the human MRE binding protein more consistently and this sequence was employed for most of our studies (69a).

The mouse and human factors detected in cell line nuclear extracts form MRE complexes (Figs. 5 and 6) that comigrate with *in vitro*-translated mouse MTF and purified human ZRF/MTF, respectively (69a). Further, the apparent molecular masses of mouse MTF (70) and human ZRF/MTF (71, 80) are comparable to the zinc-dependent mouse and human MRE binding factors detected previously in Southwestern assays (43, 48). Together, these observations suggest that the endogenous mammalian factors identified in our study are mouse and human MTF.

Although the MTF cDNAs were cloned during the course of our studies, little has been reported about the *in vivo* regulation of these factors by metals other than zinc. Cadmium strongly activates MT gene expression (1, 37) and induces MRE occupation by factors *in vivo* (40). However, we have observed that cadmium added during reactions cannot activate MRE binding of the endogenous human or mouse factor (Figs. 5 and 6) as well as purified human ZRF or *in vitro*-translated mouse MTF (69a). Other investigators have reported similar observations for HeLa cell ZRF/MTF and rat liver ZAP (55, 84, 85). In the mouse, the MRE binding activity of EDTA-chelated MEP-1, which is equivalent to mouse MTF in apparent molecular mass, is reactivated exclusively with zinc, even when studied in a cadmium-resistant cell line that has been further superinduced with cadmium treatment (86–88). The inability of cadmium to activate MTF binding activity *in vitro* suggests either that this metal cannot bind to the factor or that it binds but cannot substitute structurally for zinc in the DNA binding domain and still allow necessary base-specific contacts. The mouse and human cDNAs both contain a region with six consensus C₂H₂ zinc fingers, which is required for DNA binding (89). Sp1 also contains three such C₂H₂ zinc fingers (36), and yet cadmium

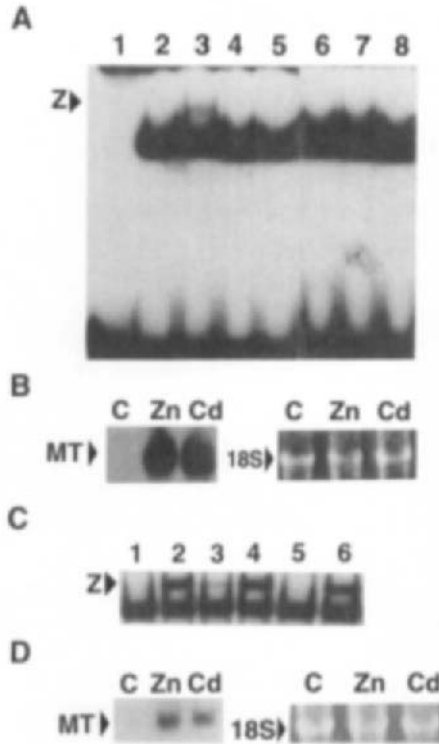


FIG. 5. The effect of metal on the interaction of human MRE binding factors with tMREa. (A) The HepG2 nuclear zinc-responsive complex with tMREa is indicated (z). The binding reactions contained tMREa (lanes 1 to 5) or the tMREa-60A mutant (lanes 6 to 8) and the following extracts: lane 1, no extract; lane 2, control extract; lanes 3 and 6, nuclear extract from cells treated with 100 μM ZnCl_2 for 4 hr; lanes 4 and 7, control extract with 10 μM CdCl_2 ; lanes 5 and 8, extract from cells treated with 10 μM CdCl_2 for 4 hr. (B) Northern blot analysis of total HepG2 RNA probed with the human MT-II-processed pseudogene. The hybridized MT message is indicated (MT). The RNA was isolated in parallel with nuclear extracts from untreated cells (C), cells treated with 100 μM ZnCl_2 for 4 hr (Zn), or cells treated with 10 μM CdCl_2 for 4 hr (Cd). (C) The HeLa nuclear zinc-responsive complex with tMREa is indicated (z). The binding reactions contained no added metal (lanes 1, 3, and 5) or 100 μM ZnCl_2 (lanes 2, 4, and 6) as well as the following extracts: lanes 1 and 2, control extract; lanes 3 and 4, extract from cells treated with 100 μM ZnCl_2 for 4 hr; lanes 5 and 6, extract from cells treated with 10 μM CdCl_2 for 4 hr. (D) Northern blot analysis of total HeLa cell RNA probed with the human MT-II-processed pseudogene. The hybridized MT message is indicated (MT). The RNA was isolated in parallel with nuclear extracts from untreated cells (C), cells treated with 100 μM ZnCl_2 for 4 hr (Zn), or cells treated with 10 μM CdCl_2 for 4 hr (Cd).

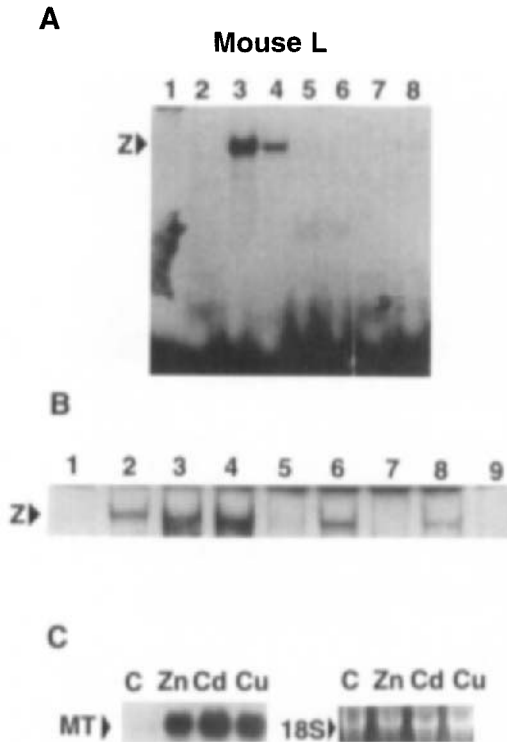


FIG. 6. The effect of *in vivo* metal treatment on the formation of the tMREa complex with mouse L cell factors. The mouse L cells were untreated (control) or were treated with 100 μM ZnCl₂, 10 μM CdCl₂, or 400 μM CuCl₂ in the media for 4 hr. (A) The zinc-responsive complex with tMREa is indicated (z). The binding reactions contained the tMREa probe (lanes 1 to 6) or the tMREa/-60A mutant probe (lanes 7 to 8) and the following nuclear extracts: lane 1, no extract; lanes 2 and 7, control extract; lanes 3 and 8, extract from zinc-treated cells; lane 4, control extract with 100 μM ZnCl₂; lane 5, extract from cadmium-treated cells; lane 6, control extract with 10 μM CdCl₂. (B) Each reaction contained either no metal (lanes 1, 3, 5, and 7) or 100 μM ZnCl₂ (lanes 2, 4, 6, 8, and 9) with the tMREa probe (lanes 1 to 8) or the tMREa/-60A mutant probe (lane 9). The nuclear extracts added to each binding reaction were from lanes 1 and 2, control cells; lanes 3, 4, and 9, zinc-treated cells; lanes 5 and 6, cadmium-treated cells; lanes 7 and 8, copper-treated cells. (C) Northern blot analysis of total RNA from control mouse L cells (C) or mouse L cells treated with zinc (Zn), cadmium (Cd), or copper (Cu). The human MT-II-processed pseudogene was used as a probe.

ions readily substitute for zinc within a three-finger peptide that still retains its ability to bind specifically to the Sp1 recognition sequence (90). Two other zinc finger types, found in GAL4 (C₆) and in the glucocorticoid receptor (C₂C₂), or "zinc twist", also can be constituted with cadmium while main-

taining specific DNA sequence recognition (91, 92). In addition, the molar equivalent of cadmium required to reconstitute DNA binding activity is similar to zinc (92). With these studies in mind, it is possible that the isolated zinc fingers of MTF potentially could be folded around cadmium and could bind MREs, but other protein domains are negatively affected by cadmium and prevent the interaction.

If MTF is unresponsive to cadmium *in vitro*, other factors could be responsible for cadmium induction of MRE activity. Although not detected by our assays, it is possible that MRE-protein interactions of cadmium-responsive factors are unstable and detected only under specific conditions (93). Alternatively, MTF may be the only factor responsible for metal induction of MRE activity. This is partly supported by the observation that allelic disruption of MTF in mouse embryonic cells abolishes basal MT expression and does not allow endogenous MT genes to respond to any metal tested (51). However, in this null MTF cell line, transfected MT promoter activity was not tested (or reported) with the various metal treatments explored for the endogenous MT genes, and the rescue of MT promoter activity with expressed MTF is presented only in response to zinc (51). If MTF is the sole metal regulatory factor, then the limitation of MTF binding activation to zinc *in vitro* suggests that other metals indirectly activate this factor *in vivo* by displacing zinc from subcellular pools, which then directly activates MTF (77). In testing this hypothesis, we found that *in vivo* cadmium treatment does not cause the formation of the zinc-responsive HepG2, HeLa, mouse L, and mouse Hepa-I cell MRE complexes (Figs. 5 and 6). Also, *in vivo* treatment of human and mouse cells with copper does not activate MTF binding to form MRE complexes (Fig. 6) (S. Samson and L. Gedamu, unpublished data). This result is difficult to reconcile with the idea that MTF is the only factor involved in metal regulation. Also, the human MT genes are differentially regulated by metals in HepG2 cells (94), which is also at odds with the concept of a single factor. For example, zinc, cadmium, and copper induce nearly equivalent levels of human MT-IIA transcript accumulation (94). However, the MT-IF gene responds poorly to copper whereas the MT-IG gene responds poorly to cadmium (94). If only one type of promoter element, MREs, and a single MRE binding factor is responsible for induction by these three metals, then each metal should have the same order of potency for inducing every MT gene within one cell type. Although the existence of auxiliary metal regulatory factors is tenuous at the moment, they cannot be ruled out completely, based on the current literature on MRE binding factors. MTF may be essential for all MT promoter activity (51), but auxiliary factors (i.e., cadmium responsive) could be present in the cell, which would modulate the transcription response depending on the metal inducer. Such factors might require interaction with a basal level of active MTF to affect MT transcription.

Through all of these interpretations, however, we have assumed that the level of endogenous MTF MRE-binding activity that is observed here with *in vivo* zinc treatment is the level required for MT gene activation. Perhaps the factor is activated indirectly by cadmium treatment *in vivo*, but the amount of factor that is required for cadmium induction is below the threshold detected by our assays.

C. Cotransfection Studies on Human and Mouse MTF cDNAs

1. IN MOUSE AND HUMAN CELL LINES

In our hands, mammalian MRE binding factors in untreated cell nuclear extracts are always responsive to zinc *in vitro*, arguing against constitutive activation of MRE binding factors. However, in cotransfection experiments in mouse and human cell lines, expression of human or mouse MTF is sufficient for MT promoter activation without zinc treatment (Fig. 7). One possible explanation for the discrepancy between the zinc dependence of the endogenous factor in mammalian cells and the near constitutive activity of expressed MTF is that, at artificially high concentrations, it titrates a limiting metal regulatory component, such as an inhibitor of MTF (70, 77) analogous to the inhibition of NF- κ B by I- κ B (95, 96) or the glucocorticoid receptor by heat-shock protein 90 (97). However, we are not completely comfortable that this is the sole mechanism of regulation of MTF. In the above examples, the transcription factors are cytoplasmically anchored by the inhibitory protein (98). Because we have exclusively employed nuclear extracts, we are unable to comment on whether a portion of MTF is localized to the cytoplasm. However, in our experiments, the endogenous MTF is always present in the nucleus and most of it is in an inactive form. Therefore it seems unlikely that association with metal ions is required for dissociation of an inhibitor and transport to the nucleus (77). Further, the inactive nuclear form of MTF is easily activated with *in vivo* or *in vitro* zinc, which can be reversed by metal chelation so that the DNA-binding activity of nuclear MTF appears to be solely regulated by association with zinc. However, this does not preclude that other events may regulate its transcription activity. Both mouse and human MTFs have a region rich in serine and threonine residues, possible candidates for phosphorylation, and deletion of this region devastates activity (89). Also, the unique 78-amino acid extension of human MTF is especially rich in these residues and this region may make important contributions to transcription activity (89; S. Samson and L. Gedamu, unpublished data). In this case, it is also possible that a "titrated" component could be an enzyme, such as a kinase or a phosphatase, which is affected by metals, and, with expression of

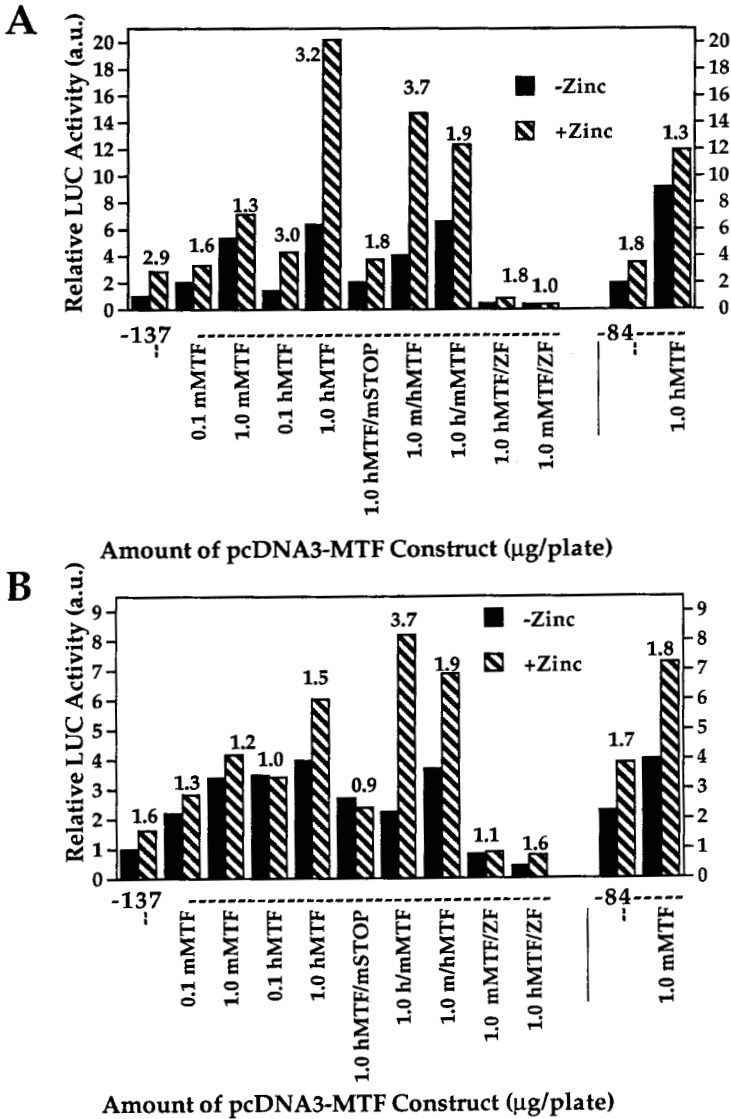


FIG. 7. Coexpression of mouse and human MTF expression constructs with the tMT-B promoter in HepG2 and mouse L cells. HepG2 (A) and mouse L (B) cells were transfected with the indicated amounts (micrograms) of MTF-pcDNA3 construct per 6-cm plate in concert with the trout MT-B proximal promoter (-137/+8) in pMEV35R. LUC activity is expressed as arbitrary units (a.u.) relative to the basal activity of the tMT-B promoter (= 1.0). Transfected cells were either untreated (-Zinc) or treated with 100 μ M ZnCl₂ (+Zinc) and the zinc fold inductions are given above each zinc activity bar.

the factor in the cell, a large portion of the exogenous factor is not appropriately modified to mediate higher transcription with metal treatment.

Perhaps the most realistic explanation for the constitutive activity of expressed MTF is that the ratio of active MTF to inactive MTF is dependent on the affinity of the DNA binding domain for zinc and, therefore, the free zinc concentration in the cell. With expression of additional factor, the absolute level of active MTF in the cell increases but the ratio of active to inactive factor is not altered substantially. In this case, zinc induction is not observed with expression of high amounts of exogenous factor because the transfection system is already saturated with active factor without zinc treatment. In considering the constitutive activity of mouse MTF, Schaffner and co-workers have supported an argument of zinc concentration dependence with their past observations that MTF requires higher zinc concentration for activation and a lower chelator concentration for inactivation compared to Sp1 (51, 70, 85).

One experiment is significant in its absence from the report of allelic disruption of MTF in mouse cells: the researchers demonstrate that both basal and metal inductions of the endogenous MT genes are completely dependent on MTF, but then proceed to present only that expressed MTF rescues the activity of a transfected MT promoter (51). In our work, expression of exogenous human MTF does not increase the basal level of the endogenous MT genes in HepG2 cells, although it increases transfected MT promoter activity. Therefore, other events must be required for the endogenous MT gene promoters to become accessible to the expressed factor. The lack of response of endogenous genes to exogenous factor has been mentioned in a discussion of gene locus control regions (LCRs) (99). Palmiter and co-workers have reported that large regions flanking the mouse MT-I and MT-II genes function as LCRs analogous to those that direct tissue-appropriate expression of globin and adenosine deaminase genes (63, 100). Functional MT genes are present as linked multicopy families (21, 22, 44, 101), so MT LCRs should also be at work in other species. LCRs are notorious for a concentration of transcription factor binding domains (63) and it is not difficult to imagine that MREs could be present in an MT LCR. Thus, the abrogation of mouse MT-I and MT-II gene expression with allelic disruption of MTF could be due to MT gene silencing at the level of the LCR rather than the promoter cis-acting elements. Perhaps under control conditions, expressed MTF is able to activate transfected MT promoters, but requires additional metal-regulated modifications to become competent to affect the endogenous MT genes from an enhancer position in a LCR. In this sense, a requirement of MTF for MT gene induction by all metals does not necessarily mean that MTF is the only metal regulatory factor, but instead suggests that it is indispensable for action by additional auxiliary metal regulatory factors. In the future, we will be convinced that metal regulation is exclusive to MTF should it be demon-

strated that the transcription activity of expressed MTF in the null MTF mouse cell line is increased by all metals, although such metals cannot activate MRE-directed transcription alone. Even so, this still would not completely rule out the possibility of auxiliary factors that are dependent on MTF for their contributions to transcription.

2. IN A RAINBOW TROUT CELL LINE

The basic MRE core consensus TGC(G/A)CNC is highly conserved, even among highly divergent species of invertebrates and vertebrates (58, 59, 102), suggesting that some degree of conservation should exist among MRE binding factors. In addition, the basic transcription apparatus is fairly tolerant of interchanges of basic transcription factors and positive activators among different species (103, 104). To determine if mouse and human MTFs could increase MT promoter activity in a nonmammalian system, the cDNAs were coexpressed with the tMT-B promoter in RTH-149 cells. In three trials with the -137/+8 metal-responsive fragment, both mammalian MTF homologs were able to increase reporter gene activity (Fig. 8). Cotransfection of the mMTF-pcDNA3 construct increased the basal level of -137 promoter activity 70- to 80-fold. Zinc-induced levels were also increased but were lower than the corresponding activities from non-metal-treated cells. Expression of the human factor also increased promoter activity, although not to the level observed with the mouse factor (Fig. 8). Assuming that the factors are produced equivalently, this suggests that mMTF is better equipped for transcriptional activation in the trout system whereas hMTF appears to have greater activity in mammalian cell lines (Fig. 7). With high amounts of hMTF expression construct (5 μ g), the zinc-induced -137-LUC activity is reduced compared to untreated cells, as was observed at the lower amounts of mMTF per plate (Fig. 8). The observation of decreased MT promoter activity with zinc treatment is consistent with the phenomenon of "squenching," whereby an unusually high concentration of transcription activator titrates indispensable transcription components (62, 105, 106). This usually has been described as a nonspecific event for the *in vitro* transcription activity of a promoter that does not contain the factor binding site. However, self-squenching has been described as well (106). If the zinc in the cell is limiting, it is possible that a basal level of MTF is activating transcription to the maximum level that can be handled by our experimental design, and specifically the amount of promoter used. The treatment of the cells with zinc then activates the extra apo factor, allowing it to interact with transcription components. However, because the amount of promoter is limiting, the excess factor cannot form productive complexes with the template and expression is reduced rather than increased. The squenching phenomenon has been described for strong acidic activation domains, such as is found in the viral activator VP-

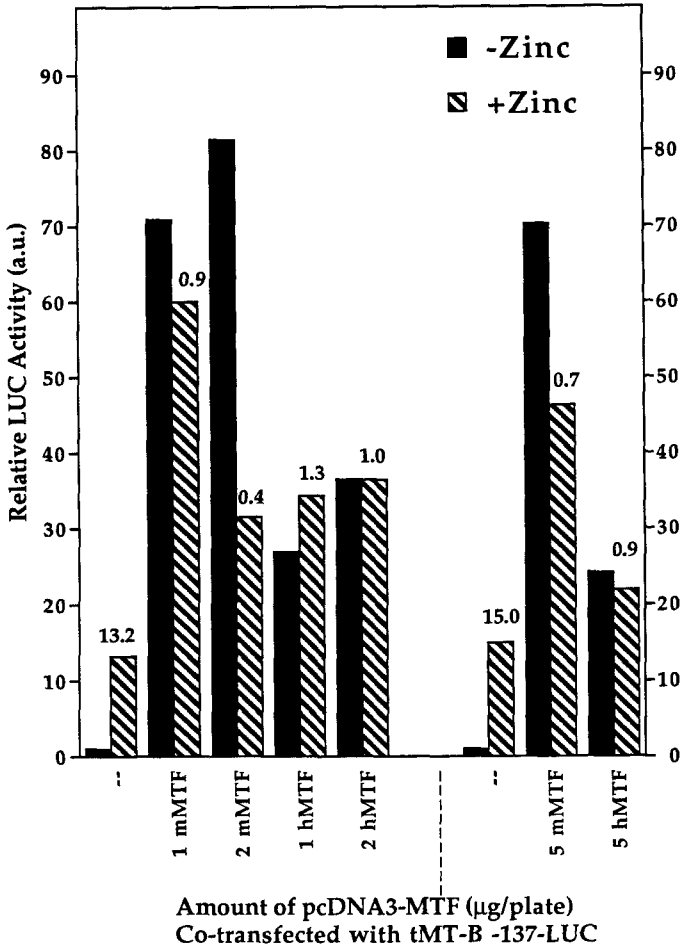


FIG. 8. Coexpression of mouse and human MTFs with the tMT-B proximal promoter in RTH-149 cells. RTH-149 cells were cotransfected with the indicated amounts (micrograms) of mMTF-pcDNA3 or hMTF-pcDNA3 with -137-LUC. The LUC activities are expressed as arbitrary units (a.u.) relative to the basal level of -137 (= 1.0) without coexpressed MTF (dashed line). Transfected cells were either untreated (-Zinc) or were treated with 100 μ M ZnCl₂ (+Zinc) and the zinc fold inductions (+Zinc/-Zinc) are given above each zinc activity bar.

16 (62). Radtke *et al.* (89) have recently shown that the acidic domain of mouse MTF is two-thirds as potent as that of VP-16 when fused to a GAL4 DNA binding domain and expressed in HeLa cells, so it is reasonable that it should be capable of squelching.

V. Conclusions and Suggestions for Further Research

Although the metal-responsive element consensus was defined more than a decade ago, advances in our understanding of metal regulatory mechanisms have been modest compared to the accumulation of data for other transcription regulatory pathways. Our study has focused on the functional and physical interactions of MRE binding proteins with metallothionein promoters. In the context of native promoter sequences, MREs contribute to basal and metal-induced promoter activity and functionally interact in a synergistic manner to elicit a significant response to zinc. In general, the zinc-induced activity of MT promoters can be correlated with an increase in the MRE binding activity of zinc-dependent mammalian nuclear factors, the characteristics of which are consistent with the cloned zinc finger protein MTF. Further analysis of the DNA-binding regulation of MRE binding factors by potent MT metal inducers has raised questions about whether mammalian MTF mediates MT transcription induction by metals other than zinc. The available MTF cDNAs and the null MTF cell line will be important experimental tools for answering such questions and it is hoped that they will fuel more rapid advances in the field of metal regulation. Future investigations should attempt to clarify the regulatory mechanisms employed by all important MT metal inducers. Further, elucidation of the mechanisms that regulate the transcription activity of MTF, in addition to its DNA binding activity, will be essential for understanding MT gene regulation and also will contribute to the current repertoire of known transcription factor regulatory pathways. Finally, the possibility of MRE involvement in MT gene activation by oxidative stress also provides an interesting problem for future studies.

ACKNOWLEDGMENTS

Studies from this laboratory were supported by funding from the Medical Research Council and Natural Sciences and Engineering Research Council of Canada to L. Gedamu. The Alberta Heritage Foundation for Medical Research has kindly provided studentship for S. L.-A. Samson. The contributions by W. Paramchuk are particularly appreciated. We thank Steven Schieman for constructive discussions. We also thank Dr. Walter Schaffner for the generous gift of the mouse and human MTFs. Finally, we are grateful to Linda Marin for her time and effort in typing.

REFERENCES

1. J. H. R. Kagi and A. Schaffner, *Biochemistry* **27**, 8509 (1988).
2. D. J. Thiele, *Nucleic Acids Res.* **20**, 1183 (1992).
3. L. Gedamu, R. Foster, N. Jahroudi, S. Samson, N. Shworak, and M. Zafarullah, in "Metal-

- lothionein III" (K. T. Suzuki, N. Imura, and M. Kimura, eds.), p. 363. Birkhauser Verlag, Basel, 1993.
4. G. K. Andrews, M. T. McMaster, S. K. De, B. C. Paria, and S. K. Dey, in "Metallothionein III" (K. T. Suzuki, N. Imura, and M. Kimura, eds.), p. 351. Birkhauser Verlag, Basel, 1993.
 5. G. K. Andrews, E. D. Adamson, and L. Gedamu, *Dev. Biol.* **103**, 294 (1984).
 6. N. Jahroudi, R. Foster, J. Price-Haughey, G. Beitel, and L. Gedamu, *J. Biol. Chem.* **265**, 6505 (1990).
 7. J. Price-Haughey, K. Bonham, and L. Gedamu, *Biochim. Biophys. Acta* **908**, 158 (1987).
 8. B. A. Masters, E. J. Kelly, C. J. Quaife, and R. L. Brinster, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 584 (1994).
 9. J. S. Lazo, Y. Kondo, D. Dellapiazza, A. E. Michalska, K. H. A. Choo, and B. R. Pitt, *J. Biol. Chem.* **270**, 5506 (1995).
 10. L. R. Beach and R. D. Palmiter, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2110 (1981).
 11. M. Sato, M. Sasaki, and H. Hojo, *Arch. Biochem. Biophys.* **316**, 738 (1995).
 12. K. T. Tamai, E. B. Gralla, L. M. Ellerby, J. S. Valentine, and D. J. Thiele, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8013 (1993).
 13. P. J. Thornally and M. Vasak, *Biochim. Biophys. Acta* **827**, 36 (1985).
 - 13a. J. H. R. Kagi, *Meth. Enzymol.* **205(B)**, 613 (1991).
 14. M. Karin and R. I. Richards, *Nature (London)* **299**, 797 (1982).
 15. C. J. Schmidt, M. F. Jubier, and D. H. Hamer, *J. Biol. Chem.* **260**, 7731 (1985).
 16. U. Varshney and L. Gedamu, *Gene* **31**, 135 (1984).
 17. R. D. Palmiter, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6333 (1992).
 18. R. I. Richards, A. Heguy, and M. Karin, *Cell* **37**, 263 (1984).
 19. A. Heguy, A. West, R. I. Richards, and M. Karin, *Mol. Cell. Biol.* **6**, 2149 (1986).
 20. U. Varshney, N. Jahroudi, R. Foster, and L. Gedamu, *Mol. Cell. Biol.* **6**, 26 (1986).
 21. M. Karin, R. L. Eddy, W. M. Henry, L. L. Haley, M. G. Byers, and T. B. Shows, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5494 (1984).
 22. C. J. Schmidt, D. H. Hamer, and O. W. McBride, *Science* **224**, 1104 (1984).
 23. M. Zafarullah, K. Bonham, and L. Gedamu, *Mol. Cell. Biol.* **8**, 4469 (1988).
 24. C. Sadhu and L. Gedamu, *J. Biol. Chem.* **263**, 2679 (1988).
 25. N. W. Shworak, T. O'Conner, N. C. W. Wong, and L. Gedamu, *J. Biol. Chem.* **268**, 24460 (1993).
 26. M. Zafarullah, P.-E. Olsson, and L. Gedamu, *Biochim. Biophys. Acta* **1049**, 318 (1990).
 27. M. Zafarullah, P.-E. Olsson, and L. Gedamu, *Gene* **83**, 85 (1989).
 28. K. Bonham, M. Zafarullah, and L. Gedamu, *DNA* **6**, 519 (1987).
 29. C. Sadhu and L. Gedamu, *Mol. Cell. Biol.* **9**, 5738 (1989).
 30. R. Foster and L. Gedamu, *J. Biol. Chem.* **266**, 9866 (1991).
 31. R. Foster, N. Jahroudi, U. Varshney, and L. Gedamu, *J. Biol. Chem.* **263**, 11528 (1988).
 32. D. H. Hamer, *Annu. Rev. Biochem.* **55**, 913 (1986).
 33. D. J. Thiele, *Nucleic Acids Res.* **20**, 1183 (1992).
 34. J. M. Hempe and R. J. Cousins, *J. Nutr.* **119**, 1179 (1989).
 35. M. Karin, R. D. Andersen, E. Slater, K. Smith, and H. R. Herschman, *Nature (London)* **286**, 295 (1980).
 36. J. T. Kadonaga, K. R. Carner, F. R. Masiarz, and R. Tjian, *Cell* **51**, 1079 (1987).
 37. J. Imbert, V. C. Culotta, P. F. Furst, L. Gedamu, and D. H. Hamer, *Adv. Inorg. Biochem.* **8**, 140 (1990).
 38. L. Gedamu, R. Foster, N. Jahroudi, S. Samson, N. Shworak, and M. Zafarullah, in "Metallothionein III" (K. T. Suzuki, N. Imura, and M. Kimura, eds.), p. 363. Birkhauser Verlag, Basel, 1993.
 39. R. L. Friedman, and G. R. Stark, *Nature (London)* **314**, 637 (1985).

40. R. D. Andersen, S. J. Taplitz, S. Wong, G. Bristol, B. Larkin, and H. Herschman, *Mol. Cell Biol.* **71**, 3754 (1987).
41. T. Dalton, R. D. Palmiter, and G. K. Andrews, *Nucleic Acids Res.* **22**, 5016 (1994).
42. T. H. Rushmore, M. R. Morton, and C. B. Pickett, *J. Biol. Chem.* **266**, 11632 (1991).
43. S. L.-A. Samson and L. Gedamu, *J. Biol. Chem.* **270**, 6864 (1995).
44. P. F. Searle, B. L. Davison, G. W. Stuart, T. M. Wilkie, G. Norstedt, and R. D. Palmiter, *Mol. Cell Biol.* **4**, 1221 (1984).
45. P. R. Searle, G. W. Stuart, and R. Palmiter, *Metallothionein II, Experientia (Basel)* **52** (suppl.), 407 (1987).
46. V. C. Culotta and D. H. Hamer, *Mol. Cell Biol.* **9**, 1376 (1989).
47. G. W. Stuart, P. F. Searle and R. D. Palmiter, *Nature (London)* **317**, 828 (1985).
48. S. L.-A. Samson, W. J. Paramchuk, N. W. Shworak, and L. Gedamu, *J. Biol. Chem.* **270**, 25194 (1995).
49. J. D. Parvin, R. J. McCormick, P. A. Sharp, and D. E. Fisher, *Nature (London)* **373**, 724 (1995).
50. P. R. Mueller, S. J. Salsler, and B. Wold, *Genes Dev.* **2**, 412 (1988).
51. R. Heuchel, F. Radtke, O. Georgiev, G. Stark, M. Aguet, and W. Schaffner, *EMBO J.* **13**, 2870 (1994).
52. P.-E., Olsson, S. J. Hyllner, M. Zafarullah, T. Andersson, and L. Gedamu, *Biochim. Biophys. Acta* **1049**, 78 (1990).
53. R. W. Carthew, L. A. Chodosh, and P. A. Sharp, *Genes Dev.* **1**, 973 (1987).
54. J. E. Coleman, *Annu. Rev. Biochem.* **61**, 897 (1992).
55. P. F. Searle, *Nucleic Acids Res.* **18**, 4683 (1990).
56. P. Kille, J. Kay, and G. E. Sweeney, *Biochim. Biophys. Acta* **1216**, 55 (1993).
57. P.-E. Olsson, P. Kling, L. J. Erkell, and P. Kille, *Eur. J. Biochem.* **230**, 344 (1995).
58. G. Maroni, E. Otto, and D. Lastowski-Perry, *Genetics* **112**, 493 (1986).
59. P. Harlow, E. Watkins, R. D. Thornton, and M. Nemer, *Mol. Cell Biol.* **9**, 5445 (1989).
60. M. Karin, A. Haslinger, A. Heguy, T. Dietlin, and T. Cooke, *Mol. Cell Biol.* **7**, 606 (1987).
- 60a. S. L.-A. Samson and L. Gedamu, in preparation (1997).
61. L. P. Fernando and G. K. Andrews, *Gene* **81**, 177 (1989).
62. M. Ptashne and A. A. F. Gann, *Nature (London)* **346**, 329 (1990).
63. G. Felsenfeld, *Nature (London)* **355**, 219 (1992).
64. R. Schleif, *Annu. Rev. Biochem.* **61**, 199 (1992).
65. T. H. Rushmore, M. R. Morton, and C. B. Pickett, *J. Biol. Chem.* **266**, 11632 (1991).
66. T. Prestora and P. Talalay, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8965 (1995).
67. M. A. Schwarz, J. S. Lazo, J. C. Yalowich, I. Reynolds, V. E. Kagan, V. Tyurin, Y.-M. Kim, S. C. Watkins, and B. R. Pitt, *J. Biol. Chem.* **269**, 15236 (1994).
68. J. S. Lazo, Y. Kondo, D. Dellapiazza, A. E. Michalska, K. H. A. Choo, and B. R. Pitt, *J. Biol. Chem.* **270**, 5506 (1995).
69. T. Nguyen, T. H. Rushmore, and C. B. Pickett, *J. Biol. Chem.* **269**, 13656 (1994).
- 69a. S. L.-A. Samson *et al.*, submitted (1997).
70. F. Radtke, R. Heuchel, O. Georgiev, M. Hergersberg, M. Gariglio, Z. Dembic, and W. Schaffner, *EMBO J.* **12**, 1355 (1993).
71. E. Brugnera, O. Georgiev, F. Radtke, F. Heuchel, E. Baker, G. R. Sutherland, and W. Schaffner, *Nucleic Acids Res.* **22**, 3167 (1994).
72. X. Wu, N. H. Bishopric, D. J. Discher, B. J. Murphy, and K. A. Webster, *Mol. Cell Biol.* **16**, 1035 (1996).
73. H. Fliss and M. Menard, *Arch. Biochem. Biophys.* **293**, 195 (1992).
74. W. Maret, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 237 (1994).

75. M. Karin, R. D. Andersen, E. Slater, K. Smith, and H. R. Herschman, *Nature (London)* **286**, 295 (1980).
76. M. Zafarullah, P.-E. Olsson, and L. Gedamu, *Biochim. Biophys. Acta* **1049**, 318 (1990).
77. R. D. Palmiter, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1219 (1994).
78. C. Seguin, B. K. Felber, A. D. Carter, and D. H. Hamer, *Nature (London)* **312**, 781 (1984).
79. H. Scholer, A. Haslinger, A. Heguy, H. Holtgreve, and M. Karin, *Science* **232**, 76 (1986).
80. F. Otsuka, A. Iwamatsu, K. Suzuki, M. Ohsawa, D. Hamer, and S. Koizumi, *J. Biol. Chem.* **269**, 23700 (1994).
81. S. Koizuma and F. Otsuka, in "Metallothionein III" (K. T. Suzuki, N. Imura, and M. Kimura, eds.), p. 457. Birkhauser Verlag, Basel, 1993.
82. R. D. Palmiter and S. D. Findley, *EMBO J.* **14**, 639 (1995).
83. S. Koizumi, K. Suzuki, and F. Otsuka, *J. Biol. Chem.* **267**, 18659 (1992).
84. S. Koizumi, H. Yamada, K. Suzuki, and F. Otsuka, *Eur. J. Biochem.* **210**, 555 (1992).
85. G. Westin and W. Schaffner, *EMBO J.* **7**, 3763 (1988).
86. C. Seguin, *Gene* **97**, 295 (1991).
87. S. Labbe, J. Prevost, P. Remondelli, A. Leone, and C. Seguin, *Nucleic Acids Res.* **19**, 4225 (1991).
88. S. Labbe, L. Larouche, D. Mailhout, and C. Seguin, *Nucleic Acids Res.* **21**, 1549 (1993).
89. F. Radtke, O. Georgiev, H.-P. Muller, E. Brugnera, and W. Schaffner, *Nucleic Acids Res.* **23**, 2277 (1995).
90. J. Kuwahara and J. E. Coleman, *Biochemistry* **29**, 8627 (1990).
91. L. P. Freedman, B. F. Luisi, Z. R. Korszun, R. Basavappa, P. B. Sigler, and K. Yamamoto, *Nature (London)* **334**, 543 (1988).
92. T. Pan and J. E. Coleman, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3145 (1989).
93. R. D. Andersen, S. J. Taplitz, A. M. Oberauer, K. L. Calame, and H. Herschman, *Nucleic Acids Res.* **18**, 6049 (1990).
94. C. Sadhu and L. Gedamu, *J. Biol. Chem.* **263**, 2679 (1988).
95. P. A. Baeuerle and D. Baltimore, *Science* **242**, 540 (1988).
96. P. A. Baeuerle and D. Baltimore, *Cell* **53**, 211 (1988).
97. M. Danielsen, in "Nuclear Hormone Receptors" (M. G. Parker, ed.), p. 63. Academic Press, New York, 1991.
98. T. Hunt, *Cell* **59**, 949 (1989).
99. R. D. Palmiter, E. P. Sandgren, D. M. Koeller, and R. L. Brinster, *Mol. Cell. Biol.* **13**, 5266 (1993).
100. B. J. Aronow, C. A. Ebert, M. T. Valerius, S. S. Potter, D. A. Wiginton, D. B. Witte, and J. J. Hutton, *Mol. Cell. Biol.* **15**, 1123 (1995).
101. M. G. Peterson, F. Hannan, and J. F. B. Mercer, *Eur. J. Biochem.* **174**, 417 (1988).
102. J. Imbert, M. Zafarullah, V. C. Culotta, L. Gedamu, and D. Hamer, *Mol. Cell. Biol.* **9**, 5315 (1989).
103. S. Buratowski, S. Hahn, P. A. Sharp, and L. Guarente, *Nature (London)* **334**, 77 (1988).
104. M. Meisterernst, M. Horikoshi, and R. C. Roeder, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9153 (1990).
105. S. L. Berger, W. D. Cress, A. Cress, S. J. Triezenberg, and L. Guarente, *Cell* **61**, 1199 (1990).
106. R. J. Kelleher III, P. M. Flanagan, and R. D. Kornberg, *Cell* **61**, 1209 (1990).
107. A. Haslinger and M. Karin, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8572 (1985).
108. N. W. Shworak, Ph.D. dissertation. Department of Biological Sciences, University of Calgary, Canada, 1990.
109. C. Seguin and D. H. Hamer, *Science* **235**, 1383 (1987).
110. M. Czupryn, W. E. Brown, and B. L. Valee, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10395 (1992).
111. C. Inouye, P. Remondelli, M. Karin, and S. Elledge, *DNA Cell Biol.* **13**, 731 (1994).
112. L. Minichiello, P. Remondelli, S. Cigliano, S. Bonatti, and A. Leone, *Gene* **143**, 289 (1994).

Transcriptional Regulation of the Steroid Receptor Genes

M. VIJAY KUMAR AND
DONALD J. TINDALL

*Departments of Urology and
Biochemistry/Molecular Biology
Mayo Foundation
Rochester, Minnesota 55905*

| | |
|---|-----|
| I. Structure of a Steroid Receptor Gene | 290 |
| A. DNA Binding Domain | 290 |
| B. Steroid Binding Domain | 291 |
| II. Molecular Mechanism of Transcription | 291 |
| A. Role of Enhancers in Transcription | 292 |
| B. Role of Suppressors/Repressors in Transcription | 292 |
| III. Regulation of the Androgen Receptor Gene | 293 |
| A. Isolation and Characterization of the 5' Flanking Region of the AR Gene | 294 |
| B. Positive Regulation of the AR Gene | 295 |
| C. Negative Regulation of the AR Gene | 295 |
| IV. Regulation of the Glucocorticoid Receptor Gene | 298 |
| A. Characterization of the 5' Flanking Region | 299 |
| B. Down-Regulation of the GR Gene | 300 |
| V. Regulation of the Progesterone Receptor Gene | 300 |
| A. Regulation of the PR Gene by Two Distinct Promoters | 301 |
| B. Estrogen Regulation of the PR Gene | 301 |
| VI. Regulation of the Estrogen Receptor Gene: Characterization of the 5' Flanking Region | 301 |
| VII. Concluding Remarks | 303 |
| References | 304 |

Steroid hormones, via their binding to specific receptors, are involved in the development, differentiation, and physiological response of cells to diverse stimuli. Activation by hormonal ligands induces conformational change in the receptor, enabling interaction with the target genes. The steroid receptor superfamily includes androgen, glucocorticoid, mineralocorticoid, progesterone, estrogen, thyroid, vitamin D, retinoic acid, and orphan receptors. This review will focus on the classic steroid receptors, i.e., the androgen, glucocorticoid, progesterone, and estrogen receptors, with emphasis on their transcriptional regulation. Readers are directed to several authoritative reviews for further details of steroid receptors (I-II). © 1998 Academic Press

I. Structure of a Steroid Receptor Gene

Elucidation of the structure of the steroid receptors was greatly advanced by the cloning of the steroid receptor cDNAs and genes. The steroid receptor genes are organized into three main functional domains: the trans-activation domain, the DNA binding domain (DBD), and the steroid binding domain (SBD). Though the size of the steroid receptors ranges from 427 amino acids (vitamin D₃ receptor; VDR) to 984 amino acids (mineralocorticoid receptor; MR), resulting in proteins ranging from about 50,000 to 110,000 Da, all the receptors demonstrate the typical domain structure. Because the sizes of the DBD and SBD are fairly similar in all the steroid receptors, the molecular size of the receptor is greatly dependent on the size of the trans-activation domain. The trans-activation domain ranges from about 24 amino acids (in the VDR) to 603 amino acids in the human MR. This region contains several polymeric stretches of amino acids, including polyglutamine, polyproline, and polyglutamic acid. The function of the polymeric amino acids is not known, although they have been implicated in trans-activation. The length of the polymeric sequences seem to be important, as seen in disease conditions such as X-linked spinal bulbar muscular atrophy and some cases of androgen insensitivity syndrome (9).

A. DNA Binding Domain

The DNA binding domain is located toward the carboxy terminus of the trans-activation domain. It consists of about 66–68 amino acids and is the most highly conserved region among the steroid receptor family. It comprises 20 invariant amino acids that form two zinc finger DNA binding motifs, providing the basis for the protein–DNA interactions in the regulator sequences of the target genes (5, 6, 8, 10, 11). Steroid receptors contain nine conserved cysteine residues, of which eight interact in a coordinated fashion to form two separate tetrahedral metal-binding complexes, or fingers. The four cysteine residues in each finger bind one Zn²⁺ molecule, which permits interaction with DNA. Recent experiments have indicated that in addition to the cysteines, amino acids adjacent to the cysteines are critical in defining the specificity of binding the cognitive response elements. Elegant mutational analysis of the three amino acids at the base of the first zinc finger demonstrated that replacement of the amino acids Glu, Gly, and Ala found in the estrogen receptor (ER) with Gly, Ser, and Val, as in the glucocorticoid receptor (GR), resulted in an ER that no longer recognized the estrogen response element (ERE), but instead recognized the glucocorticoid response element (GRE). The functional role of the second zinc finger is not clear, though it has been suggested that it aids in the stability of the binding of the protein to the DNA (10).

Following the DBD is the hinge region, which is involved in homodimer formation and contains sequences that are responsible for the nuclear localization of the steroid receptor. Earlier studies suggested that steroid receptors are localized in the cytoplasm and are transported to the nucleus after being bound by the ligand. However, recent studies indicate that the nuclear localization signal present in the hinge region aids in the internalization of the receptor even in the absence of binding by the ligand, whereas similar signals in the SBD of the receptor are responsible for ligand-mediated entry of the receptor into the nucleus (6, 8).

B. Steroid Binding Domain

The major functional domain at the carboxy terminus is the SBD. The size of the SBD varies from 224 amino acids (thyroid receptor; T₃R) to 254 amino acids (progesterone receptor; PR). The specificity of binding of glucocorticoid, mineralocorticoid, progesterone, and androgen is easily altered by mutations of a nucleotide leading to single amino acid change. The specific amino acids of GR and ER that are in contact with the ligand have been identified. However, experiments with antiestrogens, which bind to the same site as the estrogens, indicate that the contact points of the ligand may be less important than the protein structural alterations induced by binding of the ligand (3). Binding of the ligand to the receptor leads to a change in the conformation of the protein, exposing other functional domains. The activated receptor is then capable of binding to the cognate response element, resulting in trans-activation.

II. Molecular Mechanism of Transcription

The promoters that drive transcription of eukaryotic genes can be divided into two main groups: those possessing the consensus sequence TATAA (TATA box) and those not possessing the TATA box sequences. The TATA box is normally located 25–30 nucleotides upstream from the transcriptional start site that binds the transcription factor TFIID. TFIID is a multisubunit complex that is composed of the TATA-binding protein (TBP) and TBP-associated factors (TAFs). The binding of TFIID to the TATA box helps direct efficient assembly of RNA polymerase II along with the other transcription factors (TFIIA, TFIIB, TFIIE, TFIIF, TFIIIC) that are required for basal transcription initiation (12). ATP is one of the components required to melt the DNA at the transcription start site prior to transcription (13). The TATA box also helps position the start of initiation, usually from a single site.

Genes that lack TATA boxes are often called housekeeping genes. Their promoters, are usually GC rich, do not appear to bind TFIID directly, and

often have multiple sites of transcription initiation. An important component that is often found in both TATA-containing and TATA-negative promoters is the initiator sequence (PyPyCAPyPyPyPyPy consensus), which helps designate the transcription start site. The initiator can function as a minimal promoter for TATA-negative genes. It appears to be involved in transcription control because mutations in this region can cause a reduction of promoter strength and use of alternative start sites (14). The Sp1 (GC box binding protein), initiator, TFIID, and associated factors are required for transcription from some TATA-negative promoters (15). It is possible that TFIID may be stabilized in the transcription complex by these associated factors. Initiator-driven promoters also require ATP during the melting of the DNA, which is essential for formation of a functional transcription complex (13). The fact that both types of promoters require many of the same transcription factors, ATP, and a transient opening of the DNA strands indicates that the TATA-box and TATA-negative genes have transcription initiation pathways that probably converge.

A. Role of Enhancers in Transcription

In addition to the minimal components described above, many eukaryotic genes possess enhancer elements. These enhancers are bound by proteins that can interact with the basal transcription factors to up-regulate transcription. Enhancers are considered to be distinct from promoters because their activity is independent of position and orientation (16). Enhancers and enhancer-like elements are often thought to help provide tissue-specific gene expression (17). It should be noted that there are also elements whose characteristics encompass both enhancer-like and promoter-like activities, thus blurring the differences between them. One such element has been demonstrated to function as an enhancer or as a repressor by affecting the rate of preinitiation complex formation (18).

B. Role of Suppressors/Repressors in Transcription

DNA elements that down-regulate transcription have been identified. These elements are generally referred to as either suppressors or repressors with little or no functional difference associated with the two names. The elements cause a decrease in the amount of successful transcription and are less characterized than the enhancer elements. The suppressor or repressor elements are bound by proteins that seem to down-regulate transcription in several different ways. The most common mechanism is by steric hindrance at an enhancer element. The suppressor/repressor protein either binds directly to the same sequence as the enhancer protein or it binds to a nearby, possibly overlapping sequence (19). Another mechanism is similar but opposite to that of the enhancer. The suppressor/repressor binds to a sequence

that is not position dependent and then interacts with the transcription complex and causes a decrease in transcription, possibly by blocking interactions of other transcription factors with the complex (20). A third mechanism is that the suppressor/repressor can function by blocking transcription elongation. This can result either in a stable complex, which can quickly resume elongation (21), or in complete attenuation and disassociation of the complex (22). Many variations on these mechanisms of transcriptional suppression/repression probably exist, but further study is needed before they can be fully understood.

III. Regulation of the Androgen Receptor Gene

Expression of the androgen receptor is regulated by several agents, including androgens. It is well documented that castration of the rat results in increased expression of AR mRNA in organs such as prostate, seminal vesicles, epididymis, coagulating gland, kidney, and the brain. Following castration, the AR mRNA levels increased 2- to 3-fold within 24 hr after castration and continued to increase for the next 4 days to about 9- to 11-fold that of the control (23-26). Administration of testosterone 24 hr after castration decreased the AR mRNA levels to those of the control level, suggesting that the expression of AR mRNA is under the control of androgens. In order to demonstrate that the autoregulation of AR mRNA is mediated through the AR, similar experiments were conducted in rats exhibiting testicular feminization syndrome (Tfm). Tfm is a genetic defect wherein the genes for rat (26), mouse (27, 28), and human (29) ARs contain mutations that eliminate the biological functions of the protein. It was observed that castration of the Tfm rat failed to increase AR mRNA, indicating that the autologous down-regulation of the AR mRNA by androgens is mediated through a functional AR (26). Similar down-regulation of the AR mRNA by androgens was demonstrated in a prostate cancer cell line, LNCaP (26) and a human hepatoma cell line, HepG2 (25). Furthermore, nuclear run-on analysis demonstrated that androgen treatment cause a 75% reduction in transcription initiation (30, 31), suggesting that the down-regulation of AR mRNA is due to a decrease in transcription initiation.

In addition to androgens, the expression of AR mRNA is influenced by agents utilizing the protein kinase A, protein kinase C, and calcium-regulated pathways. Treatment of prostate cancer cells (LNCaP) with 10^{-6} M calcium ionophore A23187 resulted in a significant decrease in the AR mRNA (32). Similar results were obtained when the cells were treated with ionomycin (10^{-6} M) and the intracellular ATPase inhibitor thapsigargin (10^{-7} M). Time course experiment using A23187 and thapsigargin indicated that the

AR message was down-regulated after 6–8 hr of incubation. Western blot analysis using anti-AR antibodies indicated that both A23187 and thapsigargin reduced AR protein levels.

Using a highly purified rat Sertoli cell preparation, it was demonstrated that treatment with follicle-stimulating hormone (FSH) resulted in significant increase in the AR mRNA and the number of androgen receptors (33), suggesting the involvement of the protein kinase A pathway in the regulation of the AR. Furthermore, treatment of LNCaP cells with tumor-promoting agent (TPA) resulted in a decrease in levels of the AR mRNA (34), suggesting the involvement of the protein kinase C pathway in the regulation of AR.

A. Isolation and Characterization of the 5' Flanking Region of the AR Gene

The 5' flanking regions of the human (35–37), mouse (38, 39), and rat (40, 41) AR genes have been isolated. Computer analysis of the sequences indicated that the AR promoter region lacks both TATA and CAAT boxes in the promoter region. Further analysis of the promoter region indicated that it is a GC-rich region, including several Sp1 binding motifs (35–37, 39). Upstream of the Sp1 binding site is a homopurine stretch containing several copies of the sequence GGGGA. The lack of TATA and CAAT boxes and the presence of GC-rich elements are characteristic of promoters of several housekeeping genes (see discussion above). It has been demonstrated that the Sp1 site at –37 to –46 in the human 5' flanking region is responsible for the initiation of transcription of AR gene (42).

Transcription of the androgen receptor gene has been demonstrated to be initiated at multiple sites. In the human AR the two major sites are 13 bp apart (35, 36, 42). The 5' site has been named the AR transcription initiation site I (AR-TIS I) and the 3' site is the AR-TIS II (36, 42). Furthermore, it has been demonstrated that the transcription at AR-TIS II is regulated by the Sp1 site at –37/–46, whereas transcription at AR-TIS I is dependent on sequences between positions –5 and +57 (42). Similar experiments in the mouse AR indicated the presence of multiple transcription initiation sites that are about 13 bp apart and are regulated by a promoter located upstream of the transcription start site (36). Recent results demonstrated initiation of transcription of mouse AR at sites located about 162 and 170 bp downstream of the sites described earlier (43). Furthermore, a second promoter was identified as being responsible for the initiation of transcription at this site. Northern blot analysis showed that the two promoters are differentially expressed in several tissues and cell lines. Though it has been demonstrated that mRNAs transcribed by both the promoters are under androgen regulation, the importance of two promoters regulating the transcription of the AR gene at multiple sites is not clear.

B. Positive Regulation of the AR Gene

Computer analysis of the 5' flanking region of the AR gene indicated the presence of several potential enhancer elements that are capable of up-regulating the expression of AR (37, 39). One of the key regulatory elements characterized was a cyclic-AMP response element (CRE). The regulation of a gene through the CRE could be due to a direct/primary effect via the binding of CRE binding protein (CREB) and/or indirect/secondary effect via induction of other transcription factors, such as AP1 proteins. Using mutational analysis, transient transfection, and gel retardation assays, functional CREs have been demonstrated in human (37) and mouse (44) AR. In addition to cAMP, the CRE in the 5' flanking region of the AR gene seems to be regulated by several intricate mechanisms. The human AR contains a CRE sequence with a 2-bp mismatch compared to the consensus CRE exemplified by the somatostatin CRE. Though the AR CRE and somatostatin CRE exhibited similar responses in transient transfection experiments, gel retardation assays demonstrated that they formed different protein complexes, suggesting that the regulation of the CRE in somatostatin and AR could be different (37). Experiments combining androgens and cAMP-inducing agents indicated that increasing amounts of AR resulted in a dihydrotestosterone (DHT)-dependent suppression of both basal and forskolin-induced activity of a reporter construct containing the CRE from the mouse AR (44). It was suggested that the effect of the AR was probably mediated through an androgen response element (ARE) half-site located 12 bp downstream of the CRE in the mouse AR. These studies suggest that the regulation of the CRE in the AR gene is a complex process involving interactions of more than one transcription factor.

C. Negative Regulation of the AR Gene

In addition to up-regulation of the AR gene, several elements that are capable of down-regulation of AR have been described. Analysis of the 5' flanking region demonstrated the presence of suppressor elements in the rat (45) and mouse (39, 43, 46) androgen receptors. Deletion analysis of the 5' flanking region of the rat AR suggested the presence of a negative regulatory element. DNase I footprinting of the region -574 to -554 of the rat androgen receptor gene indicated the presence of a 21-bp protected region that includes a 10-bp consensus binding sequence for nuclear factor- κ B (NF- κ B) (45). Further investigation using a gel retardation assay indicated that the negative regulatory element is bound by the p50/p50 homodimer of the NF- κ B protein similar to the inhibitory effect of p50/p50 homodimer in other genes. Furthermore, it has been suggested that NF- κ B and an age-dependent factor act together to coordinate the tissue-specific down-regulation of the rat AR gene during aging (17).

Deletion (47) and transient transfection analysis of the 5' flanking region of the mouse AR indicated the presence of a novel suppressor element at -486 to -351 (39). The suppressor was functional with reference to the natural AR promoter and also a heterologous thymidine kinase promoter. Site-specific mutations demonstrated the presence of a novel GC-rich suppressor element in this region. The suppressor element contains two sequences, next to each other, which have been shown to bind separate suppressor proteins. Thus, the negative regulation of the AR gene is mediated through the binding of one or more suppressor proteins to the suppressor elements (39).

Further analysis of the 5' flanking region of the mouse AR indicated the presence of a suppressor in the 5' untranslated region. DNase I footprinting and gel retardation assays indicated specific binding of a nuclear protein in this region. Point mutation and transient transfection analyses confirmed the presence of a 20-bp suppressor element at +861 to +880 (43). The suppressor appears to exert its influence prior to initiation of translation (46). Analysis of the half-life of the AR mRNA indicated that the suppressor does not affect androgen receptor RNA degradation. Detailed analysis of the gel retardation assays indicated the presence of several species of protein-DNA complexes. Further investigation indicated that several proteins in the nuclear extract were capable of binding single-stranded DNA and RNA probes corresponding to the putative suppressor element in the 5' untranslated region (46). These studies demonstrated that the G-rich strand of the suppressor was bound by a doublet of 33- and 35-kDa proteins and that the homologous C-rich strand was bound by a 52-kDa protein. Based on these observations, a model was proposed for the negative regulation of the AR gene by the suppressor element in the 5' untranslated region. According to this model, at least two different single-stranded binding proteins bind to the suppressor-mARSbpC to the C-rich sense strand and mARSbpG to the G-rich antisense strand. In addition, mARSbpC is also capable of binding to the sense strand of AR mRNA. When both mARSbpC and mARSbpG bind to the two strands of DNA, the transcription of the gene is inhibited. As the ratio of AR DNA to mRNA becomes lower, the amount of AR DNA that is not bound by mARSbpC increases, resulting in reduction of transcription attenuation/blockage (see Fig. 1).

POSTTRANSCRIPTIONAL REGULATION OF THE AR GENE

In several tissues and cell lines, in addition to transcriptional regulation, the AR gene has been shown to be regulated by posttranscriptional events. In COS-1 and LNCaP cells that express the AR gene, the AR mRNA was down-regulated by treatment with androgens (48). In cells that do not express the AR gene (e.g., PC3 and DU 145), transfection of AR cDNA and treatment

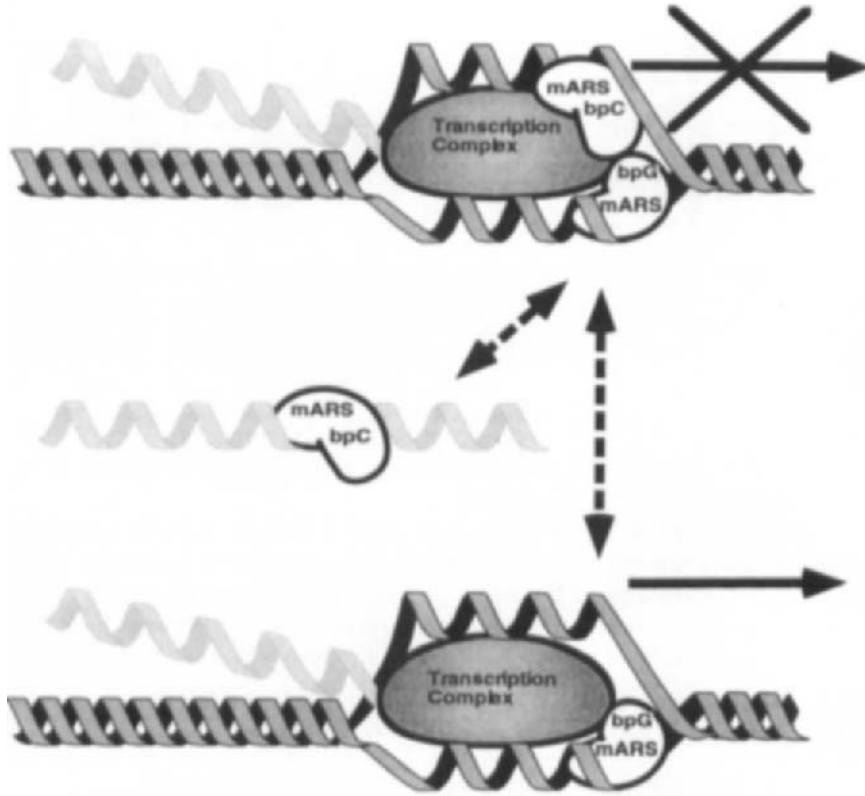


FIG. 1. Model for the action of the suppressor element in the 5' flanking region of the androgen receptor (AR) gene. The mARSbpC represents the single-strand binding protein preferentially binding mARS DNA (double-stranded dark ribbon) and single-stranded light gray mARS RNA. The mARSbpG protein binds the G-rich antisense strand of the mARS-1. The dashed, two-headed arrows indicate the potential for mARSbpG sequestration by mARS RNA and subsequent transcription elongation (solid arrow). The solid arrow with the X represents elongation blockage.

with androgens resulted in up-regulation of the AR (49), indicating that the down- and up-regulation of the AR gene by androgens is mediated by sequences in the AR cDNA. Furthermore, it has been demonstrated that the up-regulation of the AR cDNA involved the binding of the AR protein to the sequences within the AR cDNA (50). Transient transfections utilizing chloramphenicol acetyl transferase (CAT) assays indicated the presence of two putative androgen regulatory elements in the AR cDNA, similar to those pres-

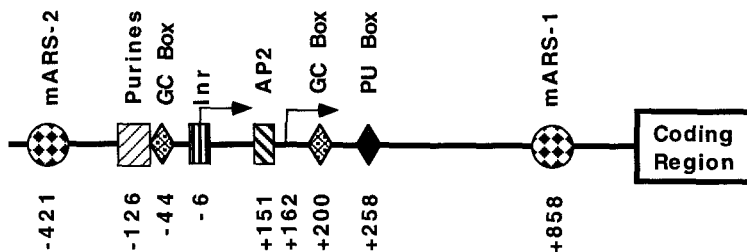


FIG. 2. Diagram of the 5' flanking region of the mouse androgen receptor (AR) gene. Transcription start sites regulated by two promoters are indicated as arrows. The transcriptional elements are listed above the schematic, and their positions are indicated below the schematic. Inr, Initiator sequence; purines, homopurine sequences; mARS-1 and mARS-2, the two suppressor elements.

ent in the probasin gene (51, 52). Disruption of either of the AREs results in about 75% reduction in CAT activity.

Thus transcription of the AR gene is regulated by several enhancer and suppressor elements present on either side of the transcription start site, suggesting the possibility that the rate of transcription of the AR is controlled by overall interaction of the regulatory elements (see Fig. 2 for a diagrammatic representation of response elements in the 5' flanking region of the AR gene).

IV. Regulation of the Glucocorticoid Receptor Gene

The components of a hormone response unit in the 5' flanking region of several genes include combinations of ligand-activated receptors, accessory factors, and coregulators, which interact with the basal transcription unit, resulting in activation or repression of the gene (53). One of the best known examples of this type of regulation is the glucocorticoid receptor (GR) gene. Several studies have demonstrated that the proliferin regulatory region contains consensus binding sites for cJun/cFos and a binding site for GR. It has been demonstrated that the nature of regulation of the proliferin gene is determined by the presence of homodimers of cJun, GR, and their heterodimers (54–56), suggesting that the interactions of factors at the composite response elements provide a mechanism by which a single factor can regulate both positively and negatively (56). Although the interaction of GR and cJun has been extensively investigated, the molecular characterization of the GR promoter region has not been pursued vigorously. Nonetheless, the available information is reviewed here.

A. Characterization of the 5' Flanking Region

The 5' flanking region of the human GR was first isolated and characterized by Zong *et al.* (57). A genomic clone of about 3.0 kb was isolated and sequenced. Computer analysis of the sequence indicated that, similar to the androgen receptor gene, no TATA or CAAT boxes are present in the promoter region of the GR gene. However, just as in other TATA-negative genes, multiple copies of GC box sequences (about 18) were identified, suggesting the involvement of this motif in the regulation of transcription of the GR gene. Another feature common to the TATA-negative genes is the initiation of transcription at several sites, resulting in several transcripts of different lengths. Using RNase protection and primer extension techniques, two to three transcription start sites were identified in two different cell lines (57). These results were confirmed in another independent study; however, S1 nuclease analysis indicated only one major transcription start site for the GR gene (58). Examination of the 5' flanking region of the human GR gene indicated the presence of only one promoter (57–59). These results have been confirmed in a study using several deletion mutants of the 5' flanking region that were cloned into a luciferase vector (60). These experiments identified a single promoter in the region –374 to –183 bp of the human GR that was functional in the context of the natural GR promoter and a heterologous thymidine kinase promoter. Further indirect evidence for the involvement of this region in the regulation of transcription was provided by DNase I footprint analysis and gel mobility shift assays (60).

Unlike the human GR, the mouse GR gene seems to be regulated by at least three promoters (61). Comparison of the structure of the mouse gene with that of the human indicates similarity in exons 2–9, whereas exon 1 demonstrates variations. Further comparison of the DNA sequences at the 5' end of the mouse GR gene with the cDNAs isolated from various tissues of the mice suggests the presence of at least three transcripts with alternate 5' ends that are spliced to the same splice acceptor site, 12 bp upstream of the ATG in exon 2. A combination of PCR amplification and RNase protection assays indicated the presence of three promoters regulating the three transcripts of the GR in different mouse tissues. Promoter 1A is cell-type specific in its activity and has been found to be active in T lymphocytes, whereas the two downstream promoters, 1B and 1C, vary in their activities in different cell lines and tissues (61).

Computer analysis has identified several enhancer elements in the 5' flanking region of the GR gene. However, the functional capabilities of these enhancer elements in the context of the GR promoter have not been demonstrated.

B. Down-Regulation of the GR Gene

Although glucocorticoid receptors are ligand-dependent transcription factors, no glucocorticoid response elements capable of up-regulating the GR gene transcription have been demonstrated in the 3-kb sequence immediately upstream of the ATG (58). However, the presence of two negative response elements have been documented in this region. Using deletion and transfection analysis of the GR promoter region, a sequence between -250 and -750 was implicated in the down-regulation of the human GR gene (59). Furthermore, it was demonstrated that this region was bound by a 95-kDa protein, GRF-1. The cDNA sequence encoding GRF-1 has been isolated and characterized (62), indicating that GRF-1 contains an open reading frame of 2505 nucleotides encoding a protein of about 94 kDa. The deduced amino acid sequence indicated the presence of three zinc finger motifs that could be involved in DNA binding, as suggested by gel retardation and Southwestern analysis. Furthermore, it was noted that cotransfection of a GRF-1 expression vector resulted in transcriptional repression of human GR-chloramphenicol acetyl transferase fusion constructs.

It was demonstrated that the orphan receptor ERR2 (estrogen receptor-related 2) is capable of functioning as a potent repressor of transcriptional activity of the GR gene (63). Transient transfection experiments with GRE-reporter constructs revealed that GR-mediated transcriptional activity is suppressed by cotransfection of an expression vector containing ERR2 cDNA. Furthermore, it was demonstrated that suppressor activity is cell and receptor specific, because progesterone receptor-mediated trans-activation was not suppressed by ERR2 (63).

In addition to the implication of some novel proteins in the repression of GR gene transcription, there is evidence to indicate the involvement of the coding region of the GR in the autoregulation of the gene (64-66). Analysis of the human GR cDNA sequence revealed the presence of sequences with homology to both the positive and negative regulatory elements. Binding of GR to its cDNA indicated specific binding, suggesting its importance in the regulation of the gene. Based on transient transfection experiments it was suggested that the GR is capable of down-regulating its expression by binding to its own cDNA. Furthermore, it was shown that *de novo* synthesis of protein was not required for the down-regulation (66). Deletion analysis indicated that the cDNA susceptible to the down-regulation is present in the 5' half.

V. Regulation of the Progesterone Receptor Gene

The molecular biology of the progesterone receptor (PR) has been extensively investigated and reviewed. Here we review the characterization of the 5' flanking region of the PR gene and the regulation of the gene.

A. Regulation of the PR Gene by Two Distinct Promoters

The progesterone receptor is expressed in two isoforms: the larger B isoform and an N-terminally truncated natural variant A isoform. Among the species examined, only rabbit PR consists solely of the B form. The 5' flanking regions of the chicken, human, rabbit, rat, and mouse PR genes have been isolated and sequenced (2, 67–69). Similar to the androgen and glucocorticoid receptors, the PR gene has no consensus TATA box sequence. However, a consensus CAAT box is present near the transcription start site of the form B PR along with one to two GC box sequences. Using a combination of S1 nuclease analysis, nuclear run-on assay, and transient transfections of deletion constructs, it was demonstrated that the form A and form B PR transcripts are regulated by two distinct promoters (67–70). In the human PR gene a functional promoter regulating the transcription of form B PR was located in the 5' noncoding region, whereas the second promoter, regulating the form A PR, was located in the region +464 to +1105. Both of these promoters were capable of trans-activation of the progesterone response element (PRE) reporter constructs (70). Similar promoter elements have been characterized in rabbit (71) and rat (67) cells.

B. Estrogen Regulation of the PR Gene

The PR gene is expressed predominantly in the female genital tract, the breast, and the brain. Its concentration is hormonally regulated—it is induced by estrogens and down-regulated by progestins. Analysis of the 5' flanking region, 5' untranslated region, and exon 1 identified four estrogen-responsive enhancers (EREs) in the rat PR gene (72). When linked together as an artificial cassette, these five EREs formed a strong estrogen-responsive enhancer. However, individually their response was weak. The strongest individual response was obtained by an ERE in the proximal promoter region and another at the end of the first exon. Furthermore, it was demonstrated that the flanking sequences on either side of the ERE affected the estrogen response by these elements (72). Thus, the expression of the PR gene by estrogens is regulated in a special and unique manner.

VI. Regulation of the Estrogen Receptor Gene: Characterization of the 5' Flanking Region

Estrogens, acting via the estrogen receptor, play an important role in the growth, differentiation, and function of the uterus, vagina, ovary, oviduct, and mammary gland. The genomic organization of the ER gene, including the promoter region, has been well characterized.

The 5' flanking region of the ER gene was isolated and sequenced as early as 1986 (73). Characterization of the 5' flanking region in the human ER gene indicated the presence of two promoters, the proximal promoter A and the distal promoter B (73–76). Promoter B is located about 2.0 kb upstream of promoter A, separated by an intron (76, 77). Analysis of the genomic structure and the transcript RNAs of the human ER indicated the presence of at least two species, mRNA1 and mRNA2, which are different in their 5' untranslated region (73–75, 78–80). Localization of the DNase I hypersensitive sites in the 5' flanking region of the human ER gene indicated differences in the utilization of the two promoters. Three different hypersensitive sites were located in the vicinity of the two promoters when nuclei from the ER-expressing cell line MCF-7 were utilized. However, a similar experiment with another ER-positive cell line, ZR-75-1, showed hypersensitive sites only in the region of the downstream promoter. Detection of the mRNAs transcribed by these two promoters indicated the presence of transcripts from the downstream promoter in both cell lines, whereas the mRNA transcribed by the upstream promoter was present only in the MCF-7 cell (77). The significance of the two different ER mRNA transcripts is not clear, because both transcripts code for identical proteins. However, it has been suggested that the two transcripts could lead to differences in the translational efficiency and/or RNA stability, probably due to formation of different secondary structures in their unique 5' ends (77). Using a reverse transcriptase polymerase chain reaction (RT-PCR) for the detection and quantitation of the ER mRNAs, it has been demonstrated that the transcripts regulated by the proximal promoter are present in osteoblast cells, breast, and uterus, whereas the transcripts of the distal promoter are present only in osteoblast cells. Furthermore, the transcripts from the proximal promoter were 6–37 times greater in mammary and endometrial cancer cell lines, compared to normal breast and uterus (81). Thus, the two promoters in the 5' flanking region of the ER gene seem to be responsible for tissue-specific expression of different species of mRNA. A recent study has suggested the presence of a third promoter at about 10 kb upstream of the two known promoters in the human ER gene (75). This is analogous to the presence of at least three promoters in the mouse GR gene (see above). However, there is no indication of the presence of more than one promoter in the mouse, rat, or chicken ER genes, suggesting either species differences or the lack of a detailed examination of the promoters in these species.

A technique using rapid amplification of cDNA ends and PCR (RACE-PCR) enabled identification of the transcription start site of the proximal (B) promoter of the human ER gene (75), allowing accurate analysis of promoters A and B. The results indicate that, similar to the other steroid hormones previously discussed (AR, GR, and PR), the human ER contains no TATA or

CAAT box in the two promoter regions. Similar to other genes with no defined TATA or CAAT boxes (see above for discussion), the mRNA for ER is transcribed at multiple sites (75, 82).

VII. Concluding Remarks

Many aspects of the mechanism by which steroids regulate target genes have been elucidated. Steroids are known to bind to their cognate receptors, which are then activated. The activated steroid receptor is then capable of binding to the respective hormone response element to influence the transcription of the target gene. However, only recently, studies have begun to elucidate the regulation of the steroid receptor genes by steroids or other factors.

All the steroid receptors discussed herein (AR, GR, PR, and ER) belong to a class of genes containing neither a TATA box nor a CAAT box in the promoter region. Typically these gene sequences are GC rich and contain multiple copies of GC boxes, which have been suggested to be responsible for regulation of transcription of the steroid receptor genes. Another characteristic of these genes is the presence of multiple transcripts originating from multiple sites, suggesting the presence of more than one promoter. Indeed, all the steroid receptor genes discussed herein contain at least two promoters, and the mouse GR and the human ER genes have been shown to contain at least three. The significance of the presence of multiple promoters regulating several species of transcripts is not clear. Several suggestions have been made: (1) the presence of sequences that bestow cell- and tissue-specific expression in this region may be significant and (2) formation of different secondary structures could affect the RNA transcription and/or its stability.

The steroid receptors are regulated by several factors, including members of the protein kinase A pathway, protein kinase C pathway, and calcium. Though the present literature indicates the capability of individual enhancers to regulate transcription, multiple enhancers could be working in tandem or through composite regulatory elements.

One of the important characteristic features of the steroid receptor genes is the down-regulation by their respective steroids, called autoregulation. Autoregulation may occur at the promoter level by the binding of the receptor to hormone response elements. Indeed, cognate hormone response elements have been identified in the 5' flanking region of some of the receptor genes. It is interesting that transcription of the AR gene has been shown to be suppressed by two unique elements that are located on either side of the transcription start site. However, recent research indicates that, in addition to

transcriptional autoregulation, the down-regulation of a steroid receptor gene may be achieved through changes in RNA stability and half-life. Furthermore, the coding regions of the glucocorticoid and estrogen receptor genes have been implicated in autoregulation.

It is clear that the regulation of transcription of the steroid receptors is a complex process involving elements for both negative and positive regulation in the 5' flanking region. Furthermore, as seen in the case of androgen and glucocorticoid receptors, transcription is also influenced by sequences in the coding region. Further studies will be needed to elucidate the complex mechanism by which these genes are regulated.

REFERENCES

1. R. M. Evans, *Science* **240**, 889 (1988).
2. J. F. Savouret, M. Misrahi, H. Loosfelt, M. Atger, A. Bailly, M. Perrot-Appianat, M. T. Vu Hai, A. Guiochon-Manter, A. Jolivet, F. Lorenzo, F. Logeat, M. F. Pichon, P. Bouchard, and E. Milgrom, *Recent Prog. Hormone Res.* **45**, 65 (1989).
3. M. A. Carson-Jurica, W. T. Schrader, and B. W. O'Malley, *Endocrinol. Rev.* **11**, 201 (1990).
4. B. M. Forman and H. H. Samuels, *Mol. Endocrinol.* **4**, 1293 (1990).
5. B. O'Malley, *Mol. Endocrinol.* **4**, 363 (1990).
6. M. G. Parker, *Seminars Cancer Biol.* **1**, 81 (1990).
7. V. Laudet, C. Hanni, J. Coll, F. Catzeflis, and D. Stehelin, *EMBO J.* **11**, 1003 (1992).
8. J. P. Lander and T. C. Spelsberg, *Crit. Rev. Eukaryot. Gene Exp.* **2**, 19 (1992).
9. J. Lindzey, M. V. Kumar, M. Grossman, C. Young, and D. J. Tindall, *Vitamins Hormones* **49**, 383 (1993).
10. M.-J. Tsai and B. W. O'Malley, *Annu. Rev. Biochem.* **63**, 451 (1994).
11. B. W. O'Malley, W. T. Schrader, S. Mani, C., Smith, N. L. Weigel, O. M. Conneely, and J. H. Clark, *Recent Progr. Hormone Res.* **50**, 333 (1995).
12. B. F. Pugh and R. Tjian, *Genes Dev.* **5**, 1935 (1991).
13. Y. Jiang, S. T. Smale, and J. D. Gralla, *J. Biol. Chem.* **268**, 6535 (1993).
14. S. T. Smale and D. Baltimore, *Cell* **57**, 103 (1989).
15. B. F. Pugh and R. Tjian, *Cell* **61**, 1187 (1990).
16. C. C. Fraizer, Y. J. Wu, S. M. Hewitt, T. Maity, C. C. Ton, V. Huff, and G. F. Saunders, *J. Biol. Chem.* **269**, 8892 (1994).
17. P. C. Supakar, H. J. Myeong, C. S. Song, B. Chatterjee, and A. K. Roy, *J. Biol. Chem.* **270**, 837 (1995).
18. F. B. Johnson and M. A. Krasnow, *Gene Dev.* **6**, 2177 (1992).
19. D. G. Skalnik, E. C. Strauss, and S. H. Orkin, *J. Biol. Chem.* **266**, 16736 (1991).
20. S. N. Agoff, J. Hou, D. I. Linzer, and B. Wu, *Science* **259**, 84 (1993).
21. N. J. C. Lamb, A. Fernandez, N. Tourkine, P. Jeanteur, and J. Blanchard, *Cell* **61**, 485 (1990).
22. S. Wright, L. F. Mirels, M. C. Calayag, and J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11383 (1991).
23. J.-A. Tan, D. R. Joseph, V. E. Quarmbly, D. B. Lubahn, M. Sar, F. S. French, and E. M. Wilson, *Mol. Endocrinol.* **2**, 1276 (1988).
24. D. B. Lubahn, J.-A. Tan, V. E. Quarmbly, M. Sar, D. R. Joseph, F. S. French, and E. M. Wilson, *Ann. N.Y. Acad. Sci.* **564**, 48 (1989).

25. L.-X. Shan, M. C. Rodriguez, and O. A. Janne, *Mol. Endocrinol.* **4**, 1636 (1990).
26. V. E. Quarmby, W. G. Yarbrough, D. B. Lubahn, F. S. French, and E. M. Wilson, *Mol. Endocrinol.* **4**, 22 (1990).
27. W. W. He, M. V. Kumar, and D. J. Tindall *Nucleic Acids Res.* **19**, 2373 (1991).
28. W. W. He, J. K. Lindzey, J. L. Prescott, M. V. Kumar, and D. J. Tindall, *Receptor* **4**, 121 (1994).
29. M. Marcelli, S. Zoppi, P. B. Grino, J. E. Griffin, J. D. Wilson, and M. J. McPhaul, *J. Clin. Invest.* **87**, 1123 (1991).
30. D. A. Wolf, T. Herzinger, H. Hermeking, D. Blaschke, and W. Horz, *Mol. Endocrinol.* **7**, 924 (1993).
31. L. J. Blok, J. W. Hoogerbrugge, A. P. N. Themman, W. M. Baarends, M. Post, and J. A. Grootegoed, *Endocrinology* **131**, 1343 (1992).
32. Y. Gong, L. J. Blok, J. E. Perry, J. K. Lindzey, and D. J. Tindall, *Endocrinology* **136**, 2172 (1995).
33. L. J. Blok, P. Mackenbach, J. Trapman, A. P. N. Themmen, A. O. Brinkmann, and J. A. Grootegoed, *Mol. Cell. Endocrinol.* **63**, 267 (1989).
34. P. Henttu and P. Vihko, 75th Annu. Mt. Endocrine Soc. #1254, 1993.
35. W. D. Tilley, M. Marcelli, and M. J. McPhaul, *J. Biol. Chem.* **265**, 13776 (1990).
36. P. W. Faber, H. C. J. Van Rooij, H. A. G. M. Van der Korput, W. M. Baarends, A. O. Brinkmann, J. A. Grootegoed, and J. Trapman, *J. Biol. Chem.* **266**, 10743 (1991).
37. A. Mizokami, S.-Y. Yeh, and C. Chang, *Mol. Endocrinol.* **8**, 77 (1994).
38. P. W. Faber, A. King, H. C. J. Van Rooij, A. O. Brinkmann, N. J. De Both, and J. Trapman, *Biochem. J.* **278** 269 (1991).
39. M. V. Kumar, E. A. Jones, M. E. Grossmann, M. D. Blexrud and D. J. Tindall, *Nucleic Acids Res.* **22**, 3693 (1994).
40. W. M. Baarends, A. P. N. Themmen, L. J. Blok, P. Mackenbach, A. O. Brinkmann, D. Meijer, P. W. Faber, J. Trapman, and J. A. Grootegoed, *Mol. Cell. Endocrinol.* **74**, 75 (1990).
41. P. C. Supaker, C. S. Song, M. H. Jung, M. A. Slomczynska, J.-M. Kim, R. L. Vellanoweth, B. Chatterjee, and A. K. Roy, *J. Biol. Chem.* **268**, 26400 (1993).
42. P. W. Faber, H. C. J. Van Rooij, H. J. Schipper, A. O. Brinkmann, and J. Trapman, *J. Biol. Chem.* **268**, 9296 (1993).
43. M. E. Grossmann, J. Lindzey, L. Blok, J. E. Perry, M. V. Kumar, and D. J. Tindall, *Biochemistry* **33**, 14594 (1994).
44. J. Lindzey, M. Grossmann, M. V. Kumar, and D. J. Tindall, *Mol. Endocrinol.* **7**, 1530 (1993).
45. C. S. Song, M. H. Jung, P. C. Supakar, S. Chen, R. L. Vellanoweth, B. Chatterjee, and A. K. Roy, *Ann. N.Y. Acad. Sci.* **761**, 97 (1995).
46. M. E. Grossmann and D. J. Tindall, *J. Biol. Chem.* **270**, 10968 (1995).
47. M. V. Kumar and D. J. Tindall, *Meth. Neurosci.* **26**, 324 (1995).
48. K. L. Burnstein, C. A. Maiorino, J. L. Dai, and D. J. Cameron, *Mol. Cell. Endocrinol.* **115**, 177 (1995).
49. J. L. Dai, C. A. Maiorino, P. J. Gkonos, and K. L. Burnstein, *Steroids* **61**, 531 (1996).
50. J. L. Dai and K. L. Burnstein, *Mol. Endocrinol.* **10**, 1582 (1996).
51. P. S. Rennie, N. Bruchovsky, K. J. Leco, P. C. Sheppard, S. A. McQueen, H. Cheng, R. Snoek, A. Hamel, M. E. Bock, B. S. MacDonald, B. E. Nickel, C. Chang, S. Liao, P. A. Cattini, and R. J. Matusik, *Mol. Endocrinol.* **7**, 23 (1993).
52. S. Kasper, P. S. Rennie, N. Bruchovsky, P. C. Sheppard, H. Cheng, L. Lin, R. P. C. Shiu, R. Snoek, and R. J. Matusik, *J. Biol. Chem.* **269**, 31763 (1994).
53. E. C. Guido, E. O. Delorme, D. L. Clemm, R. B. Stein, J. Rosen, and J. N. Miner, *Mol. Endocrinol.* **10**, 1178 (1996).
54. R. Schule, P. Rangarajan, S. Kliewer, L. J. Ransone, J. Bolado, N. Yang, I. M. Verma, and R. M. Evans, *Cell* **62**, 1217 (1990).

55. J. N. Miner, M. I. Diamond, and K. R. Yamamoto, *Cell Growth Diff.* **2**, 525 (1991).
56. J. N. Miner and K. R. Yamamoto, *Genes Dev.* **6**, 2491 (1992).
57. J. Zong, J. Ashraf, and E. B. Thompson, *Mol. Cell. Biol.* **10**, 5580 (1990).
58. I. J. Encio and S. D. Detera-Wadleigh, *J. Biol. Chem.* **266**, 7182 (1991).
59. S. LeClerc, R. Palanissami, B. Xie, and M. V. Govindan, *J. Biol. Chem.* **266**, 17333 (1991).
60. Y. Nobukuni, C. L. Smith, G. L. Hager, and S. D. Detera-Wadleigh, *Biochemistry* **34**, 8207 (1995).
61. U. Strahle, A. Schmidt, G. Kelsey, A. F. Stewart, T. J. Cole, W. Schmid, and G. Schutz, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6731 (1992).
62. S. LeClerc, B. Xie, R. Roy, and M. V. Govindan, *J. Biol. Chem.* **266**, 8711 (1991).
63. T. Trapp and F. Holboer, *J. Biol. Chem.* **271**, 9879 (1996).
64. K. L. Burnstein, C. M. Jewell, and J. A. Cidlowski, *J. Biol. Chem.* **265**, 7284 (1990).
65. K. L. Burnstein, D. L. Bellingham, C. M. Jewell, F. E. Powell-Oliver, and J. A. Cidlowski, *Steroids* **56**, 52 (1991).
66. K. L. Burnstein, C. M. Jewell, M. Sar, and J. A. Cidlowski, *Mol. Endocrinol.* **8**, 1764 (1994).
67. W. L. Kraus, M. M. Montano, and B. S. Katzenellenbogen, *Mol. Endocrinol.* **7**, 1603 (1993).
68. M. Misrahi, P. Y. Venencie, P. Saugier-Verber, S. Sar, P. Dessen, and E. Milgrom, *Biochim. Biophys. Acta* **1216**, 289 (1993).
69. K. Hagihara, X. W. Wu-Peng, T. Funabashi, J. Kato, and D. W. Pfaff, *Biochem. Biophys. Res. Commun.* **205**, 1093 (1994).
70. P. Kastner, A. Krust, B. Turcotte, U. Stropp, L. Tora, H. Gronemeyer, and P. Chambon, 1990, *EMBO J.* **9**, 1603 (1990).
71. M. Misrahi, H. Leesfelt, M. Atger, C. Meriel, V. Zerah, P. Dessen, and E. Milgrom, *Nucleic Acids Res.* **16**, 5459 (1988).
72. W. L. Kraus, M. M. Montano, and B. S. Katzenellenbogen, *Mol. Endocrinol.* **8**, 952 (1994).
73. S. Green, P. Walter, V. Kumar, A. Krust, J. M. Bornert, P. Argos, and P. Chambon, *Nature (London)* **320**, 134 (1986).
74. M. Keaveney, J. Klug, M. T. Dawson, P. V. Nestor, J. Neilan, R. C. Forde, and F. Gannon, *J. Mol. Endocrinol.* **6**, 111 (1991).
75. K. Grandien, *Mol. Cell. Endocrinol.* **116**, 207 (1996).
76. R. Piva, R. Ganbarj, F. Zorzato, L. Kumar, and L. Del Senno, *Biochem. Biophys. Res. Commun.* **183**, 996 (1992).
77. K. F. H. Grandien, A. Berkenstam, S. Nilsson, and J.-A. Gustafsson, *J. Mol. Endocrinol.* **10**, 269 (1993).
78. V. Kumar, S. Green, A. Staub, and P. Chambon, *EMBO J.* **5**, 2231 (1986).
79. V. Kumar, S. Green, G. Stack, M. Berry, J. R. Jin, and P. Chambon, *Cell* **51**, 941 (1987).
80. M. Ponglikitmongkol, S. Green, and P. Chambon, *EMBO J.* **7**, 3385 (1988).
81. K. Grandien, M. Backdahl, O. Ljunggren, J.-A. Gustafsson, and A. Berkenstam, *Endocrinology* **136**, 2223 (1995).
82. R. White, J. A. Lees, M. Needham, J. Ham, and M. Parker, *Mol. Endocrinol.* **1**, 735 (1987).

Molecular Evolution of Snake
Toxins: Is the Functional
Diversity of Snake Toxins
Associated with a Mechanism
of Accelerated Evolution?

M. OHNO,[†] R. MÉNEZ,*
T. OGAWA,[†] J. M. DANSE,*
Y. SHIMOHIGASHI,[†] C. FROMEN,*
F. DUCANCEL,* S. ZINN-JUSTIN,*
M. H. LE DU,* J.-C. BOULAIN,*
T. TAMIYA,[†] AND A. MÉNEZ*.¹

*CEA

*Département d'Ingénierie et d'Etudes
des Protéines*

91191 Gif-sur-Yvette, France

[†]*Department of Chemistry Faculty of
Science*

Fukuoka 812, Japan

[†]*Sophia University*

Department of Chemistry

Tokyo 102, Japan

| | |
|---|-----|
| I. About Snake Toxins | 309 |
| II. Snake Toxins with a Phospholipase A ₂ -Type Fold | 311 |
| A. Structures of Snake Venom Phospholipase A ₂ Isozymes | 311 |
| B. Diverse Physiological Functions of Snake Venom Phospholipase A ₂ Isozymes | 317 |
| C. Structures of cDNAs and Genes Encoding Viperidae Snake Venom Gland Phospholipase A ₂ Isozymes | 324 |
| D. Darwinian-Type Accelerated Evolution of Viperidae Snake Venom Gland Phospholipase A ₂ Isozymes | 327 |
| III. Snake Toxins with a Three-Fingered Fold | 339 |
| A. Structures of Three-Fingered Toxins | 339 |
| B. Functions of Three-Fingered Toxins | 343 |
| C. On the cDNAs Encoding Three-Fingered Toxins | 344 |
| D. Functional Sites of Three-Fingered Toxins | 348 |
| E. Nontoxic Proteins with Three-Fingered Structures | 354 |
| F. On the Functional Diversity Associated with the Three-Fingered Fold | 356 |
| IV. General Conclusion on the Evolution of Snake Toxins | 356 |
| References | 357 |

¹ Author to whom correspondence should be addressed.

Recent studies revealed that animal toxins with unrelated biological functions often possess a similar architecture. To tentatively understand the evolutionary mechanisms that may govern this principle of functional prodigality associated with a structural economy, two complementary approaches were considered. One of them consisted of investigating the rates of mutations that occur in cDNAs and/or genes that encode a variety of toxins with the same fold. This approach was largely adopted with phospholipases A₂ from *Viperidae* and to a lesser extent with three-fingered toxins from *Elapidae* and *Hydrophiidae*. Another approach consisted of investigating how a given fold can accommodate distinct functional topographies. Thus, a number of topologies by which three-fingered toxins exert distinct functions were investigated either by making chemical modifications and/or mutational analyses or by studying the three-dimensional structure of toxin-target complexes. This review shows that, although the two approaches are different, they commonly indicate that most if not all the surface of a snake toxin fold undergoes natural engineering, which may be associated with an accelerated rate of evolution. The biochemical process by which this phenomenon occurs remains unknown. © 1998 Academic Press

Various phyla of the animal kingdom contain venomous species, including the coelenterates, flatworms, annelids, echinoderms, mollusks, arthropods, and chordates. The venomous animals included in these phyla elaborate a diversity of toxins with which they subdue their prey and/or protect themselves from predators. Animal toxins, therefore, exert immobilizing and/or killing functions toward a considerable variety of creatures. After nearly three decades of studies, the structural basis that is associated with the diversity of functions exerted by animal toxins is being elucidated. Chemical analyses have initially shown that animal toxins are often proteins, frequently characterized by a small size (10 to 150 amino acids) and a high density of disulfide bonds. Structural investigations based on X-ray crystallography and/or nuclear magnetic resonance (NMR) spectroscopy have revealed that animal toxins with unrelated functions can nevertheless adopt a similar fold (*I*). At present, however, it is unclear as to whether this principle of functional prodigality associated with a structural economy results from any specific biochemical and/or evolutionary process. With the tentative view to shed some light on this question, two complementary approaches have been considered. One of them consists in studying the rates of mutations that occur in cDNAs and/or genes that encode a variety of toxins with the same fold. This approach has been principally adopted with snake phospholipases A₂ and to a lesser extent with three-fingered snake toxins. Another approach consists in investigating how a given fold can accommodate distinct functional topographies. Thus, the topologies by which three-fingered toxins exert distinct functions have been investigated using either chemical modifi-

cations and/or mutational analyses or by studying the three-dimensional structure of toxin–target complexes. Though different, the two approaches converge to indicate that most if not all the surface of a snake toxin fold undergoes natural engineering that, furthermore, may be associated with an accelerated rate of evolution.

I. About Snake Toxins

Approximately 2700 species of snakes have been identified, among which nearly 400 species, essentially belonging to five families, are venomous. These families are the Elapidae (cobras, mambas, kraits, and coral snakes), the Hydrophiidae (sea snakes), the Viperidae (vipers and crotals), the Atractaspididae (burrowing asps), and to a lesser extent the Colubridae. Throughout this article, the genus names of the members of these families are often abbreviated. The reader may wish to refer to the following table for the complete scientific nomenclature.

| | |
|--|-----------------------------------|
| <i>Agkistrodon piscivorus piscivorus</i> | <i>Laticauda semifasciata</i> |
| <i>Aipysurus Laevis</i> | <i>Naja naja atra</i> |
| <i>Bitis gabonica</i> | <i>Naja naja siamensis</i> |
| <i>Bitis nasicornis</i> | <i>Naja naja sputatrix</i> |
| <i>Bothrops asper</i> | <i>Naja nigricollis</i> |
| <i>Bungarus multicinctus</i> | <i>Notechis scutatus scutatus</i> |
| <i>Crotalus atrox</i> | <i>Trimeresurus flavoviridis</i> |
| <i>Crotalus durissus</i> | <i>Trimeresurus gramineus</i> |
| <i>Crotalus scutulatus scutulatus</i> | <i>Trimeresurus okinavensis</i> |
| <i>Dendroaspis angusticeps</i> | <i>Vipera ammodytes ammodytes</i> |
| <i>Dendroaspis jamesonii kaimosae</i> | <i>Vipera russelli formensis</i> |
| <i>Dendroaspis polylepis polylepis</i> | |

Venomous snakes synthesize a variety of toxins that have been mainly isolated by two different approaches. The first and most common approach consists in successively fractionating venoms and screening the isolated fractions exerting a given activity, including lethality and/or a capacity to interfere with an *in vitro* assay, such as binding to a receptor, an ion channel, an enzyme, or a membrane. This strategy leads, apparently endlessly, to the discovery of toxins acting on an incredible variety of targets (1). A second approach consists in systematically analyzing the chemical structure of all proteins that are anticipated to be potential toxins and then in searching for the function of the identified protein (2, 3). Such a strategy is possible because most snake toxins share a number of chemical and structural features, including a small

size, making them readily identifiable in a chromatography profile. Thus, many small proteins have been isolated from snake venoms and their biological activity has sometimes been elucidated later (2), but in most cases, their actual function still remains to be discovered.

Snake toxins can be classified into at least seven distinct groups. Thus, there are toxins with (1) 21 residues with 2 disulfides (sarafotoxins), (2) 43–45 residues with 3 disulfides (myotoxic peptides), (3) 57–60 residues with 3 disulfides (dendrotoxins), (4) 47–83 residues with 4–7 disulfides (viper disintegrins); (5) 60–74 residues with usually 4–5 disulfides (curaremimetic toxins and others), (6) approximately 120 residues with 6–8 disulfides (phospholipases A_2), and (7) 200–800 residues with several disulfides (hemorrhagins). That toxins from these different families are structurally unrelated has been inferred from the observations that (i) alignments of their amino acid sequences have failed; (ii) their secondary structures, as deduced from circular dichroism analyses, are different, and (iii) their architectures, as elucidated by X-ray crystallography and/or NMR spectroscopy, are also different. Thus, snake toxins can be classified into at least seven distinct folds.

Some of these folds are presently associated with a single known toxic function. These are the folds adopted by sarafotoxins, myotoxic peptides, viper disintegrins, and hemorrhagins. The sarafotoxins are small peptides structurally and functionally similar to endothelins. They induce cardiotoxic effects in mice, act as potent vasoconstrictors of the aorta and coronary blood cells, and bind to endothelin receptors from heart and brain (4). The myotoxic peptides cause vacuolation, lysis, and necrosis of skeletal muscle cells; however, their molecular mode of action remains unclear (5). The disintegrins are inhibitors of platelet aggregation. They interact with glycoprotein (Gp) IIb IIIa, a glycoprotein also called integrin, which is the major receptor for fibrinogen on activated platelets (6, 7). The hemorrhagins are metalloproteinases that cause bleeding in victims of crotals and viper envenomation. They are complex proteins with molecular masses ranging from 20 to 90 kDa and are composed of a metalloproteinase domain followed, sometimes, by a disintegrin-like domain, followed in turn by a cysteine-rich region and a lectin domain (8).

The other folds are associated with different activities. One type of fold is shared by two types of mamba toxins: the dendrotoxins, which act on voltage-gated potassium channels (9), and toxins such as calcicludine, which recognize calcium channels (10). Two other folds, the phospholipase A_2 -type fold and the three-fingered fold, are associated with many more activities. Toxins adopting either of these two folds have been extensively studied and therefore offer a solid basis to investigate the questions raised above.

II. Snake Toxins with a Phospholipase A₂-Type Fold

Hereafter, we describe the structures, functions, gene structures, and evolutionary aspects of Viperidae snake venom phospholipases A₂.

A. Structures of Snake Venom Phospholipase A₂ Isozymes

1. PRIMARY STRUCTURES AND CLASSIFICATION

Phospholipases A₂ (PLA₂s) (EC 3.1.1.4) catalyze the hydrolysis of the 2-acyl ester linkage of 1,2-diacyl-3-*sn*-phosphoglycerides, with the requirement of Ca²⁺. PLA₂s occur in snake venoms (Elapidae and Viperidae) in fairly large quantity, but also in mammalian tissues and cells such as pancreas (11, 12), stomach (13), intestine (14), liver (15), spleen (16), platelets (17, 18), and macrophages (19). The study of *Trimeresurus flavoviridis* (Crotalinae) snake venom gland by means of *in situ* hybridization and immunohistochemical techniques showed that PLA₂ isozymes are abundantly expressed in the cytosol of venom gland cells (25). The primary structure of more than 170 PLA₂s have been determined (26). They consist of 120–130 amino acid residues and most of them contain seven disulfide bonds (27). PLA₂s are classified into two major groups based on structural characteristics. They share six disulfide bonds but the seventh is different in the two groups. Group I PLA₂s (Elapidae snake venoms and mammalian pancreas) possess an additional disulfide bond formed between half-cystines 11 and 77. Group II PLA₂s (Viperidae snake venoms, mammalian platelets, and liver) are characterized by a distinct, short (six residues) extension at the COOH terminus that terminates in Cys-133 linked to Cys-50. The evolutionary tree constructed from the amino acid sequences justifies this classification (28). A new type of group I PLA₂s that lack the disulfide bond between Cys-11 and Cys-77 has been found in human and rat heart (29, 30). These PLA₂s seem to be placed between PLA₂s of groups I and II. In some viperid species such as *Bitis gabonica*, *Bitis nasicornis*, and *Trimeresurus gramineus* (34), PLA₂s lacking the Cys-61–Cys-91 disulfide bond have also been found. An alignment of a set of amino acid sequences of PLA₂s (groups I and II) is shown in Fig. 1.

Group II PLA₂s of viperid snake venoms are further classified into two groups, depending on whether their position 49 (numbered according to the numbering of PLA₂s from various sources), which is part of the Ca²⁺ binding site (37), is occupied by an aspartate or a lysine: [Asp⁴⁹]PLA₂s and [Lys⁴⁹]PLA₂s in which Asp-49 is replaced by lysine (28–31, 37–41). [Lys⁴⁹]PLA₂s are characterized by lower and distinct PLA₂ activity, lower re-

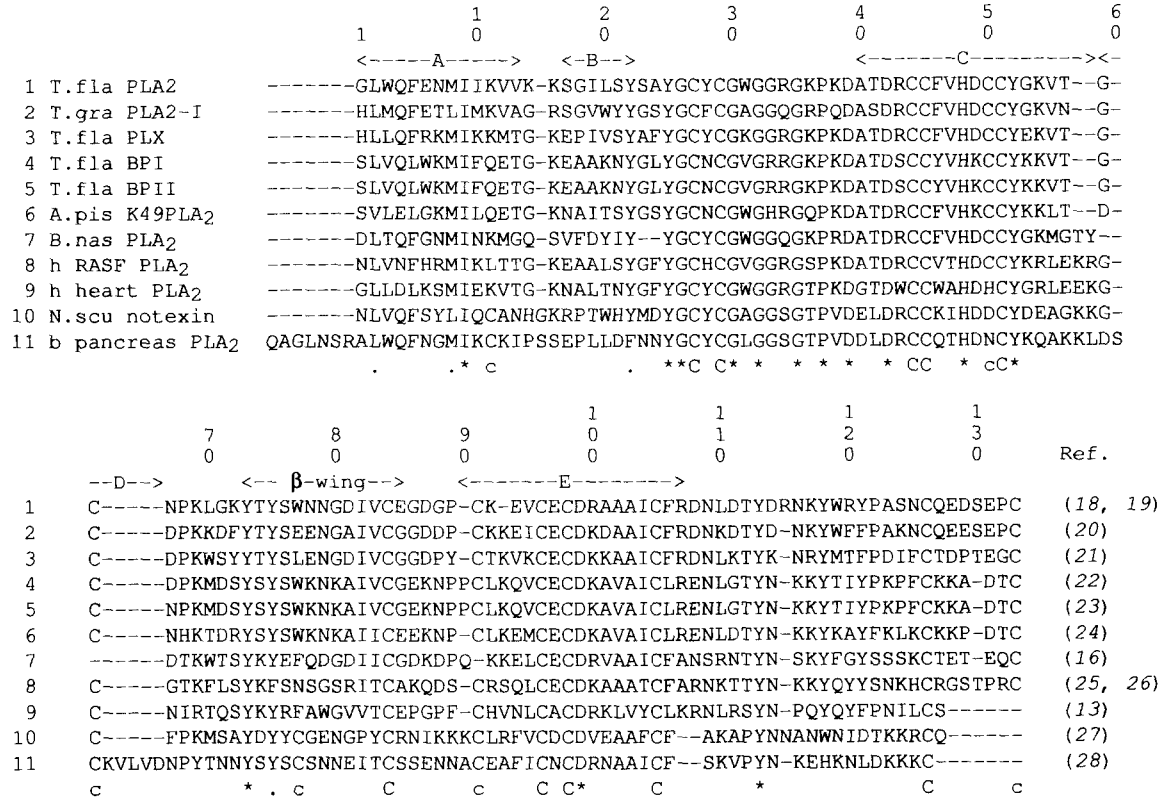


FIG. 1. Aligned amino acid sequences of selected PLA₂s from snake venom and mammalian sources. The alignment was made by the CLUSTAL W program (36). The nonconserved half-cystine residues are denoted by the lowercase letter c. Asterisks indicate the consensus residues. The five helices (A, B, C, D, and E) and the β-wing are indicated at the top of the sequences. T.fla, *Trimeresurus flavoviridis*; T.gra, *Trimeresurus gramineus*; A.pis, *Agkistrodon piscivorus piscivorus*; B.nas, *Bitis nasicornis*; h RASF, human rheumatoid arthritis synovial fluid; N.scu, *Notechis scutatus scutatus*; b, bovine.

activity of the catalytic His-48 toward *p*-bromophenacyl bromide, and weaker affinity to Ca^{2+} when compared with [Asp⁴⁹]PLA₂s (28–30).

2. SECONDARY AND TERTIARY STRUCTURES AND CATALYTIC MECHANISM

X-Ray crystallographic analyses showed that [Asp⁴⁹]PLA₂s from bovine pancreas (group I) (37) and from *Crotalus atrox* (42, 43) and *T. flavoviridis* venoms (44) (group II) adopt virtually superimposable tertiary structures, except around the extra loop between residues 62 and 66 (helix D) in group I [Asp⁴⁹]PLA₂s (Fig. 2). Most PLA₂s have a similar secondary structure comprising four helices (A, B, C, and E) and a β structure, termed the “ β -wing,” that are present in all PLA₂s (Figs. 1 and 2). Helix A is located near the catalytic site, whereas the long and parallel helices C and E are located in the center of the proteins. The residues in helices C and E are more conserved than in other parts, suggesting their possible structural importance. Helix D is present only in pancreatic and some elapid [Asp⁴⁹]PLA₂s and is called the “pancreatic loop” or the “elapid loop.” The helix D-deleted mutants of pancreatic [Asp⁴⁹]PLA₂s show an enhanced catalytic activity toward micellar phospholipids and an altered specificity (45, 46), but no significant change in affinity to M-type PLA₂ receptor (47).

The spatial structures of [Asp⁴⁹]PLA₂s complexed with transition-state analog inhibitors (48–51) and with *p*-bromophenacyl bromide, a specific alkylating reagent of the active site His-48 of PLA₂s (52) suggest a mechanism for catalysis of [Asp⁴⁹]PLA₂s (Fig. 3). A water molecule, activated by a hydrogen bonding to the imidazole of His-48, is proposed to attack the carbonyl carbon at the *sn*-2 position of phosphoglycerides. Ca^{2+} seems to be essential for the binding of the substrate and for catalysis. This ion is heptacoordinated by a pentagonal bipyramidal c. g. e, which includes the carbonyl oxygen of Gly-30 (for electrophilic stabilization of the oxyanion in the active site), the nonbridging oxygen of the *sn*-3 phosphate, and the carbonyl oxygen of the *sn*-2 ester of the substrate. Asp-49, Tyr-28, Gly-30, and Gly-32 interact with Ca^{2+} and constitute a so-called Ca^{2+} binding loop. However, the *T. gramineus* [Asp⁴⁹]PLA₂ isozyme has a phenylalanine in place of Tyr-28, with no significant effect on lipolytic activity (25).

[Lys⁴⁹]PLA₂s differ from [Asp⁴⁹]PLA₂s by some systematic substitutions, particularly in the Ca^{2+} binding loop. Thus, in association with substitution of Asp-49 by lysine, Tyr-28 and Trp-31 are replaced by asparagine and valine, respectively. An X-ray crystallographic analysis of [Lys⁴⁹]PLA₂ from *Agkistrodon piscivorus piscivorus* venom (53, 54) showed that these substituted amino acid residues are in contact via hydrogen bonds and that the tertiary structure of the enzyme can superimpose with that of group II [Asp⁴⁹]PLA₂s.

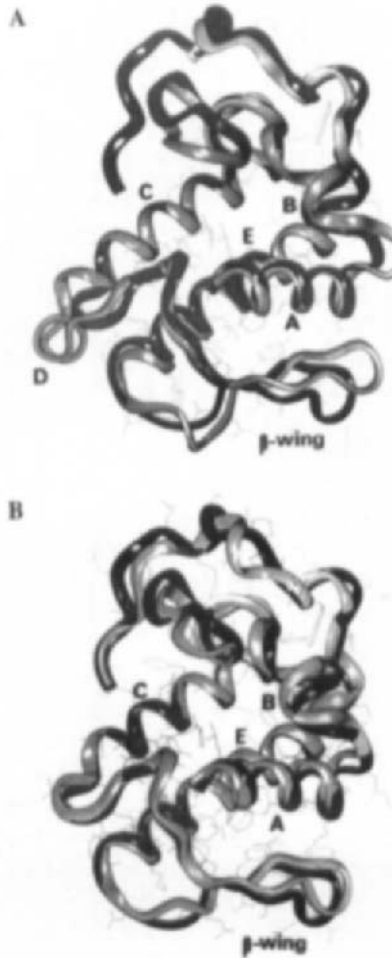


FIG. 2. (A) Overlapped tertiary structures (showing helices A–E and the β -wing) of *T. flavoviridis* venom [Asp⁴⁹]PLA₂ (group II) (black) and bovine pancreatic [Asp⁴⁹]PLA₂ (group I) (grey). (B) Overlapped tertiary structures of *T. flavoviridis* venom [Asp⁴⁹]PLA₂ (black) and *A. piscivorus piscivorus* venom [Lys⁴⁹]PLA₂ (group II) (grey). Homologous segments of the backbone were superimposed by a least-square fit of the corresponding C α atoms. Five helices (A–E) and the β -wing are indicated.

3. QUATERNARY STRUCTURES

PLA₂s can adopt different quaternary structures, including monomeric, dimeric (homo and hetero), and trimeric architectures. For example, crotoxin from *Crotalus durissus terrificus* venom is a heterodimer with an acidic

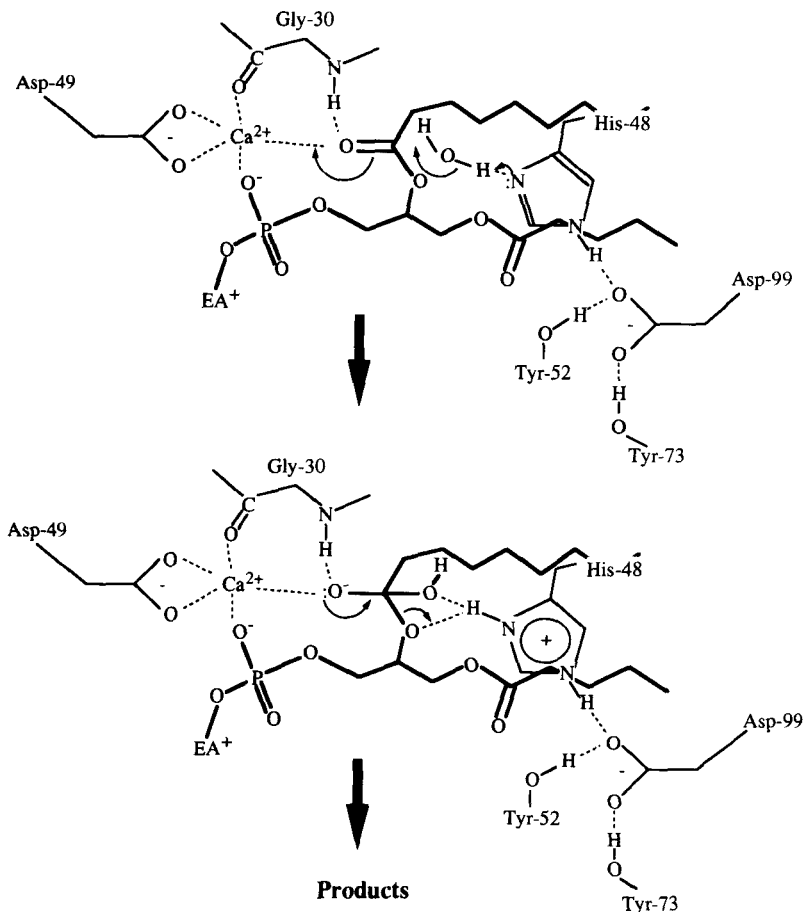


FIG. 3. Proposed catalytic mechanism of [Asp⁴⁹]PLA₂s based on their tertiary structures and chemical evidence. The phosphoglyceride is in bold face.

nontoxic subunit and a basic subunit (55), the dimer exerting more potent presynaptic neurotoxicity than the basic subunit alone. The same observation can be made for both mojave toxin (heterodimeric neurotoxic PLA₂s) from *Crotalus scutulatus scutulatus* venom (56) and neurotoxic PLA₂s (RV-4 and RV-7) from *Vipera russelli formosensis* venom (57). Therefore, the quaternary structure of PLA₂s seems to be important for expression of activity. The X-ray structures of the monomeric (53, 54, 58), dimeric (42–44, 59), and trimeric (60) PLA₂s of snake venom origin have been determined. *Crotalus atrox* [Asp⁴⁹]PLA₂ (42, 43) and *T. flavoviridis* [Asp⁴⁹]PLA₂ (44) are dimeric pro-

teins with two identical interacting subunits and their active sites embedded in the interior of the dimer.

4. CHEMICAL MODIFICATION OF PARTICULAR AMINO ACID RESIDUES

To elucidate the role of particular residues, a number of chemical modifications have been carried out with the dimeric, lipolytically highly active *T. flavoviridis* [Asp⁴⁹]PLA₂ isozyme, which is named PLA2. The combined methodologies of the stepwise oxidation of tryptophan residues by *N*-bromosuccinimide (61) and selective separation of tryptophan-containing peptides (62) revealed (1) that Trp-31 is located in close proximity to the Ca²⁺ binding site and is critical for activity (63), in agreement with what was anticipated from crystallographic analyses (44), and (2) that Trp-31 penetrates into the substrate binding site of the opposite subunit and appears to play an important role for construction of dimer. Treatment with cyanogen bromide (CNBr) cleaved an octapeptide from the N-terminal moiety, causing a reduction of the lipolytic activity to 10%. Modification of His-48 in the active site with methyl *p*-nitrobenzene sulfonate was conducted to identify the orientation of its imidazole group (64). Reaction at pH below its pK_a yielded the imidazole N¹-methylated derivative with no activity, whereas at a pH above its pK_a the resulting imidazole N¹,N³-dimethylated derivative had no activity. No N³-methylated derivative was formed throughout the entire pH range tested. These results indicate that the N-1 position of the imidazole of His-48 is oriented toward the active site cavity, again in agreement with crystallographic data (44). Treatment with tetranitromethane lead to a specific nitration of Tyr-76 (65). When Ca²⁺ was bound to the active site, nitrated Tyr-76 moved to a more polar site with the decrease of pK_a value of its hydroxyl group. When phospholipid was further added, nitrated Tyr-76 moved from a polar environment to a less polar environment with the rise in the pK_a. The environmental polarity of nitrated Tyr-76 changes in the presence or absence of ligands and depends on the kind of ligands that are used. Because Tyr-76 is located in the β-strand region that is separated from the active site cleft by the two helical rods, A and E, it is assumed that conformational change induced by the binding of ligands extends to the region remote from the active site in *T. flavoviridis* PLA₂. Such experiments provide evidence of long-range diffusional coupling between remote sites in the noncooperative globular protein (66).

The tertiary structure of dimeric PLA2 (44) reveals that the side chain of Lys-69 penetrates into the Ca²⁺ binding loop of the opposite subunit, suggesting that Lys-69 also contributes to dimer formation. This view was confirmed by chemical modification experiments with 2,4,6-trinitrobenzene sul-

fonate in the presence or absence of Ca^{2+} (67). These experiments further indicated that dimerization is not necessarily essential for expression of activity (67).

5. REFOLDING

The reduced form of *T. flavoviridis* PLA₂, in which seven disulfide bonds were reduced by dithiothreitol, easily returns to its native form using mild oxidizing conditions (68). In contrast, the reduced N-terminal octapeptide-removed PLA₂ failed to fold to its native form (68), indicating that the N-terminal moiety, which interacts with other portions of the molecule (69), is critical for construction of the native tertiary structure. Reduced [Asp⁴⁹] PLA₂ from *Naja naja atra* venom was also observed to refold and to acquire its native disulfide bonds (70). A similar experiment was carried out with a [Lys⁴⁹]PLA₂ (basic protein I) from *T. flavoviridis* venom (71). The extracellular PLA₂s can refold from the reduced state to a nativelike form.

B. Diverse Physiological Functions of Snake Venom Phospholipase A₂ Isozymes

1. ENZYMATIC ACTIVITY OF PHOSPHOLIPASE A₂ ISOZYMES

Several models have been proposed to explain the functional properties of toxic PLA₂s (72, 73). Their enzymatic activity has been largely studied using the pH stat titration method, the egg yolk assay being recommended as a simple but respected method (74). When three major PLA₂ isozymes of *T. flavoviridis* venom were studied in this assay, there was a prominent difference in the potencies of [Asp⁴⁹]PLA₂ (PLA2) and two [Lys⁴⁹]PLA₂s, basic protein I (BPI) and basic protein II (BPPII) (75). BPI and BPPII showed only 0.2–0.3% of the enzymatic activity of their PLA₂ isozyme (Table I). Using mixed micelles of various synthetic phospholipids and Triton X-100, it has been further confirmed that [Lys⁴⁹]PLA₂s (BPI and BPPII) are enzymatically extremely weak as compared with [Asp⁴⁹]PLA₂ (PLA2) (Table I). This is a rather common feature of [Lys⁴⁹]PLA₂s. Myotoxin II, a [Lys⁴⁹]PLA₂ isolated from *Bothrops asper* venom, was also described to be quite a weak enzyme, with only 0.1% of the PLA₂ activity of an [Asp⁴⁹]PLA₂ called myotoxin III from the same venom (39). [Lys⁴⁹]PLA₂ (App-K49) from *A. piscivorus piscivorus* venom (54, 76–78) exhibited a low specific activity (less than 0.03% of that of [Asp⁴⁹]PLA₂) (76, 77). It was claimed that this residual activity was due to contamination by an active PLA₂.

All snake venom PLA₂ isozymes contain the essential catalytic His-48 and Asp-99 (78), raising the question as to why [Lys⁴⁹]PLA₂s are enzymatically so weak. If they are genuine lipolytic enzymes, their weak enzymatic poten-

TABLE I
 SPECIFIC PLA₂ ACTIVITY OF *T. flavoviridis* [Asp⁴⁹]PLA₂ (PLA2) AND [Lys⁴⁹]PLA₂S
 (BPI AND BPII) AGAINST MIXED MICELLES OF VARIOUS PHOSPHOLIPIDS

| Substrate ^a | Specific activity (mmol min ⁻¹ μmol ⁻¹ of protein) ^b | | |
|------------------------|---|--------------------------|---------------------------|
| | PLA2 | BPI | BPII |
| Egg yolk ^c | 179 ± 17 ^d (100%) | 0.455 ± 0.075 (0.25%) | 0.373 ± 0.146 (0.21%) |
| DLPC | 36.3 ± 1.0 (100%) | 0.116 ± 0.002 (0.32%) | 0.199 ± 0.003 (0.55%) |
| DLPE | 31.0 ± 4.3 (100%) | 0.406 ± 0.044 (1.31%) | 0.844 ± 0.140 (2.72%) |
| DLPG | 2.73 ± 0.54 (100%) | 0.152 ± 0.025 (5.57%) | 0.307 ± 0.021 (11.25%) |
| ASPC | 11.0 ± 0.90 (100%) | 0.073 ± 0.004 (0.66%) | 0.097 ± 0.007 (0.88%) |

^aDLPC, 1,2-Dilauroyl-L-3-phosphatidylcholine; DLPE, 1,2-dilauroyl-L-3-phosphatidylethanolamine; DLPG, 1,2-dilauroyl-L-3-phosphatidylglycerol; ASPC, 2-arachidonoyl-1-stearoyl-L-3-phosphatidylcholine.

^bThe specific activity values of BPI and BPII as percentages of the corresponding PLA2 values are shown in parentheses.

^cMixed micelles were prepared with Triton X-100 (deoxycholate for egg yolk) in a 3:2 molar ratio in 50 mM Tris-HCl (pH 8.0) containing 10 mM Ca²⁺.

^dResults are means ± SE.

cy must be attributable to inadequate selection of substrates and/or analytical procedures in the assay. The rate of hydrolysis of phospholipids usually increases by several orders of magnitude on passing from monomolecularly dispersed form to micellar state, due to the presence of the water-micelle interface recognition site (79). Kinetics analyses were thus performed with a number of [Asp⁴⁹]PLA₂s (67), including *T. flavoviridis* [Asp⁴⁹]PLA₂ (PLA2) (80). *Trimeresurus flavoviridis* [Lys⁴⁹]PLA₂s (BPI and BPII) are weak at hydrolyzing micellar substrates and are almost completely inactive toward monomeric phospholipids (81). However, studies on the contraction of the smooth muscle preparation of guinea pig ileum have suggested that [Lys⁴⁹]PLA₂s liberate arachidonate from tissue membranes (81). When assayed with artificial liposomes made of 2-arachidonoyl-1-stearoyl-L-3-phosphatidylcholine (ASPC), BPII was immediately identified as a potent enzyme that hydrolyzes the C-2 ester bond (81). Although BPI and BPII liberated only traces of arachidonic acid from monomeric ASPC (Fig. 4A), there was a distinct liberation of arachidonic acid (20 and 45% after 30 min, respectively) as detected by high-performance liquid chromatography (HPLC) (Fig.

4B). After 3 hr, the liberation reached about 30% for BPI and 75% for BPII. Other evidence that BPI and BPII are membrane-acting lipolytic enzymes was obtained from the experiments using ASPC liposomes encapsulating carboxyfluorescein. The dye leakage was observed in a concentration-dependent manner for BPI or BPII (Fig. 5). This leakage was judged to be closely related to the lipolytic activity of $[\text{Lys}^{49}]\text{PLA}_2$ s, in that the levels of dye leakage and of the release of arachidonic acid coincided well with each other when compared at the same enzyme concentration. Without Ca^{2+} , no hydrolysis and no dye leakage occurred for either $[\text{Asp}^{49}]\text{PLA}_2$ or $[\text{Lys}^{49}]\text{PLA}_2$ s, as shown in Fig. 6. $[\text{Lys}^{49}]\text{PLA}_2$ s are therefore genuine Ca^{2+} -dependent lipolytic enzymes acting on phospholipid membranes.

Myotoxin II, a $[\text{Lys}^{49}]\text{PLA}_2$, is strongly myotoxic but phospholipase activity seems not to be necessary for expression of myotoxicity (39). However, myotoxin II is quite active at hydrolyzing ASPC liposomes (M. Ohno *et al.*,

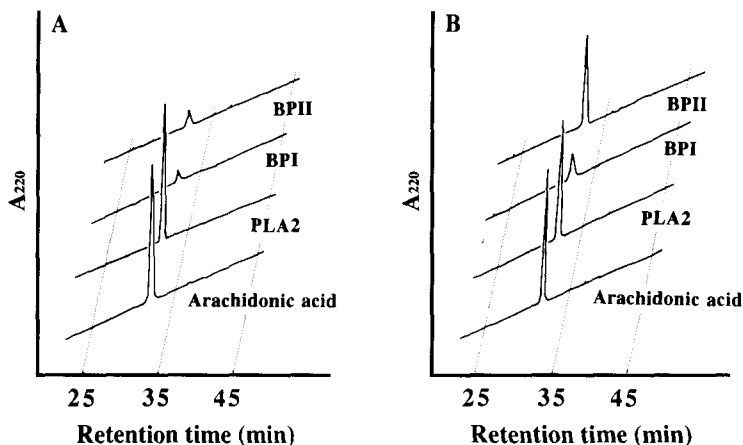


FIG. 4. HPLC elution profiles of arachidonic acid released from 2-arachidonoyl phosphatidylcholine monomers and liposomes incubated with *T. flavoviridis* $[\text{Asp}^{49}]\text{PLA}_2$ (PLA2) and $[\text{Lys}^{49}]\text{PLA}_2$ s (BPI and BPII). (A) PLA_2 (1.0×10^{-6} M final concentration) in 10 mM Tris-HCl containing 10 mM Ca^{2+} (pH 8.0, 20 μl) was added to 1.7 mM (final) 2-arachidonoyl phosphatidylcholine (ASPC) in ether-ethanol (95:5 v/v, 200 μl) and incubated for 5 hr at 25°C. After evaporation, the residue was dissolved in ethanol (150 μl) and the solution was centrifuged (12,000 rpm for 5 min). The supernatant (100 μl) was analyzed by HPLC. (B) PLA_2 (1.0×10^{-5} M final) and ASPC liposomes (1.5 mM) were incubated in 10 mM Tris-HCl (pH 8.0, 220 μl) containing 10 mM CaCl_2 and 100 mM NaCl for 30 min at 37°C. HPLC elution conditions were as follows: column, TSK-gel ODS-120T (46 \times 250 mm); gradient, linear gradient of acetonitrile (50% for initial 5 min, 50–85% for 20 min, and then 85–90% for 30 min) in 0.1% trifluoroacetic acid; flow rate, 1 ml/min.

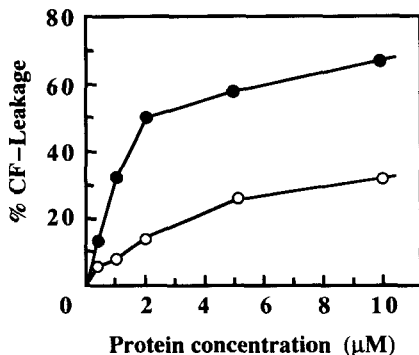


FIG. 5. Comparison of concentration-dependent dye-leakage activity between *T. flavoviridis* [Lys⁴⁹]PLA₂s (BPI and BPII). Leakage experiments were carried out by using 4.4×10^{-5} M ASPC liposomes encapsulating carboxyfluorescein: $\lambda_{\text{ex}} = 470$ nm and $\lambda_{\text{em}} = 515$ nm. Incubation was conducted in 10 mM Tris-HCl (pH 8.0) containing 10 mM CaCl₂ and 100 mM NaCl at 37°C. ●, BPII; ○, BPI.

unpublished data). [Lys⁴⁹] PLA₂ from *Agkistrodon piscivorus piscivorus* reported to be inactive as PLA₂ was active on ASPC liposomes. The enzymatic activity of PLA₂s probably should be reassessed more carefully because it is especially important for the evaluation of the physiological functions of

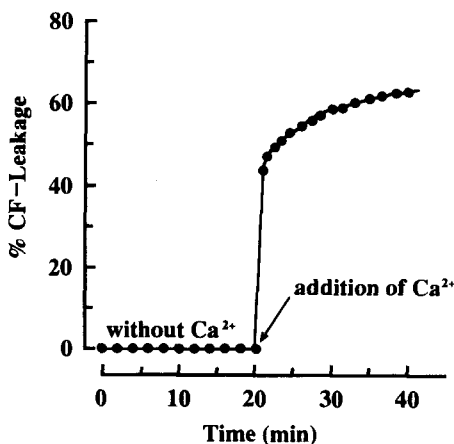


FIG. 6. Effect of Ca²⁺ on dye-leakage activity of *T. flavoviridis* [Lys⁴⁹]PLA₂ (BPII). Leakage experiments were initiated by using 4.4×10^{-5} M 2-arachidonoyl phosphatidylcholine liposomes in the absence of Ca²⁺. Other conditions were as described in the legend for Fig. 5. After 20 min incubation without Ca²⁺, 1.0 M CaCl₂ (20 µl) was added to the reaction mixture (final concentration of 10 mM).

PLA₂s. It is not yet clear whether the enzyme activity of [Lys⁴⁹]PLA₂s is involved in myotoxicity.

Recently, we isolated an edema-inducing PLA₂ and a neurotoxic PLA₂ from *T. flavoviridis* venom; both are basic [Asp⁴⁹]PLA₂s. Though their lipolytic activities were weak (10–30%) in the egg yolk assay, as compared with ordinary [Asp⁴⁹]PLA₂s, they are potent at hydrolyzing ASPC in micelles and liposomes (M. Ohno *et al.* unpublished data).

2. PHYSIOLOGICAL EFFECTS

Snake venom PLA₂s exhibit a variety of physiological activities, including cardiotoxicity (82, 83), presynaptic neurotoxicity (84, 85), myotoxicity (86, 87), and platelet (88) and anticoagulant effects (89). Kini and Evans (73) have postulated that after initial binding to their target site, PLA₂s induce various pharmacological effects by mechanisms that are either dependent or independent of enzymatic activity. Various group II PLA₂s have been found in mammalian cells and tissues (26, 90, 91). Mammalian group II PLA₂s appear to be involved not only in regulation of eicosanoid generation in various normal cells and tissues (91, 92), but also in inflammation (32, 90).

The specific activities of *T. flavoviridis* PLA₂ isozymes show that each isozyme possesses a specific function and that PLA₂ isozymes form a family of multiple functions for efficient expression of venom toxicity. When three PLA₂ isozymes of *T. flavoviridis*, PLA2, BPI, and BPII, were injected intramuscularly in the left hind thigh of mice, all induced similar histological changes of myonecrosis in muscular fibers around the injection site (93). Other changes observed in mouse skeletal muscles were infiltration of polymorphonuclear cells and edema in stromal tissues. No particular difference in myotoxicity was apparent among these three PLA₂ isozymes, and their myotoxicity was reported to be associated with their enzymatic activity (94). Indeed, when the active site His-49 was alkylated with *p*-bromophenacyl bromide, PLA2 was devoid of both enzymatic and myolytic activities (95). On the other hand, it is now evident, as described above, that [Lys⁴⁹]PLA₂s are membrane-acting lipolytic enzymes. It is important to clarify how the enzymatic activity of PLA₂ isozymes contributes to their myotoxicity.

Assessment of the creatine kinase (CK) level, an index for myolysis, in mouse plasma following intramuscular injection of PLA2, BPI, and BPII revealed distinctly different responses (93). The CK levels induced by BPI and BPII were several times greater than that caused by PLA2, whereas they were comparable with that provoked by the crude venom (Fig. 7). If CK release is associated with myotoxicity of PLA₂s, a difference in the CK levels would reflect a difference in mechanisms. This possibility was supported by the observation that BPI and BPII are capable of depolarizing muscle membranes

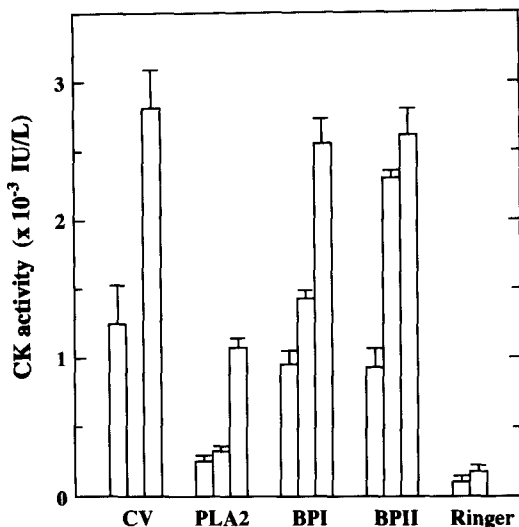


FIG. 7. Effects of [Asp⁴⁹]PLA₂ (PLA2), [Lys⁴⁹]PLA₂s (BPI and BPII), and crude *T. flavoviridis* venom (CV) on plasma creatine kinase (CK) level in mice. Each sample was injected at doses of 0.5, 1.0, and 2.0 µg/g (the crude venom was given at doses of 0.5 and 2.0 µg/g only) intramuscularly into mice. The three bars for PLA2, BPI, and BPII represent the mean CK levels of a group of four mice 1 hr after injection. Results are shown as mean ± S.E. CK values before and after injection of physiological saline (Ringer) are also shown.

whereas PLA2 is not, even at high doses (Fig. 8). It should be noted that BPI and BPII produced no depolarization in the absence of Ca²⁺, suggesting that their enzymatic activity is a major factor in the depolarization of muscle membranes.

Trimeresurus flavoviridis [Asp⁴⁹]PLA₂ (PLA2) induced contraction of smooth muscle preparations of guinea pig ileum and artery (96), primarily due to the capacity of PLA2 to liberate arachidonic acid from phospholipid membranes. A similar effect was observed in the smooth muscle of the rat stomach (97).

When BPI and BPII were examined in the guinea pig ileum, a clear activity difference was observed (81). BPII was 10–100 times more potent than BPI at causing guinea pig ileum contraction (Fig. 9). This might reflect the observation that BPII is more potent than BPI with respect to dye leakage and arachidonic acid liberation in assays using ASPC liposomes (81). Because the amino acid sequences of BPI and BPII are identical except that Asp-67 in BPI is replaced by Asn in BPII, it is likely that the activity differ-

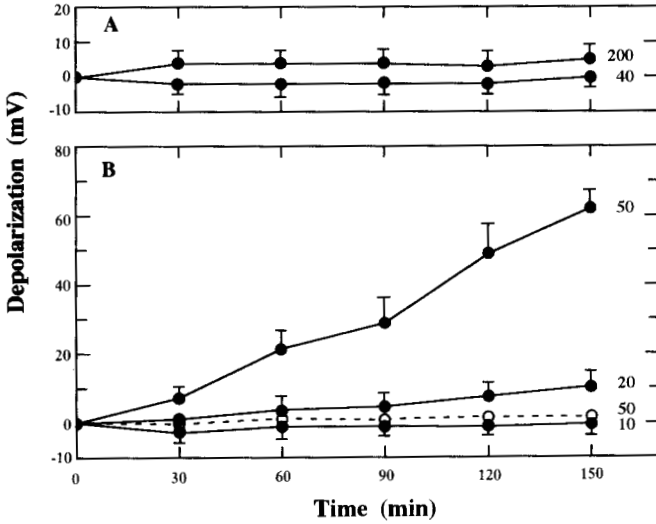


FIG. 8. Time-dependent changes of magnitudes of depolarization induced by (A) [Asp⁴⁹]PLA₂ (PLA₂) and (B) [Lys⁴⁹]PLA₂ (BPI). Numbers to the right of the curves indicate the concentrations of the enzymes (μg/ml). Each point represents the mean of 40 individual fibers from three muscle preparations; vertical bars indicate the standard deviations. Membrane potentials were measured in the presence (●) and absence (○) of 1.5 mM Ca²⁺.

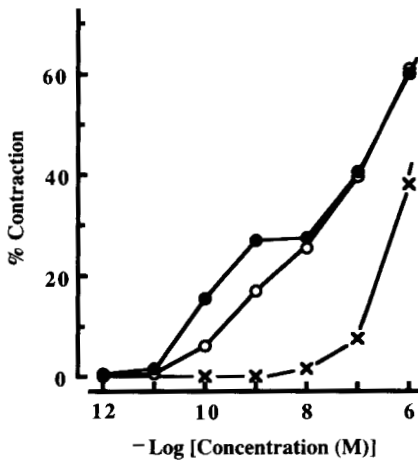


FIG. 9. Concentration-dependent contraction activity of *T. flavoviridis* PLA₂ isozymes on the smooth muscle of guinea pig ileum. The activities are expressed as a percentage of the maximum contraction induced by 1.0 × 10⁻⁵ M carbachol. (●) PLA₂; (○) BPII; and (X) BPI.

ence between BPI and BPII is due to the different ability of residue 67 to interact with phospholipid bilayers.

As mentioned, *T. flavoviridis* venom contains three major PLA₂s, one [Asp⁴⁹]PLA₂ and two [Lys⁴⁹]PLA₂s. Recently, we found two other basic [Asp⁴⁹]PLA₂ isozymes in the same venom. One of them strongly induces edema in mice whereas the other is neurotoxic. Each PLA₂ isozyme clearly has an independent functional and physiological role.

C. Structures of cDNAs and Genes Encoding Viperidae Snake Venom Gland Phospholipase A₂ Isozymes

1. STRUCTURES OF cDNAs

Snake venom PLA₂s from group II are classified into two subgroups, [Asp⁴⁹]PLA₂s and [Lys⁴⁹]PLA₂s, with high and low lipolytic activity, respectively (24). Only viperid snake venom contains [Lys⁴⁹]PLA₂s. [Asp⁴⁹]PLA₂s and [Lys⁴⁹]PLA₂s are homologous, with more than 50% identity in amino acid sequence and similar tertiary structures (26, 37, 42–44, 51, 53, 54, 59, 98, 99). The cDNAs and genes encoding a number of venom gland PLA₂ isozymes of viperid and elapid snakes have been cloned and sequenced (Table II).

Five cDNAs encoding *T. flavoviridis* venom gland PLA₂ isozymes were cloned and their nucleotide sequences were determined (24, 100) (Fig. 10). Northern blot analysis showed that the corresponding mRNAs occur in venom glands but not in other tissues (24, 100). These cDNAs contain an open reading frame of 414 nucleotides coding for 138 amino acid residues. The 5' and 3' untranslated regions (UTRs) are more homologous than the protein-coding region, with nucleotide identities of 98% for the 5' UTR, 67% for the protein-coding region, and 89% for the 3' UTR. Such greater sequence homology in the UTRs as compared to the protein-coding region has not been reported in cDNAs for other isozyme systems. Usually, lower homology is observed in the UTRs, as in various isoforms of the G-protein α -subunit family (125, 126) and the protein kinase C family (127, 128). The signal peptide-coding domain encoding 16 amino acid residues is exceptionally conserved when compared with other portions of the protein-coding region. The high conservation may suggest that the signal peptides share common roles in membrane penetration and susceptibility to a common signal peptidase. Examination of the nucleotide sequences of the protein-coding region revealed that substitutions have occurred in the first, second, and third positions of the triplet codons at similar rates: 32.1, 30.0, and 28.6%, respectively. This seems unusual compared with other cases in which the silent (third) sites of codons

are more variable than the first and second positions, whose mutations cause amino acid change (129). Presumably, *T. flavoviridis* venom gland PLA₂ isozymes have evolved with low evolutionary constraint.

2. STRUCTURES OF GENES

To better understand the above-mentioned evolutionary phenomena, six genes encoding *T. flavoviridis* venom gland PLA₂ isozymes were isolated from its liver genomic DNA library and sequenced (101). Four genes (*pgPLA 1a*, *pgPLA 2a*, *pgPLA 1b*, and *pgPLA 2b*) encoded [Asp⁴⁹]PLA₂ isozymes and two genes (*BP-I* and *BP-II*) encoded [Lys⁴⁹]PLA₂ isozymes (24) (Fig. 10). The transcription initiation site was assigned to the adenosine residue located 204 nucleotides upstream from the translation initiation codon (ATG) and the TATA-like sequence (CATAAA) was found 34 nucleotides upstream from the transcription initiation site. Each gene spanned 1.9 kb, contained four exons and three introns, and encoded 138 amino acid residues, including the signal sequence of 16 amino acid residues (Fig. 11). Four genes encoding *T. gramineus* venom gland PLA₂ isozymes were also isolated from a liver genomic DNA library and sequenced (22). Genes *gPLA₂-I* and *gPLA₂-VI* encoded [Asp⁴⁹]PLA₂ isozymes and genes *gPLA₂-V* and *PLA₂-VII* encoded [Lys⁴⁹]PLA₂ isozymes. Again, all genes spanned 1.9 kb, contained four exons and three introns, and encoded 138 amino acid residues, including the highly conserved signal sequence of 16 amino acid residues. The exon–intron structures of *T. gramineus* and *T. flavoviridis* PLA₂ isozyme genes were similar.

The genes encoding human pancreatic [Asp⁴⁹]PLA₂ (130) and mojave toxin (heterodimeric [Asp⁴⁹]PLA₂s) from *C. scutulatus scutulatus* venom (108) also contain four exons. However, the genes coding for human rheumatoid synovial fluid (nonpancreatic) [Asp⁴⁹]PLA₂ (32, 33) and ammodytoxin C from *Vipera ammodytes ammodytes* venom (115) have five exons. The highly conserved GT (donor site)/AG (acceptor site) structure commonly noted for introns (131) was seen in all cases. When the nucleotide sequences of *T. flavoviridis* and *T. gramineus* PLA₂ isozyme genes (22, 101) and other snake venom gland PLA₂ isozyme genes (108–115) and those of human pancreatic (130) and nonpancreatic PLA₂ genes (32, 33) were compared by taking account of their amino acid sequences, it became evident that the exon–intron junction sites are completely conserved among these genes (26).

The nucleotide sequences and the derived amino acid sequences of the first and second exons of four *T. gramineus* and six *T. flavoviridis* PLA₂ isozyme genes are shown in Fig. 12 (22). The first intron is clearly more homologous than the second exon. This is true for other introns and the protein-coding regions of other exons except for the signal sequence-coding domain of the first exon. For example, sectional homologies between *pgPLA 1a*

TABLE II
cDNAs AND GENES ENCODING SNAKE VENOM GLAND PLA₂ ISOZYMES
CLONED AND SEQUENCED

| Species | PLA2 | Accession number | | Ref. |
|----------------------------------|-------------------|------------------|---------------|--|
| | | cDNA | Gene | |
| Group II | | | | |
| <i>Trimeresurus</i> | | | | |
| <i>T. flavoviridis</i> | T.fla PLA2 | D10070 | D10722/D10724 | (24, 101) |
| | T.fla T37PL | D10720 | — | — |
| | T.fla PLX' | D10721 | — | (100) |
| | T.fla PL1b/2b | — | D10723/D10725 | — |
| | T.fla BPI | D10718 | D13383 | — |
| | T.fla BPII | D10719 | D13384 | — |
| | T.fla PLA2-I | D31774 | D31780 | (22) |
| <i>T. gramineus</i> | T.gra PLA2-V | D31775 | D31781 | — |
| | T.gra PLA2-VI | — | D31778 | — |
| | T.gra PLA2-VII | — | D31779 | — |
| <i>T. okinavensis</i> | T.oki D49/1 | D49388 | D49390 | (102) |
| | T.oki D49/2 | — | D49392 | — |
| | T.oki K49/3 | D49389 | D49391 | — |
| <i>T. mucrosquamatus</i> | T.muc PLA2 | X77088 | — | (103) |
| | T.muc Neu | X77645 | — | — |
| | T.muc Ant | X77646 | — | — |
| | T.muc PLHOM | X77647 | — | — |
| <i>Bothrops</i> | | | | |
| <i>B. jararaca</i> | B.jar BJPLA | X76289 | — | (104) |
| <i>Deinagkistrodon</i> | | | | |
| <i>D. acutus</i> | D.acu PLA2 | X77650 | — | — |
| | D.acu Neu | X77649 | — | — |
| | D.acu Ant | X77648 | — | — |
| | D.acu PLHOM | X77651 | — | — |
| <i>Agkistrodon</i> | | | | |
| <i>A. contortrix laticinctus</i> | A.con K49 | U21335 | — | (105) |
| <i>Crotalus</i> | | | | |
| <i>C. durissus terrificus</i> | C.dur crotoxin A | X12606 | — | (106, 107) |
| | C.dur crotoxin B1 | X12603 | — | — |
| | C.dur crotoxin B2 | X16100 | — | — |
| <i>C. scutulatus scutulatus</i> | C.scu MTXa | — | U01027 | (108) |
| | C.scu MTXb | — | U01026 | — |
| | C.scu psiMTX | — | U23186 | T. R. John and I. I. Kaiser, unpub- lished (1995) |
| <i>Vipera</i> | | | | |
| <i>V. ammodytes ammodytes</i> | V.amm atoxA | X53471 | — | (109) |
| | V.amm atoxB | X52241 | — | (110) |
| | V.amm atoxC | X15138 | X76731 | (111, 114, 115) |
| | V.amm atinI | X56878 | — | (112) |
| | V.amm atinL | X53036 | — | (113) |
| <i>V. russelli formosensis</i> | V.rus RV-7B | X68386 | — | (57) |
| | V.rus RV-4A | X68385 | — | — |

(continued)

TABLE II (Continued)

| Species | PLA ₂ | Accession number | | Ref. |
|-----------------------------|------------------|------------------|------|---|
| | | cDNA | Gene | |
| Group I | | | | |
| <i>Aipysurus</i> | | | | |
| <i>A. laevis</i> | A.lae PLA2 | X12604 | — | (116) |
| <i>Laticauda</i> | | | | |
| <i>L. laticaudata</i> | L.lat PLA2R | Y00377 | — | (117) |
| <i>Bungarus</i> | | | | |
| <i>B. multicinctus</i> | B.mul PLA2 | X53406 | — | (118) |
| | β-Bun A2 | X53407 | — | (119) |
| | B.mul PLA2A4 | X53408 | — | (120) |
| | β-Bun A5 | Z54225 | — | (121) |
| | β-Bun A6 | X98205 | — | L. Chang and P. Wu, un- published (1996) |
| <i>Notechis</i> | | | | |
| <i>N. scutatus scutatus</i> | N.scu PLA2 | X12605 | — | (122) |
| | N.scu PLA242 | X14043 | — | F. Ducancel, unpublished (1989) |
| <i>Naja</i> | | | | |
| <i>N. naja atra</i> | N.naj PHOA2 | X73225 | — | (123) |
| | N.naj PLA2 | X77755 | — | (124) |
| <i>N. naja sputatrix</i> | N.spu PLA2/1 | L42006 | — | K. Jeyaseelan <i>et al.</i> , un- published (1995) |
| | N.spu PLA2/2 | L42004 | — | |
| | N.spu PLA2/3 | L42005 | — | — |

and *BP-I* are 98.5% for the 5' UTR and 97.5% for the signal peptide-coding domain (first exon), 93.8% (first intron), 67.7% (second exon), 92.8% (second intron), 82.1% (third exon), 96.9% (third intron), and 75.2% for the protein-coding region and 91.5% for the 3' UTR (fourth exon). It is generally accepted that the evolutionary rates of introns are greater than those of the protein-coding regions (or exons) (132), so that the structural features of *T. flavoviridis* and *T. gramineus* PLA₂ isozyme genes are not in accord with those of general (nonvenomous) genes.

D. Darwinian-Type Accelerated Evolution of Viperidae Snake Venom Gland Phospholipase A₂ Isozymes

The evolutionary significance associated with the nucleotide sequences of *T. flavoviridis* and *T. gramineus* venom gland PLA₂ isozyme genes was analyzed. For each of the pairwise comparisons of the nucleotide sequences of

BPII

AGCCCAAGTCTCCA

BPII

GTACTTTTGTCTACTAATAAAGAGAAGCTGCTAACTTCAAGCTGAAGCATGGCCGAAAGAAGGCCATTA

T-PL

ATTTTCCCTGCCCGGC

BPI

BPII

PLX'

GTGACCTCGAGGAGCAGAGGAGCCTGACAGGTGTGAATAGCCACATCGTTG C.....C.....T...
.....C.....

T-PL

BPI

BPII

PLX'

TTCCTCTTCTGATCCTTGCCCTACAGGTTATCCTTGACTTACAACCGTTTGTTTAGTGACTGTTCTAAGGAC
.....C.....C.....
.....T.....CA.....

PLA2

T-PL

BPI

BPII

PLX'

5'UT →
CTTTTCACCAGCGGAGGCAATTAACGGGGTCTGTCCATTCCCAGGTCTGGATTCCGGGAGG
CATTTCAGAA.....G.....
.....G.....
.....G.....

← coding region

signal sequence →

PLA2

T-PL

BPI

BPII

PLX'

ATG AGG ACT CTC TGG ATA ATG GCC GTG TTG CTG GTG GGC GTC GAT GGG GGC CTG
Met Arg Thr Leu Trp Ile Met Ala Val Leu Leu Val Gly Val Asp Gly Gly Leu
.....
..... C.. .. A.. ..
..... Leu .. Ser ..
..... .. A.. ..
..... Ser ..
..... T .. T .. G .. CA. ...
..... Glu .. His ..

PLA2

T-PL

BPI

BPII

PLX'

TGG CAA TTC GAG AAT ATG ATC ATT AAA GTG GTG AAG AAA AGC GGT ATA CTT TCG
Trp Gln Phe Glu Asn Met Ile Ile Lys Val Val Lys Lys Ser Gly Ile Leu Ser
.....
GTCG TG. ..G T.C C.. ..A. AC. GG. ... GAA .C. GCT AAA AAC
Val Leu Trp Lys Phe Gln Glu Thr Gly Glu Ala Ala Lys Asn
GTCG TG. ..G T.C C.. ..A. AC. GG. ... GAA .C. GCT AAA AAC
Val Leu Trp Lys Phe Gln Glu Thr Gly Glu Ala Ala Lys Asn
CT.AG. ..GAGA. AC. GGA ... GAG CC. ..T G.G ..C
Leu Arg Lys Lys Met Thr Gly Glu Pro Val

PLA2

T-PL

BPI

BPII

PLX'

TAC AGT GCT TAC GGA TGC TAC TGC GGC TGG GGG GGC CGA GGC AAG CCA AAG GAC
Tyr Ser Ala Tyr Gly Cys Tyr Cys Gly Trp Gly Gly Arg Gly Lys Pro Lys Asp
.....
..... G.C TTA A.. .. GT. ... A.G
Gly Leu Asn Val Arg
..... G.C TTA A.. .. GT. ... A.G
Gly Leu Asn Val Arg
..... GCC TT. AA.
Ala Phe Lys

PLA2

T-PL

BPI

BPII

PLX'

GCC ACC GAC CGC TGC TGT TTT GTG CAC GAC GTC TGC TGT TAC GGA AAA GTG ACC GGC
Ala Thr Asp Arg Cys Cys Phe Val His Asp Cys Cys Tyr Gly Lys Val Thr Gly
.....
.....T ..TA.. ..C ..A. Lys .. AA.
..... Ser Tyr Lys Lys
.....T ..TA.. ..C ..A. Lys .. AA.
..... Ser Tyr Lys Lys
..... ..CA.

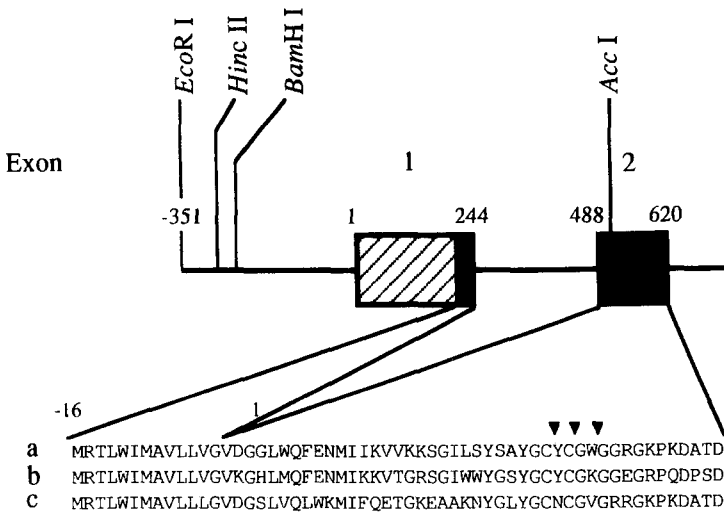


FIG. 11. Schematic representation of a common structure of six *T. flavoviridis* venom gland PLA_2 isozyme genes, *pgPLA 1a*, *pgPLA 2a*, *pgPLA 1b*, *pgPLA 2b*, *BP-I*, and *BP-II*, and the amino acid sequences corresponding to the coding regions of *pgPLA 2a* (line a), *pgPLA 1b* (line b), and *BP-I* (line c). Four exons are indicated by boxes in which the UTRs are hatched. The nucleotide position numbers represent those for *pgPLA 1a* and *pgPLA 2a*. The four coding regions and their corresponding amino acid sequences are correlated by lines. ▼, Residues involved in the Ca^{2+} binding site; ▽, residues involved in the catalytic site; *, the residue at position 58 (67 according to the aligned numbering).

these genes, the numbers of nucleotide substitutions per site (K_N) for the non-coding regions, including introns, and the numbers of nucleotide substitutions per synonymous site (K_S) and per nonsynonymous site (K_A) for the protein-coding regions were computed as described (133 or 134) with correlations for multiple substitutions (132). A synonymous site is a site of a codon at which base substitution causes no amino acid change. A nonsynonymous site is a site of a codon at which base substitution causes an amino acid change. Table III shows the values for three pairs of *T. flavoviridis* PLA_2 isozyme genes, revealing several features. First, as in the 5' and 3' UTRs, K_N values for introns are approximately one-fourth of K_S values for all the pairs of the genes compared. This indicates that the introns are unusually conserved as compared to the protein-coding regions. The high homology of introns may suggest that there are some functionally important constraints in the introns. However, the regions corresponding to introns of precursor RNAs involved no significant secondary structure when analyzed by the GENAS system (135) according to the method of Zuker and Steigler (136) (data not shown). Thus, it may be reasonable to consider that the introns have no significant role. The absence of an apparent role for the introns suggests

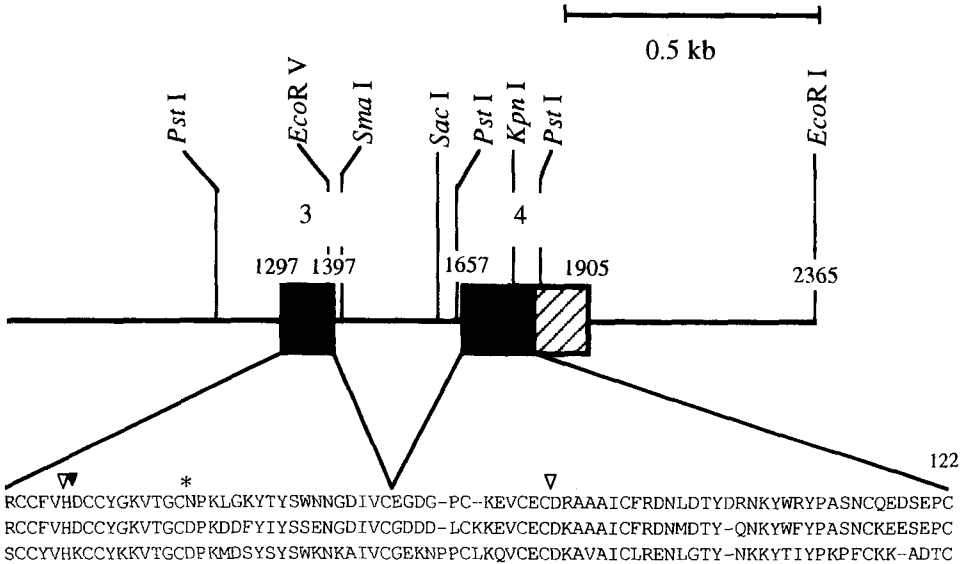


FIG. 11. (Continued)

that the protein-coding regions of the exons have evolved at greater substitution rates than the introns. Second, K_A/K_S values are greater than those reported for other isoprotein genes, which are estimated to be about 0.2 (129, 132). Though synonymous sites are more variable than nonsynonymous sites because of less functional constraint (129, 132), this is not the case in the protein-coding regions of *T. flavoviridis* PLA₂ isozyme genes. The K_A/K_S values of the protein-coding regions for pair of *pgPLA 1a* and *pgPLA 1b* or *pgPLA 2b*, and of *pgPLA 1a* and *BP-I* or *BP-II*, are close to 1, and those for pairs of *pgPLA 1b* and *BP-I* or *BP-II* are about 1.8. Such a high degree of substitutions in nonsynonymous sites suggests that *T. flavoviridis* venom gland PLA₂ isozyme genes have evolved to bring about accelerated amino acid substitutions.

Table IV shows the K_N/K_S and K_A/K_S values for pairs of PLA₂ isozyme genes when compared within species (*T. gramineus* vs. *T. gramineus*) and between species (*T. gramineus* vs. *T. flavoviridis*) (22). The values computed for pairs within *T. gramineus* PLA₂ isozyme genes are similar to those obtained for pairs within *T. flavoviridis* PLA₂ isozyme genes. Comparison between two species gave similar values, although the exceptionally low K_A/K_S value (0.253) obtained for the pair *gPLA₂-VI* and *pgPLA 1b* remains unclear.

FIG. 12. Nucleotide sequences of introns 1 and exons 2 of *T. gramineus* (*gPLA2-I*, *gPLA2-V*, *gPLA2-VI*, and *gPLA2-VII*) and *T. flavoviridis* venom gland PLA₂ isozyme genes (*pgPLA 1a*, *pgPLA 2a*, *pgPLA 1b*, *pgPLA 2b*, *BP-I*, and *BP-II*) with the predicted amino acid sequences for exons 2. The nucleotide sequences of exons are capitalized.

gPLA2-I 229 gtgagtgaagcaaaactgtaaaatgggcagctaattc-tgctccctttgcagaaggtgaaacggaggg
gPLA2-V 223 ..a.....t.....
gPLA2-VI 229t.....c.....a..
gPLA2-VII 234c.....t.....
pgPLA1a 245t.....
pgPLA2a 245t.....
pgPLA1b 245t.....g...t.
pgPLA2b 245t.....g.....
BP-I 245t.....g.....
BP-II 245t.....g.....

298 gttacaggttgctttggctttagtggccgacatcgg-acagg-aaattcaccgttaagcagggcatcgtgtcaccacat
298a.....a.....g.....g.....a.....c.....
301c.....a.....g.....g.....c.....g..
291a.....a.....g.....g.....a.....c.....
312a.....a.....g.....g.....
312a.....a.....g.....g.....
312t.....t.....g...aa.....
312t.....t.....g...aa.....
311a.g.....a.....t.g.....t.....
311a.g.....a.....t.g.....t.....g.....

376 cttcctttgctgtggtcgttaag-cgaggactgccagcatctgccattaaccctacagagaagccgagggaggctgacac
378a.....
380c.....g.....t.....
371g.....a.....
392 .a.g.....a.....a.....a.....
392 .a.g.....a.....a.....a.....
392 .acg.....t.....a.....a.....
392 .acg.....t.....a.....a.....
391 ...g...c.....t.....g.....t.....
391 ...g...c.....t.....g.....t.....

455 tttctgtctttttccagTCGAGGGGCACCTGATGCAATTGAGACACTGATCATGAAAGTGCGGGGAGAAGCGGTGTTT
 alGluGlyHisLueMetGlnPheGluThrLeuIleMetLysValAlaGlyArgSerGlyValT
 457T.....AG.G...CG...G.G..AGA....T.CC...A.AT...A..A..C.C..A
 * * * SerValIleGluLeuGlyLysMet * PheGlnGluThe * LysAsnProAlaT
 459ATA.....A.....A.....G.....A...
 • * * * * * * * * * AsnMet * Lys • * Thr * • * • Ile
 451 -----T.....GG.G...CG...GAC..AGA....G.CC...A.AT...A..A..C.C..A
 • • • GlyValIleGluLeuThrLysMet * ValGlnGluMet * LysAsnAlaLeuT
 471a.....GG...TG.....C...ATA.....T.....T.AA..A.....A.AC
 * * * Gly * Trp • * * AsnMet * Ile * * ValLysLys • • IleL
 471a.....GG...TG.....C...ATA.....T.....T.AA..A.....A.AC
 * Asp * Gly * Trp * * * AsnMet * Ile * * ValLysLys * * IleL
 471A.....C...ATA.....A.....A.....G..T...A...
 * Lys • • • • • * * AsnMet * Lys * * Thr * * * • Ile
 471A.....C...ATA.....A.....A.....G..T...A...
 * Lys * * • • • • • * AsnMet * Lys * • Thr * • • * • Ile
 470c.....T...AG...G.C...GTG..AGA....T.CC...A.A.....A.GAA.C..C A
 • Asp * Ser * Val * LeuTrpLysMet * PheGlnGluThr * LysGluAlaAlaL
 470c.....T...AG...G.C...GTG..AGA....T.CC...A.A.....A.GAA.C..C A
 • Asp • Ser • Val * LeuTrpLysMet • PheGlnGluThr * LysGluAlaAlaL

 535 GGTACTACGGCTCTTACGGATGCTTCTCGGCGCGGGGGGCCAAGGCCGGCCACAGGACGCCAGTGACCG
 rpTyrTyrGlySerTyrGlyCysPheCysGlyAlaGlyGlyGlnGlyArgProGlnAspAlaSerAspAr
 537 CA.C.....CT...T.....AA...T...C.....G.C.TAA...A.....CC....
 hrSer * * Leu • * • Asn * * Pro * * ArgArgLys * Lys * * Thr * * *
 539 ...GG.....T..C.....A.....AA.....T...C.....T...C.....T..
 * Trp * * * * * Tyr * * Lys * * * * * Leu * * • * * * * •
 527 CA.C...A. CT...T.....AA...T...C.....G.C.TAA...AT.....CC...A.
 hrSer • SerLeu • * • Asn * * Pro * * ArgArgLys * Lys * * Thr * Se
 551 TT.CG...A.TG.....A.....TG.....G...AA...A.....CC....
 euSer * SerAla * * • Tyr • • Trp * * Arg * Lys * Lys * * Thr * *
 551 TT.CG...A.TG.....A.....TG.....G...AA...A.....CC....
 euSer * SerAla • * * Tyr * * Trp • • Arg • Lys * Lys * * Thr * *
 551 ...GG.....A.....AA...A.TG.....C.....
 * Trp * • * • * * Tyr * * Lys * * Glu * * * * * Pro * * •
 551 ...GG.....A.....AA...A.TG.....C.....
 * Trp * * * * * Tyr * * Lys * * Glu * * * * * Pro * * *
 550 AAA.....TA.....AA.....T...A.G.G...AA...A.....T.C...A.
 ysAsn * * Leu * * * Asn • • Val * ArgArg * Lys * Lys • • Thr * Se
 550 AAA.....TA.....AA.....T...A.G.G...AA...A.....T.C...A.
 ysAsn • • Leu * * * Asn * * Val * ArgArg * Lys * Lys * * Thr * Se

TABLE III
 K_N/K_S AND K_A/K_S VALUES BETWEEN PAIRS OF *T. flavoviridis* VENOM GLAND PLA₂ ISOZYME GENES

| Region | <i>pgPLA 1a</i> vs. <i>pgPLA 1b</i> | | | <i>pgPLA 1a</i> vs. <i>BP-1</i> | | | <i>pgPLA 1b</i> vs. <i>BP-1</i> | | |
|------------------------------------|-------------------------------------|-------|-----------|---------------------------------|-------|-----------|---------------------------------|-------|-----------|
| | K_N | K_S | K_N/K_S | K_N | K_S | K_N/K_S | K_N | K_S | K_N/K_S |
| 5' flanking region | 0.051 | 0.224 | 0.228 | 0.035 | 0.253 | 0.138 | 0.029 | 0.183 | 0.158 |
| 5' UTR | 0.020 | 0.224 | 0.089 | 0.015 | 0.253 | 0.059 | 0.025 | 0.183 | 0.137 |
| Intron 1 | 0.042 | 0.224 | 0.186 | 0.064 | 0.253 | 0.253 | 0.069 | 0.183 | 0.377 |
| Intron 2 | 0.084 | 0.224 | 0.375 | 0.076 | 0.253 | 0.300 | 0.058 | 0.183 | 0.317 |
| Intron 3 | 0.032 | 0.224 | 0.143 | 0.032 | 0.253 | 0.126 | 0.032 | 0.183 | 0.175 |
| All introns | 0.063 | 0.224 | 0.281 | 0.063 | 0.253 | 0.249 | 0.054 | 0.183 | 0.295 |
| 3' UTR | 0.059 | 0.224 | 0.263 | 0.059 | 0.253 | 0.233 | 0.065 | 0.183 | 0.355 |
| 3' flanking region | 0.060 | 0.224 | 0.268 | 0.023 | 0.253 | 0.091 | 0.060 | 0.183 | 0.328 |
| All noncoding regions ^a | 0.057 | 0.224 | 0.254 | 0.055 | 0.253 | 0.217 | 0.050 | 0.183 | 0.273 |
| | K_A | K_S | K_A/K_S | K_A | K_S | K_A/K_S | K_A | K_S | K_A/K_S |
| Coding region of exon 1 | 0.000 | 0.116 | 0.000 | 0.039 | 0.000 | — | 0.039 | 0.113 | 0.341 |
| Exon 2 | 0.296 | 0.452 | 0.654 | 0.441 | 0.379 | 1.16 | 0.525 | 0.286 | 1.84 |
| Exon 3 | 0.166 | 0.081 | 2.05 | 0.218 | 0.203 | 1.07 | 0.251 | 0.136 | 1.84 |
| Coding region of exon 4 | 0.101 | 0.174 | 0.579 | 0.279 | 0.316 | 0.855 | 0.279 | 0.166 | 1.68 |
| All coding regions | 0.168 | 0.224 | 0.752 | 0.296 | 0.253 | 1.17 | 0.321 | 0.183 | 1.75 |

^aThe K_S values for estimation of K_N/K_S for the noncoding regions are those for the complete coding regions of the corresponding pairs.

TABLE IV
 K_N/K_S AND K_A/K_S VALUES FOR PAIRS OF VENOM GLAND PLA₂ ISOZYME GENES
 WITHIN SPECIES (*T. gramineus* vs. *T. gramineus*) AND BETWEEN SPECIES
 (*T. gramineus* vs. *T. flavoviridis*)

| Pair of genes | K_N | K_S | K_N/K_S | K_A | K_A/K_S |
|--|--------|-------|-----------|--------|-----------|
| <i>T. gramineus</i> vs. <i>T. gramineus</i> | | | | | |
| <i>gPLA2-I</i> <i>gPLA2-VI</i> | 0.0635 | 0.174 | 0.365 | 0.103 | 0.594 |
| <i>gPLA2-I</i> <i>gPLA2-VII</i> | 0.0657 | 0.301 | 0.218 | 0.331 | 1.10 |
| <i>gPLA2-I</i> <i>gPLA2-V</i> | 0.0550 | 0.263 | 0.209 | 0.296 | 1.13 |
| <i>gPLA2-VI</i> <i>gPLA2-VII</i> | 0.0620 | 0.371 | 0.167 | 0.324 | 0.873 |
| <i>gPLA2-VI</i> <i>gPLA2-V</i> | 0.0539 | 0.325 | 0.166 | 0.290 | 0.902 |
| <i>gPLA2-VII</i> <i>gPLA2-V</i> | 0.0317 | 0.024 | 0.133 | 0.063 | 2.63 |
| <i>T. gramineus</i> vs. <i>T. flavoviridis</i> | | | | | |
| <i>gPLA2-I</i> <i>pgPLA 1a</i> | 0.0697 | 0.272 | 0.257 | 0.173 | 0.635 |
| <i>gPLA2-I</i> <i>pgPLA 1b</i> | 0.0634 | 0.184 | 0.345 | 0.101 | 0.551 |
| <i>gPLA2-I</i> <i>BP-I</i> | 0.0640 | 0.238 | 0.269 | 0.309 | 1.30 |
| <i>gPLA2-V</i> <i>pgPLA 1a</i> | 0.0639 | 0.324 | 0.197 | 0.292 | 0.903 |
| <i>gPLA2-V</i> <i>pgPLA 1b</i> | 0.0567 | 0.285 | 0.199 | 0.312 | 1.10 |
| <i>gPLA2-V</i> <i>BP-I</i> | 0.0477 | 0.156 | 0.306 | 0.138 | 0.885 |
| <i>gPLA2-VI</i> <i>pgPLA 1a</i> | 0.0743 | 0.203 | 0.367 | 0.178 | 0.877 |
| <i>gPLA2-VI</i> <i>pgPLA 1b</i> | 0.0565 | 0.169 | 0.253 | 0.0427 | 0.253 |
| <i>gPLA2-VI</i> <i>BP-I</i> | 0.0644 | 0.231 | 0.278 | 0.301 | 1.30 |
| <i>gPLA2-VII</i> <i>pgPLA 1a</i> | 0.0747 | 0.357 | 0.209 | 0.312 | 0.872 |
| <i>gPLA2-VII</i> <i>pgPLA 1b</i> | 0.0693 | 0.329 | 0.210 | 0.345 | 1.04 |
| <i>gPLA2-VII</i> <i>BP-I</i> | 0.0618 | 0.166 | 0.372 | 0.165 | 0.995 |

That the nucleotide sequences of *T. flavoviridis* and *T. gramineus* venom gland PLA₂ isozyme genes show the same characteristics reveals that accelerated evolution is a rational assumption for crotalinae snake venom gland PLA₂ isozyme genes. This is supported by observations that the genes encoding (1) the heterodimeric mojave toxin from *C. scutulatus scutulatus* (crotalinae) venom (108) and (2) the ammodytoxin C from *V. ammodytes ammodytes* venom (115) are also more conserved in the noncoding regions than in the protein-coding regions.

Southern blot analysis with appropriate probes shows that genes encoding *T. flavoviridis* and *T. gramineus* venom gland PLA₂ isozymes form multi-gene families (22, 101). In contrast, the *Trimeresurus okinavensis* genome contains only three venom gland PLA₂ isozyme genes (*gPLA₂-o1*, *gPLA₂-o2*, and *gPLA₂-o3*), although they consist of four exons and three introns and show the same structural characteristics as *T. flavoviridis* and *T. gramineus* venom gland PLA₂ isozyme genes (102). The K_N/K_S and K_A/K_S values for three pairs of genes are listed in Table V. The K_A/K_S values are higher than unity, indicating that these genes have evolved via accelerated evolution. A phylogenetic tree constructed from the nucleotide sequences of the mature

TABLE V
 K_N/K_S AND K_A/K_S VALUES FOR PAIRS OF THREE *T. okinavensis*
 VENOM GLAND PLA₂ ISOZYME GENES

| Pair of genes | K_N | K_S | K_A | K_N/K_S | K_A/K_S |
|---|--------|-------|-------|-----------|-----------|
| <i>gPLA₂-01</i> vs. <i>gPLA₂-02</i> | 0.0712 | 0.234 | 0.334 | 0.334 | 1.426 |
| <i>gPLA₂-01</i> vs. <i>gPLA₂-03</i> | 0.0594 | 0.237 | 0.285 | 0.285 | 1.203 |
| <i>gPLA₂-02</i> vs. <i>gPLA₂-03</i> | 0.0491 | 0.175 | 0.268 | 0.280 | 1.531 |

protein-coding region of cDNAs or genes encoding *T. flavoviridis*, *T. gramineus*, and *T. okinavensis* venom gland PLA₂ isozymes classified them as acidic [Asp⁴⁹]PLA₂s, including *T. okinavensis* PLA₂-01, which is encoded by *gPLA₂-01*; basic [Asp⁴⁹]PLA₂s, including *T. okinavensis* PLA₂-02, which is encoded by *gPLA₂-02*; and [Lys⁴⁹]PLA₂s, including *T. okinavensis* PLA₂-03, which is encoded by *gPLA₂-03*. Presumably, the three *T. okinavensis* PLA₂ isozymes code for PLA₂ species with different physiological activities. The variability of the mature protein-coding regions of *T. okinavensis* venom gland PLA₂ isozyme genes may have been brought about by natural selection for point mutation rather than by gene conversion.

Table VI presents the mean values of K_A and K_S when compared within 13 viperidae (viperinae and crotalinae) snake venom PLA₂ cDNAs and within mammalian PLA₂ cDNAs (138). The K_A/K_S value within snake venom PLA₂ cDNAs is large (~1.2), whereas within mammalian PLA₂ cDNA it is 0.29. Mammalian PLA₂s are assumed to evolve under neutrality (132), suggesting that the protein-coding regions of Viperidae snake venom PLA₂ isozyme genes have evolved via accelerated evolution.

Evolution of Viperidae snake venom PLA₂ cDNAs and mammalian PLA₂ cDNAs was analyzed by constructing phylogenetic trees using various sequence components (138). The evolutionary tree derived from the combined sequences of the 5' and 3' UTRs and the signal peptide region (Fig. 13) re-

TABLE VI
 MEAN NUMBERS OF NUCLEOTIDE SUBSTITUTIONS PER NONSYNONYMOUS SITE (K_A)
 AND PER SYNONYMOUS SITE (K_S) IN GROUP II PLA₂ cDNAs COMPUTED
 WITHIN SPECIES, AND THEIR RATIOS

| Within species | K_A | K_S | K_A/K_S |
|--|-----------------|-----------------|-----------|
| Viperidae venom PLA ₂ cDNAs ^a | 0.4328 ± 0.0255 | 0.3634 ± 0.0430 | 1.191 |
| Mammalian PLA ₂ cDNAs | 0.2170 ± 0.0287 | 0.7493 ± 0.1354 | 0.290 |

^aThe numbers were calculated by the method of Nei and Jin (137) using the SEND program.

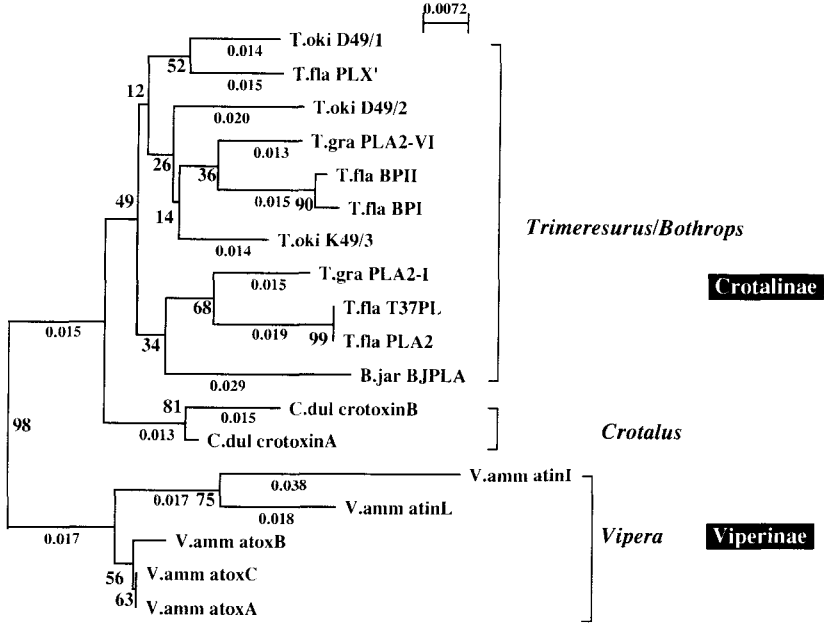


FIG. 13. Phylogenetic relationships of 18 snake venom gland group II PLA_2 s based on the combined sequences of the 5' and 3' UTRs and the signal peptide region of cDNAs. The tree was constructed with the one-parameter method (140) and the neighbor-joining method (141). The numbers at the nodes represent the bootstrap probability (142) determined by using DNA-BOOT in Felsenstein's PHYLIP package. The horizontal branch lengths are drawn to scale and represent the numbers of nucleotide substitutions per site.

veals the evolution of two groups, i.e., crotalinae PLA_2 s and viperinae PLA_2 s. Crotalinae PLA_2 s can further be split into *Crotalus* and *Trimeresurus/Bothrops* PLA_2 isozymes. This tree agrees with the taxonomic relationships of mitochondrial rRNA genes from crotalinae and viperinae snakes (139). In contrast, the tree constructed from the mature protein-coding region sequences of cDNAs leads to a pattern that is incompatible with a division of PLA_2 s into crotalinae and viperinae subspecies. Instead, it suggests three subgroups (acidic $[Asp^{49}]PLA_2$ s, basic $[Asp^{49}]PLA_2$ s, and $[Lys^{49}]PLA_2$ s) that reflect their molecular properties (data not shown). Mammalian PLA_2 s are totally separated from snake venom PLA_2 s, reflecting the diversity of snake venom PLA_2 isozymes. The above analyses as well as the trees constructed from the amino acid sequences and from the nucleotide substitutions at each of three positions of codons led to the conclusion that Viperidae snake venom PLA_2 isozyme genes have evolved via accelerated evolution.

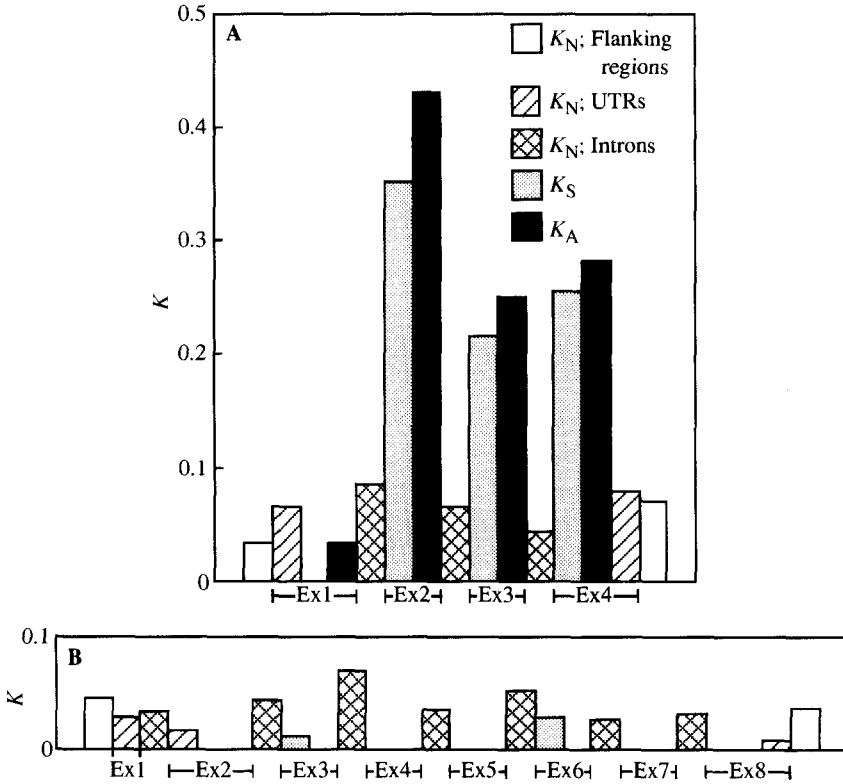


FIG. 14. Schematic representation of sectional K_N , K_S , and K_A values for (A) a pair of $gPLA_2-VI$ and $BP-I$ genes and (B) $T. gramineus$ and $T. flavoviridis$ TBP genes. Ex, Exons; introns are portions sandwiched between exons. Each K is expressed by different patterns, as indicated in the key in A. In PLA_2 isozyme genes (A), K_S for exon 1 is zero. In TBP genes (B), exon 1 contains the UTR only. In TBP genes, K_A for exons 3 and 6 and K_S and K_A for exons 2, 4, 5, 7, and 8 are all zero. It is evident from A that in venom gland PLA_2 isozyme genes, K_S and K_A values are much greater than K_N values.

To shed light on the evolutionary feature of snake venom gland PLA_2 isozyme genes in more detail, the evolutionary rates of introns of venom gland PLA_2 isozyme genes were compared with those of introns of general (nonvenomous) genes. We therefore determined the nucleotide sequences of $T. flavoviridis$ and $T. gramineus$ TATA box binding protein (TBP) genes (17 kb) (143), which contain eight exons and seven introns and which are assumed to be a general gene ubiquitously present in all cell types. The K_N , K_S , and K_A values for the pair $gPLA_2-VI$ and $BP-I$, for example, were compared

with those for *T. flavoviridis* and *T. gramineus* TBP genes (Fig. 14). The K_N values for venom gland PLA₂ isozyme genes are comparable to those of TBP genes, indicating that the introns of venom gland PLA₂ isozyme genes and the introns of TBP genes have evolved at similar rates, presumably under neutrality (129, 132). Some functional constraints in the introns of venom gland PLA₂ isozyme genes, which had been considered because of the high homology of introns compared with the protein-coding regions, can be ruled out. Thus, the highly accelerated substitutions that occur only in the protein coding regions, except for the signal sequence domain, are responsible for the Darwinian-type accelerated evolution of venom gland PLA₂ isozyme genes, producing new physiological functions. Acquisition of diverse venom PLA₂ isozymes by accelerated evolution may confer a strong selective advantage for disrupting the physiological integrity of prey or for defense against predators.

Accelerated amino acid substitutions have been reported to occur in the active site regions of serine protease inhibitors (144) and porcine elafin family (145). Also, the putative antigen recognition sites, but not other parts of the major histocompatibility complex multigene family, have evolved via positive Darwinian selection (146). In contrast, nonsynonymous substitutions are spread over the protein-coding region in Viperidae snake venom gland PLA₂ isozyme genes except for the region coding for the signal peptide. Therefore, we suggest that venom gland PLA₂ isozyme genes produced new functional proteins by introducing mutations throughout the entire molecule of the mature proteins.

III. Snake Toxins with a Three-Fingered Fold

A. Structures of Three-Fingered Toxins

Elapid and hydrophid snakes produce potent venoms that contain a variety of small proteins with approximately 60–70 amino acids and usually 4–5 disulfide bonds. Irrespective of whether these proteins are functionally different (see below), their amino acid sequences can be readily aligned, using eight half-cystines as conserved elements, suggesting the presence of common structural features. Figure 15 shows a sequence alignment of toxins belonging to various functional groups. That such proteins share similar folds has been confirmed by analysis of their secondary structure, using circular dichroism (147). The spectra of the toxins display a negative dichroic signal centered around 210–220 nm and a positive signal around 197 nm (not shown), suggesting the presence of a substantial β -sheet structure with probably little, if any, helical conformation. The presence of a predominant β -sheet structure was subsequently demonstrated by X-ray crystallography

| Loop I | | | | | | | | | | Loop II | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------|---|---|---|---|---|---|---|---|---|---------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * | | | | | | | | | | * | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | R | I | C | F | | N | H | Q | S | S | Q | P | Q | T | T | K | T | C | S | P | G | E | S | S | C | Y | H | K | Q | W | S | D | F | | R | G | T | I | I | E | R | G | C | | | | | | | | | | |
| 2 | - | - | - | - | | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | N | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | | | | | |
| 3 | - | - | - | - | | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | | | | |
| 4 | R | - | - | - | | Q | - | - | E | - | - | N | S | P | - | - | N | - | - | R | - | - | - | - | R | - | H | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | | | |
| 5 | L | T | - | - | | C | Q | - | - | - | K | - | T | D | - | A | D | N | - | - | K | M | T | - | R | - | H | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | | | |
| 6 | L | T | - | - | | Q | - | - | - | - | K | - | T | D | - | A | D | D | - | - | K | - | T | - | K | - | H | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| 7 | L | K | - | - | | | K | L | V | - | L | P | Y | - | - | P | A | K | N | L | - | - | K | M | F | M | V | A | T | P | K | V | P | V | K | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| 8 | L | T | - | V | | T | G | K | - | I | G | G | I | S | - | E | E | - | A | A | - | Q | K | I | - | F | K | - | W | T | K | M | G | - | K | L | Y | D | V | S | - | - | - | - | - | - | - | - | - | - | - | - | |
| 9 | L | T | - | V | | T | T | K | - | I | G | G | V | - | - | E | D | - | P | A | - | Q | N | V | - | F | K | R | W | H | Y | V | T | - | K | N | Y | D | - | I | K | - | - | - | - | - | - | - | - | - | - | - | |
| 10 | T | M | - | Y | | S | - | T | T | S | R | A | I | L | T | N | - | - | N | - | - | R | - | S | R | R | H | P | - | K | M | V | L | G | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 11 | - | - | - | Y | | S | - | K | L | L | - | A | K | - | - | - | - | E | N | - | - | K | R | S | L | P | K | I | - | L | I | - | G | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| 12 | M | - | - | Y | | S | - | K | T | P | - | S | A | - | I | - | - | E | K | T | - | - | K | - | S | V | R | K | L | - | A | I | V | A | G | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| X | - | - | - | Y | | - | - | L | G | T | K | - | P | - | E | - | - | Q | - | D | - | - | K | N | I | - | T | F | D | - | N | - | - | R | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| Y | - | - | - | Y | | I | - | K | A | - | L | - | R | A | - | - | - | V | - | N | T | - | - | K | M | F | I | R | T | Q | - | E | Y | - | S | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 13 | K | T | - | L | | K | T | P | - | - | - | - | - | - | - | - | - | P | Q | - | Q | D | I | - | F | L | - | V | S | C | E | Q | - | C | - | I | - | - | P | V | - | - | Q | - | - | - | - | - | - | - | - | | |
| 14 | - | T | - | L | | I | S | P | - | - | - | - | - | - | - | - | - | - | F | N | - | Q | D | I | - | F | R | - | A | Q | - | D | N | - | - | H | S | - | - | P | V | - | - | Q | - | - | - | - | - | - | - | - | |
| 15 | M | K | - | K | I | - | H | F | D | T | C | R | A | G | E | L | - | V | - | A | S | - | - | K | Y | - | F | K | E | S | - | R | E | A | - | - | - | - | R | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| 16 | I | - | - | R | T | R | D | T | Y | Q | I | - | I | - | F | T | N | - | E | E | - | H | V | - | - | K | Y | S | T | T | E | T | - | N | - | - | I | L | - | H | - | - | - | - | - | - | - | - | - | - | - | | |

FIG. 15. Amino acid sequences of sixteen selected toxins or toxinlike peptides known or presumed to adopt a three-fingered fold. Toxins 1–4 are erabutoxins c, a, b, and d from the sea snake *L. semifasciata* (197, 198). Toxins 5 and 6 are toxins b and d from the sea snake *A. laevis* (199). Toxin 7 is a cardiotoxin from the cobra *N. n. sputatrix*. Toxins 8–12 are, respectively, the synergistic-type toxin, muscarinic toxin 2 (200), fasciculin 1 (201), Da C₁₃S₁C₁1, and F8, all from the green mamba *D. angusticeps*. X is dendroaspasin from *D. jamesonii kaimosae* (202) and Y is calciceptin (157) from *D. polylepis polylepis*. Toxins 13–16 are, respectively, neurotoxin Kappa 2, neurotoxin Kappa 3 (210), CE (211), and NL1 from *B. multicinctus*. The stars indicate the half-cystines that are conserved in all toxins and which have served for sequence alignments. The positions of the three loops are also indicated. The nucleotide sequences of toxins numbered from 1 to 16 are shown in Fig. 17. The cDNAs encoding the toxins designed as X and Y are unknown.

studies and/or NMR spectroscopy analyses of various toxins. As an illustration, Fig. 16 shows the architectures of a member of each toxin family: toxin α from *Naja nigricollis* (148), α-cobratoxin from *Naja naja siamensis* (149), κ-neurotoxin from *Bungarus multicinctus* (150, 151, 152), toxin γ from *N. nigricollis* (153, 154), muscarinic toxin 2 from *Dendroaspis angusticeps* (155), FS2 from *Dendroaspis polylepis polylepis* (156) [which is homologous to calciceptine (157)], fasciculin 1 from *D. angusticeps* (158), and dendroaspasin from *Dendroaspis jamesonii kaimosae* (159a,b). Most clearly, all toxins adopt the same fold with three loops, rich in β-pleated sheet, which emerge from a globular core where four invariant disulfides are located.

However, closer inspection of the individual toxin structures reveals that a number of deviations are accommodated by the fold (see Figs. 15 and 16):

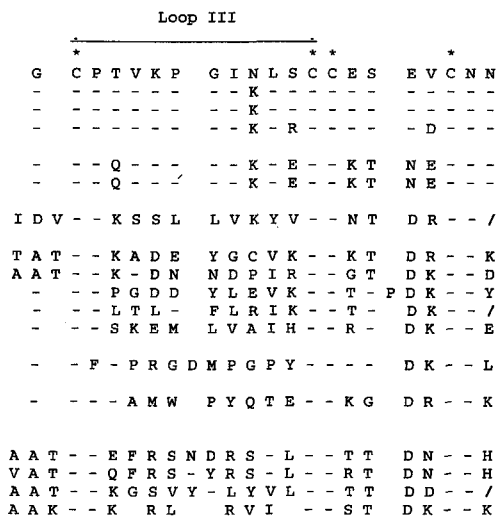


FIG. 15. (Continued)

1. A variability of loop sizes; for example, the first loop is substantially smaller in long-chain toxins (Fig. 16A, b) than in short-chain toxins (Fig. 16A, a).
2. Accommodation of extra loops; long-chain toxins (Fig. 16A, b and d) possess a small loop, usually but not always cyclized by a disulfide bond, at the very top of the central loop.
3. Introduction of bulges in a β -strand as, for instance, in the central loop of cardiotoxins (Fig. 16B, a).
4. A variability in the twisting capacity of a loop; the tip of the central loop is much more twisted in muscarinic toxin (Fig. 16A, c) as compared with short-chain curaremimetic toxins (Fig. 16A, a).
5. Major differences in the relative orientations of the loops; the central and third loops of most toxins are tightly associated with each other whereas there is a rearrangement in dendroaspin (Fig. 16B, c) with its third loop, which markedly diverges from the plane of the β -sheet of the central loop.
6. The capacity of the fold to act as a monomer, as in most toxins, or as a dimer, as in neuronal toxins (Fig. 16A, d).

Clearly, the three-fingered scaffold can accommodate a wide diversity of structural changes, enabling various toxin functions.

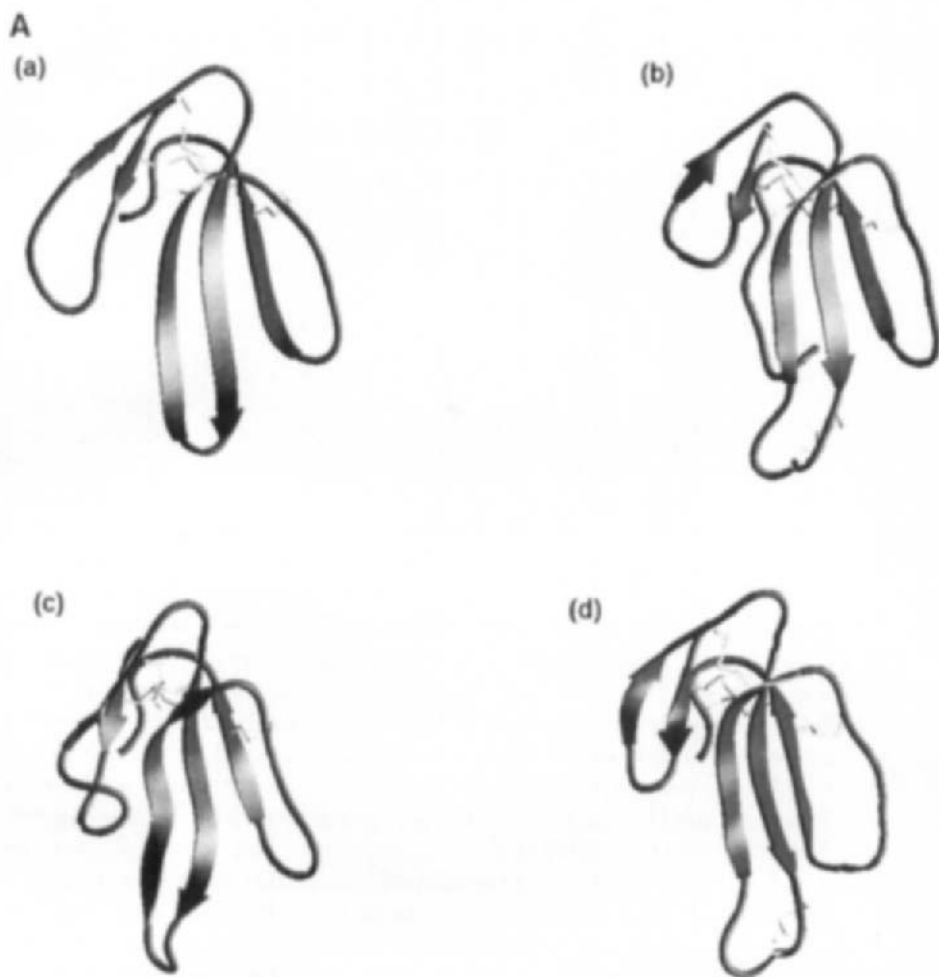


FIG. 16. Three-dimensional structures of various toxins that adopt a three-fingered fold. (A) The solution structure of (a) the short-chain curaremimetic toxin α from *N. nigricollis* (148); the solution structure (b) of the long-chain curaremimetic α cobratoxin from *N. n. siamensis* (149); the solution structure of (c) muscarinic toxin 2 from *D. angusticeps* (155); the crystal structure (d) of the monomer of the dimeric neuronal κ -bungarotoxin from *B. multicinctus* (151). (B) The solution structure (a) of cardiotoxin γ from *N. nigricollis* (153); the crystal structure (b) of fasciculin 1 from *D. angusticeps* (158), a blocker of acetylcholinesterase; the solution structure (c) of dendroaspin from *D. angusticeps* (159a), a blocker of the integrin receptor Gp IIb IIIa; the solution structure (d) of FS2 from mamba (156), an inhibitor of calcium channels. The representations of toxins structures have been made with MOLSCRIPT.

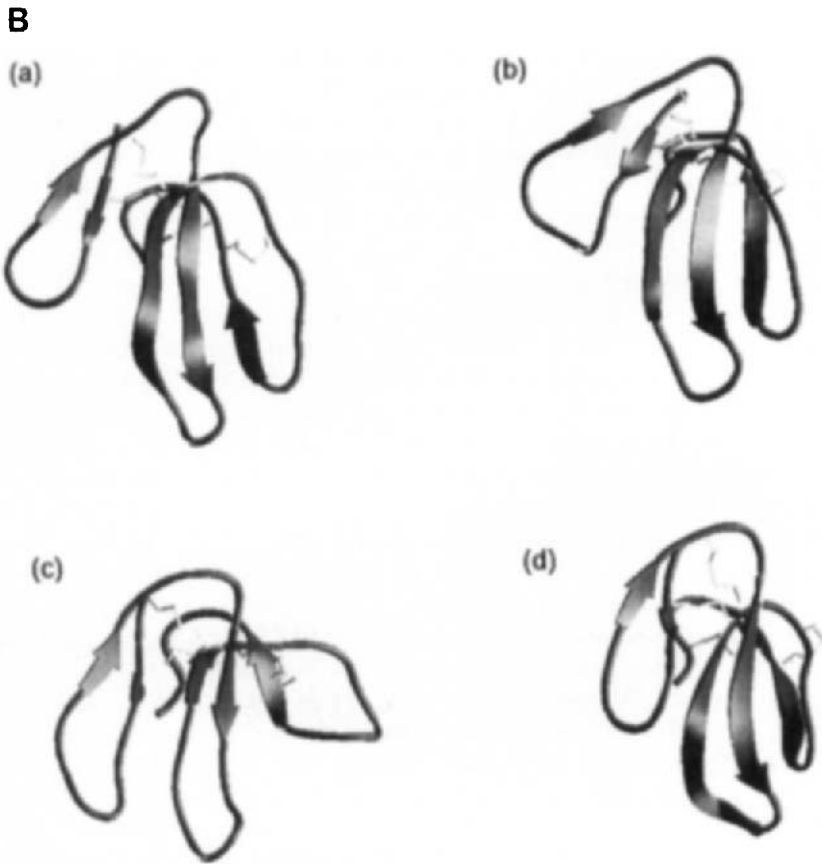


FIG. 16. (Continued)

B. Functions of Three-Fingered Toxins

The first three-finger-type toxins to be isolated were the curaremimetic toxins, also called α -neurotoxins, found in most venoms from elapid and hydrophid snakes. The amino acid sequences of more than 80 curaremimetic toxins are presently known. Both short-chain and long-chain curaremimetic toxins recognize the peripheral nicotinic acetylcholine receptors from muscles with high affinities, characterized by K_d values often close to 10 pM. They inhibit the binding of acetylcholine to these receptors, which are thus

maintained in a resting state. The affected muscles, including especially the diaphragm, remain in flaccid paralysis, thus causing respiratory failure. In sharp contrast, only long-chain curaremimetic toxins have high affinities for $\alpha 7$, a neuronal acetylcholine receptor, short-chain curaremimetic toxins having nearly four orders of magnitude lower affinities for this receptor (159c). This differential affinity has been specifically attributed to residues that are located around the small extra loop cyclized by the fifth disulfide bond (D. Servent and A. Ménez, unpublished data). This example offers a clear illustration of the capacity of the three-fingered fold to accommodate various deviations and hence to exert discriminating activities. The toxins from other families further illustrate this general rule. Three other three-fingered toxins bind with high affinity to endogeneous proteins, other than the nicotinic acetylcholine receptor; notwithstanding, they are naturally targeted by acetylcholine. These are the neuronal toxins, the muscarinic toxins, and fasciculins. The neuronal toxins, or κ -neurotoxins, present in venoms from kraits, bind to nicotinic acetylcholine receptors that are not recognized by curaremimetic toxins and that are present, for instance, in mammalian or avian autonomic ganglia (152). Muscarinic toxins have been discovered in mamba venoms (160). A considerable variety of such toxins bind to various subclasses of G-protein-coupled muscarinic receptors, some of them acting as agonists. The fasciculins, also found in mamba venoms, bind to acetylcholinesterase, an enzyme that recognizes and hydrolyzes acetylcholine. Fasciculins block the capacity of the enzyme to hydrolyze acetylcholine (161).

Other toxins that adopt a three-fingered type structure have targets that are unrelated to acetylcholine. Cardiotoxins, found in cobra venoms, cause lysis of various excitable cells (162). It is often considered that phospholipids are the targets of cardiotoxins; however, further experimental evidence is required to substantiate this. Calciceptine, discovered in mamba venom, recognizes L-type calcium channels (157). Dendroaspin, also from mamba venoms, is a disintegrin that acts as an inhibitor of platelet aggregation by binding to the integrin, Gp IIb IIIa (159b). There are, in addition, a number of functionally unidentified components present in various venoms, especially from mambas; these components are characterized by primary structures that can be aligned with those of any of the three-finger-type toxins and which are also likely to adopt this fold (161).

Snake toxins adopting the same overall three-fingered fold have a variety of unrelated functions. A major question is how this rather simple fold can accommodate such a functional diversity.

C. On the cDNAs Encoding Three-Fingered Toxins

Figure 17 shows the sequences of cDNA encoding 16 three-fingered toxins or analogs. Not all three-dimensional structures of the encoded toxins

have been individually elucidated; however, because they all possess the consensus sequence --C-----C---C-----GC--C---CC---C---, they are inferred to adopt the same three-fingered fold.

The cDNAs contain an open reading frame of approximately 280 nucleotides encoding 81–86 residues, among which 21 constantly form a signal peptide. The first six sequences encode curaremimetic toxins found in two sea snake species, *Laticauda semifasciata* and *Aipysurus laevis*. The nucleotide sequence numbered 7 encodes a cardiotoxin from an elapid snake. The next five nucleotide sequences encode toxins from the same mamba species, *D. angusticeps*, and the following sequences encode toxins from the krait *B. multicinctus*. A simple look at Fig. 17 clearly reveals that the 5' and 3' UTRs are more conserved than are the protein-coding regions. For example, five sequences, numbered 8–12 in Fig. 17, encode different toxins from the same mamba species, and the average nucleotide identities are 96, 91.5 and 80.8% for the 5' UTR, 3' UTR, and the protein-coding region, respectively. Moreover, within the protein-coding region, the signal peptide-coding domain is highly conserved, with 95.5% homology, implying that the region coding for the mature protein is characterized by only 66.3% conservation. For toxins from kraits (sequences 13–16) this value is even smaller, with a conservation of 66.3% for the region coding for the mature proteins. These observations clearly reveal that a higher rate of mutations occurs within the regions that code for the mature protein adopting the three-fingered fold. It is remarkable that these values are closely similar to those that were found for snake PLA₂ cDNAs.

The evolutionary significance associated with differences in nucleotide sequences of three-fingered toxins has been investigated. For each pair of cDNAs encoding toxins from the same mamba species (*D. angusticeps*), we computed the number of nucleotide substitutions per site (K_N) for the non-coding regions and the number of nucleotide substitutions per synonymous (K_S) and nonsynonymous (K_A) sites for the protein-coding regions. These parameters were as previously defined in Section II.D. Table VII shows K_A , K_S , K_N , K_N/K_S , and K_A/K_S values for the mamba toxins. The most striking observation that emerges from this table concerns the particularly high values that characterize the ratio K_A/K_S . All values are above 1 and can approach 8, indicating that the toxins undergo a high degree of substitution in non-synonymous sites; this suggests that the corresponding genes have evolved

FIG. 17. Alignment of the nucleotide sequences of 16 cDNAs encoding three-fingered toxins (203–211). The cDNA encoding erabutoxin c is used as a reference (203). The region that codes for the signal peptide is underlined. The encoded toxins are those whose amino acid sequences are numbered from 1–16 in Fig. 15. The putative polyadenylation signal, AATAAA, is indicated in bold face.

```

      10      20      30      40      50      60      70      80      90
1 Erabu c      ACTTTGCAGGCTCCAGAAAAGATCGCAAG ATG AAA ACT CTG CTG CTG ACC TTG GTG GTG ACA ATC GTG TGC CTG GAC TTA GGA TAC ACC
2 Erabu a      -----G-----
3 Erabu b      -----
4 Erabu d      -----
5 Alp.1.B      AC-----G-----T-----
6 Alp.1.D      AC-----G-----T-----
7 Cardio      CC--C-----C-C--GA-----G-----T--T AC-----T--C-----T--A-----T
8 Syn.1.      AC-----G-----
9 Musca. 2     -----G--C-----T-----
10 Fas. 1      AC-----T--G-----T-----C-----A-----C-----
11 Da C13S1Cl 1. AC-----G-----C-----G-----C-----G-----C--C-----
12 Da F8 1.     AC-----G-----C-----C-----C-----C-----C-----
13 Kappa 2     -----G-----T-----T-----T-----
14 Kappa 3     -----G-----T-----T-----
15 CE          -----G--C--A-----
16 NL 1        -----G--C--A-----

```

```

      100      110      120      130      140      150      160      170      180      190      200
1  AGG ATA TGT      TTT AAC CAT CAG TCA TCG CAA CCG CAA ACC ACT AAA ACT TGT TCA CCT GGG GAG AGC TCT TGC TAT CAC AAG CAA TGG AGC GAT TTC
2  -----
3  -----
4  -A -G----- -C-----A-----A G--T-----A-----G-----C-----A-----AGA-----G-----A-----CA-
5  TT- C- --- -GC ---A-----T A-----C- GA---G-- --T A-----A-A -T ACT ---G --- CA-
6  TT- C- --- -GC ---A-----T A-----C- GA---G-- --T GA-----A-A --ACT ---AG --- CA-
7  CT- -AG --C      --- A-- CTG GT- --C -TT TT- TAC --- --- C-G G-C --C A-A -A- CT- ---C A-A -T- -TC -TG -T- GCC AC-
8  TT- C- --- G-C -CA GGC A-- --- ATT GGG GGT AT- -G- --- G-G GAA ---G-- G-- ---C-- -AG ATA -T -T- A-A --- --- ACA AAG A-G GGT
9  TT- C- --- G-C -CA ACC A-- --- ATA GGG GGT GT- --- G-G GA---C-- G-- ---C-- -A- GTA -T -T- A-A -G- --- C-- TA- G-T AC-
10 -C- -G --- -AC -G- AC- A-- A-T TC- -GT GC- -T CT- -C- -A--- --A ---A-----AGA ---CT C-G CG- CAT CC-
11 ----- -AC -G- A-- -T- CT- ---G-T A-----A-- -A-----A-- -A--G- -CT CTG CC- AAA A--
12 -T----- -AC -G- A-- A-- C-----T TC- G-----T-----A-- ---AA A-----A-A --- -CT GTG CG- AAA --G
13 -A- C- --C      C-C --A ACA -CT -T -T --- C- C-G --A --- C-- AA ---G-- GA- ATA ---T- TA ---GTT -C- TGT --G GAA ---T--
14 -----C- --C      C-C -TA TCA -CT -T -T --- C- C-G --A --- C-- AA ---G-- GA- ATA ---T- GA ---GCT CA- TGT ---AA---T--
15 -T- -A --- AAG ATA -G- C-- TTC G-T A-T -GT -GT G-C GGT GAA CTC ---GT- ---G- AG- --A ---AG -A---T- A-A G-G TCT ---C-T --G GCA
16 -TA --- ---CGC -C- CG- G-- ACC T-T CAG ATA -CT ATC ACC TT- --C -T ---GA- GAA --A C-C GTA ---A-A T-T AGT ACT -CA -G ACT
      210      220      230      240      250      260      270      280      290      300      310
1  CGT GGA ACT ATA ATT GAA AGG GGA TGT GGT TGC CCC ACA GTG AAG CCC GGT ATT AAC CTT AGT TGT TGC GAA TCA GAG GTC TGC AAC AAT
2  -----
3  -----
4  -----
5  -----G-----
6  -----G-----T CAG-----T CAG-----A-- GAA-----A-- A-- A-- A-C -AA-----
7  CCG AAG GT- C-- G-C A-- C-C --G --C ATT -AC GTA --- --AG AGC TC- -T- CTG G-G --A TA- GTG --C ---A-C A-C --C AGG ---

```

```

8  --A --A -T- TAT GAC G-C -GC  --- --- --- -C- -C- ACT --- -T -A- -C- G-C GAA  TA- GG- TG- G-- -AA --- --- A- A--  --C AGA --- --- --A
9  --C --A AA- TAT GAC  -C -C  -A- --- --- GC- -C- AC- --- -T -A- -G-T AA-  AA- GA- CC- A-- CGC --- --- -G- A--  --C AAA --- --- G--
10 --T --A ATG  G-- C-C -G- --- --- --- --- -T C-- -GT G-C GA-  TA- C-- G-A G-- -AA --- --- AC- --- --- CCA  --C AAA --- --- T-C
11 --T CTA AT-  --- -C -G- --- --- --- --- -T TT- AC- CT- ---  TT- C-- CGA A-- -AA --- --- AC- --- --- -C AAA --- --- G-A
12 --T GCA AT-  G-- GCC -G- --- --- --- --- --- -T T-- AA- G-- ATG  CT- G-- GCA A-- CAA --- --- AG- --- --- -C AAA --- --- G-A
13 --C -TC A-A --- C-- G-- -C --- CAA --- --- GC- -C- ACC --- -T GA- T-T -GA T-- AAT -A- -GA TCT --- CTC --- --- AC- A--  --C AAT --- -T C-C
14 -AC -GC A-A --- C-- G-- -C --- CAA --- --- G-- -C- ACC --- -T CA- T-T -GA T-- --- TA- -GA TCT --- CTC --- --- AG- A--  --C AAT --- --- C-C
15 --- --- --- -G- -C --- --- --- --- --- GC- -C- AC- --- -T -A- -G- -GC GT- T-- --G C-C T-T G-- CTC --- --- AC- A--  --C -AT --- --- ---
16 --A --T  A-- -TC C-C -A C-C --- --- --- --- --- GC- -C- AAG --- -T -A-  AGA CTA C-- G-- -TA  --- --- -T TC- A--  --C AAA --- --- --A

```

```

310      320      330      340      350      360      370      380      390      400      410      420      430      440
1  TAG  CTCTACGAGTGGCTAAATTCCTTGAGTTTACTCTCATTCATCAAGGACCACTCTCAAAA  TGTATGC  TTCGGCCTTTACC  ACCACATGGTCCATCATCCCGCTCTCCCTGCTGTCTTTGACACCTCAAC
2  ---
3  ---
4  ---
5  ---
6  ---
7  ---
8  ---
9  ---
10 ---
11 TAA ---
12 ---
13 ---
14 ---
15 TGA ---
16 ---

```

```

450      460      470      480      490      500      510      520      530
1  ATCTTTCCCTTTCTCTGTGATCTGTAAGTTTCCCTCTGCTAGTTCCTGAGTTTGAGAAATCAAAACCTCAGCATTC  (?)
2  --- (A27)
3  --- (A10)
4  --- (A9)
5  --- (A5)
6  ---
7  ---
8  --- AATTG (A7)
9  ---
10 --- (A8)
11 ---
12 --- (A6)
13 --- (A19)
14 ---
15 --- (A16)
16 --- (A16)

```

TABLE VII
NUCLEOTIDE SUBSTITUTIONS IN *D. angusticeps* TOXIN cDNA

| Toxin pair ^a | Toxin no. | K_N | K_S | K_N/K_S | K_A | K_A/K_S |
|---|-----------|-------|-------|-----------|-------|-----------|
| Syn 1/Musc 2 | 8/9 | 0.065 | 0.077 | 0.849 | 0.227 | 2.956 |
| Fasc 1/Syn 1 | 10/8 | 0.107 | 0.265 | 0.404 | 0.475 | 1.792 |
| Fasc 1/Musc 2 | 10/9 | 0.128 | 0.381 | 0.336 | 0.451 | 1.184 |
| Fasc 1/DaC ₁₃ S ₁ C ₁ 1 | 10/11 | 0.092 | 0.082 | 1.121 | 0.232 | 3.949 |
| Fasc 1/Da F81 | 10/12 | 0.061 | 0.061 | 1.002 | 0.325 | 5.363 |
| Da C ₁₃ S ₁ C ₁ 1/F8 | 11/12 | 0.060 | 0.029 | 2.066 | 0.230 | 7.958 |
| Da C ₁₃ S ₁ C ₁ 1/Syn 1 | 11/8 | 0.108 | 0.368 | 0.293 | 0.455 | 1.236 |
| Da C ₁₃ S ₁ C ₁ 1/Musc 2 | 11/9 | 0.110 | 0.339 | 0.324 | 0.466 | 1.375 |
| Da F81/Syn 1 | 12/8 | 0.092 | 0.347 | 0.265 | 0.462 | 1.331 |
| Da F81/Musc 2 | 12/9 | 0.116 | 0.283 | 0.410 | 0.520 | 1.837 |

^aSyn 1, Synergistic-type toxin; Musc 2, muscarinic toxin 2; Fasc 1, fasciculin 1.

through accelerated amino acid substitutions. This preliminary description is in complete agreement with the extensive observations that have been reported for the PLA₂ isozymes from viper venom glands, suggesting that snake venom toxins with similar folds (PLA₂ and three-fingered folds) all undergo a higher rate of mutation, perhaps indicating accelerated evolution, exclusively limited to the region coding for the protein.

D. Functional Sites of Three-Fingered Toxins

Sequences in Fig. 17, as well as more precise calculations (not shown) did not indicate any specific area of the protein-coding region that could be associated with a higher rate of mutation, suggesting that functional topographies of the different three-fingered toxins may be localized on variable regions of the fold. One way to investigate the validity of this hypothesis consists in elucidating the functional topographies by which toxins with the same three-fingered fold can recognize different targets. Previously, our laboratory identified a number of critical residues involved in the functional sites of a cardiotoxin and a curaremimetic toxin, both isolated from the same snake. This identification was achieved by modifying, one at a time, as many residues as possible of each toxin and by investigating the consequences of the introduced changes on the biological properties of the two toxins.

I. CHEMICAL MODIFICATIONS ON CURAREMIMETIC TOXIN AND CARDIOTOXIN

Naja nigricollis (whose name may not be the most appropriate choice regarding the actual zoological nomenclature of African snakes) is the source of toxin α and toxin γ , identified as a curaremimetic toxin and a cardiotoxin,

respectively (162). Toxin α (61 residues and 4 disulfides) was made radioactive with tritium and was shown to bind to the nicotinic acetylcholine receptor (AcChoR) from the electric organ of the fish *Torpedo marmorata* with great specificity and high affinity ($K_d = 20$ pM) (171). Its three-dimensional structure (Fig. 16A) has been solved by NMR spectroscopy and molecular modelling (148). Modification of two aromatic residues and five amino groups (four lysines and the α -amino group of the N-terminal residue) was achieved and competition experiments revealed that modifications at Trp-29 as well as Lys-27 and Lys-47 (using the numbering of curaremimetic toxins having 62 residues) caused a substantial decrease in the affinity of the toxin for AcChoR, whereas modifications of the α -amino group of Leu-1 or the ϵ -amino group of Lys-15 and Lys-51 had little or no effect on the stability of the toxin–AcChoR complex (163). These data agree with others reported for various toxins (164). Thus, three residues (K27, W29, and K47), which belong to one of the two faces of the fold, as defined by the plane of the β -sheet, are functionally important, whereas residues located on the other face (K51) or in the core (N terminus and K15) are not. Thus, it was concluded that the AcChoR binding site of the toxin may essentially belong to the most concave face of the toxin, with functional residues belonging to the second and third loops.

Toxin γ has 60 residues and 4 disulfides and its three-dimensional structure (see Fig. 16B) has been elucidated by both NMR spectroscopy (153) and X-ray crystallography (154). This toxin has the capacity to be lethal and cardiotoxic, causing depolarization of excitable cells and cell lysis (162). To investigate whether these activities are related to each other, our laboratory has performed a number of chemical modifications and has examined whether the introduced modifications caused any correlative changes in the different activities. Trp-11, Tyr-22, and Tyr-51 were individually modified (165, 166) as well as Leu-1, Lys-2, Lys-12, Lys-18, Lys-23, and Lys-35 (167). Modification at a given residue caused parallel effects on the various activities, suggesting that a similar functional topography is associated with the different activities. The data showed that the most important functional residues are located on the edge of the first loop of the toxin, with the side chains pointing on both sides of the plane of its β -sheet. In particular, Lys-12 and to a lesser extent Trp-11 act as very important functional residues. In sharp contrast, modification of Leu-1 or Lys-2, which both also belong to the first loop, had no effect on the activities of toxin γ . Additional residues, including Tyr-22, Lys-23, and Lys-35, located at the base of the second loop, also play some functional role, though to a lesser extent as compared to the critical Lys-12. These data agree with various observations from other laboratories. In particular, it was shown that a cyclic peptide with the amino acid sequence of the first loop was sufficient to cause a number of the effects of cardiotoxin, though with a substantially lower potency (168). Also, structural experiments

suggested that a cluster of hydrophobic residues organized around Tyr-22 and close to the functionally important Lys-12 may play some recognition role toward membrane phospholipids (153). The functional site of a cardiotoxin appears, therefore, to be composed both of specific elements located on the first toxin loop, contributing substantially to the expression of cardiotoxin activities, and of less specific elements that may provide cardiotoxin with an ability to recognize cell membranes.

In 1990 it was observed that two toxins adopting a three-fingered fold had distinct functions and topographically different functional sites (162, 167). The curaremimetic toxin recognizes its receptor using residues located on the central loop, whereas cardiotoxin function is specified by residues located at the edge of the first loop, suggesting that the functional sites of the two three-fingered toxins are topographically distinct. However, this conclusion was based on site delineations that were incomplete. Additional evidence was needed to support this conclusion.

More precise delineations of the functional sites have since been achieved from studies with two other toxins—erabutoxin a, a curaremimetic toxin from the sea snake *L. semifasciata*, and fasciculin, an inhibitor of acetylcholinesterase that also possesses a three-fingered fold and which was discovered in the venom of *D. angusticeps*.

2. DELINEATION OF THE FUNCTIONAL SITE OF A CURAREMIMETIC TOXIN

Erabutoxin a (Ea), from *L. semifasciata* and toxin α from *N. nigricollis* share similar primary structures, with approximately 75% identity, and bind to AcChoR with comparable affinities ($K_d = 20$ and 70 pM, respectively) (171). The cDNA of Ea was cloned, fused to the gene encoding protein A, and the hybrid was produced in the periplasm of *Escherichia coli* in an active form and then cleaved by cyanogen bromide (169). The recombinant Ea was structurally and functionally indistinguishable from the original snake toxin. Using this production system, a large panel of mutants was engineered, and if one excludes the eight half-cystines whose mutations are anticipated to cause dramatic structural perturbations, the functional role of more than 80% of the remaining positions of Ea has now been explored individually (170–172), providing a clear picture as to which residues of the toxin are sensitive to mutations and hence are likely to be involved in the binding of the toxin to AcChoR. These data confirmed the conclusions drawn from chemical derivatization that Lys-27, Trp-29, and Lys-47 are functionally important. In addition, this indicates that Asp-31, Arg-33, Ile-36, and Glu-38 on loop II are also important. Even more interestingly, perhaps, Gln-7, Ser-8, and Gln-10 on the tip of loop I were also part of the functional site of Ea. The residues surrounding these functional amino acids could be mutated without signifi-

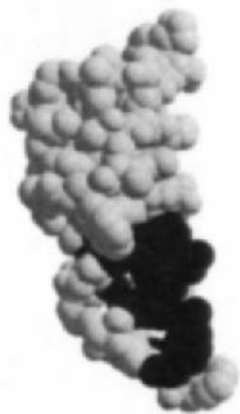
cantly altering the affinity of the toxin for AcChoR, confirming the clear delimitation of the functional site of the Ea. Thus this site includes 10 amino acids, with three of them located at the tip of loop I, six others on loop II, and one residue on loop III. It is particularly striking that all but one of these functional residues belong on the concave face of the toxin (see bottom of Fig. 18). The only exception is Ser-8, whose side chain lies virtually in the plane of the β -sheet, within loops I and II. Apparently, the convex side of the toxin establishes no contact with the receptor.

3. DELINEATION OF THE SITE BY WHICH FASCICULIN INTERACTS WITH ACETYLCHOLINESTERASE

Fasciculin was discovered in mamba venoms (161). It is a potent inhibitor of acetylcholinesterase (AChE), the enzyme that terminates the action of the neurotransmitter acetylcholine by hydrolyzing it into choline and acetate, in the cleft of the cholinergic synapses, in the central and peripheral nervous systems. Fasciculin contains 61 residues, including four disulfide bonds and its tertiary structure is highly similar to that of curare-mimetic toxins (see Fig. 16B). Identification of the fasciculin residues that are in contact with AChE was recently obtained from elucidation of the crystal structure of the toxin-AChE complex (173, 174). Fourteen toxin residues establish predominant interactions. They are located on loop I (Tyr-4, His-6, Thr-8, Thr-9, Arg-11, and Ala-12), loop II (Arg-27, His-29, Pro-30, Pro-31, Lys-32, Met-33, and Arg-37), and C-terminal Tyr-61. As shown in the upper part of Fig. 18, the surface formed by these residues belongs to both faces of the plane defined by the β -sheet.

4. COMPARISON OF THE FUNCTIONAL SITES OF A CURAREMIMETIC TOXIN AND FASCICULIN

Ea and fasciculin are the only snake toxins for which functional sites have been sufficiently studied to allow a comparison. These sites have been identified by different experimental approaches, i.e., mutational analysis and X-ray structure of the toxin-target complex, thus it can be argued that the two approaches do not provide exactly the same type of information. A structural study reveals the residue that are in contact between the toxin and AChE, without telling us which residues are most critical in terms of binding energy. In contrast, mutational analysis reveals the residues of Ea that are energetically critical for binding to AcChoR. Comparative studies concerning various protein-protein complexes, using both X-ray crystallography and mutational analyses, have indicated a good correlation between the two methods (175-177). Therefore, comparison between the AcChoR binding site of Ea, based on mutational analysis, and the AChE binding site of fasciculin, as elucidated from a structural analysis, appears to be relevant.

A**B****C**

Clearly, the AcChoR and AchE binding sites display at least two major similarities. First, their sizes are comparable, being 10 and 12 residues, respectively. The rather large number of contacts that can be made by Ea and fasciculin residues, respectively, with AcChoR and AchE, offers an explanation as to the high affinities of both toxins for their targets, i.e., $K_d = 70$ and 100 pM , respectively (158, 171). Second, the first two loops are major factors for both toxins to recognize their targets. This is all the more interesting because a remarkable difference emerges when the two functional sites are compared (Fig. 18). The functional residues of curaremimetic toxin have their side chains orientated toward one of the two faces of the β -sheet whereas in fasciculin they belong to both faces. In other words, a glance at Fig. 18 reveals that the site of fasciculin is nearly totally observable when the toxin structure is seen from the edge of loop I, whereas the functional site of curaremimetic toxin is best observed by looking at one face of the fold. In conclusion, though slightly overlapping within the first loop, the two functional sites are clearly topographically different.

5. FUNCTIONAL SITE OF DENDROASPIN

Dendroaspin is present in mamba venom. It acts as a potent inhibitor of platelet aggregation by binding to the fibrinogen receptor, a glycoprotein called Gp IIb IIIa. It adopts a three-fingered fold (Fig. 16B). Like other protein compounds, including fibrinogen, that bind to Gp IIb IIIa, dendroaspin possesses an RGD sequence that plays a critical role for recognition of the receptor (159*a,b*). This recognition triplet is located on the third loop of the toxin. Furthermore, a cyclic peptide containing all residues of the third loop has a similar antagonist activity as compared to the whole toxin, indicating that most, if not all, functional residues of dendroaspin are located on its third loop (159*a,b*). Therefore, the receptor binding site of Gp IIb IIIa shares little, in terms of topography, with the functional sites of either a curaremimetic toxin, a fasciculin or a cardiotoxin.

←

FIG. 18. Views of the binding sites of fasciculin, an inhibitor of acetylcholinesterase (top) and of erabutoxin a (Ea), a potent antagonist of the peripheral nicotinic acetylcholine receptor (bottom). The binding elements of fasciculin were deduced from the crystal structure of the fasciculin-acetylcholinesterase complex (173) and the functional residues of Ea were determined from mutational analysis (170-172). The residues of fasciculin that make contact with acetylcholinesterase and the residues of Ea whose mutation caused a decrease in affinity for acetylcholine receptor by a factor higher than 10 are in black. A, B, and C show views of fasciculin and Ea oriented similarly. In A, the concave face of Ea faces the viewer and clearly, most functional residues of Ea are visible whereas only some of the binding residues of fasciculin are observable. An opposite situation occurs in B, where the two toxins rotate by 90° , leading to a virtually total disappearance of the functional site of Ea, whereas a substantial proportion of the fasciculin binding site remains visible.

6. CONCLUSION

The three-fingered fold is not only capable of exerting a wide variety of toxic functions, but the functional sites that are responsible for expression of these activities are located on variable regions of its surface. To further illustrate the considerable functional potentiality of this fold, we will now describe many more of its possible functions elaborated in tissues that are unrelated to venom glands.

E. Nontoxic Proteins with Three-Fingered Structures

In 1980, it was noted that the promoter of wheat germ agglutinin, a plant lectin of 41 residues, adopts a folding pattern that is grossly analogous to that of curaremimetic toxin (178). Subsequently, procolipase, a cofactor of a lipase, a protein implicated in the hydrolysis of lipids, was observed to include a domain that also resembles a three-fingered toxin, as judged from its crystallographic structure (179). More recently, CD59 has also been shown to adopt a three-fingered fold, as judged from NMR studies (180–182). CD59, previously known as MEM-43 antigen, MIRL, H19, MACIF, HRF20, and protectin, is a protein found at the surface of various cells; it inhibits complement activation by binding in a glycosylation-dependent manner to C5b–C8 and/or C9, thus preventing the formation of the membrane attack complex (183). The structure of CD59 is shown in Fig. 19. As compared to the structure of the snake toxins shown in Fig. 16, one may note both clear



FIG. 19. The solution structure of the extracellular domain of the nontoxic protein CD59, which also adopts a three-fingered fold (181). This protein, which is a member of the so-called Ly-6 family, inhibits the activation of complement. As for snake toxins, five β -sheet strands are clearly observable. Also, four disulfides are globally organized, as in snake toxins. However, in contrast to toxins, the presence of a short helical stretch is found in the third loop and an additional disulfide bond is present in the first loop.

analogies and specific deviations. Regarding the analogies, one can readily recognize the presence of three adjacent loops with five β -sheet strands. One may also observe the presence of the four conserved disulfide bonds that are located in the globular core of the fold. Two major deviations, however, can be noted. The first is the presence of a fifth disulfide bond that is located in the first loop; however, this is not a unique feature of CD59 because it was recently observed that a neurotoxin from *Bungarus* also possesses two additional half-cystines in the first loop (184) and that the cDNA coding for such an analog (211) is also known (Figs. 15 and 17). Considering their positions in the amino acid sequence, we anticipate that they form a disulfide bond that is similar to that observed in CD59. The second deviation that occurs in CD59 structure concerns the presence of a small helix turn in the third loop. These findings further illustrate the great structural variability that can occur within the three-fingered fold.

Several other proteins (185) have been observed or predicted to share the same overall three-fingered fold. However, the similarities that have been proposed in a number of the studies have not always been based on a precise structural comparison of resolved architectures. Usually they rely on the presence of the eight half-cystines that can be readily aligned with a structurally well-characterized three-fingered toxin. Therefore, in the absence of available structural data, one must consider these descriptions with caution. Among the anticipated proteins adopting a three-fingered fold are xenoxins, which have been isolated from skin secretions of *Xenopus laevis*, with 66 residues and eight half-cystines (186). The function of xenoxins remains to be elucidated. A number of cell surface proteins, known as the Ly-6 superfamily, are also likely to adopt the three-fingered fold. These include Ly-6A/E and Ly-6C antigens, implicated in T lymphocyte activation (187–189); Sgp-2, a squid glycoprotein of unknown function (190); and Ly-6SF (191) as well as CD59 (183), whose three-dimensional structure is known to indeed adopt a three-fingered fold (Fig. 19). Each of these molecules consists of a single cysteine-rich domain of 70–92 residues attached to the cell surface by GPI anchors. The Ly-6 superfamily also includes the urokinase plasminogen activator receptor, which has three domains that individually probably adopt a three-finger fold (192); the mouse antigen ThB (193); and SCIP-1, a 94-kDa protein produced by the parasite *Schistosoma mansoni* to escape the host-complement system. SCIP-1 shares functional and antigenic similarities with CD59 (194). Like the latter, SCIP-1 inhibits the formation of the complement membrane attack complex, probably by binding to C8 and C9 of the complement terminal pathway. Recently, phospholipase A₂ inhibitors, found in the plasma of crocals (195) and Thailand cobras (196), were shown to share similarities in terms of primary structures with members of the Ly-6 superfamily and hence to possess a domain that adopts a three-fingered fold.

F. On the Functional Diversity Associated with the Three-Fingered Fold

As previously discussed, the three-fingered fold is found in snake venom glands, where it exerts a variety of functionally unrelated toxic functions. It also commonly occurs in other tissues, where it exerts a multiplicity of non-toxic activities. This fold, therefore, is a remarkable natural template for nature, including snakes, to engineer a diversity of functions. The related question is how such engineering occurs.

Analysis of the cDNAs encoding the three-fingered toxins revealed that the rate of mutations is higher in the region that codes for the fold as compared to the regions coding for the signal peptide or to those corresponding to the noncoding regions, in both the 5' and 3' extremities. However, the region coding for the fold has not revealed any specific region with a particular higher rate of mutations that could account for all the known functions. Only the codons that encode the eight half-cystines and few other positions (see Fig. 17) are conserved in snake toxins, whereas most other, if not all, sequence positions seem to undergo mutations. It could be argued that the number of cDNA sequences that is available is still limited and indeed one cannot exclude the possibility that some regions of the fold are more susceptible to mutations than others. However, for these positions to be identified, the number of analyzed cDNAs needs to be considerably increased. Presently, it appears that the whole fold is capable of accommodating mutations adapted to the expression of various activities. This proposal is largely supported by studies associated with delineation of the functional sites of toxins having different biological properties. Clearly, these sites are spread on either side as well as on the edge of the β -sheet of the fold. One may argue that the surface region corresponding to the core of the fold, i.e., the region that includes the base of the three loops and the large turn that joins the first and second loops, is apparently not involved in a functional site, in contrast to the loops themselves and especially their tips. However, the number of identified functional sites is still too small to conclude that this particular region is less likely to be part of a functional site.

IV. General Conclusion on the Evolution of Snake Toxins

Toxins with PLA₂ type and three-fingered folds share a number of characteristics regarding their possible evolutionary development. Analyses of cDNAs and/or genes suggest that Viperidae snake venom gland PLA₂s and three-fingered toxins from Elapidae snake venom glands have evolved via ac-

celerated evolution. Accelerated amino acid substitutions have been previously observed to occur selectively in the active site region of serine protease inhibitors (144), the porcine elafin family (145), and the major histocompatibility complex multigene family (146). In the case of PLA₂s and three-fingered toxins, accelerated substitutions seem to occur not at a specific site but over whole protein-coding regions, except for their signal peptide, which does not undergo accelerated substitutions. It is anticipated that genes encoding PLA₂ isozymes from Viperidae venom gland and three-fingered toxins from Elapidae snake venom glands are producing new functional venom protein components by introducing rapid mutations throughout the entire mature protein.

In agreement with this is the observation that different functional sites of three-fingered toxins are spread all over most, if not all, of the fold. Presumably, a similar situation holds for PLA₂ isozymes. Therefore, the folds of snake toxins appear to constitute remarkable templates for accelerated engineering of novel functions. The processes, if any, by which such engineering occurs remain to be identified.

Snake venom PLA₂ isozymes and three-fingered toxins have various nontoxic counterparts in other tissues. It is particularly striking that enzyme inhibitors with a three-fingered fold are found in snake plasma and that PLA₂s are present in the digestive apparatus of the snake. Similar observations can be made for other snake toxins. For example, sarafotoxins synthesized in the venom gland of burrowing asps are structurally and functionally similar to endothelins that are produced in minute amounts in vessels. Intriguingly, the α/β fold that is uniquely exploited by scorpions to exert all their toxic functions is also adopted by antibacterial defensins present in the hemolymph of scorpions (1). Therefore, there may be a general evolutionary link between toxins synthesized in venom glands and nontoxic counterparts present in distinct physiological systems of venomous animals. Whether this link is at the origin of the emergence of toxins and their functional diversity is unknown.

In conclusion, it appears that snake venom glands are capable of producing toxins with a small number of folds and that these have the potential to undergo accelerated evolution toward appropriate functions. Undoubtedly, further studies are needed to understand how these glands work as efficient factories for producing multiple-function proteins.

REFERENCES

1. A. Ménez, F. Bontems, C. Roumestand, B. Gilquin, and F. Toma, *Proc. Royal Soc. Edinburgh, Sec. B (biolog. sci.)* **99**, 83 (1992).
2. F. J. Joubert, and N. Taljaard, *Biochim. Biophys. Acta* **579**, 228 (1979).

3. P. Lee Ho, M. B. Soares, T. Yamane, and I. Raw, *J. Toxicol. Toxin Rev.* **14**(3), 327 (1995).
4. A. Bdolah, Z. Wollberg, and E. Kochva, in "Snake Toxins" (A. L. Harvey, ed.), p. 415. Pergamon, Oxford, 1991.
5. D. Mebs and C. L. Ownby, *Pharmacol. Ther.* **48**, 223 (1990).
6. Z.-R. Gan, R. J. Gould, J. W. Jacobs, P. A. Friedman, and M. A. Polokoff, *J. Biol. Chem.* **263**, 198927 (1988).
7. M. S. Dennis, W. J. Henzel, R. M. Pitti, M. T. Lipari, M. A. Napier, T. A. Deisher, S. Bunting, and R. A. Lazarus, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2471 (1989).
8. J. B. Bjarnasson and J. W. Fox, *Pharmacol. Ther.* **62**, 325 (1990).
9. A. L. Harvey and A. J. Anderson, in "Snake Toxins" (A. L. Harvey, ed.), p. 131. Pergamon, Oxford, 1991.
10. H. Schweitz, C. Heurteaux, P. Bois, D. Moinier, G. Romey, and M. Lazdunski, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 878 (1994).
11. W. C. Puijk, H. M. Verheij, and G. H. de Haas, *Biochim. Biophys. Acta* **492**, 254 (1977).
12. E. A. M. Fleer, H. M. Verhaij, and G. H. de Haas, *Eur. J. Biochem.* **82**, 261 (1978).
13. T. Sakata, E. Nakamura, Y. Tsuruta, M. Tamaki, H. Teraoka, H. Tojo, T. Ono, and M. Okamoto, *Biochim. Biophys. Acta* **1007**, 124 (1989).
14. R. Verger, F. Ferrato, C. M. Mansbach, and G. Pieroni, *Biochemistry* **21**, 6883 (1982).
15. Y. Natori, K. Kawasaki, H. Arai, Y. Tamori-Natori, and S. Nojima, *J. Biochem. (Tokyo)* **93**, 631 (1983).
16. H. Tojo, T. Ono, S. Kuramitsu, H. Kagamiyama, and M. Okamoto, *J. Biol. Chem.* **263**, 5724 (1988).
17. L. A. Loeb and R. W. Gross, *J. Biol. Chem.* **261**, 10467 (1986).
18. D. K. Kim, I. Kudo, and K. Inoue, *J. Biochem. (Tokyo)* **104**, 492 (1988).
19. C. C. Leslie, D. R. Voelker, J. Y. Channon, M. M. Wall, and P. T. Zelarney, *Biochim. Biophys. Acta* **963**, 476 (1988).
20. J. Chen, S. J. Engle, J. J. Seilhamer, and J. A. Tischfield, *J. Biol. Chem.* **269**, 2365 (1994).
21. F. J. Joubert, G. S. Townshend, and D. P. Botes, *Z. Physiol. Chem.* **364**, S1717 (1983).
22. K. Nakashima, I. Nobuhisa, M. Deshimaru, M. Nakai, T. Ogawa, Y. Shimohigashi, Y. Fukumaki, M. Hattori, Y. Sakaki, S. Hattori, and M. Ohno, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5605 (1995).
23. S. Tanaka, N. Mohri, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **99**, 281 (1986).
24. N. Oda, T. Ogawa, M. Ohno, H. Sasaki, Y. Sakaki, and H. Kihara, *J. Biochem. (Tokyo)* **108**, 816 (1990).
25. N. Oda, H. Nakamura, S. Sakamoto, S.-Y. Liu, H. Kihara, C.-C. Chang, and M. Ohno, *Toxicon* **29**, 157 (1991).
26. J. M. Danse, S. Gasparini, and A. Ménez, Venom phospholipases A₂ enzymes: Structure, function and mechanism, in "Molecular Biology of Snake Venom Phospholipases A₂" (R. M. Kini, ed.), pp. 29–71. Wiley, Chichester, 1997.
27. R. M. Kini, S. Kawabata, and S. Iwanaga, *Toxicon* **24**, 1117 (1986).
28. K. Yoshizumi, S.-Y. Liu, T. Miyata, S. Saita, M. Ohno, S. Iwanaga, and H. Kihara, *Toxicon* **28**, 43 (1990).
29. S.-Y. Liu, K. Yoshizumi, N. Oda, M. Ohno, F. Tokunaga, S. Iwanaga, and H. Kihara, *J. Biochem. (Tokyo)* **107**, 400 (1990).
30. J. M. Maraganore and R. L. Heimrikson, *J. Biol. Chem.* **261**, 4797 (1986).
31. J. M. Maraganore and R. L. Heimrikson, *J. Biol. Chem.* **268**, 6064 (1993).
32. J. J. Seilhamer, W. Prouzanski, P. Vadas, S. Plant, J. A. Miller, J. Kloss, and L. K. Johnson, *J. Biol. Chem.* **264**, 5335 (1989).
33. R. M. Kramer, C. Hession, B. Johansen, G. Hayes, P. McGray, E. P. Chow, R. Tizard, and R. B. Pepinsky, *J. Biol. Chem.* **264**, 5768 (1989).

34. J. Halpert and D. Eaker, *J. Biol. Chem.* **250**, 6990 (1975).
35. E. A. Fleer, H. M. Verheij, and G. H. de Haas, *Eur. J. Biochem.* **82**, 261 (1978).
36. J. D. Thompson, D. G. Higgins, and T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
37. B. W. Dijkstra, K. H. Kalk, W. G. J. Hol, and J. Drenth, *J. Mol. Biol.* **147**, 97 (1989).
38. J. M. Maraganore, G. Merutk, W. Cho, W. Welches, F. J. Kézdy, and R. L. Heinrikson, *J. Biol. Chem.* **259**, 13839 (1984).
39. B. Francis, J. A. Coffield, L. L. Shimpson, and I. I. Kaiser, *Arch. Biochem. Biophys.* **284**, 352 (1991).
40. M. Nakai, K. Nakashima, T. Ogawa, Y. Shimohigashi, S. Hattori, C.-C. Chang, and M. Ohno, *Toxicon* **33**, 1469 (1995).
41. D. S. Dhillon, S. Condrea, J. M. Maraganore, R. L. Heinrikson, S. Benjamin, and P. Rosenberg, *Biochem. Pharmacol.* **36**, 1723 (1987).
42. C. Keith, D. S. Feldman, S. Deganello, J. Glick, K. B. Ward, E. O. Jones, and P. B. Sigler, *J. Biol. Chem.* **256**, 8602 (1981).
43. S. Brunie, J. Bolin, D. Gewirth, and P. B. Sigler, *J. Biol. Chem.* **260**, 9742 (1985).
44. A. Suzuki, E. Matsueda, T. Yamane, T. Ashida, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **117**, 730 (1995).
45. O. P. Kuipers, M. M. G. M. Thunnissen, P. de Ceus, B. W. Dijkstra, J. Drenth, H. M. Verheij, and G. H. de Haas, *Science* **244**, 82 (1989).
46. S. Kimura, T. Tanaka, I. Shimada, Y. Shiratori, S. Nakagawa, H. Nakamura, F. Inagaki, and Y. Ota, *Agric. Biol. Chem. (Tokyo)* **54**, 633 (1990).
47. G. Lambeau, P. Ancian, J.-P. Nicolas, S. H. W. Beiboer, D. Moinier, H. Verheij, and M. Lazdunski, *J. Biol. Chem.* **270**, 5534 (1995).
48. M. M. G. M. Thunnissen, E. AB, K. H. Kalk, J. Drenth, B. W. Dijkstra, O. P. Kuipers, R. Dijkman, G. H. de Haas, and H. M. Verheij, *Nature (London)* **347**, 689 (1990).
49. D. L. Scott, S. P. White, Z. Otwinowski, W. Yuan, M. H. Gelb, and P. B. Sigler, *Science*, **250**, 1541 (1990).
50. D. L. Scott, Z. Otwinowski, M. H. Gelb, and P. B. Sigler, *Science* **250**, 1563 (1990).
51. S. P. White, D. L. Scott, Z. Otwinowski, M. H. Gelb, and P. B. Sigler, *Science* **250**, 1560 (1990).
52. R. Renetseder, B. W. Dijkstra, K. Huizinga, K. H. Kalk, and J. Drenth, *J. Mol. Biol.* **200**, 181 (1988).
53. D. R. Holland, L. L. Clamecy, S. W. Muchmore, T. J. Ryde, H. M. Einspahr, B. C. Finzel, R. L. Heinrikson, and K. D. Watenpaugh, *J. Biol. Chem.* **265**, 17649 (1990).
54. D. L. Scott, A. Achari, J. C. Vidal, and P. B. Sigler, *J. Biol. Chem.* **267**, 22645 (1992).
55. G. Faure, V. Choumet, C. Bouchier, L. Camoin, J.-L. Guillaume, B. Monegier, M. Vuilhorgne, and C. Bon, *Eur. J. Biochem.* **223**, 161 (1994).
56. S. D. Aird, W. G. Kruggel, and I. I. Kaiser, *Toxicon* **28**, 669 (1990).
57. Y.-M. Wang, P.-J. Lu, C.-L. Ho, and I.-H. Tsai, *Eur. J. Biochem.* **209**, 635 (1992).
58. K. Tomoo, H. Ohishi, M. doi, T. Ishida, M. Inoue, K. Ikeda, Y. Hata, and Y. Samejima, *Biochem. Biophys. Res. Commun.* **184**, 137 (1992).
59. R. K. Arni, R. J. Ward, J. M. Gutierrez, and A. Tulinsky, *Acta Crystallogr.* **D51**, 311 (1995).
60. D. H. Fremont, D. H. Anderson, I. A. Wilson, E. A. Dennis, and N.-H. Xuong, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 342 (1993).
61. T. F. Spande and B. Witkop, *Meth. Enzymol.* **11**, 498 (1967).
62. S. Tanaka, N. Mohri, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **97**, 1377 (1985).
63. N. Mohri, S. Tanaka, T. Miyajima, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **100**, 883 (1986).
64. T. Kohzuma, N. Oda, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **106**, 1054 (1989).
65. H. Sakai, N. Oda, T. Kohzuma, S.-Y. Liu, H. Kihara, and M. Ohno, *J. Mol. Recogn.* **1**, 124 (1988).

66. M. Ohno, A. Honda, S. Tanaka, N. Mohri, T.-C. Shieh, and H. Kihara, *J. Biochem. (Tokyo)* **96**, 1183 (1984).
67. S. Nakamura, M. Nakai, K. Nakashima, T. Ogawa, Y. Shimohigashi, M. Ohno, H. Kihara, T. Yamane, and T. Ashida, *J. Mol. Recogn.* **9**, 23 (1996).
68. S. Tanaka, Y. Takahashi, N. Mohri, H. Kihara, and M. Ohno, *J. Biochem. (Tokyo)* **96**, 1443 (1984).
69. B. W. Dijkstra, K. H. Kalk, J. Drenth, G. H. de Haas, M. R. Egmond, and A. J. Slotboom, *Biochemistry* **23**, 2759 (1984).
70. S.-H. Wu, C.-J. Chen, and K.-T. Wang, *Anal. Biochem.* **129**, 345 (1983).
71. S. Tsuno, T. Ogawa, K. Nakashima, N. Oda, S. Lee, Y. Shimohigashi, H. Aoyagi, and M. Ohno, *Bull. Chem. Soc. Jpn.* **65**, 2655 (1992).
72. J. M. Maraganore, R. A. Poorman, and R. L. Henrikson, *J. Protein Chem.* **6**, 173 (1987).
73. R. M. Kini and H. J. Evans, *Toxicol.* **27**, 613 (1989).
74. H. M. Verheij, A. J. Slotboom, and G. H. de Haas, *Rev. Physiol. Biochem. Pharmacol.* **91**, 91 (1981).
75. Y. Shimohigashi, A. Tani, Y. Yamaguchi, T. Ogawa, and M. Ohno, *J. Mol. Recogn.* (1996). In press.
76. C. J. van den Bergh, A. J. Slotboom, H. M. Verheij, and G. H. de Haas, *Eur. J. Biochem.* **176**, 353 (1988).
77. C. J. van den Bergh, A. J. Slotboom, H. M. Verheij, and G. H. de Haas, *J. Cell. Biochem.* **39**, 379 (1989).
78. E. Condrea, *Toxicol.* **27**, 705 (1989).
79. B. W. Dijkstra, J. Drenth, and K. H. Kalk, *Nature (London)* **289**, 604 (1981).
80. T. Miyake, S. Inoue, K. Ikeda, K. Teshima, Y. Samejima, and T. Omori-Satoh, *J. Biochem. (Tokyo)* **105**, 565 (1989).
81. Y. Shimohigashi, A. Tani, H. Matsumoto, K. Nakashima, Y. Yamaguchi, N. Oda, Y. Takano, H. Kamiya, J. Kishino, H. Arita, and M. Ohno, *J. Biochem. (Tokyo)* **118**, 1037 (1995).
82. J. E. Fletcheer, C. C. Yang, and P. Rosenberg, *Toxicol. Appl. Pharmacol.* **66**, 39 (1982).
83. W. C. Chang, M. L. Lee, and T. B. Lo, *Toxicol.* **21**, 163 (1983).
84. J. Fohlman, D. Eaker, E. Karlsson, and S. Thesleff, *Eur. J. Biochem.* **68**, 457 (1976).
85. J. B. Harris, *Pharmacol. Ther.* **31**, 79 (1985).
86. D. Mebs, in "Toxins as Tools in Neurochemistry" (F. Hucho and Y. A. Ovchinnikov, eds.), p. 337. Walter de Gruyter, Berlin, 1983.
87. R. M. Kini and S. Iwanaga, *Toxicol.* **24**, 895 (1986).
88. C. M. Teng, Y. H. Chen, and C. Ouyang, *Biochim. Biophys. Acta* **772**, 393 (1984).
89. H. M. Verheij, M. C. Boffa, C. Rothen, M. C. Bryckaert, R. Verger, and G. H. de Haas, *Eur. J. Biochem.* **112**, 25 (1980).
90. I. Kudo, M. Murakami, S. Hara, and K. Inoue, *Biochim. Biophys. Acta* **1170**, 217 (1993).
91. R. J. Mayer and L. A. Marshall, *FASEB J.* **7**, 339 (1993).
92. S. Hara, I. Kudo, and K. Inoue, *J. Biochem. (Tokyo)* **110**, 163 (1991).
93. H. Kihara, R. Uchikawa, S. Hattori, and M. Ohno, *Biochem. Int.* **28**, 895 (1992).
94. J. M. Gutierrez, C. L. Ownby, and G. V. Odell, *Toxicol.* **22**, 115 (1984).
95. H. Kihara, S. Terashi, M. Ohno, and S. Hashimura, *Snake* **18**, 84 (1986).
96. H. Matsumoto, Y. Shimohigashi, S. Nonaka, R. Saito, Y. Takano, H. Kamiya, and M. Ohno, *Biochem. Int.* **24**, 181 (1991).
97. K. Kamata, Y. Arai, A. Sakamoto, Y. Kasuya, and Y. Samejima, *Life Sci.* **44**, 137 (1989).
98. R. Renetseder, S. Brunie, B. W. Dijkstra, J. Drenth, and P. B. Sigler, *J. Biol. Chem.* **260**, 11627 (1985).
99. J.-P. Wery, R. W. Schevitz, D. K. Clawson, J. L. Bobbitt, E. R. Dow, G. Gamboa, T. Good-

- son, R. B. Hermann, Jr., R. M. Kramer, D. B. McClure, E. D. Mihelich, J. E. Putnum, J. D. Sharp, D. H. Stark, C. Teater, M. W. Warrick, and N. D. Jones, *Science* **250**, 1560 (1991).
100. T. Ogawa, N. Oda, K. Nakashima, H. Sasaki, M. Hattori, Y. Sakaki, H. Kihara, and M. Ohno, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8557 (1992).
101. K. Nakashima, T. Ogawa, N. Oda, M. Hattori, Y. Sasaki, H. Kihara, and M. Ohno, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5964 (1993).
102. I. Nobuhisa, K. Nakashima, M. Deshimaru, T. Ogawa, Y. Shimohigashi, Y. Fukumaki, Y. Sakaki, S. Hattori, H. Kihara, and M. Ohno, *Gene* **172**, 267 (1996).
103. I.-H. Tsai, P.-J. Lu, Y.-M. Wang, C.-L. Ho and L.-L. Liaw, *Biochem. J.* **311**, 895 (1995).
104. A. M. Moura-da-Silva, M. J. I. Paine, M. R. V. Diniz, R. D. G. Theakston, and J. M. Cramp-ton, *J. Mol. Evol.* **41**, 174 (1995).
105. H. S. S. de Araujo, S. P. White, and C. L. Ownby, *Arch. Biochem. Biophys.* **326**, 21 (1996).
106. C. Bouchier, F. Ducancel, G. Guignery-Frelat, C. Bon, J.-C. Boulain, and A. Ménez, *Nucleic Acids Res.* **16**, 9050 (1988).
107. C. Bouchier, J.-C. Boulain, C. Bon, and A. Ménez, *Biochim. Biophys. Acta* **1088**, 401 (1991).
108. T. R. John, L. A. Smith, and I. I. Kaiser, *Gene* **139**, 229 (1994).
109. J. Pungercar, D. Kordis, B. Strukelj, N.-S. Liang, and F. Gubensek, *Toxicon* **29**, 269 (1991).
110. D. Kordis, J. Pungercar, B. Strukelj, N.-S. Liang, and F. Gubensek, *Nucleic Acids Res.* **18**, 4016 (1990).
111. J. Pungercar, D. Kordis, R. Jerala, M. Trstenjak-Prebanda, M. Dolinar, V. Curin-Serbec, R. Komer, and F. Gubensek, *Nucleic Acids Res.* **17**, 4367 (1989).
112. I. Krizaj, N.-S. Liang, J. Pungercar, B. Strukelj, A. Ritonja, and F. Gubensek, *Eur. J. Biochem.* **204**, 1057 (1992).
113. J. Pungercar, N.-S. Liang, B. Strukelj, and F. Gubensek, *Nucleic Acids Res.* **18**, 4601 (1990).
114. D. Kordis and F. Gubensek, *Nature Genet.* **10**, 131 (1995).
115. D. Kordis and F. Gubensek, *Eur. J. Biochem.* **240**, 83 (1996).
116. F. Ducancel, G. Guignery-Frelat, C. Bouchier, A. Ménez, and J.-C. Boulain, *Nucleic Acids Res.* **16**, 9048 (1988).
117. G. Guignery-Frelat, F. Ducancel, A. Ménez, and J.-C. Boulain, *Nucleic Acids Res.* **15**, 5892 (1987).
118. J. M. Danse, *Nucleic Acids Res.* **18**, 4608 (1990).
119. J. M. Danse, J.-M. Garnier, and J. Kempf, *Nucleic Acids Res.* **18**, 4610 (1990).
120. J. M. Danse, J.-L. Toussaint, and J. Kempf, *Nucleic Acids Res.* **18**, 4609 (1990).
121. L. Chang, and P. Wu, *Biochem. Biophys. Res. Commun.* **221**, 328 (1996).
122. F. Ducancel, G. Guignery-Frelat, C. Bouchier, A. Ménez, and J.-C. Boulain, *Nucleic Acids Res.* **16**, 9049 (1988).
123. F.-M. Pan, W.-C. Chang, and S.-H. Chiou, *Biochem. Mol. Biol. Int.* **33**, 187 (1994).
124. F.-M. Pan, M.-S. Yeh, W.-C. Chang, C.-C. Hung, and S.-H. Chiou, *Biochem. Biophys. Res. Commun.* **199**, 969 (1994).
125. M. Matsuoka, H. Itoh, T. Kozasa, and Y. Kajiro, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5384 (1988).
126. M. Strathmann and M. L. Simon, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9113 (1990).
127. S. Ohno, H. Kawasaki, S. Imajoh, K. Suzuki, M. Inagaki, H. Yokokura, T. Satoh, and H. Hidaka, *Nature (London)* **325**, 161 (1987).
128. K. Kubo, S. Ohno, and K. Suzuki, *FEBS Lett.* **223**, 138 (1987).
129. M. Nei, "Molecular Evolutionary Genetics." Columbia Univ. Press, Irvington-on-Hudson, New York, 1987.
130. J. J. Seilhamer, T. L. Randall, M. Yamanaka, and L. K. Johnson, *DNA* **5**, 512 (1986).

131. R. Breathnach, C. Benoist, K. O'Hara, F. Gannon, and P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4953 (1978).
132. M. Kimura, "The Neutral Theory of Molecular Evolution." Cambridge Univ. Press, Cambridge, England, 1983.
133. T. Miyata and T. Yasunaga, *J. Mol. Evol.* **16**, 23 (1980).
134. M. Nei and T. Cojobori, *Mol. Biol. Evol.* **3**, 418 (1986).
135. S. Kuhara, F. Matsumoto, S. Futamura, A. Fujita, T. Shimohara, T. Takagi, and Y. Sakaki, *Nucleic Acids Res.* **12**, 88 (1984).
136. M. Zuker and P. Steigler, *Nucleic Acids Res.* **9**, 133 (1981).
137. M. Nei and L. Jin, *Mol. Biol. Evol.* **6**, 290 (1989).
138. T. Ogawa, M. Kitajima, K. Nakashima, Y. Sakaki, and M. Ohno, *J. Mol. Evol.* **41**, 867 (1995).
139. P. J. Heise, L. R. Maxson, H. G. Dowling, and S. B. Hedges, *Mol. Biol. Evol.* **12**, 258 (1995).
140. T. H. Jukes and C. R. Cantor in "Mammalian Protein Metabolism" (H. N. Munro, ed.), p. 21. Academic Press, New York, 1969.
141. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
142. J. Felsenstein, *Evolution* **39**, 783 (1985).
143. K. Nakashima, I. Nobuhisa, M. Deshimaru, T. Ogawa, Y. Shimohigashi, Y. Fukumaki, M. Hattori, Y. Sakaki, S. Hattori, and M. Ohno, *Gene* **152**, 209 (1995).
144. R. E. Hill and N. D. Hastie, *Nature (London)* **326** 96 (1987).
145. I. Tamechika, M. Itakura, Y. Saruta, M. Furukawa, A. Kato, S. Tachibana, and S. Hirose, *J. Biol. Chem.* **271**, 7012 (1996).
146. A. L. Hughes, and M. Nei, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 958 (1989).
147. M. J. Dufton and R. C. Hider, *CRC Crit. Rev. Biochem.* **14**, 113 (1983).
148. S. Zinn-Justin, C. Roumestand, B. Gilquin, F. Bontems, A. Ménez, and F. Toma, *Biochemistry* **31**, 11335 (1992).
149. R. Legoas, S. R. Laplante, A. Mikou, M. A. Delsuc, E. Guittet, M. Robin, I. Charpentier, and J. Y. Lallemand, *Biochemistry* **31**, 4867 (1992).
150. J. C. Dewan, G. A. Grant, and J. C. Sacchettini, *Biochemistry* **33**, 13147 (1994).
151. R. E. Oswald, M. J. Sutcliffe, M. Bamberger, R. H. Loring, E. Braswell, and C. M. Dobson, *Biochemistry* **30**, 4901 (1991).
152. V. A. Chiappinelli, in "Snake Toxins" (A. L. Harvey, ed.), p. 223. Pergamon, Oxford, 1991.
153. B. Gilquin, C. Roumestand, S. Zinn-Justin, A. Ménez, and F. Toma, *Biopolymers* **33**, 1659 (1993).
154. A. Bilwes, B. Rees, D. Moras, R. Ménez, and A. Ménez, *J. Mol. Biol.* **239**, 122 (1994).
155. I. Ségalas, C. Roumestand, S. Zinn-Justin, B. Gilquin, R. Ménez, A. Ménez, and F. Toma, *Biochemistry* **34**, 1248 (1995).
156. J. P. Albrand, M. J. Blackledge, F. Pascaud, M. Hollecker, and D. Marion, *Biochemistry* **34**, 5923 (1995).
157. J. R. deWeille, H. Schweitz, P. Maes, A. Tartar, and M. Lazdunski, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2437 (1991).
158. M.-H. le Du, P. Marchot, P. E. Bougis, and J. Fontecilla-Camps, *J. Biol. Chem.* **267**, 22122 (1992).
- 159a. M. J. Sutcliffe, M. Jaseja, E. I. Hyde, X. Lu, and J. A. Williams, *Nature (Struct. Biol.)* **1**, 802 (1994).
- 159b. J. A. Williams, X. Lu, S. Rahman, C. Keating, and V. Kakkar, *Biochem. Soc. Trans.* **21**, 735 (1992).
- 159c. V. Winckler-Dietrich, Thèse de l'Université Paris XI—Orsay, France, 1996.
160. A. Adem, A. Åsblom, G. Johansson, P. M. Mbugua, and E. Karlsson, *Biochim. Biophys. Acta* **968**, 340 (1988).

161. C. Cervenansky, F. Dajas, A. L. Harvey, and E. Karlsson, in "Snake Toxins" (A. L. Harvey, ed.), p. 303. Pergamon, Oxford, 1991.
162. A. Ménez, E. Gatineau, C. Roumestand, A. L. Harvey, L. Mouawad, B. Gilquin, and F. Toma, *Biochimie* **72**, 575 (1990).
163. G. Faure, J.-C. Boulain, F. Bouet, Th. Montenay-Garestier, P. Fromageot, and A. Ménez, *Biochemistry* **22**, 2068 (1983).
164. T. Endo and N. Tamiya, in "Snake Toxins" (A. L. Harvey, ed.), p. 165. Pergamon, Oxford, 1991.
165. E. Gatineau, F. Toma, Th. Montenay-Garestier, M. Takechi, P. Fromageot, and A. Ménez, *Biochemistry* **26**, 8046 (1987).
166. J.-M. Grognet, A. Ménez, A. Drake, K. Hayashi, I. E. G. Morrison, and R. C. Hider, *Eur. J. Biochem.* **172**, 383 (1988).
167. E. Gatineau, M. Takechi, F. Bouet, P. Mansuelle, H. Rochat, A. L. Harvey, Th. Montenay-Garestier, and A. Ménez, *Biochemistry* **29**, 6480 (1990).
168. P. Marchot, P. E. Bougis, B. Ceard, J. Van Rietschoten, and H. Rochat, *Biochem. Biophys. Res. Commun.* **153**, 642 (1988).
169. F. Ducancel, C. Bouchier, T. Tamiya, J.-C. Boulain, and A. Ménez, in "Snake Toxins" (A. L. Harvey, ed.), p. 385. Pergamon, Oxford, 1991.
170. M. Hervé, L. Pillet, P. Humbert, O. Trémeau, F. Ducancel, C. Hirth, and A. Ménez, *Eur. J. Biochem.* **208**, 125 (1992).
171. L. Pillet, O. Trémeau, F. Ducancel, P. Drevet, S. Zinn-Justin, S. Pinkasfeld, J.-C. Boulain, and A. Ménez, *J. Biol. Chem.* **268**, 909 (1993).
172. O. Trémeau, C. Lemaire, P. Drevet, S. Pinkasfeld, F. Ducancel, J.-C. Boulain, and A. Ménez, *J. Biol. Chem.* **270**, 9362 (1995).
173. P. Bourme, P. Taylor, and P. Marchot, *Cell* **83**, 503 (1995).
174. M. Harel, G. J. Kleywegt, R. B. G. Ravelli, I. Silman, and J. L. Sussman, *Structure* **3**, 1355 (1995).
175. B. C. Cunningham and J. A. Wells, *J. Mol. Biol.* **234**, 554 (1993).
176. L. Prasad, S. Sharma, M. Vandonselaar, W. J. Quail, J. S. Lee, E. B. Waygood, K. S. Wilson, Z. Dauter, and L. T. J. Delbaere, *J. Biol. Chem.* **268**, 10705 (1993).
177. W. Dall'Acqua, E. R. Goldman, E. Eisenstein, and R. A. Mariuzza, *Biochemistry*, **33**, 9667 (1996).
178. J. Drenth, B. W. Low, J. S. Richardson, and C. S. Wright, *J. Biol. Chem.* **255**, 2652 (1980).
179. H. van Tilbeurgh, L. Sarda, R. Verger, and C. Cambillau, *Nature (London)* **359**, 159 (1992).
180. B. Kieffer, P. C. Driscoll, I. D. Campbell, A. C. Willis, P. A. van der Merwe, and S. J. Davis, *Biochemistry* **33**, 4471 (1994).
181. C. M. Fletcher, R. A. Harrison, P. J. Lachmann, and D. Neuhaus, *Structure*, **2**, 185 (1994).
182. C. M. Fletcher, R. A. Harrison, P. J. Lachmann, and D. Neuhaus, *Protein Science* **2**, 2015 (1993).
183. H. Ninomya and P. J. Sims, *J. Biol. Chem.* **267**, 13675 (1992).
184. S. D. Aird, G. C. Womble, J. R. Yates, and P. R. Griffin, IV Simposio Da Sociada de Braziliara de Toxinologia (abstract) 1995.
185. B. Rees and A. Bilwes, *Chem. Res. Toxicol* **6**, 385, (1993).
186. H. V. J. Kolbe, A. Huber, P. Cordier, U. B. Rasmussen, B. Bouchon, M. Jaquinod, R. Vlasak, E. C. Délot, and G. Kreil, *J. Biol. Chem.* **268**, 16458 (1993).
187. K. P. Leclair, R. G. E. Palfree, P. M. Flood, U. Hämmerling, and A. Bothwell, *EMBO J.* **5**, 3227 (1986).
188. R. G. E. Palfree, S. Sirlin, F. J. Dumont, and U. Hämmerling, *J. Immunol.* **140**, 305 (1988).
189. T. J. Fleming, C. O'hUigin, and T. R. Malek, *J. Immunol.* **150**, 5379 (1993).

190. A. F. Williams, A. C.-D. Tse, and J. Gagnon, *Immunogenetics* **27**, 265 (1988).
191. A. F. Williams, *Cell Biol. Int. Rep.* **15**, 769 (1991).
192. M. Ploug and V. Ellis, *FEBS Lett.* **349**, 163 (1994).
193. T. P. Gumley, I. F. C. McKenzie, C. A. Kozak, and M. S. Sandrin, *J. Immunol.* **149**, 2615 (1992).
194. M. Parizade, R. Arnon, P. J. Lachmann, and Z. Fishelson, *J. Exp. Med.* **1799**, 1625 (1994).
195. C. L. Forte Diaz, Y. L. J. Ewell, C. R. Diniz, and T.-Y. Liu, *J. Biol. Chem.* **269**, 15646 (1994).
196. N. Ohkura, S. Inoue, K. Ikeda, and K. Hayashi, *Biochem. Biophys. Res. Commun.* **204**, 1212 (1994).
197. N. Tamiya and H. Abe, *Biochem. J.* **130**, 547 (1972).
198. S. Nishida, Y. Kokubun, and N. Tamiya, *Biochem. J.* **226**, 879 (1985).
199. N. Maeda and N. Tamiya, *Biochem. J.* **153**, 79 (1976).
200. E. Karlsson, C. Risinger, M. Jolkkonen, C. Wernstedt, and A. Adem, *Toxicol.* **29**, 521 (1991).
201. E. Karlsson, P. M. M'Bugua, and D. Rodriguez-Ithurralde, *Pharmacol. Ther.* **30**, 259 (1985).
202. R. S. McDowell, M. S. Dennis, A. Louie, M. Shuster, M. G. Mulkerrin, and R. A. Lazarus, *Biochemistry* **31**, 4766 (1992).
203. N. Fuse, T. Tsuchiya, Y. Nonomura, A. Ménez, and T. Tamiya, *Eur. J. Biochem.* **193**, 629 (1990).
204. T. Tamiya, A. Lamouroux, J.-F. Julien, B. Grima, J. Mallet, P. Fromageot, and A. Ménez, *Biochimie* **67**, 185 (1985).
205. K. Obara, N. Fuse, T. Tsuchiya, Y. Nonomura, A. Ménez, and T. Tamiya, *Nucleic Acids Res.* **17**, 10490 (1989).
206. F. Ducancel, G. Guignery-Frelat, J.-C. Boulain, and A. Ménez, *Toxicol.* **28**, 119 (1990).
207. M. S. L. Yeo, K. Jeyselán, M. C. M. Chung, P. Gopalakrishnakone, C. N. Tan, and H. A. Wong, *Toxicol.* **31**, 53 (1993).
208. E. G. Rowan, F. Ducancel, Y. Doljansky, A. L. Harvey, J.-C. Boulain, and A. Ménez, *Nucleic Acids Res.* **18**, 1639 (1990).
209. F. Ducancel, E. G. Rowan, E. Cassar, A. L. Harvey, A. Ménez, and J.-C. Boulain, *Toxicol.* **29**, 516 (1991).
210. J. M. Danse and J.-M. Garnier, *Nucleic Acids Res.* **18**, 1050 (1990).
211. J. M. Danse, J.-M. Garnier, and J. Kempf, *Nucleic Acids Res.* **18**, 1045 (1990).

Index

A

- ADRP, *see* Autosomal dominant retinitis pigmentosa
- α - ϵ model, *see* Elongation, protein
- Affinity chromatography, *see specific deoxynucleotide kinases*
- Allosteric three-site model, *see* Elongation, protein
- Androgen receptor (AR)
 - gene structure
 - DNA binding domain, 290–291
 - overview, 290
 - steroid binding domain, 291
 - posttranscriptional regulation of gene, 296–298
 - transcription
 - autoregulation, 293, 303–304
 - characterization of 5' flanking region, 294–296
 - cyclic-AMP response element, 295
 - enhancers, 291, 298
 - negative regulation, 295–296
 - second messengers in regulation, 293–294
 - suppressors/repressors, 292–293, 298
 - trans-acting factors, 291–292
- AR, *see* Androgen receptor
- Autosomal dominant retinitis pigmentosa (ADRP), rhodopsin mutation, 28–29

B

- Bacillus subtilis*, membrane–chromosome replication
 - advantages of *Bacillus subtilis* in study, 36
 - chromosomal membrane attachment sites, 44–45, 49, 51
 - chromosome initiation mutants, 39
 - chromosome replication initiation *in vitro*, 46–47
 - comparison to *Escherichia coli*, 36–38, 47, 49

configurations of chromosome binding to membrane, 47–48

dnab

- encoded protein properties, 41, 44
 - locus, 41
 - open reading frames, 41, 44
 - role in origin–membrane binding, 40–41, 43–44, 46, 48–49
 - history of study and early evidence for membrane–chromosome association, 36–38
 - membrane attachment to terminus, 47
 - membrane sliding models of chromosome separation, 49–50
 - origin–membrane complex preparation, 39–40
 - terminus–membrane complex preparation, 39–40
- Bacteriorhodopsin, structural homology to other opsins, 25

C

- Casein kinase II (CKII)
 - regulation, 97–98
 - Saccharomyces cerevisiae*, *see* Casein kinase II, *Saccharomyces cerevisiae*
 - Schizosaccharomyces pombe* enzyme, 99
 - substrates, 96, 98
 - subunits, 96–97
 - tissue distribution in mammals, 98
 - transgenic mice, 98–99
- Casein kinase II, *Saccharomyces cerevisiae*
 - catalytic subunit structure, 101
 - Cdc37 activation, 112–113
 - conditional allele analysis, 102, 104
 - functions
 - cell cycle progression, 111–113, 116, 127–128
 - cell polarity, 113–117
 - flocculation, 109–110

- Casein kinase II (*cont.*)
 gene expression and chromatin structure, 110–111, 128–129
 ion homeostasis, 117–118, 128
 signal transduction, 127–128
 heat shock protein, 90 association, 113
 multicopy suppression screens
 alleles in screens, 104
CDC37, 105–106, 112
 controls, 105
SUN2, 107–109
ZDS1, 106–107, 116
ZDS2, 106–107, 116
 null allele analysis, 102–104
 regulation, 129
 regulatory subunit structure, 101
 substrates
 criteria for identification, 119–120
 eIF-2 α , 120
 Nopp140, 121
 prediction by recognition site, 121–127
 Srp40, 121
 topoisomerase II, 120–121
 subunits and genes, 100
CD, *see* Circular dichroism
CDC37, multicopy suppression of casein kinase yeast mutants, 105–106, 112
 Chromosome replication, *see Bacillus subtilis*, membrane–chromosome replication
 Circular dichroism (CD)
 guanine quadruplex, 82, 84–85
 pyr*pur*pyr triplex, 65
 three-fingered fold snake toxins, 339
 CKII, *see* Casein kinase II
 Cone cell
 chloride ion and pigment binding, 12–13
 heredity of color vision defects, 11–12
 iodopsin, 4–5
 spectral properties of recombinant cone pigments, 11–13
- D
- Dendroaspin, functional site elucidation, 353
 Deoxyadenosine kinase, *Lactobacillus acidophilus* R-26
 affinity labeling, 244–247
 amino acid composition, 244, 246
 assay, 212, 214
 conformational probing with limited proteolysis, 243–244
 gene cloning
 colony screening by polymerase chain reaction, 234–235
 hybridization probes, 232, 234
 identification as deoxyguanosine kinase gene
 mass determination, 240
 protein sequencing, 238–240, 251
 kinetic mechanism, 224–226
 pH dependence, 224
 purification
 Blue Sepharose pseudoaffinity chromatography, 214–217
 deoxynucleoside derivatives for affinity chromatography
 base modifications, 218–219
 bisubstrate analogs, 219–220
 deoxyribose modification, 217–218
 dNTP-Sepharose chromatography, 220–223
 recombinant kinases, 223–224
 UDP-Sepharose, 217
 regulation, 251–252
 sequencing
 gene, 235–237
 homology with other nucleoside kinases, 235, 238
 protein, 230, 232
 site-directed mutagenesis
 deoxynucleoside-binding site, 248–249
 heterodimer interface, 249–250
 thermostability, 214
 Deoxycytidine kinase, *Lactobacillus acidophilus* R-26
 affinity labeling, 244–247
 amino acid composition, 244, 246
 assay, 212, 214
 conformational probing with limited proteolysis, 243–244
 gene cloning
 colony screening by polymerase chain reaction, 234–235
 hybridization probes, 232, 234
 identification as deoxyguanosine kinase gene
 mass determination, 240
 protein sequencing, 238–240, 251

- recombinant protein engineering evidence, 241–243
- inhibition by dNTP end products, 228–230
- kinetic mechanism, 224–226
- pH dependence, 224
- purification
 - Blue Sepharose pseudoaffinity chromatography, 214–217
 - deoxynucleoside derivatives for affinity chromatography
 - base modifications, 218–219
 - bisubstrate analogs, 219–220
 - deoxyribose modification, 217–218
 - dNTP-Sepharose chromatography, 220–223
 - recombinant kinases, 223–224
 - UDP-Sepharose, 217
- regulation, 251–252
- sequencing
 - gene, 235–237
 - homology with other nucleoside kinases, 235, 238
 - protein, 230, 232, 239
- site-directed mutagenesis
 - deoxynucleoside-binding site, 248–249
 - heterodimer interface, 249–250
- Deoxyguanosine kinase, *Lactobacillus acidophilus* R-26
 - affinity labeling, 244–247
 - amino acid composition, 244, 246
 - assay, 212, 214
 - conformational probing with limited proteolysis, 243–244
 - gene cloning
 - colony screening by polymerase chain reaction, 234–235
 - hybridization probes, 232, 234
 - inhibition by dNTP end products, 228–230
 - kinetic mechanism, 224–226
 - pH dependence, 224
 - purification
 - Blue Sepharose pseudoaffinity chromatography, 214–217
 - deoxynucleoside derivatives for affinity chromatography
 - base modifications, 218–219
 - bisubstrate analogs, 219–220
 - deoxyribose modification, 217–218
 - dNTP-Sepharose chromatography, 220–223
 - recombinant kinases, 223–224
 - UDP-Sepharose, 217
 - regulation, 251–252
 - sequencing
 - gene, 235–237
 - homology with other nucleoside kinases, 235, 238
 - protein, 230, 232, 239
 - site-directed mutagenesis
 - deoxynucleoside-binding site, 248–249
 - heterodimer interface, 249–250
- DNA quadruplex, *see* Guanine quadruplex
- DNA triplex
 - discovery, 56–57
 - hydrogen bonding, 58–59, 63–64
 - ionic strength dependence, 58
 - ligand interactions
 - intercalators, 77–79
 - minor groove binding agents, 77
 - molecularity, 60
 - pH dependence, 65
 - pur*pur*pyr triplex
 - melting transition, 67–68
 - metal ion dependence, 67
 - nuclear magnetic resonance, 77
 - sequence dependence, 68
 - ultraviolet spectroscopy mixing curves, 67
 - pyr*pur*pyr triplex
 - circular dichroism, 65
 - Fourier transform infrared spectroscopy, 76
 - nuclear magnetic resonance, 61, 63, 65, 76–77
 - polyacrylamide gel electrophoresis, 61–63, 66

DNA triplex (*cont.*)
 sequence dependence, 67
 ultraviolet spectroscopy mixing curves,
 61–63
 pur/pyr•pur•pyr triplex properties, 68–69
 sequence dependence, 57–58, 65, 67–68
 strand orientation, 59–60
 thermodynamic analysis
 differential scanning calorimetry, 71–73,
 75
 ultraviolet spectroscopy melting curves,
 63–64, 67, 69–71
 types, overview, 57
 DOCK, guanine quadruplex ligand design
 principle, 89–90
 screening of compounds, 90–91
 DODC, *see* Diethyloxadicarbocyanine
 DSC, *see* Differential scanning calorimetry

E

Elongation, protein
 models
 α - ϵ model, 184–188, 200–202
 allosteric three-site model, 180–182
 hybrid-site model, 182–184
 ribosome shape, determination by scatter-
 ing
 angular envelope function, 191–193,
 198
 coherent scattering, 190–191
 comparison with electron microscopy,
 197–199
 contrast, 188
 direct shape restoration, 192–193
 hybrid ribosome studies, 193–194
 model fitting and nonlinear refinement,
 188–189, 195–197
 parameter determination, 189
 scattering length, 189–190
 small-angle scattering intensities from
 four-phase systems, 194–195
 steps
 decoding, 178–179
 peptide bond formation, 178–179
 translocation, 178–179
 transfer RNA localization on ribosomes
 with spin-dependent scattering,
 199–201

Erabutoxin a, functional site elucidation,
 350–351
 Estrogen receptor
 gene structure
 DNA binding domain, 290–291
 overview, 290
 steroid binding domain, 291
 transcription
 autoregulation, 303–304
 characterization of 5' flanking region,
 301–303
 enhancers, 291
 promoter types, 302–303
 suppressors/repressors, 292–293
 trans-acting factors, 291–292
 Ethidium, DNA triplex binding, 77–79

F

Fasciculin, functional site elucidation, 351
 FGF, *see* Fibroblast growth factor
 FGFRHS, *see* Fibroblast growth factor re-
 ceptor heparan sulfate
 FGFRTK, *see* Fibroblast growth factor re-
 ceptor tyrosine kinase
 Fibroblast growth factor (FGF)
 extraction from receptor complexes, 152
 heparin-binding domain, 151–152
 hepatocytes
 abnormalities in hepatoma cells,
 164–166
 normal cell properties, 166–167
 resting cell properties, 167–168
 homologous factors and sequences, 136,
 140
 nomenclature of peptides, 136–137
 prostate
 proliferation role
 normal cells, 168–173
 PAP tumors, 170–171
 receptor isoforms in malignant tumors,
 172–173
 tumor progression model, 173
 receptor dimer–ligand ternary complex
 conformational models, 159–163
 fibroblast growth factor receptor tyro-
 sine kinase ectodomain and growth
 factor binding, 157–158, 164
 modeling of structure, 155–157

- mutation studies, 159–164
 - phosphorylation role, 158–161
 - sequences, 136, 138–139, 142
 - signal sequences, 142
 - three-dimensional structures of FGF-1 and FGF-2, 142–143
 - Fibroblast growth factor receptor heparan sulfate (FGFRHS)
 - assembly of fibroblast growth factor receptor dimeric complex, 155–156
 - fibroblast growth factor binding, 154
 - fibroblast growth factor receptor tyrosine kinase binding and divalent cation requirement, 154–155
 - receptor dimer–ligand ternary complex conformational models, 159–163
 - fibroblast growth factor receptor tyrosine kinase ectodomain and growth factor binding, 157–158, 164
 - modeling of structure, 155–157
 - mutation studies, 159–164
 - phosphorylation role, 158–161
 - signal transduction overview, 136–137, 141–142
 - Fibroblast growth factor receptor tyrosine kinase (FGFR TK)
 - assembly of fibroblast growth factor receptor dimeric complex, 155–156
 - autophosphorylation sites, 149–150
 - carboxy-terminal tail, 151
 - extracellular domain
 - amino-terminal loop I and FGFR ρ isoform, 144–145
 - interloop I/II acidic box, 145–146
 - interloop II/III and loop II as secondary FGF binding sites, 146–149
 - invariant loop II as binding site, 146
 - fibroblast growth factor binding, 146–149, 152–154
 - fibroblast growth factor receptor complex
 - fibroblast growth factor receptor heparan sulfate binding, 154–155
 - genes, 136, 143
 - hepatocytes
 - normal cell properties, 166–167
 - receptor abnormalities in hepatoma cells, 164–166
 - resting cell properties, 167–168
 - isoforms and structures, 136, 141, 143–151
 - juxtamembrane domains, 149
 - kinase domain, 149–151
 - mutations in disease, 159, 161, 163
 - prostate
 - isoforms in malignant tumors, 172–173
 - proliferation role
 - normal cells, 168–173
 - PAP tumors, 170–171
 - tumor progression model, 173
 - receptor dimer–ligand ternary complex conformational models, 159–163
 - fibroblast growth factor receptor tyrosine kinase ectodomain and growth factor binding, 157–158, 164
 - modeling of structure, 155–157
 - mutation studies, 159–164
 - phosphorylation role, 158–161
 - signal transduction overview, 136–137, 141–142, 150
 - transmembrane domain, 149
 - Fourier transform infrared spectroscopy (FTIR)
 - pyr*pur*pyr triplex, 76
 - rhodopsin, light-induced conformational change, 14–15, 19, 23–24, 29
 - FTIR, *see* Fourier transform infrared spectroscopy
- G
- Glucocorticoid receptor
 - gene structure
 - DNA binding domain, 290–291
 - overview, 290
 - steroid binding domain, 291
 - transcription
 - autoregulation, 303–304
 - characterization of 5' flanking region, 299
 - downregulation, 300
 - enhancers, 291, 299
 - suppressors/repressors, 292–293
 - trans-acting factors, 291–292, 298
 - G protein-coupled receptor, *see* Rhodopsin
 - Guanine quadruplex
 - biological roles, 80–81
 - cation dependence, 81–82, 84–85
 - circular dichroism, 82, 84–85

Guanine quadruplex (*cont.*)
 equilibrium ultracentrifugation and concentration profile, 83–84
 guanine quartet structure, 79–80
 kinetics of formation, 82
 ligand interactions
 diethyloxadycarbocyanine, 90–91
 DOCK method in design
 principle, 89–90
 screening of compounds, 90–91
 types of ligands, 89
 nuclear magnetic resonance, structure determination, 83–84, 86–88
 strand configurations, 80, 88–89
 thermodynamic analysis, 85–86
 X-ray crystallography, 86, 88

H

H-DNA, *see* DNA triplex
 Heparan, *see* Fibroblast growth factor receptor heparan sulfate
 Hepatocyte, fibroblast growth factor receptor complex
 abnormalities in hepatoma cells, 164–166
 normal cell properties, 166–167
 resting cell properties, 167–168
 Hybrid-site model, *see* Elongation, protein

I

Infrared spectroscopy, *see* Fourier transform infrared spectroscopy

L

Lactobacilli, deoxynucleoside salvage deoxynucleoside kinases, 208–209
Lactobacillus acidophilus R-26
 deoxynucleoside kinases, *see* Deoxyadenosine kinase; Deoxycytidine kinase; Deoxyguanosine kinase
 deoxynucleoside metabolism, contrasts with other bacteria, 209–212
 growth requirements, 207
 mass determination, 240

nucleoside, 2-deoxyribosyltransferases, 208–209
 thymidylate synthase, 210–212

M

Membrane–chromosome replication, *see* *Bacillus subtilis*, membrane–chromosome replication
 Metallothionein (MT)
 cis-acting elements, overview, 261
 functions, 258, 261
 gene structure conservation between humans and trout, 258–259
 induction
 metal specificity, 259–261, 275, 279–280
 non-metals, 259
 metal-responsive element
 basal roles
 human MT-IG MREs, 263–266
 trout MT-B MREs, 266–268
 consensus sequence, 261
 distal and proximal element interactions, 268–270
 synergistic interactions in response to zinc, 263
 metal-responsive element trans-acting factors
 assays, 274, 276
 MTF-1
 binding assays, 276
 cotransfection studies in human, mouse, and trout cell lines, 280, 282–284
 gene cloning, 274–275
 metal specificity, 275, 279–280, 285
 nomenclature, 275
 zinc fingers, 276, 278–279
 table, 272–273
 oxidative stress response of trout MT-B promoter, 270–271, 274
 Metal-responsive element (MRE)
 basal roles
 human MT-IG MREs, 263–266
 trout MT-B MREs, 266–268
 consensus sequence, 261
 distal and proximal element interactions, 268–270

- synergistic interactions in response to zinc, 263
- trans-acting factors
assays, 274, 276
- MTF-1
binding assays, 276
cotransfection studies in human, mouse, and trout cell lines, 280, 282–284
gene cloning, 274–275
metal specificity, 275, 279–280, 285
nomenclature, 275
zinc fingers, 276, 278–279
table, 272–273
- MRE, *see* Metal-responsive element
- MT, *see* Metallothionein
- MTF-1
binding to metal-responsive element, assays, 276
cotransfection studies in human, mouse, and trout cell lines, 280, 282–284
gene cloning, 274–275
metal specificity, 275, 279–280, 285
nomenclature, 275
zinc fingers, 276, 278–279
- N
- Neutron scattering
ribosome shape determination
angular envelope function, 191–193, 198
coherent scattering, 190–191
comparison with electron microscopy, 197–199
contrast, 188
direct shape restoration, 192–193
hybrid ribosome studies, 193–194
model fitting and nonlinear refinement, 188–189, 195–197
parameter determination, 189
scattering length, 189–190
small-angle scattering intensities from four-phase systems, 194–195
transfer RNA localization on ribosomes with spin-dependent scattering, 199–201
- NMR, *see* Nuclear magnetic resonance
- Nuclear magnetic resonance (NMR)
guanine quadruplex, 83–84, 86–88
pur*pur*pyr triplex, 77
pyr*pur*pyr triplex, 61, 63, 65, 76–77
- P
- PAGE, *see* Polyacrylamide gel electrophoresis
- Phospholipase A₂, snake venom
catalytic mechanism, 313–315, 318
chemical modification studies, 316–317
classification, 311, 313
complementary DNA structures, 324, 326
Darwinian-type accelerated evolution
intron analysis, 338–339
numbers of nucleotide substitutions per nonsynonymous site, 330–331, 335–336, 338
numbers of nucleotide substitutions per site, 327, 330–331, 335–336, 338
numbers of nucleotide substitutions per synonymous site, 330–331, 335–336, 338
phylogenetic relationships using untranslated regions of complementary DNA, 336–337
gene structures, 326–327
physiological effects, 321–322, 324
primary structures, 311–312
quaternary structures, 315–316
refolding studies, 317
substrate specificities, 317–321
three-dimensional structures, 313–315
- Polyacrylamide gel electrophoresis (PAGE),
pyr*pur*pyr triplex, 61–63, 66
- Progesterone receptor
gene structure
DNA binding domain, 290–291
overview, 290
steroid binding domain, 291
transcription
autoregulation, 303–304
enhancers, 291
estrogen regulation, 301
promoter types, 301
suppressors/repressors, 292–293
trans-acting factors, 291–292

Prostate, fibroblast growth factor receptor complex
 proliferation role
 normal cells, 168–173
 PAP tumors, 170–171
 receptor isoforms in malignant tumors, 172–173
 tumor progression model, 173
 Protein elongation, *see* Elongation, protein
 Protein fold, *see* Three-fingered fold

Q

Quadruplex, *see* Guanine quadruplex

R

Rhodopsin
 absorption spectroscopy
 intermediates, 2, 19–20
 mutants, 10–11
 active conformation, molecular switches and determinants, 20–22
 chromophores
 binding site, 7–9
 isomerization, 3–4, 13–14, 20–22
 types, 2
 cytoplasmic domain
 rhodopsin kinase interactions, 9
 transducin interactions, 9–10, 22
 extracellular domain
 cysteine roles, 6
 glycosylation, 7
 mutation analysis, 6–7
 light-induced conformational change
 coupling to transducin activation, 22–25
 cytoplasmic surface changes, 18–20
 electron paramagnetic resonance analysis, 19
 fluorescence studies of tryptophan, 15–16
 Fourier transform infrared spectroscopy analysis, 14–15, 19, 23–24, 29
 functional interactions between transmembrane helices 3 and 6, 15–18
 metal ion binding site introduction in analysis, 18
 photolysis pathways, 13

membrane-embedded domain and chromophore-binding site, 7–9
 mutations in disease
 autosomal dominant retinitis pigmentosa, 28–29
 congenital night blindness, 29
 phosphorylation, 5–6, 9
 purification from cattle, 5
 structure
 peptide characterization, 28
 projection mapping of structure, 25–28
 secondary structure, 2–3, 5
 Ribosome, *see* Elongation, protein
 Rod cell
 opsin, *see* Rhodopsin
 photosensitivity, 3–4

S

Snake toxin, *see also* Phospholipase A₂, snake venom; Three-fingered fold
 classification, 310
 protein folds, 310–311
 species producing toxins, 309
 structural overview, 308, 310
 Steroid receptor, *see also* Androgen receptor; Estrogen receptor; Glucocorticoid receptor; Progesterone receptor
 gene structure
 DNA binding domain, 290–291
 overview, 290
 steroid binding domain, 291
 transcription
 autoregulation, 303–304
 enhancers, 291
 suppressors/repressors, 292–293
 trans-acting factors, 291–292
 SUN2, multicopy suppression of casein kinase yeast mutants, 107–109

T

Three-fingered fold
 functional diversity engineering, 353
 nontoxic protein examples, 354–355
 snake toxins
 complementary DNA

- evolutionary analysis, 345, 348, 356–357
 - sequence homology, 344–345
 - functional site identification
 - chemical modification studies, 348–350
 - comparison of curaremimetic toxin and fasciculin sites, 351, 353
 - dendroaspin, 353
 - erabutoxin a, 350–351
 - fasciculin, 351
 - mutagenesis studies, 350–351
 - toxin α , 348–350
 - toxin γ , 348–350
 - X-ray crystallography, 351
 - receptor specificity, 343–344
 - structures, 339–341
 - Toxin, *see* Snake toxin
 - Transducin
 - light-induced conformational change coupling to transducin activation with rhodopsin, 22–25
 - rhodopsin interactions, 9–10, 22
 - Transfer RNA (tRNA), *see* Elongation, protein
 - Translation, *see* Elongation, protein
 - Triple helix, *see* DNA triplex
 - tRNA, *see* Transfer RNA
 - Tyrosine kinase, *see* Fibroblast growth factor receptor tyrosine kinase
- U
- Ultraviolet spectroscopy, melting curves and analysis of DNA
 - guanine quadruplex, 86
 - triple helices, 63–64, 67, 69–71
- V
- Venom, *see* Snake toxin
- Z
- ZDS1, multicopy suppression of casein kinase yeast mutants, 106–107, 116
 - ZDS2, multicopy suppression of casein kinase yeast mutants, 106–107, 116

ISBN 0-12-540059-4



9 780125 400596

90065