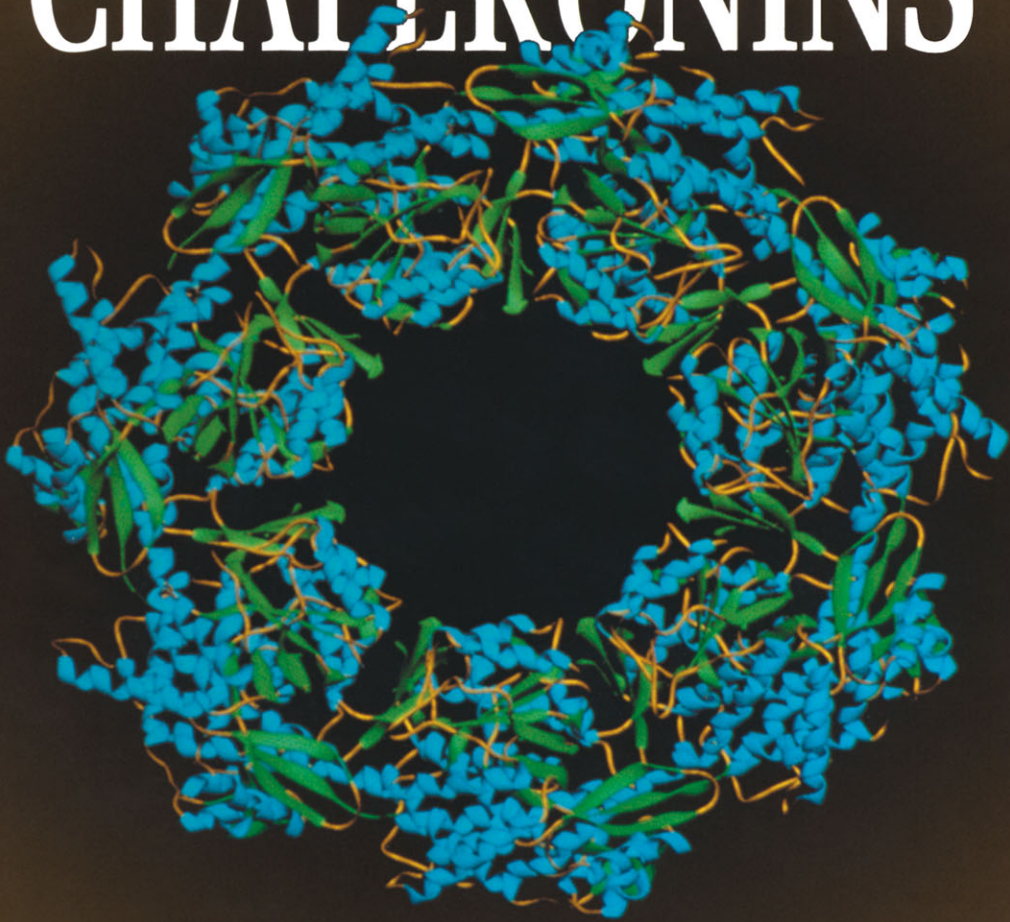
THE CHAPERONINS



Edited by

R. JOHN ELLIS

CELL BIOLOGY *A Series of Monographs*

The Chaperonins

CELL BIOLOGY: A Series of Monographs

EDITORS

D. E. BUETOW
Department of Physiology and Biophysics
University of Illinois
Urbana, Illinois

I. L. CAMERON
Department of Cellular and Structural Biology
The University of Texas
Health Science Center at San Antonio
San Antonio, Texas

G. M. PADILLA
Department of Cell Biology
Duke University Medical Center
Durham, North Carolina

A. M. ZIMMERMAN
Department of Zoology
University of Toronto
Toronto, Ontario, Canada

Volumes published since 1986

Reginald M. Gorczynski (editor). RECEPTORS IN CELLULAR RECOGNITION AND DEVELOPMENTAL PROCESSES, 1986

Govindjee, Jan Amesz, and David Charles Fork (editors). LIGHT EMISSION BY PLANTS AND BACTERIA, 1986

Peter B. Moens (editor). MEIOSIS, 1986

Robert A. Schlegel, Margaret S. Halleck, and Potu N. Rao (editors). MOLECULAR REGULATION OF NUCLEAR EVENTS IN MITOSIS AND MEIOSIS, 1987

Monique C. Braude and Arthur M. Zimmerman (editors). GENETIC AND PERINATAL EFFECTS OF ABUSED SUBSTANCES, 1987

E. J. Rauckman and George M. Padilla (editors). THE ISOLATED HEPATOCYTE: USE IN TOXICOLOGY AND XENOBIOTIC BIOTRANSFORMATIONS, 1987

Heide Schatten and Gerald Schatten (editors). THE MOLECULAR BIOLOGY OF FERTILIZATION, 1989

Heide Schatten and Gerald Schatten (editors). THE CELL BIOLOGY OF FERTILIZATION, 1989

Anwar Nasim, Paul Young, and Byron F. Johnson (editors). MOLECULAR BIOLOGY OF THE FISSION YEAST, 1989

Mary P. Moyer and George Poste (editors). COLON CANCER CELLS, 1990

Gary S. Stein and Jane B. Lian (editors). MOLECULAR AND CELLULAR APPROACHES TO THE CONTROL OF PROLIFERATION AND DIFFERENTIATION, 1991

Vitąuts I. Kalnins (editor). THE CENTROSOME, 1992

Carl M. Feldherr (editor). NUCLEAR TRAFFICKING, 1992

Christer Sundqvist and Margareta Ryberg (editors). PIGMENT–PROTEIN COMPLEXES IN PLASTIDS: SYNTHESIS AND ASSEMBLY, 1993

David H. Rohrbach and Rupert Timpl (editors). MOLECULAR AND CELLULAR ASPECTS OF BASEMENT MEMBRANES, 1993

Danton O'Day (editor). SIGNAL TRANSDUCTION DURING BIOMEMBRANE FUSION, 1993

John K. Stevens, Linda R. Mills, and Judy E. Trogadis (editors). THREE-DIMENSIONAL CONFOCAL MICROSCOPY: VOLUME INVESTIGATION OF BIOLOGICAL SPECIMENS, 1994

Lawrence I. Gilbert, Jamshed R. Tata, and Burr G. Atkinson (editors). METAMORPHOSIS: POSTEMBRYONIC REPROGRAMMING OF GENE EXPRESSION IN INSECT AND AMPHIBIAN CELLS, 1996

R. John Ellis (editor). THE CHAPERONINS, 1996

The Chaperonins

Edited by

R. John Ellis
Department of Biological Sciences
University of Warwick
Coventry, United Kingdom

ACADEMIC PRESS
San Diego New York Boston London Sydney Tokyo Toronto

Front cover photograph: Ribbon diagram of the end view of chaperonin 60 (GroEL) from *Escherichia coli*. Reprinted with permission from the front cover of *Nature* (K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, and P. B. Sigler: The crystal structure of the bacterial chaperonin GroEL at 2.8A, *Nature*, 371, 578–586, 1994). Copyright (1994) Macmillan Magazines Limited.

This book is printed on acid-free paper. ♾

Academic Press, Inc.
A Division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495

United Kingdom Edition published by
Academic Press Limited
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

The chaperonins / edited by R John Ellis.
p. cm. -- (Cell biology series)
Includes index.
ISBN 0-12-237455-X (case: alk. paper)
1. Molecular chaperones. I. Ellis R. J. (Reginald John), date.
. II. Series.
QP552.M64C48 1996
574.19'245--dc20 95-44810
CIP

PRINTED IN THE UNITED STATES OF AMERICA
96 97 98 99 00 01 EB 9 8 7 6 5 4 3 2 1

To Christian B. Anfinsen

This Page Intentionally Left Blank

Contents

4 Chaperonin-Mediated Folding and Assembly of Proteins in Mitochondria

Thomas Langer and Walter Neupert

5 Structure and Function of Chaperonins in Archaebacteria and Eukaryotic Cytosol

Keith R. Willison and Arthur L. Horwich

6 Regulation of Chaperonin Gene Expression

Saskia M. van der Vies and Costa Georgopoulos

7 Kinetic and Energetic Aspects of Chaperonin Function

Anthony R. Clarke and Peter A. Lund

8 Role of Prokaryotic Chaperonins in Protein Folding

Jörg Martin, Mark Mayhew, and F.-Ulrich Hartl

9 Chaperonin Structure and Conformational Changes

Helen R. Saibil

10 Immunological Aspects of Chaperonins

Anthony R. M. Coates

This Page Intentionally Left Blank

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

Anthony R. Clarke (167), Department of Biochemistry, University of Bristol, Bristol B58 1TD, United Kingdom

Anthony R. M. Coates (267), Division of Molecular Microbiology, Department of Medical Microbiology, St. George's Hospital Medical School, London SW17 ORE, United Kingdom

R. John Ellis (1), Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

Anthony A. Gatenby (65), BioProcess Development Center, Central Research and Development, Dupont, Experimental Station, Wilmington, Delaware 19880

Costa Georgopoulos (137), Département de Biochimie Médicale, Centre Médical Universitaire, Université de Genève, 1211 Genève, Switzerland

Radhey S. Gupta (27), Department of Biochemistry, McMaster University, Hamilton, Canada L8N 3Z5

F.-Ulrich Hartl (213), Howard Hughes Medical Institute and Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Arthur L. Horwich (107), Howard Hughes Medical Institute, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, Connecticut 06536

Thomas Langer (91), Institut für Physiologische Chemie der Ludwig-Maximilians-Universität, D-80336 München, Germany

Peter A. Lund (167), School of Biological Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

Jörg Martin (213), Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, Rhode Island 02912

Mark Mayhew (213), Howard Hughes Medical Institute and Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Walter Neupert (91), Institut für Physiologische Chemie der Ludwig-Maximilians-Universität, D-80336 München, Germany

Helen R. Saibil (245), Department of Crystallography, Birkbeck College London, London WC1E 7HX, United Kingdom

Saskia M. van der Vies (137), Département de Biochimie Médicale, Centre Médical Universitaire, Université de Genève, 1211 Genève, Switzerland

Keith R. Willison (107), CRC Centre for Cell and Molecular Biology, Institute of Cancer Research, Chester Beatty Laboratories, London SW3 6JB, United Kingdom

Preface

"Most impediments to scientific understanding are conceptual locks, not factual lacks." The history of the discovery of the chaperonins and their role in protein folding provides a good illustration of this aphorism by Stephen J. Gould (1991). The pioneering experiments of Christian B. Anfinsen in the 1950s and 1960s revealed the remarkable finding that some pure denatured proteins will refold spontaneously into their biologically active three-dimensional structures when the denaturant is removed. They will do this in the absence of either an energy source or any other macromolecules. Thus was born an important corollary to the Central Dogma of molecular biology: the principle of self-assembly, according to which all the information required for each aminoacyl sequence to fold into its functional conformation resides within the sequence itself. It is hard to exaggerate the importance of this principle for our current understanding of the molecular basis of the living state. Just imagine the difficulties if each newly synthesized polypeptide chain required steric information from preexisting cellular components before it would fold correctly!

The contribution of Anfinsen was recognized by the joint award, with S. Moore and W. H. Stein, of the Nobel Prize for chemistry in 1972. A glance at modern textbooks of biochemistry reveals the importance placed on self-assembly as an organizing principle in biology. So why is the remark of Gould relevant to what seems a satisfactory and pleasing concept?

The answer lies in the extraordinary complexity of the environment in which proteins fold inside cells and in the differing attitudes of the many researchers that have followed Anfinsen in the study of protein refolding *in vitro*. Some have pursued his basic type of experiment by applying ever more sophisticated techniques to the refolding of pure

proteins in the test tube. It was observed that success in such experiments for some proteins requires the use of low protein concentrations and low temperatures to avoid the production of aggregated or misfolded structures, but this observation did not stimulate enquiry into how these problems are countered in the cell. This work continues to be driven by the intellectual challenge of discovering the precise rules that enable a given aminoacyl sequence to fold into a particular conformation, a challenge still to be met. The "protein folding problem" is regarded by some as the most challenging intellectual problem in biology (those wanting to know the molecular basis of consciousness would disagree). The difficulty of this task, combined with the technical simplicity of the conditions under which protein refolding is studied, has discouraged many protein chemists from considering whether other processes are involved in protein folding in the living cell. In my experience protein chemists do not react well to the suggestion that they make the refolding conditions much more complex in trying to mimic intracellular conditions, and who can blame them? Why complicate the situation when the protein will regain its biological activity on its own, albeit under unphysiological conditions? But in the end we have to ask why protein chemists are doing this type of work. Is it to understand how proteins fold inside test tubes or inside living cells, and are these the same problems?

It is interesting in hindsight to note that a paper published from Anfinsen's laboratory (Epstein *et al.,* 1963) contained the prophetic remark that "Another large molecule (e.g., an antibody, other protein, or possibly even the same protein) could influence the folding process by intermolecular reactions," but this suggestion did not spark much enthusiasm among protein chemists to search for such molecules. Cell biologists and biochemists, on the other hand, are happy to work with undefined cellular soups, but then they do not aim to achieve the precision sought by protein chemists. It is from a serendipitous observation made during the study of protein synthesis in cell-free extracts and isolated organelles that a more complex story about protein folding inside cells started to emerge. The folding of some proteins did not appear to be spontaneous, but to involve interactions with preexisting proteins that act as molecular chaperones and hydrolyze ATP. Chapter 1 (R. John Ellis) discusses the origins of this new view of protein folding. The chaperonins are one major group of ubiquitous molecular chaperones that interact with folding proteins, and the aim of this book is to provide a snapshot of recent research in this exciting new field.

Chaperonins occur in all types of cell and fall into two related subfamilies called the GroE chaperonins and the TCP-1 chaperonins. Chapter

2 by Radhey S. Gupta discusses possible evolutionary relationships of members of these subfamilies. Two chapters follow on the properties and functions of the GroE chaperonins found inside chloroplasts and photosynthetic bacteria (Anthony A. Gatenby) and mitochondria (Thomas Langer and Walter Neupert), with which much of the early work on the chaperonins was carried out. The TCP-1 chaperonins found in archaebacteria and eukaryotic cytosol are discussed in Chapter 5 by Keith R. Willison and Arthur L. Horwich. The most detailed knowledge of the regulation of GroE chaperonin gene expression comes from studies on eubacteria, especially *Escherichia coli.* It was the analysis of mutants of this bacterium which were unable to replicate some bacteriophages that led to the identification of the first bacterial chaperonin called GroEL in the 1970s, but the "conceptual lock" of self-assembly discouraged attempts to determine the role of GroEL in the uninfected cell. These genetic aspects are considered in Chapter 6 by Saskia M. van der Vies and Costa Georgopoulos.

Much of the current excitement in the chaperonin field is being generated by the analysis of "neo-Anfinsen" experiments in which the refolding of denatured pure proteins is studied in the presence of chaperonins added to the refolding buffer. Biophysical aspects of these experiments are discussed in Chapter 7 by Anthony R. Clarke and Peter A. Lund and the biochemical aspects in Chapter 8 by Jörg Martin, Mark Mayhew, and F.-Ulrich Hartl. Complementing these approaches are structural studies of the chaperonins by both electron microscopy and X-ray diffraction (Chapter 9 by Helen R. Saibil). The last chapter by Anthony R. M. Coates discusses a less well-publicized aspect of the GroE chaperonins: the fact that they are the dominant immunogens in human bacterial infections and stimulate cytokine production when added to some animal cells in culture. The chaperonins may have additional functions, such as the protection of messenger RNA against degradation and obscure surface roles in some bacterial and animal cells. The chaperonin story is just beginning.

It is important to realize that the new wave of research discussed in this book casts no doubt on the validity of the principle of protein self-assembly. The chaperonins do not provide steric information for protein folding, rather they permit self-assembly to proceed with a higher probability of success within the highly concentrated and complex intracellular milieu. The mechanisms by which they achieve this improvement are under debate, but a currently favored model suggests that the chaperonins essentially act as sequestering agents inside which protein chains fold

by self-assembly in isolation from one another. I have coined the term "Anfinsen cage" for this model since it implies that protein folding *per se* occurs in the same way inside cells and test tubes. If this view is correct, chaperonin research will not help solve the protein folding problem, but it will illuminate some of the ancillary events that enable self-assembly to proceed more effectively inside the cell, and may permit the trapping of productive folding intermediates. Thus the notion of spontaneous self-assembly is being replaced by the idea that, within the cell, the self-assembly of some proteins is *assisted* by preexisting proteins such as the chaperonins. It follows that the approach of the cell biologist is complementary to that of the protein chemist, since it provides additional information about events accompanying protein folding in nature.

This book is dedicated to the memory of Christian B. Anfinsen for his brilliant insight into the nature of protein folding. He sadly did not live to receive a copy of this book, but I am glad to record the enthusiasm and interest he expressed to me about the chaperonins in his last years.

References

Gould, S. J. (1991). "Bully for Brontosaurus," p. 256. Penguin Books.

Epstein, C. J., Golberger, R. F., and Anfinsen, C. B. (1963). The genetic control of tertiary protein structure: Studies with model systems. *Cold Spring Harbor Symp. Quant. Biol.* **28,** 439–446.

R. John Ellis

1

Chaperonins: Introductory Perspective

R. JOHN ELLIS

Department of Biological Sciences
University of Warwick
Coventry CV4 7AL, United Kingdom

What are the chaperonins, how were they discovered, and what are the functions that their structures have evolved to perform inside all living cells? This volume presents the views of some of the active researchers in this field about possible answers to these questions. The purpose of this introductory chapter is to provide a general historical and conceptual framework within which these views can be considered. The chaperonins are at the center of a new and rapidly expanding field of biochemical research into the biological roles of molecular chaperones. The results of this effort are changing our views about how proteins

achieve, maintain, and lose their functional conformations in all types of cell. The contents of this volume provide a snapshot of research into arguably the single most important process in biology: protein folding.

I. ORIGINS

It is useful to consider the origins of a new subject for two reasons. First, it can be instructive; the history of science provides sobering take-home messages about the importance of not ignoring observations that do not fit the prevailing conceptual paradigm, and about the value of thinking laterally, in case apparently unrelated phenomena conceal common principles. Second, once a new idea has become accepted there is often a tendency to believe that it was obvious all along—hindsight is a wonderful thing, but the problem is that it is never around when you need it! The history of the discovery of the chaperonins contains examples of both these points.

A. Origin of Chaperonin and Molecular Chaperone

The term chaperonin was coined by Sean Hemmingsen to describe one family of highly sequence-related proteins acting as molecular chaperones found in chloroplasts, mitochondria, and eubacteria such as *Escherichia coli* (Hemmingsen *et al.*, 1988). Newcomers to this field may well wonder why there is such emphasis on the term molecular chaperone and its derivatives, and suspect that, at best, the term is used as a mere metaphor or, at worst, is a dreadful example of academic whimsy. In my view it is neither of these, but is a precise description of a ubiquitous and essential function operating within all types of cell. What is this function?

The term molecular chaperone was used first by Laskey *et al.* (1978) to describe the properties of nucleoplasmin, an acidic nuclear protein required for the correct assembly of nucleosome cores from DNA and histones in extracts of *Xenopus* eggs. Why is the word chaperone appropriate to describe the properties of nucleoplasmin? These properties are unusual because they do not fall into any of the obvious categories. Thus the role of nucleoplasmin is not to provide steric information essential for

nucleosome assembly, nor is it a structural component of the assembled nucleosome. Rather nucleoplasmin binds transiently to the histones, thereby reducing their positive charge density, and so inhibits the tendency for unspecific aggregates to form between the positively charged histones and the negatively charged DNA—in other words, nucleoplasmin acts as an electrostatic filter. This transient inhibitory role of nucleoplasmin allows the self-assembly properties of the histones with DNA to predominate over the incorrect interactions generated by their high density of opposite charges. The molecular details of how this is achieved are still obscure, but it is clear that the role of nucleoplasmin in nucleosome assembly is transitory, and does not involve the formation or breakage of covalent bonds. It can thus be detected only if nondenaturing techniques are used to study the early stages of nucleosome formation. Later work revealed an additional role of nucleoplasmin in decondensing sperm chromatin on fertilization of the egg, resulting in the replacement of the protamine proteins of the sperm nucleosomes by the histone proteins of the zygote (Laskey *et al.*, 1993).

It should be clear from these properties that nucleoplasmin is a precise molecular analog of the human chaperone. The traditional role of the latter is to prevent incorrect interactions between pairs of human beings, without either providing the steric information necessary for their correct interaction or being present during their subsequent married life—but often reappearing at divorce and remarriage! So the term is a precise description of an essential function that we now know all cells require in order to increase the probability of correct macromolecular interactions (see Section II).

B. Discovery of Chaperonins

The term molecular chaperone was used by Laskey *et al.* (1978) to describe the properties of one protein, nucleoplasmin, found in the nucleus of amphibian eggs. It was later extended to describe the properties of a different protein implicated in the assembly of the photosynthetic carbon dioxide-fixing enzyme ribulose-bisphosphate carboxylase (EC4.1.1.39, ribulosebisphosphate carboxylase/oxygenase, Rubisco) in the chloroplasts of higher plants (Musgrove and Ellis, 1986). Chloroplast Rubisco is an oligomeric protein consisting of large subunits synthesized inside the chloroplast and small subunits imported into the chloroplast

after synthesis in the cytosol. The key observation was that an abundant chloroplast protein binds transiently and noncovalently to radioactively labeled large subunits of Rubisco synthesized inside intact chloroplasts isolated from the leaves of *Pisum sativum;* time course experiments showed that these large subunits are subsequently transferred to the holenzyme (Barraclough and Ellis, 1980). Because isolated large subunits of chloroplast Rubisco have a strong tendency to form insoluble aggregates in the absence of denaturing agents, it was proposed that this binding protein might be an obligatory step in the assembly of chloroplast Rubisco that served to keep the large subunits soluble inside the chloroplast prior to assembly (Barraclough and Ellis, 1980). Although nucleoplasmin acts to reduce problems produced by the high concentrations of charged residues in the nucleus, and the chloroplast binding protein acts to reduce problems caused by the high concentration of hydrophobic residues in the chloroplast, they nevertheless share a commonality of action in that they reduce incorrect interactions of one kind or another. Figure 1 shows a repeat of the key experiment that identified the chloroplast protein as a polypeptide binding protein.

It is chastening to record that this binding phenomenon was first observed in the author's laboratory in 1973, and noted repeatedly thereafter, but was interpreted incorrectly until 1980 because the low stoichiometry of binding did not allow the complex of binding protein with radioactive Rubisco large subunits to be resolved from the bulk of the nonradioactive binding protein. Because the radioactive Rubisco large subunits migrate exactly with the chloroplast binding protein during electrophoresis on nondenaturing polyacrylamide gels (see Fig. 1), and during centrifugation on sucrose density gradients, the conclusion seemed obvious between 1973 and 1980 that what we now know is a polypeptide binding protein is simply an oligomeric form of Rubisco large subunits that the chloroplast accumulates as part of its Rubisco assembly pathway. The message from this story is that an interpretation that seems obvious and rational can nevertheless turn out to be incorrect. A detailed history of the discovery of the chloroplast binding protein is available (Ellis, 1990a).

The proposal that the chloroplast binding protein could be thought of as a molecular chaperone (Musgrove and Ellis, 1986) did not meet with much enthusiasm at the time because of the general acceptance of the principle of protein self-assembly, which proposes that both the folding and the association of newly synthesized polypeptides are spontaneous processes requiring no macromolecules other than the

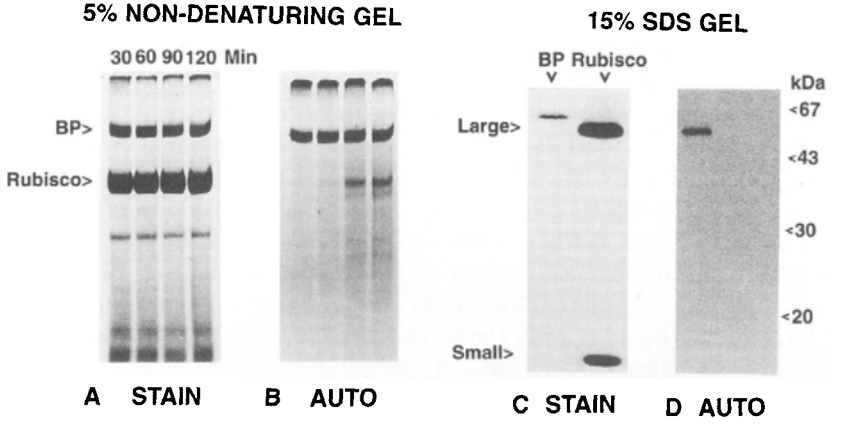


Fig. 1. Discovery of the chloroplast binding protein (BP). Intact chloroplasts were isolated from young seedlings of *Pisum sativum* and illuminated at 20°C in a medium containing sorbitol as osmoticum and [^{35}S]methionine as labeled precursor; the illumination serves as an energy source for protein synthesis by ribosomes inside the intact chloroplasts. Samples were removed at intervals, the chloroplasts were centrifuged down and lysed in hypotonic buffer, and the soluble fraction was electrophoresed on a 5% nondenaturing polyacrylamide gel (A and B). The gel was stained in Coomassie blue (A) and an autoradiograph made (B). The stained bands of Rubisco holoenzyme and BP were excised from the 30-min track and analyzed separately on a 15% SDS–polyacrylamide gel (C and D). The SDS gel was stained (C) and an autoradiograph made (D). Note that labeled Rubisco large subunits comigrate exactly with the staining band of the binding protein (compare the precise shapes of the bands in A and B). These large subunits can be visualized by their radioactivity but not by staining, since the chemical amount made in this system is very small (compare C and D). The binding protein oligomer (BP) is visible as a prominent stained band, as are its 60-kDa subunits, but these are not radioactive since they are made in the cytosol (compare C and D). BP, Rubisco subunit binding protein; Rubisco, holoenzyme of ribulose bisphosphate carboxylase-oxygenase; large and small, large and small subunits of Rubisco, respectively. Reprinted from Johnson (1987) with permission.

components of the assembled protein. This conclusion is an important corollary of the central dogma of molecular biology which states that the information transferred from nucleic acids to proteins is linear in nature. Protein self-assembly implies that no informational factors other than the linear sequence of aminoacyl residues are necessary for that sequence to fold into a specific conformation. There was thus no compelling reason to search for factors that might affect this folding process.

The self-assembly principle is based on two types of observation. Some pure proteins denatured *in vitro* will often refold correctly in a spontaneous manner when the denaturing agent is removed, in the total absence of either any other macromolecules or the expenditure of energy (Anfinsen, 1973). Some oligomeric structures such as tobacco mosaic virus will reassemble correctly when the separated components of the assembled structure are mixed together (Caspar and Klug, 1962). It is interesting that some pioneers of the self-assembly hypothesis speculated that "another large molecule (e.g., an antibody, other protein, or possibly even the same protein) could influence the folding process by intermolecular interactions" (Epstein *et al.,* 1963). This prophetic remark, however, did not turn out to be seminal, since its potential implications were largely ignored by protein chemists until the chaperonins were discovered. Thus it is only the most recent editions of biochemistry text books that refer to the role of molecular chaperones in assisting protein folding (e.g., Alberts *et al.,* 1994).

It was initially thought that nucleoplasmin and the chloroplast binding protein were special cases evolved to deal with certain oligomeric proteins whose assembly presents particular difficulties. Speculations by Pelham (1986) about a more general role of yet other proteins (heat shock 70 and 90 proteins) in assembly and disassembly processes prompted the broader suggestion that proteins acting as molecular chaperones may be widespread, and are commonly required for both protein folding and protein oligomerization to proceed correctly (Ellis, 1987). Thus the general concept of molecular chaperones was borne from the realization that several unrelated discoveries in biochemical research could be regarded as examples of a more fundamental phenomenon.

Support for this idea soon appeared. Sequence determinations of cDNAs for the subunits of the chloroplast binding protein revealed unexpected aminoacyl identities of about 50% with the GroEL protein of *E. coli,* and with the common antigen found in many human bacterial infections (Hemmingsen *et al.,* 1988). The GroEL protein was identified genetically in four laboratories in 1972–1973 as a bacterial protein required for the replication of bacteriophages such as lambda (λ)—"Gro" refers to phage growth and the suffix "E" refers to the observation that the phage lambda growth defect is overcome when the phage carries a mutation in the head gene E (Sternberg, 1973; see Chapter 6). This GroEL protein is required for phage-encoded protein subunits to assemble correctly into the oligo-

meric connector that joins the head to the tail of phage lambda, but it is not a component of the final structure.

In the 1970s most attention was paid to the role of this bacterial protein in phage assembly rather than to its role in the uninfected cell, and this remained the case until the high sequence identity of GroEL with the chloroplast binding protein was discovered in 1988. Independent work at about the same time revealed an immunological similarity between GroEL and a heat shock protein of mitochondria from plant and animal sources (McMullin and Hallberg, 1988); the role of this protein in assisting the folding of newly imported proteins by yeast and *Neurospora* mitochondria was soon defined (Cheng *et al.*, 1989; Ostermann *et al.*, 1989; see Chapter 4). The work reported by Ostermann *et al.* (1989) was of particular significance for two reasons. First, it concerned the folding of a monomeric protein dihydrofolate reductase (DHFR), whereas other studies used oligomeric proteins where it was not definitively established whether the chaperonin (cpn)60 was assisting folding and/or association. Second this protein refolds spontaneously *in vitro* with high efficiency, but nevertheless was found to interact with the mitochondrial cpn60 *in vivo*. This finding was the first to indicate that the fact that a protein refolds spontaneously *in vitro* with no difficulty does not necessarily mean that it folds spontaneously *in vivo*. In similar vein, two proteins newly synthesized in cell-free extracts of *E. coli* were found to bind transiently to GroEL, and the transport competence of one of them was prolonged by this binding (Bochkareva *et al.*, 1988).

It was not long before the pioneer experiments of Anfinsen on the refolding of denatured proteins were being repeated with the modification that GroEL was added to the refolding buffer (Goloubinoff *et al.*, 1989). This modification improves the yield of correctly folded bacterial Rubisco in an ATP-dependent process and reduces its tendency to aggregate when diluted from the denaturant. Goloubinoff *et al.* (1989) reported the first example of a continuing wave of similar "neo-Anfinsen" experiments being carried out on a variety of other proteins in several laboratories. The precise means by which GroEL assists correct protein refolding in these "neo-Anfinsen" experiments and exactly how quantitatively important the chaperonins and other molecular chaperones are at assisting protein folding *in vivo* are the subject of much current debate, some of which is recorded in Chapters 7–9.

The period 1987–1989 marks a watershed in research on protein folding. The original view that protein folding in the cell is a spontaneous energy-independent process continues to be challenged by a different

view. This view retains the principle of protein self-assembly, but emphasizes the importance of the assistance of this process by preexisting proteins acting as molecular chaperones, some of which expend energy in the form of ATP hydrolysis. The idea that self-assembly requires assistance may seem contradictory, but one of the originators of the self-assembly principle has commented that "chaperonage of self-assembly is not an oxymoron" (Caspar, 1991). The evidence for this new view was initially obtained from studies with chloroplasts, eubacteria, and mitochondria, but was extended to the archaebacteria and the eukaryotic cytosol in 1990–1991 when the relationship between the GroE chaperonins and the *t*-complex polypeptide 1 (TCP-1) chaperonins was uncovered (see Section III and Chapter 5). Thus the general concept of molecular chaperones has triggered a wave of fresh research on protein folding in systems from all types of cell that has yet to break, and whose final biological significance has yet to emerge.

II. GENERAL CONCEPT OF MOLECULAR CHAPERONES

This topic has been discussed extensively (Ellis and Hemmingsen, 1989; Rothman, 1989; Ellis, 1990a,b; Ellis and van der Vies, 1991; Gething and Sambrook, 1992; Ellis, 1993; Hendrick and Hartl, 1993), so only a summary of key points is given here.

A. Definitions of Molecular Chaperones

Molecular chaperones are currently defined as a class of unrelated families of protein that have in common the ability to assist the noncovalent assembly of other protein-containing structures *in vivo,* but which are not permanent components of these structures when they are performing their normal biological functions (Ellis and Hemmingsen, 1989; Ellis, 1990a,b, 1993). In this definition "assembly" is used in a broad sense to encompass not only the folding of newly synthesized polypeptide chains and any association into oligomers that may occur subsequently, but also any changes in the degree of either folding or association that may occur when proteins cross membranes, perform their normal functions, or are repaired or removed after damage by stresses such as heat shock.

The term noncovalent is used to exclude those proteins that carry out covalent co- or post-translational modifications; such modifications are often important for protein assembly, but the enzymes that catalyze them are not the proteins being defined here. Thus protein disulfide-isomerase is not a molecular chaperone in terms of its isomerase activity, although it may contain a molecular chaperone activity in other parts of its structure (Lamantia and Lennarz, 1993; Puig and Gilbert, 1994). An alternative form of wording that expresses essentially the same idea has been proposed: a molecular chaperone is a "protein that binds to and stabilizes an otherwise unstable conformer of another protein, and by controlled binding and release of the substrate protein facilitates its correct fate *in vivo,* be it folding, oligomeric assembly, transport to another subcellular compartment or controlled switching between active and inactive conformations" (Hendrick and Hartl, 1993).

It should be noted that both these definitions are functional rather than structural in nature, but contain no constraints on the mechanisms by which molecular chaperones function. This is the reason for the use of the imprecise words "assist" and "facilitate." Thus molecular chaperones are not defined by a common mechanism of action nor by sequence similarity. One convenient way of classifying the different types of molecular chaperone is to group them into families on the basis that within each family there is high sequence similarity, but between families there is not (Ellis and van der Vies, 1991). The chaperonins are thus just one family of sequence-related molecular chaperones.

If the above definitions are accepted, only two criteria need to be fulfilled in order to call a protein a molecular chaperone: it must in some sense assist or facilitate the assembly of other protein-containing structures, the mechanism by which it does this being irrelevant, but it must not be a component of those structures when they are performing their normal biological functions.

Both the above definitions suggest that molecular chaperones function not only during normal conditions but also under stress conditions. The molecular chaperone concept can accommodate at least some aspects of the stress response by supposing that the need for these molecular chaperones increases when proteins are damaged by stress. If the cell is to survive stress, damaged proteins need to be either refolded correctly or removed by proteolysis, while undamaged proteins need to be protected against subsequent stresses that may create incorrect assemblies. Thus the stress response can be viewed in certain respects as an amplification of a basic molecular chaperone function that all cells require

under normal conditions, rather than as a unique function required only under stress conditions. It remains possible of course that some stress proteins have functions that are uniquely required in the stress situation. Indeed, if the first living cells originated in high temperature environments, it is conceivable that molecular chaperones functioned initially under stress conditions, and only later in evolution were retained to function at the lower temperatures that most cells experience today.

The view that some stress proteins act as molecular chaperones has gained increasing experimental support (Morimoto *et al.*, 1994), but it is important to note that not all stress proteins act as molecular chaperones and, conversely, that not all molecular chaperones accumulate after stress. Only 5 of the 20 or so stress proteins of *E. coli* whose expression is controlled by the sigma (σ) 32 transcription factor are known to act as molecular chaperones (Georgopoulos *et al.*, 1990), whereas nucleoplasmin and the TCP-1-containing complexes in the eukaryotic cytosol are not stress proteins (see Chapter 5).

B. Why Molecular Chaperones Exist

An interesting question to consider is why molecular chaperones exist at all. The principle of protein self-assembly is well supported by many *in vitro* experiments, so why is it apparently insufficient to describe the *in vivo* situation? A possible answer stems from the observation that many cellular processes involve the transient exposure of interactive surfaces to the intracellular environment, and so run the risk that some of these surfaces may interact incorrectly to generate nonfunctional structures. Interactive surfaces are defined in this context as any regions of intra- or intermolecular contact that are important in maintaining the structure in its biologically active conformation. The interactive surfaces that are chaperoned in the nucleus by nucleoplasmin are charged, but in the case of all other molecular chaperones so far studied, the interactions are hydrophobic.

It has long been known that the success of an *in vitro* protein refolding experiment of the type pioneered by Anfinsen (1973) increases as the temperature and protein concentration decrease, because these conditions reduce the probability of incorrect interactions between hydropho-

bic surfaces transiently exposed during the refolding process (see Chapters 7 and 8). Such refolding experiments are often performed at protein concentrations in the 1 μg–1 mg/ml range, but protein concentrations *in vivo* are much higher, in the range 200–300 mg/ml, so the concentration of exposed interactive surfaces is also higher *in vivo*. Most ribosomes are active in protein synthesis in growing *E. coli* cells, so the concentration of nascent chains will be in the same range as the concentration of ribosomes, i.e., 30–50 μM, or 1–2 mg/ml for a protein of 40 kDa. This concentration range of nascent chains does not appear to greatly exceed that used *in vitro*, but the effective concentration of nascent chains *in vivo* is greatly increased in terms of thermodynamic activity because of the effect of macromolecular crowding, an effect that is absent *in vitro* (see Section V,B).

Whereas the self-assembly principle implies that all the interactions that occur *in vivo* are both necessary and sufficient to produce the correct structures, the molecular chaperone concept suggests that on the contrary there is an intrinsic risk of incorrect interactions producing abnormal structures because of the high concentration of interacting surfaces. It is also possible that under the conditions existing *in vivo* some proteins will become kinetically trapped in nonnative conformations and be unable to proceed on their own to the native conformation at a sufficient rate (see Chapter 7). The proportion of cellular proteins that requires assistance by molecular chaperones in order to assemble correctly in amounts sufficient for cellular needs is one of the many unresolved questions in this field (see Section V).

We can summarize all these considerations into a simple unifying principle: all cells require a molecular chaperone function that serves to prevent or to reverse incorrect interactions between transiently exposed surfaces by the binding of molecular chaperone proteins to these surfaces. Whether molecular chaperones exist that have a more active role, e.g., by altering the range of pathways of folding open to certain polypeptides or even by providing steric information, remains to be seen. The discovery that diluting from denaturant a mixture of the scrapie prion protein with its normal conformational isomer causes the conversion of some of the latter to the former may be an example of a process whereby one protein alters the folding pathway of a related protein (Bessen *et al.*, 1995). What is clear is that the list of proteins that can be regarded as molecular chaperones is growing steadily. This volume is concerned with the best-studied family: the chaperonins.

III. DEFINITIONS AND NOMENCLATURE

One reason the term chaperonin was proposed was to try to simplify the existing complex nomenclature for different members of this family of proteins whose close similarity had just been realized (Hemmingsen *et al.*, 1988). Present knowledge suggests that one way to define the chaperonins is as a family of sequence-related molecular chaperones that function to assist protein folding in all types of cell by binding to nonnative forms of proteins. The family contains two distinct subclasses (Ellis, 1992, and Table I): (1) the GroE subclass found in chloroplasts and other plastids, mitochondria, and eubacteria and (2) the TCP-1 subclass found in archaebacteria and the eukaryotic cytosol. These subclasses have also been called Group I and II chaperonins, respectively (see Chapter 5).

The sequence identity between members of the GroE subclass is around 50%, as is the sequence identity between members of the TCP-1 subclass, but the identity between the two subclasses is much less. This grouping is consistent with the view that chloroplasts and mitochondria originated as endosymbionts from free-living eubacteria, whereas the

TABLE I

Nomenclature of the Chaperonins

Generic name	Useful abbreviations	Other names
GroE subclass		
Eubacterial chaperonin 60	Eu cpn60	GroEL (*E. coli*), 65-kDa antigen
Eubacterial chaperonin 10	Eu cpn10	GroES (*E. coli*), cochaperonin
Mitochondrial chaperonin 60	Mt cpn60	hsp60, mitonin, HuCha60
Mitochondrial chaperonin 10	Mt cpn10	hsp10, cochaperonin
Chloroplast chaperonin 60	Ch cpn60	Rubisco subunit binding protein
Chloroplast chaperonin 10	Ch cpn10 Ch cpn21	Cochaperonin, chaperonin 21
TCP-1 subclass		
Cytosolic chaperonin 60	Cyt cpn60 TCP-1 CCT TRiC	*t*-complex polypeptide 1, chaperonin-containing TCP-1, TCP-1 ring complex
Archaebacterial chaperonin 60	Ar cpn60 TF55	Thermophilic factor 55, thermosome

archaebacteria contributed toward the eukaryotic cytosol. Possible evolutionary relationships within and between these subclasses are discussed in Chapters 2 and 5.

The above definition also describes the heat shock protein 70 family of molecular chaperones, although their role in protein folding is different from that of the chaperonins (see Chapter 8), so it is necessary to define the structure of the chaperonins. The chaperonins all have in common a large oligomeric structure consisting of two stacked rings of subunits, each about 60kDa in size, surrounding a central cavity within which the protein substrate binds; these subunits catalyze the hydrolysis of ATP to ADP. The GroE chaperonins have seven subunits per ring, whereas the TCP-1 chaperonins have eight or nine.

A convenient abbreviation for chaperonin(s) is cpn(s), a terminology introduced by Lubben *et al.* (1989) and Goloubinoff *et al.* (1989) and used throughout this volume. Thus the subunits of the rings can be referred to as cpn60 subunits and the oligomer itself, cpn60. The GroE subclass also contains another type of oligomeric ring made of seven smaller 10-kDa subunits, cpn10. The cpn60 oligomer from *E. coli* is the best studied chaperonin of this subclass and is often called GroEL, whereas the cpn10 oligomer from *E. coli* is often called GroES. In this volume the terms GroEL and GroES refer to the proteins from *E. coli* only. Note that it is tautologous to use the expression GroEL chaperonin 60 protein, since GroEL is the chaperonin 60 of *E. coli.* GroEL and GroES oligomers bind to each other in the presence of adenine nucleotides, and this binary complex is believed to play an important part in chaperonin function; thus the genes encoding both GroEL and GroES are essential for the survival of *E. coli* (see Chapters 7–9).

The numbers 60 and 10 are rounded off approximations of the actual molecular sizes; Chapter 2 should be consulted for the sequences. The chloroplast cpn10 is unusual in that the protein subunit consists of two copies of the cpn 10 sequence joined head to tail, and thus is often referred to as cpn21 (see Chapter 3). Cpn10 oligomers are sometimes called cochaperonins, but this term is not used in this volume since it is confusing to use similar but nonidentical names for the same protein; moreover, cpn10 sequences show some similarity to part of the cpn60 sequence and thus are true chaperonins by definition (see Chapter 2). The term cochaperonin should be reserved for proteins that modulate chaperonin activity, but that are unrelated in sequence to the chaperonins (for possible examples see Chapter 5). The TCP-1 oligomers found in the eukaryotic cytosol are much more variable than the GroE oligomers

in that the latter contain only one or two subunit sequences while the former contain up to nine; for this reason TCP-1 oligomers are also called CCT complexes, for chaperonin-containing TCP-1 complexes, or TRiC, for TCP-1 ring complexes (see Chapter 5). The TCP-1 chaperonins appear not to include cpn10 oligomers.

The cpn60 and cpn10 proteins of *E. coli* are encoded by a single operon, but in other eubacteria and some archaebacteria more than one type of cpn60 gene occurs, only some of which are located in operons with cpn10 (see Chapters 2 and 6). A nomenclature has been proposed to cover the latter situation. When either a second cpn60 or a second cpn10 gene is found within a given species, the genes that are either adjacent or operon-located should be called *cpn60-1* and *cpn10-1,* and subsequent chaperonin genes should be called *cpn60-2* or *-3* and *cpn10-2* or *-3* and so on (Coates *et al.,* 1993). Whether the different types of cpn60 sequence occur in the same protein oligomer in these bacteria is unknown, but in the case of the chloroplast cpn60, it is likely that both sequences (called α and β) are present in the same oligomer (see Chapter 3). The multiple TCP-1 subunits of eukaryotes appear not to be encoded in operons, but it is probable that up to nine types of subunit occur in each individual oligomer (see Chapter 5).

Some authors use the term chaperonin interchangeably with heat shock protein 60 (hsp60). However, the terms are not synonymous because it is only cpn60 in bacteria and mitochondria and TCP-1-related proteins in archaebacteria that are strongly heat-inducible—the chloroplast cpn60 and the TCP-1 complexes in eukaryotes are not heat shock proteins. Moreover, the heat shock terminology, although useful in the earlier studies of these proteins, places undue emphasis on just one specialized feature of these proteins whose more fundamental role is to act as molecular chaperones under nonstress conditions. In this volume the term hsp60 is restricted to the cpn60 oligomer found inside mitochondria.

The word molecular is used in the term molecular chaperone because in normal usage the word chaperone refers to a person. This may seem a point too obvious to make, but the literature contains several examples in reputable journals of use of the nonsense term molecular chaperonin! This abuse arises because of the confusion of the term molecular chaperone with the term chaperonin. It should be clear from the above discussion that the chaperonins are just one family from a large and steadily growing list of unrelated families of proteins that act as molecular chaperones.

IV. FUNCTIONS OF CHAPERONINS

A. Validity of *in Vitro* Experiments

Most of this volume discusses recent observations and interpretations of the structure and functions of the chaperonins, but a summary of key points may be useful here to orient the reader new to this field. A distinction should be drawn between those experiments performed *in vitro* with either defined components or isolated organelles and crude cell-free extracts, and those performed *in vivo,* usually by genetic means. The former experiments greatly outnumber the latter and have the advantage that they allow sophisticated techniques to be used to analyze, for example, the kinetics of protein refolding, the order of complex interactions, and the structures of partially folded forms of polypeptides. However, *in vitro* experiments suffer from the disadvantage that the conditions under which they are performed are different in important respects from those operating within the intact cell (see Section V,B). Since the chaperonins evolved to assist protein folding inside the living cell, and not in the test tube, this is not a trivial distinction.

It is my view that biochemistry is a branch of biology and not a branch of chemistry, since its declared aim is to understand how organisms work. On this basis I suggest that the ultimate aim of research into protein folding should be to understand how this process operates *in vivo.* This view appears not to be shared by all workers in this field, but I suspect that had protein chemists in the past considered more closely the intracellular environment in which proteins have evolved to fold, the chaperonins could have been discovered much sooner. One encouraging trend that has become evident recently, and one that I predict will increase in importance, is that more attention is being paid to trying to mimic *in vitro* some of the conditions that may influence how the chaperonins function *in vivo,* conditions such as the presence of other molecular chaperones acting on folding polypeptides (see Chapter 8) and the degree of macromolecular crowding (see Section V).

B. Three Main Functions of Chaperonins

The majority of published evidence about chaperonin function concerns the GroEL and GroES chaperonins of *E. coli* and the chaperonins

of yeast mitochondria; much less information is available about the functions of the TCP-1 chaperonins (see Chapter 5 for the latter). This evidence supports the provisional conclusion that these proteins serve at least three functions: (1) They prevent the aggregation of partially folded polypeptides that have either just been released by the ribosome or just emerged from a membrane after transport from the cytosol (see Chapters 4, 7–9). (2) They bind partially folded polypeptides that may be kinetically trapped in a conformation unable to proceed spontaneously to the functional conformation; this binding and/or subsequent release causes the alteration of some noncovalent interactions that allow correct folding to proceed (see Chapter 7). (3) They protect correctly folded proteins from aggregation after denaturation by stresses such as high temperatures and trap denatured proteins so that they can be degraded; these activities account for the observation that in many, but not all, cell types the GroE and TCP-1 chaperonins are induced by stress (see Chapter 8).

These functions are not mutually exclusive, but complementary, and all involve the binding of the cpn60 component to exposed hydrophobic surfaces of the protein substrate. The protein binding sites of GroEL lie within the top of each ring (see Chapter 9). This location serves to encapsulate the protein substrate to some degree and prevents GroEL oligomers from binding to one another. The binding is reversed by ATP hydrolysis, and this reversal releases the substrate protein in a manner whereby it is more likely to avoid aggregation and fold correctly. The precise manner by which these ends are achieved is the subject of intense current debate (see Chapters 7–9). Two aspects of these issues that relate to chaperonin function *in vivo* are discussed in Section V.

The majority of the evidence for the above conclusions is derived from *in vitro* experiments, but experiments in which cpn60 function is disrupted *in vivo* by genetic means support the view that cpn60 prevents the aggregation of a significant number of newly imported proteins in yeast mitochondria (Cheng *et al.*, 1989; Hallberg *et al.*, 1993) and newly synthesized proteins in *E. coli* (Gragerov *et al.*, 1992; Horwich *et al.*, 1993).

The biological significance of the chaperonins does not end with their roles in protein assembly. The GroE chaperonins are some of the most potent stimulators of the immune system yet discovered, and indeed are the dominant immunogens in all human bacterial infections studied. There is thus considerable medical interest in the possible role of the chaperonins in human disease, especially in protection against infection, cancer, and autoimmune disease (see Chapter 10).

There is some indication that the chaperonins may also play extracellular roles in animals. It is claimed that the early pregnancy factor, a protein detected by bioassay in the maternal serum of mammals within 24 h of fertilization, is identical to mitochondrial cpn10 (Cavanagh and Morton, 1994). However, there has been no report of the detection of cpn10 in serum by chemical methods, perhaps because the concentration of cpn10 may be very low. One of the cpn60 proteins from mycobacteria stimulates cytokine secretion by human monocytes, whereas cpn10 from *Mycobacterium tuberculosis* induces apoptosis in human T lymphocytes. These and related observations have been used to formulate the idea that the chaperonins are multiplex antigens; i.e., they act simultaneously on a range of cell types in stress situations so that immune messages such as cytokines cause immune effector cells to adapt to cope with the stress (see Chapter 10). There are several reports that cpn60 occurs at sites outside the mitochondria in some animal cells, including at the cell surface (Kaur *et al.*, 1993). These studies use indirect immunological methods as the main means of detection, and more detailed characterization is required before the significance of these observations becomes clearer. Perhaps the peptide binding abilities of cpn60 enable it to play an extracellular role in binding peptides for transport or detoxification.

V. PROBLEMS

The aim of this section is to stimulate discussion and experimentation on two aspects of chaperonin function inside cells. These aspects are the quantitative importance of the chaperonins in assisting protein folding *in vivo* and the effects of the highly concentrated intracellular milieu on their mechanism of action.

A. Number of Proteins Using GroEL to Fold *in Vivo*

There appears to be little or no specificity for the type of aminoacyl sequence that can bind to GroEL *in vitro* and be released productively, so it is possible in principle that the folding of the majority of proteins is assisted by GroEL in *E. coli.* Moreover, the genes encoding both GroEL and GroES are essential for viability and the proteins are

required for growth at all temperatures. One genetic study employed a temperature-sensitive mutation to shut off the production of functional GroEL *in vivo* and concluded that a minimum of about 30% of the soluble proteins of *E. coli* requires GroEL to fold correctly (Horwich *et al.*, 1993). However, it is difficult to rule out that only a small number of proteins actually interact with GroEL, the folding of the remainder being disturbed by indirect pleiotropic effects. For example, if one of the proteins folded by GroEL is DnaK, the hsp70 homolog of *E. coli*, any decline in DnaK after the temperature shift could affect the folding of other proteins; this possibility arises since it is known that DnaK binds to many nascent chains and prevents their aggregation (see Chapter 8). How important this effect is will depend on the rate at which DnaK turns over in the dying cells. So the gene for GroEL could be essential because GroEL is required to fold a small number of essential proteins.

Another way to approach this problem is to calculate from available data on the cellular content of GroEL and rates of protein synthesis the fraction of proteins that could interact with GroEL under defined growth conditions. From the information provided by Neidhardt (1978) and by Bremer and Dennis (1987) for the growth of *E. coli* strain B/r on a glucose minimal medium at 37°C, and by Herendeen *et al.* (1979) for the content of GroEL, the following values emerge: (1) doubling time is 40 min; (2) total number of polypeptide chains per cell (average size 40 kDa) is 2.35×10^6; (3) cytoplasmic volume of cell is 0.6 μm^3; and (4) GroEL content per cell is 1.65% of total soluble protein or 2.75 mg/ml or 3.4 μM or 1224 oligomers per cell. The percentage varies less than 20% over doubling time changes of fivefold (Pedersen *et al.*, 1978) so that the concentration of GroEL remains approximately constant as the cells change size with doubling time.

Thus each cell under the above growth conditions makes about 10,000 polypeptide chains every 10 s and contains about 1000 molecules of GroEL. If all these proteins were folded by GroEL this would require each GroEL to fold one chain every second. However, the actual requirement will be less than this because many proteins do not fold in the cytosol. There are four intracellular compartments known where proteins fold, but that do not appear to contain any chaperonin, i.e., the endoplasmic reticulum, intermembrane mitochondrial space, and intrathylakoid chloroplast lumen of eukaryotic cells, and the periplasmic space of bacteria. The exported proteins of *E. coli* identified in the study of Horwich *et al.* (1993) do not appear to require GroEL function. The proteins of *E. coli* that constitute the periplasmic space and the outer and inner

membranes are estimated to be about one-third of total protein (Goodsell, 1991), so if we assume these fold independently of GroEL, the folding requirement reduces to two-thirds of 10,000 or 6666 chains folded per 10 s.

These values need to be compared with those for the activity of GroEL at refolding polypeptides *in vitro*. An immediate problem is that these studies are almost always conducted at around 25°C rather than at 37°C, so they are likely to be too low by a factor of two- to threefold. If we assume that the average polypeptide chain requires three rounds of ATP hydrolysis to be correctly folded by one GroEL oligomer, this takes about 60 s at 25°C (F-U. Hartl, personal communication, 1995). If we assume that it takes 20 s for one GroEL to fold one chain at 37°C, then the 1000 molecules of GroEL could fold 500/6666 or 7.5% of the total chains folding in the cytoplasm.

Despite the uncertainties in this calculation the results suggest that GroEL folds only a minority of proteins in *E. coli*—probably in the range 7.5–30%. This does not mean that any proteins fold spontaneously, however, since the cytoplasm contains types of molecular chaperone other than GroEL that bind to nascent chains, and in higher concentrations. For example, the DnaK content of *E. coli* is reported to be about 5000 molecules per cell (Neidhardt and VanBogelen, 1987). Another possibility is that GroEL functions significantly faster *in vivo* than it does in the *in vitro* conditions that have so far been tried. There is one striking aspect of the *in vivo* situation whose significance for models of chaperonin action is only just starting to be appreciated—the phenomenon of macromolecular crowding.

B. Macromolecular Crowding and Models for Chaperonin Action

The basic assumption of biochemistry as a discipline is that molecular properties determined using either cell-free extracts or pure components derived from these reflect to a significant degree properties relevant to the function of these molecules inside the living cell. Proteins are flexible and sensitive molecules, so the properties of proteins vary with the environment within which they are studied. It is relatively easy to use *in vitro* protein concentrations, temperatures, pH values, and concentrations of components such as metal ions, nucleotides, and other known effectors that are similar to those in which the proteins under study

function *in vivo*. In the case of the chaperonins, the concentrations of GroEL oligomers commonly employed in protein refolding studies *in vitro* are in the range 1–2 μM. This range is close to the reported concentration of GroEL inside the cytoplasms of *E. coli* cells growing under defined conditions (3.4 μM, see Section V,A). Whether this correspondence is due to luck or good judgment is not clear, but the fact that most measurements of GroEL activity are performed at 25°C suggests that the former is correct! However, there is one aspect of the *in vivo* situation that is rarely mimicked in biochemical studies, but that has a large effect on certain properties relevant to understanding how the chaperonins function *in vivo*—this is the macromolecular crowding (or excluded volume) effect.

Macromolecular crowding is the term used to describe the fact that the concentration of total macromolecules inside living cells is so high that a significant proportion of the cellular volume is physically occupied by molecules, and is thus unavailable to other molecules. Thus the total RNA and protein inside *E. coli* occupy a substantial fraction of the total volume (340 g/liter). In the words of Zimmerman and Minton (1993) "it is not generally appreciated that, from a quantitative point of view, biochemical rates and equilibria in a living organism bear scant resemblance to those measured in a bath of solvent." Both theory and experiment have shown that the thermodynamic activity of each macromolecular species in a crowded environment exceeds the activity of that species at an identical concentration in an uncrowded solution. For example, it has been estimated that the ratio of thermodynamic activity to concentration (i.e., the activity coefficient) for a spherical protein of 30 nm radius inside *E. coli* is in the range 100–1000 (Zimmerman and Trach, 1991). By contrast, the effect of high volume occupancy on the activity coefficient of small ions and other solutes is small. Figure 2 attempts to illustrate the difference between the uncrowded environment in which GroEL functions *in vitro* and the crowded environment in which it functions *in vivo*.

There are two main effects of macromolecular crowding. First, the more crowded the environment the larger the association constants of interacting macromolecules; macromolecular association constants for proteins in *E. coli* are predicted to exceed those in dilute solution by several orders of magnitude (Zimmerman and Minton, 1993). Second, crowding will reduce the diffusion rates of macromolecules; the diffusion coefficient decreases about 10-fold as the total protein concentration increases to 300 g/liter. Why are these considerations relevant to models for chaperonin action?

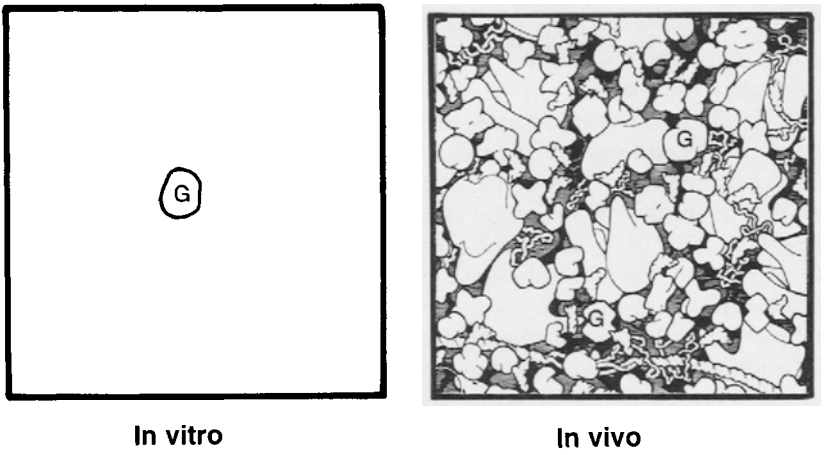


Fig. 2. Macromolecular crowding or the airport terminal effect (after Goodsell, 1991). Each square represents the face of a cube with an edge 100 nm in length. The left-hand square (*in vitro*) illustrates the 1.5 μM concentration of GroEL, similar to that used in many protein refolding experiments; there is one GroEL oligomer per cube (labeled G and approximately to scale). The right-hand square (*in vivo*) illustrates the 3 μM concentration of GroEL in the cytoplasm of *E. coli;* there are about two oligomers per cube, and the whole cytoplasm consists of about 600 such cubes. The other structures in the right-hand square represent the sizes and concentrations of the macromolecular components in the cytoplasm that create crowding by occupying so much space; thus each cube contains in addition to two GroEL molecules (G) about 30 ribosomes, 340 tRNA molecules, and 500 other protein molecules. The right-hand square is reprinted with permission from Goodsell (1991).

A study of Chapters 7–9 will reveal that there are two main classes of model proposed for the ability of GroEL/ES to increase the yield of correctly refolded protein after dilution from denaturant. Both models propose that binding to GroEL reduces the probability of aggregation by segregating individual partially folded protein chains inside the central cavity of GroEL, and may also alter some noncovalent interactions that hold the chain in a kinetically trapped conformation, but they differ in their view of subsequent events. According to one view, ATP hydrolysis releases the bound chain into the free solution where it can fold spontaneously; if it fails to bury all its hydrophobic residues before it encounters another GroEL molecule, it will bind again, and the process will be repeated until folding is complete. The problem with this model is that it does not offer any protection against aggregation when the chain is

jumping between GroEL oligomers, but we know that aggregation is a real problem *in vivo* (Horwich *et al.*, 1993). The alternative model proposes that ATP hydrolysis releases the chain into the internal cavity of GroEL where it can internalize its hydrophobic groups in a sequestered environment and so avoid aggregation; this type of model (Martin *et al.*, 1993) is called the Anfinsen cage model by this author to emphasize the point that folding as observed in an Anfinsen refolding experiment is postulated to initiate inside the GroEL cavity (Ellis, 1994). The problem with this model is that once the chain unbinds from GroEL it is free to diffuse away into the free solution unless it is constrained in some way, such as being located in the cavity under the GroES component. The diffusion rate *in vitro* is fast compared to the rate of folding, but *in vivo* diffusion is reduced by macromolecular crowding, and folding need not proceed to completion, but only to the point where the danger of aggregation has passed. So one basic unresolved question is whether the sole action of GroEL is to bind sticky folding intermediates and so reduce self-aggregation, or whether, in addition, GroEL increases the time for internalizing the sticky sites by a sequestration action. It can be seen that time is of the essence in distinguishing between these possibilities.

It can be calculated that a compact partially folded polypeptide chain 50 kDa in size takes 0.5 ms to diffuse 100 nm, which is the average distance between GroEL oligomers in a typical protein refolding experiment (see Fig. 2). This is a much shorter time than it takes the average protein to refold, and even allowing for the probability that binding is not diffusion-limited and that several transfers could take place, these values suggest that proteins will not have time to fold appreciably while jumping between GroEL oligomers, as one study suggests (Weismann *et al.*, 1994). In the crowded environment found *in vivo*, diffusion will be slowed by 10-fold, but a more important difference is that the association of both GroES and the partially folded chain with GroEL will be greatly favored. Both these factors suggest that *in vivo* the folding chain will complete its folding in close association with the same GroEL/ES complex to which it first binds.

It is possible to create a highly crowded environment *in vitro* by adding large amounts of certain high-molecular-weight synthetic polymers (Zimmermann and Minton, 1993). Use of such polymers with GroEL shows that the crowding restricts the jumping of folding intermediates between GroEL oligomers observed in their absence (F-U. Hartl, personal communication, 1995). Future research will increasingly use such methods to try to elucidate the extraordinary ability of the chaperonins to assist protein folding in the intact cell.

ACKNOWLEDGMENT

I express my thanks to F-U. Hartl for stimulating discussion.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (eds.) (1994). "Molecular Biology of the Cell," 3rd ed., pp. 213–215. Garland Publishing Co., New York/London.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181,** 223–230.

Barraclough, R., and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesised large subunits into ribulose bisphosphate carboxylase in isolated intact chloroplasts. *Biochim. Biophys. Acta* **608,** 19–31.

Bessen, R. A., Kocisko, D. A., Raymond, G. J., Nandan, S., Lansbury, P. T., and Caughey, B. (1995). Non-genetic propagation of strain-specific properties of scrapie prion protein. *Nature* (*London*) **375,** 698–700.

Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. (1988). Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. *Nature* (*London*) **336,** 254–257.

Bremer, H., and Dennis, P. D. (1987). Modulation of chemical composition and other parameters of the cell by growth rate. *In* "*Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology" (F. C. Neidhardt, ed.), Vol. 2, pp. 1527–1542. American Society for Microbiology, Washington DC.

Caspar, D. L. D. (1991). Self-control of self-assembly. *Curr. Biol.* **1,** 30–32.

Caspar, D. L. D., and Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27,** 1–24.

Cavanagh, A. C., and Morton H. (1994). The purification of early-pregnancy factor from human platelets and identification as chaperonin 10. *Eur. J. Biochem.* **222,** 551–560.

Cheng, M. Y., Hartl, F-U., Martin, J., Pollock, R. A. Kalousek, F., Neupert, W. Hallberg, E. M., Hallberg, R. L., and Horwich, A. R. (1989). Mitochondrial heat shock protein hsp60 is essential for assembly of proteins imported into yeast mitochrondria. *Nature* (*London*) **337,** 620–625.

Coates, A. R. M., Shinnick, T. M., and Ellis, R. J. (1993). Chaperonin nomenclature. *Mol. Microbiol.* **8,** 787.

Ellis, R. J. (1987). Proteins as molecular chaperones, *Nature* (*London*) **328,** 378–379.

Ellis, R. J. (1990a). Molecular chaperones: The plant connection. *Science* **250,** 954–959.

Ellis, R. J. (1990b). The molecular chaperone concept. *Semin. Cell Biol.* **1,** 1–9.

Ellis, R. J. (1992). Cytosolic chaperonin confirmed. *Nature* (*London*) **358,** 191–192.

Ellis, R. J. (1993). The general concept of molecular chaperones. *In* "Molecular Chaperones" (R. J. Ellis, R. A. Laskey, and G. H. Lorimer, eds.), pp. 1–5. Chapman and Hall, London (published for the Royal Society).

Ellis, R. J. (1994). Opening and closing the Anfinsen cage. *Curr. Biol.* **4,** 633–635.

Ellis, R. J., and Hemmingsen, S. M. (1989). Molecular chaperones: Proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem. Sci.* **14,** 339–342.

Ellis, R. J., and van der Vies, S. M. (1991). Molecular chaperones. *Annu. Rev. Biochem.* **60,** 321–347.

Epstein, C. J., Goldberger, R. F., and Anfinsen, C. B. (1963). The genetic control of tertiary protein structure: Studies with model systems. *Cold Spring Harbor Symp. Quant. Biol.* **28,** 439–446.

Georgopoulos, C., Ang, D., Liberek, K., and Zylicz, M. (1990). Properties of the *Escherichia coli* heat shock proteins and their role in bacteriophage lambda growth. *In* "Stress Proteins in Biology and Medicine" (R. I. Morimoto, A. Tissieres, and C. Georgopoulos, eds.) pp. 191–221. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Gething, M-J., and Sambrook, J. (1992). Protein folding in the cell. *Nature* (*London*) **355,** 33–45.

Goloubinoff, P., Christeller, J. P., Gatenby, A. A., and Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and ATP. *Nature* (*London*) **342,** 884–889.

Goodsell, D. S. (1991). Inside a living cell. *Trends Biochem. Sci.* **16,** 203–206.

Gragerov, A., Nudler, E., Komissarova, N., Gaitanaris, G. A., Gottesman, M. E., and Nikiforov, V. (1992). Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **89,** 10,341–10,344.

Hallberg, E. M., Shu, Y., and Hallberg, R. L. (1993). Loss of mitochondrial hsp60 function: Nonequivalent effects on matrix-targeted and intermembrane targeted proteins. *Mol. Cell. Biol.* **31,** 3050–3057.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, G. C., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* (*London*) **333,** 330–334.

Hendrick, J. P., and Hartl, F-U. (1993). Molecular chaperone functions of heat shock proteins. *Annu. Rev. Biochem.* **62,** 349–384.

Herendeen, S. L., VanBogelen, R. A., and Neidhardt, F. C. (1979). Levels of major proteins of *Escherichia coli* during growth at different temperatures. *J. Bacteriol.* **139,** 185–194.

Horwich, A. L., Low, K. B., Fenton, F. A., Hirshfield, I. N., and Furtak, K. (1993). Folding *in vivo* of bacterial cytoplasmic proteins: Role of GroEL. *Cell* **74,** 909–917.

Johnson, R. A. (1987). Ph.D thesis, University of Warwick, UK.

Kaur, I., Voss, S. D., Gupta, R. S., Schell, K., Fisch, P., and Sondel, P. M. (1993). Human peripheral gamma delta T cells recognise hsp60 molecules on Daudi Burkitt's lymphoma cells. *J. Immunol.* **150,** 2046–2055.

Lamantia, M., and Lennarz, W. J. (1993). The essential function of yeast protein disulfide isomerase does not reside in its isomerase activity. *Cell* **74,** 899–908.

Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* (*London*) **275,** 416–420.

Laskey, R. A., Mills, A. D., Philpott, A., Leno, G. H., Dilworth, S. M., and Dingwall, C. (1993). The role of nucleoplasmin in chromatin assembly and disassembly. *Phil. Trans. R. Soc. Lond. B* **339,** 263–269.

Lubben, T. H., Donaldson, G. K., Viitanen, P. V., and Gatenby, A. A. (1989). Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone. *Plant Cell* **1,** 1223–1230.

Martin, J., Mayhew, M., Langer, T., and Hartl, F-U. (1993). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature (London)* **366,** 228–233.

McMullin, T. W., and Hallberg, R. L. (1988). A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli* GroEL gene. *Mol. Cell. Biol.* **8,** 371–380.

Morimoto, R., Tissieres, A., and Georgopoulos. C. (eds.) (1994). "The Biology of Heat Shock Proteins and Molecular Chaperones," pp. 1–610. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Musgrove, J. E., and Ellis, R. J. (1986). The rubisco large subunit binding protein. *Phil. Trans. R. Soc. Lond. B* **313,** 419–428.

Neidhardt, F. C. (1987). Chemical composition of *Escherichia coli. In "Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology" (F. C. Neidhardt, ed.). Vol. 1, pp. 3–6. American Society for Microbiology, Washington, DC.

Neidhardt, F. C., and VanBogelen, R. A. (1987). Heat shock response. *In "Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology" (F. C. Neidhardt, ed.), Vol. 2, pp. 1334–1345. American Society for Microbiology, Washington DC.

Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F-U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature (London)* **341,** 125–130.

Pedersen, S., Bloch. P. L., Reeh, S., and Neidhardt, F. C. (1978). Patterns of protein synthesis in *E. coli:* A catalogue of the amount of 140 individual proteins at different growth rates. *Cell* **14,** 179–190.

Pelham, H. R. B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46,** 956–961.

Puig, A., and Gilbert, H. F. (1994). Protein disulfide isomerase exhibits chaperone and antichaperone activity in the oxidative refolding of lysozyme. *J. Biol. Chem.* **269,** 7764–7771.

Rothman, J. E. (1989). Polypeptide chain binding proteins: Catalysts of protein folding and related processes in cells. *Cell* **59,** 591–601.

Sternberg, N. (1973). Properties of a mutant of *Escherichia coli* defective in bacteriophage lambda head formation (groE). *J. Mol. Biol.* **76,** 1–23.

Weismann, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78,** 693–702.

Zimmerman, S. B., and Minton, A. P. (1993). Macromolecular crowding: Biochemical, biophysical and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **22,** 27–65.

Zimmermann, S. B., and Trach, S. O. (1991). Estimation of macromolecular concentrations and excluded volume effects for the cytoplasm of *Escherichia coli. J. Mol. Biol.* **222,** 599–620.

This Page Intentionally Left Blank

2

Evolutionary Relationships of Chaperonins

RADHEY S. GUPTA

Department of Biochemistry
McMaster University
Hamilton, Canada L8N 3Z5

I. INTRODUCTION

Chaperonin homologs are found in all extant species, including eubacteria, archaebacteria, and eukaryotes. In bacteria and eukaryotic cell organelles such as mitochondria and chloroplasts, chaperonin (cpn) 10 and cpn60 are present (related proteins also referred to as hsp10 and

hsp60, or GroES and GroEL, respectively). In contrast to the above, no proteins closely related to cpn60 (or cpn10) have been found in the archaebacteria and the eukaryotic cell cytosol thus far. These species instead contain members of a protein family referred to as TCP-1 (*t*-complex polypeptide 1), which is distantly related to the cpn60 family of proteins (see Ellis, 1990; Gupta, 1990; Horwich and Willison, 1993). This chapter will examine the cpn10, cpn60, and TCP-1 families of proteins from an evolutionary perspective to understand how these proteins have evolved, and their possible relationship to one another. The sequence data on these proteins also provide valuable information concerning the origins of the eukaryotic cell and eukaryotic cell organelles (namely, the mitochondria and chloroplasts). Some of these aspects have also been reviewed elsewhere (Gupta, 1995).

II. CHAPERONIN 60 AND CHAPERONIN 10 GENE FAMILIES

A. Sequence Characteristics of Chaperonin 60 and Chaperonin 10 Families

Since the publication of the first cpn60 sequence from *Mycobacterium leprae* (referred to as 65-kDa antigen; Mehra *et al.*, 1986) a large number of cpn60 genes have been cloned and sequenced from various species (see Table I for a partial list). A global alignment of available cpn60 sequences was carried out and excerpts from it for representative species covering the main groups are shown in Fig. 1. Based on the global alignment of sequences, the following characteristics of cpn60 sequences have been noted: (i) Most of the cpn60 sequences contain one or more repeats of the sequence GGM (i.e., Gly–Gly–Met) at their extreme C-terminal ends (see Hemmingsen *et al.*, 1988; Picketts *et al.*, 1989). The exceptions to this are the chloroplast sequences and the sequences of one or more cpn60 homologs from species that contain more than one cpn60 gene (e.g., *Streptomyces albus, Mycobacterium leprae, Rhizobium meliloti*). McLennan *et al.* (1993) have examined the effect of removal of this repeat motif on the biological activity of *Escherichia coli* cpn60, which is 547 amino acids long. In their study a truncated cpn60 ($cpn60_{tr}$), which lacked the 16 C-terminal residues (i.e., GAAGGMGGMGGMGGMM), was found to behave in a manner very similar to that of the full-length

cpn60, in a variety of *in vitro* and *in vivo* tests. These tests included assembly of $cpn60_{tr}$ into characteristic oligomeric structures, presence in it of ATPase activity that was sensitive to inhibition by cpn10, binding of the truncated protein to newly synthesized pre-β-lactamase and facilitating its transport into periplasm, and ability of the $cpn60_{tr}$ to complement a temperature-sensitive mutant of cpn60, for growth at nonpermissive temperature. In an experiment where the truncated cpn60 was the only cpn60 gene present in the cells, it supported normal cellular growth as well as multiplication of various bacteriophages (namely, λ, T_4, and T_5, which require cpn60 for successful assembly) (Georgopoulos *et al.*, 1973). These observations indicate that the GGM repeat motif in cpn60 is a dispensable feature that does not seem to have any important biological function. McLennan *et al.* (1994) have reported that the forms of cpn60 that are lacking up to 27 residues from the C-terminal end are fully functional. However, further removal of even a single amino acid causes almost complete loss of activity by greatly reducing the ability of the protein to assemble into oligomeric structures. (ii) The mitochondrial homologs in all of the eukaryotic species examined are encoded by nuclear genes and contain a characteristic N-terminal leader sequence (rich in basic and hydroxylated amino acids such as Arg, Lys, Ser, and Thr) that is required for the mitochondrial import of the protein (Schatz, 1987; Hartl and Neupert, 1990; Singh *et al.*, 1990). An N-terminal leader sequence is also present in some of the chloroplast homologs but its role in chloroplast import has not yet been demonstrated. (iii) Several bacterial species have been shown to contain more than one cpn60 homolog. This includes a number of gram-positive bacteria (e.g., *S. albus, M. leprae,* or *M. tuberculosis*) and cyanobacteria (see Mazodier *et al.*, 1991; Kong *et al.*, 1993; Lehel *et al.*, 1993). Two cpn60 homologs are also known to be present in plant chloroplasts (Martel *et al.*, 1990; see Chapter 3). In addition, several gram-negative bacteria of the Rhizobiaceae family, which can either grow in a free-living state or as endosymbiont in the plant root nodules (namely, *R. meliloti, Bradyrhizobium japonicum, R. leguminosarum*), contain multiple (between five and six) cpn60 homologs (Fischer *et al.*, 1993; Rusanganwa and Gupta, 1993; Wallington and Lund, 1994). In *R. meliloti,* which contains two large megaplasmids that harbor many of the genes involved in nitrogen fixation (see Jordan, 1984; Banfalvi *et al.*, 1985), only one of the cpn60 genes is located on the main chromosome (Rusanganwa and Gupta, 1993). The remaining cpn60 genes are localized on the megaplasmids and are presumably involved in specialized functions related to root nodule formation and nitrogen-

TABLE I

Species Names and Accession Numbers of Sequences

Species name[a]	Accession no.[b] Cpn60	Cpn10	Refs.
Gram-negative bacteria			
Ec. coli	P06139	P05380	Hemmingsen *et al.* (1988)
Ch. vinosum	P31293	P31295	Ferreyra *et al.* (1993)
Y. enterocolitica	X68526[c]		GenBank
Ha. ducreyi	P31294	P31296	Parsons *et al.* (1992)
Ps. aeruginosa	P30718	P30720	Sipos *et al.* (1991)
L. pneumophila	P26878	P26879	Sampson *et al.* (1990)
Co. burnetii	P19421	P19422	Vodkin and Williams (1988)
N. flavescens	Z22955[c]		GenBank
N. gonorrhoeae	Z23009[c]		GenBank
R. meliloti (A)	M94192[c]	M94192[c]	Rusanganwa and Gupta (1993)
R. meliloti (B)	M94190[c]		Rusanganwa and Gupta (1993)
R. meliloti (C)	M94191[c]	M94191[c]	Rusanganwa and Gupta (1993)
R. leguminosarum	L20775[c]		Wallington and Lund (1994)
Ag. tumefaciens	P30779	P30780	Segal and Ron (1993)
Ri. tsutsugamushi	P16625	P16626	Stover *et al.* (1990)
Brad. japonicum 3A	Z22603[c]	Z22603[c]	Fischer *et al.* (1993)
E. chaffeensis	L10917[c]	L10917[c]	Sumner *et al.* (1993)
Z. mobilis	L11654[c]	L11654[c]	GenBank
Bru. abortus	P25967	P25968	Lin *et al.* (1992)
Chlamydiae and Spirochetes			
He. pylori	X73840[c]		Macchia *et al.* (1993)
C. trachomatis	P17203	P17204	Morrison *et al.* (1990)
C. pneumoniae	P31681	P31682	Kikuta *et al.* (1991)
B. burgodferi	P27575		Shanafelt *et al.* (1991)
Le. interrogans	L14682[c]	L14682[c]	Ballard *et al.* (1993)
T. pallidum	S33213[d]		Houston *et al.* (1990)
Gram-positive bacteria			
Cl. perfringens	P26821	P26822	Rusanganwa *et al.* (1992)
Cl. acetobutylicum	P30717	P30719	Narberhaus and Bahl (1992)
Th. bacteria (PS-3)	P26209	P26210	Tamada *et al.* (1991)
Ba. subtilis	P28598	P28599	Li and Wong (1992)
Ba. stearothermophilus	L10132[c]	L10132[c]	Schon and Schumann (1993)
S. albus 1	Q00767[c]	Q00769[c]	Mazodier *et al.* (1991)
S. albus 2	Q00768[c]		Mazodier *et al.* (1991)
S. coelicolor	X75206[c]	X75206[c]	GenBank
St. aureus	JN0601[d]	JN0600[d]	Ohta *et al.* (1993)
La. lactis	X71132[c]	X71132[c]	Kim and Batt (1993)
M. leprae 1	P09239	P24301	Mehra *et al.* (1986)
M. leprae 2	S25181[d]		Rinke de Wit *et al.* (1992)
M. tuberculosis 1	P06806	P09621	Shinnick (1987)
M. tuberculosis 2	X60350[c]		Kong *et al.* (1993)
Cyanobacteria			
Sy. species PCC6803	P22034		Chitnis and Nelson (1991)

(*continues*)

TABLE I

(*Continued*)

Species name[a]	Accession no.[b] Cpn60	Cpn10	Refs.
Syn. species PCC6301	P12834		Cozens and Walker (1987)
Syn. species PCC7942	P22879	P22880	Webb *et al.* (1990)
Mitochondria and Chloroplasts			
Sa. cerevisiae (m)	P19882		Reading *et al.* (1989)
Human (m)	P10809		Jindal *et al.* (1989)
Chi. hamster (m)	P18687		Picketts *et al.* (1989)
Mouse (m)	P19226		Venner and Gupta (1990a)
	P19227	D26772[c]	Venner and Gupta (1990b);
Rat (m)			Hartman *et al.* (1992)
H. virescens (m)	P25420		Miller and Leclerc (1990)
Cucurbita (m)	X70868[c]		Tsugeki *et al.* (1992)
Maize (m)	P29185		Prasad and Stewart (1992)
Br. napus (m)	Z27165[c]		GenBank
Tr. cruzi (m)	L08791[c]		GenBank
A. thaliana (m)	P29197	L02843[c]	Prasad and Stewart (1992)
Bovine (m)		S29974[d]	Pilkington and Walker (1993)
Cy. caldarium (c)	P28256		Maid *et al.* (1992)
Wheat (c)	P08823		Hemmingsen *et al.* (1988)
A. thaliana (c)	P21240		Zabaleta *et al.* (1992)
Br. napus α (c)	P21239		Martel *et al.* (1990)
Br. napus β (c)	P21241		Martel *et al.* (1990)
Spinach (c)		Q02073[c]	Bertsch *et al.* (1992)
TCP-1 Sequences			
Chi. hamster	P18279		Ahmad and Gupta (1990)
Human	P17987		Kirchhoff and Willison (1990)
Rat	P28480		Kubota *et al.* (1991)
Mouse 1a	JH0475[d]		Kubota *et al.* (1991)
Mouse 1b	P11983		Willison *et al.* (1986)
A. thaliana	P28769		Mori *et al.* (1992)
D. melanogaster	P12613		Ursic and Ganetzky (1988)
Su. shibatae	P28488		Trent *et al.* (1992)
Mouse (matricin)	L20509[c]		Joly *et al.* (1993)
Avena sativa	X75777[c]		Ehmann *et al.* (1993)
Sa. cerevisiae	P12612		Ursic and Culbertson (1991)

[a] *A, Arabidopsis; Ag, Agrobacterium; B, Borrelia; Ba, Bacillus; Br, Brassica; Bra, Bradyrhizobium; Bru, Brucella; C, Chlamydia; Ch, Chromatium;* Chi, Chinese; *Cl, Clostridium; Co, Coxiella; Cy, Cyanidium; D, Drosophila; E, Ehrlichia; Ec, Escherichia; H, Heliothis; Ha, Haemophilus; He, Heliobacter; L, Legionella; La, Lactococcus; Le, Leptospira; M, Mycobacterium; N, Neisseria; P, Porphyromona; Ps, Pseudomonas; R, Rhizobium; Ri, Rickettsia; S, Streptomyces; Sa, Saccharomyces; St, Staphylococcus; Su, Sulfolobus; Sy, Synechocystis; Syn, Synechococcus; T, Treponema; Th, Thermophillic; Tr, Trypanosoma; Y, Yersinia; Z, Zymomonas;* (c) chloroplast; (m) mitochondria, other numbers or letters after the species name refer to a particular gene.

[b] All accession numbers are from the SWISSPROT database unless noted otherwise.

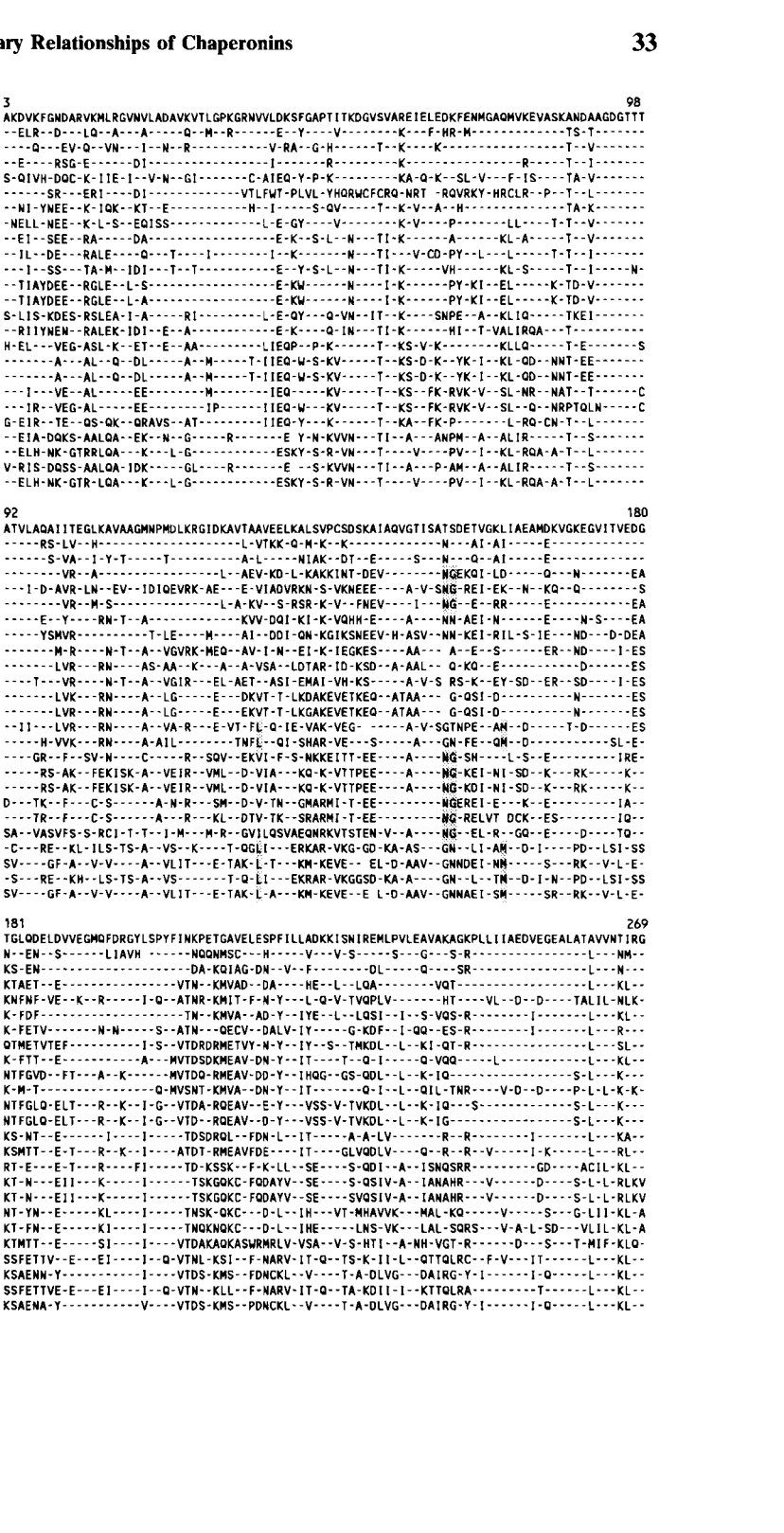
[c] GenBank database.

[d] Protein Identification Resource (PIR) database.

fixation processes (see Govezensky *et al.*, 1991; Rusanganwa and Gupta, 1993; Fischer, 1994). When multiple cpn60 genes are present in a species, the sequences generally show close evolutionary relationship to one another (see Section II,D). (iv) The global alignment of cpn60 sequences indicates that a large number of positions in the sequence are highly conserved among all species. The pairwise amino acid identity/similarity matrix of cpn60 sequences indicates that the minimum identity that is observed between any two cpn60 sequences is 42% over their entire length (Table II). In addition, between 18 and 20% of the amino acid substitutions are conservative replacements. (v) The global alignment of cpn60 sequences has identified a number of signature sequences that are uniquely shared (*a*) among members of gram-negative α-purple subdivision of bacteria and mitochondrial homologs and (*b*) between cyanobacteria and chloroplast homologs (see Fig. 1). The significance of this observation regarding the origin of mitochondria and chloroplasts is discussed later (Section II,D). In addition, all cpn60 homologs from the gram-positive species can be distinguished from other species by the lack of a single amino acid at a position corresponding to residue 153 in the *E. coli* cpn60 sequence (Fig. 1). Whether this difference is due to a deletion in the gram-positive species, or insertion in other species, is not clear at present. It should be noted in this context that similar studies with the hsp70/DnaK family of sequences have identified a large insert (23–27 residues) that is present in the N-terminal quadrants of all homologs from gram-negative bacteria and eukaryotes but is not seen in any of the hsp70s from archaebacteria and gram-positive bacteria (Gupta and Golding, 1993; Gupta *et al.*, 1994). The presence of signature sequences that are uniquely shared by homologs from different species

---

Fig. 1. Multiple alignment of cpn60 sequences from representative prokaryotic and eukaryotic species. The alignment was carried out as described in our earlier work (Picketts *et al.*, 1989; Gupta and Singh, 1992). The numbers at the top indicate the position in the *Escherichia coli* cpn60 sequence. Dashes (-) indicate residues identical to that present in the top row. Small portions of the sequence from both N- and C-terminal ends, which could not be properly aligned in all sequences, are not shown. The shaded (or hatched) residues in a column identify signature sequences that are uniquely shared either between cyanobacteria and chloroplast homologs (c), or between α-purple bacteria and mitochondria (m). Species names are as indicated in Table I.


3
98
Ec.coli AKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTT
L.pneumophila --ELR--D---LQ--A---A-----Q--M--R------E--Y----V--------K---F-HR-M-------------T-------
N.gonorrhoeae ----Q---EV-Q--VN---I--N--R-----------V-RA--G-H------T--K----K----------------T--V-------
Rh.meliloti(A) --E----RSG-E------DI-----------------I-------R---------K----------------R-----T--I-------
R.tsutsugam. S-QIVH-DQC-K-IIE-I--V-N--GI-------C-AIEQ-Y-P-K---------KA-Q-K--SL-V---F-IS----TA-V-------
Z.mobilis ------SR---ERI----DI-------------VTLFWT-PLVL-YHQRWCFCRQ-NRT -RQVRKY-HRCLR--P--T--L-------
C.trachomatis --NI-YNEE--K-IQK--KT--E-----------N--I-----S-QV-----T--K-V--A--H----------TA-K-------
T.pallidum -NELL-NEE--K-L-S--EQISS-------------L-E-GY----V--------K-V----P--------LL----T-T--V-------
Ba.subtilis(1) --EI--SEE--RA------DA-----------------E-K--S-L--N---TI-K------A------KL-A-----T--V-------
S.albus(1) --IL--DE---RALE----Q---T----I--------I--K-------N---TI---V-CD-PY--L---L-----T-T--I-------
La.lactis ---I--SS---TA-M--IDI---T--T----------E--Y-S-L--N---TI-K-----VH------KL-S-----T--I-----N-
M.leprae(1) --TIAYDEE--RGLE--L-S-----------------E-KW------N----I-K------PY-KI--EL-----K-TD-V-------
M.tubercul.(2) --TIAYDEE--RGLE--L-A-----------------E-KW------N----I-K------PY-KI--EL-----K-TD-V-------
Sy.spe.6803 S-LIS-KDES-RSLEA-I-A-----RI----------L-E-QY---Q-VN--IT--K----SNPE--A--KLIQ-----TKEI-------
Syn.spe.7942 --RIIYNEN--RALEK-IDI--E--A-----------E-K----Q-IN---TI-K------HI--T-VALIRQA---T----------
Sa.cerev.(m) H-EL---VEG-ASL-K--ET--E--AA----------LIEQP--P-K------T--KS-V-K--------KLLQ-----T-E-------S
Human(m) -------A---AL--Q--DL-----A--M-----T-IIEQ-W-S-KV-----T--KS-D-K--YK-I--KL-QD--NNT-EE-------
Mouse(m) -------A---AL--Q--DL-----A--M-----T-IIEQ-W-S-KV-----T--KS-D-K--YK-I--KL-QD--NNT-EE-------
Maize(m) ---I---VE--AL-----EE---------M--------IEQ----KV-----T--KS--FK-RVK-V--SL-NR--NAT--T------C
Br.napus(m) ---IR--VEG-AL-----EE--------IP------IIEQ-W---KV-----T--KS--FK-RVK-V--SL--Q--NRPTQLN-----C
Tr.cruzi(m) G-EIR--TE--QS-QK--QRAVS--AT---------IIEQ-Y---K------T--KA--FK-P-------L-RQ-CN-T--L-------
Wheat(c) --EIA-DQKS-AALQA--EK--N--G-----R-------E Y-N-KVVN---TI--A---ANPM--A--ALIR-----T--S-------
A.thaliana(c) --ELH-NK-GTRRLQA---K---L-G-----------ESKY-S-R-VN--T----V----PV--I--KL-RQA-A-T--L-------
Br.napus α(c) V-RIS-DQSS-AALQA-IDK-----GL----R-------E --S-KVVN---TI--A---P-AM--A--ALIR-----T--S-------
Br.napus β(c) --ELH-NK-GTR-LQA---K---L-G-----------ESKY-S-R-VN---T----V----PV--I--KL-RQA-A-T--L-------
92
180
Ec.coli ATVLAQAIITEGLKAVAAGMNPMDLKRGIDKAVTAAVEELKALSVPCSDSKAIAQVGTISATSDETVGKLIAEAMDKVGKEGVITVEDG
L.pneumophila -----RS-LV--H----------L-VTKK-Q-M-K--K-----------N---AI-AI-----E----------
N.gonorrhoeae ------S-VA--I-Y-T-----T----------A-L-----NIAK--DT--E----S---N---Q--AI-----E----------
Rh.meliloti(A) --------VR--A----------L--AEV-KD-L-KAKKINT-DEV-------NGEKQI-LD-----Q---N-------EA
R.tsutsugam. ---I-D-AVR-LN--EV--IDIQEVRK-AE---E-VIAQVRKN-S-VKNEEE----A-V-SNG-REI-EK--N--KQ--Q--------S
Z.mobilis --------VR--M-S----------L-A-KV--S-RSR-K-V--FNEV-----I---NG--E--RR-----E----------EA
C.trachomatis -----E--Y----RN-T--A----------KVV-DQI-KI-K-VQHH-E----A----NN-AEI-N------E----N-S----EA
T.pallidum -----YSMVR----------T-LE----M----AI--DDI-QN-KGIKSNEEV-H-ASV---NN-KEI-RIL-S-IE---ND---D-DEA
Ba.subtilis(1) -------M-R----N-T--A--VGVRK-MEQ--AV-I-N--EI-K-IEGKES----AA--- A--E--S------ER--ND----I-ES
S.albus(1) -------LVR---RN----AS-AA--K---A--A-VSA--LDTAR-ID-KSD--A-AAL-- Q-KQ--E----------D------ES
La.lactis ----T---VR----N-T--A--VGIR---EL-AET--ASI-EMAI-VH-KS-----A-V-S RS-K--EY-SD--ER--SD----I-ES
M.leprae(1) -------LVK---RN----A--LG-----E---DKVT-T-LKDAKEVETKEQ--ATAA--- G-QSI-D----------N-------ES
M.tubercul.(2) -------LVR---RN----A--LG-----E---EKVT-T-LKGAKEVETKEQ--ATAA--- G-QSI-D----------N-------ES
Sy.spe.6803 ---II---LVR---RN----A--VA-R---E-VT-FL-Q-IE-VAK-VEG- -----A-V-SGTNPE--AM--D-----T-D------ES
Syn.spe.7942 -----H-VVK---RN----A-AIL-------TNFL--QI-SHAR-VE---S-----A---GN-FE--QM--D----------SL-E-
Sa.cerev.(m) ----GR--F--SV-N----C-----R--SQV--EKVI-F-S-NKKEITT-EE----A----NG-SN----L-S--E---------IRE-
Human(m) -----RS-AK--FEKISK-A--VEIR--VML--D-VIA---KQ-K-VTTPEE----A----NG-KEI-NI-SD--K---RK-----K--
Mouse(m) -----RS-AK--FEKISK-A--VEIR--VML--D-VIA---KQ-K-VTTPEE----A----NG-KDI-NI-SD--K---RK-----K--
Maize(m) D---TK--F---C-S------A-N-R---SM--D-V-TN--GMARMI-T-EE---------NGEREI-E---K--E---------IA--
Br.napus(m) ----TR--F---C-S------A---R---KL--DTV-TK--SRARMI-T-EE--------NG-RELVT DCK--ES--------IQ--
Tr.cruzi(m) SA--VASVFS-S-RCI-T-T--I-M---M-R--GVILQSVAEQNRKVTSTEN-V--A----NG--EL-R--GQ--E----D----TQ--
Wheat(c) -C---RE--KL-ILS-TS-A--VS--K----T-QGLI---ERKAR-VKG-GD-KA-AS---GN--LI-AM--D-I----PD--LSI-SS
A.thaliana(c) SV----GF-A--V-V----A--VLIT---E-TAK-L-T---KM-KEVE-- EL-D-AAV--GNNDEI-NN-----S---RK--V-L-E-
Br.napus α(c) -S---RE--KH--LS-TS-A--VS-------T-Q-LI---EKRAR-VKGGSD-KA-A----GN--L--TM--D-I-N--PD--LSI-SS
Br.napus β(c) SV----GF-A--V-V----A--VLIT---E-TAK-L-A---KM-KEVE--E L-D-AAV--GNNAEI-SM-----SR--RK--V-L-E-
181
269
Ec.coli TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATAVVNTIRG
L.pneumophila N--EN--S------LIAVH ------NQQNMSC---H-----V---V-S-----S---G---S-R----------L---NM--
N.gonorrhoeae KS-EN----------DA-KQIAG-DN--V--F--------QL-----Q----SR----------L---N---
Rh.meliloti(A) KTAET--E----------VTN--KMVAD--DA----HE--L--LQA--------VQT----------L---KL--
R.tsutsugam. KNFNF-VE--K--R-----I-Q--ATNR-KMIT-F-N-Y---L-Q-V-TVQPLV-------HT----VL--D--D----TALIL-NLK-
Z.mobilis K-FDF----------TN--KMVA--AD-Y--IYE--L--LQSI--I--S-VQS-R--------I-------L---KL--
C.trachomatis K-FETV-------N-N-----S--ATN---QECV--DALV-IY-----G-KDF--I-QQ--ES-R--------I-------L---R---
T.pallidum QTMETVTEF----------I-S--VTDRDRMETVY-N-Y--IY--S--TMKDL--L--KI-QT-R----------L---SL--
Ba.subtilis(1) K-FTT--E----------A---MVTDSDKMEAV-DN-Y--IT----T--Q-I-----Q-VQQ-----L----------L---KL--
S.albus(1) NTFGVD--FT---A--K------MVTDQ-RMEAV-DD-Y--IHQG--GS-QDL--L--K-IQ----------S-L---K---
La.lactis K-M-T----------Q-MVSNT-KMVA--DN-Y--IT-------Q-I--L--QIL-TNR----V-D--D----P-L-L-K-K-
M.leprae(1) NTFGLQ-ELT---R--K--I-G--VTDA-RQEAV--E-Y---VSS-V-TVKDL--L--K-IQ---S----------S-L---K---
M.tubercul.(2) NTFGLQ-ELT---R--K--I-G--VTD--RQEAV--D-Y---VSS-V-TVKDL--L--K-IG----------S-L---K---
Sy.spe.6803 KS-NT--E------I----I-----TDSDRQL--FDN-L--IT-----A-A-LV-------R--R--------I-----L---KA--
Syn.spe.7942 KSMTT--E-T---R--K--I----ATDT-RMEAVFDE----IT----GLVQDLV----Q--R--R--V-----I-K-----L---RL--
Sa.cerev.(m) RT-E---E-T---R----FI-----TD-KSSK--F-K-LL--SE----S-QDI--A--ISNQSRR---------GD----ACIL-KL--
Human(m) KT-N---EII---K-----I------TSKGQKC-FQDAYV--SE----S-QSIV-A--IANAHR---V------D----S-L-L-RLKV
Mouse(m) KT-N---EII---K-----I------TSKGQKC-FQDAYV--SE----SVQSIV-A--IANAHR---V------D----S-L-L-RLKV
Maize(m) NT-YN--E-----KL----I-----TNSK-QKC---D-L--IH---VT-MHAVVK---MAL-KQ-----V-----S---G-LII-KL-A
Br.napus(m) KT-FN--E-----KI----I-----TNQKNQKC---D-L--IHE-----LNS-VK---LAL-SQRS---V-A-L-SD---VLIL-KL-A
Tr.cruzi(m) KTMTT--E-----SI----I----VTDAKAQKASWRMRLV-VSA--V-S-HTI--A-NH-VGT-R------D---S---T-MIF-KLQ-
Wheat(c) SSFETTV--E---EI----I--Q-VTNL-KSI--F-NARV-IT-Q--TS-K-II-L--QTTQLRC--F-V---IT-------L---KL--
A.thaliana(c) KSAENN-Y----------I----VTDS-KMS--FDNCKL--V----T-A-DLVG---DAIRG-Y-I------I-Q-----L---KL--
Br.napus α(c) SSFETTVE-E---EI----I--Q-VTN--KLL--F-NARV-IT-Q--TA-KDII-I--KTTQLRA--------T------L---KL--
Br.napus β(c) KSAENA-Y----------V----VTDS-KMS--PDNCKL--V----T-A-DLVG---DAIRG-Y-I------I-Q-----L---KL--

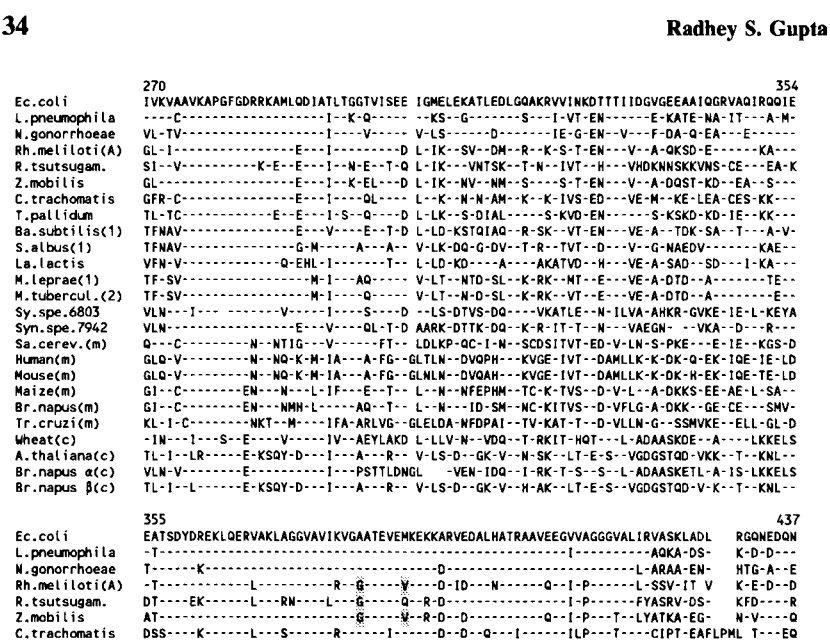


Fig. 1. *Continued.*

TABLE II

Amino Acid Identity/Similarity Matrix of cpn60 Sequences[a]

Species	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A. *Ec. coli*		76.1	74.1	68.4	50.5	53.8	55.1	55.5	49.1	59.7	56.0	52.2	47.0	51.5	46.8	50.9	58.6	58.6
B. *Co. burnetii*	83.6		72.6	67.5	52.0	52.5	56.2	56.3	50.5	61.4	57.2	53.0	48.3	52.5	47.7	50.9	58.6	59.3
C. *N. gonorrhoeae*	83.9	81.8		67.8	51.8	53.1	56.6	56.9	51.9	62.3	57.1	52.1	48.8	53.2	48.2	53.1	59.4	58.4
D. *R. meliloti* (A)	79.6	79.1	79.8		53.4	58.2	60.7	60.2	54.9	61.4	55.0	51.5	48.4	52.8	49.0	51.9	59.7	59.2
E. Human (m)	68.4	69.6	69.2	72.1		56.1	57.8	57.4	51.8	50.9	46.3	45.7	41.8	43.6	44.0	42.6	50.1	46.8
F. *Sa. cerevisiae* (m)	69.1	67.7	68.3	72.7	72.4		56.1	58.6	52.8	53.5	49.9	48.5	42.9	44.1	43.5	43.2	51.2	47.7
G. Maize (m)	68.6	69.2	70.8	74.8	74.4	71.7		85.6	55.7	54.4	48.8	47.0	44.8	45.1	43.1	43.8	49.7	52.3
H. *A. thaliana* (m)	69.7	69.9	71.9	76.2	73.1	73.5	93.6		54.1	55.7	49.0	47.8	44.4	45.2	42.9	44.5	51.4	51.8
I. *Tr. cruzi* (m)	65.3	66.4	67.4	70.5	68.5	69.5	71.5	71.5		49.6	47.7	46.4	41.8	42.0	42.8	41.6	48.1	47.7
J. *C. trachomatis*	73.9	74.1	75.1	73.5	68.2	67.8	69.1	70.0	65.4		54.0	56.7	47.9	50.2	50.4	49.8	59.4	57.1
K. *Sy.* sp. PCC6803	73.5	74.6	74.7	72.8	66.3	67.7	66.8	66.8	65.7	71.6		56.9	55.4	56.4	56.3	54.4	56.4	56.9
L. *Cy. caldarium* (c)	70.6	71.2	70.6	69.1	65.2	65.5	64.2	66.7	64.6	71.3	75.1		52.6	51.7	55.2	51.5	54.5	54.5
M. Wheat (c)	66.3	67.8	65.1	67.6	59.5	63.3	61.9	62.3	59.1	65.9	73.1	70.6		50.6	80.7	49.9	49.7	48.6
N. *A. thaliana* (c)	69.0	69.2	69.7	68.4	60.5	61.6	63.5	62.7	60.7	67.7	71.7	71.0	67.9		49.4	93.7	53.4	53.1
O. *Br. napus* α (c)	64.6	67.0	64.8	66.6	62.0	62.0	60.2	61.1	60.0	66.4	72.3	70.7	89.5	67.2		49.5	50.3	51.4
P. *Br. napus* β (c)	67.7	68.1	68.6	67.3	58.8	61.2	60.8	62.5	57.6	66.7	69.1	70.2	67.9	95.7	68.2		52.9	52.9
Q. *Cl. perfringens*	72.7	73.8	73.7	74.4	68.1	67.5	67.5	68.5	65.0	72.6	73.5	72.7	66.6	68.2	66.6	67.2		60.9
R. *M. leprae*	71.9	73.3	73.6	73.8	64.4	64.9	68.6	68.6	65.1	72.6	72.2	71.2	66.3	68.1	66.6	67.9	75.7	

[a] The amino acid identity (upper triangle) or similarity (lower triangle) between pairs of sequences was determined using the PALIGN program of PCGENE package (Myers and Miller, 1988), using the structure genetic matrix and open gap and unit gap costs of 7 and 2, respectively. The species names are the same as indicated in Table I.

provides important information concerning their evolutionary relationships (see Rivera and Lake, 1992; Gupta *et al.*, 1994). (vi) A global alignment of available cpn10 sequences has also been carried out (Fig. 2). These sequences (minimum identity ≃30–35%) are more divergent than the cpn60 sequences and no unique sequence feature distinctive of any major group of species has thus far been identified. However, two observations on gene organization that are of interest may be noted. First, although cpn10 gene sequences in bacteria form part of an operon with a downstream cpn60 gene, several species that contain multiple cpn60 genes have no cpn10 sequences upstream of the cpn60 genes (Mazodier *et al.*, 1991; Kong *et al.*, 1993; Rusanganwa and Gupta, 1993). The significance of this observation is unclear at present, but it suggests that, in contrast to the presumed specialized functions of different cpn60 homologs, the same cpn10 protein is able to function with different cpn60 partners. The second interesting observation is that the cpn10-related protein in higher plant chloroplasts has an apparent M_r of ≃24,000 (Bertsch *et al.*, 1992). The gene for this larger protein has undergone a gene duplication event, involving head-to-tail joining of two cpn10 sequences (Bertsch *et al.*, 1992). The functional significance of this gene duplication event and whether it is a characteristic of all plant chloroplast cpn10 sequences remain unclear at present (see Chapter 3).

B. Sequence Similarity between Chaperonin 10 and Chaperonin 60

Chaperonin 60 and cpn10 both exhibit a sevenfold axis of symmetry and function as a team in the protein folding/assembly processes (Chandrashekhar *et al.*, 1986; Hendrick and Hartl, 1993). Although these two molecules differ greatly in size (≃60 kDa vs 11 kDa), we have noticed that a segment of the larger chaperonin may be related to the cpn10 molecule (Gupta, 1994). Figure 3 presents an alignment of a segment of cpn60 sequences with the cpn10 sequences from representative groups of species. With the introduction of a small number of gaps, the cpn10 sequences show reasonably good alignment with the indicated region of cpn60 molecules. Many of the positions in these two groups of proteins are highly conserved. In the alignment between the *E. coli* cpn60 and the cpn10 from *Bacillus subtilis* (94 amino acid overlap), 31 identical and an additional 18 conservative substitutions were observed. The observed alignment score of these sequences differed by 4.2 standard deviations

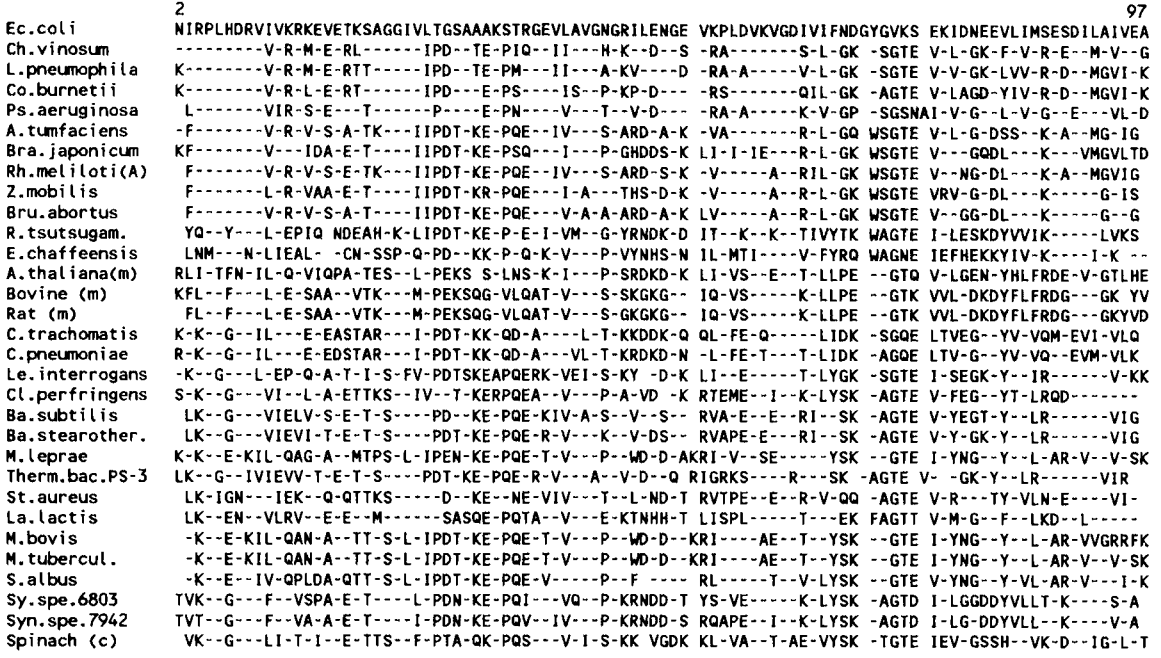


Fig. 2. Multiple alignment of cpn10 sequences from representative prokaryotic and eukaryotic species. The residues identical to that in the top row are indicated by dashes (-).

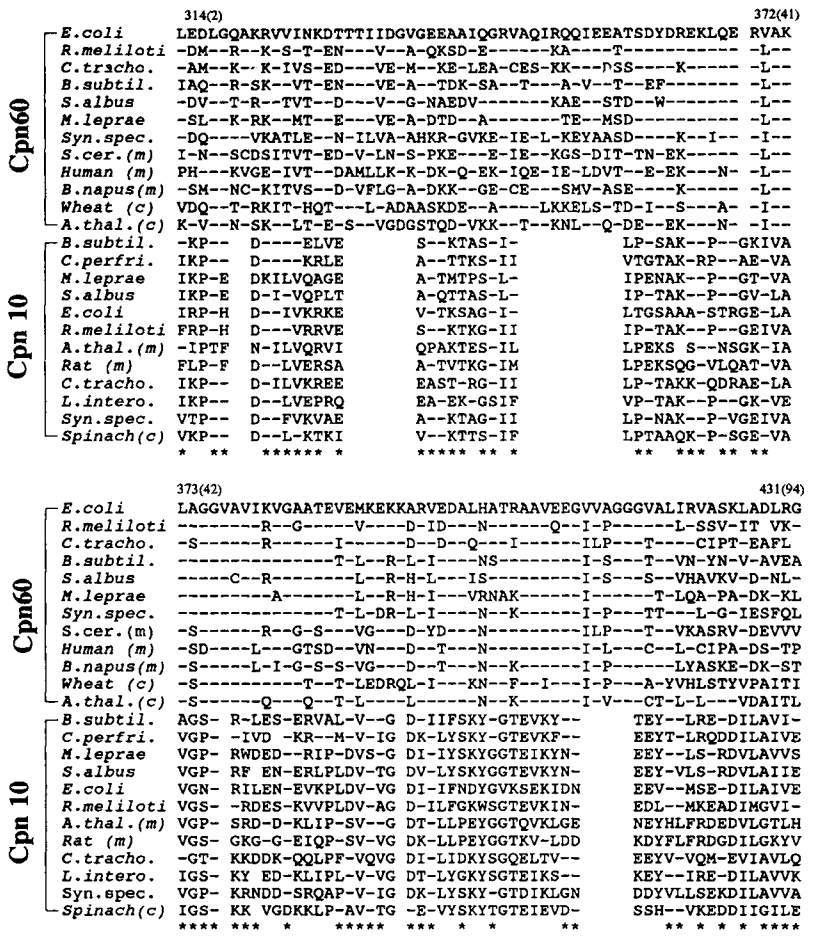


Fig. 3. Sequence similarity between cpn10 and a region of the cpn60 molecule. Sequence alignment for representative cpn10 and cpn60 sequences is shown. The numbers at the top indicate the position of residues in *E. coli* cpn60 or *B. subtilis* cpn10 (in parentheses) sequences. An asterisk (*) below a row indicates that either identical or conserved residues are present in at least three sequences of each kind. The conserved amino acids were determined according to PAM120 substitution matrix (Altschul, 1991).

from the average of the randomized (100 times) scores of the same sequences, indicating that the observed similarity between these two sequences is significant ($p < 0.0001$) (Pearson, 1990). Although the bio-

logical significance of the observed similarity remains to be determined, it is possible that this region may contain information for assembly into structures exhibiting sevenfold rotational symmetry and for interaction with diverse proteins—two characteristics that are common and essential for both these proteins (Hendrick and Hartl, 1993). It should also be mentioned in this regard that Martel *et al.* (1990) have previously suggested that a different region of the cpn60 family of proteins (residues 114 to 232 in *E. coli*) shows limited sequence similarity to the cpn10 sequences. Although we do not find this sequence similarity to be significant, it is possible that this region has diverged much more from the original ancestral sequence. Such a case would suggest the possibility of a gene duplication event in the evolution of the cpn60 gene family.

C. Sequence Similarity between Chaperonin 60 and Other Proteins

1. Other Antigenic Proteins

In addition to the cpn10 and TCP-1 families, limited sequence similarities between cpn60 homologs and a number of other proteins (e.g., penicillin-binding proteins, Pol polyprotein, cholera enterotoxin, immunoglobulin G (IgG) delta (δ) chain, myosin light chain, thyroglobulin, DNA-binding protein, cytokeratin, neurofilament triplet protein, glutamate decarboxylase) have also been observed (Dudani and Gupta, 1989; Jones *et al.*, 1993). Since many of these proteins correspond to known autoantigens, the observation is suggestive that molecular mimicry between cpn60 and these proteins could be playing a role in the development of autoimmune diseases (Kaufmann, 1990; Young *et al.*, 1990; Cohen, 1991; Brudzynski *et al.*, 1992; Jones *et al.*, 1993).

2. Relationship to mRNA Stability Gene: ams$^+$

Chanda *et al.* (1985) have previously reported the cloning of an *E. coli* gene (*ams*$^+$) that complemented the *ams-1* mutation affecting mRNA stability in such cells. The *ams*$^+$ gene was reported to be contained within a 595-base-pair fragment that harbored an open reading frame of 450 nucleotides encoding for a protein of M_r 17,000. Interestingly, the first 117 residues of this protein show nearly a perfect match (116 out of 117 identical) with the amino acids 307–423 in the *E. coli* GroEL or cpn60

sequence. However, the C-terminal portion of the sequence is unrelated to cpn60 (Chanda *et al.*, 1985).

To examine the relationship of *ams*$^+$ protein to cpn60 sequences, the nucleotide sequence of the 595 nucleotide *ams*$^+$ fragment was translated in all three reading frames and compared with the *E. coli* cpn60 sequence. Results of these analyses are presented in Fig. 4. The translation of the 5′-end 378 nucleotides in the first frame shows a perfect match with the *E. coli* GroEL, whereas beyond this point, the translation in the third frame corresponds to the GroEL. The observed frame shift is due to a

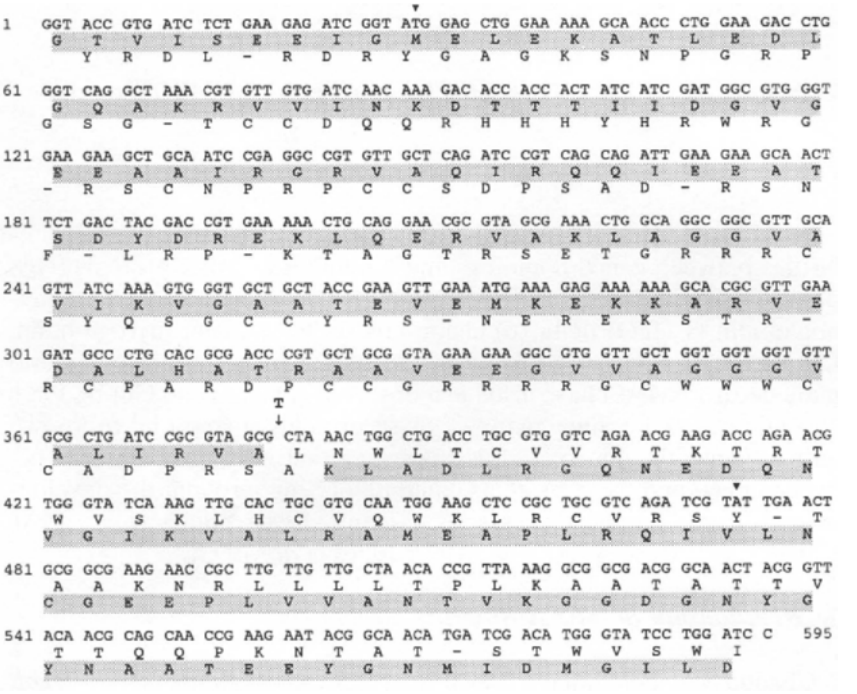


Fig. 4. Translation of the *ams*$^+$ gene fragment (Chanda *et al.*, 1985) in the first and third reading frames. The shaded regions identify amino acids corresponding to those of the *E. coli* cpn60 sequence (Hemmingsen *et al.*, 1988). The arrowheads indicate the beginning and end of the postulated *ams*$^+$ protein (Chanda *et al.*, 1985). The arrow after nucleotide 378 indicates the position where a missing base (T) causes the apparent frame shift in the reported sequence. The protein sequence encoded by this fragment corresponds to residues 298–495 in *E. coli* cpn60.

single missing nucleotide (T after position 378) in the reported sequence. Upon inclusion of the missing nucleotide, the sequence reported by Chanda *et al.* (1985) encodes for a portion of the *E. coli* GroEL or cpn60 from residues 298 to 495. In subsequent work, Claverie-Martin *et al.* (1989) have observed that the gene fragment cloned by Chanda *et al.* (1985) does not complement the *ams-1* mutation as claimed in the earlier work (Chanda *et al.*, 1985). The *ams*$^+$ gene cloned and characterized by Claverie-Martin *et al.* (1989) encodes a protein with M_r 110,000 and it shows no sequence similarity to the cpn60 from *E. coli* or other species. Thus the sequence for *ams*$^+$ gene reported earlier (Chanda *et al.*, 1985) corresponds to a fragment of the *E. coli* cpn60 gene, and there is no sequence similarity between the *ams*$^+$ protein and cpn60 sequences.

D. Phylogenetic Analysis Based on Chaperonin 60 and Chaperonin 10 Sequences

One of the utilities of gene/protein sequence data is in inferring evolutionary relationships between various organisms (see Fitch and Margoliash, 1967; Felsenstein, 1982). The concept that "molecules or molecular sequences constitute documents of evolutionary history" was first put forth clearly by Zuckerkandl and Pauling (1965). Its central tenet that changes in gene/protein sequences occur in a time-dependent manner has led to the concept of the "molecular or evolutionary clock." Due to the recent developments in molecular biology that have greatly facilitated the delineation of gene/protein sequences from various species, fundamental advances in our understanding of the evolutionary relationships between extant species are occurring (see Woese, 1987; Gogarten *et al.*, 1989; Iwabe *et al.*, 1989; Woese *et al.*, 1990; Gupta and Singh, 1994). The cpn60 sequences, because of their large size, high degree of sequence conservation, and presence in all eubacteria and eukaryotic organelles, provide a good molecular model to investigate the evolutionary relationship within and between these groups. cpn10 sequences, although smaller in size and information content, are also useful in this regard.

The evolutionary relationship based on cpn60 sequences was examined for a group of 48 sequences covering all main groups of bacteria and eukaryotic organelles (namely, mitochondria and chloroplasts). A global alignment of various sequences was carried out, and only those regions

of the alignment (corresponding to those shown in Fig. 1) that could be aligned without any ambiguity in all species were employed for phylogenetic analysis. A neighbor-joining distance matrix tree based on these sequences is presented in Fig. 5. The tree based on cpn60 sequences was rooted using the TCP-1 (or TF55) sequence from the archaebacteria *Sulfolobus shibatae.* TCP-1 is a distant homolog of cpn60 (see Section III,A). A comparable neighbor-joining tree was also constructed based on cpn10 sequences. Since the gram-positive group of bacteria are indicated to be an ancestral lineage (see Gupta and Golding, 1993, and Fig. 5), this tree was rooted using a member from this group (*M. leprae*). From comparison of Figs. 5 and 6, it is apparent that the orders of branching of various species in both cpn60 and cpn10 trees are very similar, which independently shows that both of these gene sequences give similar results. From these trees a number of inferences concerning the relationships between eubacterial species, as well as organellar homologs, can be made. (i) Based on cpn60 and cpn10 trees (Figs. 5 and 6), the various homologs examined could be arranged in a number of distinct groups, namely, gram-positive bacteria, cyanobacteria and chloroplasts, chlamydiae and spirochetes, gram-negative purple bacteria, and eukaryotic mitochondrial homologs. These groups correspond to the main divisions of bacteria and eukaryotic organelles that have been identified based on small subunit (SSU) rRNA and other gene sequences (see Fox *et al.,* 1980; Woese, 1987; Cedergren *et al.,* 1988; Woese, 1991). After the present work was completed, Viale and co-workers have independently reported phylogenetic analysis based on cpn60 sequences (Viale and Arkaki, 1994; Viale *et al.,* 1994). The inferences reached by these authors are very similar to those reported here and in our other work (Gupta, 1995). (ii) In the cpn60 tree, homologs from the gram-positive group of bacteria show the deepest branching (Fig. 5). Within this group, the species with high and low G + C contents form separate clades. The monophyletic grouping of high and low G + C gram-positive species is shown to be reliable by bootstrap analysis of the sequence data (Viale *et al.,* 1994; Gupta and Singh, 1994; R. S. Gupta, unpublished results, 1994). The ancient nature of gram-positive bacteria is also supported by phylogeny based on other gene sequences, including rRNA and hsp70 (Woese, 1987; Gupta and Golding, 1993). Recent studies based on hsp70 gene sequences strongly indicate that the homologs from the gram-positive group of bacteria that do not contain a large insert in the N-terminal quadrant constitute the ancestral form of the protein from which all other homologs are derived (Gupta and Singh, 1992; Gupta

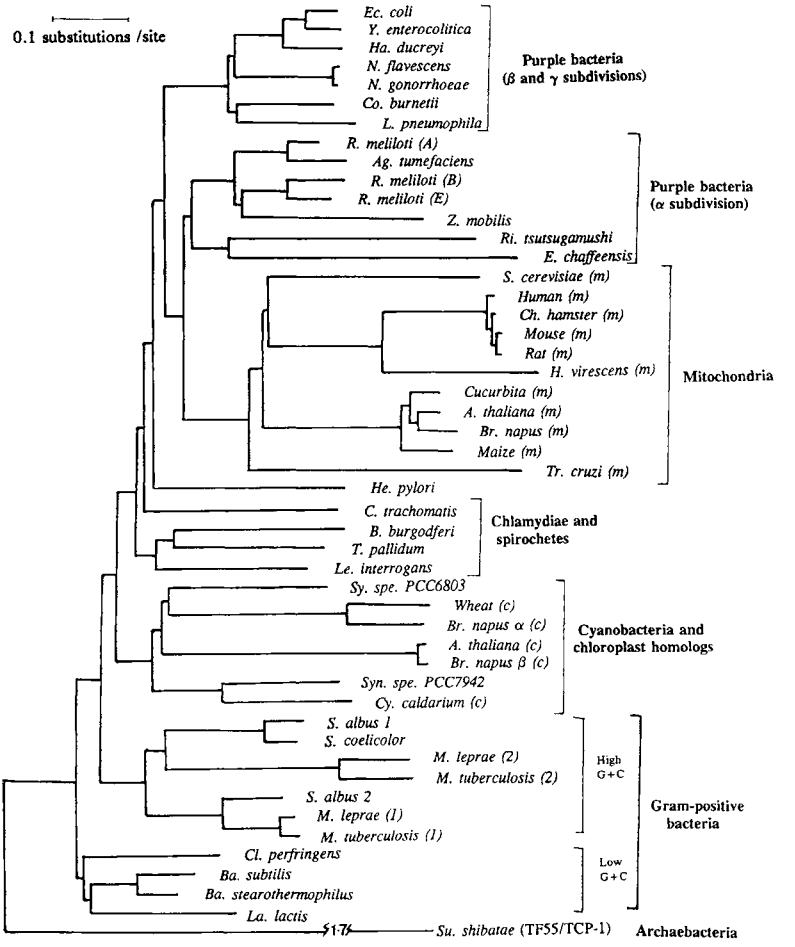


Fig. 5. Evolutionary tree based on cpn60 sequences. The tree shown is a neighbor-joining distance matrix tree constructed using the PHYLIP version 3.5C program package (Saitou and Nei, 1987; Felsenstein, 1991). The phylogenetic analysis was carried out on the sequence region and alignment similar to that shown in Fig. 1. The tree was rooted using the TCP-1 homolog (TF55) from *Sul. shibatae* as the outgroup. The branching of mitochondrial homologs with the α-purple subdivision of bacteria and of chloroplasts with the cyanobacteria should be noted.

and Golding, 1993). Following the gram-positive group of bacteria, the order of branching of other groups, namely cyanobacteria, chlamydiae, spirochetes, and purple bacteria, in both cpn60 and cpn10 trees is again

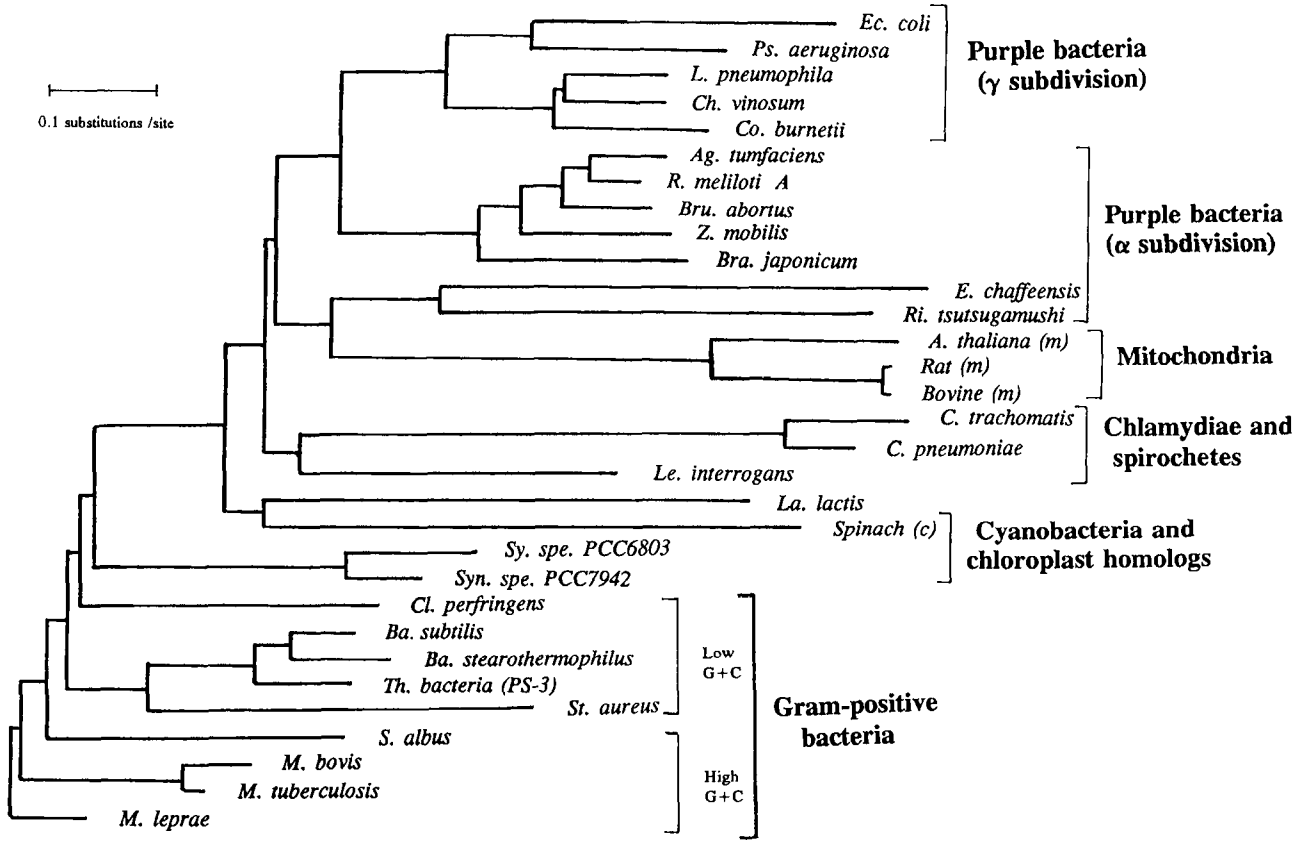


Fig. 6. Phylogenetic tree based on cpn10 sequences. The tree was constructed as described in Fig. 5 using a sequence alignment similar to that shown in Fig. 2.

similar to that based on other gene phylogenies (e.g., hsp70, SSU rRNA) (Woese, 1991; see Gupta and Golding, 1993). (iii) The phylogenetic tree based on cpn60 sequences indicates that multiple cpn60 homologs, when present in a species or a group of species, generally show a clustering in the tree (see Fig. 5). This observation suggests that the multiple cpn60 homologs arose by independent gene duplication events that took place within different branches of the tree. On the basis of the known cpn60 sequences and their positions in the phylogenetic tree, three independent gene duplication events can at present be identified. One of these is postulated to have occurred in the common ancestor to *Mycobacteria* and *Streptomyces* group of gram-positive bacteria. A second such event in the cyanobacterial lineage could account for the two cpn60 homologs present in cyanobacteria and chloroplasts (Martel *et al.*, 1990; Lehel *et al.*, 1993). The third gene duplication event appears to have occurred in the most recent common ancestor to the Rhizobiaceae group of bacteria. These groups of bacteria, which are involved in nitrogen fixation (e.g., *R. meliloti, B. japonicum, R. leguminosarum*), contain either 5 or 6 cpn60 genes (Fischer *et al.*, 1993; Rusanganwa and Gupta, 1993; Wallington and Lund, 1994). In *R. meliloti* all of these genes, with the exception of one that is present on the main chromosome, are located on the two megaplasmids that are present in this species (Rusanganwa and Gupta, 1993). In view of their megaplasmid localization, this gene duplication event presumably occurred in the most recent common ancestor of these species, after it had acquired the megaplasmids.

1. Mitochondrial Origin

As can be seen from Fig. 5, all of the mitochondrial cpn60 homologs form a monophyletic cluster in the phylogenetic tree. Within this cluster the various plant and animal species form distinct subgroups and the deepest branching is observed for the protist species, *Trypanosoma cruzi*. The branching order of various eukaryotic species within the mitochondrial cluster is in accordance with that based on other gene sequences (see Cedergren *et al.*, 1988; Iwabe *et al.*, 1989; Gupta *et al.*, 1994). One interesting observation that is apparent from Fig. 5 is that the mitochondrial group shares a most recent common ancestor with the members of the α-subdivision of the gram-negative purple bacteria (or proteobacteria). A close and specific relationship between cpn60 homologs from the α-purple subdivision of bacteria and mitochondria is also indicated by

the observation that the homologs from these groups contain a number of unique sequence characteristics (namely, amino acid substitutions) that are not found in other cpn60 sequences (Fig. 1). Within the α-subdivision of gram-positive bacteria, a closer affiliation of mitochondrial homologs was observed for the species *Ehrlichia chaffeensis* and *Rickettsia tsutsugamushi* (Figs. 5 and 6; see also Viale and Arakaki, 1994), which are obligate intracellular pathogens (Moulder, 1985). Bootstrap analysis of cpn60 sequence data shows that a specific relationship between these two groups of homologs is observed in about 75–85% of the bootstrap trees (Gupta, 1995; Viale and Arakaki, 1994), indicating that this affiliation is reasonably robust. Viale and Arakaki (1994) have noted several other characteristics of the *Ehrlichia* and *Rickettsia* groups of species which tend to support this. These observations indicate that the last common eubacterial ancestor from which the mitochondrion originated was related to the above subgroup (composed of *Ehrlichia/Rickettsia* species) of purple bacteria. The phylogenetic tree based on cpn10 sequences (Fig. 6) independently supports this inference. A similar inference concerning the origin of mitochondria from α-purple bacteria has been reached in earlier studies based on the size characteristics of cytochrome-*c* sequences (Schwartz and Dayhoff, 1978; Dickerson, 1980) and phylogenies based on SSU rRNA and hsp70 sequences (Gray and Doolittle, 1982; Yang *et al.,* 1985; Gray, 1992; Gupta and Golding, 1993; Falah and Gupta, 1994). The detailed phylogenetic analyses of hsp70 sequences also point to a monophyletic origin of mitochondria (Falah and Gupta, 1994)—an inference that is supported by the cpn60 sequence data.

2. Chloroplast Origin

In the cpn60 tree (Fig. 5) the branching of chloroplast homologs with the cyanobacterial group supports the inference from earlier studies regarding the endosymbiotic origin of chloroplasts from this group of bacteria (see Margulis, 1970; Schwartz and Dayhoff, 1978; Gray and Doolittle, 1982; Taylor, 1987; Gupta *et al.,* 1989; Gray, 1992; Gupta and Golding, 1993). A specific relationship between plant chloroplasts and cyanobacterial homologs is also indicated by other observations: (*a*) In Table II, the maximum amino acid identity of chloroplast homologs is observed with the cyanobacterial group of species; (*b*) the cpn60 homologs from both these groups share a number of sequence characteristics not present in other sequences (see Fig. 1). The tree based on cpn10

sequences is not reliable in this regard since the only sequence information currently available is for a single chloroplast cpn10 homolog, which has undergone gene duplication and is found as a larger M_r protein (Bertsch *et al.*, 1992). The branching of both the α and β subunits of chloroplast cpn60 sequences within the cyanobacterial/chloroplast cluster suggests that both of these homologs evolved by a gene duplication event that took place in the common ancestor to chloroplast cpn60. The identification of two different cpn60 homologs in a cyanobacterial species supports this inference (Lehel *et al.*, 1993).

III. *t*-COMPLEX POLYPEPTIDE 1 GENE FAMILY

A. Sequence Similarity between Chaperonin 60 and *t*-Complex Polypeptide 1 Chaperonins

The first indication that TCP-1 may be a cytosolic chaperonin was provided by the sequence similarity searches that indicated significant sequence similarity between the cpn60 family of proteins and the mouse TCP-1 protein (Ellis, 1990; Gupta, 1990). The significance of sequence similarity between these sequences has been examined using a number of different programs, namely, RDF2 program of the FASTA PIR package (Gupta, 1990), FASTP (Ellis, 1990), and Relate (Trent *et al.*, 1992), and all except the last program indicate significant similarity between these protein families. In Table III, the alignment scores obtained by the Needleman and Wunsch (1970) method between representative TCP-1 and cpn60 sequences from the major groups of species are presented. In this program, which is part of the PCGENE package, alignment score values >3 are considered significant ($p < 0.0001$) and those above 6 are regarded highly significant, indicative of homology due to common ancestory (see Needleman and Wunsch, 1990; Pearson, 1990). As seen from Table III, all of the alignment scores between these two groups of proteins are >3, and most are >6, supporting the inference that these proteins are structurally related and shared a common ancestor. The complete alignment of TCP-1 sequences with the cpn60 sequences has been previously published (Ahmad *et al.*, 1990; Gupta, 1990) and, therefore, is not described here. The inference from sequence similarity studies

TABLE III

Amino Acid Similarity between cpn60 and TCP-1 Sequences[a]

TCP-1 → Cpn60 ↓	Human	*D. melanogaster*	*Sa. cerevisiae*	*A. thaliana*	*Su. shibatae*	*A. sativa*	Mouse
Ec. coli	37.0 (6.7)	33.4 (5.2)	33.8 (5.5)	36.5 (6.1)	34.4 (9.0)	35.5 (3.3)	35.7 (9.5)
S. cerevisiae (m)	34.7 (6.0)	33.9 (5.4)	35.1 (7.1)	35.9 (6.1)	32.4 (8.1)	35.7 (6.7)	34.4 (7.3)
Human (m)	34.5 (5.2)	35.6 (5.9)	35.0 (6.6)	39.8 (5.6)	37.3 (9.4)	37.2 (6.7)	34.8 (7.5)
Wheat (c)	36.1 (7.1)	35.0 (4.9)	35.9 (5.5)	35.7 (6.5)	36.1 (6.4)	33.3 (2.0)	34.0 (6.4)
Cl. perfringens	38.0 (10.2)	36.7 (7.3)	39.1 (5.9)	40.0 (7.1)	38.9 (10.1)	37.4 (5.4)	36.9 (9.2)
M. leprae	37.9 (6.5)	36.9 (6.2)	37.0 (5.4)	38.2 (5.6)	37.2 (7.5)	39.2 (3.3)	38.2 (8.8)

[a] The amino acid similarity (i.e., percentage identical plus conserved residues) between the pairs of sequences was determined using the PALIGN program as described in Table I. The numbers in parentheses indicate the alignment scores for the pairs of sequences computed as described previously (Gupta, 1990; Gupta and Singh, 1992) according to Pearson (1990) for 50 random runs. The alignment scores above 3 indicate significant similarity between the pairs of sequences. Abbreviations as in Table I.

that the TCP-1 family of proteins may be carrying out a similar function in eukaryotic cell cytosol as cpn60 within the organelles has now been amply confirmed by different studies (see Ellis, 1992; Gao *et al.*, 1992; Horwich and Willison, 1993; Yaffe *et al.*, 1992; see Chapter 5).

B. Identification of Two Subfamilies of *t*-Complex Polypeptide 1 Sequences

To understand the evolutionary relationships among known TCP-1 homologs, a global alignment of available TCP-1 sequences was carried out. The alignment indicated interesting differences between various sequences, and excerpts from it for representative TCP-1 sequences are shown in Fig. 7. The top three lines, which include sequences from animal (mouse), plant (*Arabidopsis thaliana*), and yeast, contain numerous common amino acids (indicated by asterisks) that are not present in other TCP-1 sequences. The sequences from other animal species (namely, human, rat, Chinese hamster, *Drosophila*) are quite similar to these (not shown). In contrast to these sequences, the sequences in the last three lines, which include a TCP homolog from archaebacteria (TF55), a mouse homolog referred to as matricin, and a homolog from a plant species (*Avena sativa*), contain numerous other shared sequence characteristics (shaded regions) that are not present in the former group (Fig. 7). It should be noted that the differences noted above are not species-related, as sequences from the same (namely, mouse) or closely related plant species (namely, *A. sativa* and *A. thaliana*) are present in both groups.

The evolutionary relationships of TCP-1 sequences were examined by constructing a neighbor-joining tree based on these sequences (Fig. 8A). In this figure, distance from the nodes along the horizontal axis indicates the distance between the species. The group I sequences form a coherent closely related group, wherein all of the animal TCP-1 sequences form a separate clade. In contrast, the group II TCP-1 sequences show completely distinct branching, and the observed large distance between these two groups of sequences is suggestive of their distant evolutionary relationship. More detailed investigation of the sequence data was carried out by means of the bootstrap methods (Felsenstein, 1985; Saitou and Nei, 1987). A neighbor-joining bootstrap consensus tree based on these sequences is shown in Fig. 8B. To obtain this tree the sequences were bootstrapped 100 times, for each a neighbor-joining tree was found,

Fig. 7. Identification of two TCP-1 subfamilies based on sequence characteristics. An alignment of TCP-1 sequences from representative homologs is shown. The asterisks mark the residues that are either identical or conserved in the top three lines. Other members of this subfamily (namely, from rat, human, Chinese hamster, *D. melanogaster*) also contain similar residues. The shaded residues in the bottom three lines identify residues that are either identical or conserved in the second group of TCP-1 homologs, but which differ from that in the first group. The dashes (-) indicate residues identical to those in the top line. The numbers at the top refer to those in mouse TCP-1 sequence.

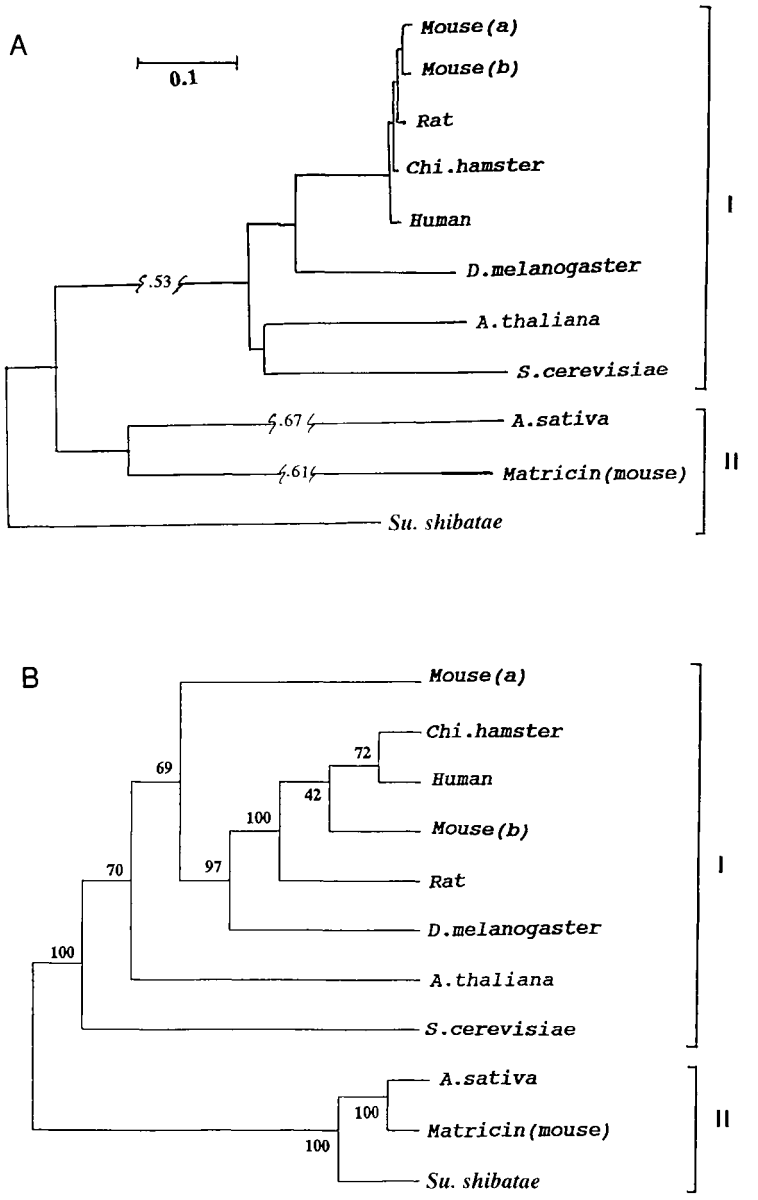


Fig. 8. Phylogenetic trees based on TCP-1 sequences. (A) The neighbor-joining distance matrix tree was constructed as described in Fig. 5. The distance along the horizontal axis indicates the distances between the species. The bar shown corresponds to 0.1 substitution per amino acid. I and II denote the two subfamilies. (B) A neighbor-joining boot-strap consensus tree based on TCP-1 sequences. The numbers at the forks indicate the number of times the species (or homologs), which are to the right of that fork, group together out of 100 boot-strapped trees.

and finally a consensus tree based on these analyses was obtained. The numbers at the fork indicate the number of times the sequences that are to the right of that fork group together out of 100 trees. There is a clear distinction (100 out of 100) between the group I and II sequences (Fig. 8B), which strongly supports their division into two groups. The above analyses provide strong evidence that the TCP-1 sequences in the two groups (each of which contain members from the plant and animal species) diverged from each other very early and have since then evolved independently. To account for the presence of two different groups of TCP-1 sequences in eukaryotic (e.g., plant and animal) species, the simplest explanation would be that they are derived from different TCP-1 homologs present in the archaebacterial parent. The available evidence indicates that similar to the eukaryotic TCP-1 complex, the archaebacterial TCP-1 complex is also a heterooligomer made up of several different subunits (see Horwich and Willison, 1993, and Chapter 5). It is likely that the sequence of one of these subunits will bear closer similarity to the group I TCP-1 sequences, and possess many of the sequence characteristics of this group.

After the present analysis was completed, Kubota *et al.* (1994; see Chapter 5) reported the complete sequences of six additional TCP-1 related proteins from mouse cells. The newly described TCP-1 homologs (designated CCT_{β}, CCT_{γ}, CCT_{δ}, $CCT\varepsilon$, CCT_{ζ}, and CCT_{η}; the original TCP-1 referred to as $TCP\alpha$) are shown to be highly divergent, exhibiting between 25 and 39% amino acid identity to the other homologs. On the basis of sequence comparisons, the mouse matricin protein considered here (Joly *et al.*, 1994) corresponds to the CCT_{γ} subunit of the TCP complex. In view of the highly divergent nature of the TCP-1 subunits (Kubota *et al.*, 1994), it is likely that more than two subfamilies of protein sequences (as described here) will be found within this group.

IV. EVOLUTIONARY RELATIONSHIP OF *t*-COMPLEX POLYPEPTIDE 1 AND CHAPERONIN 60 GENE FAMILIES AND ORIGIN OF EUKARYOTIC CELLS

As indicated earlier, cpn60 homologs have been found only in bacterial species and in eukaryotic cell organelles that have originated from eubacteria via endosymbiosis. In contrast, TCP-1-related chaperonins are present in archaebacteria and eukaryotic cell cytosol. To understand the

species distribution of these two families of proteins and their relationship to each other, it is important to consider the evolutionary relationship among archaebacteria, eubacteria, and eukaryotic species.

Phylogenetic analysis based on SSU rRNA and a few other gene sequences has led to the concept of the division of all extant organisms into three primary urkingdoms or domains comprising archaebacteria, eubacteria, and eukaryotic species (Woese *et al.,* 1990). Each of these species domains is proposed to be monophyletic and completely distinct from the others. Although this concept has gained wide acceptance, sequence data for a number of highly conserved proteins, including hsp70, glutamine synthetase (glutamate–ammonia ligase), glutamate dehydrogenase, etc., do not support this view (Benachenhou-Lahfa *et al.,* 1993; Gupta and Golding, 1993; Tiboni *et al.,* 1993; Golding and Gupta, 1995). These latter studies indicate a closer relationship between archaebacteria and gram-positive bacteria, and this group is postulated to be the ancestral lineage (Gupta and Singh, 1992; Gupta and Golding, 1993). For the origin of eukaryotic cells, phylogenetic studies based on different gene sequences have led to two main types of model. The first type of model assumes progressive evolution of the eukaryotic cell from a prokaryotic ancestor (see Fig. 9A and Woese *et al.,* 1990). On the basis of the sequence data for EF-1α, EF-2, F- and V-type ATPases, and RNA polymerase II and III subunits, where greater similarity is observed between the archaebacterial and the eukaryotic homologs, it has been postulated that the eukaryotic cells have evolved from an archaebacterial ancestor (Gogarten *et al.,* 1989; Iwabe *et al.,* 1989; Pühler *et al.,* 1989), most likely belonging to the thermoacidophilic group of archaebacteria, referred to as eocytes (Rivera and Lake, 1992). In contrast to the above genes, the sequence features and phylogenies based on a number of other gene sequences, the most striking of which is hsp70, strongly indicate that the eukaryotic homologs have evolved from a gram-negative eubacterium (Gupta *et al.,* 1994; Gupta and Singh, 1994). In the case of hsp70, a relatively conserved insert of 23–27 residues is present in the same position in all hsp70s from eukaryotic species and gram-negative eubacteria, but is not found in any of the hsp70s from gram-positive bacteria or archaebacteria, pointing to the evolution of eukaryotic cells from a gram-negative bacterium. To account for the conflicting phylogenies obtained using different gene/protein sequences, a second type of model that postulates a chimeric origin of the eukaryotic cell nucleus has been proposed (Zillig, 1991; Gupta and Golding, 1993; Gupta and Singh, 1994; Irwin, 1994; Golding and Gupta, 1995).

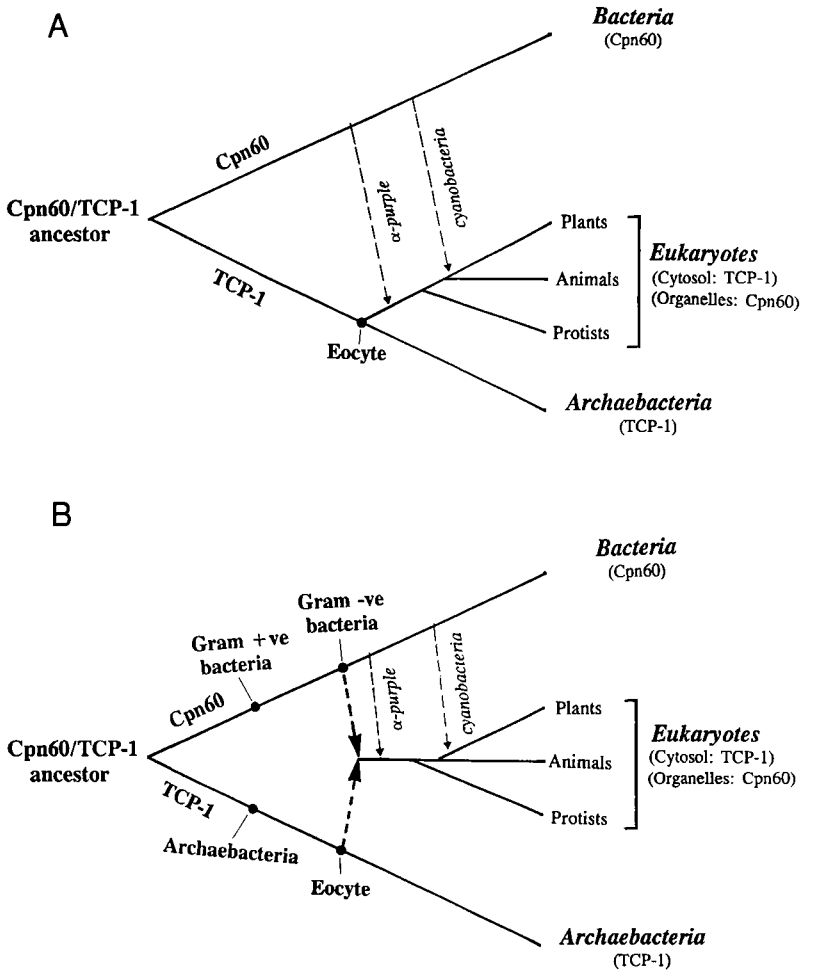


Fig. 9. Evolutionary relationship between cpn60 and TCP-1 sequences. Two alternative models based on current views regarding the origin of eukaryotic cells are presented. Both models assume that the cpn60 and TCP-1 proteins evolved from an ancestral protein that was present in the universal ancestor of all organisms. Changes in this protein at an early stage in evolution led to the evolution of cpn60 and TCP-1. In model (A), referred to as archaebacterial model (Woese *et al.*, 1990), the eukaryotic cell nucleus is presumed to have evolved from an archaebacterial (eocyte group) ancestor, which contained TCP-1 homolog(s). In the chimeric model (B), early stages of divergence from the universal ancestor gave rise to archaebacteria which contained TCP-1 and gram-positive bacteria containing cpn60. From these two primary lineages other groups of archaebacteria and eubacteria evolved. At a later time, a primary fusion event between an eocyte group of archaebacteria and a gram-negative bacterium gave rise to a proeukaryotic cell. During its establishment into a eukaryotic cell, this cell retained different genes from the two fusion partners. From the cpn60/TCP-1 group, only TCP-1 homologs were retained. The thin dotted arrows indicate the endosymbiotic capture of an α-purple bacterium and a cyanobacterium resulting in the origin of mitochondria and chloroplasts, respectively.

According to the chimeric model, the eukaryotic cell nucleus is a chimera that originated by the primary fusion between a gram-negative eubacterium and an eocyte group of archaebacteria (see Fig. 9B). This primary fusion event could be akin to endosymbiosis, where one of the fusion partners served as the host (e.g., gram-negative eubacterium) and the other, which was engulfed, was the guest (namely, eocyte), followed by mixing of their genes in a common pool surrounded by the membrane of the guest species (Gupta *et al.,* 1994; Lake and Rivera, 1994). During the early stages of the evolution of this proeukaryotic cell, it is postulated that an assortment or selection of genes from the two fusion partners took place, and the genes that were not selected were either lost or so drastically modified that no similarity to the original gene(s) is currently observed. As a result, all extant eukaryotic species contain genes that are descendant from either archaebacteria or gram-negative bacteria but not from both species (see Gupta and Golding, 1993; Gupta and Singh, 1994).

The sequence data for the cpn60/TCP-1 family of proteins are consistent with and could be explained by either of these two models. However, of these two, we prefer the latter model because it explains all of the other gene phylogenies as well (see Gupta and Singh, 1994; Irwin, 1994; Golding and Gupta, 1995). To understand the relationship between these two families of proteins, one needs to assume the existence of an ancestral protein in the universal ancestor of all organisms from which both cpn60 and TCP-1 evolved (Gupta, 1990). On the basis of the hsp70 sequence data, it is suggested that this universal ancestor was related to the lineage comprising archaebacteria and gram-positive bacteria (Gupta and Golding, 1993). Since many of the known archaebacteria grow optimally at high temperatures (>100°C) in anaerobic environments and utilize energy from the reduction of sulfur and hydrogen, conditions that are probably akin to what may have existed in the primitive earth's environment (Pace, 1991), it is possible that the universal ancestor was more closely related to this group.

During the early period of evolution (which could have lasted up to 1 billion years), this universal ancestor gave rise to two main groups of organisms, archaebacteria, and gram-positive bacteria. Due to the rapidly changing environment, extensive divergence took place by this time for some genes/proteins for which functional constraints were not rigid (e.g., EF-1α, EF-2, cpn60), such that the ancestral cpn60/TCP-1 related protein gave rise to the TCP-1 family of proteins in archaebacteria and cpn60 derivatives in gram-positive bacteria. In archaebacteria, due to altered

functional requirements, duplication of the TCP-1 gene probably also took place at an early stage, giving rise to multiple isoforms. In contrast to the above genes, for a number of other highly conserved and constrained functions (namely, hsp70, glutamate dehydrogenase, glutamine synthetase), the close relationship between these two groups of organisms was retained and can still be observed (Benachenhou-Lahfa *et al.*, 1993; Gupta and Golding, 1993; Tiboni *et al.*, 1993; Brown *et al.*, 1994; Golding and Gupta, 1995). In the period that followed (during which the environment changed from predominantly anaerobic to aerobic) other groups of archaebacteria and eubacteria (including the gram-negative group) are postulated to have evolved from these ancestral groups. The evolution of archaebacterial and eubacterial groups was followed by the postulated fusion between an archaebacterium and a gram-negative eubacterium that gave rise to the proeukaryotic cell. During the assortment of genes from the two fusion parents, only the TCP-1-related genes but not cpn60-related genes were retained in the eukaryotic nucleus. [This is in contrast to hsp70 where the homolog from the gram-negative bacteria was selected (Gupta *et al.*, 1994).] It is not clear at present at what stage the cpn10 chaperonin evolved. If a related protein was present in the universal ancestor before the divergence of archaebacteria and bacteria, then it is expected that a cpn10 homolog should also be found in archaebacteria and eukaryotic cell cytosol, unless its function has become redundant in the TCP-1 complex. On the other hand, if it evolved at a later stage (in concert with the evolution of cpn60 in the bacterial lineage) then no cpn60 homolog will be found in archaebacteria or eukaryotic cell cytosol. At later stages in the evolution of eukaryotic cells, other endosymbiotic events that gave rise to mitochondria (from an α-purple gram-negative bacterium) and chloroplasts (from cyanobacteria) took place, as depicted in Fig. 9B.

The model proposed here explains all of the characteristics of the cpn60/TCP-1 gene families as well as other gene/protein phylogenies. In addition, it makes a number of predictions that are experimentally testable. These include the following: (i) The model proposes that TCP-1 homologs should not be found in bacteria nor the cpn60 homologs in the eukaryotic cell cytosol (i.e., of nuclear origin). (ii) It predicts that different TCP-1 subunits in eukaryotic cells originated from different genes present in the archaebacteria. Therefore one should expect to find multiple TCP-1 genes in archaebacteria and they should bear closer relationship to different TCP-1 subfamilies in eukaryotic homologs. (iii) The model also suggests that for some of the highly conserved

proteins that are found in all eukaryotic species (e.g., actin, tubulin) a related protein may also be found in prokaryotic species (namely, either archaebacteria or gram-negative bacteria). (iv) The ancient nature of gram-positive bacteria also provides a clue to why the homologs from some of the gram-positive bacteria (mycobacteria) are highly antigenic in mammalian systems (Kaufmann, 1990; Young *et al.*, 1990; see Chapter 10). In comparison to the eukaryotic cells, they represent the most divergent group of species and it is predicted that some of the other members of this group will also be highly antigenic. (v) The phylogeny based on cpn10 and cpn60 sequences strongly supports the origin of mitochrondria from the α-purple subdivision of gram-negative bacteria, and of chloroplasts from the cyanobacteria—an inference in accordance with the earlier studies (see Yang *et al.*, 1985; Gray, 1992; Falah and Gupta, 1994). The analysis based on cpn60 sequences further suggests that the eubacterial ancestor from which mitochondrion evolved was evolutionarily closely related to the species belonging to the *Ehrlichia/Rickettsia* group, which live symbiotically within eukaryotic cells. The analysis presented here thus provides valuable insight into the origin of eukaryotic cells and organelles and it enables us to understand the relationship between different members of the chaperonin family of proteins from a broader perspective.

ACKNOWLEDGMENT

The research work from the author's laboratory has been supported by a research grant from the Medical Research Council of Canada.

REFERENCES

Ahmad, S., and Gupta, R. S. (1990). Cloning of a Chinese hamster protein homologous to the mouse *t*-complex protein TCP-1: Structural similarity to the ubiquitous "chaperonin" family of heat-shock proteins. *Biochim. Biophys. Acta* **1087,** 253–255.

Altschul, S. F. (1991). Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* **219,** 555–565.

Arakere, G., Kessel, M., Nguyen, N., and Frasch, C. E. (1993). Characterization of a stress protein from group B *Neisseria meningitidis*. *J. Bacteriol.* **175,** 3664–3668.

Ballard, S. A., Segers, R. P., Bleumink-Pluym, N., Fyfe, J., Faine, S., and Adler, B. (1993). Molecular analysis of the hsp (groE) operon of *Leptospira interrogans* serovar copenhageni. *Mol. Microbiol.* **8,** 739–751.

Banfalvi, Z., Kondorosi, E., and Kondorosi, A. (1985). *Rhizobium meliloti* carries two megaplasmids. *Plasmid* **13,** 129–138.

Benachenhou-Lahfa, N., Forterre, P., and Lebeday, B. (1993). Evolution of glutamate dehydrogenase genes: Evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. *J. Mol. Evol.* **36,** 335–346.

Bertsch, U., Soll, S., Seetharam, R., and Viitanen, P. V. (1992). Identification, characterization, and DNA sequence of a functional "double" groES-like chaperonin from chloroplasts of higher plants. *Proc. Natl. Acad. Sci. USA* **89,** 8696–8700.

Brown, J. R., Masuchi, Y., Robb, F. T., and Doolittle, W. F. (1994). Evolutionary relationships of bacterial and archael glutamine synthetase genes. *J. Mol. Evol.* **38,** 566–576.

Brudzynski, K., Martinez, V., and Gupta, R. S. (1992). Secretory granule autoantigen in insulin-dependent diabetes mellitus is related to 62 kDa heat-shock protein (hsp60). *J. Autoimmun.* **5,** 453–463.

Cedergren, R., Gray, M. W., Abel, Y., and Sankoff, D. (1988). The evolutionary relationships among known life forms. *J. Mol. Biol.* **28,** 98–112.

Chanda, P. K., Ono, M., Kuwano, M., and Kung, H.-F. (1985). Cloning, sequence analysis, and expression of alteration of the mRNA stability gene (*ams*$^+$) of *Escherichia coli*. *J. Bacteriol.* **161,** 446–449.

Chandrashekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* **261,** 12414–12419.

Chitnis, P. R., and Nelson, B. (1991). Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC 5803. *J. Biol. Chem.* **266,** 58–65.

Claverie-Martin, F., Diaz-Torres, M. R., Yancey, S. D., and Kushner, S. R. (1989). Cloning of the altered mRNA stability (*ams*) gene of *Escherichia coli* K-12. *J. Bacteriol.* **171,** 5479–5486.

Cohen, I. R. (1991). Autoimmunity to chaperonins in the pathogenesis of arthritis and diabetes. *Annu. Rev. Immunol.* **9,** 567–589.

Cozens, A. L., and Walker, J. E. (1987). The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301: Support for an endosymbiotic origin of chloroplasts. *J. Mol. Biol.* **194,** 359–383.

Dickerson, R. E. (1980). Evolution and gene transfer in purple photosynthetic bacteria. *Nature* (*London*) **283,** 210–212.

Dudani, A. K., and Gupta, R. S. (1989). Immunological characterization of a human homolog of the 65-kilodalton mycobacterial antigen. *Infect. Immun.* **57,** 2786–2793.

Ehmann, B., Krenz, M., Mummert, E., and Schaefer, E. (1993). Two TCP-1 related but highly divergent gene families exist in oat encoding proteins of assumed chaperone function. *FEBS Lett.* **336,** 313–316.

Ellis, R. J. (1990). Molecular chaperones: The plant connection. *Science* **250,** 954–959.

Ellis, R. J. (1992). Cytosolic chaperonin confirmed. *Nature* (*London*) **358,** 191.

Falah, M., and Gupta, R. S. (1994). Cloning of the hsp70 (*dnaK*) genes from *Rhizobium meliloti* and *Psuedomonas cepacia:* phylogenetic analyses of mitochondrial origin based on a highly conserved protein sequence. *J. Bacteriol.* **176,** 7748–7753.

Felsenstein, J. (1982). Numerical methods for inferring evolutionary trees. *Q. Rev. Biol.* **57,** 379–404.

Felsenstein, J. (1985). Confidence limits in phylogenies: An approach using the bootstrap. *Evolution* **39,** 783–791.

Felsenstein, J. (1991). "PHYLIP manual, ver. 3.3. Herbarium." University of California, Berkeley.

Ferreyra, R. Soncini, F., and Viale, A. M. (1993). Cloning, characterization and fuctional expression in *Escherichia coli* of chaperonin (*gro*ESL) genes from the phototropic sulfur bacterium *Chromatium vinosum. J. Bacteriol.* **175,** 1514–1523.

Fischer, H. M. (1994). Genetic regulation of nitrogen fixation in Rhizobia. *Microbiol. Rev.* **58,** 352–386.

Fisher, H. M., Babst, M., Kaspar, T., Acuna, G., Arigoni, F., and Hennecke, H. (1993). One member of a gro-ESL-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. *EMBO J.* **12,** 2901–2912.

Fitch, W. M., and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* **155,** 279–284.

Fox, G. E., Stackerbrandt, E., Hespell, R. B., Gibson, H. J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N., and Woese, C. R. (1980). The phylogeny of prokaryotes. *Science* **209,** 457–463.

Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H., and Cowan, N. J. (1992). A cytoplasmic chaperonin that catalyzes β-actin folding. *Cell* **69,** 1043–1050.

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, I., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., Oshima, T., Konishi, J., Denda, K., and Yoshida, M. (1989). Evolution of the vacuolar H^+-ATPases: Implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA* **89,** 6661–6665.

Golding, G. B., and Gupta, R. S. (1995). Protein-based phylogenics support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* **12,** 1–6.

Govezensky, D., Greener, T., Segal, G., and Zamir, A. (1991). Involvement of GroEL in *nif* gene regulation and nitrogenase assembly. *J. Bacteriol.* **173,** 6339–6346.

Gray, M. W. (1992). The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141,** 233–357.

Gray, M. W., and Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiol Rev.* **46,** 1–42.

Gupta, R. S. (1990). Sequence and structural homology between a mouse *t*-complex protein TCP-1 and the "chaperonin" family of bacterial (GroEL, 60–65 kDa heat shock antigen) and eukaryotic proteins. *Biochem. Int.* **20,** 833–841.

Gupta, R. S. (1994). Presence of a GroES (cpn10)-related sequence motif in the GroEL (cpn60) chaperonins. *Biochem. Mol. Biol. Int.* **33,** 591–595.

Gupta, R. S. (1995). Evolution of the chaperonin families (HSP60, HSP10 and TCP-1) of proteins and the origin of eukaryotic cells. *Mol. Microbiol.* **15,** 1–11.

Gupta, R. S., Aitken, K., Falah, M., and Singh, B. (1994). Cloning of *Giardia lamblia* hsp70 homologs: Implications regarding origin of eukaryotic cells and of endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA* **91,** 2895–2899.

Gupta, R. S., and Golding, G. B. (1993). Evolution of the HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. *J. Mol. Evol.* **37,** 573–582.

Gupta, R. S., and Singh, B. (1992). Cloning of the HSP70 gene from *Halobacterium marismortui:* Relatedness of archaebacterial HSP70 to its eubacterial homologs and a model for the evolution of the HSP70 gene. *J. Bacteriol.* **174,** 4594–4605.

Gupta, R. S., and Singh, B. (1994). Phylogenetic analysis of 70 kDa heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Current Biol.* **4,** 1104–1114.

Gupta, R. S., Picketts, D. J., and Ahmad, S. (1989). A novel ubiquitous protein 'chaperonin' supports the endosymbiotic origin of mitochondrion and plant chloroplast. *Biochem. Biophys. Res. Commun.* **163,** 780–787.

Hartl, R., and Neupert, W. (1990). Protein sorting to mitochondria: Evolutionary conservations of folding and assembly. *Science* **247,** 930–934.

Hartman, D. J., Hoogenraad, N. J., Condron, R., and Hoj, P. B. (1992). Identification of a mammalian 10-kDa heat shock protein, a mitochondrial chaperonin 10 homologue essential for assisted folding of trimeric ornithine transcarbamoylase *in vitro. Proc. Natl. Acad. Sci. USA* **89,** 3394–3398.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* (*London*) **333,** 330–334.

Hendrick, J. P., and Hartl, F.-U. (1993). Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62,** 349–384.

Horwich, A. L., and Willison, K. R. (1993). Protein folding in the cell: Functions of two families of molecular chaperone, hsp60 and TF55-TCP1. *Phil. Trans. R. Soc. Lond. B:* **339,** 313–326.

Houston, L. S., Cook, R. G., and Norris, S. (1990). Isolation and characterization of a *Treponema pallidum* major 60-kilodalton protein resembling the groEL protein of *Escherichia coli. J. Bacteriol.* **172,** 2862–2870.

Irwin, D. M. (1994). Who are the parents of eukaryotes? *Current Biol.* **4,** 1115–1117.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86,** 9355–9359.

Jindal, S., Dudani, A. K., Singh, B., Harley, C. B., and Gupta, R. S. (1989). Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. *Mol. Cell. Biol.* **9,** 2279–2283.

Joly, E. C., Sevigny, G., Todorov, I. T., and Bibor-Hardy, V. (1994). cDNA encoding a novel TCP-1 related protein. *Biochem. Biophys. Acta* **1217,** 224–226.

Jones, D. B., Coulson, A. F., and Duff, G. W. (1993). Sequence homologies between hsp60 and autoantigens. *Immunol. Today* **14,** 115–118.

Jordan, D. C. (1984). *Rhizobiaceae. In* "Bergey's Manual of Systematic Bacteriology" (N. R. Kreig and J. G. Holt, eds.), Vol. 1, pp. 234–255. Williams & Wilkins, London.

Kaufmann, S. H. E. (1990). Heat shock proteins and the immune response. *Immunol. Today* **1,** 129–136.

Kikuta, L. C., Puolakkainen, M., Kuo, C. C., and Campbell, L. A. (1991). Isolation and sequence analysis of the *Chlamydiae pneumoniae* GroE operon. *Infect. Immun.* **59,** 4665–4669.

Kim, S. G., and Batt, C. A. (1993). Cloning and sequencing of the *Lactococcus lactis* subsp. lactis groESL operon. *Gene* **127,** 121–126.

Kirchhoff, C., and Willison, K. R. (1990). Nucleotide and amino-acid sequence of human testis-derived TCP1. *Nucleic Acids Res.* **18,** 4247.

Kong, T. H., Coates, A. R., Butcher, P. D., Hickman, C. J., and Shinnick, T. M. (1993). *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc. Natl. Acad. Sci. USA* **90,** 2608–2612.

Kubota, H., Morita, T., Nagata, T., Takemoto, Y., Nozaki, M., Gachelin, G., and Matsushiro, A. (1991). Nucleotide sequence of mouse Tcp-1a cDNA. *Gene* **105,** 269–273.

Kubota, H., Hynes, G., Carne, A., Ashworth, A., and Willison, K. (1994). Identification of six Tcp-1-related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr. Biol.* **4,** 89–99.

Lake, J. A., and Rivera, M. C. (1994). Was the nucleus the first endosymbiont? *Proc. Natl. Acad. Sci. USA* **91,** 2880–2881.

Lehel, C., Los, D., Wada, H., Gyorgyei, J., Horvath, I., Kovacs, E., and Murata, N. (1993). A second groEL-like gene, organized in a groESL operon, is present in the genome of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **268,** 1799–1804.

Li, M., and Wong, S. L. (1992). Cloning and characterization of the groESL operon from *Bacillus subtilis*. *J. Bacteriol.* **174,** 3981–3992.

Lin, J., Adams, L. G., and Ficht, T. A. (1992). Characterization of the heat shock response in *Brucella abortus* and isolation of the genes encoding the GroE heat shock proteins. *Infect. Immun.* **60,** 2425–2431.

Macchia, G., Massone, A., Burroni, D., Covacci, A., Censini, S., and Rappuoli, R. (1993). The Hsp60 protein of *Helicobacter pylori:* Structure and immune response in patients with gastroduodenal diseases. *Mol. Microbiol.* **9,** 645–652.

Maid, U., Steinmuller, R., and Zetsche, K. (1992). Structure and expression of a plastid-encoded groEL homologous heat-shock gene in a thermophilic unicellular red alga. *Curr. Genet.* **21,** 521–525.

Margulis, L. (1970). "Origin of Eukaryotic Cells." Yale University Press, New Haven, CT.

Martel, R., Cloney, L. P., Pelcher, L. E., and Hemmingsen, S. M. (1990). Unique composition of plastid chaperonin-60: α and β polypeptide-encoding genes are highly divergent. *Gene* **94,** 181–187.

Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C. J. (1991). Characterization of the *groEL*-like genes in *Streptomyces albus*. *J. Bacteriol.* **173,** 7382–7386.

McLennan, N. F., Girshovich, A. S., Lissin, N. M., Charter, Y., and Masters, M. (1993). The strongly conserved carboxyl-terminus glycine-methionine motif of the *Escherichia coli* GroEL chaperonin is dispensable. *Mol. Microbiol.* **7,** 49–58.

McLennan, N. F., McAteer, S., and Masters, M. (1994). The tail of a chaperonin: The C-terminal region of *Escherichia coli* GroEL protein. *Mol. Microbiol.* **14,** 309–321.

Mehra, V., Sweetser, D., and Young, R. A. (1986). Efficient mapping of protein antigenic determinants. *Proc. Natl. Acad. Sci. USA* **83,** 7013–7017.

Miller, S. G., and Leclerc, R. F. (1990). Identification and characterization of a testis-specific isoform of a chaperonin in a moth, *Heliothis virescens*. *J. Mol. Biol.* **214,** 407–422.

Mori, M., Murata, K., Kubota, H., Yamamoto, A., Matsushiro, A., and Morita, T. (1992). Cloning of a cDNA encoding the Tcp-1 (*t*-complex polypeptide 1) homologue of *Arabidopsis thaliana*. *Gene* **122,** 381–382.

Morrison, R. P., Su, H., Lyng, K., and Yuan, Y. (1990). The *Chlamydia trachomatis hyp* operon is homologous to the *groE* stress response operon of *Escherichia coli*. *Infect. Immun.* **58,** 2701–2705.

Moulder, J. W. (1985). Comparative biology of intracellular parasitism. *Microbiol. Rev.* **49,** 298–337.

Myers, E. W., and Miller, W. (1988). Optimal alignments in linear space. *Comput. Appl. Biosci.* **4,** 11–17.

Narberhaus, F., and Bahl, H. (1992). Cloning, sequencing, and molecular analysis of the groESL operon of *Clostridium acetobutylicum. J. Bacteriol.* **174,** 3282–3289.

Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–453.

Ohta, T., Honda, K., Kuroda, M., Saito, K., and Hayashi, H. (1993). Molecular characterization of the gene operon of heat shock proteins HSP60 and HSP10 in methicillin-resistant *Staphylococcus aureus. Biochem. Biophys. Res. Commun.* **193,** 730–737.

Pace, N. R. (1991). Origin of life—Facing up to the physical setting. *Cell* **65,** 531–533.

Parsons, L. M., Waring, A. L., and Shayegani, M. (1992). Molecular analysis of the *Haemophilus ducreyi* groE heat shock operon. *Infect. Immun.* **60,** 4111–4118.

Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *In* "Methods in Enzymology," Vol. 183, pp. 62–98. Academic Press, San Diego.

Picketts, D. J., Mayanil, C. S., and Gupta, R. S. (1989). Molecular cloning of a Chinese hamster mitochondrial protein related to the "chaperonin" family of bacterial and plant proteins. *J. Biol. Chem.* **264,** 12001–12008.

Pilkington, S. J., and Walker, J. E. (1993). Complementary DNA sequence of bovine cpn10 (HSP10), a chaperone protein from mitochondria. *DNA Seq.* **3,** 291–295.

Prasad, T. K., and Stewart, C. R. (1992). cDNA clones encoding *Arabidopsis thaliana* and *Zea mays* mitochondrial chaperonin HSP60 and gene expression during seed germination and heat shock. *Plant Mol. Biol.* **18,** 873–885.

Pühler, G., Leffersy, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. A., and Zillig, W. (1989). Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Natl. Acad. Sci. USA* **86,** 4569–4573.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature (London)* **337,** 655–659.

Rinke de Wit, T. F., Bekelie, S., Osland, A., Miko, T. L., Hermans, P. W., vanSoolingen, D., Drijfhout, J. W., Schoningh, R., Janson, A. A., and Thole, J. E. (1992). Mycobacteria contain two groEL genes: The second *Mycobacterium leprae* groEL gene is arranged in an operon with groES. *Mol. Microbiol.* **6,** 1995–2007.

Rivera, M. C., and Lake, J. A. (1992). Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257,** 74–76.

Rusanganwa, E., and Gupta, R. S. (1993). Cloning and characterization of multiple groEL chaperonin-encoding genes in *Rhizobium meliloti. Gene* **126,** 67–75.

Rusanganwa, E., Singh, B., and Gupta, R. S. (1992). Cloning of HSP60 (GroEL) operon from *Clostridium perfringens* using a polymerase chain reaction based approach. *Biochim. Biophys. Acta* **1130,** 90–94.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method of reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

Sampson, J. S., O'Conner, S. P., Holloway, B. P., Plikaytis, B. B., Carlone, G. M., and Mayer, L. W. (1990). Nucleotide Sequence of *htpB,* the *Legionella pneumophila* gene encoding the 58-kilodalton (kDa) common antigen, Formerly designated the 60-kDa common antigen. *Infect. Immun.* **58,** 3154–3157.

Schatz, G. (1987). Signals guiding proteins to their correct locations in mitochondria. *Eur. J. Biochem.* **165,** 1–6.

Schon, U., and Schumann, W. (1993). Molecular cloning, sequencing, and transcriptional analysis of the groESL operon from *Bacillus stearothermophilus. J. Bacteriol.* **175,** 2465–2469.

Schwartz, R. M., and Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts: A perspective is derived from protein and nucleic acid sequence data. *Science* **199,** 395–403.

Segal, G., and Ron, E. Z. (1993). Heat shock transcription of the groESL operon of *Agrobacterium tumefaciens* may involve a hairpin-loop structure. *J. Bacteriol.* **175,** 3083–3088.

Shanafelt, M. C., Hindersson, P., Soderberg, C., Mensi, N., Turck, C. W., Webb, D., Yssel, H., and Peltz, G. (1991). T cell and antibody reactivity with the *Borrelia burgdorferi* 60-kDa heat shock protein in Lyme arthritis. *J. Immunol.* **146,** 3985–3992.

Shinnick, T. M. (1987). The 65-kilodalton antigen of *Mycobacterium tuberculosis. J. Bacteriol.* **169,** 1080–1088.

Singh, B., Patel, H. V., Ridley, R. G., Freeman, K. B., and Gupta, R. S. (1990). Mitochondrial import of the human chaperonin (HSP60) protein. *Biochem. Biophys. Res. Commun.* **169,** 391–396.

Sipos, A., Klocke, M., and Frosch, M. (1991). Cloning and sequencing of the genes coding for the 10- and 60-kDa heat shock proteins from *Pseudomonas aeruginosa* and mapping of a species-specific epitope. *Infect. Immun.* **59,** 3219–3226.

Stover, C. K., Marana, D. P., Dasch, G., and Oaks, E. V. (1990). Molecular cloning and sequence analysis of the Sta58 major antigen gene of *Rickettsia tsutsugamushi:* Sequence homology and antigenic comparison of Sta58 to the 60-kilodalton family of stress proteins. *Infect. Immun.* **58,** 1360–1368.

Sumner, J. W., Sims, K. G., Jones, D. C., and Anderson, B. E. (1993). *Ehrlichia chaffeensis* expresses an immunoreactive protein homologous to the *Escherichia coli* GroEL protein. Infect. Immun. **61,** 3536–3539.

Tamada, H., Ohta, T., Hamamoto, T., Otawara-Hamamoto, Y., Yanagi, M., Hiraiwa, H., Hirata, H., and Kagawa, Y. (1991). Gene structure of heat shock proteins 61 kDa and 12 kDa (thermophilic chaperonins) of thermophilic bacterium PS3. *Biochem. Biophys. Res. Commun.* **179,** 565–571.

Taylor, F. J. R. (1987). An overview of the status of evolutionary cell symbiosis theories. *Ann. N.Y. Acad. Sci.* **503,** 1–16.

Tiboni, O., Cammarano, P., and Sanangelantoni, A. M. (1993). Cloning and sequencing of the gene encoding glutamine synthetase I from the archaeum *Pyrococeus woesi:* Anomalous phylogenies inferred from analysis of archael and bacterial glutamine synthetase I sequences. *J. Bacteriol.* **175,** 2961–2969.

Trent, J. D., Nimmesgern, E., Wall, J. S., Hart, F., and Horwich, A. L. (1992). A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein *t*-complex polypeptide-1. *Nature* (*London*) **354,** 490–493.

Tsugeki, R., Mori, H., and Nishimura, M. (1992). Purification, cDNA cloning and Northern-blot analysis of mitochondrial chaperonin 60 from pumpkin cotyledons. *Eur. J. Biochem.* **209,** 453–458.

Ursic, D., and Culbertson, M. R. (1991). The yeast homolog to mouse Tcp-1 affects microtubule-mediated processes. *Mol. Cell. Biol.* **11,** 2629–2640.

Ursic, D., and Ganetzky, B. (1988). A *Drosophila melanogaster* gene encodes a protein homologous to the mouse *t*-complex polypeptide 1. *Gene* **68,** 267–274.

Venner, T. J., and Gupta, R. S. (1990a). Nucleotide sequence of mouse HSP60 (chaperonin, GroEL homolog) cDNA. *Biochim. Biophys. Acta* **1087,** 336–338.

Venner, T. J., and Gupta, R. S. (1990b). Nucleotide sequence of rat hsp60 (chaperonin, GroEL homolog) cDNA. *Nucleic Acids Res.* **18,** 5309.

Viale, A. M., and Arakaki, A. K. (1994). The chaperone connection to the origins of the eukaryotic organelles. *FEBS Lett.* **341,** 146–151.

Viale, A. M., Arakaki, A. K., Soncini, F. C., and Ferreyra, R. G. (1994). Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int. J. Systemat. Bacteriol.* **44,** 527–533.

Vodkin, M. H., and Williams, J. C. (1988). A heat shock operon in *Coxiella burnetii* produces a major antigen homologous to a protein in both mycobacteria and *Eschericha coli*. *J. Bacteriol.* **170,** 1227–1234.

Wallington, E. J., and Lund, P. A. (1994). *Rhizobium leguminsarum* contains multiple chaperonin (cpn60) genes. *Microbiology* **140,** 113–122.

Webb, R., Reddy, K. J., and Sherman, L. A. (1990). Regulation and sequence of the *Synechococcus* sp. strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. *J. Bacteriol.* **172,** 5079–5088.

Willison, K. R., Dudley, K., and Potter, J. (1986). Molecular cloning and sequence analysis of a haploid expressed gene encoding T-complex polypeptide 1. *Cell* **44,** 727–738.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51,** 221–271.

Woese, C. R. (1991). The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria. *In* "Evolution at the Molecular Level" (R. K. Selander, A. G. Clark, and T. S. Whittmay, eds.), pp. 1–24. Sinauer Associates, Sunderland, MA.

Woese, C. R. Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.

Yaffe, M. B., Farr, G. W., Miklos, D., Horwich, A. L., Sternlicht, M. L., and Sternlicht, H. (1992). TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature (London)* **358,** 245–248.

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., and Woese, C. R. (1985). Mitochondrial origins. *Proc. Natl. Acad. Sci. USA* **82,** 4443–4447.

Young, D. B., Mehlert, A., and Smith, D. F. (1990). Stress proteins and infectious diseases. *In* "Stress Proteins in Biology and Medicine" (R. I. Morimoto, A., Tissieres, and C. Georgopoulos, eds.), pp. 131–165. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Zabaleta, E., Oropeza, A., Jiminez, B., Salerno, G., Crespi, M., and Herrera-Estrella, L. (1992). Isolation and characterization of genes encoding chaperonin 60β from *Arabidopsis thaliana*. *Gene* **111,** 175–181.

Zillig, W. (1991). Comparative biochemistry of Archae and Bacteria. *Curr. Opinion Genet. Dev.* **1,** 544–551.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theoret. Biol.* **8,** 357–366.

3

Chaperonins of Photosynthetic Organisms

ANTHONY A. GATENBY

BioProcess Development Center
Central Research and Development
DuPont, Experimental Station
Wilmington, Delaware 19880

I. INTRODUCTION

From a historical perspective, the early observations on the nature and function of chloroplast chaperonin 60 (cpn60) were auspicious (Barraclough and Ellis, 1980; Roy *et al.*, 1982; Bloom *et al.*, 1983). The

significance of these papers was not fully appreciated at the time, perhaps because the role of "folding helpers" was not generally apparent. A number of the features more recently described for bacterial and mitochondrial chaperonins can be found in the early literature on chloroplast chaperonins, such as binding of nascent polypeptides to the chaperonin oligomer and the requirement for MgATP to release the bound target polypeptide. Because these studies described the binding of the large (L) subunit of ribulose-bisphosphate carboxylase-oxygenase (EC 4.1.1.39, Rubisco) to an oligomeric "assembly" protein, this oligomer was termed the large subunit binding protein (LSBP). The name was later modified to the Rubisco subunit binding protein when it was realized that the small (S) subunit of Rubisco could also bind to the LSBP (Gatenby *et al.*, 1988; Ellis and van der Vies, 1988). With the demonstration that a wide range of proteins imported into isolated chloroplasts would also form a complex with the binding protein, it was given its current name of chaperonin 60 (cpn60) (Lubben *et al.*, 1989) to reflect both its functional role as a molecular chaperone, and the size of the cpn60 protomers (60 kDa). cpn60 is now frequently used as a generic name for the related chaperonins in mitochondria and bacteria. An example of the binding of newly synthesized maize Rubisco L subunits to *Escherichia coli* cpn60 is shown in Fig. 1.

Viewed by electron microscopy (Hendrix, 1979; Pushkin *et al.*, 1982; McMullin and Hallberg, 1988; Saibil *et al.*, 1993; see Chapter 9) cpn60 from different organisms contains fourteen 60 kDa subunits arranged in two stacked rings of seven subunits, forming a double-toroid structure with a central cavity. The complex molecular architecture of chaperonins clearly has important mechanistic significance, since the pattern of sevenfold rotational symmetry is repeated for cpn60 homologs isolated from chloroplasts and mitochondria (Pushkin *et al.*, 1982; McMullin and Hallberg, 1988). Among the most intensively studied of these proteins are *E. coli* GroEL, yeast mitochondria hsp60, and cpn60 from plant chloroplasts. These are all highly conserved proteins at the level of amino acid sequence similarity (Hemmingsen *et al.*, 1988; Ellis and van der Vies, 1991; Zeilstra-Ryalls *et al.*, 1991). In contrast to bacteria and mitochondria, which contain only a single type of cpn60 subunit (Hemmingsen *et al.*, 1988; Reading *et al.*, 1989), chloroplasts contain two distinct cpn60 polypeptides (α and β) that are present in roughly equal amounts (Hemmingsen and Ellis, 1986; Musgrove *et al.*, 1987; Martel *et al.*, 1990).

Members of a second chaperonin subfamily that functionally interact with cpn60 are also oligomeric proteins, but are significantly smaller than cpn60, and usually contain identical subunits of about 10 kDa. This

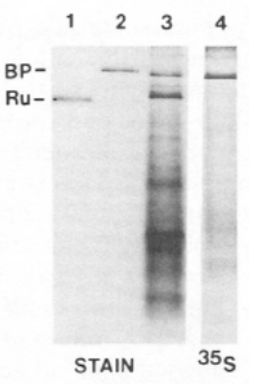


Fig. 1. Formation of a stable binary complex between the maize Rubisco L subunit and *E. coli* cpn60 (GroEL). A cloned maize *rbcL* gene (Gatenby *et al.*, 1981) was expressed in an *E. coli* cell-free coupled transcription/translation lysate in the presence of [^{35}S]methionine. Samples of the *in vitro* translation reaction mixture were analyzed by nondenaturing gel electrophoresis using 6% polyacrylamide gels. Samples in lanes 1–3 were stained with Coomassie brilliant blue, and lane 4 is an autoradiograph of lane 3. A complex collection of stained protein bands can be seen in lane 3, as expected for an *E. coli* S30 lysate. In contrast, the autoradiograph (lane 4) of the stained sample shown in lane 3 reveals that most of the synthesized Rubisco L subunits migrate as a discrete band near the top of the gel. By comparison with stained markers, the radioactive band in lane 4 does not have the same mobility as purified Rubisco (Ru) (lane 1), but does comigrate with pea Rubisco subunit binding protein (BP) (lane 2), now referred to as cpn60. A band of similar mobility to BP is present in the *E. coli* S30 lysate (lane 3), and subsequent experiments showed that it is the bacterial cpn60 (GroEL) which has bound to newly synthesized Rubisco L subunits. It was experiments like these, together with analogous approaches using isolated chloroplasts and direct comparison of DNA sequences, that led to the recognition of chaperonins in bacteria and chloroplasts that could assist in the folding of polypeptide chains by interacting with nonnative (partially folded) structures. This experiment was performed by the author.

smaller chaperonin protein is known as GroES in *E. coli,* and is more generally referred to as chaperonin 10 (cpn10), again reflecting its subunit size. Homologs of cpn10 are found in numerous prokaryotic species (Zeilstra-Ryalls *et al.,* 1991), mammalian mitochondria (Lubben *et al.,* 1990; Hartman *et al.,* 1992b), and higher plant chloroplasts (Bertsch *et al.,* 1992), and a functionally related protein called Gp31 is also expressed from a bacteriophage T4 gene during growth in *E. coli* (van der Vies *et al.,* 1994; see Chapter 6). The available evidence suggests that like cpn60, bacterial (Chandrasekhar *et al.,* 1986) and mitochondrial (Hartman *et*

al., 1992b) cpn10 oligomers possess toroidal structures with sevenfold symmetry. Electron micrographs of the chloroplast homolog of cpn10 also reveal a toroidal structure (Baneyx *et al.,* 1995). Intuitively, this observation makes sense since cpn60 and cpn10 form a stable complex with each other in the presence of certain adenine nucleotides. Therefore, the unusual sevenfold symmetry of cpn60 is mirrored in the structure of cpn10 molecules that must bind to cpn60 to promote efficient discharge and folding of the target polypeptide. The term cpn10 is actually somewhat misleading for the chloroplast chaperonin, because a functional chloroplast cpn10 homolog comprises two similar but nonidentical halves of about 10 kDa fused together (Bertsch *et al.,* 1992). The two halves of the chloroplast protein show 40% amino acid identity and are 70% homologous if conservative changes are also considered. Since the molecular mass of the fused cpn10 domains is about 21 kDa (Bertsch *et al.,* 1992; Baneyx *et al.,* 1995), we refer to this protein as cpn21.

II. CHLOROPLAST CHAPERONINS

There are many similarities between the transcriptional (Bradley and Gatenby, 1985) and translational (Gatenby *et al.,* 1989) mechanisms of chloroplasts and *E. coli,* which is no doubt a reflection of the essentially prokaryotic-like nature of these photosynthetic organelles. It is also apparent that factors influencing the final stages of gene expression, namely protein folding and assembly, are also conserved (Gatenby and Viitanen, 1994). Molecular chaperones in *E. coli* were initially identified using genotype screening for mutations altering their function, and subsequently followed by purification and biochemical analysis of the various proteins. With chloroplasts a genetical approach is rarely available, and so much of the characterization has been achieved by analysis of the purified proteins. An exception to this comment is the successful expression of an active chloroplast cpn10 homolog (cpn21) in *E. coli* (Baneyx *et al.,* 1995), which has permitted analysis of functional features using the approaches of bacterial genetics and mutagenesis.

A. Chaperonin 60

Chloroplast cpn60 was first discovered during studies on the biosynthesis of Rubisco in isolated chloroplasts. Although the majority of chloro-

plast proteins are synthesized in the cytosol and translocated into the developing organelle, a significant number, most notably the L subunit of Rubisco, are synthesized in the chloroplast stroma. The Rubisco L subunit polypeptide is by far the major product of *in organello* protein synthesis, and it assembles with imported S subunits into a holoenzyme with an L8S8 composition (Gatenby and Ellis, 1990). It was observed, following electrophoresis on nondenaturing polyacrylamide gels, that although Rubisco L subunits synthesized in isolated chloroplasts can assemble into the holoenzyme, a significant proportion of the L subunits are stably associated with a large oligomeric protein (Barraclough and Ellis, 1980). This oligomeric protein is greater than 600 kDa in size, and contains subunits of 60 kDa (Barraclough and Ellis, 1980; Hemmingsen and Ellis, 1986). The results of time course experiments during radiolabeling demonstrated that as radioactive L subunits become assembled into Rubisco holoenzyme, the radioactivity in the cpn60 oligomer declines. These observations raised the possibility that nascent Rubisco L subunits are specifically associated with cpn60 prior to assembly into holoenzyme, and that the cpn60 · L subunit binary complex is an obligatory intermediate in the assembly of Rubisco (Barraclough and Ellis, 1980; see Chapter 1).

Independent experiments also demonstrated that Rubisco L subunits synthesized *in vivo* or *in organello* can be recovered from intact chloroplasts in the form of two different sedimentation complexes of 7S and 29S, as determined on sucrose gradients (Roy *et al.,* 1982). The 29S complex contains unassembled Rubisco L subunits associated with cpn60, and the 7S complex may represent Rubisco dimers. When chloroplasts are incubated in the light, it is found that the newly synthesized L subunits present in both the 7S and the 29S complexes disappear and are subsequently located in the assembled 18S Rubisco holoenzyme (Roy *et al.,* 1982). This post-translational assembly of Rubisco is accelerated in chloroplast extracts by the addition of ATP, but the 29S cpn60 oligomer remains intact (Bloom *et al.,* 1983). However, in the presence of magnesium, ATP causes dissociation of the 29S cpn60 molecule, whereas a nonhydrolyzable analog of ATP has no effect (Bloom *et al.,* 1983; Musgrove *et al.,* 1987; Roy *et al.,* 1988). Although dissociation of the chloroplast cpn60 occurs at physiological concentrations of ATP, the low concentrations of chaperonins in these experiments would in themselves favor oligomer dissociation. As noted by the authors, the *in vivo* conditions might be quite different, and the higher concentration of chaperonins could permit oligomers to maintain their structure even in the pres-

ence of ATP (Musgrove *et al.*, 1987). In contrast to the effect of ATP on chloroplast cpn60 oligomer stability, the bacterial and mitochondrial cpn60 oligomers are not dissociated to a large extent by ATP. A complex set of reactions was proposed by Bloom *et al.* (1983) that requires nucleotides, magnesium, cpn60, and putative intermediates in the assembly of the Rubisco holoenzyme. Purified chloroplast cpn60 has been used to refold successfully a denatured bacterial Rubisco *in vitro* using defined biochemical components. Under the conditions used, where chemically denatured Rubisco fails to revert spontaneously to its native state, the successful refolding of bacterial Rubisco is obtained in the presence of chloroplast cpn60 and *E. coli* cpn10 in a reaction that requires ATP hydrolysis (Goloubinoff *et al.*, 1989b). Some additional complexities in the refolding and assembly of a chloroplast enzyme were noted in the studies by Chen and Jagendorf (1994), who obtain reconstitution of an active multisubunit coupling factor CF_1 core in the presence of molecular chaperones and MgATP. Successful folding and assembly of a catalytically active $\alpha_3\beta_3\gamma$ core from urea-denatured subunits do not occur in the presence of either chloroplast cpn60 and cpn21 or *E. coli* cpn60 and cpn10, but require a mixture of additional chaperones to be added. These additional components have not been fully characterized, but apparently include hsp70.

As already noted, chloroplast cpn60 is composed of two types of subunit of 61 and 60 kDa, known respectively as the α and β subunits (Musgrove *et al.*, 1987). The two cpn60 subunits are highly divergent in their predicted amino acid sequences (Martel *et al.*, 1990). It is not known whether the α and β subunits reside in the same or different cpn60 tetradecamers, which would result in either hetero- or homooligomers, respectively. Both cpn60 subunits are encoded by nuclear genes and are imported into chloroplasts following synthesis of the precursor form by cytosolic ribosomes (Hemmingsen and Ellis, 1986). Historically, it was the isolation and analysis of cDNA for the α subunit from plants that revealed a high degree of sequence similarity to the *E. coli* GroEL protein, and ultimately led to the identification of the family of proteins that are now called chaperonins (Hemmingsen *et al.*, 1988). Although the chloroplast cpn60 was originally demonstrated to be involved in Rubisco assembly (Barraclough and Ellis, 1980; Roy *et al.*, 1982; Bloom *et al.*, 1983), it undoubtedly plays a more general role in chloroplast biogenesis, and probably also in the development of other types of plastid.

B. Chaperonin 21

A wide range of purified proteins in their nonnative states can interact with cpn60; examples include Rubisco (Goloubinoff *et al.*, 1989b; Baneyx and Gatenby, 1992; van der Vies *et al.*, 1992), pre-β-lactamase (Laminet *et al.*, 1990), rhodanese (Mendoza *et al.*, 1991); Martin *et al.*, 1991), dihydrofolate reductase (Martin *et al.*, 1991; Viitanen *et al.*, 1991), citrate synthase (Buchner *et al.*, 1991), lactate dehydrogenase (Badcoe *et al.*, 1991), glutamine synthetase (Fisher, 1992), and phytochrome (Grimm *et al.*, 1993). From these, and other studies (Höll-Neugebauer *et al.*, 1991; Martin *et al.*, 1993), it has become apparent that chaperonins regulate protein folding by stabilizing folding intermediates, thereby influencing the kinetic partitioning between aggregated (misfolded) and correctly folded proteins (see Chapters 7 and 8). The release of target proteins bound to cpn60, and subsequent progression to the native state, occurs through interactions with the cpn10 chaperonin and MgATP. Even where cpn10 is not essential for release, its presence usually facilitates the discharge reaction (Laminet *et al.*, 1990; Martin *et al.*, 1991; Viitanen *et al.*, 1991; Fisher, 1992). This role for cpn10 in effective dissociation of the cpn60 · target protein complex suggested that chloroplasts might also contain a cpn10 homolog to facilitate protein folding by interacting with the chloroplast cpn60 (Lubben *et al.*, 1990).

A functionally related cpn10 protein was eventually identified in pea and spinach chloroplasts (Bertsch *et al.*, 1992). Surprisingly, the chloroplast cpn10 homolog is a polypeptide of about 21 kDa (cpn21), which is twice the size of bacterial and mitochondrial cpn10, and thus different from any cpn10 described previously. The chloroplast cpn21 (previously called ch-cpn10) comprises two distinct cpn10-like domains fused in tandem to give a binary structure (Bertsch *et al.*, 1992; see Fig. 2). On the basis of sequence similarities between each domain in cpn21 and other published cpn10 sequences, a potential oligopeptide linker is located in the mature protein at T103 (Bertsch *et al.*, 1992) with the sequence TDDVKD. This oligopeptide has a composition highly favorable for domain linkage (Argos, 1990), and is effectively positioned midway along the cpn21 polypeptide to provide an internal symmetry axis around which a pseudosymmetric structure could form following collapse and folding of the two domains. Each of the two fused domains possesses several highly conserved amino acid residues that are also

```
          1                                                  50
  GroES   MNIRPLHDRVIVKRKEVETKSAGGIVLTGSAAAKSTRGEVLAVGNGRILE
  9-104   TSVK  G   LI T I  E TTS  F PTA QK PQS   V I S KK--
107-202   KDLK  N  LLI VA   N TS  LL AE SKE PSF T V T P VLD
              ** *                                  * *

          51                                                 100
  GroES   NGE-VKPLDVKVGDIVIFNDGYGVKSEK-IDNEEVLIMSESDILAIVEA
  9-104   V DKKL VA  T AE VYSK- TGTEIE-V GSSH  VK D  IG L TD
107-202   E N-RI  P CS NT LYSK- AGNDF GV GSDYMVLRV  VM VLS
           *       *  *
```

Fig. 2. Each half of the mature spinach chloroplast cpn21 shows similarity to *E. coli* GroES. The GroES amino acid sequence is shown on the top lines numbered from 1 to 100. Beneath GroES is the sequence for amino acids 9–104 of the N-terminal domain of cpn21, and beneath it is the sequence for amino acids 107–202 of the C-terminal domain of cpn21. Blank spaces correspond to residues in cpn21 that are identical to GroES, and alignment gaps are marked with (-). The asterisks beneath eight of the residues indicate positions that are identical in cpn10 sequences deposited in GenBank. Based on data from Hemmingsen *et al.* (1988), Bertsch *et al.* (1992), and the GenBank database.

encoded in many other *groES* genes examined, suggesting that each domain could be functional. The two-domain structure of chloroplast cpn21 is also present in a broad range of distantly related photosynthetic eukaryotes, as judged by immunoblotting of plant extracts using antisera raised against cpn21. An immunoreactive protein of similar size is present in liverworts, mosses, club mosses, ferns, gymnosperms, monocots, and dicots (Baneyx *et al.*, 1995; see Fig. 3). The evolutionary divergence represented by this group of species covers a time span of 5×10^8 or more years. Clearly, the double-domain structure of the chloroplast cpn10 evolved early, and its unusual features have been maintained during the course of diverse speciation. Domain duplication presumably occurred during the early evolution of the more complex and differentiated chloroplasts following endosymbiosis, since the postulated progenitors of chloroplasts, namely photosynthetic cyanobacteria, possess the single-domain type of cpn10 (Webb *et al.*, 1990; Lehel *et al.*, 1993; see Chapter 2).

To test the biological activity of the double-domain cpn21 and each of its two separate single domains, the complete and partial coding sequences have been expressed in *E. coli.* Spinach cpn21 and the N-terminal (cpn10-N) and C-terminal (cpn10-C) domains were synthesized in a *groES* mutant strain that fails to support bacteriophage assembly.

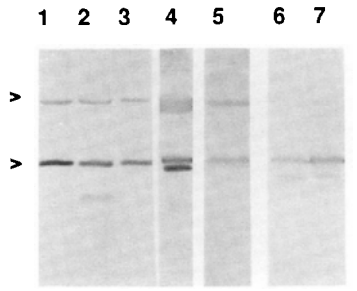


Fig. 3. Immunoblot of soluble leaf extracts resolved by SDS–PAGE and incubated with anti-spinach cpn21. Lane 1, *Juniperus horizontalis;* lane 2, *Taxus media;* lane 3, *Ginkgo biloba;* lane 4, *Dicranum flagellare;* lane 5, *Selaginella krausiania;* lane 6, *Zea mays;* lane 7, *Hordeum vulgare.* The lower arrow indicates the position of cpn21, and the upper arrow indicates cpn60 cross-reacting with contaminating anti-cpn60 in the serum. Proteins were electroblotted onto nitrocellulose, and sequentially incubated with rabbit anti-cpn21 and goat anti-rabbit IgG conjugated to alkaline phosphatase (Baneyx *et al.,* 1995).

The cpn21 protein and each separate domain support bacteriophage growth and can substitute for bacterial cpn10 in virus morphogenesis (Baneyx *et al.,* 1995). cpn21 and cpn10-N, but not cpn 10-C, also suppress the growth defect of a temperature-sensitive *groES* mutant strain. It has been established that in the presence of either ATP or ADP a stable complex can be formed between *E. coli* cpn60 and cpn10 (Chandrasekhar *et al.,* 1986; Lissin *et al.,* 1990; Viitanen *et al.,* 1990). cpn10 from mitochondria, cpn21 from chloroplasts, and Gp31 from bacteriophage T4 will each form a complex with *E. coli* cpn60 in the presence of adenine nucleotides (Lubben *et al.,* 1990; Bertsch *et al.,* 1992; Rospert *et al.,* 1993a; Burt and Leaver, 1994; van der Vies *et al.,* 1994; Baneyx *et al.,* 1995; see Fig. 4), indicating considerable functional conservation in the structure of the interactive cpn60/cpn10 surfaces. Purified spinach cpn21 inhibits the Mg^{2+}-dependent ATPase activities of both *E. coli* cpn60 (Baneyx *et al.,* 1995) and spinach chloroplast cpn60, the latter also possessing a Ca^{2+}-dependent ATPase activity (G. G. Chen, A. A. Gatenby, and A. T. Jagendorf, unpublished). Chloroplast cpn21 will effectively substitute for bacterial cpn10 in the chaperonin-facilitated refolding of denatured bacterial Rubisco (Baneyx *et al.,* 1995). Electron micrographs of cpn21 reveal a ring-like organization that is very similar to the surface views of the toroidal cpn10 protein from *E. coli* (Chandrasekhar *et al.,* 1986). The shape of the oligomer indicates that the cpn21 subunits are

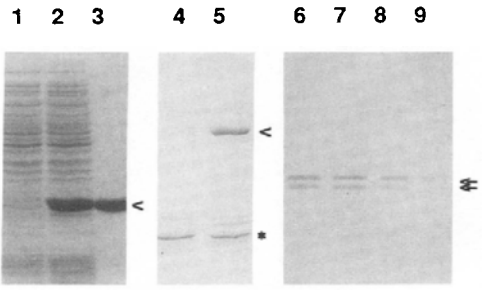
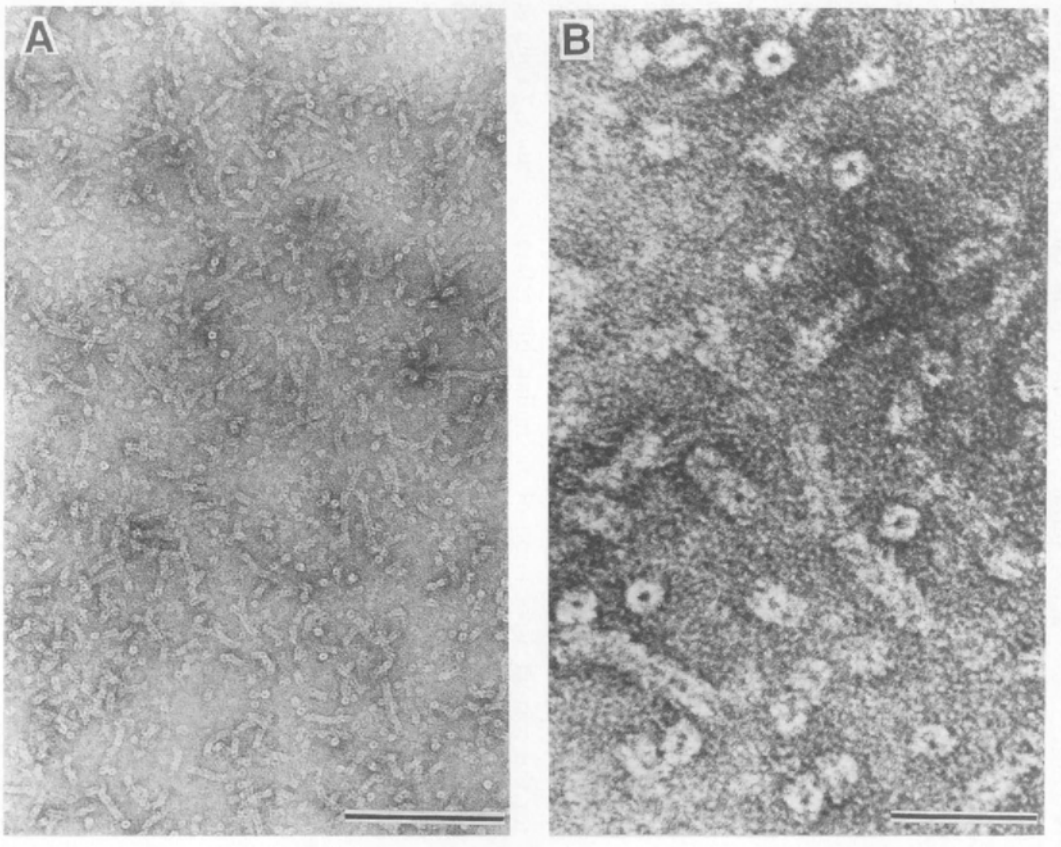


Fig. 4. Adenine nucleotide-dependent binding of *E. coli* and chloroplast cpn60 to immobilized cpn21. Plasmid pCK28 encodes an N-terminal fusion between six histidine residues and cpn21. The His-cpn21 protein was synthesized in *E. coli,* bound to nickel (Ni-NTA) columns and washed free of contaminating proteins. Purified *E. coli* cpn60, or soluble spinach chloroplast proteins, were incubated with the His-cpn21 resin ± adenine nucleotide, washed, eluted with imidazole, and analyzed by SDS–PAGE. Lane 1, JM105(pCK28) total preinduced proteins; lane 2, total induced JM105(pCK28) proteins; lane 3, His-cpn21 bound and eluted from Ni-NTA column (the arrow for lane 3 marks the position of His-cpn21); lanes 4 and 5 show protein eluted from column following passage of *E. coli* cpn60 -ADP (lane 4) and + 1 m*M* ADP (lane 5) (the arrow for lane 5 marks the position of cpn60 and the star the position of His-cpn21); lanes 6–9 show chloroplast extracts where 6 is the column preload, sample 7 is cpn60 bound to His-cpn21 in presence of 1 mM ATP, sample 8 is cpn60 bound in presence of 1 m*M* ADP, and sample 9 is pretreated with apyrase before column loading in the absence of either ADP or ATP. The two bands marked by arrows in lanes 6–9 are the α (top) and β (bottom) of spinach cpn60. Lanes 1–5 were stained with Coomassie brilliant blue, and lanes 6–9 immunoblotted, reacted with rabbit anti-cpn60, and goat anti-rabbit IgG conjugated to alkaline phosphatase (Baneyx *et al.,* 1995).

arranged with rotational symmetry around an axis through the center of the toroid (Baneyx *et al.,* 1995; see Fig. 5).

Because each domain of cpn21 is functional, this raises the possibility that the α and β cpn60 subunits in chloroplasts require different domain interactions for maximal activity, thus implying that the two cpn21 domains have preferred binding sites either on different cpn60 molecules

Fig. 5. Electron micrographs of purified spinach cpn21 synthesized in *E. coli.* The protein was fixed with 1% glutaraldehyde and negatively stained with 1% uranyl acetate. Two magnifications are shown; in (A) the bar represents 100 nm and in (B) the bar is equivalent to 20 nm. Electron micrographs were kindly provided by Jan van Breemen (Baneyx *et al.,* 1995).


A
B

or on different faces of the same molecule. Perhaps the fused-domain structure of the chloroplast cpn21 satisfies this requirement by presenting two differentially active surfaces to cpn60 in the target polypeptide discharge reaction. Binding experiments with histidine-tagged cpn21 reveal that both the α and the β subunits are present in cpn60 molecules that interact with cpn21 in the presence of ADP or ATP, but binding does not occur in the absence of ADP or ATP (Baneyx *et al.*, 1995; see Fig. 4). Irrespective of whether the two cpn10-like domains of cpn21 have different specificities in their interactions with cpn60 α or β, why should the chloroplast cpn21 maintain two functional cpn10 domains fused together, rather than using separate cpn10 subunits? Many other proteins have apparently also evolved by domain duplication and fusion (McLachlan, 1987). For example, domains are autonomous cooperative folding units, and it has been found that domain fusion enhances the rate of folding and can improve stability by reducing the entropy of the unfolded state (Liang *et al.*, 1993). This improvement occurs because the two chains are no longer independent of each other, thus reducing the translational and rotational entropy of the unfolded polypeptide. Perhaps the most compelling reason for domain fusion is that it provides a mechanism for the correct association of dissimilar, but related, subunits (Tang *et al.*, 1979). As each domain collapses to its native-like structure it is tethered to the other domain by a peptide linker, thus ensuring a high local concentration of the fused subunits with a fixed polarity. This polarity may be important in the overall assembly of the cpn21 oligomer to achieve close packing in a sevenfold rotationally symmetrical molecule. This hypothesis implies a protein folding mechanism that requires selectivity in the binding of distinct surfaces of cpn21 to the α and β forms of cpn60, to give optimum interactions and efficiency in the polypeptide discharge reaction. Clearly, a better understanding of the manner in which chloroplast cpn21 interacts with chloroplast cpn60 will require additional structural information, together with the isolation and analysis of mutated plastid cpn21. Such studies may provide an explanation for the unique double-domain cpn10 proteins found in chloroplasts.

C. Interaction of Chloroplast Chaperonins with Imported Proteins

The first examples of imported proteins becoming associated with chaperonins were reported for the Rubisco L and S subunits following their uptake into isolated pea chloroplasts (Gatenby *et al.*, 1988; Ellis and

van der Vies, 1988). The L subunit of Rubisco from the cyanobacterium *Anacystis nidulans* was fused to a soybean chloroplast transit peptide and, following import, was observed to interact with the plastid cpn60 protein (Gatenby *et al.,* 1988). This interaction is transient, and during a time course experiment the number of Rubisco L subunits that are associated with cpn60 decreases at a similar rate as L subunits become assembled into Rubisco holoenzyme. Similar kinetics are observed in the presence of chloramphenicol, indicating that active protein synthesis is not required for the interaction of imported large subunits with cpn60 and their subsequent transfer into holoenzyme. These data, showing imported Rubisco L subunits interacting with cpn60 before association into holoenzyme, are similar to results obtained by Barraclough and Ellis (1980), except that the latter authors studied L subunits synthesized within chloroplasts. The two experimental approaches indicate that chaperonin interactions and Rubisco assembly follow similar kinetics, whether the L subunit polypeptide is in a nonnative state due to either import through the chloroplast membrane or synthesis in the stromal compartment. Imported S subunits of Rubisco are also associated with cpn60 to form stable complexes that can be resolved by nondenaturing gel electrophoresis, although not to the same extent observed for large subunits (Gatenby *et al.,* 1988; Ellis and van der Vies, 1988).

A more general role for chloroplast cpn60 in chloroplast development is indicated by the observation that a wide range of unrelated proteins become associated with this chaperonin after import into isolated chloroplasts (Lubben *et al.,* 1989). These imported proteins include the β subunit of ATP synthase, glutamine synthetase, chloramphenicol acetyltransferase, the light-harvesting chlorophyll *a/b* binding protein (LHCP), and pre-β-lactamase. The observed degree of interaction, however, varies significantly for each target protein. The complexes formed between cpn60 and imported proteins are dissociated by the addition of MgATP, but not when a nonhydrolyzable analog is substituted. Binding of these imported proteins to cpn60 is probably an early event in the folding of chloroplast polypeptides, and occurs before they can progress to their native states. This broad specificity of chloroplast cpn60 for a range of imported proteins is very similar in principle to the interaction of a wide range of proteins with the cpn60 protein of *E. coli* (Van Dyk *et al.,* 1989; Viitanen *et al.,* 1992). Other studies have also shown an interaction between imported proteins and chloroplast chaperonins (Tsugeki and Nishimura, 1993; Madueno *et al.,* 1993). By following the kinetics of association after import of ferredoxin-$NADP^+$ reductase, Tsugeki and

Nishimura (1993) detected an initial interaction between the imported protein and hsp70. This initial pathway is similar to the examples of imported mitochondrial proteins that first associate with hsp70 prior to a transfer to cpn60 (Hendrick and Hartl, 1993; see Chapter 4). In contrast, it appears that the imported Rieske iron–sulfur protein of chloroplasts interacts with cpn60 before forming a complex with hsp70 (Madueno *et al.*, 1993), and it was suggested that these two chloroplast molecular chaperones operate at different points of the import pathway.

D. Alterations to Chaperonin Composition of Chloroplasts

The chaperonin content of chloroplasts can be modified in transgenic plants where the "transforming" DNA encodes a transit peptide targeting sequence fused to a chaperonin gene. Wu *et al.* (1993) fused a pea Rubisco S subunit transit peptide coding sequence to *E. coli groEL*, and positioned the resulting gene fusion downstream of a tandem 35S cauliflower mosaic virus (CaMV) promoter. Expression of the chaperonin gene fusion in transgenic tobacco plants leads to production of modified bacterial cpn60, which is targeted and imported into chloroplasts. Depending on the particular transgenic plant analyzed, bacterial cpn60 accumulates to high or low levels, and the imported chaperonin becomes resistant to proteolytic degradation when isolated intact chloroplasts are incubated with thermolysin (Wu *et al.*, 1993). Interestingly, the N-terminal transit peptide is not removed from all of the imported cpn60 molecules, with about 20% of the protein retaining the transit peptide. This incomplete processing is similar to that observed when bacterial β-lactamase is fused to a transit peptide and imported into isolated pea chloroplasts (Lubben *et al.*, 1989), and these intact precursors may represent folded forms in which the processing sites become inaccessible to chloroplast proteases. The retention of an N-terminal transit peptide does not appear to impair bacterial cpn60 oligomerization. High levels of foreign cpn60 in transgenic tobacco plants represent 1–2% of total soluble leaf protein, which is a concentration in excess of the endogenous levels of chloroplast cpn60 (Wu *et al.*, 1993). The major chaperonin species present in high bacterial cpn60 plants appear to be novel hybrid tetradecameric molecules containing both *E. coli* and tobacco cpn60 polypeptides. In contrast, transgenic plants accumulating low levels of *E. coli* cpn60 possessed as predominant species both authentic chloroplast cpn60 and bacterial cpn60 tetradecamers. It was suggested that the authentic tetradecamers formed because of the combined influence of lower levels of *groEL* expression and the

differential temporal regulation of the natural and foreign genes, plant cpn60 expression being diurnal and that of *groEL* using a CaMV 35S promoter being constitutive (Wu *et al.*, 1993). In all cases the growth and development of transgenic and control plants were indistinguishable.

A different strategy to modify the chaperonin content of chloroplasts is to use antisense technology to inactivate the expression of cpn60 genes. Zabaleta *et al.* (1994) used this approach to reduce the levels of the β subunit of cpn60 in transgenic tobacco plants and found that abnormal phenotypes develop. The modified plants grow more slowly, have delayed flowering, are stunted, and possess chlorotic leaves. The most extreme effect in antisense cpn60-β plants is lethality. It is not clear from these data whether the abnormal phenotypes occur because of a reduction in total cpn60 concentration in chloroplasts or whether the phenotypes are a reflection of specific functions attributable to the β subunit, which cannot be compensated for by the residual α subunits.

III. CHAPERONIN 60 AND CHAPERONIN 10 IN MITOCHONDRIA FROM ANGIOSPERMS

Compared to the two cpn60 types (α and β) and the "double" cpn21 proteins found in chloroplasts, those identified so far in plant mitochondria appear to have structural properties similar to those of chaperonins located in the mitochondria of other eukaryotic species and in prokaryotes. Very little characterization of these plant proteins at the biochemical level has been reported, but the limited data available conform to the functional properties of mitochondrial chaperonins from other kingdoms (see note added in proof).

A. Chaperonin 60

The cpn60 protein present in plant mitochondria was initially characterized by Prasad and Hallberg (1989). Analysis of the maize mitochondrial cpn60 protein revealed that its morphological structure is very similar to that of *E. coli* cpn60 with 12.5-nm-diameter particles exhibiting sevenfold axial symmetry, and with each of the 14 subunits having a relative molecular mass of 62 kDa. Sequencing of cDNA clones of mitochondrial cpn60 from *Arabidopsis thaliana* and maize indicates that the

predicted proteins are about 55–60% identical to the homologous proteins from *E. coli* and yeast mitochondria. Plant mitochondrial cpn60 is nuclear-encoded, and it appears to possess transit peptide sequences for mitochondrial targeting (Prasad and Stewart, 1992). Cross-reacting polypeptides of 62 kDa were also identified by immunoblotting mitochondrial extracts from barley, carrot, cauliflower, mungbean, pea, radish, tomato, and wheat (Prasad and Hallberg, 1989). There is some degree of heat shock response, with a two- to threefold increase in both transcript levels and cpn60 polypeptide accumulation (Prasad and Hallberg, 1989; Prasad and Stewart, 1992). A developmental response was also noted by these authors, who found that the overall concentration of maize cpn60 was two- to threefold higher at early stages of germination than at later ones, suggesting that during periods of increased mitochondrial biogenesis there may be a demand for elevated chaperonin levels.

In a functional study of the maize cpn60 protein, Prasad *et al.* (1990) were able to demonstrate that two radioactive proteins synthesized in isolated mitochondria form stable complexes with cpn60 oligomers. One of these proteins was identified as the α subunit of F_1-ATPase, and its association with cpn60 is sufficiently stable to withstand sucrose gradient centrifugation and electrophoresis on nondenaturing gels. These results are similar in concept to the observation that Rubisco L subunits synthesized in chloroplasts form a stable complex with plastid cpn60 (Barraclough and Ellis, 1980; Bloom *et al.*, 1983), and indicate that in both plant mitochondria and chloroplasts at least some of the newly synthesized organellar proteins interact with cpn60 during progression to their native states.

B. Chaperonin 10

Although a preliminary report suggested the presence of a cpn10 homolog in barley mitochondria (Hartman *et al.*, 1992a), the unambiguous identification and characterization of the plant mitochondrial cpn10 was achieved only recently (Burt and Leaver, 1994). Mitochondrial extracts were prepared from potato (*Solanum tuberosum*) tubers and incubated in the presence or absence of MgATP. In the presence of MgATP it was assumed that a stable binary complex would form between the endogenous mitochondrial cpn60 and cpn10, which could subsequently be isolated by affinity chromatography on a column of immobilized potato cpn60 antibodies. This, indeed, turned out to be the case and a

8.6-kDa polypeptide was found to bind to cpn60 only when MgATP was present (Burt and Leaver, 1994). Amino acid sequence analysis of a trypsin-digestion fragment revealed a sequence match in 10 out of 11 positions when compared with rat mitochondrial cpn10. The plant mitochondrial cpn10 protein is, therefore, similar in overall size to the homologous protein found in mammalian (Lubben *et al.*, 1990; Hartman *et al.*, 1992b) and fungal (Rospert *et al.*, 1993a,b) mitochondria, rather than the fused double-domain cpn21 structure located in chloroplasts (Bertsch *et al.*, 1992).

IV. CHAPERONIN 60 AND CHAPERONIN 10 IN PHOTOSYNTHETIC PROKARYOTIC ORGANISMS

It is generally accepted that the chloroplasts of plants and algae originated as endosymbionts, with those organisms having plastids surrounded by two membranes (rhodophytes, chlorophytes, and land plants) evolving from a progenitor phagotrophic eukaryotic host plus cyanobacterium (Douglas *et al.*, 1991). It is, therefore, interesting to consider the structure and function of chaperonins from photosynthetic prokaryotes in terms not only of the protists themselves but also of their possible contribution to the chaperonins of chloroplasts (see also Chapter 2). Analysis of chaperonins from photosynthetic prokaryotes has primarily focused on the cyanobacteria (blue-green algae), the purple nonsulfur bacterium *Rhodobacter sphaeroides,* and the purple sulfur bacterium *Chromatium vinosum.*

Localization experiments with the nitrogen-fixing cyanobacterial strain *Anabaena* PCC 7120 reveal the presence of cpn60 in both vegetative cells and heterocysts, with antisera raised against *E. coli* cpn60 cross-reacting with a 65-kDa protein (Jäger and Bergman, 1991). This strain of *Anabaena* forms filaments with photosynthetic vegetative cells and nitrogen-fixing heterocysts. The immunochemical studies of vegetative cells indicate that cpn60 is present both in the polyhedral carboxysomes, the apparent site of Rubisco in strain 7120, and in the cytoplasm, with an enrichment in the thylakoid region. An even distribution of cpn60 was observed in heterocysts (Jäger and Bergman, 1991). Clues concerning the genetic arrangement of the *groE* operon in cyanobacteria were reported by Cookson *et al.* (1989) and Gupta *et al.* (1989), who pointed out that the unknown reading frames URF-4 and URF-3 (partial) of *Synechococ-*

cus PCC 6301 described by Cozens and Walker (1987) are highly homologous to *groES* and *groEL* from *E. coli.* A more detailed analysis of the *groESL* operon and its regulation was subsequently described for *Synechococcus* PCC 7942 (Webb *et al.,* 1990). The gene order for strain PCC 7942 is identical to that found in strain PCC 6301 and *E. coli,* with *groES* preceding *groEL,* and separated by 46 bp. The derived proteins from PCC 7942 show identity to *E. coli* cpn10 (44%), *E. coli* cpn60 (55%), yeast hsp60 (48%), and wheat cpn60 (50%), and the cyanobacterial cpn60 homolog is recognized by antisera against *E. coli* cpn60 (Webb *et al.,* 1990). Transcription of the *groESL* operon of *Synechococcus* PCC 7942 increases dramatically for 20 min following heat shock at 45°C, and produces levels of mRNA accumulation of 120-fold over that present in cells grown at 30°C. The response is transient and levels decline to normal after 60 min. This pattern of expression, however, does lead to increased accumulation of chaperonins in cells under stress conditions that can destabilize protein folding intermediates, at which time it may be important to have increased concentrations of cpn60 to capture the labile intermediates.

Chitnis and Nelson (1991) initially identified the gene for cpn60 from the cyanobacterium *Synechocystis* PCC 6803 and found that its transcription increases in response to heat shock, oxidative stress, and ultraviolet exposure. In addition, the heat shock-induced synthesis of cpn60 and its subsequent purification were used to identify the cpn60 homolog from *Synechocystis* PCC 6803 by others (Lehel *et al.,* 1992). The HSP64 protein induced by a temperature shift from 30°C to 40–45°C cross-reacts with antisera against *E. coli* cpn60, and N-terminal sequence analysis reveals 76% identity to *Synechococcus* cpn60 (Lehel *et al.,* 1992) and 56–58% identity to other bacterial homologs (Chitnis and Nelson, 1991). An interesting feature of the *groEL* gene isolated by the latter authors is the absence of a closely linked *groES* gene. Further analysis, however, of the chaperonin genes from PCC 6803 subsequently revealed the presence of a second *groEL* gene that has a *groES* homolog closely positioned at its 5′ end (Lehel *et al.,* 1993). N-terminal sequence analysis of the purified chaperonin proteins shows that these regions are identical to the predicted sequences of the two proteins in the *groESL* operon. Clearly, *Synechocystis* PCC 6803 has an unusual arrangement of chaperonin genes with two copies of *groEL,* only one of which is transcriptionally linked to a *groES* gene. This type of gene arrangement is not unique, and multiple copies of *groEL* genes have been identified in species of *Streptomyces* (Mazodier *et al.,* 1991) and mycobacteria (Kong *et al.,*

1993), suggesting that the multiple cpn60 proteins may fulfill physiologically distinct roles. Needless to say, the presence of two copies of *groEL* in certain cyanobacteria may have some significance to the occurrence of two related forms of cpn60 (α and β subunits) in chloroplasts (Musgrove *et al.*, 1987), particularly considering the probable evolution of chloroplasts from cyanobacteria.

Chaperonins from the photosynthetic bacteria *C. vinosum* and *R. sphaeroides* have also been characterized. A homolog of cpn60 was initially identified in *C. vinosum* by cross-reaction of a 700-kDa protein composed of 60-kDa subunits with antisera against pea chloroplast cpn60 (Torres-Ruiz and McFadden, 1988). The purified protein also exhibits some N-terminal sequence similarity to pea cpn60. Subsequently, a cpn10 homolog was also purified from *C. vinosum* and shown to interact with the cpn60 protein to form a binary complex in the presence of MgATP (Torres-Ruiz and McFadden, 1992). As observed for other cpn10 · cpn60 complexes, the *C. vinosum* cpn60/cpn10 interaction results in the inhibition of the ATPase activity of cpn60; this effect has also been reported for the *R. sphaeroides* proteins (Terlesky and Tabita, 1991). The *groES* and *groEL* genes from *C. vinosum* were isolated using a functional expression strategy based on the failure of bacteriophage λ to grow on a *groEL140 E. coli* strain. By plating a *C. vinosum* genomic library constructed in bacteriophage λ with *groEL140* cells, and selecting for plaque formation, Ferreyra *et al.* (1993) successfully cloned the *groESL* operon. The two genes are linked on a contiguous fragment of DNA, with *groES* preceding *groEL*. It was also observed that expression of the *C. vinosum* chaperonin genes in *E. coli*, together with cloned Rubisco genes from different phototrophic bacteria, results in enhanced assembly of the latter enzymes (Ferreya *et al.*, 1993). Although the genetic organization of the *groESL* genes from *R. sphaeroides* has not yet been established, this prokaryote too contains characteristic chaperonin proteins that have been purified (Terlesky and Tabita, 1991). The cpn60 and cpn10 proteins of *R. sphaeroides* are oligomers with 61 and 12.7 kDa subunits, respectively, and copurify together in the presence of 1 m*M* ATP.

V. CONCLUDING COMMENTS

It is ironic that studies initiated in the 1980s to understand the complexity of higher plant chloroplast Rubisco assembly (Barraclough and Ellis,

1980; Gatenby *et al.*, 1981; Roy *et al.*, 1982; Bloom *et al.*, 1983; Hemmingsen and Ellis, 1986; Musgrove *et al.*, 1987; Gatenby *et al.*, 1988) should have crossed paths with studies on bacteriophage assembly (Georgopoulos *et al.*, 1973; Hendrix, 1979; Tilly *et al.*, 1981), to contribute instead to the discovery and characterization of the chaperonin family of molecular chaperones in bacteria (Hemmingsen *et al.*, 1988; Bochkareva *et al.*, 1988; Van Dyk *et al.*, 1989; Goloubinoff *et al.*, 1989a,b; Laminet *et al.*, 1990), chloroplasts (Hemmingsen *et al.*, 1988; Lubben *et al.*, 1989; Gupta *et al.*, 1989; Bertsch *et al.*, 1992; Baneyx *et al.*, 1995), and mitochondria (McMullin and Hallberg, 1988; Cheng *et al.*, 1989; Prasad and Hallberg, 1989; Reading *et al.*, 1989; Ostermann *et al.*, 1989; Lubben *et al.*, 1990). It should also be recalled that structural similarities were identified between the GroEL protein of *E. coli* and a pea leaf protein as early as 1982 in a paper that was not generally noticed, and remarked on at the time (Pushkin *et al.*, 1982).

Although we now know many of the details of chaperonin interactions with folding intermediates, and how these interactions facilitate protein folding (reviewed by Hendrick and Hartl, 1993; Gatenby and Viitanen, 1994; see Chapters 7 and 8), the folding and assembly of the L8S8 form of Rubisco from higher plants is still elusive when attempted either *in vitro* using purified components or in *E. coli* using expressed genes (Gatenby *et al.*, 1987; Gatenby and Ellis, 1990). This failure is a reflection of our incomplete understanding of how this major component of the photosynthetic pathway in chloroplasts is assembled, and is a puzzle since the analogous L8S8 Rubisco from cyanobacteria is assembled in *E. coli* (Gatenby *et al.*, 1985) using the GroE chaperonins (Goloubinoff *et al.*, 1989a). Even with the knowledge that chaperonins facilitate cyanobacterial Rubisco assembly in *E. coli*, reconstruction of the folding pathway *in vitro* in the presence of chaperonins has failed to yield a functional hexadecamer (Gatenby, unpublished). The problem appears to reside at the stage of L subunit folding and assembly into a soluble core octomer, and insolubility of plant Rubisco L subunits is well known (Gatenby, 1984). Once a stable L8 core is formed from plant subunits, the binding of eight S subunits to the core should proceed relatively well since it is known that plant S subunits remain soluble when expressed in *E. coli* (Gatenby *et al.*, 1987), and they also possess the ability to bind to a cyanobacterial L8 core *in vivo* to give a functional hybrid Rubisco enzyme (van der Vies *et al.*, 1986). Clearly there are many problems remaining to be solved in our understanding of the biogenesis of proteins in photosynthetic organisms, and how this process is influenced by molecu-

lar chaperones, and especially the chaperonins. Our knowledge to date is undoubtedly based on the more abundant and easily isolatable chaperonins from plants and other photosynthetic organisms, and there are sure to be many surprises and delights as we gradually unravel the complexity and mystery of protein folding in living cells.

NOTE ADDED IN PROOF: The following references were omitted from the review and should be consulted for information about two forms of cpn60 in pumpkin mitochondria (Tsugeki *et al.*, 1992) and a recent detailed study of cpn21 and cpn60 from chloroplasts (Viitanen *et al.*, 1995).

Tsugeki, R., Mori, H., and Nishimura, M. (1992). Purification, cDNA cloning and Northern-blot analysis of mitochondrial chaperonin 60 from pumpkin cotyledons. *Eur. J. Biochem.* **209,** 453–458.

Viitanen, P. V., Schmidt, M., Buchner, J., Suzuki, T., Vierling, E., Dickson, R., Lorimer, G. H., Gatenby, A. A., and Soll, J. (1995). Functional characterization of the higher plant chloroplast chaperonins. *J. Biol. Chem.* **270,** 18158–18164.

REFERENCES

Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol.* **211,** 943–958.

Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J. J., Lund, P., and Clarke, A. R. (1991). Binding of a chaperonin to the folding intermediates of lactate dehydrogenase. *Biochemistry* **30,** 9195–9200.

Baneyx, F., and Gatenby, A. A. (1992). A mutation in GroEL interferes with protein folding by reducing the rate of discharge of sequestered polypeptides. *J. Biol. Chem.* **267,** 11637–11644.

Baneyx, F., Bertsch, U., Kalbach, C. E., van der Vies, S. M., Soll, J., and Gatenby, A. A. (1995). Spinach chloroplast cpn21 co-chaperonin possesses two functional domains fused together in a toroidal structure, and exhibits nucleotide-dependent binding to plastid chaperonin 60. *J. Biol. Chem.* **270,** 10695–10702.

Barraclough, R., and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated pea chloroplasts. *Biochim. Biophys. Acta* **608,** 19–31.

Bertsch, U., Soll, J., Seetharam, R., and Viitanen, P. V. (1992). Identification, characterization, and DNA sequence of a functional "double" groES-like chaperonin from chloroplasts of higher plants. *Proc. Natl. Acad. Sci. USA* **89,** 8696–8700.

Bloom, M. V., Milos, P., and Roy, H. (1983). Light-dependent assembly of ribulose 1,5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* **80,** 1013–1017.

Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. (1988). Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. *Nature (London)* **336,** 254–257.

Bradley, D., and Gatenby, A. A. (1985). Mutational analysis of the maize chloroplast ATPase-β subunit gene promoter: The isolation of promoter mutants in *E. coli* and

their characterization in a chloroplast *in vitro* transcription system. *EMBO J.* **4,** 3641–3648.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kiefhaber, T. (1991). GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30,** 1586–1591.

Burt, W. J. E., and Leaver, C. J. (1994). Identification of a chaperonin-10 homologue in plant mitochondria. *FEBS Lett.* **339,** 139–141.

Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* **261,** 12414–12419.

Chen, G. G., and Jagendorf, A. T. (1994). Chloroplast molecular chaperone-assisted refolding and reconstitution of an active multisubunit coupling factor CF_1 core. *Proc. Natl. Acad. Sci. USA* **91,** 11497–11501.

Cheng, M., Hartl, F., Martin, J., Pollock, R., Kalousek, F., Neupert, W., Halberg, E., and Horwich, A. (1989). Mitochondrial heat-shock protein Hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* (*London*) **337,** 620–625.

Chitnis, P. R., and Nelson, N. (1991). Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp PCC 6803. *J. Biol. Chem.* **266,** 58–65.

Cookson, M. J., Baird, P. N., Hall, M. C., and Coates, A. R. (1989). Identification of two unknown reading frames in *Synechococcus* 6301 as homologues of the 10k and 65k antigen genes of *Mycobacterium tuberculosis* and related heat shock genes in *E. coli* and *Coxiella burnetii*. *Nucleic Acids Res.* **17,** 6392.

Cozens, A. L., and Walker, J. E. (1987). The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. Support for an endosymbiotic origin of chloroplasts. *J. Mol. Biol.* **194,** 359–383.

Douglas, S. E., Murphy, C. A., Spencer, D. F., and Gray, M. W. (1991). Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* (*London*) **350,** 148–151.

Ellis, R. J., and van der Vies, S. M. (1988). The Rubisco subunit binding protein. *Photosynth. Res.* **16,** 101–115.

Ellis, R. J., and van der Vies, S. M. (1991). Molecular chaperones. *Annu. Rev. Biochem.* **60,** 321–347.

Ferreyra, R. G., Soncini, F. C., and Viale, A. M. (1993). Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the phototrophic sulfur bacterium *Chromatium vinosum*. *J. Bacteriol.* **175,** 1514–1523.

Fisher, M. T. (1992). Promotion of the *in vitro* renaturation of dodecameric glutamine synthetase from *Escherichia coli* in the presence of GroEL (chaperonin-60) and ATP. *Biochemistry* **31,** 3955–3963.

Gatenby, A. A. (1984). The properties of the large subunit of maize ribulose bisphosphate carboxylase/oxygenase synthesized in *Escherichia coli*. *Eur. J. Biochem.* **144,** 361–366.

Gatenby, A. A., and Ellis, R. J. (1990). Chaperone function: The assembly of ribulose bisphosphate carboxylase-oxygenase. *Annu. Rev. Cell Biol.* **6,** 125–149.

Gatenby, A. A., and Viitanen, P. V. (1994). Structural and functional aspects of chaperonin-mediated protein folding. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45,** 469–491.

Gatenby, A. A., Castleton, J. A., and Saul, M. W. (1981). Expression in *E. coli* of maize and wheat chloroplast genes for large subunit of ribulose bisphophate carboxylase. *Nature* (*London*) **291,** 117–121.

Gatenby, A. A., van der Vies, S. M., and Bradley, D. (1985). Assembly in *E. coli* of a functional multi-subunit ribulose bisphosphate carboxylase from a blue-green alga. *Nature (London)* **314,** 617–620.

Gatenby, A. A., van der Vies, S. M., and Rothstein, S. J. (1987). Co-expression of both the maize large and wheat small subunit genes of ribulose-bisphosphate carboxylase in *Escherichia coli. Eur. J. Biochem.* **168,** 227–231.

Gatenby, A. A., Lubben, T. H., Ahlquist, P., and Keegstra, K. (1988). Imported large subunits of ribulose bisphosphate carboxylase/oxygenase, but not imported β-ATP synthase subunits, are assembled into holoenzyme in isolated chloroplasts. *EMBO J.* **7,** 1307–1314.

Gatenby, A. A., Rothstein, S. J., and Nomura, N. (1989). Translational coupling of the maize chloroplast *atpB* and *atpE* genes. *Proc. Natl. Acad. Sci. USA* **86,** 4066–4070.

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989a). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature (London)* **342,** 884–889.

Goloubinoff, P., Gatenby, A. A., and Lorimer, G. H. (1989b). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli. Nature (London)* **337,** 44–47.

Grimm, R., Donaldson, G. K., van der Vies, S. M., Schäfer, E., and Gatenby, A. A. (1993). Chaperonin-mediated reconstitution of the phytochrome photoreceptor. *J. Biol. Chem.* **268,** 5220–5226.

Gupta, R. S., Picketts, D. J., and Ahmad, S. (1989). A novel protein "chaperonin" supports the endosymbiotic origin of mitochondrion and plant chloroplast. *Biochem. Biophys. Res. Commun.* **163,** 780–787.

Hartman, D. J., Dougan, D., Hoogenraad, N. J., and Hoj, P. B. (1992a). Heat shock proteins of barley mitochondria and chloroplasts. Identification of organellar hsp10 and 12: putative chaperonin 10 homologues. *FEBS Lett.* **305,** 147–150.

Hartman, D. J., Hoogenraad, N. J., Condron, R., and Hoj, P. B. (1992b). Identification of a mammalian 10-kDa heat shock protein, a mitochondrial chaperonin 10 homologue essential for assisted folding of trimeric ornithine transcarbamoylase *in vitro. Proc. Natl. Acad. Sci. USA* **89,** 3394–3398.

Hemmingsen, S. M., and Ellis, R. J. (1986). Purification and properties of ribulosebisphosphate carboxylase large subunit binding protein. *Plant Physiol.* **80,** 269–276.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature (London)* **333,** 330–334.

Hendrick, J. P., and Hartl, F.-U. (1993). Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62,** 349–384.

Hendrix, R. W. (1979). Purification and properties of groE, a host protein involved in bacteriophage lambda assembly. *J. Mol. Biol.* **129,** 375–392.

Höll-Neugebauer, B., Rudolph, R., Schmidt, M, and Buchner, J. (1991). Reconstitution of a heat shock effect *in vitro:* Influence of GroE on the thermal aggregation of α-glucosidase from yeast. *Biochemistry* **30,** 11,609–11,614.

Jäger, K. M., and Bergman, B. (1991). Localization of a multifunctional chaperonin (GroEL protein) in nitrogen-fixing *Anabaena* PCC7120. *Planta* **183,** 120–125.

Kong, T. H., Coates, A. R. M., Butcher, P. D., Hickman, C. J., and Shinnick, T. M. (1993). *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc. Natl. Acad. Sci. USA* **90,** 2608–2612.

Laminet, A. A., Ziegelhoffer, T., Georgopoulos, C., and Plückthun, A. (1990). The *Escherichia coli* heat shock proteins GroEL and GroES modulate the folding of the β-lactamase precursor. *EMBO J.* **9,** 2315–2319.

Lehel, C., Wada, H., Kovacs, E., Török, Z., Gombos, Z., Horvath, I., Murata, N., and Vigh, L. (1992). Heat shock protein synthesis of the cyanobacterium *Synechocystis* PCC 6803: Purification of the GroEL-related chaperonin. *Plant Mol. Biol.* **18,** 327–336.

Lehel, C., Los, D., Wada, H., Györgyei, J., Horvath, I., Kovacs, E., Murata, N., and Vigh, L. (1993). A second *groEL*-like gene, organized in a *groESL* operon is present in the genome of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **268,** 1799–1804.

Liang, H., Sandberg, W. S., and Terwilliger, T. C. (1993). Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc. Natl. Acad. Sci. USA* **90,** 7010–7014.

Lissin, N. M., Venyaminov, S. Y., and Girshovich, A. S. (1990). (Mg-ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature* (*London*) **348,** 339–342.

Lubben, T. H., Donaldson, G. K., Viitanen, P. V., and Gatenby, A. A. (1989). Several proteins imported into chloroplasts form stable complexes with the groEL-related chloroplast molecular chaperone. *Plant Cell* **1,** 1223–1230.

Lubben, T. H., Gatenby, A. A., Donaldson, G. K., Lorimer, G. H., and Viitanen, P. V. (1990). Identification of a groES-like chaperonin in mitochondria that facilitates protein folding. *Proc. Natl. Acad. Sci. USA* **87,** 7683–7687.

Madueno, F., Napier, J. A., and Gray, J. C. (1993). Newly imported Rieske iron-sulfur protein associates with both cpn60 and hsp70 in the chloroplast stroma. *Plant Cell* **5,** 1865–1876.

Martel, R., Cloney, L. P., Pelcher, L. E., and Hemmingsen, S. M. (1990). Unique composition of plastid chaperonin-60: α and β polypeptide-encoding genes are highly divergent. *Gene* **94,** 181–187.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. *Nature* (*London*) **352,** 36–42.

Martin, J., Mayhew, M., Langer, T., and Hartl, F.-U. (1993). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature* (*London*) **366,** 228–233.

Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C. J. (1991). Characterization of the *groEL*-like genes in *Streptomyces albus*. *J. Bacteriol.* **173,** 7382–7386.

McLachlan, A. D. (1987). Gene duplication and the origin of repetitive protein structures. *Cold Spring Harbor Symp. Quant. Biol.* **52,** 411–420.

McMullin, T. W., and Hallberg, R. L. (1988). A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli groEL* gene. *Mol. Cell. Biol.* **8,** 371–380.

Mendoza, J. A., Rogers, E., Lorimer, G. H., and Horowitz, P. M. (1991). Chaperonins facilitate the *in vitro* folding of monomeric mitochondrial rhodanese. *J. Biol. Chem.* **266,** 13,044–13,049.

Musgrove, J. E., Johnson, R. A., and Ellis, R. J. (1987). Dissociation of the ribulosebisphosphate-carboxylase large-subunit binding protein into dissimilar subunits. *Eur. J. Biochem.* **163,** 529–534.

Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F.-U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature (London)* **341,** 125–130.

Prasad, T. K., and Hallberg, R. L. (1989). Identification and metabolic characterization of the *Zea mays* mitochondrial homolog of the *Escherichia coli* groEL protein. *Plant Mol. Biol.* **12,** 609–618.

Prasad, T. K., and Stewart, C. R. (1992). cDNA clones encoding *Arabidopsis thaliana* and *Zea mays* mitochondrial chaperonin HSP60 and gene expression during seed germination and heat shock. *Plant Mol. Biol.* **18,** 873–885.

Prasad, T. K., Hack, E., and Hallberg, R. L. (1990). Function of the maize mitochondrial chaperonin hsp60: Specific association between hsp60 and newly synthesized F1-ATPase alpha subunits. *Mol. Cell. Biol.* **10,** 3979–3986.

Pushkin, A. V., Tsuprun, V. L., Solovjeva, N. A., Shubin, V. V., Evstigneeva, Z. G., and Kretovich, W. L. (1982). High molecular weight pea leaf protein similar to the groE protein of *Escherichia coli. Biochim. Biophys. Acta* **704,** 379–384.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature (London)* **337,** 655–659.

Rospert, S., Glick, B. S., Jenö, P., Schatz, G., Todd, M. J., Lorimer, G. H., and Viitanen, P. V. (1993a). Identification and functional analysis of chaperonin 10, the groES homolog from yeast mitochondria. *Proc. Natl. Acad. Sci. USA* **90,** 10967–10971.

Rospert, S., Junne, T., Glick, B. S., and Schatz, G. (1993b). Cloning and disruption of the gene encoding yeast mitochondrial chaperonin 10, the homolog of *E. coli* groES. *FEBS Lett.* **335,** 358–360.

Roy, H., Bloom, M., Milos, P., and Monroe, M. (1982). Studies on the assembly of large subunits of ribulose bisphosphate carboxylase in isolated pea chloroplasts. *J. Cell Biol.* **94,** 20–27.

Roy, H., Hubbs, A., and Cannon, S. (1988). Stability and dissociation of the large subunit Rubisco binding protein complex *in vitro* and *in organello. Plant Physiol.* **86,** 50–53.

Saibil, H. R., Zheng, D., Roseman, A. M., Hunter, A. S., Watson, G. M. F., Chen, S., auf der Mauer, A., O'Hara, B. P., Wood, S. P., Mann, N. H., Barnett, L. K., and Ellis, R. J. (1993). ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Curr. Biol.* **3,** 265–273.

Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., and Blundell, T. L. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature (London)* **271,** 618–621.

Terlesky, K. C., and Tabita, F. R. (1991). Purification and characterization of the chaperonin 10 and chaperonin 60 proteins from *Rhodobacter sphaeroides. Biochemistry* **30,** 8181–8186.

Tilly, K., Murialdo, H., and Georgopoulos, C. (1981). Identification of a second *Escherichia coli groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl. Acad. Sci. USA* **78,** 1629–1633.

Torres-Ruiz, J. A., and McFadden, B. A. (1988). A homolog of ribulose bisphosphate carboxylase/oxygenase-binding protein in *Chromatium vinosum. Arch. Biochem. Biophys.* **261,** 196–204.

Torres-Ruiz, J. A., and McFadden, B. A. (1992). Purification and characterization of chaperonin 10 from *Chromatium vinosum. Arch. Biochem. Biophys.* **295,** 172–179.

Tsugeki, R., and Nishimura, M. (1993). Interaction of homologues of hsp70 and cpn60 with ferredoxin-NADP$^+$ reductase upon its import into chloroplasts. *FEBS Lett.* **320,** 198–202.

van der Vies, S. M., Bradley, D., and Gatenby, A. A. (1986). Assembly of cyanobacterial and higher plant ribulose bisphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in *Escherichia coli. EMBO J.* **5,** 2439–2444.

van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H., and Jaenicke, R. (1992). Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry* **31,** 3635–3644.

van der Vies, S. M., Gatenby, A. A., and Georgopoulos, C. (1994). Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature (London)* **368,** 654–656.

Van Dyk, T. K., Gatenby, A. A., and LaRossa, R. A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature (London)* **342,** 451–453.

Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P., and Lorimer, G. H. (1990). Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K^+ dependent. *Biochemistry* **29,** 5665–5671.

Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H., and Gatenby, A. A. (1991). Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. *Biochemistry* **30,** 9716–9723.

Viitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1,** 363–369.

Webb, R., Reddy, K. J., and Sherman, L. A. (1990). Regulation and sequence of the *Synechococcus* sp Strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. *J. Bacteriol.* **172,** 5079–5088.

Wu, H. B., Feist, G. L., and Hemmingsen, S. M. (1993). A modified *Escherichia coli* chaperonin (groEL) polypeptide synthesized in tobacco and targeted to the chloroplasts. *Plant Mol. Biol.* **22,** 1087–1100.

Zabaleta, E., Oropeza, A., Assad, N., Mandel, A., Salerno, G., and Herrera-Estrella, L. (1994). Antisense expression of chaperonin 60β in transgenic tobacco plants leads to abnormal phenotypes and altered distribution of photoassimilates. *Plant J.* **6,** 425–432.

Zeilstra-Ryalls, J., Fayet, O., and Georgopoulos, C. (1991). The universally conserved GroE (Hsp60) chaperonins. *Annu. Rev. Microbiol.* **45,** 301–325.

4

Chaperonin-Mediated Folding and Assembly of Proteins in Mitochondria

THOMAS LANGER AND WALTER NEUPERT

Institut für Physiologische Chemie
der Ludwig-Maximilians-Universität
80336 München, Germany

I. INTRODUCTION

Molecular chaperone proteins bind to nonnative protein structures which expose hydrophobic regions that are buried in the interior of a completely folded protein, and so prevent the aggregation of these

structures. Within a cell, these interactive surfaces are exposed during synthesis when polypeptide chains emerge from ribosomes and at later stages of the folding pathway. Organellar proteins face additional risks of engaging irreversible off-reactions as they have to traverse membranes to reach their final location. Membrane translocation requires an unfolded, "translocation-competent" conformation of polypeptide chains. This implies, on one hand, that cytosolic forms of the proteins have to be unfolded on membrane translocation or that folding has to be suppressed until the final location is reached. On the other hand, after membrane translocation has been achieved, folding of organellar proteins into the native conformation must occur. On the basis of these considerations it can be expected that the biogenesis of organellar proteins strongly depends on molecular chaperone proteins. Indeed, during recent years studies on intracellular protein sorting have revealed important insights into the mechanism of molecular chaperone function within the cell.

Mitochondria turned out to provide an extremely useful system in this context. The overwhelming majority of mitochondrial proteins are nuclear-encoded. After synthesis by polyribosomes in the cytosol, in most cases as precursor proteins with an N-terminal targeting sequence, they are imported into mitochondria and sorted into the various subcompartments, i.e., matrix space, inner and outer membrane, and intermembrane space. The post-translational import of proteins can be reconstituted *in vitro* using isolated mitochondria and radiolabeled precursor proteins. The availability of this powerful *in vitro* system has allowed a detailed characterization of the mitochondrial import machinery and has unravelled basic principles underlying the membrane translocation of proteins (Hannavy *et al.,* 1993; Kiebler *et al.,* 1993; Stuart *et al.,* 1994, for reviews). Most relevant for the topic of this review, the *in organello* translocation system allows us to analyze the folding of newly imported proteins in a "cellular" environment. In contrast to *in vivo* studies in intact cells, however, experimental conditions can be manipulated more easily and interfering effects of protein synthesis on the folding process can be excluded. Using this approach, direct experimental evidence was obtained that folding and assembly of newly imported proteins are mediated by molecular chaperones. Mitochondrial (mt) heat shock protein (hsp) 70 and the chaperonin (cpn) hsp60 (mt cpn60) were identified as central components of the folding machinery. Here we will review the current understanding of the function of mt cpn60. Other molecular chaperone proteins are dealt with only in respect to their functional interaction with mt cpn60. A detailed discussion of the role of chaperone proteins in mitochondrial biogenesis can be found elsewhere (Stuart *et al.,* 1994; Langer and Neupert, 1994).

II. MITOCHONDRIAL CHAPERONIN 60

Mitochondrial cpn60 belongs to a family of highly conserved proteins, termed chaperonins (Hemmingsen *et al.*, 1988; Ellis, 1993), that occur in prokaryotes and in mitochondria and chloroplasts of eukaryotic cells (McMullin and Hallberg, 1987, 1988; Jindal *et al.*, 1989; Mizzen *et al.*, 1989; Picketts *et al.*, 1989; Waldinger *et al.*, 1989; Barraclough and Ellis, 1980; Martel *et al.*, 1990). Mitochondrial cpn60 is encoded by an essential gene in yeast whose expression is induced two- to threefold upon heat shock (Reading *et al.*, 1989; Johnson *et al.*, 1989). Under these conditions mt cpn60 represents 1–2% of total mitochondrial protein. Several features define mt cpn60 as a chaperonin, including its tertiary structure and its ATPase activity, which is modulated by another protein called cpn10. Mitochondrial cpn60 is a homooligomeric protein consisting of 14 subunits of 60 kDa. These subunits are arranged in two heptameric layers forming a barrel-like structure (Hutchinson *et al.*, 1989). Recombinant mammalian mt cpn60 was found to form and be functional as a heptameric toroid, suggesting that the minimal functional unit of this chaperonin is a single ring (Viitanen *et al.*, 1992). This notion is in agreement with recent *in vitro* studies on the mechanism of GroEL function, the prokaryotic cpn60 homolog (Todd *et al.*, 1993; Martin *et al.*, 1993; see Chapters 7–9). As other chaperonins, mt cpn60 exhibits an ATPase activity by which the binding of substrate proteins is regulated. On ATP hydrolysis, release of bound substrate proteins occurs. Essential for the function of prokaryotic chaperonin GroEL is the coordination of the ATPase activities of its subunits by the cpn10 called GroES (Chandrasekhar *et al.*, 1986; Martin *et al.*, 1991; Todd *et al.*, 1993; Jackson *et al.*, 1993). Within mitochondria, proteins highly homologous to GroES, designated mt cpn10 or hsp10, have been identified in various organisms (Lubben *et al.*, 1990; Hartman *et al.*, 1992a,b, 1993; Rospert *et al.*, 1993a,b; Höhfeld and Hartl, 1994), indicating that the basic mechanism of chaperonin function is conserved from prokaryotes to mitochondria.

A. Mitochondrial Chaperonin 60-Mediated Assembly of Oligomeric Protein Complexes

First evidence for the function of mt cpn60 in the biogenesis of mitochondria was obtained by a classical genetic approach with yeast. The

yeast *mif4* mutant was identified by screening a library of temperature-sensitive mutants for a phenotype that results from a deficiency of assembly of mitochondrial proteins (Cheng *et al.*, 1989). It turned out that in this mutant the *HSP60* gene, which was identified in the yeast genome at the same time (Reading *et al.*, 1989), was affected. Subsequent biochemical characterization of the phenotype of the *mif4* mutant strain provided direct evidence for a function of mt cpn60 in the assembly of matrix-localized and inner membrane proteins. At nonpermissive temperatures mt cpn60 is functionally inactivated and becomes insoluble. Under these conditions newly imported β subunits of the F_1-ATPase failed to assemble into the F_1-ATPase complex and maturation of the Rieske Fe/S protein was impaired. A number of newly imported matrix-localized proteins form aggregates in the absence of functional mt cpn60, including ornithine transcarbamoylase, lipoamide dehydrogenase, and the β subunit of the matrix processing peptidase (Cheng *et al.*, 1990; Glick *et al.*, 1992). Interestingly, newly imported mt cpn60 subunits also were found to require preexisting functional mt cpn60 oligomers to assemble into tetradecameric complexes (Cheng *et al.*, 1990). These findings have been confirmed by genetic depletion experiments (Hallberg *et al.*, 1993). Yeast strains with a disrupted *HSP60* gene were rescued by expression of mt cpn60 from a galactose-inducible promotor. Growth of this strain on glucose medium resulted in subsequent depletion of mt cpn60 from mitochondria. In the virtual absence of mt cpn60 matrix-localized proteins were targeted to mitochondria and processed as in wild-type cells, but accumulated as insoluble aggregates.

These results point to a general role of mt cpn60 in the assembly of newly imported, matrix-localized proteins. In addition, mt cpn60 function appears to be required also for the assembly of proteins that are encoded by the mitochondrial genome and translated on mitochondrial ribosomes. The identification of a complex between mt cpn60 and the mitochondrially encoded α-subunit of the F_1-ATPase in maize mitochondria suggests that mt cpn60 plays a role in the assembly of this oligomeric complex (Prasad *et al.*, 1990). In the yeast *Saccharomyces cerevisiae*, mitochondrial DNA codes for eight proteins, which, with a single exception, are subunits of respiratory chain and ATP synthase complexes in the inner membrane. The only translation product that is soluble in the mitochondrial matrix space is the Var1p, a component of the small subunit of mitochondrial ribosomes. In contrast to wild-type cells, in *mif4* mutant mitochondria Var1p was found to aggregate under restrictive conditions (Horwich *et al.*, 1992). Although indirect effects of mt cpn60

on other ribosomal subunits cannot be excluded, this observation suggests a role of mt cpn60 in the assembly of mitochondrial ribosomes. The translation process itself appears not to be affected by mt cpn60 (Horwich *et al.*, 1992).

B. Mitochondrial Chaperonin 60-Mediated Polypeptide Folding

The experiments discussed so far established an essential role of mt cpn60 in the biogenesis of various mitochondrial proteins. Since in all cases oligomeric protein complexes are affected by a mt cpn60 dysfunction, however, it remained unclear on the basis of these studies whether mt cpn60 promotes the assembly of already folded subunits or whether polypeptide chain folding itself is mediated by mt cpn60. To address these questions, a fusion protein was imported into mitochondria isolated from *Neurospora crassa;* this protein consists of amino terminal regions of the precursor of ATP synthase subunit 9 (amino acids 1–69) fused to dihydrofolate reductase (DHFR) [pSu9(1–69)–DHFR; Ostermann *et al.*, 1989]. The intrinsic protease resistance of native DHFR can be exploited to monitor the folding state of the newly imported polypeptide chain from which the presequence (residues 1–66) has been removed by the matrix processing peptidase. Folding of the DHFR domain within mitochondria was found to be ATP-dependent and to occur at a considerable slower rate than the import reaction. At reduced ATP levels, which still allowed efficient import, folding of the DHFR domain was inhibited. Under these conditions newly translocated Su9(1–69)–DHFR was found in a high-molecular-weight complex with mt cpn60. On addition of MgATP, DHFR folded to the protease-resistant, native structure. Interestingly, after partial purification of the mt cpn60–DHFR complex by gel chromatography, DHFR remained associated with mt cpn60 upon addition of MgATP and exhibited a partial resistance toward added protease. This observation indicates that different conformations of a polypeptide chain during folding can be stabilized by mt cpn60 and that mt cpn60 mediates ATP-dependent folding of the bound DHFR moiety (Ostermann *et al.*, 1989). Indeed, various *in vitro* studies using prokaryotic cpn60 proteins have provided compelling evidence that at least partial folding of polypeptide chains can occur in association with chaperonins (Martin *et al.*, 1991; van der Vies *et al.*, 1992; Hayer-Hartl *et al.*, 1994; see Chapters 7 and 8). On the other hand, the partial inhibition

of DHFR folding after enrichment of the mt cpn60–DHFR complex points to a yet unidentified component required for mt cpn60 function. The mitochondrial chaperonin 10 (mt cpn10 or hsp10) is a likely candidate to cooperate with mt cpn60 in this reaction. However, in a mutant strain of *S. cerevisiae,* carrying a temperature-sensitive allele of *HSP10,* DHFR folding was not affected at a nonpermissive temperature (Höhfeld and Hartl, 1994; see below).

A more general function of mt cpn60 in the folding of newly imported proteins is suggested by the observation of complexes between several newly imported authentic mitochondrial proteins with mt cpn60. The β subunit of the F_1-ATPase, the Rieske Fe/S protein (Ostermann *et al.,* 1989), the α subunit of the matrix processing peptidase (Manning-Krieg *et al.,* 1991), and medium-chain acyl-CoA dehydrogenase (Saijo *et al.,* 1994) were detected in association with mt cpn60 using coimmunoprecipitation or native gel electrophoresis as a tool to demonstrate a direct physical interaction. Although folding of these polypeptide chains cannot be easily assessed under cellular conditions, it appears to be very likely that mt cpn60 mediates the folding rather than the assembly step. A detailed examination of the chaperonin-dependent folding and assembly of trimeric ornithine transcarbamylase (OTC) *in vitro* using the prokaryotic chaperonin system GroE revealed no evidence for a function of the chaperonin in the assembly of this protein (Zheng *et al.,* 1993). Rather, folded monomers of OTC dissociate from the chaperonin, which assemble spontaneously after release.

III. MITOCHONDRIAL CHAPERONIN 60 AS STRESS PROTEIN

Physiological stress results in denaturation of proteins within a cell. Under these conditions heat shock proteins, acting as molecular chaperones, are thought to diminish damage to the cell. Within mitochondria, at high temperatures a variety of proteins were observed to become associated with mt cpn60 (Martin *et al.,* 1992). Using DHFR as a relatively thermolabile model protein, mt cpn60 was shown to prevent protein denaturation under heat stress conditions in an ATP-dependent manner (Martin *et al.,* 1992). After import into mitochondria *in vivo,* DHFR was detected in a complex with mt cpn60 only when cells were grown at 37°C. In the absence of functional mt cpn60, heat treatment of cells resulted in aggregation of newly imported as well as preexisting DHFR.

As demonstrated *in vitro,* mt cpn60 stabilizes the native conformation of DHFR by repeated cycles of binding and release from the polypeptide chain (Martin *et al.,* 1993). MgATP and cpn10 are required for the maintenance of DHFR activity. Interestingly, *in vivo,* other mitochondrial chaperone proteins, e.g., mitochondrial hsp70 (mt hsp70), cannot replace mt cpn60 in the stabilization of DHFR at high temperatures. As suggested by studies of the thermal inactivation of firefly luciferase in *Escherichia coli* under heat stress, hsp70 proteins act in a manner different from that of cpn60 (Schröder *et al.,* 1993). Inactivation of proteins at high temperatures appears to be unaffected by hsp70. However, irreversible aggregation of (partially) unfolded polypeptide chains is prevented, allowing the hsp70-dependent reactivation under normal growth conditions. The function of mitochondrial hsp70 under heat stress remains to be determined.

IV. REGULATION OF MITOCHONDRIAL CHAPERONIN 60 FUNCTION BY MITOCHONDRIAL CHAPERONIN 10

Similar to other chaperonins, mitochondrial mt cpn60 function is modulated by mt cpn10 (hsp10). The principles of the functional interaction of both components were first recognized in *E. coli,* and subsequently exploited to identify mitochondrial homologs in various organisms including mammals and yeast (Lubben *et al.,* 1990; Hartman *et al.,* 1992a,b, 1993; Rospert *et al.,* 1993a,b; Höhfeld and Hartl, 1994). Mitochondrial cpn10s are highly conserved in both structural and functional terms. Yeast mt cpn10 is 36.5% identical in amino acid sequence to *E. coli* GroES and 43.6% identical to rat liver mt cpn10 (Höhfeldt and Hartl, 1994). Like the prokaryotic homolog, mitochondrial cpn10s were found to form homooligomeric complexes, most likely consisting of seven subunits. Under stress conditions, such as high temperatures or in the presence of amino acid analogs, expression of *HSP10* is induced, as is that of *HSP60* (Hartman *et al.,* 1992a; Höhfeld and Hartl, 1994).

All available data are consistent with a function of mt cpn10 in synchronizing the ATPase activities of individual subunits of mt cpn60, as demonstrated for the prokaryotic chaperonin system (Martin *et al.,* 1991; Gray and Fersht, 1991; Todd *et al.,* 1993; see Chapters 7 and 8). Yeast mt cpn60 exhibits an ATPase activity of about six ATP molecules/minute/protomer (Rospert *et al.,* 1993a). As with *E. coli* GroES, ATP hydrolysis

is stimulated by K^+ ions. Together with the observation that the prokaryotic chaperonin GroEL was found to functionally interact with cpn10s isolated from diverse organisms (Lubben *et al.*, 1990; Hartman *et al.*, 1992a; Rospert *et al.*, 1993a; Höhfeld and Hartl, 1994), these similarities strongly suggest a highly conserved mechanism of chaperonin action (Goloubinoff *et al.*, 1989). However, minor differences may exist. Whereas GroEL is active in folding assays with either the authentic bacterial or a heterologous mitochondrial cpn10, no complex formation was observed between mitochondrial mt cpn60 and the bacterial cpn10 (Viitanen *et al.*, 1992). Interestingly, although mitochondrial mt cpn60 and mt cpn10 from yeast promote folding of model proteins in *in vitro* assays, in contrast to the situation in *E. coli,* no inhibition of the ATPase activity of mt cpn60 was observed on incubation with the mt cpn10 (Rospert *et al.*, 1993a). This observation is consistent with studies using prokaryotic GroEL and GroES, which suggest that the coordination of the cpn60 subunits by cpn10 is important for the folding activity, rather than the inhibition of the ATPase activity.

In *S. cerevisiae* mt cpn10 is encoded by an essential gene (Rospert *et al.*, 1993a; Höhfeld and Hartl, 1994). Although at the very N-terminus positively charged and hydroxylated amino acids appear to be enriched, a characteristic of mitochondrial targeting sequences, no processing was observed upon import into isolated mitochondria. Evidence for the function of mt cpn10 under physiological conditions was obtained by analyzing the phenotype of *S. cerevisiae* strains carrying temperature-sensitive alleles of the *HSP60* gene. Sequencing of the alleles revealed point mutations in a highly conserved loop region at the N-terminus of mt cpn10 (Höhfeld and Hartl, 1994). In *E. coli* GroES, this region was shown to be exposed to solvent and required for stable complex formation with GroEL (Landry *et al.*, 1993). At a nonpermissive temperature cpn60 binding of mutant mt cpn10 was strongly reduced, again emphasizing the conservation of chaperonin function. Under these conditions, folding and assembly of newly imported proteins into the mitochondrial matrix were impaired. No trimeric ornithine transcarbamylase, which is formed in a mt cpn60-dependent manner (Cheng *et al.*, 1989), was detected in mt cpn10 mutant mitochondria. Also, α-MPP, a subunit of the matrix processing peptidase known to interact with mt cpn60 after import (Manning-Krieg *et al.*, 1991), did not assemble with β-subunits into the enzymatically active, dimeric structure, but aggregated in the absence of functional mt cpn10 (Höhfeld and Hartl, 1994). However, other model proteins show a different behavior. Mitochondrial cpn60-dependent fold-

ing of newly imported DHFR was not affected at a nonpermissive temperature in mt cpn10 mutant mitochondria (Höhfeld and Hartl, 1994). This observation is in agreement with *in vitro* folding experiments demonstrating that cpn10 is dispensable for the cpn60-mediated refolding of DHFR (Martin *et al.*, 1991; Viitanen *et al.*, 1991). It remains to be determined, however, whether mt cpn10 accelerates the mt cpn60-dependent refolding of newly imported DHFR at low temperatures.

V. ROLE OF MITOCHONDRIAL CHAPERONIN 60 MACHINERY IN INTRAMITOCHONDRIAL PROTEIN SORTING

Interestingly, at restrictive temperatures in mt cpn10 mutant mitochondria the maturation of intermembrane space proteins was also affected (Höhfeld and Martl, 1994). The Rieske Fe/S protein passes through the mitochondrial matrix, where it is processed twice and an intermediate is formed. Subsequently, sorting to the outer surface of the inner membrane occurs (Hartl *et al.*, 1986; Hartl and Neupert, 1990). At a nonpermissive temperature in mt cpn10 mutant mitochondria, and similarly at a restrictive temperature in the *mif4* mutant (Cheng *et al.*, 1989), maturation of the precursor of the Rieske Fe/S protein was affected (Höhfeld and Hartl, 1994). Mt cpn60 is thought to maintain the intermediate form of this protein in a conformation allowing its efficient sorting to the intermembrane space. It appears as if, at least at high temperatures, mt cpn10 function is crucial for the assembly of Rieske Fe/S protein into complex III of the inner membrane. This requirement may also provide an explanation for the earlier observation that the release of the Rieske Fe/S protein from mt cpn60 was impaired after partial purification of the mt cpn60–Fe/S protein complex (Ostermann *et al.*, 1989).

In agreement with the effect on the Rieske Fe/S protein, the maturation of another intermembrane space protein, cytochrome b_2, was also affected kinetically in mt cpn10 mutant mitochondria (Höhfeld and Hartl, 1994). This effect is particularly interesting in view of the current debate about the sorting pathway of this protein (Koll *et al.*, 1992; Glick *et al.*, 1992). At nonpermissive temperature in *mif4* mutant mitochondria, accumulation of the intermediate form of cytochrome b_2 was observed (Cheng *et al.*, 1989). In *in vitro* experiments the association of cytochrome b_2 with mt cpn60 was reported, and taken as additional evidence for the conservative sorting model, which predicts that cytochrome b_2 traverses

the matrix space on its route to the intermembrane space (Koll *et al.*, 1992; Hartl and Neupert, 1990). This conclusion, however, was recently challenged, on the basis of the observation of a different phenotype in *mif4* mutant mitochondria and of correct sorting of cytochrome b_2 in the virtual absence of mt cpn60 (Glick *et al.*, 1992; Hallberg *et al.*, 1993; Rospert *et al.*, 1994). Although the latter results suggest that mt cpn60 may not be essential for the sorting of cytochrome b_2, in the presence of mt cpn60 cytochrome b_2 may be stabilized by the chaperonin, at least under conditions that favor import over reexport. This stabilization might explain the occurrence of intermediate-sized cytochrome b_2 when synthesized in the absence of mt cpn60 (Hallberg *et al.*, 1993), and the observed kinetic effect of a conditional mutation in the *HSP10* gene on maturation of cytochrome b_2 (Höhfeld and Hartl, 1994).

VI. COOPERATION OF MITOCHONDRIAL CHAPERONIN 60 WITH MITOCHONDRIAL HEAT SHOCK PROTEIN 70 MACHINERY

Folding of newly imported proteins requires, in addition to mt cpn60, functional mt hsp70 (Fig. 1). Incoming polypeptide chains bind first to hsp70 in the mitochondrial matrix, which subsequently mediates, by cycles of ATP-dependent binding and release, their vectorial movement across the mitochondrial membranes (Stuart *et al.*, 1994, for review). In mitochondria containing a temperature-sensitive mutant of hsp70, *ssc1-2,* protein import is impaired under restrictive conditions (Kang *et al.*, 1990). The block of translocation, however, can be circumvented in *in vitro* experiments by denaturation of the precursor protein prior to import. Under these conditions, a fusion protein consisting of a matrix targeting sequence and mouse cytosolic DHFR is efficiently imported into mitochondria (Kang *et al.*, 1990). However, hsp70 remains associated with the newly imported polypeptide chain and stabilizes it in an unfolded conformation. Folding requires the ATP-dependent release from hsp70 and transfer to mt cpn60. A sequential action of hsp70 and mt cpn60 in the mitochondrial matrix during folding is also suggested by the observation that at a restrictive temperature in *mif4* mutant mitochondria, in contrast to *ssc1-2* mitochondria, protein import was not affected (Cheng *et al.*, 1989; Kang *et al.*, 1990). Indeed, various mitochondrial proteins, including α-MPP and medium-chain acyl-CoA dehydrogenase (Manning-Krieg *et al.*, 1991; Saijo *et al.*, 1994), could be coimmunoprecipitated

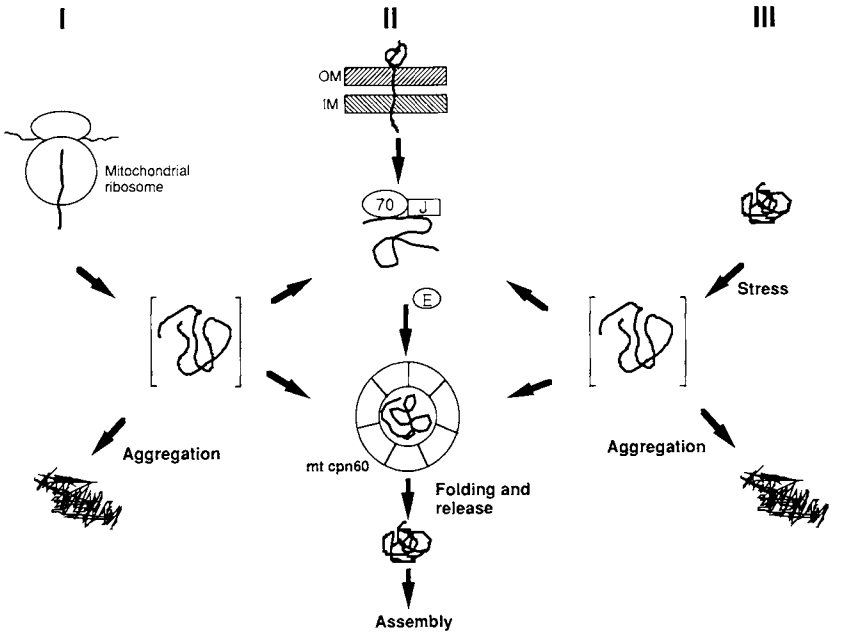


Fig. 1. Functions of mt cpn60 in mitochondrial biogenesis: (I) Folding of proteins newly synthesized within mitochondria. (II) Folding of newly imported proteins. (III) Prevention of heat denaturation of preexisting proteins in the mitochondrial matrix. See text for details. 70, mt hsp70; J, Mdj1p; E, Mge1p; OM, mitochondrial outer membrane; IM, mitochondrial inner membrane.

successively with hsp70 and mt cpn60. Interestingly, in mammalian cells both heat shock proteins can be coimmunoprecipitated with each other, suggesting a direct complex formation during transfer of the polypeptide chain (Mizzen *et al.*, 1991). In agreement with a sequential action, newly imported mt cpn60, which requires preexisting mt cpn60 for correct assembly (Cheng *et al.*, 1990), remains associated with hsp70 even in the presence of ATP in mt cpn60-depleted mitochondria (Hallberg *et al.*, 1993).

A detailed analysis of the folding reaction mediated by hsp70 and mt cpn60 was carried out *in vitro* using the corresponding highly homologous chaperones from *E. coli*, which are easily accessible from overproducing strains (Langer *et al.*, 1992). These studies revealed a

tight regulation of the functional cooperation of hsp70- and mt cpn60-like proteins. First, hsp70 and mt cpn60 differ from each other in their substrate specificity. Whereas hsp70 exhibits high affinity for polypeptide chains lacking secondary structure (Palleros *et al.*, 1991; Langer *et al.*, 1992), cpn60-type proteins stabilize folding intermediates in a compact conformation characterized by a disordered tertiary structure (Martin *et al.*, 1991; van der Vies *et al.*, 1992; Hayer-Hartl *et al.*, 1994). Second, the ATP-dependent interaction of a polypeptide chain with hsp70 is modulated by two other heat shock proteins. *E. coli* DnaJ and GrpE (Liberek *et al.*, 1991; Georgopoulos *et al.*, 1994). Homologs of these proteins, Mdj1p and Mge1p, respectively, have been identified in mitochondria (Rowley *et al.*, 1994; Ikeda *et al.*, 1994; Bollinger *et al.*, 1994; Laloraya *et al.*, 1994). In view of the high sequence similarity between the prokaryotic and the mitochondrial proteins, a conserved mode of action is an attractive possibility. Indeed, Mdj1p was found to be required for folding of newly imported and preexisting proteins (Rowley *et al.*, 1994), as is mitochondrial hsp70.

VII. PERSPECTIVES

Although the cooperation of hsp70 and mt cpn60 in folding of newly imported mitochondrial proteins is well documented, a number of questions remain unanswered. How general is this folding pathway? Is mt cpn60 only required for the folding of a subset of mitochondrial proteins, as is the prokaryotic homolog GroEL (Horwich *et al.*, 1992), and do other proteins fold spontaneously or are they assisted by other chaperone proteins? hsp70 was recently found to fulfill multiple functions during the biogenesis of mitochondria, including protein translocation, folding, assembly, and degradation. This discovery raises the intriguing question of whether mt cpn60 cooperates with the hsp70 machinery in all of these processes. Does hsp70 binding to the polypeptide chain precede mt cpn60 binding in all cases? Answers to these questions should provide a more detailed view on the mechanism of action of chaperonins and their functional interaction with other chaperone proteins *in vivo*. Again, mitochondria may turn out to represent a useful model system to unravel various aspects of chaperone function in the intact cell.

REFERENCES

Barraclough, R., and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX: Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. *Biochim. Biophys. Acta* **608,** 19–31.

Bollinger, L., Deloche, O., Glick, B. S., Georgopolous, C., Jenö, P., Kronidou, N., Horst, M., Morishima, N., and Schatz, G. (1994). A mitochondrial homolog of bacterial GrpE interacts with mitochondrial hsp70 and is essential for viability. *EMBO J.* **13,** 1998–2006.

Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* **261,** 12414–12419.

Cheng, M. Y., Hartl, F. U., Martin, J., Pollock, R. A., Kalousek, F., Neupert, W., Hallberg, E. M., Hallberg, R. L., and Horwich, A. L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature (London)* **337,** 620–625.

Cheng, M. Y., Hartl, F. U., and Horwich, A. L. (1990). The mitochondrial chaperonin hsp60 is required for its own assembly. *Nature (London)* **348,** 455–458.

Ellis, R. J. (1993). The general concept of molecular chaperones. *Phil. Trans. R. Soc. London B* **339,** 257–261.

Georgopoulos, C., Liberek, K., Zylicz, M., and Ang., D. (1994). Properties of heat shock proteins of *Escherichia coli* and the autoregulation of the heat shock response. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp. 209–249. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Glick, B. S., Brandt, A., Cunningham, K., Muller, S., Hallberg, R. L., and Schatz, G. (1992). Cytochromes c_1 and b_2 are sorted to the intermembrane space of yeast mitochondria by a stop-transfer mechanism. *Cell* **69,** 809–822.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature (London)* **342,** 884–889.

Gray, T. E., and Fersht, A. R. (1991). Cooperativity in ATP hydrolysis by GroEL is increased by GroES. *FEBS Lett.* **292,** 254–258.

Hallberg, E. M., Shu, Y., and Hallberg, R. L. (1993). Loss of mitochondrial hsp60 function: Nonequivalent effects on matrix-targeted and intermembrane-targeted proteins. *Mol. Cell. Biol.* **13,** 3050–3057.

Hannavy, K., Rospert, S., and Schatz, G. (1993). Protein import into mitochondria: A paradigm for the translocation of polypeptides across membranes. *Curr. Biol.* **5,** 694–700.

Hartl, F.-U., and Neupert, W. (1990). Protein sorting to mitochondria: Evolutionary conservations of folding and assembly. *Science* **247,** 930–938.

Hartl, F.-U., Schmidt, B., Wachter, E., Weiss, H., and Neupert, W. (1986). Transport into mitochondria and intramitochondrial sorting of the Fe/S protein of ubiquinol-cytochrome c reductase. *Cell* **47,** 939–951.

Hartman, D. J., Hoogenraad, N. J., and Hoj, P. B. (1992a). Heat shock proteins of barley mitochondria and chloroplast. Identification of organellar hsp10 and 12: Putative chaperonin 10 homologues. *FEBS Lett.* **305,** 147–150.

Hartman, D. J., Hoogenraad, N. J., Condron, R., and Hoj, P. B. (1992b). Identification of a mammalian 10-kDa heat shock protein, a mitochondrial chaperonin 10 homologue essential for assisted folding of trimeric ornithine transcarbamoylase *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 3394–3398.

Hartman, D. J., Hoogenraad, N. J., Condron, R., and Hoj, P. B. (1993). The complete primary structure of rat chaperonin 10 reveals a putative $\beta\alpha\beta$ nucleotide-binding domain with homology to $p21^{ras}$. *Biochim. Biophys. Acta* **1164,** 219–222.

Hayer-Hartl, M. Ewbank, J. J., Creighton, T. E., and Hartl, F.-U (1994). Conformational specificity of the chaperonin GroEL for the compact folding intermediates of α-lactalbumin. *EMBO J.* **13,** 3192–3202.

Hemmingsen, S. M., Woolford, C., Van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* (*London*) **333,** 330–334.

Höhfeld, J., and Hartl, F.-U. (1994). Requirement of the chaperonin cofactor HSP10 for protein folding and sorting in yeast mitochondria. *J. Cell Biol.* **126,** 305.

Horwich, A., Caplan, S., Wall, J. S., and Hartl, F.-U. (1992). Chaperonin-mediated protein folding. *In* "Membrane Biogenesis and Protein Targeting" (W. Neupert and R. Lill, eds.), pp. 329–337. Elsevier, Amsterdam.

Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H., and Leonard, K. R. (1989). Identification and electron microscopic analysis of a chaperonin oligomer from *Neurospora crassa* mitochondria. *EMBO J.* **8,** 1485–1490.

Ikeda, E., Yoshida, S., Mitsuzawa, H., Uno, I., and Toh-e, A. (1994). *YGE1* is a yeast homologue of *Escherichia coli grpE* and is required for maintenance of mitochondrial functions. *FEBS Lett.* **339,** 265–268.

Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbrook, J. J., Clarke, A. R., and Burston, S. G. (1993). Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: Implications for the mechanism of assisted protein folding. *Biochemistry* **32,** 2554–2563.

Jindal, S., Dudani, A. K., Singh, B., Harley, C. B., and Gupta, R. S. (1989). Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65-kilodalton mycobacterial antigen. *Mol. Cell. Biol.* **9,** 2279–2283.

Johnson, C., Chandrasekhar, G. N., and Georgopoulos, C. (1989). *Escherichia coli* DnaK and GrpE heat shock proteins interact both *in vivo* and *in vitro*. *J. Bacteriol.* **171,** 1590–1596.

Kang, P. J., Ostermann, J., Shilling, J., Neupert, W., Craig, E. A., and Pfanner, N. (1990). Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. *Nature* (*London*) **348,** 137–143.

Kiebler, M., Becker, K., Pfanner, N., and Neupert, W. (1993). Mitochondrial protein import: Specific recognition and membrane translocation of preproteins. *J. Membrane Biol.* **135,** 191–207.

Koll, H., Guiard, B., Rassow, J., Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F.-U. (1992). Antifolding activity of hsp60 couples protein import into the mitochondrial matrix with export to the intermembrane space. *Cell* **68,** 1163–1175.

Laloraya, S., Gambill, B. D., and Craig, E. A. (1994). A role for a eukaryotic GrpE-related protein, Mge1p, in protein translocation. *Proc. Natl. Acad. Sci. USA* **91,** 6481–6485.

Landry, S., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., and Gierasch, L. M. (1993). Characterization of a functionally important mobile domain of GroES. *Nature* (*London*) **364,** 255–258.

Langer, T., and Neupert, W. (1994). Chaperoning mitochondrial biogenesis. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp. 53–83. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., and Hartl, F.-U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature (London)* **356,** 683–689.

Liberek, K., Marszalek, J., Ang., D., Georgopoulos, C., and Zylicz, M. (1991). *Escherichia coli* DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc. Natl. Acad. Sci. USA* **88,** 2874–2878.

Lubben, T. H., Gatenby, A. A., Donaldson, G. K., Lorimer, G. H., and Viitanen, P. V. (1990). Identification of a groES-like chaperonin in mitochondria that facilitates protein folding. *Proc. Natl. Acad. Sci. USA* **87,** 7683–7687.

Manning-Krieg, U., Scherer, P. E., and Schatz, G. (1991). Sequential action of mitochondria chaperones in protein import into the matrix. *EMBO J.* **10,** 3273–3280.

Martel, R., Cloney, L. P., Pelcher, L. E., and Hemmingsen, S. M. (1990). Unique composition of plastid chaperonin-60: Alpha and beta polypeptide-encoding genes are highly divergent. *Gene* **94,** 181–187.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. *Nature (London)* **352,** 36–42.

Martin, J., Horwich, A. L., and Hartl, F.-U. (1992). Prevention of protein denaturation under heat stress by the chaperonin hsp60. *Science* **258,** 995–998.

Martin, J., Mayhew, M., Langer, T., and Hartl, F.-U. (1993). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature (London)* **366,** 228–233.

McMullen, T. W., and Hallberg, R. L. (1987). A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **7,** 4414–4423.

McMullen, T. W., and Hallberg, R. L. (1988). A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli* groEL gene. *Mol. Cell. Biol.* **8,** 371–380.

Mizzen, L. A., Chang, C., Garrels, J. I., and Welch, W. J. (1989). Identification, characterization, and purification of two mammalian stress proteins present in mitochondria, grp 75, a member of the hsp 70 family and hsp 58, a homolog of the bacterial groEL protein. *J. Biol. Chem.* **264,** 20,664–20,675.

Mizzen, L. A., Kabiling, A. N., and Welch, W. J. (1991). The two mammalian mitochondrial stress proteins, grp 75 and hsp 58, transiently interact with newly synthesized mitochondrial proteins. *Cell Regul.* **2,** 165–179.

Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F.-U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis *Nature (London)* **341,** 125–130.

Palleros, D. R., Welch, W. J., and Fink, A. L. (1991). Interaction of hsp70 with unfolded proteins: Effects of temperature and nucleotides on the kinetics of binding. *Proc. Natl. Acad. Sci. USA* **88,** 5719–5723.

Picketts, D. J., Mayanil, C. S., and Gupta, R. S. (1989). Molecular cloning of a Chinese hamster mitochondrial protein related to the "chaperonin" family of bacterial and plant proteins. *J. Biol. Chem.* **264,** 12,001–12,008.

Prasad, T. K., Hack, E., and Hallberg, R. L. (1990). Function of the maize mitochondrial chaperonin hsp60: Specific association between hsp60 and newly synthesized F_1-ATPase alpha subunits. *Mol. Cell. Biol.* **10,** 3979–3986.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature (London)* **337,** 655–659.

Rospert, S., Glick, B. S., Jenö, P., Schatz, G., Todd, M. J., Lorimer, G. H., and Viitanen, P. V. (1993a). Identification and functional analysis of chaperonin 10, the groES homolog from yeast mitochondria. *Proc. Natl. Acad. Sci. USA* **90,** 10,967–10,971.

Rospert, S., Junne, T., Glick, B. S., and Schatz, G. (1993b). Cloning and disruption of the gene encoding yeast mitochondrial chaperonin 10, the homolog of the *E. coli* groES. *FEBS Lett.* **335,** 358–360.

Rospert, S., Müller, S., Schatz, G., and Glick, B. S. (1994). Fusion proteins containing the cytochrome b_2 presequence are sorted to the mitochondrial intermembrane space independently of hsp60. *J. Biol. Chem.* **269,** 17279–17288.

Rowley, N., Prip-Buus, C., Westermann, B., Brown, C., Schwarz, E., Barrell, B., and Neupert, W. (1994). Mdj1p, a novel chaperone of the DnaJ family, is involved in mitochondrial biogenesis and protein folding. *Cell* **77,** 249–259.

Saijo, T., Welch, W. J., and Tanaka, K. (1994). Intramitochondrial folding and assembly of medium-chain acyl-CoA dehydrogenase (MCAD). *J. Biol. Chem.* **269,** 4401–4408.

Schröder, H., Langer, T., Hartl, F.-U., and Bukau, B. (1993). DnaK, DnaJ, and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J.* **12,** 4137–4144.

Stuart, R. A., Cyr, D. M., Craig, E. A., and Neupert, W. (1994). Mitochondrial molecular chaperones: Their role in protein translocation. *Trends Biochem. Sci.* **19,** 87–92.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1993). Hydrolysis of adenosine 5′-triphosphate by *Escherichia coli* GroEL: Effects of GroES and potassium ion. *Biochemistry* **32,** 8560–8567.

van der Vies, S. M., Viitanen, P. V., Gatenby, A. A. Lorimer, G., and Jaenicke, R. (1992). Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry* **31,** 3635–3644.

Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H., and Gatenby, A. A. (1991). Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. *Biochemistry* **30,** 9716–9723.

Viitanen, P. V., Lorimer, G. H., Seetharam, R., Gupta, R. S., Oppenheim, J., Thomas, J. O., and Cowan, N. J. (1992). Mammalian mitochondrial chaperonin 60 functions as a single toroidal ring. *J. Biol. Chem.* **267,** 695–698.

Waldinger, D., Subramanian, A. R., and Cleve, H. (1989). The polymorphic human chaperonin protein HuCha60 is a mitochondrial protein sensitive to heat shock and cell transformation. *Eur. J. Cell Biol.* **50,** 435–441.

Zheng, X., Rosenberg, L. E., Kalousek, F., and Fenton, W. A. (1993). GroEL, GroES, and ATP-dependent folding and spontaneous assembly of ornithine transcarbamylase. *J. Biol. Chem.* **268,** 7489–7493.

5

Structure and Function of Chaperonins in Archaebacteria and Eukaryotic Cytosol

KEITH R. WILLISON* AND ARTHUR L. HORWICH†

* CRC Centre for Cell and Molecular Biology
Institute of Cancer Research
Chester Beatty Laboratories
London SW3 6JB, United Kingdom

† Howard Hughes Medical Institute
Boyer Center for Molecular Medicine
Yale University School of Medicine
New Haven, Connecticut 06536

I. INTRODUCTION

One exciting development in the study of chaperonins has been the discovery of a second family of chaperonins distinct from GroEl/hsp60/RBP in thermophilic archaebacteria and the eukaryotic cytosol. Thermophilic factor 55 (TF55) in thermophilic archaebacteria exhibits the structural and functional hallmarks of the bacterial and organellar chaperonins, i.e., double-ring structure, ATPase activity, and an ability to bind nonnative proteins but the primary structure of the TF55 chaperonins has only weak resemblance of heat shock protein (hsp)60 members, confined mainly to the region corresponding in GroEL to the ATPase domain. Strikingly, the primary structure of TF55 bears a much closer relationship (40% identity overall) to a protein of the eukaryotic cytosol whose function had previously been unknown, *t*-complex polypeptide 1 (TCP-1). When subsequent studies established that TCP-1 and a collection of at least eight structurally related subunits comprise a double-ring structure that also exhibits ATPase activity and ability to assist folding of at least actin and tubulin, it became unmistakably clear that a second family of chaperonins had evolved (Horwich and Willison, 1993; Willison and Kubota, 1994; Kubota *et al.,* 1995b). The lines of evolution of these two families of chaperonin seem comprehensible in the light of more general understanding of evolution; since components of the eukaryotic cytosol bear a closer relation to archaebacterial relatives than to prokaryotic homologs, it seems understandable that there is a closer relation between TF55 and TCP-1-related subunits. Likewise an endosymbiotic origin of organelles, mitochondria, and chloroplasts, from prokaryotic ancestors, accounts for the origin and distinct identity of the GroEL/hsp60/RBP family (Table I).

Studies have made clear that there are features of the TF55/TCP-1 chaperonins that are different from those of the GroEL family, particularly in the case of TCP-1 complex, sharpening interest in understanding the structure and mechanics of action. (i) Only a restricted range of substrates has been identified for TCP-1 complex, most notably actin, tubulin, and homologs. A broad role in assisting folding, as in, for example, GroEL, has not been demonstrated. Although studies *in vitro* have shown binding of TCP-1 complex to neurofilament proteins (Roobol and Carden, 1993), hepatitis virus capsid proteins (Lingappa *et al.,* 1994), and chromaffin granule membranes (Creutz *et al.,* 1994) (see Table III), it seems possible that TCP-1 complex evolved solely to cope with

TABLE I

Identification of Chaperonins[a]

Chaperonin	Organism	Localization	Amino acid identity to		Rotational symmetry	Subunit species	Refs.[d]
			GroEL	TF55			
Group I[b]							
GroEL	Eubacteria	Soluble	—	Weak	7	1	*1–7*
hsp60	Eukaryotes	Mitochondria	50%	Weak	7	1	*8–10*
RBP	Plants	Chloroplasts/plastids	50%	Weak	7	2	*2, 11*
Group II[c]							
TF55	Archaebacteria (*S. shibatae, S. solfataricus*)	Soluble	Weak	—	9	2	*12–16*
Thermosome	Archaebacteria (*P. occultum, P. brockii, T. acidophilum, Archaeglobus*)	Soluble	Weak	60%	8	2	*17–19*
CCT	Eukaryotes	Cytosol/nucleus	Weak	40%	8	7–9	*20–27*

[a] Modified from Kubota *et al.* (1995b).

[b] Subfamily of eubacterial chaperonins and eukaryotic intraorganellar chaperonins.

[c] Subfamily of archaebacterial chaperonins and eukaryotic cytosolic chaperonins.

[d] *Key to references:* (*1*) Georgopoulos *et al.* (1973); (*2*) Hemmingsen *et al.* (1988); (*3*) Martin *et al.* (1991); (*4*) Weissman *et al.* (1994); (*5*) Chen, S., *et al.* (1994); (*6*) Braig *et al.* (1994); (*7*) Fenton *et al.* (1994); (*8*) Cheng *et al.* (1989); (*9*) Reading *et al.* (1989); (*10*) Koll *et al.* (1992); (*11*) Martel *et al.* (1990); (*12*) Trent *et al.* (1991); (*13*) Kagawa *et al.* (1995); (*14*) Knapp *et al.* (1994); (*15*) Marco *et al.* (1994b); (*16*) Guagliardi *et al.* (1994); (*17*) Phipps *et al.* (1991); (*18*) Phipps *et al.* (1993); (*19*) Waldmann *et al.* (1995); (*20*) Lewis *et al.* (1992); (*21*) Yaffe *et al.* (1992); (*22*) Gao *et al.* (1992); (*23*) Frydman *et al.* (1992); (*24*) Kubota *et al.* (1994); (*25*) Joly *et al.* (1994b); (*26*) Marco *et al.* (1994a); (*27*) Tian *et al.* (1995).

the folding of cytoskeletal substrates and is a specialized chaperonin. (ii) TCP-1 complex is heterooligomeric in composition, with potentially eight or nine different subunits in each ring of the double ring assembly, although eightfold rotational symmetry is the averaged value determined by electron microscopy (Marco *et al.*, 1994a). Although it may be in general a single heterooligomeric assembly, it has been observed to be heterogeneous in subunit composition under certain conditions. The nature of the functional purpose to such heterogeneity remains unclear, but it seems possible that heterogeneity evolved to handle recognition and productive release of specific substrates. (iii) Neither TF55 nor TCP-1 complex appears to require a cooperating GroES-like chaperone for assisting productive folding. Nevertheless, in the case of the substrate, β-tubulin, it appears that two cofactors play a key role in production of the native state in cooperation with TCP-1 complex. (iv) TCP-1 complex is not stress-induced and whereas its abundance in testis, leukocytes, and some tissues such as lung epithelium may be similar to that of GroEL, in other mammalian tissues and in *Saccharomyces cerevisiae,* for example, its abundance appears to be reduced by at least one order of magnitude compared with GroEL.

This chapter first discusses the general features of thermophilic archaebacterial chaperonins, then the biochemistry of TCP-1 complex (also termed CCT, for chaperonin-containing TCP-1, or TRiC for TCP-1 ring complex), and the nature of its heterogeneity. A section on analyses in yeast follows. The chapter ends with a discussion of the evolution of archaebacteria, CCT, actins, and tubulins.

II. ARCHAEBACTERIAL CHAPERONINS

A. Thermophilic Factor 55

Chaperonins have been studied in detail in three species of thermophilic archaebacteria: *Sulfolobus shibatae* (Trent *et al.*, 1991; Kagawa *et al.*, 1995), *Sulfolobus solfataricus* (Knapp *et al.*, 1994; Marco *et al.*, 1994a), and *Pyrodictium occultum* (Phipps *et al.*, 1991, 1993). Thermophilic factor 55 was originally detected as virtually the only protein synthesized in thermophilic archaebacteria under conditions of heat shock (Trent *et al.*, 1990). In particular, it was the dominant protein produced when cells

of *S. shibatae* were shifted from a normal growth temperature of 68°C to a near lethal temperature of 88°C. Following such incubation, cells could survive at a temperature that is otherwise lethal, i.e., 95°C. Because the prior expression of TF55 could be correlated with this acquisition of thermotolerance, a role in stabilizing cellular components against thermal inactivation was suggested. Classically, such action by heat-inducible proteins has involved stabilization of other proteins against misfolding and aggregation, an action characteristic of molecular chaperones. Observations *in vitro* that purified TF55 from two species of *Sulfolobus* and *Pyrodictium occultum* exhibited ATPase activity (Trent *et al.*, 1991; Phipps *et al.*, 1991; Knapp *et al.*, 1994), and that the former can selectively bind mesophilic proteins at temperatures where inactivation occurs (Trent *et al.*, 1991), provide a strong indication that TF55 functions as a molecular chaperone in the cell. Indeed, the observation that TF55 is abundant even under normal growth conditions suggested that it most likely functions as a chaperone under normal conditions, and Guagliardi *et al.* (1994) have shown interaction of a number of thermophilic enzymes with TF55.

With a basal abundance of a least 1 to 2% of total soluble protein, the abundance of TF55 resembles that of GroEL in the bacterial cytoplasm. Likewise, its quaternary structure of two stacked oligomeric rings (Phipps *et al.*, 1991, 1993; Trent *et al.*, 1991; Knapp *et al.*, 1994; Marco *et al.*, 1994b) and biochemical properties of ATPase activity and ability to bind nonnative proteins fit the definition of a chaperonin. Since GroEL has been shown to be involved in folding of at least 40% of newly translated soluble proteins in *Escherichia coli* (Viitanen *et al.*, 1992; Horwich *et al.*, 1993), it is conceivable that TF55, with its similar abundance, may also have a broad role in mediating folding of many newly translated proteins in thermophilic archaebacteria.

Thermophilic factor 55 may also have a role in protecting already native proteins from thermal inactivation. Guagliardi *et al.* (1994) have carried out experiments *in vitro* with TF55 from *S. solfataricus,* showing that its coincubation with several thermophilic enzymes could prevent them from inactivation at temperatures of nearly 100°C. Thermoprotection was associated with binding of nonnative forms and addition of ATP, after downshift of temperature led to release of biologically active protein. A similar protective role has been previously suggested for hsp60 in protecting some proteins inside the mitochondrial matrix from thermal inactivation (Martin *et al.*, 1992). Unfortunately, the extent of involvement of TF55 *in vivo* with newly translated versus already-native

proteins cannot be directly tested since means of genetically altering thermophilic archaebacteria, needed for an *in vivo* test, are lacking.

The original examination of a TF55 homolog from *P. occultum,* termed the thermosome, identified two constituent subunits to this complex, 56 and 59 kDa in size, composing two stacked eight-member rings (Phipps *et al.,* 1991). Studies of TF55 from both *S. solfataricus* and *S. shibatae* now make clear that chaperonins from these organisms are also composed of what appear to be equimolar amounts of two distinct, closely related, subunits (Knapp *et al.,* 1994; Kagawa *et al.,* 1995). Each subunit species could compose an individual homooligomeric ring, making a complex of two distinct rings, or it is possible that the rings have a mixed composition. Indeed this mixed arrangement was proposed originally for the eight-member rings of *Pyrodictium* chaperonin, but in the case of *S. shibatae* and *S. solfataricus* TF55, which have nine-member rings (Knapp *et al.,* 1994; Marco *et al.,* 1994a), a mixed composition would clearly not allow for strict symmetry within the rings.

B. Thermosome as Immediate Ancestor of *t*-Complex Polypeptide 1

Baumeister and co-workers have structurally characterized TF55 chaperonin homologues from *P. occultum, P. brockii, Thermoplasma acidophilum,* and *Archaeoglobus,* termed thermosomes (Phipps *et al.,* 1991, 1993). In all four of these species, in contrast with *S. shibatae,* eightfold symmetrical rings were observed. The ATPase activity and subunit composition of the *P. occultum* thermosome were analyzed in some detail. The ATPase activity is very temperature-dependent, with maximum activity found at 100°C. The K_m for ATP at 90°C, pH 7.5, was 5.6 μM, with a turnover rate of 5 s^{-1}, an affinity and turnover rate at least an order of magnitude greater than that of GroEL at 37°C. The thermosome complex from *T. acidophilum* is composed of subunits of 58 and 60 kDa that are 60% identical to each other (Waldmann *et al.,* 1995). The study of these components has important ramifications for the understanding of the evolutionary origin of CCT. Because CCT has eightfold symmetry (Marco *et al.,* 1994a) and all its structurally related subunits show around 40% identity to TF55 (Kubota *et al.,* 1994), it seems likely that CCT evolved from an archaebacterium of the *Pyrodictium* type with an eightfold symmetrical composition.

III. *t*-COMPLEX POLYPEPTIDE 1 IN EUKARYOTIC CYTOSOL

A. Composition of CCT or Chaperonin-Containing TCP-1

When the primary structure relationship between TF55 in thermophilic archaebacteria and TCP-1 protein in the eukaryotic cytosol was initially detected (Trent *et al.,* 1991), it seemed immediately possible that a GroEL equivalent had been identified in the eukaryotic cytosol. After all, TCP-1 had been shown previously to be highly conserved in eukaryotes (Willison *et al.,* 1986, 1987; Ursic and Ganetsky, 1988) and to be an essential gene in yeast (Ursic and Culbertson, 1991). Indeed, weak homology with GroEL chaperonins had been intimated (Gupta, 1990; Ellis, 1990). Investigations, particularly studies *in vitro* (Section III.D), leave little question that the TCP-1 complex functions as a chaperonin, but findings concerning the relative abundance of TCP-1 subunits, heterogeneity of the subunits, and substrate specificity leave a broad role, equivalent to that of GroEL, as an unresolved question.

t-complex polypeptide protein was originally identified as a product of the mouse T locus (Silver *et al.,* 1979) but more recently has been recognized as but one of a set of related proteins whose genes are dispersed throughout the genome (Kubota *et al.,* 1994, 1995a). Combined with the related subunits, the abundance of the TCP-1-related proteins in total appears to be somewhat less than that of GroEL in the basal state, making up 0.1 to 0.3% of soluble cytosolic protein, as determined in lysates of mammalian testis and reticulocyte, and as measured in *S. cerevisiae* (Lewis *et al.,* 1992; Miklos *et al.,* 1994). Most cells contain between 6×10^4 and 3×10^5 CCT complexes (Lewis *et al.,* 1992; Kubota *et al.,* 1995b). This abundance appears to be at least one order of magnitude less than the collective of hsp70 class proteins in the eukaryotic cytosol. Also, neither the TCP-1 protein, now termed TCP-1α/CCTα, nor its relatives are heat-inducible (Lewis *et al.,* 1992; Ursic and Culbertson, 1991). In animal and yeast systems, within any given cell type, the different TCP-1-related subunits appear to be present at similar levels (Lewis *et al.,* 1992; Frydman *et al.,* 1992; Kubota *et al.,* 1994). That is, no single subunit or group of subunits is substantially more abundant than any other, but there is clearly variation in relative levels of certain subunits in different preparations. This observation is most consistent with formation of a single type of heterooligomeric ring, but it does not exclude the possibility of formation of roughly equal amounts of various

homooligomeric or heterogeneously composed complexes. There are indications that complexes containing different combinations exist in cells (see Kubota *et al.*, 1994; Roobol *et al.*, 1995), but none have been biochemically characterized so far. We think it likely that a "core" CCT complex, heterooligomeric in composition, contributes the main CCT assembly in most cells, but it is also likely that other CCT subunit-containing complexes exist (K. R. Willison *et al.*, unpublished results).

The subunit TCP-1α/CCTα (Table II) has been examined in the greatest number of cell types with respect to abundance. In general, TCP-1α abundance seems to correlate with, if anything, the abundance in a given cell type of the structural component, tubulin, with greater levels found in *Schizosaccharomyces pombe* than *S. cerevisiae* (Horwich and Willison, 1993) and in mammalian testis, compared with other tissues (Silver *et al.*, 1979; Lewis *et al.*, 1992). The abundance of TCP-1α in mouse tissue varies considerably, with the mRNA raised more than 20-fold in testis versus liver (Dudley *et al.*, 1984). The TCP-1α/CCTα is abundantly expressed in lung epithelial cells, which also have a high content of actin and tubulin. Involvement in tubulin biogenesis is supported both by studies of mutant *S. cerevisiae* strains bearing either cold-sensitive (Ursic and Culbertson, 1991) or temperature-sensitive (Miklos *et al.*, 1994; Ursic *et al.*, 1994) mutations in TCP-1α and by studies in mammalian cells (Sternlicht *et al.*, 1993). The CCTγ gene of the protozoan *Tetrahymena* has been shown to be coinduced during tubulin biosynthesis required for regrowth of cilia (Soares *et al.*, 1994). Sun *et al.* (1995) have examined the expression of TCP-1α/CCTα mRNA during early development of the axolotl (*Ambystoma mexicanum*) and find that the TCP-1α/CCTα transcript level is regulated temporally and spatially during embryogenesis, especially in neural and somitic development. This regulation correlates with an abundance of tubulin in neural tissues and of actin proteins in somites. At the later tailbud stages, the anterior/posterior pattern of TCP-1α/CCTα correlates with the anterior/posterior expression of actin and myosin protein patterns. Such observations suggest that CCT is important during morphogenesis, particularly during neurogenesis, in regulating the folding and assembly of cytoskeletal components.

Roobol *et al.* (1995) have examined the behavior of CCT during nerve growth factor (NGF)-induced differentiation of the ND7/23 cell line, a hybrid formed from neonatal rat dorsal root ganglion and mouse neuroblastoma. The TCP-1α/CCTα subunit enters growing neuritic processes and is particularly noticeable at the leading edge of growth cone-like structures where it colocalizes with actin. Three other components

of CCT (CCTβ, ε, and γ), however, remain predominantly restricted to perikaryal cytoplasm. These findings suggest a heterogeneous population of chaperonin particles within single differentiated ND7/23 cells. This heterogeneity may reflect specialization of chaperonin function within different cytoplasmic compartments of an individual cell. These observations also underscore that the action of CCT-containing chaperonin molecules is at least in part likely to be post-translational, since ribosomes do not enter the growing neurites.

B. CCT-Mediated Folding: Recognition of Substrates

Experiments carried out with CCT *in vitro,* as with GroEL, have been used to study the mechanism of folding mediated by the chaperonin. As with GroEL, an action associated with rounds of polypeptide binding and ATP-driven release/folding has been identified (Gao *et al.,* 1994; Tian *et al.,* 1995). Distinctly missing from the action of both TF55 and CCT, however, is cooperation with a GroES-like ring. In particular, no small protein coinduced with TF55 has been observed in thermophilic archaebacteria, and similarly, in the case of CCT, a small associated eight- or nine-member oligomer has not been detected, although two factors have been identified as cooperating in the folding of β-tubulin (see Section III,E).

Binding of proteins in nonnative form to CCT (see Table III) has been observed *in vitro,* and in general the specificity for substrate binding *in vitro* appears to correlate with that observed *in vivo* (Gao *et al.,* 1992; Yaffe *et al.,* 1992; Frydman *et al.,* 1992; Sternlicht *et al.,* 1993; Hynes *et al.,* 1995). Binding has been observed in a number of *in vitro* systems. In reticulocyte lysate, newly synthesized actin (Gao *et al.,* 1992), tubulin (Yaffe *et al.,* 1992; Gao *et al.,* 1993; Frydman *et al.,* 1992), luciferase (Frydman *et al.,* 1992), and transducin (G. Farr, unpublished) can be observed to associate with CCT. This interaction was observable when ATP-directed release is blocked by addition after translation of either apyrase or EDTA (which releases nucleotide from chaperonin).

It seems probable that peptide binding *in vivo* follows initial interaction of nascent peptide chains with hsp70 class chaperones. This possibility has been supported both by studies with intact yeast and by studies in the reticulocyte lysate system. Nelson *et al.* (1992) observed that deficiency of the two yeast SSB class hsp70 proteins leads to slow growth, suppressible

TABLE II

Correspondence between Subunit Proteins and Genes of CCT[a]

Source	Protein							
	CCTα	CCTβ	CCTγ	CCTδ	CCTε	CCTζ	CCTη	CCTθ
Mouse testis[b]	S3	S4	S5	S9	S2	S7	S8	S1
Bovine testis[c]	P4	P1	P5	P3	—	—	P2	—
Rabbit reticulocyte[d]	5	6b	4	6a	3	1	2	7
	Gene							
	Ccta (*CCT1*)	*Cctb* (*CCT2*)	*Cctg* (*CCT3*)	*Cctd* (*CCT4*)	*Ccte* (*CCT5*)	*Cctz* (*CCT6*)	*Ccth* (*CCT7*)	*Cctq* (*CCT8*)
Mouse	*Tcp-1*[e]	pTβ2[b]	pTγ2[b] mTRiC-P5[f]	pTδ2[b] A45[y]	pTε5[b]	pTζ12[b]	pCBL80[b]	pTθ1[g]
Human	*TCP1*[h]	p383*[b]	hTRiC-P5*[i]	IB713*[j]	p384*[b]	TCP20-1*[k]	—	KIAA0002[l]
Axolotl	*Tcp-1**[m]	—	—	—	—	—	—	—
Fruitfly	Tcp-1[n]	—	—	—	—	—	—	—
Nematode	*pG1**[b,o]	—	—	pG3*[b,o]	pG4*[b,o]	pG1*[b,o]	—	—
Budding Yeast	*TCP1*[p]	*TCP1β*[q] *BIN3*[r]	*BIN2*[r]	*ANC2*[s]	p701[t]	*TCP20*[k]	—	—
Plant	*Tcp-1*[u]	—	—	—	ASTCP-K19[v]	—	—	—
Protozoa	Ma68[w]	—	TpCCTγ[x]	—	—	—	—	—

[a] The gene and protein nomenclatures for CCT subunits have been described in Kubota *et al.* (1994) (*Ccta-Ccth* genes for CCTα–CCTη proteins, respectively) and Kubota *et al.* (1995a) (*Cctq* for CCTθ). Numbers are shown for each mammalian CCT subunit protein prepared from each source and clone names are shown for *Cct* genes of each species except for mouse, human, axolotl, fruitfly, yeast, and plant *Tcp-1* genes, yeast genes *TCP1β* (*BIN3*), *BIN2, ANC2,* and *TCP20,* and *Tetrahymena* TpCCTγ. The gene nomenclatures shown in parentheses (*CCT1–CCT8*) are recommended for budding yeast (*S. cerevisiae*) genes (F. Sherman, personal communication). Asterisks indicate partial clones or partially sequenced clones. Modified from Kubota *et al.* (1995b).

[b] Kubota *et al.* (1994).

[c] Frydman *et al.* (1992).

[d] Rommelaere *et al.* (1993).

[e] Willison *et al.* (1986) and Kubota *et al.* (1992).

[f] Joly *et al.* (1994a).

[g] Kubota *et al.* (1995a).

[h] Willison *et al.* (1987) and Kirchhoff and Willison (1990).

[i] Sevigny *et al.* (1994).

[j] Khan *et al.* (1992).

[k] Segel *et al.* (1992) and Li *et al.* (1994).

[l] Nomura *et al.* (1994).

[m] Sun *et al.* (1995).

[n] Ursic and Ganetzky (1988).

[o] Waterson *et al.* (1992) (*Caenorhabditis elegans*).

[p] Ursic and Culbertson (1991).

[q] Miklos *et al.* (1994).

[r] Chen, X., *et al.* (1994).

[s] Vinh and Drubin (1994).

[t] Kim *et al.* (1994).

[u] Mori *et al.* (1992) (*Arabidopsis thaliana*).

[v] Ehmann *et al.* (1993) (*Avena sativa*).

[w] Maerker and Lipps (1994) (*Stylonychia lemnae,* this sequence may include several sequencing mistakes).

[x] Soares *et al.* (1994) (*Tetrahymena pyriformis*).

[y] Xie and Palacios (1994).

TABLE III

Proteins Known to Be Folded or Assembled by CCT and Cofactors[a]

Protein	Experiment	Refs.[b]
Actin	*In vitro, in vivo*	*1,2*
Actin-RVP	*In vitro*	*3*
Tubulin	*In vitro, in vivo*	*2–8*
Neurofilament (fragment)	*In vitro*	*9*
Luciferase	*In vitro*	*5,10*
Hepatitis B virus capsid	*In vitro*[c]	*11*

[a] Kubota *et al.* (1995b). Reproduced with kind permission from *European Journal of Biochemistry.*

[b] Key to references: (*1*) Gao *et al.* (1992); (*2*) Sternlicht *et al.* (1993); (*3*) Melki *et al.* (1993); (*4*) Yaffe *et al.* (1992); (*5*) Frydman *et al.* (1992); (*6*) Gao *et al.* (1993); (*7*) Rommelaere *et al.* (1993); (*8*) Gao *et al.* (1994); (*9*) Roobol and Carden (1993); (*10*) Frydman *et al.* (1992); (*11*) Lingappa *et al.* (1994).

[c] TCP-1 or TCP-1-like chaperonin recognized by anti-TCP-1 antibody.

by increased copy number of a gene for translation elongation factor 1α. The SSB proteins were found to be associated wtih translating ribosomes, an association disrupted by puromycin. This association suggested that SSB-class proteins interact with the nascent chain, aiding passage out of the ribosome. Using a rabbit reticulocyte lysate system, Frydman *et al.* (1994) found that immunodepletion of hsp70 blocked the production of the native active form of newly synthesized luciferase (without affecting translation efficiency). Hsp70 could be added back cotranslationally, but not posttranslationally, to rescue the native state. Interaction of hsp70 proteins with the nascent chains emerging from the ribosome may prevent premature protein folding or misfolding, allowing translation to be completed prior to global arrangement of tertiary structure.

It has also been suggested that cotranslational interaction can occur between nascent luciferase chains and CCT (Frydman *et al.*, 1994). By a similar strategy of immunodepletion, in this case of CCT, a reduced production of native luciferase was observed. However, readdition of CCT posttranslationally restored a substantial amount of luciferase activity. Thus the issue of whether luciferase folding *in vivo* occurs in a

cotranslational or posttranslational manner, or in both ways, could answer these questions. There is a lack of precedent for cotranslational interaction of nascent chains with chaperonin ring complexes. In the best-studied system, mitochondria, proteins appear to be fully imported and released from mitochondrial hsp70 before they interact with hsp60.

So far, a number of proteins have been found to associate with CCT in reticulocyte lysate (see Table III); these include α- and β-tubulin (Yaffe *et al.,* 1992; Gao *et al.,* 1993; Frydman *et al.,* 1992) and their centrosomally localized relative γ-tubulin (Melki *et al.,* 1993); actin (Gao *et al.,* 1992), and its centrosomal relative centractin (Melki *et al.,* 1993); firefly luciferase (Frydman *et al.,* 1992); a neurofilament fragment (Roobol and Carden, 1993); and, more recently, a heterotrimeric G protein family member, Gα transducin (G. Farr, unpublished). Creutz *et al.* (1994) have shown an *in vitro* stable association of CCT with bovine chromaffin granule membranes. As with the other substrates, the association was stabilized in the absence of ATP.

Some of the same collective of substrates can be bound by purified CCT upon dilution from denaturant, including actin, tubulin, and luciferase. Binding in this context prevented these proteins from forming insoluble aggregates. The general location of binding on CCT appears to be shared, as actin and tubulin could compete with each other for binding to CCT (Gao *et al.,* 1993). The kinetics of binding by CCT appears to differ from binding by GroEL (Tian *et al.,* 1995). Whereas actin diluted from denaturant was bound efficiently by GroEL within seconds, a period of minutes was required before binding by CCT was observed. The same time course of binding by CCT was observed when CCT was added at various times after actin had been diluted from denaturant, further supporting the idea that actin must proceed through conformational changes beyond the initial putative early collapse that is recognized by GroEL, before being recognizable by CCT. This distinction implies that CCT and GroEL recognize different conformations of actin and tubulin, correlating with the ability of CCT, but not GroEL, to mediate productive release/folding (Tian *et al.,* 1995).

Additional tests for broader recognition by CCT have been carried out both in reticulocyte lysate and on dilution from denaturant. In the former case, when total poly(A)$^+$ RNA from a number of tissues was translated for short periods of time prior to addition of chelator, only a small group of proteins in addition to tubulin and actin could be observed associated with CCT (G. Farr, unpublished). In addition, tests with synthesis of specific protein substrates including yeast cytosolic ornithine

transcarbamoylase, yeast cytosolic invertase, and enolase did not detect appreciable binding. In the case of dilution from denaturants, a variety of cytosolic proteins have been tested, including globin and p21-ras (Gao *et al.*, 1992) and these fail to be efficiently bound. Studies in intact mammalian cells looking for association of newly synthesized proteins with CCT (Sternlicht *et al.*, 1993) show principally tubulin and actin, and only a few other unidentified proteins, binding to CCT. Thus in the eukaryotic cytosol many proteins may not require interaction with CCT in order to reach native form. Such proteins may have evolved pathways of folding that are sufficiently rapid and efficient that a chaperonin interaction is obviated (Creighton, 1994; Sosnick *et al.*, 1994). Nevertheless, Hynes *et al.* (1995) have isolated a form of CCT (II) from intact cells that contains bound tubulins and actin but also many other unidentified proteins that could be substrates.

C. Sites on CCT Involved with Polypeptide Binding

Electron microscopic studies of CCT and GroEL suggest that both chaperonins have the same overall subunit architecture (see Chapter 9). Side views of both chaperonin cylinders reveal subunits that are similarly composed of two major globular domains, one at the terminal apical end of the cylinder and one at the equatorial position, contributing a pattern of four horizontal stripes (two per ring) (Gao *et al.*, 1992; Lewis *et al.*, 1992; Frydman *et al.*, 1992). Further microscopy studies have examined complexes with bound substrates (Marco *et al.*, 1994b). In top views of such binary complexes, peptide is observed to be localized within the respective central channels. Interestingly, the channel of CCT appears to be larger than that of GroEL, measuring 60 Å, compared with 45 Å for GroEL, the latter construed from the crystal structure. In the case of GroEL, side view studies have identified bound peptide at the level of the apical domains, at the ends of the GroEL cylinder (Braig *et al.*, 1994; Chen, S. *et al.*, 1994; see Chapter 9). Localization of peptide substrates in the central channel at the apical level is further supported by crystallographic and functional analyses of GroEl (Braig *et al.*, 1994). Residues critical to peptide binding were identified by mutational analysis and were found at the inside channel face of the apical domains (Fenton *et al.*, 1994). These residues were invariably hydrophobic and

appear to form a noncharged hydrophobic surface that may make hydrophobic interactions with bound nonnative peptides.

Comparison of primary structures of GroEL and CCT subunits reveals a high degree of similarity in the N-terminal and C-terminal regions, which together make up the equatorial ATPase domain of GroEL (Kim *et al.*, 1994). In particular, a set of invariant residues within these regions line the ATP-binding pocket in GroEL, and it seems likely that they also contribute to the ATPase site in CCT. This leaves the central regions of the primary structures apparently comprising the apical domains of the respective subunits. Comparison of the primary structures through this region reveals, in contrast with the equatorial regions, little or no significant homology between CCT and GroEL. Not only do CCT subunits not share a significant relationship with GroEL, but they also do not share central sequences with one other. By contrast, individual CCT subunits have a high degree of identity in the central portion when compared across species lines. This suggests that the apical domains of individual subunits have diverged substantially from each other but functional roles of individual apical domains have been highly conserved from yeast to humans. Such a conserved and critical function of individual subunits is supported by genetic experiments with CCT subunits carried out in the yeast *S. cerevisiae* (see below), where disruption of any individual CCT subunit proves to be lethal and where such lethality cannot be rescued by overexpression of another CCT subunit.

Conserved residues in the ATP-binding site include residues 86–91 in GroEL, including the GDGTT motif, found universally in chaperonins but also present as GxGxx(G) in the phosphate-binding loop of many type I kinases (Lewis *et al.*, 1992; Kubota *et al.*, 1994) or as GxxxxGKT of Walker consensus sites (Story and Steitz, 1992). Mutation of Asp-87 in GroEL to lysine or asparagine abrogates ATPase activity and blocks release of bound polypeptide from the chaperonin (Fenton *et al.*, 1994). Residues 30–35 in GroEL are also positioned at the ATP site and form a loop segment. Mutation of the equivalent of the conserved Gly-35 of GroEL in yeast CCTα (Gly48) leads to temperature-sensitive lethality and a phenotype of tubulin deficiency (Miklos *et al.*, 1994), as does mutation in yeast CCTα (Gly-45) (Ursic *et al.*, 1994). Also conserved around the ATP site is the hexapeptide TITxDG (residues 48–53 in GroEL), and a triple glycine loop (residues 414–416 in GroEL). Mutation of the equivalent of Gly-414 in yeast CCTα produces an actin/tubulin-deficient phenotype (Ursic *et al.*, 1994, and Table IV). There are additional stretches of identity in the equatorial region of CCT, whose roles

TABLE IV

Effects of Temperature-Sensitive Mutants of CCT Subunits in Budding Yeast[a]

Subunit species	Mutant	Mutation	Tubulin effect	Actin effect	Refs.[b]
CCTα/TCP-1	*tcp1-1*	$D_{96} \rightarrow E$	+	+	*1,2*
(Cct1p)	*tcp1-2*	$G_{425} \rightarrow D$	+	+	*2*
	tcp1-3	$G_{45} \rightarrow S$	+	+	*2*
	tcp1α -245	$G_{48} \rightarrow E$	+	−	*3*
CCTβ	*tcp1β -270*	ND	+	−	*3*
(Cct2p)	*tcp1β -326*	ND	+	−	*3*
	bin3-1	ND	+	+	*4*
	bin3-2	ND	+	+	*4*
	bin3-3	ND	+	+	*4*
	bin3-4	ND	+	+	*4*
CCTγ	*bin2-1*	ND	+	+	*4*
(Cct3p)					
CCTδ	*anc2-1*	ND	−	+	*5*
(Cct4p)					

[a] ND, not determined. The symbols + and − denote presence and absence of abnormal organization of tubulin actin in the temperature sensitive mutants, respectively. Modified from Kubota *et al.* (1995b). Note the different nomenclature for the yeast genes and proteins compared to mouse and human; the eight yeast genes encoding subunits of CCT are called *CCT1–CCT8* and their protein products are called Cct1p–Cct8p. Mutations in subunit genes are listed by their mutant allele names as described in the primary papers.

[b] *Key to references:* (*1*) Ursic and Culbertson (1991); (*2*) Ursic *et al.* (1994); (*3*) Miklos *et al.* (1994); (*4*) Chen, X., *et al.* (1994); (*5*) Vinh and Drubin (1994).

are not as obvious. They may be involved in maintaining a highly-α-helical domain fold, as in GroEL, or could be involved in maintaining subunit–subunit interactions similar to those in GroEL. However, few conserved sites correspond to the equatorial interface between GroEL rings. Similarly, there is no obvious preservation of potential side-to-side contact sites formed between subunits within a ring. Loss of conservation related to intersubunit contacts would not be surprising considering that CCT assemblies have a subunit number differing from that of GroEL, presumably as a function of specific subunit–subunit contacts.

Another region of noticeable sequence homology corresponds to a position in the intermediate domain of GroEL. This domain, like the equatorial domain, comprises two parts, an amino-terminal (ascending) segment and a carboxy-terminal (descending) segment. A region of the

loop in the ascending strand, residues 172–175, lies parallel to a stretch of α helix in the descending strand, residues 396–407. Interestingly, the conserved residues in the descending helix all lie on the face adjacent to the ascending polypeptide segment. Nevertheless, the main-chain distance between these segments in GroEL is nearly 10 Å. Only two pairs of side chains approach each other (V175 and L400 are 3.9 Å apart, and E172 and R404 are 3.6 Å apart). Thus, in the unliganded form of GroEL the two structures do not make obvious contact. However, when nucleotide is bound in the equatorial pocket, these structures could conceivably make a critical functional contact.

The lack of homology in the putative apical region of CCTs invites questions about the nature and origin of the fold of this domain. Despite the lack of sequence relatedness to GroEL, does the apical region of CCT have the same or similar secondary and tertiary fold as GroEL? Do individual CCT members share this fold or are they divergent one from the other? Whatever the fold, it seems likely that the subunits maintain their binding sites on the inside faces, as supported by an electron microscopy study showing that actin bound to CCT was localized, at least in part, inside the central channel (Marco *et al.*, 1994b).

An issue related to the structure of the CCT apical domains is why binding by CCTs should involve multiple forms of the apical domain, as opposed to the single form in GroEL. We note from structural studies that the GroEL apical domains around a ring may differ in conformation from subunit to subunit, perhaps reflecting the ability of the apical binding surfaces to flexibly conform to bind polypeptide substrates with different surface structures (D. C. Boisvert, unpublished). The differences in CCT primary structure could represent an extreme of such conformational accommodation to substrates, with specificities for substrates programmed at the level of primary structure rather than directed by adjustment of tertiary structure. This possibility would amount to genetic programming of a chaperone to accommodate its substrates.

D. ATP-Dependent Release/Folding from CCT

Both in reticulocyte lysate and in defined renaturation systems, discharge from CCT and folding of actin, tubulin, and luciferase has been shown to require hydrolyzable ATP. Experiments with GroEL suggest that ATP binding/hydrolysis promotes release of peptides in nonnative

forms that partition between folding to the native state and reaching kinetically trapped or aggregation-prone conformations that are bound by other molecules of chaperonin (Weissman *et al.*, 1994; Todd *et al.*, 1994). Multiple cycles of release/rebinding are associated with folding of input molecules to native form. Evidence for such ATP-dependent cycling from CCT has recently been presented: when GroEL or mitochondrial hsp60 was mixed with a tubulin–CCT binary complex and ATP added, a portion of the tubulin molecules could be observed to transfer to GroEL or to hsp60 over a period of minutes (Tian *et al.*, 1995).

In contrast with the reduced affinity of CCT for peptide in the presence of ATP, formation of stable binary complexes is favored in the presence of ADP, suggesting that ADP-bound CCT is likely to comprise the peptide acceptor state *in vivo* (Gao *et al.*, 1994).

E. Cofactors Required for CCT-Mediated Folding of Tubulin

When binary complexes formed in reticulocyte lysate between newly translated substrate proteins and TCP-1 complex are incubated with MgATP, the bound polypeptide is discharged and can be detected in a native conformation (Gao *et al.*, 1992; Yaffe *et al.*, 1992; Frydman *et al.*, 1992). Likewise, binary complex formed *in vitro*, between substrate diluted from denaturant and purified TCP-1 complex can be discharged on incubation with MgATP (Gao *et al.*, 1992; Frydman *et al.*, 1992). In the latter reaction both actin and luciferase can be released; by contrast, neither tubulin (Gao *et al.*, 1993) nor a recently studied G protein (G. Farr, unpublished) can be productively released by MgATP alone. Two proteins have been identified as additionally required for productive folding of β-tubulin, termed A and B, of molecular sizes 28 and 100 kDa, respectively (Gao *et al.*, 1993). These components do not appear to be seven-membered ring structures, but their role in promoting productive release may be similar to that of GroES. Factor A was purified from bovine testis and cloned cDNA derived, the sequence predicting a 108 residue hydrophilic protein with no homology to any known protein (Gao *et al.*, 1994). The factor A mRNA, like that for TCP-1, is abundant in testis, at least one order of magnitude more message being present than in other somatic tissues. Given its predicted size and migration of the purified component in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–

PAGE) as a 13-kDa species, factor A is apparently a dimer. Factor A reportedly stimulated ATPase activity of CCT by severalfold when supplied in amounts up to equimolar. By implication, it must interact with CCT. In the presence of ATP, addition of factor A leads to release (without apparent rebinding) of tubulin monomers that are nonnative (Gao *et al.,* 1993). These species migrate into native gels, to a position ahead of CCT itself, and are unable to assemble with added tubulin α/β heterodimers. The functional nature of factors A and B remains unknown. Perhaps they act as specific modifiers of a specific tubulin intermediate structure, e.g., by isomerizing specific proline residues. Alternatively, they could interact with tubulin at the point of release or beyond, influencing the pathway of oligomeric assembly (Archer *et al.,* 1995). The structure of factor B remains to be determined, but it is critical to production of the native state. Campo *et al.* (1994) independently discovered factor A acting as a release factor of β-tubulin monomers from a 300 kDa complex (C-300). In the tubulin folding pathway, C-300 is downstream of CCT, and therefore it may be the case that Factor A/p14 is acting as a chaperone for β-tubulin when it is bound to CCT or to C-300. Further work is required to define more clearly the role of protein co-factors in tubulin assembly pathways.

IV. CCT ANALYSIS IN YEAST

Analysis of CCT in *S. cerevisiae* has shown that yeast contains a heteromeric CCT complex similar to that found in mammalian cells. The yeast CCT complex purifies as a multisubunit assembly (Miklos *et al.,* 1994) and two subunits, TCP-1α and TCP-1β (Cct1p and Cct2p), have been shown to occupy the same complex by coimmunoprecipitation of heterologously tagged (*myc*$^-$ and HA) species (Miklos *et al.,* 1994). It is likely that the yeast CCT complex comprises eight subunits encoded by separate genes; six *CCT* genes have been discovered and fully sequenced so far (Table IV). All six *CCT* genes are essential (Ursic and Culbertson, 1991; Miklos *et al.,* 1994; Chen, X., *et al.,* 1994; Vinh and Drubin, 1994; Li *et al.,* 1994; Kim *et al.,* 1994). Conditional mutations have been isolated in four Cctp subunits and they produce phenotypes of deficiency of either actin or tubulin or both (Table IV). Notably Cctδ (*CCT4*) was identified by a synthetic

lethality screen with a genetic mutant affecting actin (Vinh *et al.*, 1994). Ursic *et al.* (1994) have recently shown gene- and allele-specific interactions between Tcp-1 (*CCT1*) and actin and tubulins; *tub2-402*, a mild allele of β-tubulin (Huffaker *et al.*, 1988), and *tcp1-1* produce synthetic lethality in the double mutant. Genetic interaction between Tcp-1 and actin was found in *tcp1-1 act1-1* double mutants, which grow slowly and show a high frequency of cell lysis during growth. These genetic experiments are very dependent on the alleles of each gene used; *tcp1-1* has a substitution of the absolutely conserved aspartic acid (Table IV) in the GDGTT motif involved in ATP hydrolysis (Fenton *et al.*, 1994; Kim *et al.*, 1994), whereas the *tcp1α-245* (Gly-48 → Glu) mutation appeared to display no obvious actin phenotype (Miklos *et al.*, 1994). Indeed all four mutations identified in *CCT1* (Table IV) affect ATP binding and/or hydrolysis based on the molecular modeling analysis of Kim *et al.* (1994). It will be important to find out whether defects in ATP activity of a single subunit affect all the functions of CCT, to produce a completely defective complex, or only some functions.

Our own work has focused on the Tcp1α (*CCT1*) and Tcp1β (*CCT2*) genes. Both TCP-1α- and TCP-1β-deficient yeast cells exhibit specific defects of the mitotic spindle at nonpermissive temperature, arresting as large-budded cells with replicated but nonsegregated chromosomes (Miklos *et al.*, 1994). In these cells there is either absence of antitubulin staining, or a diffuse staining pattern, suggestive of defective biogenesis of tubulin and microtubules. In contrast to this readily appreciable phenotype, when specific activities of several metabolic enzymes of the cytosol were examined in the temperature-sensitive mutant strains, no reduction of specific activity could be observed. In particular when two indigenous enzymes of *S. cerevisiae*, a cytosolic dimer, invertase, or a cytosolic homotrimer, ornithine transcarbamylase, were programmed for inducible expression at nonpermissive temperature in TCP-1α or TCP-1β temperature-sensitive mutants, no defect in specific activity was observed (D. Miklos, unpublished). By contrast, in compartments where mutant versions of hsp60 or GroEL were present, the mammalian version of ornithine transcarbamylase (OTC) was detected in an inactive form in insoluble aggregates (Cheng *et al.*, 1989; Horwich *et al.*, 1993). OTC in these latter compartments has an absolute requirement for GroE chaperonin function to reach its native state. Observations in the yeast TCP-1 mutants of normal biogenesis of cytosolic enzymes indigenous to yeast again leave open the possibility of a restricted substrate specificity of CCT.

V. EVOLUTION

A. CCT

As first pointed out by Gupta (1990), GroEL and TCP-1 (CCTα) are weakly, although significantly, related throughout their length, and others support this view (Ellis, 1990; Hemmingsen, 1992; Lewis *et al.*, 1992). The sequencing of the gene encoding TF55 from the archaebacterium *S. shibatae* showed that the predicted amino acid sequence of TF55 is 40% identical to mammalian TCP-1 (Trent *et al.*, 1991), suggesting that all eukaryotic *Tcp-1* and *CCT* genes are derived from a lineage giving rise to the Archaea (archaebacterial) and Eukarya (eukaryotes). Thus, there are two major subfamilies of chaperonins: the GroE and TCP-1 subfamilies (Ellis, 1992; Lewis *et al.*, 1992). All chaperonins are thought to have evolved from a common ancestral chaperonin in the progenote (Gupta, 1990, and Chapter 2) and to have diverged independently in the eubacterial lineage and archaebacterial–eukaryotic lineage (Fig. 1). After the first divergence, the latter lineage diverged to give rise to the archaebacterial and eukaryotic lineages. We have speculated that the sequence conservation between the two subfamilies is due to the Mg-ATP binding and ATPase activity of the chaperonins, rather than the ability to form stacked toruses or bind substrates (Kubota *et al.*, 1994, 1995b), and this idea has been confirmed by the X-ray structure of GroEL (Kim *et al.*, 1994) as discussed in Section III,C of this chapter.

Complete cDNA sequences have been determined for eight members of the mouse CCT family (Kubota *et al.*, 1994, 1995a,b). All members of the family are approximately 30% identical in pair-wise comparisons of their amino acid sequences. We think it likely that these eight *CCT* genes encode the eight subunits of the "core" CCT complex, each position in the ring being occupied by a different CCT protein. Six *CCT* genes have been found so far in yeast (Table II) but it is likely that all eukaryotes contain orthologs of the eight *CCT* genes (*Ccta–Cctq*, Table II) found in mouse. However, the situation is likely to be more complicated than the simple eight-fold symmetry/8 gene hypothesis since we have discovered tissue-specific *CCT* genes in mouse (Kubota *et al.*, unpublished results) and the subunit stoichiometry as determined by biochemical and metabolic analysis is not straightforward, that is, the presence of molar equivalents of each of the eight subunits, as discussed by Kubota *et al.* (1995b). Detailed phylogenetic analysis suggests that the

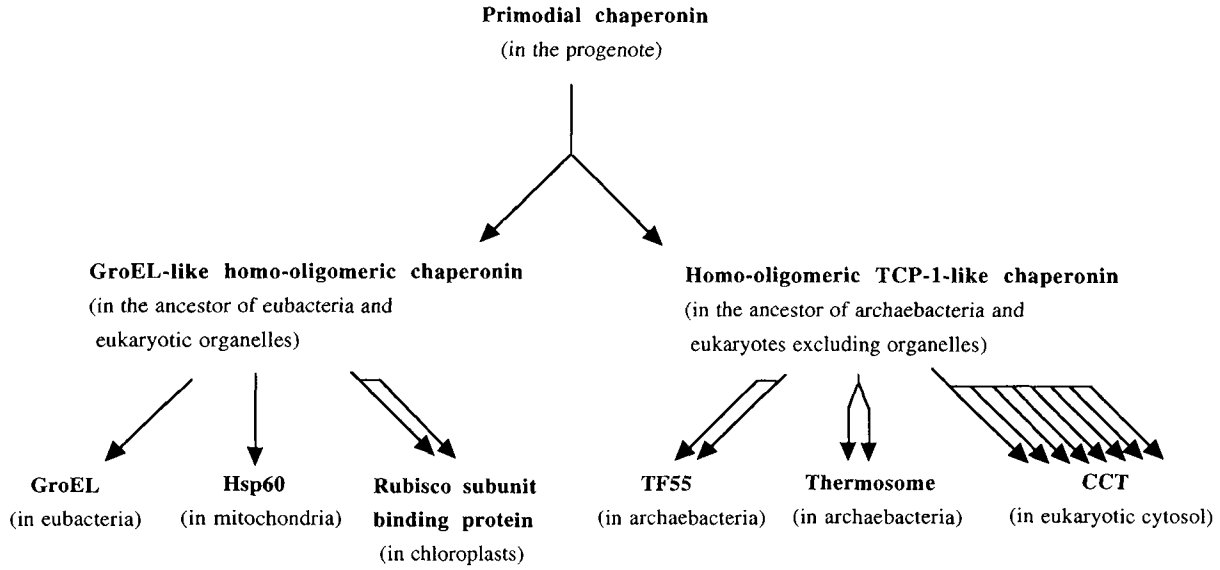


Fig. 1. Evolution of chaperonins. This evolutionary diagram was constructed by consideration of the amino acid sequence homology between chaperonin subunits (see Table I) and the endosymbiont hypothesis. An arrow indicates a subunit species. Note that it is likely that CCT evolved from an eightfold symmetrical thermosome-type chaperonin rather than from a ninefold symmetrical TF55-type chaperonin (Section II,B).

CCT family may have originated by duplications and divergence at the time of the origin of eukaryotic cells (Kubota, 1994). If so, the establishment of the cytoskeleton might have been dependent on CCT function and the rapid expansion of the *CCT* gene family to eight members from two members in the immediate archaebacterial precursor of the eukaryotic cell.

B. Actins and Tubulins

There are suggestions of a deep relationship between the heat shock proteins and the two cytoskeletal protein families, the actins and the tubulins. The ATP binding domains of hsp70, actin, and hexokinase have remarkably similar tertiary structures (Flaherty *et al.,* 1991) and it seems likely that they evolved from a common precursor that possessed an ATP binding and/or ATP hydrolysis activity. At present, it is not possible to determine whether hsp70 or hexokinase evolved first, that is, ATP hydrolysis and energy production followed by a protective hsp70 chaperone activity, or chaperone activity first in response to heat shock insults. It will be important to discover whether archaebacteria contain hsp70/dnaK-like proteins, hexokinases, and actins in order to determine the possible order of evolution of this family. The results of the *Sulfolobus* genome sequencing project are eagerly awaited. Both actins and tubulins use nucleotide binding to regulate polymerization activity, although tubulins use a GDP/GTP hydrolysis cycle. Because the crystal structure of tubulin has not yet been determined, it is not possible to speculate on the origin and nature of tubulin and its nucleotide binding domain.

Doolittle (1995) has compared the rates of evolution of the eukaryotic actins and tubulins and the related proteins of *E. coli,* ftz A (20% identical to actins) and ftz Z (20% identical to tubulins). He concludes that in eukaryotes the actin and tubulin families have both shown remarkably little sequence divergence; i.e., they are both very highly conserved proteins. Thus there is a conundrum: how did the prokaryotic precursors of actins and tubulins, presumably archaebacterial ftz A- and ftz Z-like proteins, respectively, evolve rapidly into the eukaryotic type proteins? Put another way: how did a 50% change in sequence occur in, say 1 billion years or, indeed, much less, and then changes only accumulated at the rate of 10% per billion years over the next 2 billion years? We suggest that the expansion and divergence of the CCT gene family at

2000 Myr ago (Kubota *et al.,* 1994) could have influenced the rate of change of sequence divergence. If actins and tubulins each interact specifically with multiple CCT subunits, then their structures and folding pathways could be constrained by their geometrical interaction with CCT. This model partially resolves Doolittle's conundrum because it suggests that the evolution of eight CCT genes could both have driven the increase in the rate of evolution of actin and tubulin, because they would be responding to changes occurring in up to eight genes, and then caused fixation once the folding pathway was set in place.

VI. CONCLUSIONS

If CCT functions like GroEL, why has it evolved so many subunits, each type of which has been very conserved during eukaryotic evolution? We have speculated that subunits of CCT evolved to cope with the folding of newly evolving proteins (Kubota *et al.,* 1994), and if this is the case, it seems inescapable that CCT subunits have specific binding sites for particular polypeptide structures or motifs found in proteins. Biochemical and genetic analyses are beginning to support this original model based on phylogenetic analysis (Kim *et al.,* 1994; Kubota *et al.,* 1995b). It is possible that CCT evolved solely to cope with the folding of a few cytoskeletal proteins such as actin and tubulin, and that it represents a special, unusual type of chaperonin able to do useful tasks for the folding problems encountered by this particular class of proteins. It remains to be shown that CCT is a broad-spectrum folding machine, but it is hard to envisage how other proteins (other heat shock families?) in evolving eukaryotic cells took up the general functions performed by chaperonins, whereas the CCT evolved from having a general function to having a restricted one. It may be that CCT works through a combination of general affinity for unfolded proteins and specific affinities for particular domains of certain partially folded proteins. Functional analysis of each subunit species through manipulation of the genes that encode them (Kubota *et al.,* 1994) is now under way to solve the mechanisms by which CCT facilitates protein folding in eukaryotic cytosol.

ACKNOWLEDGMENTS

K. W. thanks Dr. Hiroshi Kubota for comments on the manuscript and help with the tables. K. W. is supported by the Cancer Research Campaign.

REFERENCES

Archer, J. E., Vega, L. R., and Solomon, F. (1995). RbL2p, a yeast protein that binds to β-tubulin and participates in microtubule function *in vivo*. *Cell* **82,** 425–434.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* (*London*) **371,** 578–586.

Campo, R., Fontalba, A., Sanchez, L. M., and Zabala, J. C. (1994). A 14-kDa release factor is involved in GTP-dependent β-tubulin folding. *FEBS. Lett.* **353,** 162–166.

Chen, S., Roseman, A. M., Hunter, A. A., Wood, S. P., Burston, S. G., Ranson, N. A., Clarks, A. R., and Saibil, H. R. (1994). Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature* (*London*) **371,** 261–264.

Chen, X., Sullivan, D. S., and Huffaker, T. C. (1994). Two yeast genes wtih similarity to TCP-1 are required for microtubule and actin function *in vivo*. *Proc. Natl. Acad. Sci. USA* **91,** 9111–9115.

Cheng, M. Y., Hartl, F.-U., Martin, J., Pollock, R. A., Kalousek, F., Neupert, W., Hallberg, E. M., Hallberg, R. L., and Horwich, A. L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* (*London*) **337,** 620–625.

Craig, E. A., Baxter, B. K, Becker, J., Halladay, J., and Ziegelhoffer, T. (1994). Cytosolic hsp70s of *Saccharomyces cerevisiae:* Roles in protein synthesis, protein translocation, proteolysis, and regulation. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A., Tissieres, and C. Georgopoulos, eds.), pp. 31–52. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Creighton, T. (1994). The energetic ups and downs of protein folding. *Nature Struct. Biol.* **3,** 135–138.

Creutz, C. E., Liou, A., Snyder, S. L., Brownawell, A., and Willison, K. (1994). Identification of the major chromaffin granule-binding protein, chromobindin A, as the cytosolic chaperonin CCT (chaperonin containing TCP-1). *J. Biol. Chem.* **269,** 32035–32038.

Doolittle, R. F. (1995). The origins and evolution of eukaryotic proteins. *Phil. Trans. R. Soc. Lond. B* **349,** 235–240.

Dudley, K., Potter, J., Lyon, M. F., and Willison, K. (1984). Analysis of male sterile mutations in the mouse using haploid stage expressed cDNA probes. *Nucleic Acids Res.* **12,** 4281–4293.

Ehmann, B., Krenz, M., Mummert, E., and Schafer, E. (1993). Two *Tcp-1*-related but highly divergent gene families exist in oat encoding proteins of assumed chaperone function. *FEBS Lett.* **336,** 313–316.

Ellis, R. J. (1990). Molecular chaperones: The plant connection. *Science* **250,** 954–959.

Ellis, R. J. (1992). Cytosolic chaperonin confirmed. *Nature* (*London*) **358,** 191–192.

Fenton, W. A., Kashi, Y., Furtak, K., and Horwich, A. L. (1994). Residues in chaperonin GroEL polypeptide binding and release. *Nature* (*London*) **371,** 614–619.

Flaherty, K. M., McKay, D. B., Kabsch, W., and Holmes, K. C. (1991). Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc. Natl. Acad. Sci. USA* **88,** 5041–5045.

Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P., and Hartl, F.-U. (1992). Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11,** 4767–4778.

Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F.-U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370,** 111–117.

Gao, Y., Thomas, J. O., Chow, R. L., Lee, G.-H., and Cowan, N. J. (1992). A cytoplasmic chaperonin that catalyze β-actin folding. *Cell* **69,** 1043–1050.

Gao, Y., Vainberg, I. E., Chow, R. I., and Cowan, N. J. (1993). Two cofactors and cytoplasmic chaperonin are required for the folding of α- and β-tubulin. *Mol. Cell Biol.* **13,** 2478–2485.

Gao, Y., Melki, R., Walden, P. D., Lewis, S. A., Ampe, C., Rommelaere, H., Vandekerckhove, J., and Cowan, N. J. (1994). A novel cochaperonin that modulates the ATPase activity of cytoplasmic chaperonin. *J. Cell Biol.* **125,** 989–996.

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Guagliardi, A., Cerchia, L., Bartolucci, S., and Rossi, M. (1994). The chaperonin from the archaeon *Sulfolobus solfataricus* promotes correct refolding and prevents thermal denaturation in vitro. *Protein Sci.* **3,** 1436–1443.

Gupta, R. S. (1990). Sequence and structural homology between a mouse *t*-complex protein TCP-1 and the 'chaperonin' family of bacterial (GroEL, 60–65 kDa heat shock antigen) and eukaryotic proteins. *Biochem. Int.* **20,** 833–841.

Hemmingsen, S. M. (1992). What is a chaperonin? *Nature* (*London*) **357,** 650.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* (*London*) **333,** 330–334.

Horwich, A. L., and Willison, K. R. (1993). Protein folding in the cell: Function of two families of molecular chaperone, hsp60 and TF55-TCP1. *Phil. Trans. R. Soc. Lond. B* **339,** 313–326.

Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N., and Furtak, K. (1993). Folding *in vivo* of bacterial cytoplasmic proteins: Role of groEL. *Cell* **74,** 909–917.

Huffaker, T. C., Thomas, J. H., and Botstein, D. (1988). Diverse effects of β-tubulin mutations on microtubule formation and function. *J. Cell Biol.* **106,** 1997–2010.

Hynes, G., Kubota, H., and Willison, K. R. (1995). Antibody characterisation of two distinct conformations of the chaperonin containing TCP-1 from mouse testis. *FEBS Lett.* **358,** 129–132.

Joly, E. C., Sevigny, G., Todorov, I. T., and Bibor-Hardy, V. (1994a). cDNA encoding a novel TCP-1-related protein. *Biochim. Biophys. Acta* **1217,** 224–226.

Joly, E. C., Tremblay, E., Tanguay, R. M., Wu, Y., and Bibor-Hardy, V. (1994b). TRiC-P5, a novel TCP-1-related protein is located in the cytoplasm and in the nuclear matrix. *J. Cell Sci.* **107,** 2851–2859.

Kagawa, H. K., Osipiuk, J., Maltsev, N., Overbeek, R., Quaite-Randall, E., Joachimiak, A., and Trent, J. D. (1995). The heat shock proteins from a hyperthermophilic archaeon are related to the cytosolic chaperonin in eukaryotes. *J. Mol. Biol.,* in press.

Khan, A. S., Wilcox, A. S., Polymeropoulos, M. H., Hopkins, J. A., Stevens, T. J., Robinson, M., Orpana, A. K., and Sikela, J. M. (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* **2,** 180–185.

Kim, S., Willison, K. R., and Horwich, A. L. (1994). Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem. Sci.* **19,** 543–548.

Kirchhoff, C., and Willison, K. R. (1990). Nucleotide and amino-acid sequence of human testis-derived TCP1. *Nucleic Acids Res.* **18,** 4247.

Knapp, S., Schmidt-Krey, I., Hebert, H., Bergman, T., Jornvall, H., and Ladenstein, R. (1994). The molecular chaperonin TF55 from Archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* **242,** 397–407.

Koll, H., Guiard, B., Rassow, J., Osterman, J., Horwich, A. L., Neupert, W., and Hartl, F.-U. (1992). Antifolding activity of hsp60 couples protein import into the mitochondrial matrix with export to the intermembrane space. *Cell* **68,** 1163–1175.

Kubota, H., Willison, K., Ashworth, A., Nozaki, M., Miyamoto, H., Yamamoto, H., Matsushiro, A., and Morita, T. (1992). Structure and expression of the gene encoding *t*-complex polypeptide 1 (*Tcp1*). *Gene* **120,** 207–215.

Kubota, H., Hynes, G., Carne, A., Ashworth, A., and Willison K. (1994). Identification of six *Tcp-1*-related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr. Biol.* **4,** 89–99.

Kubota, H., Hynes, G., and Willison, K. (1995a). The eighth *Cct* gene, *Cctq*, encoding the theta subunit of the cytosolic chaperonin that contains TCP-1. *Gene* **154,** 231–236.

Kubota, H., Hynes, G., and Willison K. (1995b). The chaperonin containing *t*-complex polypeptide 1 (TCP-1): Multisubunit machinery assisting in protein folding and assembly in the eukaryotic cytosol. *Eur. J. Biochem.* **230,** 3–16.

Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H., and Willison, K. (1992). T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature* (*London*) **358,** 249–252.

Li, W.-Z., Lin, P., Frydman, J., Boal, T. R., Cardillo, T. S., Richard, L. M. Toth, D., Lichtman, M. A., Hartl, F.-U., Sherman, F., and Segel, G. B. (1994). Tcp20, a subunit of the eukaryotic TRiC chaperonin from humans and yeast. *J. Biol. Chem.* **269,** 18,616–18,622.

Lingappa, J. R., Martin, R. L., Wong, M. L., Gamen, D., Welch, W. J., and Lingappa, V. R. (1994). A eukaryotic cytosolic chaperonin is associated with a high molecular weight intermediate in the assembly of hepatitis B virus capsid, a multimeric particle. *J. Cell Biol.* **125,** 99–111.

Maercker, C., and Lipps, H. J. (1994). A macronuclear DNA molecule from the hypotrichous ciliate *Stylonychia lemnae* encoding *t*-complex polypeptide 1-like protein. *Gene* **141,** 147–148.

Marco, S., Carrascossa, J. L., and Valpuesta, J. M. (1994a). Reversible interaction of β-actin along the channel of the TCP-1 cytoplasmic chaperonin. *Biophys. J.* **67,** 364–368.

Marco, S., Urena, D., Carrascossa, J. L. Waldmann, T., Peters, J., Hegerl, R., Pfeifer, G., Sackkongehl, H., and Baumeister, W. (1994b). The molecular chaperone TF55: Assessment of symmetry. *FEBS Lett.* **341,** 152–155.

Martel, R., Cloney, L. P., Pelcher, L. E., and Hemmingsen, S. M. (1990). Unique composition of plastid chaperonin-60: α and β polypeptide-encoding genes are highly divergent. *Gene* **94,** 181–187.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. *Nature* (*London*) **352,** 36–42.

Martin, J., Horwich, A. L., and Hartl, F.-U. (1992). Role of chaperonin hsp60 in preventing protein denaturation under heat-stress. *Science* **258,** 995–998.

Melki, R., Vainberg, I. E., Chow, R. L., and Cowan, N. J. (1993). Chaperonin-mediated folding of vertebrate actin-related protein and γ-tubulin. *J. Cell Biol.* **122,** 1301–1310.

Miklos, D., Caplan, S., Martens, D., Hynes, G., Pitluk, Z., Brown, C., Barrell, B., Horwich, A. L., and Willison, K. (1994). Primary structure and function of a second essential member of heterooligomeric TCP1 chaperonin complex of yeast, *TCP1β. Proc. Natl. Acad. Sci. USA* **91,** 2743–2747.

Mori, M., Murata, K., Kubota, H., Yamamoto, A., Matsushiro, A., and Morita, T. (1992). Cloning of a cDNA encoding the *Tcp-1* (*t*-complex polypeptide 1) homologue of *Arabidopsis thaliana. Gene* **122,** 381–382.

Nelson, R. J., Heschl, M., and Craig, E. A. (1992). Isolation and characterization of extragenic suppressors of mutations in the SSA hsp70 genes of *Saccharomyces cerevisiae. Genetics* **131,** 277–285.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayashi, Y., Sato, S., Nagase, T., Seiki, N., Ishikawa, K.-I., and Tabata, S. (1994). Prediction of the coding sequences of unidentified human genes: I. The coding sequences of 40 new genes (KIAA0001–KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.* **1,** 27–35.

Phipps, B. M., Hoffmann, A., Stetter, K. O., and Baumeister, W. (1991). A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archebacteria. *EMBO J.* **10,** 1711–1722.

Phipps, B. M. Typke, D., Hegerl, R., Volker, S., Hoffmann, A., Stetter, K. O., and Baumeister, W. (1993). Structure of a molecular chaperone from a thermophilic archaebacterium. *Nature (London)* **316,** 475–477.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast *HSP60* gene coding for a mitochondrial assembly factor. *Nature (London)* **337,** 655–659.

Rommelaere, H., van Troys, M., Gao, Y., Melki, R., Cowan, N. J., Vandekerckhove, J., and Ampe, C. (1993). Eukaryotic cytosolic chaperonin contains *t*-complex polypeptide 1 and seven related subunits. *Proc. Natl. Acad. Sci. USA 90,* 11,975–11,979.

Roobol, A., and Carden, M. J. (1993). Identification of chaperonin particles in mammalian brain cytosol and *t*-complex polypeptide 1 as one of their components. *J. Neurochem.* **60,** 2327–2330.

Roobol, A., Holmes, F. E., Hayes, N. V. L., Baines, A. J., and Carden, A. J. (1995). Cytoplasmic chaperonin complex enter neurites developing *in vitro* and differ in subunit composition within single cells. *J. Cell Sci.* **107,** 1477–1488.

Segel, G. B., Boal, T. R., Cardillo, T. S., Murant, F. G., Lichtman, M. A., and Sherman, F. (1992). Isolation of a gene encoding a chaperonin-like protein by complementation of yeast amino acid transport mutants with human cDNA. *Proc. Natl. Acad. Sci. USA* **89,** 6060–6064.

Sevigny, G., Joly, E. C., Bibor-Hardy, V., and Lemieux, N. (1994). Assignment of the human homologue of the mTRiC-P5 gene (TRIC5) to band 1q23 by fluorescence in situ hybridization. *Genomics* **22,** 634–636.

Silver, L. M., Artzt, K., and Bennett, D. (1979). A major testicular cell protein specified by a mouse T/t complex gene. *Cell* **17,** 275–284.

Soares, H., Penque, D., Mouta, C., and Rodrigues-Pousada, C. (1994). *Tetrahymena* orthologue of the mouse chaperonin subunit CCTγ and its coexpression with tubulin during cilia recovery. *J. Biol. Chem.* **269,** 29,299–29,307.

Sosnik, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **3,** 149–156.

Sternlicht, H., Farr, G. W., Sternlicht, M. L., Driscoll, J. K., Willison, K., and Yaffe, M. B. (1993). The *t*-complex polypeptide 1 complex is a chaperonin for tubulin and actin *in vivo*. *Proc. Natl. Acad. Sci. USA* **90,** 9422–9426.

Story, R. M., and Steitz, T. A. (1992). Structure of the recA protein-ADP complex. *Nature* (*London*) **355,** 374–376.

Sun, H. B., Neff, A. W., Mescher, A. L., and Malacinski, G. M. (1995). Expression of the axolotl homologue of mouse chaperonin *t*-complex protein 1 during early development. *Biochim. Biophys. Acta* **1260,** 157–166.

Tian, G., Vainberg, I. E., Tap, W. D., Lewis, S. A., and Cowan, N. J. (1995). Specificity in chaperonin-mediated protein folding. *Nature* (*London*) **375,** 250–253.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. *Science* **265,** 659–666.

Trent, J. D., Osipiuk, J., and Pinkau, T. (1990). Acquired thermotolerance and heat shock in the extremely thermophilic archaebacterium *Sulfolobus* spl. strain B12. *J. Bacteriol.* **172,** 1478–1484.

Trent, J. D., Nimmesgern, E., Wall, J. S., Hartl, F.-U., and Horwich, A. L. (1991). A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein *t*-complex polypeptide-1. *Nature* (*London*) **354,** 490–493.

Ursic, D., and Culbertson, M. R. (1991). The yeast homolog to mouse *Tcp-1* affects microtubule-mediated processes. *Mol. Cell. Biol.* **11,** 2629–2640.

Ursic, D., and Ganetzky, B. (1988). A *Drosophila melanogaster* gene encodes a protein homologous to the mouse *t*-complex polypeptide 1. *Gene* **68,** 267–274.

Ursic, D., Sedbrook, J. C., Himmel, K. L., and Culbertson, M. R. (1994). The essential yeast Tcp1 protein affects actin and microtubules. *Mol. Biol. Cell* **5,** 1065–1080.

Viitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992). Purified chaperonin (GroEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1,** 363–369.

Vinh, D. B.-N., and Drubin, D. G. (1994). A yeast TCP-1 like protein is required for actin function *in vivo*. *Proc. Natl. Acad. Sci. USA* **91,** 9116–9120.

Waldmann, T., Lupas, A., Kellermann, J., Peters, J., and Baumeister, W. (1995). Primary structure of the thermosome from *Thermoplasma acidophilum*. *Biol. Chem. Hoppe-Seyler* **376,** 119–126.

Warterston, R., Martin, C., Craxton, M., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Tomas, K., Thierry-Meig, J., and Sulston, J. (1992). A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet.* **1,** 114–123.

Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78,** 693–702.

Willison, K. R., and Kubota, H. (1994). The structure, function, and genetics of the chaperonin containing TCP-1 (CCT) in eukaryotic cytosol. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A. Tissieres, and C. Georgopoulos, eds.), pp. 299–312. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Willison, K. R., Dudley, K., and Potter, J. (1986). Molecular cloning and sequence analysis of a haploid expressed gene encoding *t*-complex polypeptide 1. *Cell* **44,** 727–738.

Willison, K., Kelly, A., Dudley, K., Goodfellow, P., Spurr, N., Groves, V., Gorman, P., Sheer, D., and Trowsdale, J. (1987). The human homologue of the mouse *t*-complex

gene *TCP1* is located on chromosome 6 but not near the HLA region. *EMBO J.* **6,** 1867–1974.

Wyman, J., and Gill, S. J. (1990). "Binding and Linkage: Functional Chemistry of Biological Macromolecules." University Science Books, Mill Valley, CA.

Xie, X., and Palacios, R. (1994). Cloning and expression of a new mammalian chaperonin gene from a multipotent hematopoietic progenitor clone. *Blood* **84,** 2171–2174.

Yaffe, M. B., Farr, G. W., Miklos, D., Horwich, A. L., Sternlicht, M. L., and Sternlicht, H. (1992). TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature* (*London*) **358,** 245–248.

6

Regulation of Chaperonin Gene Expression

SASKIA M. VAN DER VIES AND COSTA GEORGOPOULOS

Département de Biochimie Médicale
Centre Médical Universitaire
Université de Genève
1211 Genève, Switzerland

I. INTRODUCTION

As is often the case in science, most progress in biology has come from the serendipity of seemingly unrelated phenomena. The chaperonin

proteins are a perfect example of this. The existence of the chaperonin functions was originally inferred from genetic studies with the *Escherichia coli*/bacteriophage λ system (reviewed in Friedman *et al.,* 1984, and Georgopoulos *et al.,* 1990). The original genetic selections were designed to detect genetic interactions between bacteria and bacteriophage-encoded gene products. These studies resulted in the identification of many bacterial genes essential for bacteriophage λ growth at all temperatures and bacterial growth at high temperatures. Two of the bacterial genes thus identified turned out to encode the universally conserved chaperonin proteins GroES (or chaperonin 10; cpn10) and GroEL (or chaperonin 60; cpn60).

The reason the selection procedure worked in the first place is that bacteriophage growth exhibits a keen requirement for "chaperonin power," especially during its morphogenetic cycle, late in the infectious process. If sufficient chaperonin power does not exist at that particular stage of bacteriophage development, sufficient morphogenesis will not take place, thus blocking DNA encapsidation and limiting the formation of viable bacteriophage progeny. The bacterial host, however, can "limp" along with reduced chaperonin power and propagate under these laboratory conditions, thus enabling the isolation of mutations in the corresponding chaperonin genes. If it were not for the use of bacteriophage as a selection agent, the discovery of the chaperonin genes would have been considerably delayed.

II. CHAPERONIN GENE ORGANIZATION

As stated above, the chaperonin-encoding genes of *E. coli* were identified more than 20 years ago, due to the fact that chaperonin gene mutations interfered with the morphogenesis of several bacteriophages (reviewed by Georgopoulos *et al.,* 1990; Zeilstra-Ryalls *et al.,* 1991). The *E. coli* chaperonin operon is composed of two genes, *groES* and *groEL* (Tilly *et al.,* 1981) and maps at 94 min on the *E. coli* chromosome (Georgopoulos and Eisen, 1974). For bacteriophage λ, *groES* or *groEL* mutations inhibit the assembly of the "preconnector," the basic oligomeric structure of 12 identical bacteriophage-encoded λB polypeptides, onto which the major λE capsid protein assembles (Georgopoulos *et al.,* 1973). In the case of bacteriophage T4, capsid protein Gp23 forms large insoluble aggregates that associate with the cell inner membrane either

when its own gene *31* is mutated or in certain *groEL* mutant hosts (Laemmli *et al.,* 1970; Georgopoulos *et al.,* 1972; Takano and Kakefuda, 1972). We have shown that the product of gene *31,* Gp31, is a functional analog of GroES (van der Vies *et al.,* 1994, see Section VI below).

Like *E. coli,* most bacteria contain one *groESL* operon (see Table I). However, multigene families of closely related *groESL*-like operons have been found in members of the *Rhizobium* genus, while certain gram-positive bacteria as well as the gram-negative cyanobacteria contain an additional monocistronic *groEL* gene without an upstream *groES* sequence (see Table II). The *Rh. meliloti* bacterium contains an unusual selection of chaperonin genes with two *groESL*-like operons and one monocistronic *groEL*-like gene. In addition, one of the *groEL* sequences is present on one of the large plasmids present in this bacterium. The purpose of these different and multicopy *groE(S)L* gene arrangements is currently not understood, but may reflect the necessity to modulate the cellular quantity and quality of specific chaperonins to meet various physiological requirements, thus allowing bacteria to respond to specific environmental stimuli (see below).

In eukaryotes, the presence of chaperonins is not just restricted to mitochondria and plastids, since a distantly related chaperonin subfamily, represented by the *t*-complex polypeptide 1 (TCP-1) protein, constitutes a key component of the cytosolic protein folding machinery. The TCP-1-related proteins show low, but significant, sequence similarity to the GroEL-like family members, suggesting that they have evolved from a common archaebacterial ancestor (see Chapters 2 and 5). Little is known about the gene organization and transcriptional control of the higher eukaryotic chaperonins, since most of the sequences have been isolated from cDNA libraries. The mitochondrial chaperonin genes in yeast are nuclear-encoded and contain their own regulatory sequences (Johnson *et al.,* 1989; Reading *et al.,* 1989; Rospert *et al.,* 1993). The two *groEL*-related genes identified in the nuclear genome of *Zea mays* contain 16 introns each that divide the entire coding sequence in approximately equal size fragments of about 100 nucleotides each (Pose, 1993). The two maize *groEL*-related genes code for an identical chaperonin that contains a typical mitochondrial targeting presequence. Although no chaperonin gene sequences have so far been identified in the mitochondrial genome, the plastid genome of the thermophilic unicellular red alga *Cyanidium caldarium* contains a single *groEL*-related gene (Maid *et al.,* 1992).

The observation that either the *groES* or the *groEL* gene of *E. coli* (Fayet *et al.,* 1989; see Table I) and their corresponding prokaryotic and

TABLE I

Single or Multiple *groESL* Operons under Different Regulatory Controls in Bacteria

Bacterium	Regulation	Stress response	Additional information	Refs.
Acrythosiphon pisum	σ^{32}	GroEL: major cellular protein under symbiotic conditions	Complementation of *E. coli groESL* mutants; Rubisco assembly *in vitro;* small IR between *groES* and *groEL;* σ^{32} promoter functional in *E. coli*	Ohtaka *et al.* (1992); Kakeda and Ishikawa (1991)
Bacillus subtilis	σ^{43} [a] 9 bp IR[b]	mRNA: 5-fold increase during 10 min at 48°C	*groESL* is essential for growth at 30° and 42°C; σ^{43} is used at both 30° and 48°C	Li and Wong (1992); Schmidt *et al.* (1992)
Bacillus stearothermophilus	σ^{43} [a] 9 bp IR[b]	mRNA: 7-fold increase during 10 min at 70°C; GroEL: major protein after 30 min at 65°C	Complementation of *E. coli groESL* mutants for bacterial and λ growth	Schön and Schumann (1993)
Brucellus abortus	σ^{70} σ^{32}	GroEL: 10-fold increase at 44°C up to 4 h	Two distinct promoters: σ^{70} at 97, and σ^{32} at 217 nucleotides upstream from initiating ATG codon	Lin *et al.* (1992); Gor and Mayfield (1992)
Clostridium acetobutylicum	σ^{43} [a,c] 11 bp IR[b]	mRNA: maximal levels at 15 min at 42°C	σ^{43} is used at both 30° and 48°C	Narberhaus and Bahl (1992)
Chromatium vinosum	9 bp IR[b]	Not determined	Complementation of *E. coli groESL* mutants for bacterial and λ growth; Rubisco assembly *in vivo*	Ferreyra *et al.* (1993)

Escherichia coli	σ^{70} [b] σ^{32} [b]	GroEL: 10-fold increase at 46°C	*groESL* is essential for growth at all temperatures (see text)	Fayet *et al.* (1989); Cowing *et al.* (1985); Hemmingsen *et al.* (1988)
Haemophilus ducreyi	σ^{70} [c] σ^{32} [c]	mRNA: 5-fold increase after 45 min at 42°C	σ^{32} promoter functional in *E. coli*	Parsons *et al.* (1992)
Legionella pneumophila	σ^{32}	Not determined	When grown in HeLa cells, GroEL on bacterial cell surface or in periplasm; σ^{32} promoter functional in *E. coli*	Hoffman *et al.* (1990)
Synechococcus PCC7942	9 bp IR[b] σ^{32} [c]	mRNA: 120-fold increase after 20 min at 45°C GroEL: 10-fold increase after 20 min at 45°C	GroEL associates with thylakoid membranes under heat shock in iron-deficient cells	Webb *et al.* (1990)
PS3	σ^{32}-like	GroEL: 2-fold increase after 10 min at 70°C		Ohta *et al.* (1993) Tamada *et al.* (1991)
Bradyrhizobium japonicum	σ^{54} [c] σ^{32} [c] σ^{70} [c]	GroEL3: 120-fold increase under anaerobic and symbiotic conditions GroEL1: 40-fold increase after 200 min at 39°C	Differential regulation of the *five* related *groESL* operons (see text)	Fischer *et al.* (1993)
Rhizobium leguminosarium	not known	GroEL: 4-fold increase after 120 min at 37°C	*Three* related *groESL* operons; no complementation of *E. coli groEL* mutants	Wallington and Lund (1994)

[a] σ^{43} in *Bacillus* is the functional analog of *E. coli*'s σ^{70}.
[b] IR indicates the presence of an inverted repeat.
[c] Promoter analysis included also determination of transcriptional start site by primer extension.

TABLE II

Multiple and Different *groE(S)L* Gene Arrangements in Bacteria

Bacterium	No. of genes		Regulation	Stress response	Additional information	Refs.
	groESL	*groEL*				
Synechocystis PCC6803	1		9 bp IR[a]	mRNA: 100-fold increase after 15 min at 42°C	GroEL1 and 2 are as related to one another as they are to *E. coli* GroEL (a.a. seq = 56% identity)[b]	Lehel *et al.* (1993)
		1	9 bp IR[a] σ32	mRNA: 30-fold increase after 90 min at 42°C	σ32 only −10 consensus present, −35 consensus missing	Chitnis and Nelson (1991)
Mycobacterium tuberculosis	1		σ70 σ32	Not determined	IR[a] between *groES* and *groEL;* GroEL1 (*groESL*) contains His-rich carboxyl terminus; GroEL1 and 2: a.a. seq = 66% identity	Shinnick (1987); Shinnick *et al.* (1989); Kong *et al.* (1993)
		1	9 bp IR[a]	Not determined		
Streptomyces albus	1		two 9 bp IR[a]	GroEL: 4-fold increase during 2 h at 41°C	IR between *groES* and *groEL; groESL* is essential for growth (see text)	Guglielmi *et al.* (1991); Mazodier *et al.* (1991); Servant *et al.* (1993)
		1	two 9 bp IR[a]			
Rhizobium meliloti	2		Not known	GroEL: 60% of total protein after 15 min at 45°C	One *groEL* present on plasmid; six *GroEL* isoforms at 45°C; no σ54 consensus sequence	Rusanganwa and Gupta (1993)
		1	Not known			

[a] IR indicates the presence of an inverted repeat.

[b] a.a. seq, amino acid sequence.

eukaryotic homologs (Reading *et al.*, 1989; Rospert *et al.*, 1993) are essential for growth of all organisms thus far examined, under all conditions tested, highlights the fundamental importance of both proteins for protein folding and cellular survival.

III. INDUCTION OF CHAPERONIN SYNTHESIS IN RESPONSE TO STRESS

Many kinds of stress such as heating, chilling, wounding, and infection either elicit or enhance the expression of specific genes. All organisms examined respond to heat (heat shock response) by transiently accelerating the rate of synthesis of a specific group of proteins, called "heat shock" or "stress" proteins (reviewed by Neidhardt and VanBogelen, 1987; Lindquist, 1986; Welch, 1993). The heat shock response is one of the best studied phenomena in biology, because of its universal character, the structural and functional conservation of the heat shock proteins, and the use of its regulation as a model for gene expression. Occasionally, stress-induced genes are found together in operons, as is the case for the *groES–groEL* and *dnaK–dnaJ* genes in *E. coli.* The effect of stress on the synthesis of chaperonins varies in duration and magnitude but invariably causes enhanced accumulation (Tables I and II). For example, at high temperatures the concentration of the plastid and mitochondrial chaperonins increases two- to threefold (McMullen and Hallberg, 1987; see Chapter 3), whereas the transient rate of synthesis of the *E. coli* chaperonins increases 20-fold. Pathogenic and endosymbiotic bacteria encounter another type of stress imposed by their hostile environment, which causes the induction of specific set of proteins, including the chaperonins (see Section V). As a first approximation, both the transient rate of chaperonin synthesis and the final accumulation are proportional to the temperature of exposure. As a consequence, synthesis of the GroEL protein accounts for 1% of the total protein synthesis in the cell under steady-state growth at 37°C, but soon after the cells are shifted to 46°C the amount of GroEL protein increases to 10% of the total cellular protein (Herendeen *et al.*, 1979). The best studied example of chaperonin regulation is presented by the *E. coli groESL* operon, which is regulated as part of the heat shock response. The regulation of the *E. coli groESL* operon is mainly at the level of transcription and is discussed below.

IV. REGULATION OF *Escherichia coli* CHAPERONIN GENES

A. Identification of *Escherichia coli* Chaperonins as Stress Proteins

The chaperonins [also known as GroEL and GroES in *E. coli,* and more generally as chaperonin 60 (cpn60) and chaperonin 10 (cpn10)] were originally identified as one of the major cellular proteins whose rate of synthesis is increased by several external factors, such as bacteriophage λ infection, heat, ultraviolet light (UV) radiation, and various chemical reagents such as ethanol and heavy metals (reviewed by Neidhardt and VanBogelen, 1987). In *E. coli* cells that carry a mutation in the *rpoH* gene, expression of the chaperonin operon is constitutively low, a key observation that suggested that this gene codes for a protein that functions as a positive regulator of the stress response. It turned out that the *rpoH* (also called *htpR* or *hin*) gene codes for a polypeptide with a M_r of 32,000 (σ^{32}), which associates with the RNA polymerase core (E) to form the $E\sigma^{32}$ holoenzyme (Grossman *et al.,* 1984; Landick *et al.,* 1984). The $E\sigma^{32}$ polymerase recognizes specific promoter sequences that consist of two conserved regions, -CCCTTGAA- and -CCCATTTA, separated by 13 to 15 bp and centered around the −35 bp and −10 bp positions respectively, upstream of the transcriptional start site (Cowing *et al.,* 1985; Gross *et al.,* 1990). These conserved −35 and −10 regions, together known as the heat shock- or the σ^{32}-promoter, are also present in the *groE(S)L* operons of myco-, cyano-, and symbiotic bacteria (Table III).

In *E. coli,* transcription of the *groESL* operon at all temperatures is ensured by two promoters, one of which is recognized by the $E\sigma^{32}$ holoenzyme and the other by the $E\sigma^{70}$ holoenzyme (Grossman *et al.,* 1987; Zhou *et al.,* 1988). The presence of the σ^{32} promoter confers on the *groESL* operon the typical regulation of heat shock gene expression. Under physiological conditions, i.e., at 30°C and above, transcription from the σ^{32} promoter occludes transcription from the σ^{70} promoter. Upon lowering the temperature, transcription from the σ^{70} promoter increases and at 17°C transcription constitutes 25% of that seen at 30°C (Zhou *et al.,* 1988).

The levels of heat shock gene expression in *E. coli* are directly proportional to the amount of σ^{32} in the cell (Straus *et al.,* 1987). For example, when the growth temperature of *E. coli* cells changes from 30 to 42°C, the amount of both σ^{32} and *dnaK*-specific mRNA increases about 17-

TABLE III

Heat Shock Promoter Sequences in *groE(S)L* Operons

Species	−35 region	Spacer	−10 region	Refs.
E. coli (*groESL*)	**CCCCCTTGAA**	GGGGCGAAGCCAT	**CCCCATTTC**	Straus *et al.* (1987)
A. pisum (*groESL*)	**TACCCTTGAA**	AGTTTTAATAAATAT	**CCCTATATT**	Ohtaka *et al.* (1992)
L. pneumophila (*groESL*)	**TACCCTTGAA**	AGTTTGTTTATTGT	**CCCCATATC**	Hoffman *et al.* (1989)
H. ducreyi (*groESL*)	**TCACCTTGAA**	AGTAATAAATTGA	**TCCCAATC**	Parsons *et al.* (1992)
B. japonicum (*groESL1*)	**ACCTCTTGAA**	CGGCCGTTTTTCGGAT	**CCTAGGTCG**	M. Babst (personal communication)
Synechococcus (groESL)	**TCGCGTTGCC**	CTCCGAGAAGGCG	**GCCCGTACA**	Webb *et al.* (1990)
Synechocystis (*groEL*)			**CCCCATTTA**	Chitnis and Nelson (1991)
M. tuberculosis (*groESL*)	**CGCCCTTGAG**	GTGCTAGCAC	**TCTCATGTA**	Kong *et al.* (1993)
PS3 (*groESL*)	**CCTACTTGCA**	AAACGGAAAGAA	**GTTCATTAA**	Ohta *et al.* (1993)
B. abortus (*groESL*)	**CCTCCTTGAC**	AAAAAACATTGCGG	**CTTCTATCT**	Lin *et al.* (1992)
σ^{32} consensus	**C - CCCTTGAA**		**- CCCATTTA**	Cowing *et al.* (1985)

fold within 5 min, followed by a gradual, concomitant decline to a steady state level of three times that seen at 30°C. Thus the synthesis of heat shock proteins is controlled by those cellular factors that determine the amount and activity of σ^{32}. Extensive genetic and biochemical work on the regulation and function of the *rpoH* gene has been carried out in the laboratories of T. Yura, F. Neidhardt, and C. Gross over the last few years. Some of the history and early contributions have been reviewed elsewhere (Neidhardt and VanBogelen, 1987; Gross *et al.*, 1990), whereas more recent studies are summarized in a review by Yura *et al.* (1993). In this section we briefly summarize the regulation of the *rpoH* gene and highlight the mechanism by which the DnaK/DnaJ/GrpE chaperone machine directly or indirectly regulates both the concentration and the activity of the σ^{32} protein.

B. Regulation of σ^{32} Transcription Factor

1. Transcriptional Regulation of rpoH Gene

Transcription of the *rpoH* gene is governed by at least four promoters that are required for the transcription of the *rpoH* gene. P1, P4, and P5 are recognized by $E\sigma^{70}$ (the P2 promoter is strain-specific), whereas P3 is recognized by the $E\sigma^{E}$ (σ^{24}) holoenzyme (reviewed by Gross *et al.*, 1990; Yura *et al.*, 1993; Fig. 1). Under physiological conditions the P1 promoter is responsible for the majority of the *rpoH* transcripts. Transcription from the P3 promoter is unusual, as it gradually increases with an increase in temperature within the normal growth range of *E. coli.* However, at extremely high temperatures, such as 51°C, where $E\sigma^{70}$-directed transcription completely ceases, only $E\sigma^{E}$-directed transcription of the *rpoH* gene is observed, thus ensuring the synthesis of sufficient amounts of σ^{32} (Erickson and Gross, 1989). The continuous presence of σ^{32} is especially required at high temperatures to guarantee the continuous transcription of the heat shock genes, thereby providing the cell with sufficient amounts of heat shock proteins to combat the increased denaturation and subsequent aggregation of proteins that result from the increased temperature.

There are two known negative controls that influence *rpoH* gene transcription. One is imposed by the DnaA replication protein of *E. coli* (Wang and Kaguni, 1989). The DnaA protein inhibits P3- and P4-

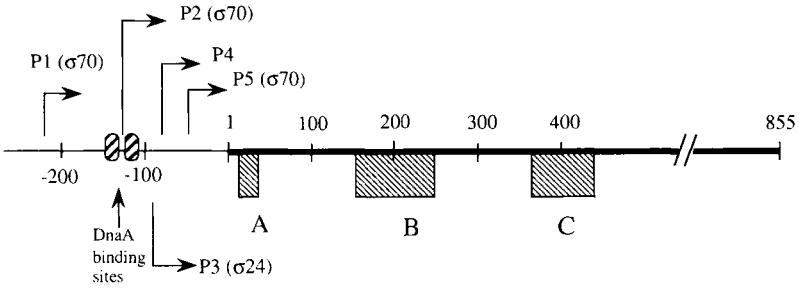


Fig. 1. Organization of the promoter and other regulatory sites of the *rpoH* gene. The five different promoters (P1, P2, P3, P4, and P5), their respective start sites, and direction of transcription are indicated by the arrows. Nucleotide numbering is with reference to the structural gene which starts at position 1 and ends at position 855. Regulatory regions A, B, and C as well as the DnaA binding sites are indicated and discussed in the text (Section IV,B).

directed transcription through binding at two specific sites in the *rpoH* promoter region (Fig. 1). This negative regulation may serve to "fine-tune" *rpoH* gene expression in the cell, since the bulk of the transcription is directed from the P1 promoter, which is not affected by the DnaA protein. The second negative control is exerted by the *htrC* gene product (Raina and Georgopoulos, 1990). The basis for this conclusion is the fact that in *htrC* null mutant bacteria, the heat shock response is turned on at all temperatures. The mechanism by which the HtrC protein exerts this negative, pleiotropic effect on heat shock gene expression is not known.

2. *Translational Regulation of rpoH Gene*

Although there is a relatively minor increase in the accumulation of *rpoH* transcripts, the amount of σ^{32} protein is dramatically increased following a temperature shift from 30 to 42°C. This increase is a consequence of both an increased rate of translation of the *rpoH* mRNA at 42°C, and a transient stabilization of the mature σ^{32} polypeptide (Straus *et al.*, 1987; reviewed in Gross *et al.*, 1990, and Yura *et al.*, 1993). An extensive deletion analysis of *rpoH–lacZ* gene fusions by the Yura laboratory led to the identification of three important regulatory regions of the *rpoH* mRNA (Fig. 1).

The first region, named A, is present immediately downstream of the *rpoH–lacZ* initiation codon (nucleotides 6–20). This region is a positive element that ensures a high rate of translation and is thought to interact with a complementary region present on the 16S rRNA, thereby stabilizing the binding of the *rpoH* mRNA to the 30S ribosomal subunit. The rate of *rpoH* translation decreases 15-fold when region A is deleted.

The second region, B, is located between nucleotides 153 and 247 and acts as a negative element, since its absence leads to increased translation of *rpoH* mRNA at all temperatures. The observation that certain point mutations in region A result in constitutive *rpoH* gene expression led Yura *et al.* (1993) to propose the existence of a secondary structure in the *rpoH* transcript, which may be responsible for the temperature-dependent translational regulation. In this model the putative regulatory secondary structure involves base-pairing of the AUG initiation codon, found in region A, with complementary sequences present in region B. In strong support of this model is the observation that point mutations designed to disrupt the predicted RNA secondary structure result in higher constitutive levels of *rpoH* gene expression. It is possible that a *trans*-acting protein is required to stabilize the *rpoH* mRNA secondary structure, thereby impeding translation. Such a putative protein factor would be expected to be temperature-labile, since the rate of *rpoH* mRNA translation increases instantaneously upon a shift to 42°C.

The third *cis*-acting region (C) is flanked by nucleotides 364 and 433 and appears to be responsible for the translational repression of the *rpoH* mRNA. A frameshift mutation changing the amino acid sequence in this region abolishes translational repression, suggesting that this polypeptide segment is directly involved in restraining translation. In addition, repression of *rpoH* mRNA translation does not occur in *dnaK* or *dnaJ* or *grpE* mutant bacteria, indicating that the DnaK/DnaJ/GrpE chaperone machine directly or indirectly exerts this translational repression effect (Yura *et al.*, 1993). It could be that members of the DnaK chaperone machine bind to the C region in the σ^{32} polypeptide, perhaps as it emerges from the ribosome, thus arresting translation directly or, alternatively, indirectly by presenting the unfinished σ^{32} polypeptide to proteases for immediate degradation.

3. Stability of Mature σ^{32} Polypeptide

The mature σ^{32} polypeptide chain is highly unstable with a half-life of 45-60 s, measured during steady-state growth of *E. coli* both at 30 and 42°C (Straus *et al.*, 1987; Tilly *et al.*, 1989). When *E. coli* cells are

shifted from 30 to 42°C, however, the σ^{32} polypeptide is transiently stabilized for a short period of time, after which its instability is restored. It had previously been observed that its half-life increases 10- to 30-fold in *dnaK* or *dnaJ* or *grpE* mutant bacteria, suggesting that the DnaK/DnaJ/GrpE chaperone machine may be directly responsible for mediating the stability of the σ^{32} polypeptide (Tilly *et al.*, 1989; Straus *et al.*, 1990). Studies with purified proteins *in vitro* have demonstrated directly the ability of σ^{32} to form stable complexes with different members of the DnaK/DnaJ/GrpE chaperone machine (reviewed by Georgopoulos and Welch, 1993; Bukau, 1993). However, the exact mechanism by which the DnaK chaperone machine determines the *in vivo* instability of σ^{32} remains to be determined (see Section IV,B,4).

4. *Autoregulation of Heat Shock Response by DnaK/DnaJ/GrpE Chaperone Machine*

There is a great deal of genetic and biochemical evidence that the DnaK/DnaJ/GrpE chaperone machine plays a key role in the regulation of synthesis and degradation of σ^{32}. It is beyond the scope of this review to discuss the supporting evidence in more detail, and the interested reader is referred to reviews by Gross *et al.* (1990), Georgopoulos *et al.* (1990), Craig and Gross (1991), Yura *et al.* (1993), Georgopoulos and Welch (1993), and Bukau (1993). Here, we highlight the various activities displayed by the DnaK chaperone machine during σ^{32} regulation.

Various members of the DnaK chaperone machine function together to regulate both the amount and the activity of σ^{32} by: (1) inhibiting the polymerase $E\sigma^{32}$ activity following its binding to a heat shock promoter; (2) sequestering the σ^{32} polypeptide, thus preventing the formation of the $E\sigma^{32}$ polymerase holoenzyme; (3) arresting *rpoH* mRNA translation; and (4) mediating the proteolysis of σ^{32}. A likely scheme of σ^{32} regulation may be envisioned as follows. Under normal conditions heat shock gene expression is inhibited because very little σ^{32} is synthesized, and in addition, whatever mature σ^{32} is present is sequestered by the DnaK chaperone machine. When the temperature is raised, various polypeptides begin to unfold, and to prevent their subsequent aggregation, the activity of the DnaK chaperone machine is required. Consequently, both σ^{32} translation increases and the "captured" σ^{32} polypeptide is liberated from the DnaK chaperone machine. The σ^{32} then associates with the RNA polymerase core to form the $E\sigma^{32}$ holoenzyme, which subsequently accelerates the rate of transcription of heat shock genes. As a result, the amount of heat shock proteins, like the DnaK chaperone machine, increases. Once

the intracellular level of the DnaK chaperone machine has increased by a sufficient amount to deal with the damaged proteins, it can again resequester σ^{32}, thus limiting its amount and activity. An indirect role in the regulation of σ^{32} may be played by the chaperonins (sometimes referred to as the GroEL chaperone machine), since they mediate the correct folding of many proteins (and hence perhaps σ^{32}), including themselves (Van Dyk *et al.,* 1989; Lissin *et al.,* 1990; Viitanen *et al.,* 1992).

V. REGULATION OF CHAPERONIN GENES IN OTHER BACTERIA, INCLUDING GRAM-POSITIVE BACTERIA

The discovery of multiple *groESL* operons and additional monocistronic *groEL* genes in various bacteria other than *E. coli* (Table II) raises two, not mutually exclusive, questions: (1) Do the different chaperonins fulfill physiologically distinct functions? (2) Do certain bacteria need differential regulation of chaperonin gene expression either at different times during their life cycle or to survive externally imposed stress conditions? For example, in the cyanobacterium *Synechocystis,* the amount of *groESL* mRNA is about 10-fold higher than the *groEL* monocistronic mRNA at 30°C, whereas both mRNA levels increase 50-fold after 30 min at 42°C (A. Glatz, personal communication). At the same sublethal temperature of 42°C, cpn60 and cpn10 are found associated with the thylakoid membrane, possibly to prevent destabilization by directly binding to the membrane, or indirectly through binding to membrane-associated proteins, such as the photosynthetic complex (Kovács *et al.,* 1994a,b). Unlike *E. coli,* most endosymbiotic bacteria examined synthesize larger quantities of cpn60 relative to cpn10, which most likely reflects their specific physiological needs. The calculated molar ratio of GroEL/GroES in *E. coli,* under steady-state cell growth, is approximately 2, based on the figures given by Neidhardt and VanBogelen (1987). Here we review the regulation of chaperonin gene expression in organisms other than *E. coli.*

A. Regulation of *groESL* Multigene Family in *Bradyrhizobium japonicum*

Under natural conditions, bacteria of the genera *Rhizobium* and *Bradyrhizobium* infect plant root cells. This infection is followed by

the differentiation of the bacterium into an intracellular symbiont (also known as a bacteroid) and involves a series of complex developmental processes, including the assembly of the nitrogenase complex (reviewed by Gussin *et al.*, 1986; Fisher and Long, 1992). This enzyme complex is essential for symbiosis, as it catalyzes the reduction of atmospheric nitrogen into ammonia, a growth requirement for both the symbiont and the host. The bacterial infection of the root cells is expected to cause considerable stress to both plant and bacterium since the initial host-defense response is aimed at elimination of the infective organism. Additional stresses, such as low oxygen concentration and nutrient limitation, may also be encountered by bacteroids. These stresses elicit or enhance the expression of specific bacterial- and host-encoded genes, including those encoding the chaperonins. The chaperonins of the soybean symbiont *B. japonicum* are encoded by a multigene family of five very similar, but not identical, *groESL*-like operons, the members of which have been shown to be under different regulatory controls.

Fischer *et al.* (1993) used a series of translational *lacZ* fusions, integrated at their homologous sites in the *B. japonicum* chromosome, to show that the *groESL1* and *groESL5* operons are transcribed very weakly, whereas *groESL2* and *groESL4* are expressed constitutively at high levels, under aerobic, anaerobic, or symbiotic conditions. The *groESL1* operon contains a typical *E. coli* $E\sigma^{32}$-dependent heat shock promoter sequence (Table III), and its expression is consequently increased about 25-fold when the temperature is shifted from 28 to 39°C. Three operons, *groESL2, groESL4,* and *groESL5* contain an inverted repeat (IR) element, previously shown to be characteristic of the regulation of heat shock gene expression in *Bacillus subtilis* and other bacteria (see Section V,B and Table IV). In all three operons, the IR element is present downstream of the transcriptional start site, is preceded by canonical σ^{70} promoter sequence, and is thought to be involved in the induction of *groESL4* and *groESL5* gene expression during heat shock. In contrast, *groESL2* is not induced during heat stress, suggesting a distinct, and yet unknown, mechanism of gene regulation.

Most interesting is the regulation of the *groESL3* operon. Under anaerobic growth, or in bacteroids, *groESL3* gene expression increases 450- and 200-fold, respectively. The induction of *groESL3* gene expression appears to be absolutely dependent on both the oxygen-responsive transcriptional activator NifA and the RpoN (σ^{54}) proteins. The NifA protein plays a key role in the transcription of different genes involved in nitrogen fixation (Fisher and Long, 1992). Under low oxygen tension, as present

TABLE IV

Inverted Repeat Sequences Located in *groE(S)L* Operons

Bacterium	Inverted repeat sequence	Refs.
B. subtilis (*groESL*)	TTAGCACTCTTTAGTGCTGAGTG CTAA	Li and Wong (1992); Schmidt *et al.* (1992)
C. acetobutylicum (*groESL*)	TTAGCACTCAAGATTAACGAGTG CTAA	Narberhaus and Bahl (1992)
Ch. vinosum (*groESL*)	TTAGCACTCGTTACAAGTGAGTG CTAA	Ferreyra *et al.* (1993)
L. lactis (*groESL*)	TTAGCACTCGTTTAATAAGAGTG CTAA	van Asseldonk *et al.* (1993)
Synechococcus (groESL)	TTAGCACTCAGGTACTGGGAGTG CTAA	Webb *et al.* (1990)
Synechocystis (*groESL*)	TTAGCACTCGTGAGGTGGGAGTG CTAA	Lehel *et al.* (1993)
L. interrogans (groESL)	TAAGCACTCTCACGAATATAGTG CTAA	Ballard *et al.* (1993)
B. japonicum (*groESL2*)	CTAGCACTCGCTAGCTTCGACTG CTAT	M. Babst (personal communication)
(*groESL4*)	CTAGCACTCGCGGGCACAGACTG CTAA	M. Babst (personal communication)
(*groESL5*)	TTAGCACTCGGAAAGCCCGATTG CTAA	M. Babst (personal communication)

M. leprae (*groEL*)	**TTAGCACTC**GCCTTAGGG**GAGTG CTAA**	Mehra *et al.* (1986)
M. tuberculosis (*groEL*)	**TT GCACTC**GGCATAGGC**GAGTG CTAA**	Shinnick (1987); Shinnick *et al.* (1989)
Synechocystis (*groEL*)	**TTAGCACTC**CACTGCCAA**GAGTGTCTAA**	Chitnis and Nelson (1991)
A. tumefaciens (*groEL*)	**CTGGCACTC**CAA CAAGG**GAGTGTCTAA**	Segal and Ron (1993)
S. albus (*groESL*)		
Untranscribed	**TTGGCACTC**CGCTTGACC**GAGTG CTAA**	Guglielmi *et al.* (1993)
Transcribed	**CTGGCACTC**CCCGCAGGA**GAGTG CCAA**	Guglielmi *et al.* (1993)
(*groEL*)		
Untranscribed	**CTTGCACTC**TCCTACCCC**GAGTG CTAA**	Guglielmi *et al.* (1993)
Transcribed	**TTAGCACTC**TCCGAGTGA**GAGTG ACAG**	Guglielmi *et al.* (1993)

in the root nodule, NifA binds to an upstream activator sequence where it interacts with the RNA polymerase $E\sigma^{54}$ thereby allowing transcription to begin (Morett and Buck, 1989; Morett *et al.*, 1991). The *groESL3* operon maps within a cluster of genes required for nitrogen fixation, and promoter elements characteristic of NifA-regulated genes are found within the *groESL3* operon, suggesting that these chaperonins may be required for nitrogen fixation. For example, GroEL3 and GroES3 may modulate the observed temperature-dependent reversible inactivation of the NifA protein, by preventing its aggregation and/or mediating the correct folding of nitrogenase itself. The observation that NifA, as well as three subunits of the nitrogenase complex, NifD, NifH, and NifK, can be individually coimmunoprecipitated in a complex with the *Rhizobium* cpn60 is in support of this idea (Govezensky *et al.*, 1991). However, analysis of mutant *B. japonicum* bacteria containing individually inactivated *groESL* operons (caused by a Tn*5* insertion) revealed that none of these operons are essential for soybean root nodulation and nitrogen fixation under symbiotic conditions, suggesting that none of the chaperonins possesses an essential highly specialized activity. The high percentage of amino acid identity between GroEL2 and GroEL3 (90–91%) suggests that certain chaperonins may be able to replace one another functionally, as shown for the SSA subfamily of hsp70-like genes in yeast (reviewed by Craig *et al.*, 1994). Four members of this family contain very similar nucleotide sequences and each of the various HSP70 proteins can functionally substitute, at least partly, for the absence of the other three. The SSA genes are differentially regulated in response to growth temperature, whereas mutations in some SSA genes influence the expression of the heat shock response itself.

B. Inverted Repeat Element: An Alternative Regulatory Mechanism

Although chaperonin synthesis is enhanced upon heat shock in all bacteria examined thus far, the use of a σ^{32}-like transcription factor appears not to be universal. For example, in gram-positive bacteria, like *B. subtilis,* the *groESL* operon contains a characteristic sequence (TTGAAA-17 bp-TATAT) that is recognized by the vegetative sigma factor σ^{43}, which mediates the initiation of transcription at all temperatures (Li and Wong, 1992; Schmidt *et al.*, 1992). However, when *B. subtilis* cells are shifted from 37 to 40°C, the amount of *groESL* mRNA

increases dramatically within 5 min, and this increase is followed by a gradual decline for the next 15 min to a steady-state level higher than that at 37°C, suggesting a regulatory mechanism for heat shock gene expression different than that described for *E. coli* (see Section IV). The DNA sequence around the transcriptional start site of the *B. subtilis groESL* operon contains a perfect IR element consisting of 9 nucleotides, separated by 9 nucleotides. Almost identical IR sequences have been found in the chaperonin *groE(S)L* operons of cyanobacteria, purple sulfur bacteria, and mycobacteria as well as the gram-negative bacterium *Agrobacterium tumefaciens* (Table IV). In addition, IR elements are present in the *dnaK* operon of many gram-positive bacteria, such as *B. subtilis* (Wetzstein *et al.,* 1992), in the *groESL* operon of *Clostridium acetobutylicum* (Narberhaus and Bahl, 1992), and in the *dnaJ* gene of *Lactococcus lactis* (van Asseldonk, 1993), suggesting that the IR element may be involved in the regulation of gene expression as part of the heat shock response.

Zuber and Schumann (1994) have reported that different point mutations in the left, the right, or both "arms" of the IR element present in the *dnaK* operon of *B. subtilis* result in constitutive high levels of both *dnaK* mRNA and DnaK protein at 30°C, suggesting that the IR element acts as a repressor of *dnaK* operon transcription. In agreement with this, when the IR element is inserted between the promoter and the start site of the structural *lacZ* gene in *B. subtilis,* the β-galactosidase activity is reduced fourfold, demonstrating that the IR element is both necessary and sufficient for repression. When a frame-shift mutation alters the nucleotide sequence, but not the secondary structure of the IR element, similar constitutive high levels of *dnaK* gene expression at 30°C are observed, suggesting that the IR element is specifically bound by a *trans*-acting factor, such as a protein. If such a protein exists, it can function at either the DNA or the RNA level and limit heat shock gene expression at 30°C through binding to this IR element.

In most cases reported, a single IR is present between the promoter sequence and the start site of the structural gene. However, a unique case is presented by the *groESL (groEL1)* operon and the single *groEL (groEL2)* gene of *Streptomyces albus* and *S. coelicolor,* which all contain a region accommodating two identical IR elements, one downstream of the transcriptional start site and the second within the promoter region (Guglielmi *et al.,* 1993; Duchêne *et al.,* 1994, and Table IV). In both bacteria three mRNA species corresponding to *groESL1, groEL2,* and *groES* are found. The latter is probably the result of either transcriptional

termination or processing of the full-length bicistronic *groESL1* mRNA at the distinct IR element located between the *groES* and the *groEL1* genes. In *S. coelicolor,* synthesis of *groESL1* and *groEL2* mRNA increases and remains high even up to 2 h after a temperature shift, whereas *groES* mRNA synthesis is induced and turned off much faster.

In contrast to the single IR element present in the *dnaK* gene of *B. subtilis* (see above), the double IR element of the *S. albus groESL* operon does not impose a heat-inducible expression of a reporter gene, suggesting that other parts of the *groESL1* gene sequence may also be required for thermal regulation (Servant *et al.,* 1994). Analysis of a series of *groEL–neo* (the *neo* gene encodes aminoglycoside 3′-phosphotransferase, APH) gene fusions, integrated at their homologous sites in the *S. albus* chromosome, revealed a nucleotide region, encoding amino acids 5 to 471 of the GroEL polypeptide that acts as a negative element. Consistent with this interpretation, when the double IR element of the *S. albus groESL–neo* fusion gene is replaced by the *E. coli lac* promoter, heat-regulated expression of the GroEL–APH fusion protein is still observed. However, heat-regulated expression of the APH protein is abolished when the *groESL–neo* transcriptional fusion contains an additional ribosome binding site in front of the *neo* gene. These observations demonstrate that an internal transcribed region of the *groEL* gene is both necessary and sufficient for repression of the *groESL1* operon of *S. albus.*

Although the double IR element is not essential for repression it may be involved in transcriptional control. Experiments using the regulatory double IR element of the *groESL1* operon of *S. coelicolor,* as well as a synthetic single IR element, have revealed that retarded electrophoretic migration of both DNA fragments occurs only in the presence of a specific protein fraction, suggesting that the IR element acts as a protein binding site (Duchêne *et al.,* 1994). It is not known whether the protein binds to one or both of the IRs present in the double IR region. Surprisingly, the same level of DNA binding activity is present in protein extracts isolated from cells before and after heat shock. This situation may be similar to that reported for the heat-shock activator factor (HSF) of yeast. Although HSF binds to the promoter region with the same efficiency in both heat-shocked and control cells, the rate of transcription is modulated by changes in the phosphorylation state of the HSF (Sorger *et al.,* 1987).

A single IR element is present upstream of the vegetative promoter sequence of the monocistronic *dnaJ* gene of *L. lactis* (van Asseldonk *et*

al., 1993). Deletion of this IR element results in constitutive high-level expression of a reporter gene at 30°C. It is noteworthy that this IR element is positioned only 3 nucleotides upstream of the TTGCCA −35 promoter region, and should therefore interfere with the efficient binding of the RNA polymerase holoenzyme, consequently influencing the rate of transcription.

In summary, the 9-bp IR element most likely acts as a negative element influencing both transcription and translation of the structural gene(s) it precedes. The quantitative and qualitative effect may vary from organism to organism and from gene to gene. Clearly, the extent and level (DNA or RNA) of regulation depend on the position and the number of IR elements as well as on the structural gene itself. Interestingly, the double IR element of the *groESL* operon of *S. coelicolor* is sufficient to direct chaperonin synthesis in *E. coli* (A. Labigne, personal communications), supporting the suggestion that a related sequence in the untranscribed IR element (Guglielmi *et al.*, 1993, and Table IV) may be recognized by the *E. coli* $E\sigma^{70}$ polymerase, and indicating that these regulatory sequences may have a common ancestor.

C. Parasitic Bacteria and Other Intracellular Survivors

Certain parasitic bacteria survive intracellularly within their host. The human pathogen *Legionella pneumophila* normally lives in freshwater protozoa, but when taken up by macrophages, present in specialized lung tissue, multiplies within a ribosome-studded membrane-bound vesicle (Horwitz, 1983). The survival of a parasite is controlled by multiple regulons, and in the case of *Salmonella typhimurium* the intracellular response pattern is similar, yet not identical, to the global stress responses induced by heat, osmotic shock, or hydrogen peroxide (Abshire and Neidhardt, 1993). In response to the intracellular environment of the macrophage, both *L. pneumophila* and *S. typhimurium* dominantly synthesize two proteins: cpn60 and DnaK (Buchmeier and Heffron, 1990; Kwaik *et al.*, 1993). The purpose of the large quantity of endosymbiotic cpn60 relative to cpn10 is not known, but may reflect specific physiological needs. In most cases it has been found that processes mediated by cpn60 *in vivo* also require the function of cpn10 (see Chapter 8). However, *in vitro* cpn60 alone is capable of maintaining precursor polypeptides in conformations compatible with subsequent transport across

biological membranes and protecting polypeptides against heat denaturation (Bochkareva *et al.*, 1988; Lecker *et al.*, 1989; Phillips and Silhavy, 1990; Ziemienowicz *et al.*, 1993). In this model, the precursor is released from the cpn60 by some component of the transport machinery rather than by cpn10. In this respect it is interesting that cpn60 associates with the bacteroidal membrane of *L. pneumophila* 24 h after infection of HeLa cells (Hoffman *et al.*, 1990), whereas Kakeda and Ishikawa (1991) have suggested that the endosymbiontic cpn60 of the gram-negative prokaryotic *Acyrthosiphon pisum* (also known as "symbionin") may be required for translocation of proteins across the bacteroidal membrane. *Acyrthosiphon pisum* lives in a specialized differentiated aphid fat body cell (Buchner, 1965) and needs cpn60 to sustain its bacteroidal structure (Ohtaka *et al.*, 1992). Large quantities of cpn60 may alternatively be required because cpn60 is secreted from the endosymbiont, as has been reported for the human pathogen *Helicobacter pylori* (Evans *et al.*, 1992). Cpn60 is one of the five surface-associated proteins that are essential for intracellular survival, and are secreted when *H. pylori* is cultured on agar plates. Hoffman *et al.* (1990) showed that in contrast to *E. coli*, where cpn60 is restricted to the cytoplasm, in *L. pneumophila* bacteria hosted by HeLa cells, cpn60 is present in the cytoplasm and either the periplasm or on the cell surface, as seen by fluorescence microscopy using antibodies raised against cpn60.

Although little is known about the regulation of chaperonin gene expression in these endosymbiotic bacteria, in *L. pneumophila* and *A. pisum* the *groESL* operon contains a nucleotide sequence that resembles the *E. coli* heat shock promoter (Table III; Hoffman *et al.*, 1990; Ohtaka *et al.*, 1992). Although these endosymbiotic promoter sequences are functional in *E. coli*, it is not known how they regulate chaperonin gene expression during symbiosis. The short IR element (7 nucleotides with a 7-nucleotide spacer), found just upstream of the ribosome binding site of the *groEL* gene of *A. pisum*, may function as a positive regulatory element to enhance the synthesis of cpn60 relative to cpn10 (Ohtaka *et al.*, 1992).

The *groESL* operon of the human pathogen *Mycobacterium tuberculosis* contains a distinctly different long-inverted repeat, which may also be involved in gene regulation. This distinct IR element consists of two 21 nucleotide sequences with a 5-nucleotide spacer and separates the *groES* from the *groEL* gene (Baird *et al.*, 1989; Kong *et al.*, 1993). Although a putative *E. coli* $E\sigma^{32}$-like heat shock promoter is present upstream of the *groES* gene, the observed mRNA contains only the

structural *groEL* gene sequence plus the preceding ribosome binding site (Kong *et al.*, 1993), thus resulting in the synthesis of large amounts of cpn60 relative to cpn10. The 5′ end of this *groEL* transcript maps just downstream of the 21-nucleotide IR element, suggesting that the *groEL* mRNA may be produced by the specific cleavage of a larger transcript. This long IR element may thus be part of a novel regulatory mechanism that controls the differential synthesis of the chaperonins by acting as a RNase enzyme cleavage site. Inhibition of *groES* translation could then subsequently be achieved by progressive degradation of the remaining shorter transcript in the 3′ to 5′ direction. Such a retroregulatory control has been shown to occur in the synthesis of the Int protein of bacteriophage λ (reviewed in Echols and Guarneros, 1983). In this case a long IR element, termed *sib* (18 nucleotides separated by a 14 nucleotide spacer), was proposed to regulate the differential expression of the *int* and *xis* genes that are cotranscribed from the P_L promoter.

VI. MODULATION OF CHAPERONIN ACTIVITY BY BACTERIOPHAGE Gp31 PROTEIN

Both cpn60 and cpn10 are required for the ATP-dependent folding of many unrelated proteins *in vivo,* since newly synthesized polypeptide chains collapse rapidly into partly folded compact intermediates that display hydrophobic regions and hence tend to aggregate. The potential to form such aggregates varies with each individual polypeptide but invariably increases with temperature and the concentration of the folding polypeptide chains. Studies with purified proteins have shown that the correct refolding of certain polypeptides is facilitated by cpn60 and ATP alone (see Chapters 7 and 8). It should be appreciated, however, that in most of these cases the presence of cpn10 does increase the yield of the refolding reaction. There is strong genetic and biochemical evidence that correct folding *in vivo* requires both cpn60 and cpn10 (Goloubinoff *et al.*, 1989; Van Dyk *et al.*, 1989; Horwich *et al.*, 1993). So far the only exception to this rule is the bacteriophage T4 capsid protein Gp23, which fails to assemble correctly in *E. coli* and requires the bacteriophage-encoded protein Gp31 instead of GroES (Laemmli *et al.*, 1970; Georgopoulos *et al.*, 1972; Takano and Kakefuda, 1972). Other bacteriophages, such as λ and T5, use the *E. coli* host cpn60 and cpn10 for the assembly of their capsid and tail structures.

We have shown that Gp31 is a functional analog of the *E. coli* cpn10 (van der Vies *et al.*, 1994). Despite the lack of obvious sequence similarity with GroES, Gp31 can substitute for GroES in the assembly of bacteriophage λ and T5 in *E. coli*, and in the cpn60-dependent folding of the bacterial enzyme ribulose-bisphosphate carboxylase, both *in vivo* and *in vitro*. The absolute requirement of Gp31 for the correct assembly of the Gp23 capsid protein *in vivo* suggests that Gp31 may possess specific cochaperonin properties that are absent in GroES, or alternatively may function similarly to GroES but with improved efficiency.

The identification of Gp31 as a cochaperonin raises the interesting question of whether other specialized chaperonins exist that handle proteins whose folding requires specific features. Although there is no other reported example of this, the observation that chaperonins are under differential transcriptional control in *B. japonicum* and in certain endosymbionts (see Section V,B and V,C) could be explained in this way.

VII. CONCLUDING REMARKS

Not only are chaperonins ubiquitous, they are also involved in a diverse array of cellular processes that make them essential for cell viability. From the data reviewed here it is clear that the regulation of chaperonin synthesis in bacteria is determined by various physiological needs and requirements, and although the mechanisms of chaperonin gene expression are different in the various bacteria examined, chaperonin synthesis is invariably enhanced during stress. Further studies on chaperonin gene expression in bacteria, especially in those possessing multiple and different chaperonin genes, will contribute to our understanding of chaperonin requirement and the mechanisms of chaperonin action and regulation.

ACKNOWLEDGMENTS

We thank A. Glatz, E. Kovács, A. Labigne, P. Lund, P. Madozier, W. Schumann, and L. Vigh for helpful discussions and access to unpublished results. Our work has been supported by the Swiss National Science Foundation (FN-31-31129-91) and the Canton of Geneva.

REFERENCES

Abshire, K. A., and Neidhardt, F. C. (1993). Analysis of proteins synthesized by *Salmonella typhimurium* during growth within a host macrophage. *J. Bacteriol.* **175,** 3734–3743.

Baird, P. N., Hall, L. M., and Coates, A. R. (1989). Cloning and sequence analysis of the 10 kDa antigen gene of *Mycobacterium tuberculosis*. *J. Gen. Microbiol.* **135,** 931–939.

Ballard, S. A., Segers, R. P. A. M., Bleumink-Pluym, N., Fyfe, J., Faine, S., and Adler, B. (1993). Molecular analysis of the *hsp (groE)* operon of *Leptospira interrogans* serovar *copenhageni*. *Mol. Microbiol.* **8,** 739–751.

Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. (1988). Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. *Nature (London)* **336,** 254–257.

Buchmeier, N. A., and Heffron, F. (1990). Induction of *Salmonella* stress proteins upon infection of macrophages. *Science* **248,** 730–732.

Buchner, P. (1965). *In* "Endosymbiosis of Animals with Plant Microorganisms." Interscience, New York.

Bukau, B. (1993). Regulation of the *E. coli* heat shock response. *Mol. Microbiol.* **9,** 671–680.

Chitnis, P. R., and Nelson, N. (1991). Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **266,** 58–65.

Cowing, D. W., Bardwell, J. C. A., Craig, E. A., Woolford, C., Hendrix, R. W., and Gross, C. A. (1985). Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc. Natl. Acad. Sci. USA* **82,** 2679–2683.

Craig, E. A., and Gross, C. A. (1991). Is hsp70 the cellular thermometer? *Trends Biochem. Sci.* **16,** 135–140.

Craig, E. A., Baxter, B. K., Becker, J., Halladay, J., and Ziegelhoffer, T. (1994). Cytosolic hsp70s of *Saccharomyces cerevisiae:* Roles in protein synthesis, protein translocation, proteolysis, and regulation. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp 31–52. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Duchêne, A.-M., Thompson, C. J., and Mazodier, P. (1994). Transcriptional analysis of *groEL* genes in *Streptomyces coelicolor* A3 (2). *Mol. Gen. Genet.* **245,** 61–68.

Echols, H., and Guarneros, B. (1983). Control of integration and excision. *In* "LAMBDA II" (R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, eds.), pp. 75–92. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Erickson, J. W., and Gross, C. A. (1989). Identification of the σ^E subunit of *Escherichia coli* RNA polymerase: A second alternate σ factor involved in high-temperature gene expression. *Genes Dev.* **3,** 1462–1471.

Evans, D. J., Jr., Evans, D. G., Engstrand, L., and Graham, D. Y. (1992). Urease-associated heat shock protein of *Helicobacter pylori*. *Infect. Immun.* **60,** 2125–2127.

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171,** 1379–1385.

Ferreyra, R. G., Soncini, F. C., and Viale, A. M. (1993). Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the phototrophic sulphur bacterium *Chromatium vinosum*. *J. Bacteriol.* **175,** 1514–1523.

Fischer, H. M., Babst, M., Kaspar, T., Acuna, G., Arigoni, F., and Hennecke, H. (1993). One member of a *groESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. *EMBO J.* **12,** 2901–2912.

Fisher, R. F., and Long, S. R. (1992). Rhizobium–plant signal exchange. *Nature* (*London*) **357,** 655–659.

Friedman, D. I., Olson, E. R., Tilly, K., Georgopoulos, C., Herskowitz, I., and Banuett, F. (1984). Interactions of bacteriophage λ and host macromolecules in the growth of bacteriophage λ. *Microbiol. Rev.* **48,** 299–325.

Georgopoulos, C., and Eisen, H. (1974). Bacterial mutants which block phage assembly. *J. Supramol. Struct.* **2,** 349–359.

Georgopoulos, C., and Welch, W. J. (1993). Role of major heat shock proteins as molecular chaperones. *Annu. Rev. Cell. Biol.* **9,** 601–635.

Georgopoulos, C., Hendrix, R. W., Kaiser, A. D., and Wood, W. B. (1972). Role of the host cell in bacteriophage morphogenesis: Effects of a bacterial mutation on T4 head assembly. *Nature New Biol.* **239,** 38–41.

Georgopoulos, C., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Georgopoulos, C., Ang, D., Liberek, K., and Zylicz, M. (1990). Properties of the *E. coli* heat shock proteins and their role in bacteriophage lambda growth. *In* "Stress Proteins in Biology and Medicine" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp. 191–221. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Goloubinoff, P., Gatenby, A. A., and Lorimer, G. H. (1989). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli. Nature* (*London*) **337,** 44–47.

Gor, D., and Mayfield, J. E. (1992). Cloning and nucleotide sequence of the *Brucella abortus groE* operon. *Biochim. Biophys. Acta* **1130,** 120–122

Govezensky, D., Greener, T., Segal, G., and Zamir, A. (1991). Involvement of GroEL in *nif* gene regulation and nitrogenase assembly. *J. Bacteriol.* **173,** 6339–6346.

Gross, C. A., Straus, D. B., Erickson, J. W., and Yura, T. (1990). The function and regulation of heat shock proteins in *Escherichia coli. In* "Stress Proteins in Biology and Medicine" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp 167–189. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Grossman, A. D., Erickson, J. W., and Gross, C. A. (1984). The *htpR* gene product of *E. coli* is a sigma factor for heat shock promoters. *Cell* **38,** 383–390.

Grossman, A. D., Straus, D. B., Walter, W. A., and Gross, C. A. (1987). σ^{32} synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli. Genes Dev.* **1,** 179–184.

Guglielmi, G., Mazodier, P., Thompson, C. J., and Davies, J. (1991). A survey of the heat shock response in four *Streptomyces* species reveals two *groEL*-like genes and three GroEL-like proteins in *Streptomyces albus. J. Bacteriol.* **173,** 7374–7381.

Guglielmi, G., Duchêne, A.-M., Thompson, C., and Mazodier, P. (1993). Transcriptional analysis of two different *Streptomyces albus groEL*-like genes. *In* "Industrial Microorganisms: Basic and Applied Molecular Genetics" (R. H. Baltz, G. D. Hegeman, and P. Ll. Skatrud, eds.), pp 17–24. American Society for Microbiology, Washington DC.

Gussin, G. N., Ronson, C. W., and Ausubel, F. M. (1986). Regulation of nitrogen fixation genes. *Annu. Rev. Genet.* **20,** 567–591.

Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* (*London*) **333,** 330–334.

Herendeen, S. L., VanBogelen, R. A., and Neidhardt, F. C. (1979). Levels of major proteins of *Escherichia coli* during growth at different temperatures. *J. Bacteriol.* **139,** 185–194.

Hoffman, P. S., Butler, C. A., and Quinn, F. D. (1989). Cloning and temperature-dependent expression in *Escherichia coli* of a *Legionella pneumophila* gene coding for a genus-common 60-kilodalton antigen. *Infect. Immun.* **57,** 1731–1739.

Hoffman, P. S., Houston, L., and Butler, C. A. (1990). *Legionella pneumophila htpAB* heat shock operon: Nucleotide sequence and expression of the 60-kilodalton antigen in *L. pneumophila*-infected HeLa cells. *Infect. Immun.* **58,** 3380–3387.

Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N., and Furtak, K. (1993). Folding *in vivo* of bacterial cytoplasmic proteins: Role of GroEL. *Cell* **74,** 909–917.

Horwitz, M. A. (1983). Formation of a novel phagosome by the Legionnaires' disease bacterium (*Legionella pneumophila*). *J. Exp. Med.* **158,** 1319–1331.

Johnson, R. B., Fearon, K., Mason, T., and Jindal, S. (1989). Cloning and characterization of the yeast chaperonin *HSP60* gene. *Gene* **84,** 295–302.

Kakeda, K., and Ishikawa, H. (1991). Molecular chaperone produced by an intracellular symbiont. *J. Biochem.* **110,** 583–587.

Kong, T. H., Coates, A. R. M., Butcher, P. D., Hickman, C. J., and Shinnick, T. M. (1993). *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc. Natl. Acad. Sci. USA* **90,** 2608 2612.

Kovács, E., Török, Z., Horváth, I., and Vigh, L. (1994a). Heat stress induces association of the GroEL-analog chaperonin with thylakoid membranes in cyanobacterium, *Synechocystis* PCC 6803. *Plant Physiol. Biochem.* **32,** 285–293.

Kovács, E., Horváth, I., Glatz, A., Török, Zs., Bagyinka, Cs., and Vigh, L. (1994b). Molecular characterization assembly and membrane association of the GroEL-type chaperonins in *Synechocystis* PCC6803. *In* "NATO ASI Series" (J. A. F. Op den Kamp, ed.), Vol. H 82, pp 253–261. Springer-Verlag, Berlin/Heidelberg.

Kwaik, Y. A., Eisenstein, B. I., and Engelberg, N. C. (1993). Phenotypic modulation by *Legionella pneumophila* upon infection of macrophages. *Infect. Immun.* **61,** 1320–1329.

Laemmli, U. K., Beguin, F., and Guyer-Kellenberger, G. (1970). A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.* **47,** 69–85.

Landick, R., Vaugh, V., Tau, E. T., VanBogelen, R. A., Erickson, J. W., and Neidhardt, F. C. (1984). Nucleotide sequence of the heat shock regulatory gene of *E. coli* suggests its protein product may be a transcription factor. *Cell* **38,** 175–182.

Lecker, S., Lill, R., Ziegelhoffer, T., Georgopoulos, C., Bassford, Jr., P. J., Kumamoto, C. A., and Wickner, W. (1989). Three pure chaperone proteins of *Escherichia coli,* secB, trigger factor and groEL, form soluble complexes with precursors proteins *in vitro. EMBO J.* **8,** 2703–2709.

Lehel, Cs., Los, D., Wada, H., Györgyei, J., Horváth, I., Kovács, E., Murata, N., and Vigh, L. (1993). A second *groEL*-like gene, organized in a *groESL* operon is present in the genome of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **268,** 1799–1804.

Li, M., and Wong, S.-L. (1992). Cloning and characterization of the *groESL* operon from *Bacillus subtilis. J. Bacteriol.* **174,** 3981–3992.

Lin, J., Adams, L. G., and Ficht, T. A. (1992). Characterization of the heat shock response in *Brucella abortus* and isolation of the genes encoding the GroE heat shock proteins. *Infect. Immun.* **60,** 2425–2431.

Lindquist, S. (1986). The heat-shock response. *Annu. Rev. Biochem.* **55,** 1151–1191.

Lindquist, S., and Craig, E. A. (1988). The heat-shock proteins. *Annu. Rev. Genet.* **22,** 631–677.

Lissin, N. M., Venyaminov, S. Yu., and Girshovich, A. S. (1990). (Mg-ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature* (*London*) **348,** 339–342.

Maid, U., Steinmüller, R., and Zetsche, K. (1992). Structure and expression of a plastid-encoded *groEL* homologous heat-shock protein gene in a thermophilic unicellular red alga. *Curr. Genet.* **21,** 521–525.

Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C. J. (1991). Characterization of the *groEL*-like genes in *Streptomyces albus*. *J. Bacteriol.* **173,** 7382–7386.

McMullin, T. W., and Hallberg, R. L. (1987). A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **7,** 4414–4423.

Mehra, V., Sweetser, D., and Young, R. A. (1986). Efficient mapping of protein antigenic determinants. *Proc. Natl. Acad. Sci. USA* **83,** 7013–7017.

Morett, E., and Buck, M. (1989). *In vivo* studies on the interaction of RNA polymerase-σ^{54} with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. *J. Mol. Biol.* **210,** 65–77.

Morett, E., Fischer, H.-M., and Hennecke, H. (1991). Influence of oxygen on DNA binding, positive control, and stability of the *Bradyrhizobium japonicum* NifA regulatory protein. *J. Bacteriol.* **173,** 3478–3487.

Narberhaus, F., and Bahl, H. (1992). Cloning, sequencing, and molecular analysis of the *groESL* operon of *Clostridium acetobutylicum*. *J. Bacteriol.* **174,** 3282–3289.

Neidhardt, F. C., and VanBogelen, R. A. (1987). Heat shock response. *In "Escherichia coli* and *Salmonella typhimurium"*: Cellular and Molecular Biology (F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, eds.), pp. 1334–1345. American Society for Microbiology, Washington DC.

Ohta, T., Honda, K., Saito, K., Hayashi, H., Tano, H., Hamamoto, T., and Kagawa, Y. (1993). Heat shock promoter of thermophilic chaperone operon. *Biochem. Biophys. Res. Commun.* **191,** 550–557.

Ohtaka, C., Nakamura, H., and Ishikawa, H. (1992). Structures of chaperonins from an intracellular symbiont and their functional expression in *Escherichia coli groE* mutants. *J. Bacteriol.* **174,** 1869–1874.

Parsons, L. M., Waring, A. L., and Shayegani, M. (1992). Molecular analysis of the *Haemophilus ducreyi groE* heat shock operon. *Infect. Immun.* **60,** 4111–4118.

Phillips, G. J., and Silhavy, T. J. (1990). Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in *E. coli*. *Nature* (*London*) **344,** 882–884.

Pose, P. S. (1993). "Molecular Characterization of Two Nuclear Genes Encoding *Zea mays* Mitochondrial Chaperonin 60." Genbank database Release 82.0 April 1994, Locus: MZECPN60A.

Raina, S., and Georgopoulos, C. (1990). The identification and characterisation of a new heat shock gene *htrC*, whose product is essential for *Escherichia coli* viability at high temperatures. *J. Bacteriol.* **172,** 3417–3426.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast *HSP60* gene coding for a mitochondrial assembly factor. *Nature* (*London*) **337,** 655–659.

Rospert, S., Junne, T., Glick, B. S., and Schatz, G. (1993). Cloning and disruption of the gene encoding yeast mitochondrial chaperonin 10, the homolog of *E. coli groES*. *FEBS Lett.* **335,** 358–360.

Rusanganwa, E., and Gupta, S. (1993). Cloning and characterization of multiple *groEL* chaperonin-encoding genes in *Rhizobium meliloti*. *Gene* **126,** 67–75.

Schmidt, A., Schiessswohl, M., Völker, U., Hecker, M., and Schumann, W. (1992). Cloning, sequencing, mapping, and transcriptional analysis of the *groESL* operon from *Bacillus subtilis*. *J. Bacteriol.* **174,** 3993–3999.

Schön, U., and Schumann, W. (1993). Molecular cloning, sequencing, and transcriptional analysis of the *groESL* operon from *Bacillus stearothermophilus*. *J. Bacteriol.* **175,** 2465–2469.

Segal, G., and Ron, E. Z. (1993). Heat shock transcription of the *groESL* operon of *Agrobacterium tumefaciens* may involve a hairpin-loop structure. *J. Bacteriol.* **175,** 3083–3088.

Servant, P., Tompson, C., and Mazodier, P. (1993). Use of new *Escherichia coli/Streptomyces* conjugative vectors to probe the functions of two *groEL*-like genes of *Streptomyces albus* G by gene disruption. *Gene* **134,** 25–32.

Servant, P., Thompson, C. J., and Mazodier, P. (1994). Post transcriptional regulation of the *groEL1* gene of *Streptomyces albus*. *Mol. Microbiol.* **12,** 423–432.

Shinnick, T. (1987). The 65-kilodalton antigen of *Mycobacterium tuberculosis*. *J. Bacteriol.* **169,** 1080–1088.

Shinnick, T. M., Plikaytis, B. B., Hyche, A. D., Van Landingham, R. M., and Walker, L. L. (1989). The *Mycobacterium tuberculosis* BCG-a protein has homology with the *Escherichia coli* GroES protein. *Nucleic Acids Res.* **17,** 1252.

Sorger, P. K., Lewis, M. J., and Pelham, H. R. B. (1987). Heat shock factor is regulated differently in yeast and HeLa cells. *Nature* (*London*) **329,** 358–360.

Straus, D. B., Walter, W. A., and Gross, C. A. (1987). The heat shock response of *E. coli* is regulated by changes in the concentration of σ^{32}. *Nature* (*London*) **329,** 348–351.

Takano, T., and Kakefuda, T. (1972). Involvement of a bacterial factor in the morphogenesis of bacteriophage capsid. *Nature New Biol.* **239,** 34–37.

Tamada, H., Ohta, T., Hamamoto, T., Otawara-Hamamoto, Y., Yanagi, M., Hiraiwa, H., Hirata, H., and Kagawa, Y. (1991). Gene structure of heat shock proteins 61 kDa and 12 kDa (thermophilic chaperonins) of thermophilic bacterium *PS3*. *Biochem. Biophys. Res. Commun.* **179,** 565–571.

Tilly, K., Murialdo, H., and Georgopoulos, C. (1981). Identification of a second *Escherichia coli groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl. Acad. Sci. USA* **78,** 1629–1633.

Tilly, K., Spence, J., and Georgopoulos, C. (1989). Modulation of the stability of *Escherichia coli* heat shock regulatory factor σ^{32}. *J. Bacteriol.* **171,** 1585–1589.

van Asseldonk, M., Simons, A., Visser, H., de Vos, W. M., and Simons, G. (1993). Cloning, nucleotide sequence, and regulatory analysis of the *Lactococcus lactis dnaJ* gene. *J. Bacteriol.* **175,** 1637–1644.

van der Vies, S. M., Gatenby, A. A., and Georgopoulos, C. (1994). Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature* (*London*) **368,** 654–656.

Van Dyk, T. K., Gatenby, A. A., and LaRossa, R. A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* (*London*) **342,** 451–453.

Viitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1,** 363–369.

Wallington, E. J., and Lund, P. A. (1994). *Rhizobium leguminosarum* contains multiple chaperonin (cpn60) genes. *Microbiology* **140,** 113–122.

Wang, Q., and Kaguni, J. M. (1989). DnaA protein regulates transcription of the *rpoH* gene of *Escherichia coli*. *J. Bacteriol.* **264,** 7338–7344.

Webb, R., Reddy, K. J., and Sherman, L. A. (1990). Regulation and sequence of the *Synechococcus* sp. strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. *J. Bacteriol.* **172,** 5079–5088.

Welch, W. J. (1993). How cells respond to stress. *Scientific American* **268,** 34–41.

Wetzstein, M., Völker, U., Dedio, J., Löbau, S., Zuber, U., Schiesswohl, M., Herget, C., Hecker, M., and Schumann, W. (1992). Cloning, sequencing, and molecular analysis of the *dnaK* locus from *Bacillus subtilis*. *J. Bacteriol.* **174,** 3300–3310.

Yura, T., Nagai, H., and Mori, H. (1993). Regulation of the heat-shock response in bacteria. *Annu. Rev. Microbiol.* **47,** 321–350.

Zeilstra-Ryalls, J., Fayet, O., and Georgopoulos, C. (1991). The universally conserved GroE chaperonins. *Annu. Rev. Microbiol.* **45,** 301–325.

Zhou, Y.-N., Kusukawa, N., Erickson, J. W., Gross, C. A., and Yura, T. (1988). Isolation and characterization of *Escherichia coli* mutants that lack the heat shock sigma factor σ^{32}. *J. Bacteriol.* **170,** 3640–3649.

Ziemienowicz, A., Skowyra, D., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., and Zylicz, M. (1993). Either of the *Escherichia coli* GroEL/GroES and DnaK/DnaJ/GrpE chaperone machines can reactivate heat-treated RNA polymerase: Different mechanisms for the same activity. *J. Biol. Chem.* **268,** 25,425–25,431.

Zuber, U., and Schumann, W. (1994). CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *bacillus subtilis*. *J. Bacteriol.* **176,** 1359–1363.

7

Kinetic and Energetic Aspects of Chaperonin Function

ANTHONY R. CLARKE* AND PETER A. LUND†

* Department of Biochemistry
University of Bristol
Bristol BS8 1TD, United Kingdom

† School of Biological Sciences
University of Birmingham
Birmingham B15 2TT, United Kingdom

A combination of genetic, cell biological, and *in vitro* renaturation studies shows that chaperonins assist the folding of proteins, an activity required in the cell both to ensure that newly synthesized proteins fold and assemble to their biologically active states and to facilitate recovery from cellular damage. To achieve these ends chaperonins bind to nonnative protein structures and use the energy derived from ATP hydrolysis to drive the formation of native states. In this discussion we take this conclusion as a starting point to examine the physical mechanisms underlying this intriguing activity, but as an opening to such a discussion it is useful to define those general phenomena for which a tenable mechanism must account.

I. GROUND RULES OF CHAPERONIN BEHAVIOR

A. Specificity and Affinity for Protein Substrates

In this chapter our considerations are limited to the homotetradecameric chaperonins and their homoheptameric coproteins, typified by the chaperonin 60 (cpn60) and chaperonin 10 (cpn10) proteins from eubacterial prokaryotes, especially *Escherichia coli.* In these cases there is no evidence of specialization of subunits for particular protein targets or cellular functions, as there may be, for instance, in the case of the TCP1-containing chaperonin complexes, which are discussed in Chapter 5. On the contrary, chaperonins such as GroEL show surprisingly little specificity in the binding of protein substrates. This conclusion is demonstrated by the burgeoning number of unrelated proteins which, when refolding *in vitro* from denatured states, interact with chaperonins to increase the yield of active protein.

In the early stages of chaperonin research it was suggested that their function may be to guide the folding of protein molecules by binding

to, and therefore stabilizing, transiently formed elements of the native structure (Rothman, 1989). Such an idea might be considered heretical by physical biochemists studying spontaneous folding, since it appears to contradict the Anfinsenian dictum that all the information necessary to fold to the native state is contained within the primary structure (Anfinsen, 1973).

In fact, the view that chaperonins may stabilize features of the native state that are unstable in intermediates in the folding process cannot, *a priori,* be ruled out. However, the features that are recognized in such a mechanism must be crude and general, owing to the fact that proteins with unrelated topologies and sequences have their folding yields improved by chaperonins (Jaenicke, 1993). Further to this argument, the members of a family of proteins with identical topologies and subunit arrangements can behave quite differently from one another with respect to their response to chaperonins during folding (Staniforth *et al.,* 1994a). Some support for the possibility that elements of secondary structure are stabilized by cpn60 was furnished by nuclear magnetic resonance (NMR) measurements which showed that a peptide capable of forming an amphipathic α helix of the type commonly encountered in globular proteins adopted a random conformation in free solution and a helical one when bound to cpn60 (Landry and Gierasch, 1991; Landry *et al.,* 1992). The general importance of this observation was later diluted by the demonstration that a protein that contained only β-sheet elements of secondary structure was a good substrate for chaperonin-enhanced folding (Schmidt and Buchner, 1992). The NMR data, however, were valuable in showing that binding is achieved by contact between hydrophobic side chains on the peptide and the chaperonin surface rather than through hydrogen bonding interactions with backbone peptide groups accessible only in extended conformations. Further evidence that chaperonins bind to protein substrates by hydrophobic forces is provided by the observation that such interactions are strengthened by high ionic concentrations and weakened by low temperatures (Hansen and Gafni 1993; Brunschier *et al.,* 1993; Schmidt *et al.,* 1994a). In addition, when a random mixture of tryptic peptides derived from a cell extract of *E. coli* is incubated with cpn60, those bound preferentially are rich in the larger aliphatic and aromatic amino acids (N. A. Ranson and A. R. Clarke, unpublished data, 1993), and even among individual amino acids, the most hydrophobic are bound tightest (Richarme and Kohiyama, 1994). Mutagenic experiments (Fenton *et al.,* 1994) interpreted in the light of the crystal structure of GroEL (Braig *et al.,* 1994) confirm this hypothesis

by demonstrating that the binding of protein substrates is impaired by substituting hydrophobic residues on the surface of the apical domain for hydrophilic ones (see Chapter 9).

These observations, and the general fact that chaperonins bind to any unfolded protein, albeit with varying affinities, lead to the conclusion that they recognize nothing more specific than the nonpolar regions of proteins which are normally buried in the native state. By this property alone they must bind to and stabilize unfolded states in preference to folded ones; dissociation constants for the interaction of cpn60 with such states are in the nanomolar (Burston *et al.*, 1992) or even the picomolar (G. H. Lorimer and P. V. Viitanen, personal communication, 1993) range. The simple binding and folding processes involved in chaperonin-mediated folding are illustrated in Fig. 1.

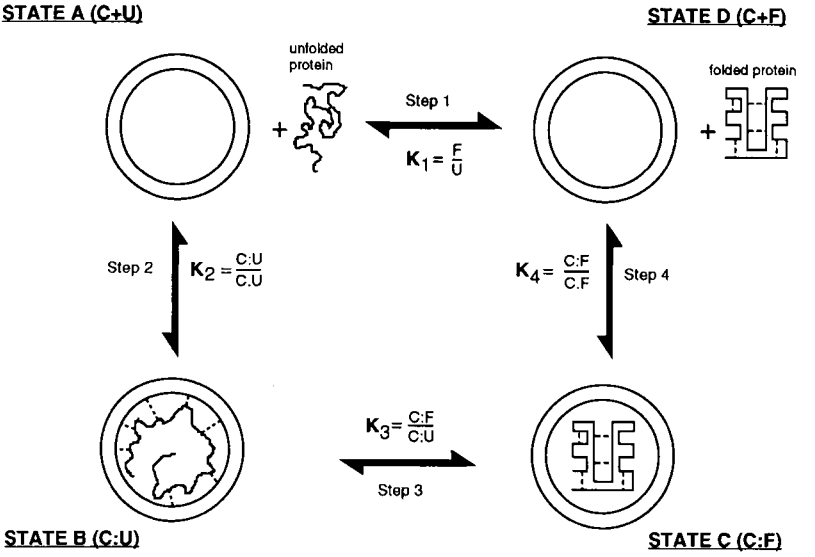


Fig. 1. Binding and folding of protein substrates: the thermodynamic cycle. The diagram depicts the basic equilibrium states of apo-cpn60 and a protein substrate. The protein substrate is shown in just two forms; unfolded (with a high affinity for the cpn60 cavity) and folded (low affinity). Step 1 represents spontaneous folding of the protein in free solution, step 2 is the tight association of the unfolded form with cpn60, step 3 shows the folding process on the surface of the cpn60, and step 4 the association/dissociation of the folded substrate. Note that binding equilibria are described by association constants. C, Chaperonin; U, unfolded protein; F, folded protein.

B. Release of Bound Proteins

In chaperonin-facilitated folding, the protein substrate must bind and dissociate, perhaps many times (Weissman *et al.,* 1994), during the process. In terms of the protein association event, it has been widely demonstrated that the most effective way of forming a strong chaperonin : protein complex is to unfold the substrate completely and mix it with cpn60 alone (Goloubinoff *et al.,* 1989; Badcoe *et al.,* 1991; Buchner *et al.,* 1991; Martin *et al.,* 1991; Gray and Fersht, 1993). All proteins when confronted with these conditions either (i) remain bound in an unfolded state (Goloubinoff *et al.,* 1989; Buchner *et al.,* 1991; Martin *et al.,* 1991) or (ii) show a much retarded rate of folding owing to the kinetic consequences of tight interaction with unfolded forms (Badcoe *et al.,* 1991; Gray and Fersht, 1993). In the scheme represented in Fig. 1, the former kind of substrate is locked in the unfolded state (B), whereas the latter is able to proceed to the free, native state (D) because the intrinsic folding energy (step 1) is sufficient to overcome the binding energy between cpn60 and the unfolded state (step 2) at the protein concentrations used in the experiment. This behavior is illustrated by two, homologous dimeric proteins—*Bacillus stearothermophilus* lactate dehydrogenase (bLDH) and mitochondrial malate dehydrogenase (mMDH). The former is extremely stable ($\Delta G = -12.5$ kcal/mol) and can fold in the presence of 50 n*M* cpn60 (Badcoe *et al.,* 1991), whereas the latter is less stable ($\Delta G = -2.5$ kcal/mol) and cannot (Staniforth *et al.,* 1994a). The native state of bLDH is 10 kcal/mol (or 2×10^7 times) more stable than that of mMDH. In order to stabilize the unfolded form of bLDH to the same degree as mMDH, the cpn60 would have to bind to it 20 million times more tightly.

Even when the folding energy is sufficient to overcome that of binding to unfolded states, the kinetics of folding in the presence of cpn60 are found to be slow. There are two general explanations for this. Either (i) the protein cannot fold while bound to cpn60 and the proportion of free, unfolded protein able to undergo the transition (Fig. 1, step 1) to the native state is small and/or the rate of dissociation is slow or (ii) the protein can fold on cpn60 (step 3) but must break intermolecular bonds in order to form the intramolecular contacts necessary for acquisition of the native structure. In this latter case the energy barrier governing folding through step 3 is higher than that through step 1 (illustrated in Fig. 2). Whether proteins are able to fold while in contact with the

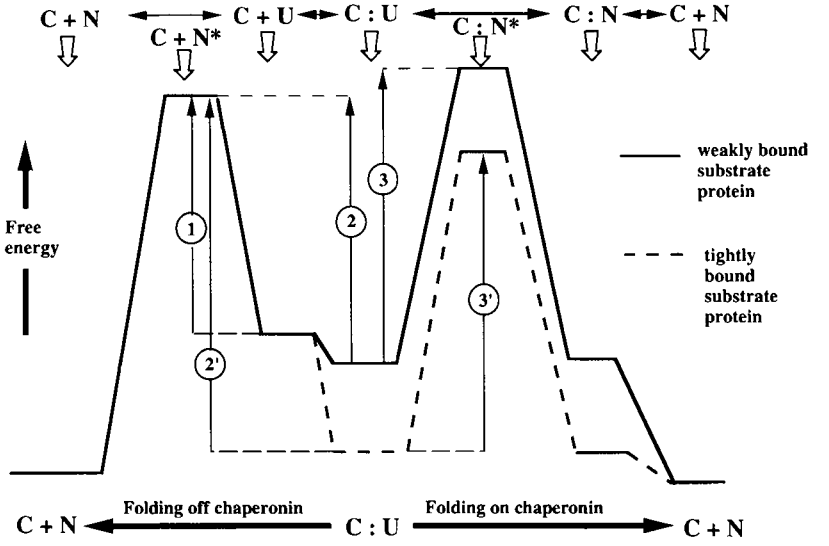


Fig. 2. The energetics of protein folding in the presence of cpn60 As shown in Fig. 1, folding can proceed either in free solution or on the surface of cpn60. The above diagram shows the free energy profiles for these processes referred to in the text. Note that, owing to the preferential stabilization of the unfolded form, the barriers to folding are always higher when folding occurs in the presence of cpn60 (2, 2′, 3, and 3′) compared to folding in free solution (1). The route from the bound unfolded state (C:U) to the free native state in the off-chaperonin process, i.e., by dissociation, has a rate governed by barriers 2 and 2′, the former for weak binding conditions, the latter for tight. The on-chaperonin route occurs at a rate determined by barriers 3 and 3′ for weak and tight binding conditions, respectively. The profiles illustrate that the preferred route (lowest barrier) will depend on the tightness of the interaction and/or the concentration of cpn60. C, cpn60; N, folded protein; U, unfolded protein; N, folding transition state.

surface of cpn60 is an often discussed point and may have a strong bearing on the mechanism. In the case of barnase (the ribonuclease enzyme from *B. amyloliquefaciens*) this process has been shown to occur by the fact that there is a slow but finite folding rate even at an extrapolated infinite concentration of cpn60. Under these conditions there is no protein substrate in free solution, so folding must occur while it is bound to cpn60 (Gray and Fersht, 1993).

To develop this point further, consider the energy profiles for folding reactions occurring both on and off cpn60 shown in Fig. 2. For the purposes of this illustration the unfolded protein(U) passes through a

transition state (N*) before forming the native conformer (N). As indicated by spontaneous folding experiments (Matouschek *et al.*, 1992a), the solvent exposure of side chains in the N* state is taken to be halfway between that in U and in N. The affinity of binding to cpn60 follows the side chain exposure (U > N* > N). Taking the C:U complex as a starting point, there are two ways of forming the free native state: either by dissociation to C + U (leftward) or by folding on the cpn60 through the bound transition state C:N* (rightward). The continuous line represents the energetics of folding of a weakly bound protein and the dotted line, those of a "sticky," tightly bound protein. Equally, one could consider the continuous line to represent the reaction at low, and the dashed line at high, cpn60 concentrations. The spontaneous folding rate is determined by the energy difference between C + U and C + N*, which represents the reaction in the absence of cpn60; i.e., the rate is determined only by the energy difference between U and N* (represented by barrier 1 in Fig. 2). In the presence of cpn60, the rate of the dissociation pathway is determined by the energy difference between C:U and C + N* (barriers 2 and 2′) since the protein must leave the cpn60 surface, to form C + U, before it can undergo the folding reaction. The rate of the "on-chaperonin" pathway is determined by the energy difference between C:U and C:N* (barriers 3 and 3′). This energetic scheme immediately illustrates two points. First, folding is necessarily slower in the presence of cpn60; i.e., barrier 1 will always be the lowest. Second, in circumstances of tight binding, on-chaperonin folding (rightward) becomes more favorable when compared to the dissociation pathway (leftward); i.e., barrier 3′ becomes lower than barrier 2′.

The effect of cofactors such as nucleotides and cpn10 is to weaken the contact between the cpn60 and its protein substrate in order to allow folding. However, the cofactor requirements are found to be different, depending on the properties of the bound protein substrate. These requirements can be described on a more quantitative basis by referring to the reaction scheme shown in Fig. 1 and making the justifiable assumption that, at equilibrium, the system contains negligible amounts of free, unfolded protein (U) or bound, folded protein (C:F). In this case, and if the concentration of cpn60 is large compared with its protein substrate (i.e., $C_o \sim [C]$) the following relationships hold:

$$\alpha_f = \frac{[F]}{[C:U] + [F]} = \frac{K_1/K_2}{K_1/K_2 + C_o},$$

where α_f is the proportion of folded protein at equilibrium, K_1 is the

equilibrium constant for folding in free solution ([F]/[U]), K_2 is the association constant for formation of the C:U complex, and C_o is the total cpn60 concentration. If an unstable protein with a K_1 of 10^2 were to fold at a cpn60 concentration of 1 μM and have an association constant (K_2) for the unfolded form of 10^{11} M^{-1}, only 0.1% of the protein molecules would reach the native state at equilibrium. If the binding of a cofactor such as a nucleotide and/or cpn10 decreased the cpn60 affinity by a factor of 10, the final concentration of native protein would still be less than 1%. If the same effect occurred with a more stable protein substrate for which K_1 had a value of 10^4, the addition of the cofactor would have a large effect on the yield at equilibrium, increasing it from 10 to 50%. From this argument it is clear that the requirements for folding will depend on the balance between the protein:cpn60 affinity (determined by K_2 and the cpn60 concentration, C_o) and the intrinsic stability of the protein (determined by K_1).

The preceding discussion applies to assisted folding reactions at equilibrium, but it is often in the kinetics of folding where the most profound effects of cpn60 are seen. This is illustrated by taking the case of a moderately stable protein with a K_1 of 10^5 [i.e., with a folding energy of 7 kcal/mol ($-RT \ln K_1$)] in which the unfolded form or forms bind to cpn60 with a K_2 of 10^8 M^{-1}. In a 1 μM solution of cpn60 about 99.9% of the molecules would attain the native state; i.e., there would be no detectable effect on the end point of folding. In fact it would require an impossibly high cpn60 concentration of 1 mM (about 860 mg/ml) to arrest folding to 50%. However, the proportion of unfolded, chaperonin-bound protein ($\alpha_{c:u}$) at any time is given by

$$\alpha_{c:u} = \frac{C:U}{C:U + U} = \frac{K_2/C_o}{1 + K_2C_o} = 0.99,$$

and the proportion of unfolded protein substrate free in solution ($\alpha_{u(free)}$) is defined by

$$\alpha_{u(free)} = \frac{U}{C:U + U} = \frac{1}{1 + K_2C_o} = 0.01.$$

These relationships show that, at any one time, only 1% of the unfolded protein is free in solution and able to fold spontaneously in these conditions. Such an interaction has a profound effect on the kinetics of folding. If the half-time for spontaneous folding is 10 s, and slow compared with

the rates of binding and dissociation from cpn60, then the observed half-time for the chaperonin-mediated process, via the dissociation route, would be about 1000 s, i.e., 100 times longer. The observed rate constant (k_{obs}) for this process is formally defined,

$$k_{obs} = k_f/(1 + K_2C_o),$$

where k_f is the rate of folding in free solution.

If the protein were able to fold on the chaperonin surface, the kinetics of the process would still be slow owing to an increase in the activation barrier (see Fig. 2). When both processes occur the observed rate of folding is approximated by

$$k_{obs} = \frac{k_bK_2C_o + k_f}{1 + K_2C_o},$$

where k_b is the rate of folding of the bound protein substrate.

As an aside to the general theme, the rates of binding and dissociation of unfolded protein substrates to cpn60—which define K_2—are clearly important factors in determining the behavior of a chaperonin-mediated folding reaction and so must color our interpretation of such events. These rates are most easily and reliably measured in protein molecules that cannot fold to the native state, thus eliminating the competing effect of the folding process. Reduced, fluorescently labeled lysozyme when rapidly mixed with cpn60 in a stopped-flow apparatus forms a complex at a rate of $4 \times 10^6\ M^{-1}s^{-1}$ with an association equilibrium constant (K_2) of $10^8\ M^{-1}$. From this measurement we can deduce a half-time of dissociation of 18 s and a half-time for formation at 1 μM cpn60 of 0.18 s (Burston and Clarke, unpublished data, 1994). The first half-time must always be a component of the observed time of refolding, and the second determines how far spontaneous folding can proceed in free solution before an encounter with cpn60. The implications of this latter process are discussed in more detail in Section III.

With regard to the more central problem of protein release, the influence of nucleotides and cpn10 on the reaction scheme shown in Fig. 1 is to decrease the value of K_2 (in the off-chaperonin pathway) or to increase K_4 (folding on-chaperonin). In measurements of the retardation of folding of bacterial LDH, the effectiveness of nucleotides in reducing

the apparent affinity of cpn60 for the enzyme is in the order ATP > AMP–PNP > ADP. The addition of cpn10, which binds in the presence of any nucleotide, reduces the affinity still further in the case of ATP and its analog but has little influence on the ADP effect (Staniforth *et al.*, 1994b).

The above arguments explain the variety of effects that cpn60, cpn10, and nucleotides have on folding reactions, both on the equilibrium level of folded product and the time course of the process. These considerations help interpret observations such as barnase needing no releasing factor (Gray and Fersht, 1993) and tryptophanase requiring only ADP (Mizobata *et al.*,1992), whereas Rubisco (Goloubinoff *et al.*, 1989), rhodanese (Martin *et al.*, 1991), and mitochondrial MDH (Miller *et al.*, 1993) need ATP and cpn10, and a host of proteins show intermediate requirements. However, these considerations, although useful in laying the crude foundations of chaperonin action, do not necessarily address the underlying mechanism by which the yield of a folding reaction is enhanced.

The crux of the physical mechanism lies in the use to which this binding energy between the cpn60 and its protein substrate is put. In a classical enzyme-catalyzed reaction, binding energy is used to stabilize the most improbable states in the progress between reactant and product, so diminishing reaction barriers, whereas in nucleotide-based energy transduction, the ability of nucleoside triphosphates and their hydrolytic products to stabilize different functional states of proteins is used to drive unfavorable processes. The cpn60 case is odd in that the folding process is energetically favorable, but to reach high efficiency for chaperonin-dependent proteins requires the hydrolysis of ATP. It is in a careful understanding of the role of the binding energy of nucleotides, protein substrate, and cpn10 with cpn60 that the basis of the molecular mechanism can be understood.

C. Yield Enhancements in Chaperonin-Assisted Folding Reactions

Thus far, we have stressed the diversity of proteins that are bound by cpn60, the range of affinities, and the variety of conditions required for the release of the protein substrate. All of these factors suggest a rather "catch-all" mechanism of action, capable of dealing with a wide spectrum of proteins. We now examine their most biologically relevant function,

that is, the ability to enhance the yield of a folding reaction. Here too, there are no transparent rules. Consider the case of three proteins: bLDH and the porcine MDHs from the mitochondrion (mMDH) and the cytosol (cMDH). These proteins are dimers with the same subunit fold and quaternary symmetry, but with completely different responses to the chaperonins. The folding time courses for the unassisted processes and those occurring in solutions of cpn60, cpn10, and ATP have all been measured (Staniforth *et al.*, 1994b) and the results are summarized in Table I. The folding of cMDH in the "full" chaperonin system is unchanged over the spontaneous process, bLDH folds slightly more slowly and the yield is increased two- to threefold, whereas mMDH in physiological ionic conditions shows an acceleration of folding combined with a threefold improvement of yield. In the case of mMDH the inclusion of cpn10 is required to enhance the yield, but for bLDH it is not. Even more oddly, if the mMDH experiment is performed in a high concentration of "structure-forming" orthophosphate ions (Jaenicke *et al.*, 1979), spontaneous folding is accelerated so that the only effect of chaperonins is to enhance the yield.

These results show that the topology of the protein dictates neither the necessity for chaperonins nor its behavior in the assisted folding reaction. As such, this conclusion is complementary to the finding that

TABLE I

Influence of Chaperonins on Folding Rates and Yields for Three Homologous Proteins[a]

					mMDH			
	cMDH		bLDH		10 m*M*[P_i]		200 m*M* [P_i]	
Chaperonin	Rate	Yield	Rate	Yield	Rate	Yield	Rate	Yield
cpn60	0.5	0.4	0.1	3.2	~0	~0	~0	~0
cpn60/ATP	0.9	0.9	0.7	2.3	0.4	0.5	0.2	0.3
cpn60/ATP/cpn10	1.0	1.0	0.9	2.3	2.8	3.0	1.1	2.0

[a] Data taken from Staniforth *et al.* (1994a,b) on the refolding rates and yields of porcine cytosolic malate dehydrogenase (cMDH), *Bacillus stearothermophilus* lactate dehydrogenase (bLDH), and porcine mitochondrial malate dehydrogenase (mMDH), the last at differing concentrations of orthophosphate. The rate referred to is for the unimolecular folding of the monomer prior to dimerization. For both rates and yields, the values given are relative to those measured for the spontaneous process (i.e., in the absence of cpn60).

proteins of quite different topologies can behave in the same way. Among this trio of dehydrogenases, one can then ask why the behavior varies. Two general molecular properties suggest themselves: the stability of the folded state and the overall hydrophobicity of the molecule. On the first criterion there is no clear relationship since the order of stability is mMDH < cMDH < bLDH, whereas the chaperonin-induced enhancement of folding is in the order mMDH > bLDH > cMDH. Moreover, the addition of citrate, an allosteric effector of mMDH, equalizes the stability of the native states of cMDH and mMDH but has no effect on assisted folding. By contrast, the hydrophobicities determined by the summed side chain properties are in the appropriate order (i.e., mMDH > bLDH > cMDH) implying, albeit on the basis of only three examples, that the requirement for chaperonins and the affinity of binding are a function of hydrophobicity. From a biological standpoint, it is notable that the two chaperonin-assisted homologs (bLDH and mMDH) fold in chaperonin-rich cellular compartments, the prokaryotic cytoplasm and the mitochondrion, whereas cMDH folds in the TCP-1 chaperonin environment found in the cytosol. It is formally possible that proteins which fold in these cpn60-rich milieux contain sequences that predispose them to interact with cpn60. In this respect, it is difficult to see how polypeptide binding sites on cpn60 can be completely indifferent to the steric properties of a ligand. However, there is no convincing evidence for specific sequence recognition.

A further, more controversial, point arises from these results. In the absence of an artifically high concentration of orthophosphate, the unimolecular rate that governs the folding of the mMDH subunit appears to be increased by chaperonins. This is at odds with the conventional view that they cannot enhance rates of folding.

These varied manifestations of cpn60 behavior have significant mechanistic consequences, especially where there are yield enhancements and either no change in folding rate or an acceleration (discussed in Sections II and VI). Such observations raise the fundamental question of the molecular processes occurring during the encounter between the folding protein and the chaperonin complex, and must be explained by any plausible model of the chaperonin mechanism. An important observation by Peralta *et al.* 1994) has shed light on the consequences of such encounters. If mitochondrial MDH is denatured by incubation in a solution of guanidinium chloride, and then diluted into a nondenaturing buffer at 36°C, the enzyme cannot fold to the native state. In these conditions it adopts a soluble, kinetically trapped state that can be refolded by the

addition of cpn60, cpn10, and ATP. This experiment emphatically demonstrates catalysis of folding from a nonnative state by chaperonins.

D. Behavior of Chaperonins *in Vivo*

Less can be learned about mechanisms of action of chaperonins through studies on their *in vivo* behavior, due to the high inherent complexity of the intracellular milieu and the impossibility of defining reaction steps and intermediates with any precision. However, any model of chaperonin action must at least be consistent with their observed *in vivo* properties, since ultimately it is their role in the cell that is of biological significance. In all cases tested, the chaperonin genes are essential for cell viability, a property that makes genetic studies dependent on the isolation of conditional mutants that can be inactivated by a change in temperature or some other manipulable parameter. Studies on such mutants in *E. coli* show that when a strain carrying a highly temperature-sensitive GroEL (cpn60) protein is shifted to the nonpermissive temperature, a large subset of proteins fail to fold correctly and either accumulate in an inactive form or are not seen at any significant level in the cell, presumably because they are turned over rapidly by proteases. Some of these proteins have been identified as key enzymes involved in metabolic pathways and protein translation (Horwich *et al.*, 1993). Similarly, global depression of the heat shock response of which the *groES* and *groEL* genes are a component leads to substantial aggregation of proteins and extreme temperature sensitivity, both of which can be suppressed by the overexpression of the *groES* and *groEL* genes alone (Gragerov *et al.*, 1992). Overexpression of both genes can also suppress temperature-sensitive mutations in a wide variety of other proteins. This is presumably because the defect in these mutant proteins is in their folding pathway rather than the final folded protein, and increased levels of chaperonins are able to increase the proportion of proteins that partition to the active state (Van Dyk *et al.*, 1989). It is worth recalling that both *groES* and *groEL* are essential for growth of *E. coli* at all temperatures (Fayet *et al.*, 1989); thus, although some GroEL-mediated reactions can proceed without GroES *in vitro*, at least one and probably several essential proteins must require both to act *in vivo*.

II. AVOIDANCE OF DEAD ENDS

It is worth introducing at this point the possible origins of yield enhancement in protein folding by considering the reaction scheme shown in Fig. 3. The pathway U → I → N is productive, but the yield of folded protein is reduced by the nonproductive paths U → M_1 and 1 → M_2 (see Fig. 3A). There are three possible ways in which the chaperonin system can enhance the yield of N, the native protein. First, it could catalyze the steps U → I and/or U → N, so increasing the flux of molecules through the productive path at the expense of the unproductive. This mechanism would require, as in conventional catalysis, the stabilization of specific transition state structures and these will be different for each protein substrate. The lack of structural similarity in chaperonin substrates and, in most reactions, the absence of rate enhancement preclude this as a general mechanism. Second, the steps leading to M_1 and M_2 can be arrested. The blocking of off-pathway reactions has been widely discussed and, in essence, requires the binding and stabilization of on-

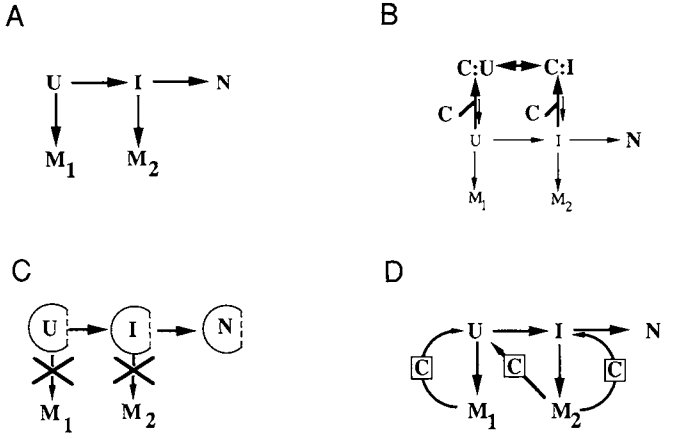


Fig. 3. Possible modes of cpn60 action. Shown in (A) is a simplified and general scheme for protein folding and misfolding. Schemes shown in (B), sequestration by binding; (C), encapsulation; and (D), recycling of misfolded forms; represent ways in which cpn60 can influence the process. For the model of sequestration by binding (B) and the model of encapsulation (C), the misfolded forms (M_1 and M_2) are aggregates since these mechanisms are based purely on the ability of cpn60 to suppress the aggregation process. A full description of these schemes is given in Section II of the text.

pathway states, i.e., U or I (shown in Fig. 3B). This is a particularly attractive proposition if M_1 and M_2 are aggregated forms of the protein, since the binding of U or I to cpn60 will prevent their interacting with each other in free solution. However, this has the disadvantage of reducing the folding rate. The preferential stabilization of U and/or I will not only reduce the flux to the misfolded forms but also that to the native state (N). This aspect is described in more quantitative detail in Section VI. An alternative means of blocking off-pathway reactions that lead to aggregated forms is to isolate the folding, monomeric species from each other to prevent their collision (shown in Fig. 3C). This is different from the binding model (Fig. 3B) in that the protein is encapsulated in an "Anfinsen cage" without interaction with the sides (Saibil *et al.*, 1992; Ellis, 1993; Martin *et al.*, 1993a). In these circumstances aggregation can be prevented without an accompanying retardation of the folding rate, although an initial interaction of the protein substrate and the chaperonin must occur in order to locate the substrate in the cage in the first place.

Third, M_1 and M_2 can be converted back to U and/or I (shown in Fig. 3D). This is an attractive mechanism since it will not inhibit the flux through the folding pathway but will increase the yield of the native state. This outcome is close to the observed behavior of a majority of chaperonin-assisted reactions, and will require the input of energy to recycle the protein substrate from misfolded states back onto the productive pathway. It is also true that, in a spontaneous folding reaction where the majority of the protein molecules form misfolded states (M_1 and M_2), and their conversion to structures on the productive folding pathway (U or I) is slow, this third mechanism could account for chaperonins accelerating the measured rate of folding (as seen by Peralta *et al.*, 1994, and Staniforth *et al.*, 1994a) without acting as a catalyst of steps in the productive pathway. The structural nature of such misfolded states is not specified by this model; they may be "scrambled" monomers with an inappropriate tertiary fold or "low order" aggregates in which the association of a small number of chains blocks their progression to the native state. In either case, if the encounter between protein and cpn60 randomly disrupts nonnative intra- or intermolecular interactions it will release the trapped state to allow folding to resume.

At its most basic, the mechanistic argument reduces to whether the transient association of the protein substrate with the chaperonin surface is used, passively, to prevent aggregation (Fig. 3B or 3C) or, actively, to induce changes in the structure of misfolded forms so facilitating progression to the native state. It is the relationship between the process

of spontaneous folding, chaperonin-substrate interactions and the use of chemical energy that will concern the remainder of this chapter.

III. SPONTANEOUS PROTEIN FOLDING

The experimental model of assisted protein folding from which almost all physical data have been derived is based on unfolding the protein substrate by use of chaotropic denaturants, pH, or heat and introducing the denatured form into native solvent conditions in the presence of chaperonins. To gain an understanding of the influence of chaperonins in such conditions, it is essential to consider what is known of the spontaneous folding process.

The least native-like states of proteins are adopted in high concentrations of guanidinium chloride or urea, where the structure is described as a random coil. In this state the polymer has no noncovalent stabilizing contacts between residues and conformation is determined only by steric hindrance. In this form the protein volume is large and the side chains and backbone rotate freely. When the denaturant is diluted out, the chain instantaneously and probably randomly contracts due to nonspecific interactions (Karplus and Shakhnovich, 1992). This contraction occurs within the dead time of the most rapid mixing experiments. During the next 1–50 ms, depending on the protein, a high proportion of the native amide hydrogen bonds are formed to generate a framework of secondary structure, but contacts between side chains are weak or absent. This leaves even the core side chains relatively free to rotate and in a "liquid" state giving rise to the term "molten globule" to describe this state or collection of states (for review, see Pittsyn, 1992). The stability of the molten globule type of intermediate over the random coil is small compared with that of the native over the intermediate (Mann and Matthews, 1993). If such intermediates are formed rapidly and are only marginally stable, then they must be able to unfold quickly and sample new conformations before undergoing the next step.

The final phase in the folding of a small monomer is slow, occurring over a period of 50 ms and longer, depending on the protein, and is characterized by a solidification of the interior of the protein brought about by the precise docking of side chains which, in addition, stabilizes the existing secondary structure and promotes the formation of the remainder. In large, multidomain proteins the evidence suggests that

preformed, native-like domains then shuffle and pair, and in oligomeric proteins well-ordered subunits must collide to bring interfaces together (Garel, 1992).

With regard to the function of chaperonins, there are several aspects of the above process that deserve consideration; these stem from the central question of what goes wrong with the process and leads to the need for these specialized helper proteins.

The first, reasonable possibility for folding errors is the aggregation of the long-lived molten globule states; these will be populated before earlier species can interact with one another. The collision of protein chains will occur at a diffusion controlled rate of 10^6–10^7 $M^{-1}s^{-1}$, and the protein concentration in most refolding experiments is between 10^{-6} and 10^{-8} M (discussed in Section IB). It will thus take between 0.1 and 100 s for the first steps in protein–protein aggregation to occur, whereas the first folding intermediates are formed on the millisecond time scale.

It is widely, but not universally, believed that the essence of chaperonin action lies in the ability to bind specifically to these molten globule states to prevent their aggregation. A consideration of the kinetics of such a mechanism is given in Section VI. For the present purposes, this model envisages a distinct and productive intermediate with a defined structure that is inherently insoluble, being prone to aggregation through the interaction of exposed hydrophobic side chains. Such an argument is not wholly convincing, since the best-characterized of these transient states has been found to have extensive intramolecular contact between such hydrophobic side chains in the core of the structure (Matouschek *et al.*, 1992b), making them unavailable for intermolecular contacts that cause aggregation. However, the intermediate states in folding that have been most extensively studied are those encountered in small, monomeric proteins that tend to fold rapidly and efficiently without detectable aggregation. In larger structures and, perhaps more particularly, those composed of multiple subunits, the tendency for folding intermediates to aggregate in a disordered manner is almost certainly greater.

Moreover, it has been demonstrated that protein–cpn60 interactions occur by hydrophobic contacts and these are most available when the protein is extensively unfolded. Given that folding intermediates in general are unstable, and in relatively rapid equilibrium with more unfolded states, it is likely that association with the chaperonin will predispose such structures to unfold to some extent (Zahn *et al.*, 1994). To say that chaperonins preferentially bind the molten globule is like saying enzymes preferentially bind the reactant state in a conventional enzymatic reaction because this is the form most often encountered.

Another source of inefficiency in folding arises from the possibility of multiple pathways. Proteins, and perhaps particularly large ones, may collapse into a series of conformers; some are productive and fold easily to the native state, some must slowly rearrange in order to proceed to the native form (Sosnick *et al.,* 1994), and others are trapped in kinetic dead ends; they are misfolded, inactive, and unable to unfold to sample more productive routes (Peralta *et al.,* 1994). This more random mechanism of folding has recently been supported by experiments performed on lysozyme, which has been shown to fold through parallel rather than sequential pathways (Radford *et al.,* 1992; Miranker *et al.,* 1993). Even in a protein as small as this, a single defined molten globule intermediate does not exist. Any two molecules may follow different paths and pass through different intermediate states. Bovine pancreatic trypsin inhibitor, the folding of which has been examined in great detail (Weissman and Kim, 1991; Creighton, 1992), is seen to pass through a number of rapidly and slowly interconverting intermediates, some of which constitute kinetic blind alleys. Inevitably some of these types of intermediates with extensive, exposed hydrophobic surfaces will aggregate to form an insoluble mass, but this is not always the case; some stable and misfolded states remain perfectly soluble (Teschner *et al.,* 1987; Peralta *et al.,* 1994; R. N. Sleigh, A. R. Clarke, and J. J. Holbrook, unpublished data, 1993), although they may still be multimeric.

IV. ENERGY TRANSDUCTION: ROLE OF ATP IN CHAPERONIN ACTIVITY

One of the most intriguing aspects of the chaperonins is their ability to capture chemical energy derived from the magnesium-dependent hydrolysis of ATP and use this to improve the yield of otherwise inefficient folding reactions. The requirement for an energy input is, at least at first sight, difficult to understand.

In muscle the free energy of ATP hydrolysis is used in the contraction of fibers, and in membrane ion pumps it is used to drive the translocation of ionic species against concentration and charge gradients. In these processes the requirement for energy is self-evident. This is not so in the case of chaperonin-assisted protein folding because the folding of protein molecules is energetically favorable. In G proteins

nucleotide turnover is used to to initiate, and to control the duration of, hormone-dependent cellular activity through the modulation of protein–protein interactions. This role in the initiation and timing of biological events is also not apparent in chaperonin activity, unless the folding protein is captured in an isolated cavity and is periodically allowed to escape. It is tempting to speculate that any energy requirement in a process that is energetically favorable is required for the type of transient reversal required to unravel mistakes. In an attempt to identify the requirement for chemical energy we now examine what is known of the ATPase cycle and the properties conferred on the chaperonins in each component state.

The diversity of chaperonin behavior, dependent on the protein substrate, has already been described and it must be stressed that some proteins require only cpn60 and ATP to achieve an improved yield of the native state on a rapid time scale (Jackson *et al.*, 1993; Mizobata *et al.*, 1992; Fisher, 1992). The coprotein cpn10, albeit necessary for the refolding of some proteins, can therefore be said to ameliorate the function of cpn60 rather than being an indispensable element of its action in all cases. As a consequence of this it is worth considering the hydrolytic reaction in the absence of cpn10.

A. Binding and Hydrolysis of ATP: Emergence of Positive and Negative Cooperativity in Chaperonin 60

Steady-state kinetic experiments, in which the velocity of the ATPase activity of cpn60 is measured as a function of ATP concentration, reveal a sigmoidal dependence characteristic of positive cooperativity that is near-saturated above 0.1 m*M* (Gray and Fersht, 1991; Yifrach and Horovitz, 1994). However, if a solution of 0.16 m*M* ATP is mixed with increasing concentrations of cpn60 and the initial reaction velocity is measured, this does not reach a maximum until 0.32 m*M* cpn60 subunits are present. In other words, only half of the subunits are catalytically active at any time (Burston *et al.*, 1995). These results imply that occupation of the first ring of cpn60 with ATP inhibits occupation of the second; i.e., there are two, distinct forms of homotropic communication in cpn60: positive cooperativity within a ring and negative cooperativity between them. The latter type of behavior is strikingly manifested by a mutant of cpn60 (GroEL R196A) that shows a strongly biphasic ATP dependence in

steady-state hydrolysis. In the first phase, at low [ATP], there is the expected increase in hydrolytic rate with increasing nucleotide concentration, whereas at higher concentrations the rate sharply declines (Yifrach and Horovitz, 1994). These observations show that ring : ring asymmetry in cpn60 is established by nucleotide interactions alone and does not require the one-sided association of cpn10 observed in structural studies.

B. Asymmetric Behavior in ADP Binding

Weak, cooperative binding of ADP to cpn60 alone can be followed by the response of an undisruptive pyrenylmaleimide probe, the fluorescence of which is sensitive to the conformation of the protein (Jackson *et al.*, 1993). This binding process is half-saturated at an ADP concentration of 2.3 m*M* and induces a 12–15% enhancement in pyrene fluorescence. However, there is an initial, high-affinity binding process occurring in the micromolar range that can be detected either by a very small (0.20–0.25%) fluorescence decrease or by conventional Scatchard analysis of free and bound ADP using ultrafiltration (Burston *et al.*, 1995). This tight binding process accounts for only seven molecules of ADP per 14-mer. Thus, both ATP and ADP occupy the first ring with high affinity and the second with low.

C. Rates of Individual Steps in ATPase Cycle of Chaperonin 60

The rate of the steady-state cycle of ATP binding, hydrolysis, and product release in the absence of cpn10 is 0.04 mol of ATP consumed per mole cpn60 subunit per second. Given that only half of the cpn60 subunits can bind and hydrolyze ATP at any time, the rate-limiting step occurs at 0.08 s^{-1}. A true single-turnover reaction in this system is achieved by mixing seven equivalents of ATP with one cpn60 oligomer (14-mer). When such a reaction is performed under saturating conditions [0.2 m*M* ATP 0.03 m*M* $(cpn60)_{14}$] and the total (free and bound) orthophosphate product is measured as a function of time, the progress of the reaction is described by a single exponential process with a rate constant of 0.08 s^{-1}. These results show that the release of products does

not limit the rate of the steady-state reaction, otherwise single-turnover hydrolysis, which requires no product release, would be quicker.

By use of the pyrene fluorescence probe and stopped-flow mixing methods, it has been established that ATP binding is highly cooperative and occurs in two stages. ATP first forms a weak collision complex with a dissociation constant of 4 m*M*. This binding event then drives a rearrangement of the oligomer at a rate of 180 s^{-1} (Jackson *et al.*, 1993). We can therefore describe the ATPase cycle of a single ring as: (1) weak ATP binding; (2) fast cooperative rearrangement to produce a conformation in which ATP is bound tightly; (3) slow, rate-limiting hydrolysis; (4) product releae (see Fig. 4).

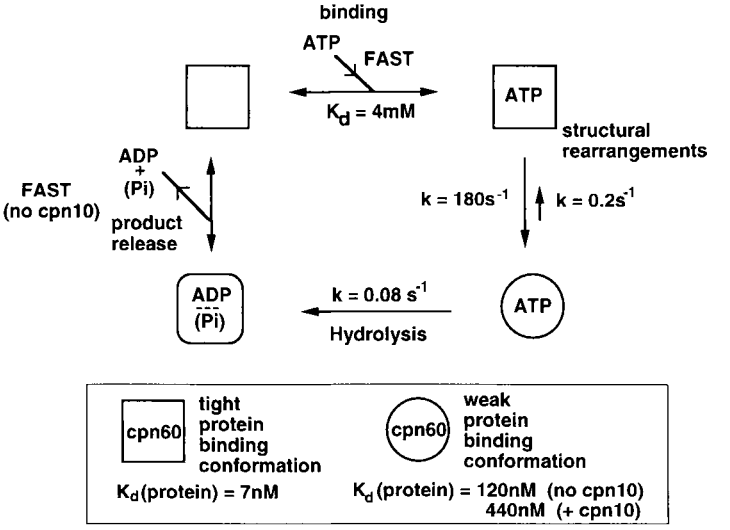


Fig. 4. The ATPase cycle of cpn60. The diagram summarizes the hydrolytic cycle of cpn60 and the coupling of the energy of ATP hydrolysis to structural rearrangements in the protein (Jackson *et al.*, 1993). The steps are described in detail in Section IV of the text. K_d values are equilibrium dissociation constants, $K_{0.5}$ is the concentration of ADP needed for half saturation of cpn60, and k values represent rate constants. Note that hydrolysis is the slowest step, making the weak protein binding conformer with associated ATP the steady-state (predominant) complex. The association of cpn10 dramatically reduces the rate of dissociation of ADP (Pi) from cpn60 subunits and so enhances its binding affinity. Protein binding affinities ($K_{d(protein)}$) are for the unfolded form of bacterial LDH (Staniforth *et al.*, 1994b).

D. Conformational Rearrangement of Chaperonin 60 and Positive Cooperativity within Ring

As described in Section I,B, the binding of ATP and analogs promotes the dissociation of bound protein substrates. Apo-cpn60 has the highest affinity for nonnative proteins; for the purposes of this discussion we call this conformation the P state, or protein-binding state. The rearrangement induced by the interaction with ATP weakens protein binding and produces a conformation, which we denote the A state, or ATP-bound state. Note that the rate of rearrangement is several thousand times faster than the rate of hydrolysis so that the A state is rapidly populated. The structural changes that occur in the transition from the P to the A state are described in detail in Chapter 9.

If the association of ATP is examined at equilibrium the binding curve is sigmoidal, showing positive cooperativity with a Hill constant of between 3 and 4. The binding of ATP is very tight with half-saturation occurring at a concentration of 10 μM. This tight binding is in line with existence of the two conformers, with the P state binding ATP weakly, and the A state tightly. In a classical Monod–Wyman–Changeux (MWC) analysis, the P and A states constitute the T and R forms of the protein, respectively. The stopped-flow experiment gives the weak binding constant to the T form ($K_{d(ATP)}$, 4 mM) and, with a cooperative unit of seven ATP sites, the binding constant to the R form ($K_{d(ATP)}$) is 5 μM, and L, the ratio of T to R forms in the unliganded state, is 860. With a cooperative unit of 14 ATP sites, the binding curve gives a poor fit to the MWC model. The cooperative mechanism is summarized in Fig. 5 [data taken from Jackson *et al.* (1993)]. The binding affinity for unfolded lactate dehydrogenase in the P state is very high, the dissociation constant (K_d) being 7 nM (Staniforth *et al.,* 1994b). When converted to the A state by association with ATP, the binding affinity is reduced by at least 17-fold (K_d, 120 nM), and on hydrolysis of the nucleotide to give the ADP (P_i) state the affinity is increased (K_d, 30 nM).

These findings are important for three reasons. First, they show that ATP binding energy is used to alter the structural properties of cpn60 to allow it to release a bound protein. Hydrolysis and product dissociation serve to return the chaperonin to its original, tight protein-binding state. This conclusion is also borne out by the observation that AMP–PNP (an unhydrolyzable analog of ATP) is able to displace proteins more effectively than is ADP or a combination of ADP and P_i (Jackson *et*

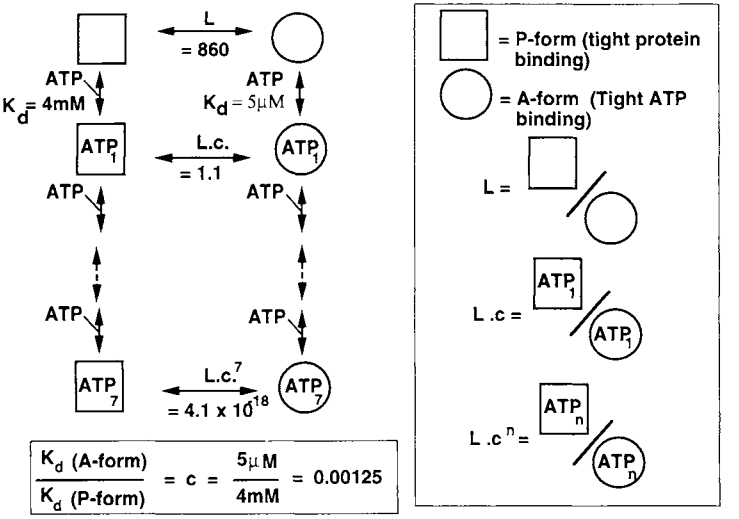


Fig. 5. Cooperativity in cpn60. As shown in Fig. 4, cpn60 exists in at least two conformational states; here they are designated the P- and A-forms. ATP binds weakly to the P-form and tightly to the A-form, thus driving the conformational equilibrium toward the latter. Cooperativity is represented as an MWC model, with the T-state being the P-form and the R-state being the A-form. The relevant MWC parameters are given on the diagram and the model is discussed in detail in Section IV. In the presence of cpn10 the cooperativity of nucleotide binding is increased. The values are calculated from data presented in Jackson *et al.* (1993).

al., 1993). Second, the rearrangement is highly cooperative, so that all subunits contributing to the toroid change structure in a concerted fashion. Third, the equilibrium ratio of the P to the A state in the absence of ATP is only 860 (the L value in an MWC model), whereas the ratio of the binding affinity of ATP to these respective states is 7.5×10^{-4} (the c value). These values mean that the binding of a single ATP molecule to the 7-mer toroid of cpn60 is sufficient to tip the structural equilibrium so that the A and P states are energetically balanced (the P:A ratio then becomes $Lc = 1.1$). Binding seven ATP molecules to the cooperative unit gives a ratio of 4.1×10^{-18} (Lc^7), i.e., massively in favor of the A state (see Fig. 5). If this ratio is compared to the binding of molecular oxygen to hemoglobin, where $c = 0.01$ and $L = 300{,}000$, three oxygen molecules must bind before the tetramer is tipped into the R

state, and even when fully saturated the T:R ratio is 3×10^{-3}. This means that in the case of the chaperonin there is a large amount of "spare" energy in the conversion of the P to the A state when saturated with ATP: presumably this is used to eject bound protein substrates that associate preferentially with the P state and therefore increase the value of *L*. The effectiveness of nucleotides in displacing the protein substrate is in the order ATP > AMP–PNP > ADP (see Section I,B). This ability reflects the degree to which these nucleotides promote a cooperative transition to the A state, which has only a weak affinity for unfolded proteins.

E. Nucleotide Asymmetry in Chaperonin 60: Implications for Hydrolytic Cycle

As described above, the binding of both ATP and ADP shows negative cooperativity with respect to the occupation of the two rings; i.e., the first ring is filled with high affinity, but this converts the second ring to a low-affinity state. Such behavior is unusual in biology and this prompts the somewhat teleological response that it is an indispensable element of the mechanism. The most compelling interpretation of ATP/ADP asymmetry is that it provides a basis for a reciprocating mechanism; i.e., the instability of complexes containing either 14 ATPs or 14 ADPs is used as a device to switch the rings, in an alternating cycle, between the P state and the A state. Thus, at any time, one ring will be occupied by ATP and the other by ADP (P_i). Hydrolysis of ATP on the first ring will then lead to displacement of ADP(P_i) on the second to allow its occupation with ATP. This type of mechanism is described in Fig. 6.

F. Behavior of Chaperonin 10 in ATPase Cycle

Although some protein substrates do not need the assistance of cpn10 for folding yields to be enhanced (Mizobata *et al.*, 1992; Fisher, 1992; Jackson *et al.*, 1993), others do. In some cases the cpn10 is essential to achieve any renaturation of the protein substrate (Goloubinoff *et al.*, 1989) and in others it has been clearly shown that the requirement

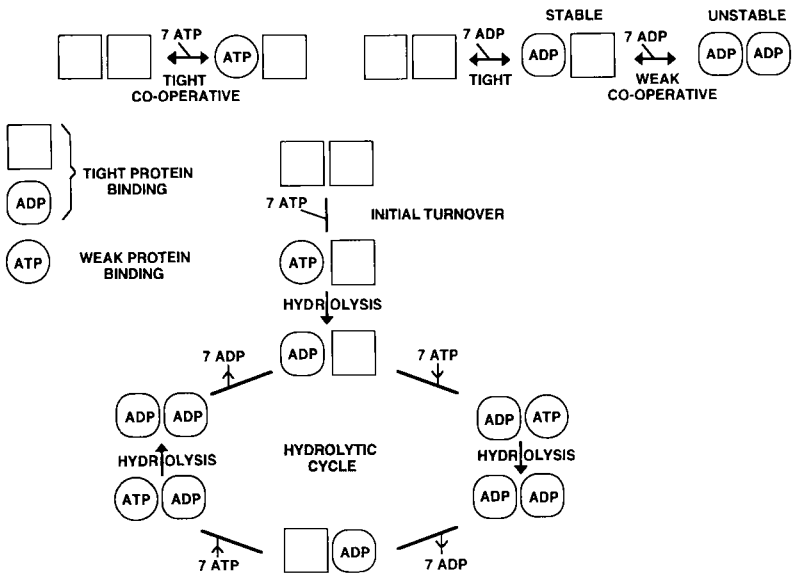


Fig. 6. Asymmetry in the binding and hydrolysis of nucleotides by cpn60. Even in the absence of cpn10, both ATP (top, left) and ADP (top, right) bind much more tightly when occupying the first ring than the second. In the diagram each ring is represented as a single unit that can adopt one of the three conformations (middle, left). A proposed catalytic cycle is shown (center) in which the initial turnover of ATP on half of the sites produces tightly bound ADP on a single ring. This forces further ATP to bind and hydrolyze on the opposite ring to give an unstable, symmetrical $ADP_7 : ADP_7$ complex. This instability then drives dissociation of ADP from one ring, so producing a more stable asymmetric form with seven free sites for ATP. The second half of the cycle then commences. We have assumed that dissociation of ADP occurs from the ring opposite to that on which ATP has just been hydrolyzed, so that the cycle alternates from side to side. The most highly populated states in the cycle are $ADP_7 : ATP_7$ and $ATP_7 : ADP_7$, meaning that each ring spends half its time in a tight protein-binding conformation and half in the weak.

for cpn10 depends as much on the solvent conditions as on the species of protein substrate (Schmidt *et al.*, 1994a). The precise role of cpn10 remains one of the more enigmatic aspects of the mechanism, but the accepted properties of the extensively studied cpn10 from *E. coli* (GroES) are as follows. It is a heptameric toroid of identical subunits which, in the presence of ATP, AMP–PNP, or ADP, forms a high-affinity complex with cpn60. In most conditions this complex is asym-

metric with a single cpn10 ring associated with one end of the cpn60 double toroid. However, with ATP, at a slightly elevated pH and at high concentrations of magnesium ions, a symmetrical complex is formed with two cpn10 rings (Schmidt *et al.*, 1994b). From such observations it is tempting to suggest that this symmetrical structure may be an intermediate in an ATPase reaction cycle that involves the association and dissociation of cpn10 on alternating sides of cpn60. Support for a cyclic displacement mechanism comes from the observation that the normally very stable and asymmetric cpn60:ADP:cpn10 complex is induced to dissociate by the addition of ATP. This effect was noted in experiments that measured the exchange of radioactively labeled components (Lorimer and Viitanen, 1993), and further demonstrated by fluorescence measurements in which the migration of cpn10 from a labeled cpn60:ADP:cpn10 complex to unlabeled cpn60 was seen to occur rapidly on addition of ATP (Burston *et al.*, 1995) but not unhydrolyzable analogs.

These complications aside, a known effect of cpn10 in chaperonin complexes is to stabilize the binding of ADP (Jackson *et al.*, 1993) by reducing its rate of dissociation from the cpn60:ADP:cpn10 complex (Todd *et al.*, 1993). If this complex is challenged with alkaline phosphatase to destroy any free ADP, the rate of decay of this species is 0.004 s^{-1}. During steady-state hydrolysis in a cpn60/ATP mixture, the addition of cpn10 reduces the rate of turnover from about 0.08 s^{-1} per active subunit to about 0.04 s^{-1} (taking into account that only half the sites can undergo hydrolysis at a time). If a mixture of fluorescently labeled cpn60, cpn10, and ATP is allowed to reach the steady state and the reaction is "flooded" with a 10-fold excess of unlabeled cpn60, the rate of migration of cpn10 to the unlabeled form is the same as the rate of ATP hydrolysis. These measurements demonstrate two characteristics of cpn10 in this complicated catalytic cycle. First, it undergoes obligatory dissociation from the complex on each round of hydrolysis and, second, this step appears to limit the turnover rate of ATP. As mentioned above, stable ATP analogs do not displace cpn10 from a cpn60:ADP:cpn10 complex, showing that it is the hydrolysis of ATP that triggers the displacement of the coprotein. In addition to these observations, it is noteworthy that exchange of cpn10 between cpn60 molecules can occur at substoichiometric ratios (e.g., 1 cpn10:4 cpn60), implying that the formation of a symmetrical complex with two cpn10 oligomers per cpn60 is not an obligatory intermediate in the exchange process (Burston *et al.*, 1995).

G. Effect of Chaperonin 10 on Cooperativity and Displacement of Protein Substrates

The cooperativity of ATP and AMP–PNP binding to their sites on cpn60 is increased in the presence of cpn10. The coprotein therefore enhances the "sharpness" of the transition between the P state and the A state (Jackson *et al.*, 1993). This increased cooperativity is also manifested in the kinetics of ATP hydrolysis (Gray and Fersht, 1991; Todd *et al.*, 1994; Yifrach and Horovitz, 1994). Chaperonin 10, in addition, reduces the binding affinity of cpn60 for its protein substrates (Martin *et al.*, 1993a; Staniforth *et al.*, 1994b). Such a combination of effects would produce a more concerted switch between the tight and weak protein-binding forms of cpn60 in the ATPase cycle and facilitate the release of the protein substrate. It has been proposed that this "all-or-nothing" switch, coordinated by cpn10, is important in expelling the substrate from all binding surfaces in a single step (Jackson *et al.*, 1993; Todd *et al.*, 1994).

The underlying mechanism by which cpn10 achieves these effects is not entirely understood, but some insight into the process has been acquired through recent structural (Chen *et al.*, 1994; Braig *et al.*, 1994) and site-specific mutagenic studies (Fenton *et al.*, 1994). The results of this work show that protein substrates and cpn10 bind to the same surface on cpn60; i.e., the site responsible for the binding of hydrophobic chains on protein substrates is the same as that which binds the hydrophobic and mobile loop on cpn10 identified by NMR spectroscopy (Landry *et al.*, 1993).

One possibility, therefore, is that cpn10 displaces the protein from its site on cpn60 by direct competition. In some models of cpn10 action, direct displacement of the protein substrate by the coprotein is thought to drive its translocation through the center of the cpn60 complex to interact with the other toroid, or simply to release the protein into an internal cage where folding can proceed (Martin *et al.*, 1993a). Despite the attractive simplicity of this mechanism, direct visualization of a complex containing cpn60, cpn10, and mMDH, made in the presence of either ATP or ADP and rapidly vitrified, shows the protein substrate and the cpn10 coprotein bound at opposite ends of cpn60. It appears that the action of cpn10 in displacing bound proteins is communicated from one ring to the other. This mechanism is consistent with much of the unorthodox negative cooperativity shown in ring-to-ring communica-

tions in cpn60. This behavior is exemplified by the asymmetric binding of ATP, ADP, cpn10, and the mMDH substrate itself (Chen *et al.*, 1994). It follows from these properties that cpn10 and protein substrates would oppose each other's association with cpn60 from opposite ends; i.e., the binding of cpn10 to one ring would be expected to promote release of protein substrate from the other. In line with this idea it has been ingeniously shown in trapping experiments (Weissman *et al.*, 1994) that proteins are ejected from the cpn60 surface and rebind many times during folding. This may mirror, in a reciprocal process, the obligatory dissociation and reassociation of cpn10 that is driven by the hydrolytic cycle. From this point of view it is interesting to note that, from a combination of both spectroscopic and cryoelectron microscopy data, it can be shown that in active refolding conditions, where unfolded mMDH is being "chaperoned" to the native state—but well before the stage of commitment has been reached—the predominant form of cpn60 has a high occupancy of cpn10 (as expected from a system in which the displacement of cpn10 is rate-limiting) but a low occupancy of the protein substrate on its binding surfaces (A. R. Clarke and H. R. Saibil, unpublished data, 1994). This means that in the steady state of an assisted folding process much of the mMDH is free in solution.

If it is true that cpn10 and protein substrates compete in their binding to cpn60, albeit from opposite ends via negative cooperativity, then it ought to be possible to find conditions in which the presence of a bound protein substrate interferes with the association of cpn10. This expectation is fulfilled by the observation that unfolded proteins can, at least partially, displace cpn10 from an otherwise stable cpn60:ADP:cpn10 complex (Martin *et al.*, 1993a; S. G. Burston and A. R. Clarke, unpublished data, 1994). Additionally, if a stable complex between cpn60 and denatured mMDH is initially formed and then challenged with ADP and cpn10, the association of the coprotein is very slow and incomplete (S. G. Burston and A. R. Clarke, unpublished data, 1994). These results with mMDH gained from spectroscopic experiments are reinforced by cryoelectron microscopy, which shows the association of mMDH at the opposite end of the cpn60 oligomer to the one at which cpn10 is associated. The contact between cpn60 and cpn10 is only tenuous in these "three-protein" complexes and the conformation of cpn60 is very "open" and distorted. In other molecules in the field the cpn10 has dissociated and only mMDH is bound to cpn60. Conversely, in the presence of ATP, cpn10 is effective in displacing mMDH from a preformed cpn60:mMDH complex.

These experiments imply that during a reciprocating catalytic cycle the ATP state favors association of cpn10 and dissociation of the protein substrate, whereas hydrolysis and the consequent formation of the ADP(P_i) state favors association of the substrate and displacement of cpn10. This hypothesis is strengthened by the data of Weissman *et al.* (1994), which show that protein substrates (rhodanese and ornithine transcarbamoylase) are discharged from the cpn60:cpn10 complex at a rate ($\sim 0.03\ s^{-1}$) similar to that of the hydrolytic cycle. A scheme summarizing these events is shown in Fig. 7.

As this discussion shows, the question of the molecular role of cpn10 is a complicated one and the properties described are only beginning to be unified to give a full and coherent explanation. Chaperonin 10 has an effect on the cooperativity of cpn60, on the symmetry of chaperonin complexes, on the affinity of nucleotide interactions, and on facilitating the displacement of tightly bound proteins. These are challenging issues that are attracting much deserved attention.

V. CONFORMATION OF BOUND PROTEIN SUBSTRATES

A major goal of physical experiments on chaperonins is to understand what happens to the conformational properties of folding proteins during their encounters with the binding surface of cpn60. At the heart of this goal lie the linked questions of how protein and nucleotide binding energies are used, and which structural state of the protein is preferentially bound, and thus stabilized, by cpn60. Such insights would be akin to the demonstration that orthodox enzymes accelerate reactions by forming preferential contacts with the transition states of chemical reactions, although, as explained earlier, the analogy between enzyme catalysis and chaperonin-assisted folding may not be strong.

Techniques for following the steps in spontaneous folding reactions are well developed. These techniques include continuous, time-resolved measurements of the acquisition of secondary structure by circular dichroism, the development of stable hydrogen bonds by amide protection NMR, the rate of formation and decay of intermediates by monitoring the transient association of fluorescent hydrophobic molecules, and the effect of mutations on the kinetics and energetics of the process (for reviews of these subjects, see Creighton, 1993). All of these techniques

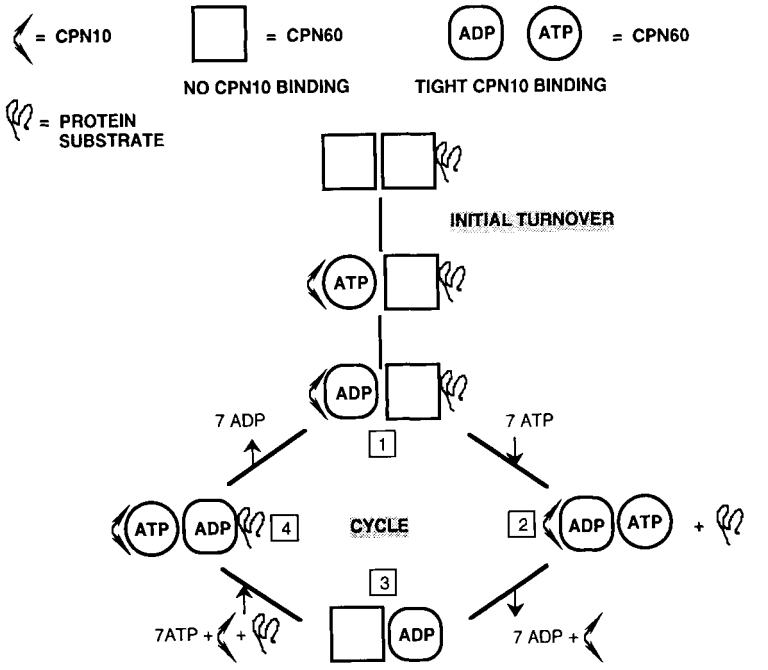


Fig. 7. Dynamics of protein substrate and cpn10 binding in the hydrolytic cycle. In the initial turnover of a tight cpn60:protein binary complex, ATP and cpn10 bind to the ring opposite to the protein substrate (Chen *et al.*, 1994) and the nucleotide is hydrolyzed to establish complex 1 in the cycle. Binding of ATP to the protein-bound ring releases the substrate from complex 2. Hydrolysis of ATP forces cpn10 and ADP to dissociate (Todd *et al.* 1994; Burston *et al.*, 1995) from the opposite ring, producing complex 3. The binding of ATP to the vacant sites followed by rapid association of cpn10 [the coprotein binds only slowly to cpn60:ADP (Jackson *et al.*, 1993)] leaves the open, ADP-ring in complex 4 free to bind the protein substrate tightly once more. Hydrolysis and ADP dissociation from the opposite ring then reset the cycle. The net result is a repetitive cycle in which both cpn10 (Todd *et al.*, 1994; Burston *et al.*, 1995) and the protein substrate (Weissman *et al.*, 1994) are forced to associate and dissociate. Complexes 2 and 4 are long-lived; complexes 1 and 3 are transitory.

become impossible, difficult, or merely awkward to interpret when chaperonins are introduced.

The early assertion that cpn60 specifically binds the molten globule intermediate was drawn from the observation that the tryptophan fluorescence properties of bound rhodanese when arrested in its folding are

intermediate between those of the native and the random coil states. In addition, cpn60-associated rhodanese is able to bind the fluorescent dye, anilino-naphthalene-sulphonate (ANS), so increasing its emission intensity (Martin *et al.*, 1991). This latter property is considered to be a diagnostic for the formation of the molten globule in free solution. The problem with such optical probes is the notorious uncertainty in interpreting the signal. The fluorescence of tryptophan may be enhanced or shifted in wavelength by its interaction with hydrophobic sites on cpn60, and ANS binding sites may be formed by pockets in the cpn60:protein complex rather than those exclusively within the bound substrate.

Contradictory evidence was initially provided by experiments on LDH in which the dimeric native state was unfolded in denaturant, first to the "native-like" monomer, then to a part-folded intermediate having the accepted properties of a molten globule, then to the fully unfolded, random coil state. Only when folding was initiated from the completely unfolded state did cpn60 interact tightly with LDH and so block the folding reaction. When folding was initiated from a productive molten globule, cpn60 had little affinity (Badcoe *et al.*, 1991). This result can be rationalized if one takes into account the observation that unfolding only as far as the molten globule before renaturation gives a high yield of active enzyme, whereas the yield of refolding from the fully denatured state is low. In the first case, a productive molten globule state refolds with little interaction with cpn60. In the second, the collapse of the random chain leads to states that are less well organized, fold less productively, and bind tightly to cpn60 with dissociation constants in the nanomolar range (Badcoe *et al.*, 1991; Staniforth *et al.*, 1994a). Experiments using rapid mixing techniques on a mutant form of barnase show that cpn60 has a higher affinity for states more unfolded than the major, long-lived intermediate (Gray and Fersht, 1993); this latter state has little tendency to bind to cpn60.

Work with α-lactalbumin has produced somewhat conflicting conclusions (Hayer-Hartl *et al.*, 1994; Okazaki *et al.*, 1994). The protein can be trapped in various states of "foldedness" by the removal of the calcium ions or by breaking or scrambling the native disulfide bridges. The former condition produces a structure that is considered to be an orthodox molten globule, but this does not bind to cpn60. However, less organized states with scrambled disulfide bridges that still maintain some secondary structure do bind, as has been confirmed by studies utilizing electrospray ionization mass spectroscopy (Robinson *et al.*, 1994). If the process of denaturation is taken to its extreme by carboxymethylating all the cys-

teine residues in the protein, binding to cpn60 is abolished. Finally, NMR amide protection experiments show that cpn60 in its tight protein-binding form stabilizes cyclophilin in a conformation that has relatively freely exchangeable amide protons (Zahn *et al.*, 1994). This observation signifies a weakening of backbone hydrogen bonds and is interpreted to mean that bound cyclophilin adopts a highly unfolded state.

One of the problems with determining the state of protein substrates prior to their interaction with cpn60, and the change in state of the released material, arises from the complicated dynamics of assisted folding. This problem is illustrated by the fact that unfolded protein substrates have to be diluted into a renaturing buffer containing the chaperonin components in order to initiate folding. In these circumstances it is probable that folding steps or aggregation processes will occur before interaction with cpn60. In addition, kinetic studies are limited to conditions in which cpn60 is in excess of the concentration of protein substrate so that most of the substrate material is not irretrievably lost. In short, the condition of the substrate for assisted folding reactions, between its encounters with the chaperoin, is not static. These difficulties are partially alleviated by working with a substrate that can be maintained in an unfolded state in a solvent that does not denature chaperonins. The most promising substrate from this point of view is mMDH (Peralta *et al.*, 1994), which can be trapped in an unfolded state by raising the temperature of the refolding reaction. It can then be released to commence folding by the addition of chaperonins and ATP.

If the mMDH refolding reaction is studied at a poised temperature (30°C) where some spontaneous folding occurs, but assisted refolding is highly efficient, the kinetic properties of the two processes can be compared. Results from these experiments show that spontaneous folding proceeds through the slow folding of individual subunits followed by their association to produce active dimers. The formation of active mMDH is made inefficient by a misfolding process that occurs in two phases. The first is a rapid, collision-controlled formation of small aggregates that is only slowly reversible; the second is an irreversible phase of aggregation. In the presence of cpn60, cpn10, and ATP, the first stage of the misfolding process is rapidly reversed. This has two effects. First, it increases (by a factor of 3–4) the observed rate constant for the productive folding of individual subunits by generating a higher concentration of viable monomers able to undergo this step (the scheme is summarized in Fig. 8). Second, it increases (by a factor of 7) the yield of active mMDH. Experiments in which the final folding yield is measured as a function

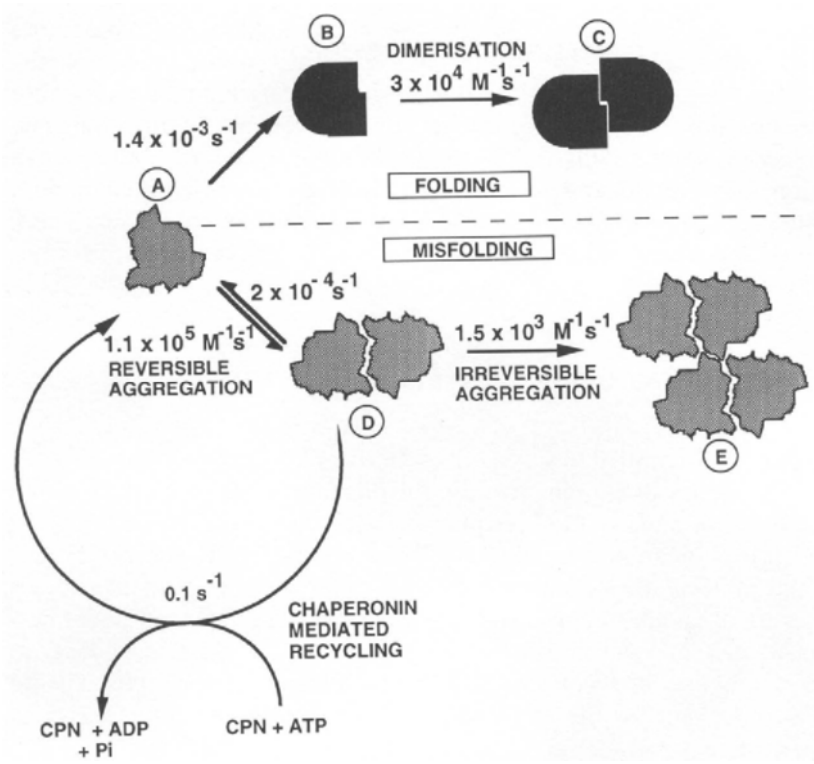


Fig. 8. Chaperonin-assisted folding of mMDH. The productive folding of mMDH follows the pathway A → B → C but is subject to competition from a misfolding pathway (A → D → E), which consists of wrong associations between chains. Early aggregates (D) can only slowly dissociate in a spontaneous folding reaction but chaperonins induce rapid dissociation to maintain a pool of folding-competent monomers (A). The slowly formed, larger aggregates (E) cannot be salvaged by the action of chaperonins. The rate constants for each step have been defined and the process is described further in Sections V and VI of the text.

of the number of turnovers of ATP allowed in the assisted reaction show that mMDH is not committed to progress to the native state until the monomers have folded to a state competent for correct dimerization.

The fact that the second, irreversible phase of aggregation is slow means that mMDH can be "rescued" from its early aggregated state for several minutes after spontaneous folding is initiated. A second

consequence is that cpn60 is effective at markedly substoichiometric ratios with respect to the substrate (1 cpn60:20 mMDH), there being sufficient time for a single cpn60 to undergo many binding and release cycles before the mMDH is lost to slow, irreversible aggregation. The conclusion of this study is that the chaperonin is able to bind to small aggregate structures and promote their dissociation to provide competent monomers for the folding reaction (A. R. Clarke, and H. R. Saibil, unpublished data, 1994).

VI. GENERAL MODELS OF CHAPERONIN ACTION

We have attempted to summarize a body of knowledge on the kinetics and energetics of protein folding, on the functional properties of the chaperonins, and on the interplay between the two. The information available is by no means complete, but it is sufficient to suggest models of the molecular mechanism by which this energy-dependent process leads to an enhancement of folding efficiency. The types of model currently envisaged fall into three categories. None requires specificity in the interaction between cpn60 and the folding protein; thus, all are able to account for the extraordinary diversity of substrates on which chaperonins act.

A. Binding of Aggregation-Prone Intermediates

One model proposes that the role of chaperonins is to sequester sticky protein folding intermediates from the bulk phase, thereby preventing their irreversible aggregation. The evidence that chaperonins prevent aggregation is undeniable (for examples, see Buchner *et al.*, 1991; Martin *et al.*, 1991), but an inevitable consequence of this mechanism is that, although it will increase the yield, the interaction of chaperonins with folding intermediates will slow down the rate of formation of the native state in proportion to the decrease in the free concentration of these intermediates (Jackson *et al.*, 1993); i.e., both the aggregation process and the folding process is inhibited. This conclusion is illustrated by the following scheme:

$$\begin{array}{c} \text{I ag.} \\ \text{fast} \uparrow \text{slow} \\ \text{U} \rightarrow \text{I} \rightarrow \text{N}, \\ \text{C} \downarrow\uparrow \\ \text{C:I} \end{array}$$

where U is unfolded protein; I is intermediate; N is folded protein; C is chaperonin; and I ag. is the aggregated form of I.

In this system the rate of folding (V_f) is defined by

$$V_f = k[\text{I}],$$

and the rate of formation of the aggregate (V_a) is given by

$$V_a = k_a[\text{I}]^n.$$

The concentration of the intermediate is raised to the power n because aggregation is a multimolecular process. The value of n is typically 2.5 (Zettlmeissl *et al.*, 1979). When the processes of folding and aggregation compete, the pathways are partitioned according to their relative velocities,

$$p = V_f/V_a = (k_f[\text{I}])/(k_a[\text{I}]^{2.5}) = (k_f/k_a)(1/[\text{I}]^{1.5}),$$

where p is the partition ratio describing the relative initial rates of the two pathways, thus reflecting the yield of native protein. This relationship shows that a decrease in the concentration of the free intermediate [I] by its binding to the chaperonin will increase the value of p and so favor folding over aggregation. However, the velocity of the folding reaction (V_f) is directly proportional to the concentration of free intermediate, so this will be reduced. In this example a 10-fold increase in the value of p will result in a folding rate 4.6 times slower.

If this model is correct, increases in the yield of native material in the presence of chaperonins will inevitably be accompanied by reductions in the rate of folding. In many cases this reciprocal relationship between folding efficency and folding rate is not observed. This fact is particularly well demonstrated by Buchner *et al.* (1991), who show that with citrate synthase, the kinetics of formation of the active enzyme are unaffected by the presence of chaperonins and Mg-ATP in experiments where the improvement of yield is 10-fold. This observation has also been made

in experiments with mMDH, where, under some conditions, the yield is improved threefold while the kinetics are unchanged (Miller *et al.*, 1993) and, more strikingly, in other conditions the observed folding rates are increased (Peralta *et al.*, 1994; Staniforth *et al.*, 1994a). Therefore, the action of the chaperonins cannot solely be due to passive binding of unfolded protein states.

B. Noninteractive Encapsulation: Anfinsen Cage

A variation on this model, in which aggregation is suppressed, not by *binding* sticky intermediates, but by *caging* them in the central cavity (Saibil *et al.*, 1993; Martin *et al.*, 1993a; Ellis, 1993, 1994), is more promising. For the reasons given above, folding would have to occur without significant interaction with the internal walls of the chaperonin toroids to avoid slowing the reaction down. This model must, nonetheless, include an initial binding of a nonnative protein substrate through hydrophobic interactions in order to locate them in the open cpn60 cavity. The association of cpn10 could then trap the protein in a closed cavity and simultaneously displace it from the walls of cpn60 to allow folding to proceed. The hydrolytic cycle must then include a step in which cpn10 is displaced from the complex to allow the folded protein to escape and another substrate to enter.

As yet there is little direct evidence for this model, but the architecture of cpn60 makes it seductive. It is also unclear whether the cavity is sufficiently voluminous to encapsulate the larger proteins that undergo chaperonin-assisted folding. The size of the cpn60 cavity, in terms of the molecular weight of a protein it could accommodate, then becomes a critical mechanistic feature. The best estimates of the shape and dimensions of the cavity in the 7-subunit toroid (Chen *et al.*, 1994; Braig *et al.*, 1994) are that it is a roughly cylindrical space with a height of 4.7 nm and a diameter of about 5 nm. The electron density corresponding to the last 24 residues is not visible in the crystal structure, but is clearly seen in cryoelectron microscopy images (Chen *et al.*, 1994) to form an equatorial partition dividing the central cavity into two spaces, one in each toroid. Condensed folding intermediates, described as molten globules, have radii of gyration 15% larger than the native structure (calculated as an average of the properties summarized in Ptitsyn, 1992). Taking this increase into account the cavity could accommodate an intermediate

that had native dimensions of 4.10-nm height and 4.35-nm diameter. These values set an upper limit for a protein substrate capable of being completely engulfed in the cpn60 cavity. Taking the density of an average native protein to be 0.0023 nm^3 Da^{-1} (derived from crystallographic analysis) gives a molecular weight of 26,500 for a perfectly fitting, dehydrated, and cylindrical folding intermediate. A spherical folding intermediate with a molecular diameter of 5 nm (native state 4.35 nm) would have a molecular weight of 19,000. The existence of tight bound hydration shells would reduce these values still further. It should be noted that the mass of 90 kDa quoted widely in the literature derives from a calculation that assumes the bound substrate to be native, perfectly cylindrical, and dehydrated, an unlikely trio of properties.

The possibility of folding, or at least critical steps in the process, occurring in a chaperonin cage looks more promising in cpn60:cpn10 complexes. Here the cavity is widened owing to a "petal-like" opening of the apical domains so that cpn10 can be neatly accommodated on the protein-binding surfaces of cpn60 (Chen *et al.*, 1994). The size of the cavity in the cpn10-capped structure is larger and could accommodate a globular folding intermediate with a diameter of 6.5 nm (native state 5.65 nm) and a molecular weight of 41,000 (see Chapter 9).

If the size limit for encapsulation were 41,000, this value would account for many cases of cpn60:cpn10-dependent protein folding. Recent work in which three physiologically relevant proteins are taken as case studies shows convincingly that cpn10 is required in solvent conditions where aggregation/misfolding predominates in the spontaneous folding process (Schmidt *et al.*, 1994c). The authors argue that cpn10 allows discharge of the protein in a state committed to folding. Prior to this study, experiments by Hartl and colleagues (Martin *et al.*, 1991) showed that casein does not interfere with chaperonin-assisted folding in the presence of cpn10. Both of these observations may be considered consistent with folding occurring in a protected, encapsulated environment when cpn10 is present.

C. Dynamic Properties of Chaperonin 60:Protein Interactions

There are several reasons why this "cpn10 encapsulation" model does not provide a consistent explanation for the behavioral properties of the chaperonins. First, for some proteins cpn10 is not required to achieve

large enhancements of the folding yield; in the refolding of these substrates it merely serves to promote a more rapid release of active protein (Fisher, 1992; Jackson *et al.*, 1993; Mizobata *et al.*, 1992). Second, some substrates appear too large to be caged during the folding process—for example, citrate synthase, a well-studied model substrate, has a molecular weight of 50,000. Third, if the folding process occurs within a chaperonin cage, then it is difficult to see how radioactively labeled Rubisco can be exchanged for unlabeled at times much shorter than those required for the unimolecular folding steps (Todd *et al.*, 1994). Fourth, when mMDH is added to a solution of cpn60, cpn10, and ATP that has already established steady-state hydrolysis, the enzyme folds with high efficiency. In this reaction the half-time for folding and for commitment to the native pathway is 700 s. Yet, when the chaperonin complex is imaged by cyro-electron microscopy, there is little electron density for bound mMDH at times when the folding reaction is less than 5% complete (A. R. Clarke and H. R. Saibil, unpublished data, 1994). Further, when the binary cpn60:mMDH complex is made first and then challenged with cpn10 and ATP, the immediately formed complex has mMDH bound in the "jaws" of the apical domains on the opposite ring to that which cpn10 is associated. Last, experiments in which release-deficient forms of cpn60 are used as "traps" for chaperonin-assisted substrates show that functional cpn60 must bind and release its substrates around 10 times during the refolding reaction; i.e., the rate of dissociation is about 10 times the folding rate (Weissman *et al.*, 1994). These last three observations suggest that interactions between the protein substrate and the chaperonin complex, which must be occurring in assisted folding, are of a transient nature and that the substrate is "bouncing" on and off during the process.

D. Active Recycling of Unproductive States

A second, general model for the activity of chaperonins proposes that cpn60 has the ability to **unfold** misfolded states and thus allow them another chance to refold to the native structure. These misfolded states may be conformationally trapped monomers or wrongly associated multimers, for example, partially folded but inactive protomers that have formed a dimer through the interaction of external hydrophobic patches. (In this context, it is interesting to note that even proteins that are active as monomers can form inactive dimers when refolded from the fully

denatured state.) Here the energy of Mg-ATP hydrolysis is used to cycle the chaperonin between states that have alternatively high and low affinities for an unfolded substrate and so produce a mechanism that entails binding, unfolding/dissociation, and then release (Jackson *et al.*, 1993; Weissman *et al.*, 1994; Staniforth *et al.*, 1994b). This mechanism (illustrated in Fig. 7) essentially allows the protein substrate to jump out of false minima in the folding reaction and prevent the formation of stable, misfolded structures (Hubbard and Sander, 1991; Weissman and Kim, 1992).

The crucial question is whether a "suppression-of-aggregation" or an active "recycling" model best explains the behavior of chaperonins. We suggest that the properties of chaperonins make it inevitable that both of these activities are integral parts of the mechanism of enhanced folding, rather than being mutually exclusive.

The argument begins by accepting that proteins, by and large, do not fold through a single pathway. The immediate collapse of a random coil leads to a range of folding intermediates, some of which are productive and easily isomerize to either the native state (see upper sequence of reactions in Fig. 8) or to a monomer competent for quaternary association. Some molecules, however, form less productive intermediates (described as "misfolded" in Fig. 8) that have a high energy barrier standing in the way of their progression to the native state. The situation is best illustrated by the experiments of Peralta *et al.* (1994), which show that a stable and misfolded state can be reactivated by chaperonins to allow folding to recommence. This observation is underlined by the ability of chaperonins to accelerate the folding of mMDH and to work at substoichiometric levels in this reaction. Thus each encounter can unlock an unproductive state and there is no need for each folding molecule to be held in its own cage.

The most straightforward means of reactivating a trapped structure, be it an incorrectly folded monomer or a wrongly associated multimer, is to break contacts within the existing structure to allow it the chance to explore a more productive pathway. To facilitate the breaking of hydrophobic contacts, the interaction with cpn60 must be tightest with the more unfolded or the more dissociated state in which hydrophobic exposure is greater. This is achieved by cpn60 providing alternative hydrophobic surfaces for these nonnative tertiary or quaternary contacts. Thus encounters between trapped states and the tight protein-binding form of cpn60 will drive unfolding and/or dissociation processes but will produce an overstable protein–protein complex. Chaperonin 60 then

has to be returned to its weak protein-binding state to allow dissociation of an untrapped state that can resume folding.

The postulated steps in such a process are illustrated in Fig. 9, which represents events occurring on one ring of cpn60 during a catalytic cycle (the accompanying ring will be in the opposite conformation). In the presence of cpn60:Mg-ADP, a spontaneously misfolded protein can bind tightly to the apical domains of cpn60 (step 1 in Fig. 9). In this tight protein-binding state, the high affinity for hydrophobic regions normally buried in folded states will then promote unfolding to a less compact conformation (step 2 in Fig. 9). The degree and nature of unfolding at this stage have not been clearly defined; it is not certain whether a large proportion of the secondary structure is lost (Zahn *et*

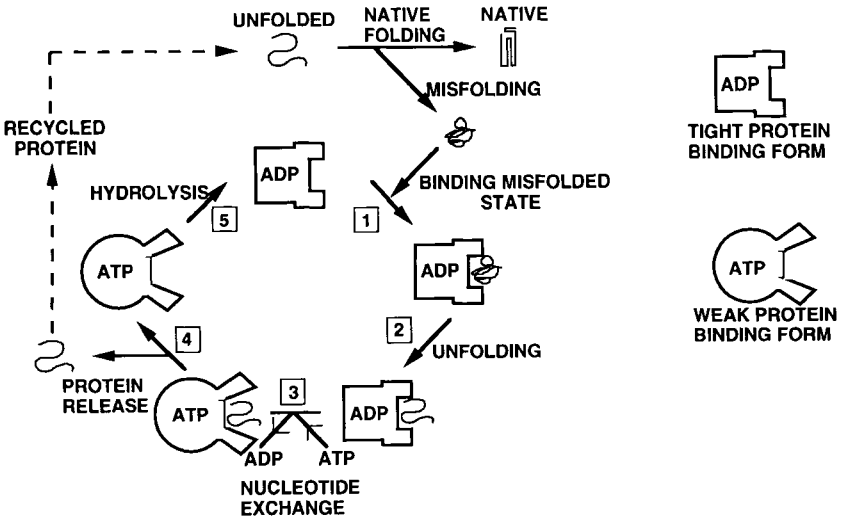


Fig. 9. A general model for the chaperonin mechanism; the binding/unfolding/release cycle. An unfolded protein can either fold to a productive intermediate, committed to fold to the native state, or to an unproductive misfolded state that may be a wrongly associated multimer (as in the case of mMDH, see Fig. 8) or a conformationally trapped monomer. Misfolded states bind to cpn60 (step 1) and are taken through the chaperonin cycle (clockwise) to be expelled in an untrapped state (step 4) in which inappropriate contacts within the substrate have been broken by the encounter with the hydrophobic surfaces of cpn60. The protein substrate is then allowed a further chance to fold correctly. The details are described in Section VI of the text.

al., 1994), whether inappropriate tertiary contacts between secondary structures are disengaged, or whether quaternary contacts between wrongly associated chains are broken. Although the diagram shows the misfolded state to be monomeric, in the case of mMDH it is a multimer (A. R. Clarke and H. R. Saibil, unpublished data, 1994).

In any case, the breaking of contacts, at least at the tertiary level, is facilitated by the fact that the hole is lined by hydrophobic binding sites acting in concert, so acting as a denaturing environment (Jackson *et al.*, 1993). The problem with the resultant complex is that the protein substrate may be so tightly bound that a large barrier to folding is produced. However, dissociation of ADP and association of ATP (step 3) weakens the protein-binding affinity to release the more unfolded or dissociated structure (step 4), which can then resume folding. Hydrolysis of ATP (step 5) reestablishes the tight protein-binding state of the ring.

As demonstrated experimentally, this cyclic mechanism is energetically costly and requires many molecules of ATP to be hydrolyzed (typically 50–200) per protein molecule folded (for examples see Martin *et al.*, 1991; Jackson *et al.*, 1993). With such a "catch-all" apparatus this is not entirely surprising and an explanation arises naturally from this model. When the protein substrate is discharged to allow folding to continue, it may again spontaneously misfold (to generate complex 2), so requiring further cycles. This mechanism is akin to a selective denaturation process. If a protein is allowed to fold spontaneously and a proportion (P) spontaneously misfolds, this material can then be bound, unfolded, and released to undergo a second folding process in which the same proportion (P) will misfold. The relationship between the efficiency (E) of folding (where $E = 1$ for a completely efficient process) and the number of cycles (n) can be described:

$$E = 1 - P^n.$$

Proteins that fold efficiently will preferentially populate productive states and have a low value for P; those that do not will predominantly populate misfolded states and have a high value. Thus the demand for ATP, reflected by the number of cycles required to reach high efficiency in the folding reaction, is dependent on the partition ratio.

The cycle depicted in Fig. 9 does not, explicitly, include the role of cpn10. It is likely that the coprotein exerts its effect most strongly in step 4 by promoting a more effective (Staniforth *et al.*, 1994b) or obligatory (Weissman *et al.*, 1994) displacement of the bound protein. A strong,

concerted release of the protein from all seven binding surfaces on a cpn60 toroid would allow it to form, unhindered, the network of cooperative intramolecular interactions necessary for the formation of a productive folding intermediate.

Although there remain difficult and unresolved elements in the mechanism of chaperonins, this tentative model proposes that they use the energy of Mg-ATP hydrolysis to drive the formation of productive intermediates by disrupting poorly folded, kinetically trapped conformers and early aggregates formed in folding pathways. With regard to the difference in molecular properties between productive intermediates and misfolded states, it is reasonable to suppose that the latter have a greater exposure of hydrophobic surface and so will be preferentially bound and disrupted by the chaperonins. Such effects, which are intrinsic to chaperonin behavior, may provide a general explanation for the coupling of Mg-ATP hydrolysis to improved protein folding in the cell.

REFERENCES

Anfinsen, C. (1973). Principles that govern the folding of protein chains. *Science* **181,** 223–230.

Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J. J., Lund, P., and Clarke, A. R. (1991). Binding of a chaperonin to the folding intermediates of lactate dehydrogenase. *Biochemistry* **30,** 9195–9200.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å *Nature (London)* **371,** 578–586.

Brunschier, R., Danner, M., and Seckler, R. (1993). Interactions of phage-P22 tailspike protein with GroE molecular chaperones during refolding *in vitro*. *J. Biol. Chem.* **268,** 2767–2772.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kiefhaber, T. (1991). GroE facilitates folding of citrate synthase by suppressing aggregation. *Biochemistry* **30,** 1586–1591.

Burston, S. G., Sleigh, R. N., Halsall, D. J., Smith, C. J., Holbrook, J. J., and Clarke, A. R. (1992). The influence of chaperonins on protein folding. *Ann. N.Y. Acad. Sci.* **672,** 1–9.

Burston, S. G., Ranson, N. A., and Clarke, A. R. (1995). The origins and consequences of asymmetry in the reaction cycle of cpn60. *J. Mol. Biol.* **249,** 138–152.

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R., and Saibil, H. R. (1994). Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryoelectron microscopy. *Nature* **371,** 261–264.

Creighton, T. E. (1992). The disulphide folding pathway of BPTI. *Science* **256,** 111–112.

Creighton, T. E. (1993). "Protein Folding" (T. E. Creighton, ed.). Freeman, New York.

Ellis, R. J. (1993). Chaperonin duet. *Nature (London)* **366,** 213–214.

Ellis, R. J. (1994). Opening and closing the Anfinsen cage. *Curr. Biol.* **4,** 633–635.

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171,** 1379–1385.

Fenton, W. A., Kashi, Y., Furtak, K., and Horwich, A. L. (1994). Residues in chaperonin GroEL required for substrate binding and release. *Nature (London)* **371,** 614–619

Fisher, M. T. (1992). Promotion of the *in vitro* folding of dodecameric glutamine synthetase from *Escherichia coli* in the presence of GroEL (chaperonin-60) and ATP. *Biochemistry* **31,** 3955–3963.

Garel, J.-R. (1992). *In* "Protein Folding" (T. E. Creighton, ed.), pp. 405–454. Freeman, New York.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends upon two chaperonin proteins and MgATP. *Nature (London)* **342,** 884–889.

Gragerov, A., Nudler, E., Komissarova, N., Gaitnaris, G., Gottesman, M. E., and Nikiforov, V. (1992). Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **89,** 10,341–10,344.

Gray, T. E., and Fersht, A. R. (1991). Cooperativity in ATP hydrolysis by GroEL is increased by GroES. *FEBS Lett.* **292,** 254–258.

Gray, T. E., and Fersht, A. R. (1993). Refolding of barnase in the presence of GroEL. *J. Mol. Biol.* **232,** 1197–1207.

Hansen, J. E., and Gafni, A. (1993). The thermal switching between enhanced and arrested reactivation of bacterial glucose-6-phosphate dehydrogenase assisted by GroEL in the absence of ATP. *J. Biol. Chem.* **268,** 21,632–21,636.

Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E., and Hartl, F.-U. (1994). Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. *EMBO J.* **13,** 3192–3202.

Horwich, A. L., Brooks Low, K., Fenton, W. A., Hirshfield, I. N., and Furtak, K. (1993). Folding *in vivo* of bacterial proteins: Role of GroEL. *Cell* **74,** 909–917.

Hubbard, T. J. P., and Sander, C. (1991). Heat shock proteins and protein folding. *Protein Engineering* **4,** 711–717.

Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbrook, J. J., Clarke, A. R., and Burston, S. G. (1993). The binding and hydrolysis of nucleotides in the chaperonin catalytic cycle—Implications for the mechanism of assisted protein folding. *Biochemistry* **32,** 2554–2563.

Jaenicke, R. (1993). Role of accessory proteins in protein folding. *Curr. Opinion Struct.* **3,** 104–112.

Jaenicke, R., Rudolph, R., and Heider, I. (1979). Quaternary structure, subunit activity and *in vitro* association of porcine malate dehydrogenase. *Biochemistry* **18,** 1217–1223.

Karplus, M., and Shakhovich, E. (1992). *In* "Protein Folding" (T. E. Creighton, ed.), pp. 127–195. Freeman, New York.

Landry, S. J., and Gierasch, L. M. (1991). The chaperonin GroEL binds a polypeptide in an α-helical conformation. *Biochemistry* **30,** 7359–7562.

Landry, S. J., Jordan, R., McMacken, R., and Gierasch, L. M. (1992). Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature* (*London*) **355,** 455–457.

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, G., and Gierasch, L. M. (1993). Genetic and biophysical identification of a functionally important mobile domain of GroES. *Nature* (*London*) **364,** 255–258.

Lorimer, G. H., and Viitanen, P. V. (1993). Personal communication.

Mann, C. J., and Matthews, C. R. (1993). Structure and stability of an early folding intermediate of *E. coli trp* aporepressor measured by far-UV stopped-flow circular dichroism and ANS binding. *Biochemistry* **32,** 5282–5290

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F. U. (1991). Chaperonin-mediated folding at the surface of groEL through a "mollen globule"-like intermediate. *Nature* (*London*) **352,** 36–42.

Martin, J., Mayhew, M., Langer, T., and Hartl, F. U. (1993a). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature* (*London*) **366,** 228–233.

Martin, J., Geromanos, S., Tempst, P., and Hartl, F.-U. (1993b). Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES. *Nature* (*London*) **366,** 279–282.

Matouschek, A., Serrano, L., and Fersht, A. R. (1992a). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224,** 819–835.

Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, B., and Fersht, A. R. (1992b). The folding of an enzyme. V. H/D exchange–NMR studies on the folding pathway of barnase: Complementarity to and agreement with protein engineering studies *J. Mol. Biol.* **224,** 837–845.

Miller, A. D., Maghlaoui, K., Alabnese, G., Kleinjan, D. A., and Smith, C. (1993). *Escherichia coli* chaperonins Cpn60 (groEL) and Cpn10 (groES) do not catalyze the refolding of mitochondrial malate dehydrogenase. *Biochem. J.* **291,** 139–144.

Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993). Detection of transient protein folding populations by mass spectrometry. *Science* **262,** 896–900.

Mizobata, T., Akiyama, Y., Ito, K., Yumoto, N., and Kawata, Y. (1992). Effects of the chaperonin GroE on the refolding of tryptophanase from *Escherichia coli:* Refolding is enhanced in the presence of ADP. *J. Biol. Chem.* **267,** 17,773–17,779.

Okazaki, A., Ikura, T., Nikaido, K., and Kuwajima, K. (1994). The chaperonin GroEL does not recognize apo-a lactalbumin in the molten globule state. *Nature Struct. Biol.* **1,** 439–446.

Peralta, D., Hartman, D. J., Hoogenraad, N. J., and Hoj, P. B. (1994). Generation of a stable folding intermediate which can be rescued by GroEL and GroES. *FEBS Lett.* **339,** 45–49.

Ptitsyn, O. B. (1992). *In* "Protein Folding" (T. E. Creighton, ed.), pp. 243–300. Freeman, New York.

Radford, S. E., Dobson, C. M., and Evans, P. A. (1992). The folding of hen lysozyme invokes partially structured intermediates and multiple pathways. *Nature* (*London*) **358,** 302–307.

Richarme, G., and Kohiyama, M. (1994). Amino acid specificity of the *E. coli* chaperone GroEL (HSP60). *J. Biol. Chem.* **269,** 7095–7098.

Robinson, C. V., Gross, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F.-U., Dobson, C. M., and Radford, S. E. (1994). Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry. *Nature (London)* **372,** 646–651.

Rothman, J. E. (1989). Polypeptide chain binding proteins: Catalysts of protein folding and related processes in cells. *Cell* **59,** 591–601.

Saibil, H. R., Zheng, D., Roseman, A. M., Hunter, A. S., Watson, G. M. F., Chen, S., Auf Der Mauer, A., O'Hara, B. P., Wood, S. P., Mann, N. H., Barnett, L. K., and Ellis, R. J. (1993). ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Curr. Biol.* **3,** 265–273.

Schmidt, M., and Buchner, J. (1992). Interaction of GroE with an all-β-protein. *J. Biol. Chem.* **267,** 16,829–16,833.

Schmidt, M., Bucheler, U., Kaluza, B., and Buchner, J. (1994a). Correlation between the stability of the GroEL–ligand complex and the release mechanism. *J. Biol. Chem.* **269,** 27,964–27,972.

Schmidt, M., Rutkat, K., Rachel, R., Pfeifer, G., Jaenicke, R., Viitanen, P., Lorimer, G., and Buchner, J. (1994b). Symmetric complexes of GroE chaperonins as part of the functional cycle. *Science* **265,** 656–659.

Schmidt, M., Buchner, J., Todd, M. J., Lorimer, G. H., and Viitanen, P. V. (1994c). On the role of GroES in the chaperonin-assisted folding reaction: Three case studies. *J. Biol. Chem.* **267,** 10,304–10,311.

Sosnick, T. R., Mayne, L., and Englander, S. W. (1994). The barriers of protein folding. *Nature Struct. Biol.* **1,** 149–156.

Staniforth, R. A., Cortes, A., Burston, S. G., Atkinson, T., Holbrook, J. J., and Clarke, A. R. (1994a). The stability and hydrophobicity of cytosolic and mitochondrial malate dehydrogenases and their relation to chaperonin-assisted folding. *FEBS Lett.* **344,** 129–135.

Staniforth, R. A., Burston, S. G., Atkinson, T., and Clarke, A. R. (1994b). The affinity of chaperonin-60 for a protein substrate and its modulation by nucleotides and chaperonin-10. *Biochem. J.* **300,** 651–658.

Teschner, W., Rudolph, R., and Garel, J.-R. (1987). Intermediates on the folding pathway of octopine dehydrogenase from Pecten-jacobaeus. *Biochemistry* **26,** 2791–2796.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1993). The hydrolysis of ATP by *Escherichia coli* GroEL: The Effects of GroES and K^+. *Biochemistry* **32,** 8560–8567.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. *Science* **265,** 659–666.

Van Dyk, T. K., Gatenby, A. A., and LaRossa, R. A. (1989). Demonstration by genetic suppression of interaction of groE products with many proteins. *Nature (London)* **342,** 451–453

Weissman, J. S., and Kim, P. S. (1991). Reexamination of the folding of BPT1—Predominance of native intermediates. *Science* **253,** 1386–1393.

Weissman, J. S., and Kim, P. S. (1992). Kinetic role of non-native species in the folding of bovine pancreatic trypsin-inhibitor. *Proc. Natl. Acad. Sci. USA* **89,** 9900–9904.

Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of non-native forms. *Cell* **78,** 693–702.

Yifrach, O., and Horovitz, A. (1994). Two lines of allosteric communication in the oligomeric chaperonin GroEL are revealed by the single mutation Arg196 → Ala. *J. Mol. Biol.* **243,** 397–401.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K., and Plückthun, A. (1994). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature* (*London*) **368,** 261–265.

Zettlmeissl, G., Rudolph, R., and Jaenicke, R. (1979). Reconstitution of lactate dehydrogenase: Non-covalent aggregation vs. reactivation. I. Physical properties and kinetics of aggregation. *Biochemistry* **18,** 5567–5571.

8

Role of Prokaryotic Chaperonins in Protein Folding

Jörg Martin, Mark Mayhew, and F.-Ulrich Hartl

Howard Hughes Medical Institute and Cellular Biochemistry & Biophysics Program
Memorial Sloan–Kettering Cancer Center
New York, New York 10021

I. INTRODUCTION

A. Development of Molecular Chaperone Concept

The important role molecular chaperones play in mediating cellular protein folding and assembly has only recently come to light. This is

remarkable given the fundamental biological significance of protein folding. However, the long-held view that both *in vitro* and *in vivo* protein folding are spontaneous processes provided an elegantly simple explanation of a very complex problem. Indeed it is astonishing that the *Escherichia coli* protein GroEL, which has been studied for over two decades (Georgopoulos *et al.*, 1973; Sternberg, 1973), has only in the past few years been demonstrated to mediate the folding of a large fraction of cytosolic proteins. Our current understanding of protein folding and oligomeric assembly has been primarily derived from three lines of investigation: (1) the analysis of the assembly of the enzyme ribulose-bisphosphate carboxylase-oxygenase (Rubisco, EC 4.1.1.39) in chloroplasts (see Chapters 1 and 3), (2) the analysis of protein folding in mitochondria, and (3) the *in vitro* reconstitution of a pathway of chaperone-mediated protein folding, which appears to occur generally in the cytosol of bacteria and eukaryotic cells.

A search for yeast mutants defective in mitochondrial protein import led Cheng *et al.* (1989) to the discovery of a temperature-sensitive lethal mutant allele with an unexpected phenotype. The test protein ornithine transcarbamoylase was imported normally into the mutant organelles; however, it was unable to form the enzymatically active trimer. Furthermore, the assembly of a number of other imported proteins was similarly affected. Analysis of the mutant strain pinpointed the mutation to the mitochondrial stress protein hsp60. Cloning and sequencing of the HSP60 gene (Reading *et al.*, 1989) revealed that hsp60 is homologous to the bacterial stress protein GroEL and the Rubisco subunit binding protein, both of which had been classified as "chaperonins" in the previous year (Hemmingsen *et al.*, 1988). Thus, mitochondria, chloroplasts, and bacteria were all shown to contain highly related forms of chaperonins that were thought to be required for the oligomeric assembly of proteins *in vivo* (Barraclough and Ellis, 1980; Ellis, 1987; Ellis and Hemmingsen, 1989).

Based on the ability of denatured proteins to spontaneously refold *in vitro* (Anfinsen, 1973), there was little doubt at the time that the folding *in vivo* of monomeric polypeptides or subunits of oligomeric proteins could occur without the assistance of other cellular components. Protein import into mitochondria provided an excellent experimental system to test whether this view was correct. It had become clear that proteins had to be unfolded (Eilers and Schatz, 1986), probably even fully extended (Rassow *et al.*, 1990), in order to cross the mitochondrial membranes. Initially, however, the folding reaction in the mitochondrial matrix was

believed to occur spontaneously, tightly coupled with translocation. This belief appeared to be consistent with the results of early studies on the folding of the cytosolic protein dihydrofolate reductase (DHFR) imported into mitochondria as a fusion protein with a cleavable N-terminal targeting sequence (Eilers and Schatz, 1988). A simple assay to measure the folding of imported DHFR was developed using the intrinsic protease resistance of DHFR as a marker. Surprisingly, folding in the mitochondrial matrix could be completely uncoupled from membrane translocation (Ostermann *et al.*, 1989). The dissociation of the two processes was most striking when the isolated organelles, prior to import, were partially depleted of ATP. Unfolded, protease-sensitive DHFR then accumulated in the matrix as a soluble complex with the 800-kDa hsp60 from which it could be chased into the folded monomer upon readdition of ATP. These findings demonstrated that the folding of a monomeric polypeptide chain in the cell is mediated by the chaperonin in an ATP-dependent reaction, despite the fact that DHFR can efficiently and spontaneously refold *in vitro*. Following the *in vitro* reconstitution of the GroEL/ES-dependent assembly of dimeric Rubisco (Goloubinoff *et al.*, 1989a), several studies using purified proteins established that the primary function of the chaperonins is indeed to fold monomeric polypeptide chains (Laminet *et al.*, 1990; Martin *et al.*, 1991; Viitanen *et al.*, 1991; Zheng *et al.*, 1993).

The use of the mitochondrial system also stimulated the development of an integrated view of the functions of the hsp70 and hsp60 (chaperonin) classes of molecular chaperones in protein folding. Work by Kang *et al.* (1990) showed that hsp70 is required for the process of membrane translocation by binding to polypeptide chains as they emerge from the inner surface of the inner mitochondrial membrane. Mitochondrial hsp70 is necessary, but not sufficient, for the folding of imported DHFR (see Chapter 4). These studies also support the notion that cytosolic hsp70 might interact with nascent polypeptide chains emerging from ribosomes, as demonstrated by Beckman *et al.* (1990). Subsequently, *in vitro* reconstitution of cellular protein folding, representative of both folding in the mitochondrial matrix after translocation and folding in the cytosol after translation, demonstrated the successive action of hsp70 and the chaperonins (Langer *et al.*, 1992a). Using the homologous chaperones of *E. coli*, DnaK (hsp70), DnaJ, GrpE, and GroEL/GroES (hsp60/hsp10), it was shown that DnaK and DnaJ are able to stabilize the mitochondrial monomeric protein rhodanese in an unfolded but folding-competent conformation. In the presence of the nucleotide-exchange factor GrpE

and ATP, the folding intermediate is transferred to GroEL and folded to the native state. These observations led to a proposed pathway of *in vivo* protein folding in which nascent polypeptide chains are bound by DnaK and DnaJ early during translation or membrane translocation, and subsequently can be transferred to a double-ring chaperonin for folding. The sequence of interactions of the various chaperones with the substrate polypeptide is proposed to be governed by their differential specificity for structural elements exposed by the polypeptide chain at different stages of the folding pathway. In support of the generality of ths concept, DnaJ homologs have been discovered in every cellular compartment that contains hsp70 (Caplan *et al.*, 1993), and very recently the first eukaryotic homolog of GrpE has been identified in mitochondria (for review, see Höhfeld and Hartl, 1994). The eukaryotic cytosol also contains a chaperonin, the *t*-complex polypeptide 1 (TCP-1) ring complex (see Chapter 5). In the mammalian cytosol, this chaperonin is indeed able to function in the domain-wise folding of nascent polypeptide chains in a coupled folding pathway together with hsp70, the DnaJ-homolog Hsp40, and perhaps other factors (Frydman *et al.*, 1994).

B. Requirement for Molecular Chaperones in Cellular Protein Folding

An unfolded polypeptide chain emerging from the ribosome is thrown into an environment that provides unfavorable conditions for folding: the concentration of total protein in the cytosol is of the order of 25–30%, resulting in excluded volume effects (Zimmerman and Trach, 1991). More importantly perhaps, the concentration of unfolded polypeptide chains is high, reaching 30–50 μM in an actively growing *E. coli* cell. Unfolded or partially folded polypeptides are intrinsically insoluble. They expose hydrophobic surfaces that make them prone to aggregation in a concentration-dependent manner. The tendency to engage in such unproductive interactions becomes more pronounced with increasing temperature, a critical factor considering the need of many cell types to fold proteins with high efficiency over a range of temperatures. Furthermore, the acquisition of stable tertiary structure requires the presence of at least a complete, independently folding, protein domain (usually 100–200 amino acid residues in length). It follows that during translation and translocation a nascent polypeptide must remain exposed in an

unfolded or partially folded state until such a polypeptide segment (or the complete chain in the case of single-domain proteins) has become available for folding.

In vitro many small proteins are capable of folding spontaneously. This led Anfinsen to propose that the amino acid sequence of a protein alone is sufficient to specify its final native structure (Anfinsen, 1973). However, an attempt to simulate cellular folding conditions in an *in vitro* folding experiment, starting with an unfolded protein diluted from denaturant, would almost inevitably lead to misfolding and aggregation. Not surprisingly therefore, *in vitro* refolding is usually performed at low concentrations of a single protein and at an individually optimized pH, salt concentration, and temperature.

Although Anfinsen's proposal still holds true today, it can only in part explain how proteins are able to fold so efficiently *in vivo* (Ellis and Hemmingsen, 1989). Molecular chaperones are probably present in all the cellular compartments in which proteins fold. The major role of these versatile helper proteins seems to be in the prevention of unproductive intra- and intermolecular interactions that can occur during folding. Molecular chaperone proteins come in different shapes and sizes (Ellis and van der Vies, 1991; Gething and Sambrook, 1992; Hendrick and Hartl, 1993; Ellis, 1994a; Hartl *et al.*, 1994). Most of them have the ability to interact with many different unfolded proteins. Some chaperones, for example, the members of the hsp70 family or the bacterial chaperone SecB, primarily stabilize polypeptide chains in an unfolded state, whereas other mediate ATP-dependent folding to the native conformation. This latter group of components, represented in *E. coli* by GroEL/GroES, are termed chaperonins. On the following pages we will describe the current view of the mechanisms involved in chaperone-mediated protein folding in bacteria with emphasis on the role of the chaperonins. We will also discuss the functions of molecular chaperones in a variety of cellular processes that are critically dependent on protein conformation, such as the proteolytic degradation of proteins.

II. PATHWAY OF CHAPERONE-ASSISTED PROTEIN FOLDING

During translation and folding, a nascent polypeptide in the bacterial cytosol first makes contact with the hsp70 homolog DnaK and its regulator, DnaJ (Hendrick *et al.*, 1993; Gaitanaris *et al.*, 1994; Kudlicki *et al.*,

1994). These two heat shock proteins are cotranscribed from a single operon and cooperate functionally as a chaperone team by directly interacting with each other (Georgopoulos, 1992). The hsp70s are known to recognize short, extended peptides enriched in hydrophobic amino acid residues (Flynn *et al.*, 1989, 1991; Landry *et al.*, 1992; Blond-Elguindi *et al.*, 1993; Gragerov *et al.*, 1994). The binding specificity of the 40-kDa chaperone DnaJ is not yet understood, but a recent study has shown that DnaJ is able to interact with a nascent polypeptide once 40 amino acid residues have emerged from the ribosome (Hendrick *et al.*, 1993). Both DnaK and DnaJ are specifically associated with translating polyribosomes (Gaitanaris *et al.*, 1994). The sequence in which DnaK and DnaJ interact with nascent chains may depend on the structural properties of the N-terminal segment of the growing polypeptide chain. Studies *in vitro* have shown that DnaK and DnaJ do not interact independently with their substrate (Schröder *et al.*, 1993; Szabo *et al.*, 1994). With unfolded firefly luciferase or rhodanese as substrate proteins, DnaJ has been shown to bind first, subsequently targeting the substrate proteins to DnaK. These observations were based on the ability of DnaJ to stimulate the hydrolysis of DnaK-bound ATP (Liberek *et al.*, 1991), thus relatively stabilizing the ADP-state of DnaK which has a high polypeptide binding affinity (Palleros *et al.*, 1993). As a result, a stable substrate protein/DnaJ/DnaK complex is formed in the presence of hydrolyzable ATP (Langer *et al.*, 1992a; Szabo *et al.*, 1994) (Fig. 1). A protein bound in this complex has certain conformational properties of a compact folding intermediate and is effectively prevented from aggregating (Langer *et al.*, 1992a). To induce dissociation of the substrate protein, a 24-kDa heat shock protein, GrpE, must interact with the complex, thereby stimulating the exchange of ADP for ATP. Subsequent ATP binding by DnaK results in the release of a protein folding intermediate, which may proceed along the folding pathway to the native state, or be transferred to GroEL/GroES, or rebind to DnaJ and DnaK in another round of ATP hydrolysis (Langer *et al.*, 1992a; Szabo *et al.*, 1994). Whether DnaK, DnaJ, and GrpE alone are sufficient to mediate protein folding *in vivo* is not yet known. It seems possible that some small, single-domain proteins with a strong propensity to fold spontaneously *in vitro* may be able to fold without further assistance after their release from DnaK/DnaJ. Many proteins, however, may have to be transferred to GroEL/GroES to achieve their native conformation (Fig. 1). The function of DnaK and DnaJ may therefore be the stabilization of nascent proteins in a nonaggregated, folding-competent conformation until the chaperonin is available.

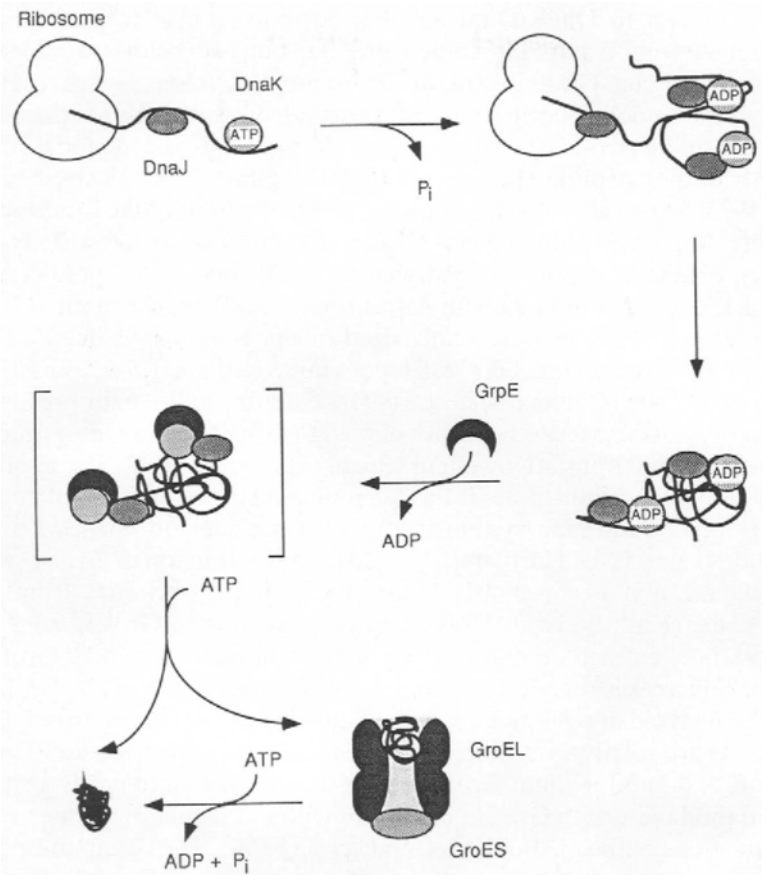


Fig. 1. Model for the pathway of protein folding in the bacterial cytosol. As the polypeptide chain emerges from the ribosome, it is bound by DnaJ and DnaK (hsp70) (Hendrick *et al.*, 1993; Gaitanaris *et al.*, 1994). The direct interaction between DnaK and DnaJ in the presence of ATP leads to the formation of a ternary complex between nascent chain, DnaK and DnaJ in which DnaK is in the ADP state. This complex is resolved upon the GrpE-dependent dissociation of ADP and the binding (not hydrolysis) of ATP (Szabo *et al.*, 1994). The protein may then fold to the native state by multiple rounds of interaction with the DnaK, DnaJ, GrpE system or is transferred for final folding to GroEL/GroES (Langer *et al.*, 1992a). We assume that a large fraction of cytosolic proteins have to interact with one or both chaperone systems to reach the native state.

In contrast to DnaK/DnaJ, GroEL not only is able to prevent the aggregation of a partially folded protein, but also allows productive folding to occur. The ability of GroEL to promote folding arises from its peculiar homooligomeric structure consisting of two stacked heptameric rings forming a large central cavity, which apparently accommodates the substrate polypeptide (Langer *et al.*, 1992b; Braig *et al.*, 1993; Saibil *et al.*, 1993; Ishii *et al.*, 1994) (see following). On the basis of the localization of the polypeptide binding site inside the cavity and its low affinity for short, extended peptides (Landry *et al.*, 1993), one would predict that GroEL could not interact with nascent polypeptide chains until at least a significant segment were synthesized or the complete chain released from the ribosome. Indeed, GroEL does not cosediment with translating polyribosomes (Gaitanaris *et al.*, 1994). Only the full-length precursor form of β-lactamase has been crosslinked to GroEL when synthesized in a bacterial S30 translation system (Bochkareva *et al.*, 1988). For another protein, bovine rhodanese, it has been demonstrated that its folding in a bacterial translation system is strictly dependent on GroEL/ES (J. Hendrick and F.-U. Hartl, unpublished). The interaction of GroEL with rhodanese may occur shortly before the protein is released from the ribosome (Kudlicki *et al.*, 1994). Immunodepletion of GroEL from the translation extract inhibits folding and addition of purified GroEL/GroES increases the yield of active enzyme (Tsalkova *et al.*, 1993). Due to the unavoidable dilution of cytosol during cell lysis, these translation extracts are relatively depleted of chaperones. As expected, addition of DnaK and DnaJ without GrpE inhibits the interaction of newly synthesized rhodanese with GroEL, consistent with the sequence of chaperone interactions outlined above (J. Hendrick and F.-U. Hartl, unpublished).

A directed pathway of protein folding from DnaK/DnaJ to GroEL has been demonstrated most directly *in vitro* with purified chaperone components and the substrate protein rhodanese (Langer *et al.*, 1992a). The chaperones DnaK and DnaJ cooperate in preventing the aggregation of the unfolded rhodanese when it is diluted from denaturant; however, they are unable to mediate its efficient folding even in the presence of GrpE and Mg-ATP. The protein cycles in a partially folded state between chaperone-bound and -free forms but does not proceed to the native state. Only after addition of GroEL and GroES does efficient folding occur. Transfer of partially folded rhodanese from DnaK/DnaJ to the chaperonin is dependent on GrpE. We envision that, in the context of translation, the folding process may be initiated cotranslationally, assisted by DnaK and DnaJ, and that the completion of folding to the

native state occurs posttranslationally in association with GroEL (Fig. 1). A careful analysis of the folding of translating polypeptide chains will be necessary to test this hypothesis.

The hierarchy and the sequence of chaperone interactions may be largely determined by the different structural specificities of DnaK/DnaJ and GroEL for extended polypeptide chains and collapsed folding intermediates, respectively. It appears likely that DnaK and DnaJ would normally be in contact with a nascent chain before GroEL, and little if any bypass should occur. Surprisingly, however, deletion mutants in *dnaK* and *dnaJ* are viable within a limited temperature range, although the cells grow slowly and rapidly acquire extragenic suppressors (Bukau and Walker, 1989). At heat shock temperatures, these strains stop growing. In contrast, GroEL is an essential protein under all growth conditions (Fayet *et al.*, 1989). The recent identification of a second DnaK-related hsp70 in *E. coli*, and of additional DnaJ homologs, may explain why cells can grow without DnaK and DnaJ at normal growth temperatures (Ueguchi *et al.*, 1994; Seaton and Vickery, 1994).

III. MECHANISM OF CHAPERONIN-MEDIATED PROTEIN FOLDING

The structure and function of the *E. coli* chaperonin GroEL are intimately related (see Chapter 9). Two heptameric rings of identical 58-kDa subunits are symmetrically stacked to form a hollow cylinder of ~800 kDa (Hendrix, 1979; Hohn *et al.*, 1979; Langer *et al.*, 1992b; Saibil *et al.*, 1993; Martin *et al.*, 1994) that is able to bind one or two molecules of unfolded substrate protein (Martin *et al.*, 1991; Bochkareva *et al.*, 1992). The physiologically relevant stoichiometry is likely to be one substrate protein bound per GroEL 14-mer. Several electron microscopy studies have presented evidence that substrate binding occurs within the central cavity of GroEL (Langer *et al.*, 1992b; Braig *et al.*, 1993; Ishii *et al.*, 1994), apparently at the level of the outer domains of the GroEL subunits (Chen *et al.*, 1994). GroEL has 14 ATP binding sites and exhibits a weak Mg^{2+}- and K^{+}-dependent ATPase activity (Viitanen *et al.*, 1990; Bochkareva *et al.*, 1992). The kinetic aspects of the GroEL ATPase are discussed in detail in Chapter 7). A second essential component of the chaperonin system is the chaperonin GroES, a heptameric ring of identical 10-kDa subunits (Chandrasekhar *et al.*, 1986), which is encoded

together with GroEL in one operon. GroES has been shown to interact with GroEL in the presence of nucleotides and to inhibit the basal ATPase activity of GroEL by 50–100%, dependent on the concentration of K^+ and the ratio of ATP to ADP concentrations (Chandrasekhar *et al.*, 1986; Viitanen *et al.*, 1990; Martin *et al.*, 1991; Jackson *et al.*, 1993; Todd *et al.*, 1993). Thus, a complete understanding of chaperonin function will require information about the structural and kinetic aspects of all the interactions between the participating components, GroEL, GroES, ATP, ADP, and substrate protein.

Our knowledge of the mechanism of GroEL/GroES-mediated protein folding is mainly derived from *in vitro* studies using purified components. Substrate binding to GroEL is usually achieved by unfolding a complete polypeptide chain in denaturant and diluting it into GroEL-containing buffer solutions (Goloubinoff *et al.*, 1989a; Laminet *et al.*, 1990; Badcoe *et al.*, 1991; Buchner *et al.*, 1991). GroES and nucleotide are added subsequently to initiate folding. These experiments bypass the translation of the proteins on ribosomes and the interaction of the elongating chain with DnaK and DnaJ. Furthermore, they do not reflect the *in vivo* situation where GroEL, GroES, and nucleotides are always present simultaneously.

Preferred substrate proteins, such as rhodanese or citrate synthase, tend to aggregate when diluted from denaturant into buffer solution alone, and only little spontaneous reactivation is achieved (Goloubinoff *et al.*, 1989a; Laminet *et al.*, 1990; Badcoe *et al.*, 1991; Buchner *et al.*, 1991; Martin *et al.*, 1991; Mendoza *et al.*, 1991). It was observed in early studies that binding of these substrate proteins to GroEL prevents aggregation, and that the addition of nucleotides and GroES leads to their release and refolding (Buchner *et al.*, 1991; Martin *et al.*, 1991; Mendoza *et al.*, 1991). It was concluded that unfolded proteins bind to GroEL in the absence of nucleotides and that ATP hydrolysis and GroES are required for their release. Soon it became clear, however, that certain proteins, such as DHFR and β-lactamase can be released from GroEL by the addition of ATP or even nonhydrolyzable ATP analogs, in the absence of GroES (Laminet *et al.*, 1990; Martin *et al.*, 1991; Viitanen *et al.*, 1991). In some cases, even addition of ADP is able to cause the release of bound substrates (Kawata *et al.*, 1994). In contrast to rhodanese or citrate synthase, these proteins have a strong tendency to fold spontaneously *in vitro*. Whether the folding of these proteins is also GroES-independent *in vivo* remains to be seen. It has to be stressed that, so far, there is no evidence that the chaperonin system directs the pathway of protein folding, or that it affects the outcome of the folding reaction.

A. Role of GroES

GroES is critically required to regulate the ATPase-dependent function of GroEL in binding and releasing the substrate protein. GroES is not necessary for ATP-dependent protein release from GroEL per se; however, it is needed to coordinate the release of many proteins from GroEL to achieve productive folding. This requirement can be illustrated by comparing the chaperonin-mediated folding of DHFR and rhodanese, both monomeric proteins but with fundamentally different folding properties: DHFR, a 20-kDa protein, spontaneously refolds with high efficiency upon dilution from denaturant, but can be trapped as a folding intermediate on GroEL in the absence of nucleotides (Martin *et al.*, 1991; Viitanen *et al.*, 1991). Addition of Mg-ATP (or nonhydrolyzable ATP analog) results in the release of the substrate protein and productive folding. GroES is not required for this reaction, although it accelerates its kinetics. In contrast, the 33-kDa protein rhodanese has a strong tendency to aggregate and does not refold spontaneously under most refolding conditions *in vitro*. Again, a folding intermediate can be trapped on GroEL, but ATP hydrolysis alone does not lead to productive refolding. Only in the presence of GroES does the efficient reactivation of rhodanese occur (Martin *et al.*, 1991; Mendoza *et al.*, 1991). However, the function of GroES is not simply to enhance the release of "more difficult" substrate proteins such as rhodanese (Martin *et al.*, 1991). When a preformed GroEL–rhodanese complex is incubated with Mg-ATP and casein, a soluble protein that exposes hydrophobic surfaces and binds to GroEL, the rhodanese folding intermediate is displaced from GroEL and subsequently aggregates. Interestingly, casein has no such effect when the same experiment is performed in the presence of GroES. It was concluded that in the presence of GroES rhodanese is released from GroEL in a conformation that is partially or completely protected against aggregation (Martin *et al.*, 1991, see also Schmidt *et al.*, 1994a). GroES apparently allows the bound polypeptide to undergo a critical folding step(s) before it is released from GroEL. Taking into account that substrate proteins bind to the interior of the chaperonin cylinder, it seems possible that in the presence of GroES the substrate protein might initially be released into the cavity, where it would be transiently sequestered from other folding proteins (Fig. 2). This should also be the case for proteins such as DHFR. Addition of casein to a DHFR–GroEL complex in the presence of Mg-ATP accelerates the rate of DHFR

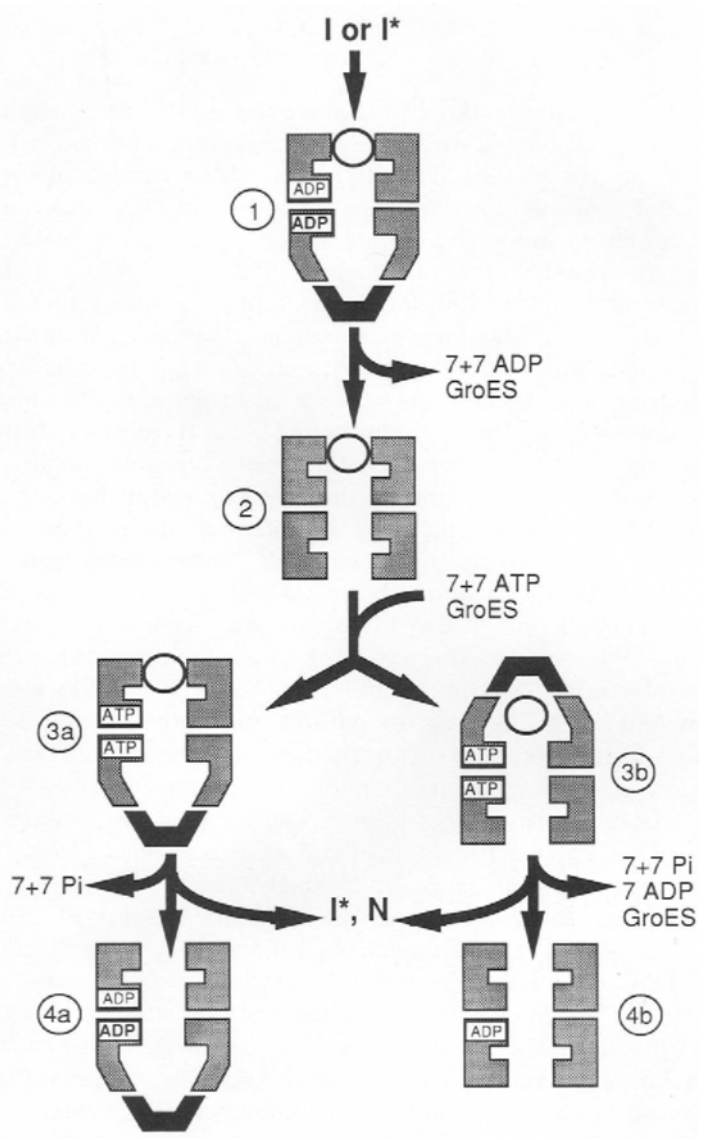


Fig. 2. Model for the ATP-dependent interaction between GroEL, GroES, and folding polypeptide. ADP (bold), the high-affinity ADP state in the seven subunits of GroEL that are bound to GroES; ADP (not bold), the lower ADP affinity of the subunits in the opposite toroid that may hydrolyze ATP; ATP, the subunits in a GroEL toroid in the ATP-bound state; I, polypeptide substrate as compact folding intermediate; I', folding

folding (Martin *et al.*, 1991). This observation indicates that DHFR is released from GroEL and can be prevented from rebinding to the chaperonin by the competitor casein, resulting in rapid spontaneous folding. Apparently, the conformation of the released DHFR is not yet native, thereby allowing it to rebind to GroEL. As expected, this effect is more pronounced at concentrations of GroEL in molar excess over DHFR. However, in the presence of GroES no competition by casein is observed. DHFR folds rapidly, and independently of the GroEL concentration, consistent to a certain degree with GroES-dependent folding on GroEL. The extent to which a substrate protein folds prior to its release from the chaperonin remains to be determined. For example, the small protein barnase has been shown to fold while closely associated with GroEL (Gray and Fersht, 1993; Gray *et al.*, 1993). Proteins with rapid folding kinetics may leave GroEL in their native state, whereas others that fold more slowly may emerge as partially folded intermediates. These intermediates are then able to rebind to the same or any other available GroEL molecule (Todd *et al.*, 1994; Weissman *et al.*, 1994). We assume, however, that these proteins may also undergo a critical folding step(s) within GroEL that strongly reduces their tendency to aggregate.

B. Reaction Cycle of GroEL and GroES

Several studies have shed light on the mechanism of GroES action (Jackson *et al.*, 1993; Martin *et al.*, 1993a,b; Todd *et al.*, 1993, 1994). GroEL and GroES exist predominantly as a 1 : 1 complex, the so-called holochaperonin. This complex dissociates upon isolation in the absence of nucleotide. It remains stable, however, in case of the chaperonin purified from *Thermus thermophilus* (Taguchi *et al.*, 1991). Typically one

intermediate part way advanced toward the native state; N, native protein. 1, Binding of polypeptide substrate facilitates the dissociation of GroES by accelerating ATP hydrolysis in the substrate-bound ring of GroEL and by lowering the affinity of ADP in the opposite ring. 2, Polypeptide is bound stably in the nucleotide-free state of GroEL. 3, ATP and GroES rebind. GroES may associate either with the free GroEL ring (3a) or with the polypeptide-containing ring (3b). 4a, ATP hydrolysis leads to polypeptide release and incompletely folded polypeptide rebinds in (1). In 3b, polypeptide is transiently enclosed in the central cavity and is free to fold. ATP hydrolysis in the GroES-bound toroid then generates the tight ADP state (not shown). ADP dissociates upon ATP hydrolysis in the opposite ring, causing dissociation of GroES and allowing polypeptide release.

GroES heptamer binds to one end of the GroEL tetradecamer, giving rise to an asymmetrical structure (Saibil *et al.*, 1991, 1993; Ishii *et al.*, 1992; Langer *et al.*, 1992b). This complex forms in the presence of Mg-ATP or Mg-ADP but is most stable in the presence of the latter (Fig. 2). Electron microscopy analysis of the asymmetrical particles demonstrates that GroES binding induces a conformational change at the opposite end of GroEL, which presumably reduces the affinity for a second GroES heptamer. Recently, conditions have been established that allow the electron microscopic visualization of apparantly symmetrical particles composed of two GroES heptamers bound to one GroEL tetradecamer (Schmidt *et al.*, 1994b; Llorca *et al.*, 1994). These complexes are not isolatable as stable entities. It has been postulated that symmetrical complexes play a key role in the GroEL/GroES reaction cycle; however, direct evidence for this remains elusive (Todd *et al.*, 1994). Interestingly, in the presence of unfolded polypeptide, symmetrical particles are not observed by electron microscopy (J. Flanagan, personal communication). Under physiological conditions, therefore, a substrate protein will most likely encounter a 1:1 GroEL-GroES complex in the presence of nucleotide. Consistent with recent electron microscopy studies, this conclusion then also suggests that, at least initially, substrate protein and GroES associate at opposite ends of the GroEL tetradecamer (Ishii *et al.*, 1994; Chen *et al.*, 1994) (Fig. 2).

It has been demonstrated that substrate protein and GroES have opposite effects on the GroEL ATPase activity. Binding substrate protein to the GroEL–GroES complex stimulates the ATP hydrolysis of GroEL, whereas GroES, while being essential for productive folding, has an inhibitory effect on the GroEL ATPase (Martin *et al.*, 1991; Jackson *et al.*, 1993). As demonstrated recently, GroES inhibits the ATPase activity of GroEL by stabilizing the holochaperonin complex in the ADP-bound state (Jackson *et al.*, 1993; Martin *et al.*, 1993b; Todd *et al.*, 1993). GroES binding strongly increases the ADP and ATP affinities of GroEL (from mM to nM) (Jackson *et al.*, 1993; Martin *et al.*, 1993b). Specifically, the seven ADP molecules in the GroEL toroid that is in direct contact with GroES are tightly bound and exchange slowly (Martin *et al.*, 1993b). The opposite toroid of GroEL is able to hydrolyze ATP (Martin *et al.*, 1993b; Todd *et al.*, 1994), resulting in a 50% inhibition of the observed ATPase activity. Thus, the asymmetrical binding of GroES confers half-of-the-sites reactivity to the GroEL double-ring. Stable binding of GroES requires the hydrolysis of seven molecules of ATP to ADP in the interacting toroid of GroEL. Further-

more, in the presence of GroES, nucleotide binding and ATP hydrolysis by the subunits of GroEL become more cooperative within the heptameric rings (Gray and Fersht, 1991; Bochkareva *et al.,* 1992; Jackson *et al.,* 1993; Martin *et al.,* 1993a). These effects are the result of significant conformational rearrangements in GroEL upon GroES binding. A mobile loop region within GroES is thought to play an important role in the interaction with GroEL (Landry *et al.,* 1993). Furthermore, the ability of GroES to bind, but not hydrolyze, ATP could also be important in the interplay between the two components (Martin *et al.,* 1993a).

In the absence of substrate protein, when the GroEL ATPase activity is 50% inhibited (i.e., at a high ratio of [ATP]/[ADP] and 100 m*M* K^+), GroES cycles between bound and free states, interacting with either side of the GroEL double-ring (Martin *et al.,* 1993b). Release of GroES is apparently coupled to ATP hydrolysis in the opposite, noninhibited ring. Substrate protein plays an active role in the cycle. The binding of unfolded polypeptide to GroEL weakens the GroES-induced ATPase inhibition by facilitating the release of tightly bound ADP from the GroEL–GroES complex (Martin *et al.,* 1993b) and by directly stimulating ATP hydrolysis itself (Martin *et al.,* 1991; Jackson *et al.,* 1993), this being the rate-limiting step of the reaction cycle (Fig. 2). As a consequence, the cycling of GroES between bound and free states is accelerated. This effect is most pronounced when cycling is initiated from a tight GroEL-GroES complex in which the GroEL ATPase activity is almost completely inhibited, despite the presence of free ATP (Martin *et al.,* 1991). The addition of substrate protein to this complex induces the dissociation of ADP and GroES, thereby allowing ATP to occupy the free nucleotide binding sites. Adenosine triphosphate binding is followed by the reassociation of GroES to either ring of GroEL, resulting in the highly cooperative hydrolysis of the bound ATP. This hydrolysis allows the release of substrate protein from its multiple attachment sites to the individual GroEL subunits (Fig. 2). The cross-linking of GroES to unfolded substrate protein during the folding cycle is consistent with the possibility that GroES transiently covers the GroEL ring where the substrate is bound (Bochkareva and Girshovich, 1992). This binding of GroES could result in polypeptide release into the cavity (see following). Alternatively, binding of GroEs to the empty toroid of GroEL may induce allosteric changes in the opposite ring that cause polypeptide release. Both mechanisms of substrate dissociation may coexist (Fig. 2).

During the short, nucleotide-free phase of the reaction cycle the substrate protein is bound by GroEL with highest affinity. The stability of

this interaction is subsequently lowered by ATP binding (in the absence of GroES, GroEL has a lower affinity for ADP than for ATP). Adenosine triphosphate hydrolysis, not just binding, seems to be required for the release and folding of tightly bound proteins. It is possible that during ATP hydrolysis, GroEL undergoes a conformational change that hides its binding sites for substrate protein. Rebinding of incompletely folded polypeptide would then occur in the ADP- and GroES-bound state, inducing another round of the reaction cycle. Approximately 130 molecules of ATP were found to be hydrolyzed by GroEL per molecule of rhodanese folded (Martin *et al.*, 1991). Assuming that 14 molecules of ATP are hydrolyzed in one cycle, the folding of a rhodanese molecule might require 8–9 cycles of GroEL binding and release. This number may vary considerably with different proteins, depending on their propensity to fold. Dihydrofolate reductase, for example, is expected to require less folding cycles than rhodanese or citrate synthase.

Cryoelectron microscopy of GroEL–GroES complexes has shown that binding of GroES causes a major conformational rearrangement in the interacting GroEL toroid, resulting in a significant increase in the size of its internal cavity (Chen *et al.*, 1994). The binding sites for GroES are at the outer domains of the GroEL subunits and seem to overlap partially with those for the substrate polypeptide. These observations suggest an attractive mechanism for the GroES-mediated release of polypeptide into the internal cavity. The central cavity of a single GroEL toroid is in principle large enough to accommodate substrate proteins up to ~50 kDa, taking into account that compact folding intermediates are about 15% larger in radius than the native state. Currently it is not known whether the central cavities of each GroEL toroid are connected, and, if so, whether substrate proteins are able to move between toroids within the GroEL cylinder. If such a movement were not possible (Saibil *et al.*, 1993), then the substrate polypeptide would be confined within the GroEL toroid to which it initially bound, until ATP hydrolysis in the opposite ring of GroEL triggers the release of GroES (Fig. 2).

A central feature of the GroEL/GroES reaction cycle is the high degree of functional coordination at the level of the individual GroEL rings achieved by GroES. Upon release of GroES from one GroEL toroid, all seven subunits of this toroid are allowed to bind ATP simultaneously, and subsequently hydrolyze it. This simultaneous hydrolysis may ensure that all substrate binding sites release their bound polypeptide segments in a coordinated fashion. Likewise, all subunits in a GroEL heptamer reach the ADP state synchronously. The time window between

polypeptide release and rebinding would therefore be of a specific magnitude, whereas in the absence of GroES, individual subunits of the same toroid might be in high- or low-affinity states for substrate binding, reducing the efficiency of complete polypeptide release and the time available for efficient folding.

IV. CONFORMATIONAL PROPERTIES OF CHAPERONIN-BOUND PROTEINS

Molecular chaperones have the ability to discriminate between unfolded and folded proteins. This remarkable capacity is based on the recognition of relatively unspecific structural features that are common to most unfolded or partially folded polypeptides but are typically absent from native protein conformations. One of the most obvious parameters that distinguishes folded and unfolded polypeptides is the exposure of hydrophobic amino acid residues in the unfolded state. Pelham proposed that the hsp70s could shield these hydrophobic residues, thereby preventing the aggregation of unfolded polypeptides (Pelham, 1986). More recent findings confirm this proposal. Various members of the hsp70 family have a high affinity for extended 7- to 9-residue-long peptides that are enriched in hydrophobic amino acids (Flynn *et al.*, 1989, 1991). These peptides might bind into a cleft similar to the one found in the MHC class I histocompatibility glycoprotein (Blond-Elguindi *et al.*, 1993; Gragerov *et al.*, 1994).

Comparatively less is known about the (poly)peptide binding specificity of the chaperonins. The spectrum of cytosolic proteins that contain potential GroEL binding sites is very broad. About 50% of the soluble proteins of *E. coli* are able to bind to GroEL upon dilution from denaturant (Viitanen *et al.*, 1992), and the folding of about 30% of cytosolic proteins was found to be affected in a temperature-sensitive *groEL* mutant strain (Horwich *et al.*, 1993). Similar results were obtained by analyzing the binding of heat-denatured proteins to hsp60 in mitochondria. A large number of different polypeptides associate with the chaperonin when the organelles are exposed to elevated temperatures up to 46°C (Martin *et al.*, 1992). In all these studies the chaperonin-bound proteins range in size between 10 and 90 kDa. It seems realistic to assume, based on the dimensions of the GroEL cylinder, that proteins up to 40–50 kDa can be accommodated in the ring cavity of the chaper-

onin. This is based on the consideration that proteins bind to the chaperonin as compact folding intermediates (see following). Larger proteins may protrude from the opening of the cylinder when bound by the chaperonin. The largest protein that has been shown to form a complex with GroEL *in vitro* is the 124-kDa phytochrome photoreceptor (Grimm *et al.*, 1993). Significantly, it is released by binding of a nonhydrolyzable ATP analog. Neither ATP hydrolysis nor GroES is required for phytochrome release, which suggests that the protein is already folded to a large extent and binds only weakly to GroEL, perhaps via an incompletely folded domain. Simply reducing the concentration of free-folding intermediate may, in this case, be sufficient to prevent aggregation and allow the completion of folding. Such a rudimentary interaction would be mechanistically distinct from the binding by GroEL of an unfolded substrate protein that has to undergo GroES and ATP hydrolysis-dependent folding.

Systematic binding studies with a wide spectrum of different peptides have not been performed with GroEL, due to the fact that short peptides display only a relatively low binding affinity for the chaperonin. However, using nuclear magnetic resonance techniques, it is possible to derive information about the conformation of peptide in the chaperonin-bound state. It has been demonstrated that a peptide derived from the N-terminus of rhodanese is bound by GroEL in an α-helical conformation (Landry and Gierasch, 1991; Landry *et al.*, 1992). The same peptide assumes an extended conformation when bound to DnaK, indicating that GroEL either induces the formation of secondary structure in these shorter peptides or, when presented with a peptide able to assume different conformations, preferentially stabilizes those with α-helical structure. On the other hand, not all peptides studied are stabilized in an α-helical conformation by GroEL. The strength of binding seems to correlate with the overall hydrophobicity of the peptide substrate (Landry and Gierasch, 1994). Nevertheless, the potential to form an α-helical structure could favor tight association, since the binding of an all β-protein to the chaperonin does not result in a very stable complex (Schmidt and Buchner, 1992). These and other data (Rosenberg *et al.*, 1993; Dessauer and Bartlett, 1994) are consistent with the chaperonin recognizing certain hydrophobic amino acid side chains, presumably in the form of hydrophobic patches or clusters. Negative charges are found to be largely absent from peptides that display chaperonin affinity. The dissociation constants of several substrate proteins have been measured recently and are found to be in the nanomolar range (Gray and Fersht, 1993; Dessauer and Bartlett, 1994; Staniforth *et al.*, 1994).

Despite the promiscuity in substrate binding there appear to be significant differences among chaperonin proteins. In this context it is interesting to note that several bacterial species express more than one form of the GroEL and GroES proteins. *Bradyrhizobium japonicum,* for example, expresses as many as five different GroEL and also different GroES genes that are regulated in a complex manner (Fischer *et al.,* 1993).

With *E. coli* GroEL it is not possible to refold firefly luciferase, although a tight binary complex can be formed. In contrast, efficient refolding of luciferase is obtained with the eukaryotic cytosolic chaperonin TRiC (Frydman *et al.,* 1992). One can speculate that chaperones and their substrate proteins have coevolved, thereby ensuring the efficient binding and release of compact folding intermediates while preventing the binding of native proteins.

A. Stabilization of Molten Globule State

A number of studies have attempted to define the conformation of the substrate protein while bound by the chaperonin. The binding of substrate protein to GroEL is usually performed by rapidly diluting the protein from denaturant into a buffer solution containing purified chaperonin. Upon dilution from denaturant it is known that proteins form secondary structural elements on a micro- or millisecond time scale. However, the binding of substrate protein to GroEL upon manual mixing occurs on a much slower time scale (seconds), suggesting that the protein conformation recognized by the chaperonin is likely to contain secondary structure. This conclusion is also consistent with the efficient interaction of GroEL with acid-denatured Rubisco, which is known to retain a considerable amount of secondary structure (van der Vies *et al.,* 1992). Substrate proteins for which various conformational parameters have been analyzed while bound to GroEL include rhodanese (Martin *et al.,* 1991), DHFR (Martin *et al.,* 1991), Rubisco (van der Vies *et al.,* 1992), α-glucosidase (Höll-Neugebauer *et al.,* 1991), pre-β-lactamase (Zahn and Plückthun, 1991), and α-lactalbumin (Hayer-Hartl *et al.,* 1994). Common features of these chaperonin-bound proteins are high protease sensitivity, a tryptophan fluorescence emission maximum between that of the folded and unfolded states, partial exposure of tryptophan residues to solvent, and significant adsorption of the hydrophobic fluorescent dye anilinonaphthalene sulfonate (ANS). On the basis of these observations it has

been proposed that the conformation of a chaperonin-bound polypeptide resembles that of the so-called molten globule (Creighton, 1991; Martin *et al.*, 1991), a compact folding intermediate that contains secondary structure but lacks stable tertiary interactions (Bychkova and Ptitsyn, 1993). Significantly, the molten globule is an early folding intermediate that exposes hydrophobic surfaces, making it prone to aggregation. The radius of gyration of a partially folded, molten globule-like protein is thought to be 15–20% more expanded than that of the native protein (Bychkova and Ptitsyn, 1993).

A recent analysis of the interaction between GroEL and various conformers of α-lactalbumin (α-LA) further supports the conclusion that substrate proteins bind to GroEL as molten globule-like intermediates (Hayer-Hartl *et al.*, 1994). α-Lactalbumin is a 15-kDa protein that is stabilized in the native state by a calcium ion and four disulfide bonds. Selective reduction of a single disulfide bond and removal of calcium allows the remaining disulfide bonds to rearrange spontaneously, giving rise to a large number of folding intermediates (rearranged [3SS]) that possess all the properties of a molten globule (Ewbank and Creighton, 1991, 1993a, b). α-Lactalbumin binds most efficiently to GroEL when it is in this molten globule-like state. The rearranged [3SS] forms display a tryptophan fluorescence emission between that of mature and unfolded α-lactalbumin, and bind ANS. These properties do not change significantly upon binding to GroEL, although the chaperonin interacts only with a subset of the large number of possible [3SS] forms (Hayer-Hartl *et al.*, 1994). Interestingly, recent deuterium exchange experiments coupled to mass spectrometry have demonstrated that GroEL-bound α-LA contains most of the secondary structure present in the free molten globule (Robinson *et al.*, 1994). It has to be considered that the secondary structure in the molten globule state is intrinsically unstable, so that sensitive methods have to be applied to detect it. This instability would explain why a similar study of the conformation of GroEL-bound cyclophilin protein failed to detect protection of backbone amide hydrogens against exchange (Zahn *et al.*, 1994).

B. Interaction of GroEL with Different Conformational States

Most of the studies described above have analyzed the conformation of a protein that is stably bound by the chaperonin in the absence of nucleotide. To understand the dynamics of the ATP-dependent folding

reaction mediated by GroEL and GroES, it is important to bear in mind that the chaperonin has the capacity to interact with different folded states of the same polypeptide substrate. Initially, evidence for this arose from analysis of changes in the tryptophan fluorescence of GroEL-bound DHFR during folding (Martin *et al.*, 1991). On addition of Mg-ATP to the stable DHFR–GroEL complex, the intensity of tryptophan fluorescence drops within about 15 s, followed by a much slower blue-shift of the emission maximum, eventually approaching the value of the native protein at a rate corresponding to that of DHFR reactivation. Interruption of the folding process at an early time point by inhibition of ATP hydrolysis (by adding Mg^{2+} chelator), and rapidly cooling the reaction mixture, gives rise to a folding intermediate that has partially progressed toward the native state. In contrast to the initially bound form of DHFR, which is exquisitely sensitive to protease, this intermediate form of DHFR, although still bound to GroEL, is more resistant to protease digestion. Interestingly, when this intermediate GroEL–DHFR complex is returned to room temperature, a large fraction of the DHFR unfolds back into the initially bound conformation (Martin *et al.*, 1991). This observation indicates that the chaperonin has the capacity to reverse partial folding in the absence of nucleotide and may thus be able to retrieve kinetically trapped folding intermediates (Badcoe *et al.*, 1991; Peralta *et al.*, 1994). This unfolding could occur within the functional reaction cycle at the stage of nucleotide exchange (ADP to ATP), where the transient nucleotide free state has the highest affinity for the substrate protein (Fig. 2).

The binding of GroEL to a series of α-lactalbumin conformers demonstrates that GroEL is indeed able to interact with a range of conformational states of a single protein (Hayer-Hartl *et al.*, 1994). Although neither the fully unfolded nor the native state of α-LA interacts with the chaperonin, folding intermediates with two or three native disulfide bonds bind to GroEL, albeit with lower affinities than the more flexible molten globule intermediate containing nonnative disulfides. These data indicate that the chaperonin is in principle able to accompany a polypeptide along the folding pathway toward the native state.

V. ROLE OF CHAPERONINS IN OLIGOMERIC PROTEIN ASSEMBLY

The chaperonins were initially suggested to be proteins assisting in the assembly of oligomeric protein complexes. Examples include the interaction of the chloroplast chaperonin with nonassembled subunits

of Rubisco (Barraclough and Ellis, 1980), the GroEL/ES-dependent formation of active dimeric cyanobacterial Rubisco both in *E. coli* and *in vitro* (Goloubinoff *et al.*, 1989a, b), and the requirement of mitochondrial hsp60 for the assembly of various proteins imported into the organelles (Cheng *et al.*, 1989). Retrospectively, it is clear that the early identification of the *groE* genes was also based on the participation of the chaperonin system in a complex protein assembly reaction (Georgopoulos *et al.*, 1973; Sternberg, 1973). Temperature-sensitive mutations in *groEL* block the growth of phages λ, T4 and T5 (see Chapter 6). Similar defects are observed in *groES* mutants, with one interesting exception: phage T4 head assembly is not affected. A mutation in gene 31 of bacteriophage T4 allows the block imposed by the mutation in *groES* to be bypassed, raising the possibility that the gene 31 product may interact with GroEL in a GroES-like manner. This hypothesis has recently been substantiated by the demonstration that the T4-encoded gp31 is indeed a specialized cofactor, not closely related in sequence to GroES but having a very similar oligomeric structure (van der Vies *et al.*, 1994; see Chapter 6). gp31 may alter the substrate specificity of GroEL, thus increasing the efficiency of phage head assembly (Ellis, 1994b).

Are the chaperonins directly involved in the assembly of oligomeric proteins? Our present understanding suggests that the chaperonins assist primarily in the folding of monomeric polypeptide chains, with the subsequent assembly steps occurring largely spontaneously. The functions of GroEL in phage λ assembly, for example, can be attributed to the assisted folding of the λB protein rather than to the subsequent oligomerization steps (Georgopoulos *et al.*, 1994). The reason for this seems clear, considering the space constraints imposed by the chaperonin cavity. As demonstrated most convincingly for the assembly of trimeric ornithine transcarbamoylase, monomeric subunits leave the chaperonin cylinder to find their partner(s) for oligomerization (Zheng *et al.*, 1993). The nonassembled subunits are probably more or less folded and have a rather low binding affinity for the chaperonin. A weaker, transient interaction with the chaperonin or with other molecular chaperones may stabilize the monomeric subunits in a nonaggregated form until assembly is complete.

VI. CHAPERONIN FUNCTION UNDER CELLULAR STRESS CONDITIONS

Proteins may have to be chaperoned even after the polypeptide chain has emerged from the *de novo* folding pathway as a native protein. The

structure of a folded protein can be perturbed by various forms of cellular stress, including heat shock and exposure to heavy metals, oxygen radicals, ethanol, and UV light (Parsell and Lindquist, 1993). Not surprisingly, many molecular chaperones are stress proteins and their concentrations are increased under these conditions. Cells may respond to a particular form of stress by increasing the synthesis of a specific subset of stress proteins. Chaperone components with a general function in protein folding, such as DnaK, DnaJ, and GrpE or GroEL and GroES, are often among the induced stress proteins. As during *de novo* folding, these chaperones may interact with proteins that have partially unfolded under stress, preventing their aggregation and mediating their subsequent refolding. Proteins of relatively low structural stability may thus be maintained in an active form or in a soluble state competent for refolding, once more favorable conditions have been reestablished.

The prevention of protein denaturation by chaperonins under heat stress *in vivo* has been demonstrated for the mitochondrial hsp60 (Martin *et al.*, 1992). Since most mitochondrial proteins are imported from the cytosol (see Chapter 4), this experimental system can easily distinguish between effects of a chaperonin either on preexistent or on newly imported proteins. In the absence of protein import, hsp60 was shown to bind a large number of preexistent organellar proteins at temperatures between 39 and 46°C, and to prevent the thermal inactivation of DHFR at 40°C that had been imported as a fusion precursor at 25°C. In contrast, the specific function of the chaperonins under stress conditions in the cytosol is more difficult to establish. The overexpression of GroEL/GroES was shown to suppress many, though not all, temperature-sensitive mutants in various *E. coli* and *Salmonella typhimurium* genes (Fayet *et al.*, 1986; Jenkins *et al.*, 1986; VanDyk *et al.*, 1989). The temperature-sensitive phenotype of the mutant strains tested may have been due either to the structural instability of a given protein at the elevated temperature or to its inability to fold correctly. Similarly, overexpression of GroEL and GroES significantly increases the functional expression of the dimeric luciferase of *Vibrio harveyi* (Escher and Szalay, 1993). This effect occurs at 42°C but not at 37°C and is apparently due to the stabilization of the labile β subunit of luciferase. Another important example of a protein that is stabilized (and folded) by GroEL *in vivo* is the short-lived UmuC protein. UmuC is a key factor in repair after UV mutagenesis, explaining why *groE* mutants are particularly sensitive to this type of stress (Donnelly and Walker, 1989, 1992).

The mechanism by which the chaperonins stabilize proteins under heat

stress has emerged from *in vitro* studies with purified components. The protective effect of mitochondrial hsp60 on DHFR in intact mitochondria can be reproduced *in vitro* using purified GroEL and GroES. In the absence of nucleotide and GroES and at elevated temperatures, GroEL interacts preferentially with DHFR, thereby shifting the folding equilibrium toward the unfolded state and lowering the temperature for the inactivation of the enzyme significantly (Martin *et al.*, 1992). Addition of Mg-ATP after temperature downshift restores the activity of DHFR but does not allow renaturation at 40°C. Interestingly, in the presence of GroES and ATP, DHFR can be maintained in active state even at 40°C. Notably, at 25°C DHFR is able to refold from the GroEL-bound state in the absence of GroES. The requirement of GroES for folding at high temperature is consistent with the view that GroES allows partial folding to occur within the protected environment of the GroEL cavity. Heat-induced binding to the chaperonin *in vitro* has been demonstrated also for a number of different substrate proteins (Ziemienowicz *et al.*, 1993; Höll-Neugebauer *et al.*, 1991; Mendoza *et al.*, 1992; Taguchi and Yoshida, 1993; Hartman *et al.*, 1993). As shown with malate dehydrogenase and DHFR (Hartman *et al.*, 1993; Martin *et al.*, 1992) even substoichiometric amounts of chaperonin are able to exert a significant stabilizing effect. However, GroEL can only stabilize a protein if it is present during the denaturation process and cannot resolve preformed protein aggregates.

The pathway of chaperone-mediated protein folding and stabilization must be intimately related to the process of protein degradation. Proteins that are no longer able to pursue a productive folding pathway due to a mutation or chemical modification are normally degraded. A well-known exception are the so-called inclusion bodies that frequently form in bacteria upon rapid expression of large amounts of a single type of polypeptide. At least in certain cases, the high level of protein expression may simply exceed the capacity of the chaperone machinery. These protein aggregates are protected from proteolysis, whereas substrate proteins bound to chaperones, but unable to fold, are accessible to proteases. Preliminary results indicate that a functional cooperation between chaperones and cellular proteases can be reproduced *in vitro*, in the course of which a substrate protein is transferred backward from GroEL to DnaK and DnaJ and is subsequently degraded by the large *E. coli* proteases La or Clp (J. Martin and F.U. Hartl, unpublished; Sherman and Goldberg, 1992). This finding is consistent with defects in proteolysis occurring in temperature-sensitive *groE* mutants (Straus *et al.*, 1988).

VII. CONCLUDING REMARKS

The past five years have seen enormous progress in understanding the mechanistic details of how the chaperonins mediate the folding of a large number of different polypeptides. Detailed models for the GroEL/GroES reaction cycle have been proposed and are now being tested experimentally. The three-dimensional structure of GroEL, solved by the Sigler and Horwich laboratories (see Chapter 9) will enable knowledge-based mutagenesis to be carried out, providing us with more specific information on the structure–function relationships within GroEL. Important questions to be addressed in the future are whether and to what extent the chaperonins can influence the pathways or even the final outcome of a folding reaction. Has the coevolution of proteins and chaperones perhaps favored certain folding pathways over others? How far does the protein proceed along the folding pathway before it leaves the chaperonin? Another interesting direction will be the analysis of the role of the chaperonins in the cellular pathway of protein folding that involves interplay with members of other chaperone families. When does GroEL first interact with newly synthesized polypeptide chain and how do the physical conditions of the cellular environment influence the chaperonin mechanism? Given the breathtaking speed at which progress is made in the chaperonin field, the answers to at least some of these questions should be available in the near future.

ACKNOWLEDGMENTS

Work in the authors' laboratory was supported by a grant from the National Institutes of Health and by the Howard Hughes Medical Institute. J.M. was a fellow of the Dr. Mildred Scheel Stiftung für Krebsforschung, and M.M. is the recipient of a postdoctoral fellowship from the Human Science Frontiers Program.

REFERENCES

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181,** 223–230.

Badcoe, I. G., Smith, C. J., Wood, S., Halsall, D. J., Holbrook, J., Lund, P., and Clarke, A. R. (1991). Binding of a chaperonin to the folding intermediates of lactate dehydrogenase. *Biochemistry* **30,** 9195–9200.

Barraclough, R., and Ellis, R. J. (1980). Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. *Biochim. Biophys. Acta* **608,** 18–31.

Beckmann, R. P., Mizzen, L., and Welch, W. (1990). Interaction of hsp70 with newly synthesized proteins: Implications for protein folding and assembly. *Science* **248,** 850–856.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F., and Gething, M.-J. H. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* **75,** 717–728.

Bochkareva, E. S., and Girshovich, A. S. (1992). A newly synthesized protein interacts with GroES on the surface of chaperonin GroEL. *J. Biol. Chem.* **267,** 25672–25675.

Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S. (1988). Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. *Nature (London)* **336,** 254–257.

Bochkareva, E. S., Lissin, N. M., Flynn, G. C., Rothman, J. E., and Girshovich, A. S. (1992). Positive cooperativity in the functioning of molecular chaperone GroEL. *J. Biol. Chem.* **267,** 6796–6800.

Braig, K., Simon, M., Furuya, F., Hainfeld, J., and Horwich, A. L. (1993). A polypeptide bound by the chaperonin groEL is localized within a central cavity. *Proc. Natl. Acad. Sci. USA* **90,** 3978–3982.

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kiefhaber, T. (1991). GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30,** 1586–1591.

Bukau, B., and Walker, G. C. (1989). Cellular defects caused by deletion of the *Escherichia coli dnak* gene indicate roles for heat shock proteins in normal metabolism. *J. Bacteriol.* **171,** 2337–2346.

Bychkova, V. E., and Ptitsyn, O. B. (1993). The molten globule *in vitro* and *in vivo*. *Chemtracts* **4,** 133–163.

Caplan, A. J., Cyr, D., and Douglas, M. G. (1993). Eukaryotic homologues of *Escherichia coli* dnaJ: A diverse protein family that functions with HSP70 stress proteins. *Mol. Biol. Cell* **4,** 555–563.

Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* **261,** 12,414–12,419.

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A., Clarke, A. R., and Saibil, H. (1994). Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature (London)* **371,** 261–264.

Cheng, M. Y., Hartl, F. U., Martin, J., Pollock, R. A., Kalousek, F., Neupert, W., Hallberg, E. M., Hallberg, R. L., and Horwich, A. L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature (London)* **337,** 620–625.

Creighton, T. E. (1991). Molecular chaperones. Unfolding protein folding. *Nature (London)* **352,** 17–18.

Dessauer, C. W., and Bartlett, S. G. (1994). Identification of a chaperonin binding site in a chloroplast precursor protein. *J. Biol. Chem.* **269,** 19,766–19,776.

Donnelly, C. E., and Walker, G. C. (1989). groE mutants of *Escherichia coli* are defective in umuDC-dependent UV mutagenesis. *J. Bacteriol.* **171,** 6117–6125.

Donnelly, C. E., and Walker, G. C. (1992). Coexpression of UmuD with UmuC suppresses the UV mutagenesis deficiency of groE mutants. *J. Bacteriol.* **174,** 3133–3139.

Eilers, M., and Schatz, G. (1986). Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. *Nature (London)* **322,** 228–232.

Eilers, M., and Schatz, G. (1988). Protein folding and the energetics of protein translocation across biological membranes. *Cell* **52,** 481–483.

Ellis, J. (1987). Proteins as molecular chaperones. *Nature (London)* **328,** 378–379.

Ellis, R. J. (1994a). Roles of molecular chaperones in protein folding. *Curr. Op. in Struct. Biol.* **4,** 117–122.

Ellis, R. J. (1994b). Opening and closing the Anfinsen cage. *Curr. Biol.* **4,** 633–635.

Ellis, R. J., and Hemmingsen, S. M. (1989). Molecular chaperones: Proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem. Sci.* **14,** 339–342.

Ellis, R. J., and van der Vies, S. M. (1991). Molecular chaperones. *Annu. Rev. Biochem.* **60,** 321–367.

Escher, A., and Szalay, A. A. (1993). GroE-mediated folding of bacterial luciferases *in vivo*. *Mol. Gen. Genet.* **238,** 65–73.

Ewbank, J., and Creighton, T. E. (1991). The molten globule protein conformation probed by disulphide bonds. *Nature (London)* **350,** 518–520.

Ewbank, J., and Creighton, T. E. (1993a). Pathway of disulfide coupled unfolding and refolding of bovine α-lactalbumin. *Biochemistry* **32,** 3677–3693.

Ewbank, J., and Creighton, T. E. (1993b). Structural characterization of the disulfide folding intermediates of bovine α-lactalbumin. *Biochemistry* **32,** 3694–3707.

Fayet, O., Louarn, J. M., and Georgopoulos, C. (1986). Suppression of the *Escherichia coli* dnaA46 mutation by amplification of the groES and groEL genes. *Mol. Gen. Genet.* **202,** 435–445.

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171,** 1379–1385.

Fischer, H. M., Babst, M., Kaspar, T., Acuña, G., Arigoni, F., and Hennecke, H. (1993). One member of a *groESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. *EMBO J.* **12,** 2901–2912.

Flynn, G. C., Chappell, T. G., and Rothman, J. E. (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science* **245,** 385–390.

Flynn, G. C., Rohl, J., Flocco, M. T., and Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature (London)* **353,** 726–730.

Frydman, J., Nimmesgern, E., Erdjument-Bromage, H., Wall, J. S., Tempst, P., and Hartl, F. U. (1992). Function in protein folding of TRiC, a cytosolic ring-complex containing TCP1 and structurally related subunits. *EMBO J.* **11,** 4767–4778.

Frydman, J., Nimmesgern, E., Ohtsuka, K., and Hartl, F. U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature (London)* **370,** 111–117.

Gaitanaris, G. A., Vysokanov, A., Hung, S. Z., Gottesman, M., and Gragerov, A. (1994). Successive action of *Escherichia coli* chaperones *in vivo*. *Mol. Microbiol.* **14,** 861–869.

Georgopoulos, C. (1992). The emergence of the chaperone machines. *TIBS* **17,** 295–299.

Georgopoulos, C., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Georgopoulos, C., Liberek, K., Zylicz, M., and Ang, D. (1994). Properties of the heat shock proteins of *Escherichia coli* and the autoregulation of the heat shock response. *In* "The Biology of Heat Shock Proteins and Molecular Chaperones" (R. I. Morimoto, A. Tissières, and C. Georgopoulos, eds.), pp. 233–240. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Gething, M.-J., and Sambrook, J. (1992). Protein folding in the cell. *Nature (London)* **355,** 33–45.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989a). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and MgATP. *Nature (London)* **342,** 884–889.

Goloubinoff, P., Gatenby, A. A., and Lorimer, G. H. (1989b). GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature (London)* **337,** 44–47.

Gragerov, A., Zeng, L., Zhao, X., Burkholder, W., and Gottesman, M. E. (1994). Specificity of DnaK-peptide binding. *J. Mol. Biol.* **235,** 848–854.

Gray, T. E., and Fersht, A. R. (1991). Cooperativity in ATP hydrolysis by GroEL is increased by GroES. *FEBS Lett.* **292,** 254–258.

Gray, T. E., and Fersht, A. R. (1993). Refolding of barnase in the presence of GroE. *J. Mol. Biol.* **232,** 1197–1207.

Gray, T. E., Eder, J., Bycroft, M., Day, A. G., and Fersht, A. R. (1993). Refolding of barnase mutants and pro-barnase in the presence and absence of GroEL. *EMBO J.* **12,** 4145–4150.

Grimm, R., Donaldson, G. K., van der Vies, S. M., Schäfer, E., and Gatenby, A. A. (1993). Chaperonin-mediated reconstitution of the phytochrome photoreceptor. *J. Biol. Chem.* **268,** 5220–5226.

Hartl, F.-U., Hlodan, R., and Langer, T. (1994). Molecular chaperones in protein folding: The art of avoiding sticky situations. *TIBS* **19,** 20–25.

Hartman, D. J., Surin, B. P., Dixon, N. E., Hoogenraad, N. J., and Høj, P. B. (1993). Substoichiometric amounts of the molecular chaperones GroEL and GroES prevent thermal denaturation and aggregation of mammalian mitochondrial malate dehydrogenase *in vitro*. *Proc. Natl. Acad. Sci. USA* **90,** 2276–2280.

Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E., and Hartl, F. U. (1994). Conformational specificity of the chaperonin GroEL for the compact folding intermediates of α-lactalbumin. *EMBO J.* **13,** 3192–3202.

Hemmingsen, S. M., Woolford, C., van der Vies, S., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, R. J. (1988). Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature (London)* **333,** 330–334.

Hendrick, J. P., and Hartl, F. U. (1993). Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62,** 349–384.

Hendrick, J. P., Langer, T., Davis, T. A., Hartl, F. U., and Wiedmann, M. (1993). Control of folding and membrane translocation by binding of the chaperone DnaJ to nascent polypeptides. *Proc. Natl. Acad. Sci. USA* **90,** 10,216–10,220.

Hendrix, R. W. (1979). Purification and properties of GroE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* **129,** 375–392.

Höhfeld, J., and Hartl, F. U. (1994). Post-translational protein import and folding. *Curr. Op. Cell Biol.* **6,** 499–509.

Hohn, T., Hohn, B., Engel, A., Wurtz, M., and P. R. Smith (1979). Isolation and characterisation of the host protein GroE involved in bacteriophage lambda assembly. *J. Mol. Biol.* **129,** 359–373.

Höll-Neugebauer, B., Rudolph, R., Schmidt, M., and Buchner, J. (1991). Reconstitution of a heat shock effect *in vitro:* Influence of GroE on the thermal aggregation of α-glucosidase from yeast. *Biochemistry* **30,** 11,609–11,614.

Horwich, A. L., Brooks Low, K., Fenton, W. A., Hirshfield, I. N., and Furtak, K. (1993). Folding *in vivo* of bacterial cytoplasmic proteins: Role of GroEL. *Cell* **74,** 909–917.

Ishii, N., Taguchi, H., Sumi, M., and Toshida, M. (1992). Structure of holo-chaperonin studied with electron microscopy. Oligomeric cpn10 on top of two layers of cpn60 rings with two stripes each. *FEBS Lett.* **299,** 169–174.

Ishii, N., Taguchi, H., Sasabe, H., and Yoshida, M. (1994). Folding intermediate binds to the bottom of bullet-shaped holo-chaperonin and is readily accessible to antibody. *J. Mol. Biol.* **236,** 691–696.

Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbrook, J. J., Clarke, A. R., and Burston, S. G. (1993). Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: Implications for the mechanism of assisted protein folding. *Biochemistry* **32,** 2554–2563.

Jenkins, A. J., March, J. B., Oliver, I. R., and Masters, M. (1986). A DNA fragment containing the groE genes can suppress mutations in the *Escherichia coli* dnaA gene *Mol. Gen. Genet.* **202,** 446–454.

Kang, P. J., Ostermann, J., Shilling, J., Neupert, W., Craig, E. A., and Pfanner, N. (1990). Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. *Nature (London)* **348,** 137–143.

Kawata, Y., Nosaka, K., Hongo, K., Mizobata, T., and Nagai, J. (1994). Chaperonin GroE and ADP facilitate the folding of various proteins and protect against heat inactivation. *FEBS Lett.* **345,** 229–232.

Kudlicki, W., Odom, O. W., Kramer, G., and Hardesty, B. (1994). Activation and release of enzymatically inactive, full-length rhodanese that is bound to ribosomes as peptidyl-tRNA. *J. Biol. Chem.* **269,** 16549–16553.

Laminet, A. A., Ziegelhoffer, T., Georgopoulos, C., and Plückthun, A. (1990). The *Escherichia coli* heat shock proteins GroEL and GroES modulate the folding of the beta-lactamase precursor. *EMBO J.* **9,** 2315–2319.

Landry, S. J., and Gierasch, L. M. (1991). The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry* **30,** 7375–7362.

Landry, S. J., and Gierasch, L. M. (1994). Polypeptide interactions with molecular chaperones and their relationship to *in vivo* protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **23,** 645–669.

Landry, S. J., Jordan, R., McMacken, R., and Gierasch, L. M. (1992). Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature (London)* **355,** 455–457.

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., and Gierasch, L. M. (1993). Characterization of a functionally important mobile domain of GroES. *Nature (London)* **364,** 255–258.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K., and Hartl, F. U. (1992a). Successive action of molecular chaperones DnaK (hsp70), DnaJ and GroEL (hsp60) along the pathway of assisted protein folding. *Nature (London)* **356,** 683–689.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F. U. (1992b). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder which accommodates the protein substrate within its central cavity. *EMBO J.* **11,** 4757–4765.

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C., and Zylicz, M. (1991). *Escherichia coli* DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc. Natl. Acad. Sci. USA* **88,** 2874–2878.

Llorca, O., Marco, S., Carrascosa, J. L., and Valpuesta, J. M. (1994). The formation of symmetrical GroEL-GroES complexes in the presence of ATP. *FEBS Lett.* **345,** 181–186.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., and Hartl, F. U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. *Nature (London)* **352,** 36–42.

Martin, J., Horwich, A. L., and Hartl, F. U. (1992). Prevention of protein denaturation under heat stress by the chaperonin hsp60. *Science* **258,** 995–998.

Martin, J., Geromanos, S., Tempst, P., and Hartl, F.-U. (1993a). Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES. *Nature (London)* **366,** 279–282.

Martin, J., Mayhew, M., Langer, T., and Hartl, F.-U. (1993b). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature (London)* **366,** 228–233.

Martin, J., Goldie, K. N., Engel, A., and Hartl, F.U. (1994). Topology of the morphological domains of the chaperonin GroEL visualized by immuno-electron microscopy. *Biol. Chem. Hoppe Seyler* **375,** 635–639.

Mendoza, J. A., Lorimer, G. H., and Horowitz, P. M. (1992). Chaperonin cpn60 from *Escherichia coli* protects the mitochondrial enzyme rhodanese against heat inactivation and supports folding at elevated temperatures. *J. Biol. Chem.* **267,** 17,631–17,634.

Mendoza, J. A., Rogers, E., Lorimer, G. H., and Horowitz, P. A. (1991). Chaperonins facilitate the *in vitro* folding of monomeric mitochondrial rhodanese. *J. Biol. Chem.* **266,** 13,044–13,049.

Ostermann, J., Horwich, A. L., Neuport, W., and Hartl, F. U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis *Nature (London)* **341,** 125–130.

Palleros, D. R., Reid, K. L., Shi, L., Welch, W. J., and Fink, A. L. (1993). ATP-induced protein–HSP70 complex dissociation requires K^+ but not ATP hydrolysis. *Nature (London)* **365,** 664–666.

Parsell, D. A., and Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: Degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27,** 437–496.

Pelham, H. R. B. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* **46,** 959–961.

Peralta, D., Hartman, D. J., Hoogenraad, J. J., and Høj, P. B. (1994). Generation of a stable folding intermediate which can be rescued by the chaperonins GroEL and GroES. *FEBS Lett.* **339,** 45–49.

Rassow, J., Hartl, F.-U., Guiard, B., Pfanner, N., and Neupert W. (1990). Polypeptides traverse the mitochondrial envelope in an extended state. *FEBS Lett.* **275,** 190–194.

Reading, D. S., Hallberg, R. L., and Myers, A. M. (1989). Characterization of the yeast *HSP60* gene coding for a mitochondrial assembly factor. *Nature (London)* **337,** 655–659.

Robinson, C. V., Grosse, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M., and Radford, S. E. (1994). Hydrogen exchange protection in GroEL-bound α-lactalbumin detected by electrospray ionisation mass spectrometry. *Nature (London)* **372,** 646–651.

Rosenberg, H. F., Ackerman, S. J., and Tenen, D. G. (1993). Characterization of a distinct binding site for the prokaryotic chaperone, GroEL, on a human granulocyte ribonuclease. *J. Biol. Chem.* **268,** 4499–4503.

Saibil, H., Dong, Z., Wood, S., and auf der Mauer, A. (1991). Binding of chaperonins. *Nature (London)* **353,** 25–26.

Saibil, H. R., Zheng, D., Roseman, A. M., Hunter, A. S., Watson, G. M. F., Chen, S., auf der Mauer, A., O'Hara, B. P., Wood, S. P., Mann, N. H., Barnett, L. K., and Ellis, R. J., (1993). ATP induces large quarternary rearrangements in a cage-like chaperonin structure. *Curr. Biol.* **3,** 265–273.

Schmidt, M., and Buchner, J. (1992). Interaction of GroE with an all β-protein. *J. Biol. Chem.* **267,** 16,829–16,833.

Schmidt, M., Buchner, J., Todd, M. J., Lorimer, G. H., and Viitanen, P. (1994a). On the role of groES in the chaperonin-assisted folding reaction. *J. Biol. Chem.* **269,** 10,304–10,311.

Schmidt, M., Rutkat, K., Rachel, R., Pfeifer, G., Jaenicke, R., Viitanen, P., Lorimer, G., and Buchner, J. (1994b). Symmetric complexes of GroE chaperonins as part of the functional cycle. *Science* **265,** 656–659.

Schröder, H., Langer, T., Hartl, F. U., and Bukau, B. (1993). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J.* **12,** 4137–4144.

Seaton, B. L., and Vickery, L. E. (1994). A gene encoding a DnaK/hsp70 homolog in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **91,** 2066–2070.

Sherman, M., and Goldberg, A. L. (1992). Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Escherichia coli. EMBO J.* **11,** 71–77.

Staniforth, R. A., Burston, S. G., Atkinson, T., and Clarke, A. R. (1994). Affinity of chaperonin-60 for a protein substrate and its modulation by nucleotides and chaperonin-10. *Biochem. J.* **300,** 651–658.

Sternberg, N. (1973). Properties of a mutant of *Escherichia coli* defective in bacteriophage λ head formation (*groE*). *J. Mol. Biol.* **76,** 25–44.

Straus, D. B., Walter, W. A., and Gross, C. A. (1988). *Escherichia coli* heat shock gene mutants are defective in proteolysis. *Genes Dev.* **2,** 1851–1858.

Szabo, A., Langer, T., Schröder, H., Flanagan, J., Bukau, B., and Hartl, F. U. (1994). The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* hsp70 system—DnaK, DnaJ and GrpE. *Proc. Natl. Acad. Sci. USA* **91,** 10,345–10,349.

Taguchi, H., Konishi, J., Ishii, N., and Yoshida, M. (1991). A chaperonin from a thermophilic bacterium, *Thermus thermophilus,* that controls refolding of several thermophilic enzymes. *J. Biol. Chem.* **266,** 22,411–22,418.

Tagushi, H., and Yoshida, M. (1993). Chaperonin from *Thermus thermophilus* can protect several enzymes from irreversible heat denaturation by capturing denaturation intermediate. *J. Biol. Chem.* **268,** 5371–5375.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1993). Hydrolysis of adenosine 5′-triphosphate by *Escherichia coli* GroEL: Effects of GroES and potassium ion. *Biochemistry* **32,** 8560–8567.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. *Science* **265,** 659–666.

Tsalkova, T., Zardeneta, G., Kudlicki, W., Kramer, G., Horowitz, P. M., and Hardesty, B. (1993). GroEL and GroES increase the specific enzymatic activity of newly-synthesized rhodanese if present during *in vitro* transcription/translation. *Biochemistry* **32,** 3377–3380.

Ueguchi, C., Kakeda, M., Yamade, H., and Mizuno, T. (1994). An analogue of the DnaJ molecular chaperone in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **91,** 1054–1058.

van der Vies, S. M., Viitanen, P., Gatenby, A. A., Lorimer, G. H., and Jaenicke, R. (1992). Conformational states of ribulosebisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry* **31,** 3635–3644.

van der Vies, S. M., Gatenby, A. A., and Georgopoulos, C. (1994). Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature (London)* **368,** 654–656.

VanDyk, T., Gatenby, A. A., and LaRossa, R. A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature (London)* **342,** 451–453.

Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H., and Gatenby, A. A. (1991). Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. *Biochemistry* **30,** 9716–9723.

Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P., and Lorimer, G. H. (1990). Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are K^+ dependent. *Biochemistry* **29,** 5665–5671.

Viitanen, P., Gatenby, A. A., and Lorimer, G. H. (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1,** 363–369.

Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78,** 693–702.

Zahn, R., and Plückthun, A. (1992). GroE prevents the accumulation of early folding intermediates of pre-β-lactamase without changing the folding pathway. *Biochemistry* **31,** 3249–3255.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K., and Plückthun, A. (1994). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature (London)* **368,** 261–265.

Zheng, X., Rosenberg, L. E., Kalousek, F., and Fenton, W. A. (1993). GroEL, GroES, and ATP-dependent folding and spontaneous assembly of ornithine transcarbamylase. *J. Biol. Chem.* **268,** 7489–7493.

Ziemienowicz, A., Skowyra, D., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C., and Zylicz, M. (1993). Both the *Escherichia coli* chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE, can reactivate heat-treated RNA polymerase. *J. Biol. Chem.* **268,** 25,425–25,431.

Zimmerman, S. B., and Trach, S. O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222,** 599–620.

9

Chaperonin Structure and Conformational Changes

HELEN R. SAIBIL

Department of Crystallography
Birkbeck College London
London WC1E 7HX, United Kingdom

I. INTRODUCTION: METHODS FOR STRUCTURAL STUDIES OF CHAPERONINS

The large size and sevenfold symmetry of the GroE chaperonins present special challenges for structural work. For X-ray crystallography, it is a major achievement to crystallize and solve a structure of this size; this has been achieved by Braig *et al.* (1994). On the other hand, low-resolution studies by electron microscopy (EM) are facilitated by the high and odd-numbered symmetry, as well as the large size and distinctive features of the chaperonin 60 (cpn60) structure. The two approaches can provide complementary information, since there are large conformational changes and transient complexes that can be captured by cryoelectron microscopy (Saibil, 1994), which provides low-resolution density envelopes into which atomic structure data can be fitted and manipulated. This combination of X-ray and cryo-EM data has been very effective in other areas, notably for actomyosin and virus–receptor complexes, and it has great potential for cpn60. Figure 1 shows a diagram of the cpn60 14-mer, with the subunits shown in outline, based on the crystal structure.

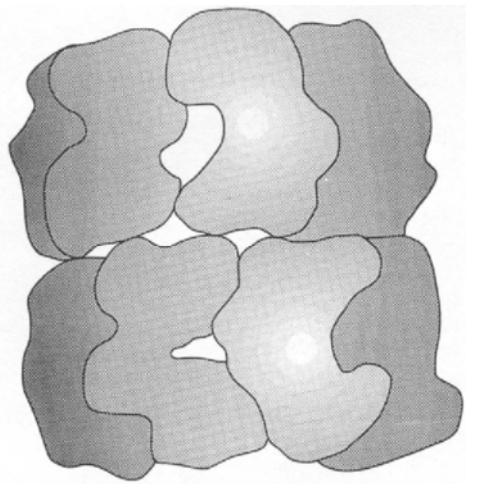


Fig. 1. The cpn60 14-mer. Subunit outlines based on the X-ray crystal structure of GroEL. Diagram adapted from Braig *et al.* (1994).

A. Electron Microscopy

The EM methods for sample preparation and imaging will be briefly explained in the following sections. Very small amounts of protein are needed for EM study. Unlike the situation in optical microscopy, the depth of focus in transmission EM is much greater than the specimen thickness, so that the observed image is a projection through the whole structure. In order to get three-dimensional (3D) information, images must be obtained of the specimen in different orientations so that the different views can be combined into a 3D density map. This is done by recording images with different tilt angles, and/or by taking advantage of molecular symmetry to generate other views. Transmission EM images are two-dimensional (2D) projections of 3D structures, and 3D reconstruction can be done by computed tomography, if enough views at different angles are available.

1. Negative Staining

In routine transmission EM, samples are imaged under high vacuum and therefore must be dehydrated. Protein structures can be imaged by allowing them to dry in a thin film of heavy metal stain, effectively encasing them in an electron-dense medium. The image is formed by the stain, which reveals the surface shape of the protein by its exclusion. Although crude and simple to apply, negative staining often gives quite a good idea of the shape and symmetry of a protein structure to about 20 Å resolution. Its main drawbacks are variable distortion due to dehydration and lack of internal structure. Drying usually causes flattening of the 3D structure, which poses particular problems for 3D reconstruction from negative stain images.

2. Cryoelectron Microscopy and Three-Dimensional Reconstruction

A technique increasingly used by structural biologists is cryo-EM. The aqueous protein sample is applied to an EM grid, but instead of staining and dehydration, a thin layer of suspension is very rapidly cooled by plunging into a cold liquid (typically ethane, cooled to liquid nitrogen temperature). This cooling traps the protein in the native, hydrated state, embedded in vitrified water. Rapid cooling prevents the formation of ice crystals, which would otherwise damage the protein by removing

water from its surface, and the protein is preserved in the native state as long as the grid is kept below about −150°C (Dubochet *et al.*, 1988). There are very significant advantages to this approach, mainly that the native protein electron density is imaged, and is directly comparable to that determined by X-ray crystallography, even though the resolution is normally much lower, of the order of 20–30 Å in single molecule work (as opposed to 2D crystal data which can go to atomic resolution). A disadvantage of cryo-EM of unstained, frozen-hydrated proteins is the extremely low image contrast. This, combined with the need to limit the electron dose because of radiation damage, means that the signal-to-noise ratio is very low. In order to extract reliable information, it is necessary to average many copies of the protein image, by computer processing. This is essential in any case for 3D reconstruction, which requires the combination of many different views.

B. X-Ray Crystallography

Crystallization and structure determination for a large protein with sevenfold symmetry are enormous tasks. Highly purified protein must be available in large quantities to find good crystallization conditions and grow suitable crystals for data collection. Wild-type GroEL is difficult to crystallize, but Braig *et al.* (1994) found that a double mutant, Arg13 → Gly, Ala126 → Val formed good crystals. The $C222_1$ crystal form contained one-half molecule per asymmetric unit; a molecular two-fold corresponded to a lattice twofold, reducing the problem of finding the molecule to that of positioning the sevenfold axis, known to be perpendicular to the twofold. Three heavy atoms per subunit were located in the heptameric ring to yield a 6-Å map by single isomorphous replacement; then sevenfold averaging was exploited to extend the phases from 6 to 2.8 Å. Because sevenfold symmetry does not fit into a crystal lattice, the seven subunits make different packing interactions with neighboring rings, and the more deformable parts of the structure deviate from exact sevenfold symmetry. Such regions are not resolved by methods that rely on sevenfold averaging. Most of the structure, with the notable exception of the termini and a region near the middle of the sequence, was clearly resolved by this procedure. Considerable work lies ahead in the refinement of the structure and in the analysis of GroEL in complex with nucleotides, GroES, and the various possible substrate

combinations. All of these structures should illuminate the molecular mechanism of chaperonin-assisted protein folding.

II. ARRANGEMENT OF SUBUNIT DOMAINS IN CHAPERONIN 60 OLIGOMER

A. Low-Resolution Studies of Chaperonins by Electron Microscopy

The distinctive structure of cpn60 was observed in the 1970s by negative stain EM. Two views, a round one with sevenfold symmetry and a rectangular or square one with four stripes of density, were reported by Hendrix (1979) and by Hohn *et al.* (1979), when GroEL was described as a factor required for bacteriophage λ assembly. Chloroplast cpn60 was shown to have similar structure (Pushkin *et al.,* 1982), and a 3D model was proposed, but the relationship between side and end views was not understood until the work of Hutchinson *et al.* (1989). By recording highly tilted views, these authors were able to show that the sevenfold axis seen in the (round) end view was perpendicular to the stripes in the side view. This established the basic layout of cpn60: in order to explain the four stripes, the 14-mer had to be a cylinder formed of two rings of seven subunits, in which each subunit was divided into two major domains, so that each ring contributed two stripes, or layers of density. The chloroplast protein, unlike GroEL, is formed of two distinct subunit types (see Chapter 3). Figures 2a and 2b show averaged cryo-EM views of GroEL. The negative stain appearance of these views is very similar. Despite the variable flattening of the structure in negative stain, a 3D reconstruction correctly showed the two major subunit domains to be connected by a thin bridge of density on the outside of the cylindrical oligomer, forming a cage-like structure with side openings (Saibil *et al.,* 1993).

1. Comparison of GroE and t-Complex Polypeptide 1 Chaperonins

Figure 2 shows averaged, cryo-EM end and side views of GroEL (Figs. 2a and 2b, respectively). Comparison with preliminary data from human chaperonin-containing *t*-complex polypeptide 1 (CCT, see Chapter 5) indicates that although they are both double-ring

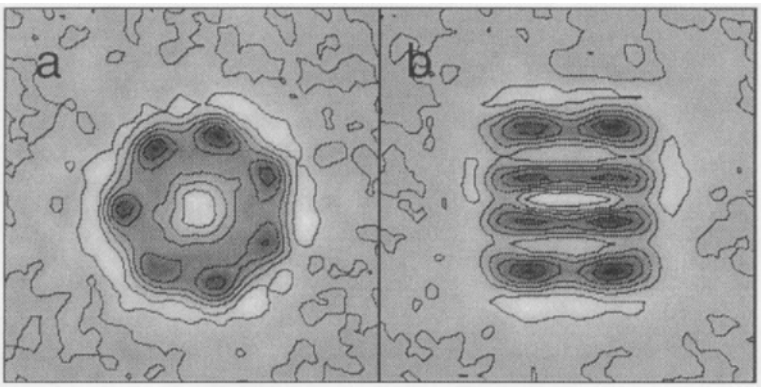


Fig. 2. Cryo-EM averaged end and side views of GroEL (a,b), from Chen *et al.* (1994).

complexes with four layers of density perpendicular to the symmetry axis, there are seven subunits forming the rings in the eubacterial chaperonin and eight in the eukaryotic cytosolic protein (A. Liou, S. Chen, H. R. Saibil, and K. R. Willison, unpublished observations). In archaebacterial members of the *t*-complex polypeptide 1 (TCP-1) subfamily of chaperonins, there are eight or nine subunits in the ring (Phipps *et al.*, 1991; Marco *et al.*, 1994b). GroEL has 14 identical subunits, whereas there are up to 9 different gene products in CCT preparations (Kubota *et al.*, 1994). These different gene products are all related in sequence, and they are likely to have very similar structures. An apparent structural difference between GroEL and CCT is the difference in relative masses of the two subunit domains, with the apical domains appearing relatively smaller than the equatorial domains in the latter case. Also, the CCT oligomer appears to have a much larger central channel. These structural features were shown by negative stain EM of the archaebacterial protein (Phipps *et al.*, 1991).

2. *Three-Dimensional Reconstructions from Cryoelectron Microscopy Images*

Cryo-electron microscopy has proved more reliable than negative stain in defining the dimensions and particularly the domain reorientations in the various cpn60 complexes. Figure 3 shows 3D reconstructions of GroEL, GroEL–ATP, and GroEL–GroES–ATP from cryo-EM.

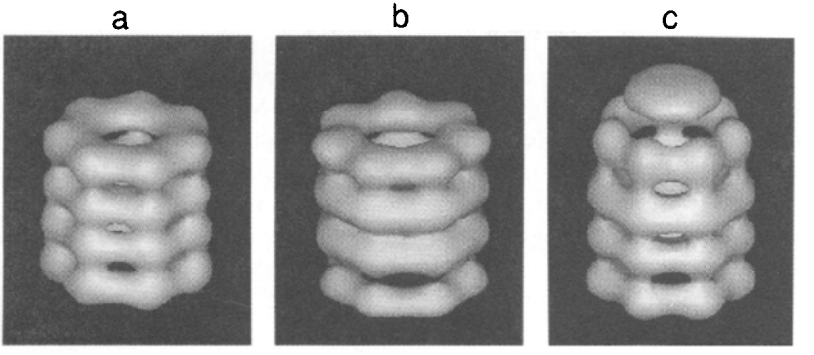


Fig. 3. Surface-rendered views of 3D reconstructions of GroEL, GroEL–ATP and GroEL–GroES–ATP from cryo-EM. The GroES ring is seen as a disk above the GroEL. Reprinted with permission from *Nature* (Chen *et al.*, 1994). Copyright 1994 Macmillan Magazines Limited.

The GroEL cylinder (a) is progressively elongated, with a small but distinct increase in length with ATP binding (b), and a very substantial elongation, particularly visible in the upper ring of GroEL in the complex with GroES (c). At the same time, the end cavities are widened. GroES is seen as a flat, disk-like object above the GroEL. With this surface contour display, the contacts between GroEL and GroES are not seen.

III. CRYSTAL STRUCTURES OF GroEL AND GroES

A. Subunit Structure and Oligomeric Contacts in GroEL

Each subunit is divided into three distinct domains (Fig. 4, color plate): the N terminus begins in the equatorial domain, and continues up through the intermediate domain (Braig *et al.*, 1994). The central part of the sequence forms the apical domain, after which the chain returns down through the intermediate domain and the crystallographically visible density terminates near the N terminus in the equatorial domain. Both N and C termini (a total of 30 residues in total are not seen at the termini,

presumably due to disorder) face the central channel, which is continuous through the 14-mer in the crystal structure. However, the missing 30 residues per subunit appear to form a central constriction to the channel as seen in cryo-EM images (Chen *et al.*, 1994). Mutation of Lys4 → Glu (sequence begining at Met-1) completely blocks oligomer assembly (Horovitz *et al.*, 1993), suggesting an important structural role for this apparently disordered region. The equatorial domain, containing N and C terminal parts of the chain, forms the backbone of the oligomeric structure. It provides most of the contacts holding the rings together (including a parallel β-strand interaction between adjacent subunits), and also the only contacts between the two rings in the 14-mer, an important route of allosteric communication. This domain is a well-defined bundle of seven α-helices.

At the top of the equatorial domain, there is a well-defined junction with the intermediate domain, a small, antiparallel domain connecting the top and bottom large domains. It comprises a pair of crossed helices and a small β-strand region that contacts the neighboring apical domain. This is followed by another junction leading to the large apical domain, which contains α and β structures and has poorly defined density in regions lining the channel, the site of residues involved in substrate and GroES binding (see Section III,C). A conspicuous feature is that much of the apical domain surface is not in close contact with neighboring domains, and that there is little barrier to its movement, both locally in some parts and in overall orientation of the domain. It is this part of the structure that deviates most from sevenfold symmetry in the crystal.

Very interesting potential hinge sites are found at the upper and lower domain junctions. These provide the possibility of large hinge movements allowing rigid body rotations (Gerstein *et al.*, 1994) of the apical and equatorial domains. In the four regions of exposed antiparallel chain, three of the four sequences contain conserved glycine residues. Figure 6a shows a schematic diagram representing the subunit domain arrangement in the oligomer, in which the potential hinge sites are marked with black dots.

B. Structure of GroES

GroES was isolated and characterized as a ring structure with approximately seven 10-kDa subunits that formed a complex with GroEL, by

Chandrasekhar *et al.* (1986) in their study of bacteriophage λ assembly. Crystallographic analysis of GroES has been completed (Hunt *et al.*, in press; Fig. 5). Each subunit consists of a β-barrel region that forms most of the contacts around the ring, and a β-hairpin pointing slightly upward and toward the center of the ring. The β-hairpin region is loosely packed, with little intersubunit contact, and forms the roof of the dome-like structure of the oligomer. Earlier NMR work showed that a mobile

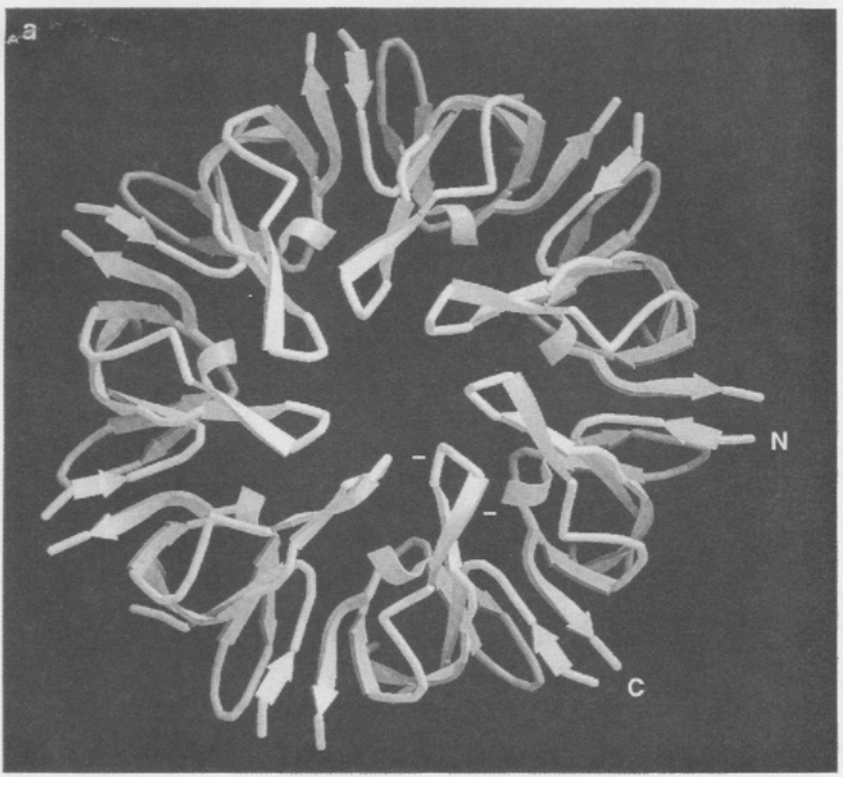


Fig. 5. Ribbon diagrams of the GroES structure viewed (a) from above and (b) from the side, showing the β-barrel structure of the subunit with a β-hairpin forming the roof of the dome-like heptamer. The N and C termini are labeled and point radially outward. In (a), the positions of two glutamic acid residues in the β-hairpin are indicated as negative charges. The broken ends of the disordered mobile loop, which is not seen in this map, are indicated by dashed arrows in (b). Images kindly provided by Dr John Hunt.

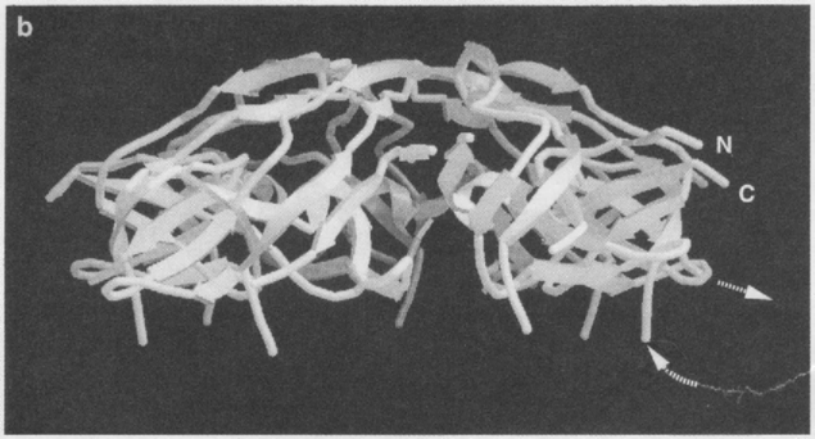


Fig. 5. *Continued*

domain in GroES became ordered upon binding to GroEL (Landry *et al.,* 1993). This region (residues 17–32) is not seen on the X-ray structure for most subunits of GroES, but is expected to form at least part of the binding contact. The visible ends of the mobile loop point downward and radially outward from the bottom of the β-barrel domain (dashed arrows, Fig. 5b). Electron microscopic reconstruction of the GroEL–GroES complex shows that the dome caps the opened apical domains of GroEL (Figs. 3 and 6). In the complex, the mobile loop is expected to be in contact with GroEL, because its accessibility to trypsin is reduced in the complex and a synthetic peptide with the loop sequence binds to GroEL (Landry *et al.,* 1993). Furthermore, all the GroES mutations originally isolated by their inability to support bacteriophage λ growth map to the mobile loop region of the sequence (Georgopoulos *et al.,* 1973).

C. Functional Sites in GroEL Defined by Mutagenesis

An extensive mutational study (Fenton *et al.,* 1994) has enabled the mapping of several important functional sites onto the atomic structure, including nucleotide, substrate, and GroES binding sites. Results from

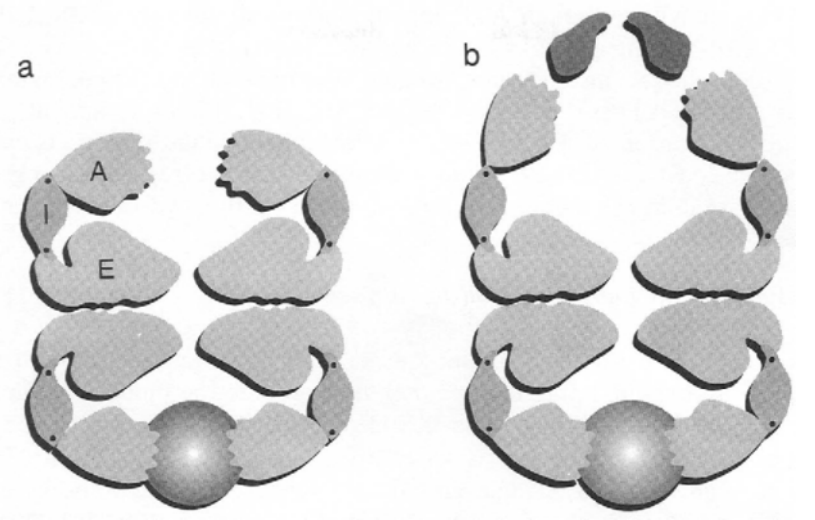


Fig. 6. Schematic diagram of the subunit arrangement in a hypothetical slice through the oligomer, showing the major functional sites. (a) GroEL, based on the crystal structure with bound substrate (shaded light to dark) as imaged by cryo-EM. A, apical domain; I, intermediate domain; E, equatorial domain. The apical domains form a ring of hand-like structures with the substrate binding sites on the fingers protruding into the central channel. Sites of potential hinge rotations are indicated by black dots. The notch in the equatorial domain represents the ATP-binding cleft. The interaction across the equatorial plane is shown as pairs of wavy surfaces. (b) The folding complex GroEL–MDH–GroES–ATP, rotating the subunit domains and adding the GroES (dark gray) and substrate (shaded) densities according to cryo EM observations.

other mutant studies (e.g., Baneyx and Gatenby, 1992; Yifrach and Horovitz, 1994) can now also be interpreted in terms of the 3D structure.

1. Nucleotide Binding Pocket

The helix bundle in the equatorial domain contains the ATP binding pocket, which is bordered by a highly conserved sequence motif containing Asp-87 and GDGTT, and lined by other stretches of highly conserved residues (Fenton *et al.*, 1994). The structure with nucleotide bound has been solved by D. C. Boisvert, J. Wang, Z. Otwinowski, A. L. Horwich, and P. B. Sigler (unpublished) and confirms that nucleotide occupies this pocket. Mutations in Asp-87 completely abolish

ATPase activity, although ATP analog binding is relatively unaffected. The ATP-binding pocket is adjacent to the lower hinge region, and mutations in the intermediate domain just beyond this hinge region also abolish ATPase activity (residues 150, 151, 152, 405, and 406), as does mutation of residue 383, near the upper hinge region. These strategic locations of key residues around the hinge regions strongly suggest that hinge movements are involved in the hydrolysis mechanism.

2. Polypeptide Substrate Binding Surface

The mutagenesis study defines a set of hydrophobic residues on the apical domain, lining the channel, that provide the binding surface for substrates, a region that is flexible in the GroEL structure. The binding surface is shown as a wavy profile on the apical domains in the diagram (Fig. 6a). Single amino acid substitutions, in some cases severe changes, of these residues (including residues 199, 203, 204, 234, 237, 259, 263, and 264) abolish binding of the substrates ornithine transcarbamylase and dihydrofolate reductase. By cryo-EM, density of the substrate malate dehydrogenase (MDH) is found in exactly the same region, held between the ends of the claw-like domains and protruding outward (Chen *et al.,* 1994). The ring of seven binding sites is easily accessible from outside, being very near the surface of the cylinder, but the inward-facing orientation may protect GroEL from self-aggregation. Mutation of residue 152, near the lower hinge region, also has a strong effect on substrate binding. The ability of monomeric cpn60, missing 78 N-terminal residues, to bind substrates and partially promote their folding is consistent with the apical location of substrate binding sites (Taguchi *et al.,* 1994).

3. GroES Binding Surface

Many of the same mutations to the apical channel-lining surface also interfere with GroES binding, suggesting that the binding sites for GroES and substrate overlap. In addition, the GroES binding surface appears to extend further over the top surface of the apical domain. Looking at the GroEL crystal structure, it would appear impossible to fit the GroES heptamer (diameter ~75 Å, Hunt *et al.,* in press) into the cavity to contact most of these binding sites. However, an explanation for this is found in the low-resolution structure of the GroEL–GroES complex

from cryo-EM (Figs. 6b, 7a, and 7c). A large rotation of at least the upper hinge region in the GroEL subunits is clearly evident, to give a 50–60° reorientation of the apical domains, placing the substrate binding surfaces adjacent to the bases of the β-barrel domains of the GroES subunits (Fig. 5b).

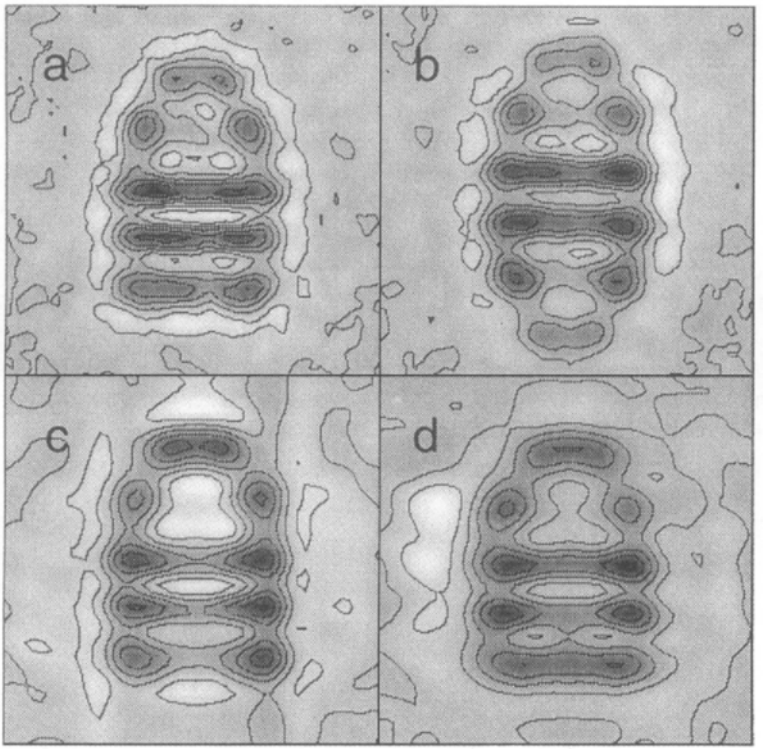


Fig. 7. (a) Side view of the GroEL-GroES complex averaged from 200 cryo EM images (Chen *et al.*, 1994). (b) Side view average of "football" complexes from cryo EM (S. Chen, A. Roseman, and H. Saibil, unpublished observations). (c) Section through a 3D reconstruction of the complex in (a) showing the large reorientation of the apical domains in contact with GroES, forming the contact with the same region as the substrate binding site. The GroEL cylinder is greatly elongated. (d) Section of a 3D reconstruction of the folding complex GroEL–MDH–GroES–ATP trapped by vitrification after 12 s of folding from cryo EM (Chen *et al.*, 1994). The MDH substrate density is found in the opposite ring to GroES.

IV. CONFORMATIONAL CHANGES IN GroEL AND ITS COMPLEXES STUDIED BY CRYOELECTRON MICROSCOPY

A. Effects of Nucleotides

Binding and hydrolysis of ATP drive the key functional cycle of cpn60: in the ADP-bound form, or without nucleotide, cpn60 has a high affinity for unfolded substrate, and in the ATP-bound form, it has lower affinity for substrate (Staniforth *et al.*, 1994). Alternation between these states has been proposed as the basis of the assisted folding mechanism (Jackson *et al.*, 1993). These two states have different conformations in cryo EM images, with a 5–10° opening of the apical domains in the presence of ATP. The superposed outlines of sections through the 3D reconstructions of GroEL and GroEL-ATP are shown in Fig. 8. The apical domains are seen to open out slightly, elongating the cylinder and widening the binding cavity. A structural change induced by ATP binding was originally observed by negative stain EM, and it appeared to cause an inward rotation of the subunits (Saibil *et al.*, 1993). The discrepancy can be partly explained by the collapse of the oligomer in negative stain, which is most evident in molecules that are noticeably flattened (larger in diameter relative to equivalent cryo-EM views). This behavior suggests that ATP binding makes the hinge regions more flexible. Nonhydrolyz-

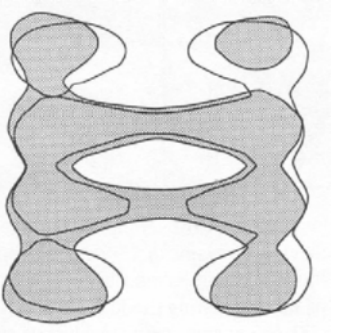


Fig. 8. Superposed outlines of sections through GroEL and GroEL–ATP (shaded) from cryo EM (Chen *et al.*, 1994), showing the slight but distinct outward rotation of the apical domains in the ATP form. These displacements have the effect of lengthening the cylinder and widening the binding cavity.

able analogs of ATP appear to cause the same type of cavity opening as ATP (A. S. Hunter, A. M. Roseman, S. P. Wood, and H. R. Saibil, unpublished observations), but ADP causes a more subtle change (Langer *et al.,* 1992). Recent work shows that ATP induces asymmetry of the GroEL 14-mer (Bochkareva and Girshovich, 1994).

B. GroEL–GroES Complexes: Hinge Rotations in GroEL

In the presence of nucleotide, a complex is formed between GroEL and GroES (Chandrasekhar *et al.,* 1986). GroES binds tightly to one ring of GroEL after ATP hydrolysis, trapping ADP in the seven sites on that ring (Todd *et al.,* 1994). Adenosine triphosphate binding by GroEL is cooperative, and this positive cooperativity is increased by GroES, suggesting that all seven sites in the opposite ring of GroEL act together. With each round of ATP hydrolysis, the GroES is released and rebinds. The biggest structural change in GroEL is seen in its complex with GroES. From negative stain EM studies of the complex it was immediately apparent that GroES binding was asymmetric and that it appeared to cause a significant expansion in GroEL (Saibil *et al.,* 1991; Ishii *et al.,* 1992; Langer *et al.,* 1992). At that time, the asymmetry was surprising, since the two rings of GroEL are originally identical. It is now known that the nucleotides required for GroES binding create asymmetry between the two rings of GroEL (Bochkareva and Girshovich, 1994; Burston *et al.,* 1995). It implies strong negative cooperativity of GroES binding, and accords well with the observation that the rate of chaperonin-assisted refolding of Rubisco *in vitro* as a function of GroES/GroEL molar ratio saturates as the GroES 7-mer concentration approaches that of the GroEL 14-mer (Azem *et al.,* 1994), although not with the interpretation of those authors. However, it was also shown that double-ended GroES binding could take place (Harris *et al.,* 1994; Azem *et al.,* 1994; Schmidt *et al.,* 1994; Todd *et al.,* 1994), and that the "football" form was favored under certain ionic conditions (high pH and Mg^{2+} seem to be important). The second GroES binds more weakly (Todd *et al.,* 1994).

Averaged side views and sections through the GroEL–GroES complex show large displacements of the apical domains, consistent with a large rotation of the apical domains in contact with GroES about the upper

hinge region (Figs. 7a and 7c). This dramatic rotation occurs at both ends of GroEL in the case of the football structures (Fig. 7b).

C. Ligand Complexes: Binary and Ternary

Substrates bound to GroEL are thought to be in a partially folded form, with native secondary structure, but with expanded and incomplete tertiary structure (the "molten globule") (Goloubinoff *et al.,* 1989; Martin *et al.,* 1991; van der Vies *et al.,* 1992). The position of substrate binding in the central channel in end projection was first suggested by negative stain EM of GroEL complexes with rhodanese and alcohol oxidase (Langer *et al.,* 1992), and by scanning transmission EM of complexes with dihydrofolate reductase labeled with colloidal gold particles (Braig *et al.,* 1993). More recently, a central location for actin binding in TCP-1 has been demonstrated by negative stain EM (Marco *et al.,* 1994a). Ishii *et al.* (1994) used antibody labeling of the substrate 3-isopropylmalate dehydrogenase to show that it could be found at the opposite end of cpn60 from the cpn 10 in the asymmetric complex from a thermophilic eubacterium. However, the substrate density is not seen in negative stain side views of GroEL complexes.

By cryo-EM, the substrate malate dehydrogenase was directly observed as additional density at one end of GroEL, held between the apical domains, in the mouth of the central channel (Chen *et al.,* 1994; diagrammed in Fig. 6a). One-sided binding was observed even in the presence of greater than a twofold molar excess of MDH, implying a negative cooperativity of substrate binding. Because the vitrification procedure traps the sample rapidly from the native state, it is possible to capture transient complexes whose lifetime is longer than the mixing, blotting, and freezing time (about 10 s with normal procedures). The transient complex GroEL–GroES–ATP–MDH was captured and imaged within 12 s of starting the folding reaction. It showed extra density, compared to the "empty" GroEL–GroES–ATP complex, in the apical cavity remote from the bound GroES (Chen *et al.,* 1994; Figs. 6b, 7c, and 7d). Samples vitrified after longer periods of folding show a progressive emptying of the binding site (A. M. Roseman, S. Chen, S. G. Burston, A. R. Clarke, and H. R. Saibil, unpublished).

V. MOLECULAR BASIS OF CHAPERONIN FUNCTION

A. Hypotheses about Mechanism of Chaperonin Action

There is general agreement that GroEL binds folding intermediates with exposed hydrophobic surfaces and suppresses aggregation, and that different substrates have varying requirements for release from cpn60–cpn10 complexes into a form committed to folding to the native state. (Gatenby and Viitanen, 1994; Hartl and Martin, 1995). It is also generally held that cycles of ATP binding and hydrolysis are central to the binding and release mechanisms (Jackson *et al.*, 1993; Todd *et al.*, 1994). The main arguments about the precise function of cpn60 and the role of cpn10 center around the conformational state of the bound polypeptide substrate and the sequence of events during folding. In particular, does the substrate fold while bound to, or trapped inside, cpn60; is it unfolded as a result of the binding interaction; and does folding take place in free solution? The requirement for cpn10 depends on the substrate and the folding conditions; ATP alone is often sufficient to induce substrate release but results in aggregation under nonpermissive folding conditions if cpn 10 is absent (Martin *et al.*, 1991; Schmidt *et al.*, 1994). Studies with the *E. coli* proteins show that cpn10 is bound and released with each round of ATP hydrolysis (Martin *et al.*, 1993; Todd *et al.*, 1994), and there is strong evidence that the substrate also goes through cycles of binding and release, since it can be trapped in a nonnative state by mutant cpn60, added after the start of the reaction, that binds it irreversibly (Weissman *et al.*, 1994). But the temporal and spatial relationships between cpn10 and substrate binding are still unclear. For example, does the substrate get enclosed in the cavity formed by cpn10 binding, or is does it only bind on the opposite ring of cpn60? The volume of the enclosed cavity is only large enough to accommodate a somewhat expanded folding intermediate of up to ~40 kDa. The requirements for folding vary greatly among different substrates, which makes it difficult to arrive at a single interpretation of the results. Broadly speaking, one school of thought holds that chaperonins are essential for preventing aggregation and that substrates fold in the protected environment (Martin *et al.*, 1991, 1993), whereas another proposes that binding is only transient and that the interaction with chaperonins serves to unfold misfolded structures (Jackson *et al.*, 1993; Zahn *et al.*, 1994), which are

then released to a fresh start to fold, in free solution, to a productive form (Todd *et al.,* 1994; Weissman *et al.,* 1994). In the latter model, they would continue to be released and rebound until a form committed to folding is reached. The time course of commitment to folding has been measured for rhodanese and glutamine synthetase by removing the GroEL by immunoprecipitation at different times during refolding (Fisher and Yuan, 1994). These experiments show that the GroE system is needed for subunit folding, but is not required during oligomeric assembly of folded monomers to produce the active, multimeric protein (Fisher, 1994). For rhodanese, the GroE system is needed throughout the folding period, until full regain of enzyme activity.

B. Mutational Probes of Allosteric Interactions

A very striking property of cpn60 is the negative cooperativity between the rings. Adenosine triphosphate, ADP, cpn10, and protein substrates all bind asymmetrically: once a ligand is bound, the two rings are in some way different.

In the crystal structure, Arg-197 in the apical domain is shown to be involved in a contact with Glu-386 in the neighboring intermediate domain. Mutation of Arg197 → Ala reduces both the positive cooperativity of the ATPase within rings and its negative cooperativity between rings, leading to the proposal, now backed up by the crystal structure, that it is involved in both intra- and intersubunit allosteric interactions (Yifrach and Horovitz, 1994). These residues are near the upper (apical) hinge, whereas the ATP binding site is next to the lower hinge.

Other mutations, around the hinge regions, and at a site at the equatorial plane that forms a contact between the two rings, influence GroES binding. Glu-461, which forms a charge pair contact between the two rings with Arg-452, is particularly remote from the apical domain, and likely to be important in negative cooperativity.

A set of particularly interesting mutations points to allosteric sites affecting GroES binding and ATP hydrolysis. In addition to the sites that are likely to be in direct contact with these ligands, mutations of residues throughout the intermediate domain interfere with GroES binding and ATPase activity. The intermediate domain with its two potential hinges is thus heavily implicated in the allosteric interactions at all stages of the reaction cycle. Not surprisingly, mutations that affect

GroES binding and ATP hydrolysis also affect polypeptide release and folding. Another line of communication, via a direct contact between apical and equatorial domains, is suggested by the effect of mutating Gly-45, a residue in the equatorial domain that comes quite close to the apical domain. Mutating Gly45 → Glu causes defective release without reducing ATPase or GroES binding. Intriguingly, other mutations with these effects are also found near the top of the apical domain (Glu-238), and toward the bottom of the equatorial domain (Asp-25). It appears that binding and release are blocked by very different mutations. Binding may be relatively straightforward to understand but release is complex, involving ATPase and GroES mechanisms. The wide distribution of sites affecting substrate release suggests that global rearrangements accompany the release step. A fascinating feature of the GroEL–GroES system is the high degree of coordinated structural rearrangement that occurs during the functional cycle; this network of cooperative interactions will certainly complicate the interpretation of site-directed mutagenesis experiments.

ACKNOWLEDGMENTS

I thank Drs. A. L. Horwich and G. Laxer for helpful comments on the manuscript and Drs. K. Braig and J. F. Hunt for providing figures.

REFERENCES

Azem, A., Kessel, M., and Goloubinoff, P. (1994). Characterization of a functional $GroEL_{14}(GroES_7)_2$ chaperonin hetero-oligomer. *Science* **265,** 653–656.

Baneyx, F., and Gatenby, A. A. (1992). A mutation in GroEL interferes with protein folding by reducing the rate of discharge of sequestered polypeptides. *J. Biol. Chem.* **267,** 11,637–11,644.

Bochkareva, E. S., and Girshovich, A. S. (1994). ATP induces non identity of two rings in chaperonin GroEL. *J. Biol. Chem.* **269,** 23,869–23,871.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å *Nature (London)* **371,** 578–586.

Braig, K., Simon, M., Furuya, F., Hainfeld, J. F., and Horwich, A. L. (1993). A polypeptide bound by the chaperonin groEL is localized within a central cavity. *Proc. Natl. Acad. Sci. USA* **90,** 3978–3982.

Burston, S. G., Ranson, N. A., and Clarke, A. R. (1995). The origins and consequences of asymmetry in the chaperonin reaction cycle. *J. Mol. Biol.* **249,** 138–152.

Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R., and Georgopoulos, C. (1986). Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* **261,** 12,414–12,419.

Chen, S., Roseman, A. M., Hunter, A., Wood, S. P., Burston, S. G., Ranson, N., Clarke, A. R., and Saibil, H. R. (1994). Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature (London)* **371,** 261–264.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A. W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21,** 129–228.

Fenton, W. A., Kashi, Y., Furtak, K., and Horwich, A. L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. *Nature (London)* **371,** 614–619.

Fisher, M. T. (1994). The effect of groES on the groEL-dependent assembly of dodecameric glutamine synthetase in the presence of ATP and ADP. *J. Biol. Chem.* **269,** 13,629–13,636.

Fisher, M. T., and Yuan, X. (1994). The rates of commitment to renaturation of rhodanese and glutamine synthetase in the presence of the GroE chaperonins. *J. Biol. Chem.* **269,** 29598–29601.

Gatenby, A. A., and Viitanen, P. V. (1994). Structural and functional aspects of chaperonin-mediated protein folding. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45,** 469–491.

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R., and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* **76,** 45–60.

Gerstein, M., Lesk, A. M., and Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry* **33,** 6739–6749.

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature (London)* **342,** 884–889.

Harris, J. R., Zahn, R., and Plückthun, A. (1994). Transmission electron microscopy of GroEL, GroES and the symmetrical GroEL/ES complex. *J. Struct. Biol.* **112,** 216–230.

Hartl, F.-U., and Martin, J. (1995). Molecular chaperones in cellular protein folding. *Curr. Opinion Struct. Biol.* **5,** 92–102.

Hendrix, R. (1979). Purification and properties of GroE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* **129,** 375–392.

Hohn, T., Hohn, B., Engel, A., Wurtz, M., and Smith, P. R. (1979). Isolation and characterization of the host protein GroE involved in bacteriophage lambda assembly. *J. Mol. Biol.* **129,** 359–373.

Horovitz, A., Bochkareva, E. S., and Girshovich, A. S. (1993). The N terminus of the molecular chaperone GroEL is a crucial structural element for its assembly. *J. Biol. Chem.* **268,** 9957–9959.

Hunt, J. F., Weaver, A. J., Landry, S., Gierasch, L., and Deisenhofer, *J. Nature* (*London*), in press.

Hutchinson, E. G., Tichelaar, W., Hofhaus, G., Weiss, H., and Leonard, K. (1989). Identification and electron microscopic analysis of a chaperonin oligomer from *Neurospora crassa* mitochondria. *EMBO J.* **8,** 1485–1490.

Ishii, N., Taguchi, H., Sumi, M., and Yoshida, M. (1992). Structure of holochaperonin studied with electron microscopy. *FEBS Lett.* **299,** 169–174.

Ishii, N., Taguchi, H., Sasabe, H., and Yoshida, M. (1994). Folding intermediate binds to the bottom of bullet-shaped holo-chaperonin and is readily accessible to antibody. *J. Mol. Biol.* **236,** 691–696.

Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbrook, J. J., Clarke, A. R., and Burston, S. G. (1993). Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: Implications for the mechanism of assisted protein folding. *Biochemistry* **32,** 2554–2563.

Kubota, H., Hynes, G., Carne, A., Ashworth, A., and Willison, K. (1994). Identification of six *Tcp-1* related genes encoding divergent subunits of the TCP-1-containing chaperonin. *Curr. Biol.* **4,** 89–99.

Landry, S. J., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, G., and Gierasch, L. M. (1993). Characterization of a functionally important mobile domain of GroES. *Nature (London)* **364,** 255–258.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F. U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J.* **11,** 4757–4765.

Marco, S., Carrascosa, J. L., and Valpuesta, J. M. (1994a). Reversible interaction of β-actin along the channel of TCP1 cytoplasmic chaperonin. *Biophys. J.* **67,** 364–368.

Marco, S., Ureña, D., Carrascosa, J. L., Waldmann, T., Peters, J., Hegerl, R., Pfeifer, G., Sack-Kongehl, H., and Baumeister, W. (1994b). The molecular chaperone TF55: Assessment of symmetry. *FEBS Lett.* **341,** 152–155.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A., and Hartl, F.-U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a "molten globule"-like intermediate. *Nature (London)* **352,** 36–42.

Martin, J., Mayhew, M., Langer, T., and Hartl, F. U. (1993). The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature (London)* **366,** 228–233.

Phipps, B. M., Hoffman, A., Stetter, K. O., and Baumeister, W. (1991). A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *EMBO J.* **10,** 1711–1722.

Pushkin, A. V., Tsuprun, V. L., Solojeva, N. A., Shubin, V. V., Evstigneeva, Z. G., and Kretovich, W. L. (1982). High molecular weight pea leaf protein similar to the groE protein of *Escherichia coli. Biochim. Biophys. Acta* **704,** 379–384.

Saibil, H. R. (1994). How chaperonins tell wrong from right. *Nature Struc. Biol.* **1,** 838–842.

Saibil, H. R., Zheng, D., Wood, S. P., and auf der Mauer, A. (1991). Binding of chaperonins. *Nature (London)* **353,** 25–26.

Saibil, H. R., Zheng, D., Roseman, A. M., Hunter, A. S., Watson, G. M. F., Chen, S., auf der Mauer, A., O'Hara, B. P., Wood, S. P., Mann, N. H., Barnett, L. K., and Ellis, R. J. (1993). ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Curr. Biol.* **3,** 265–273.

Schmidt, M., Buchner, J., Todd, M. J., Lorimer, G. H., and Viitanen, P. V. (1994). On the role of GroES in the chaperonin-assisted folding reaction. *J. Biol. Chem.* **267,** 10304–10311.

Schmidt, M., Rutkat, K., Rachel, R., Pfeiffer, G., Jaenicke, R., Viitanen, P., Lorimer, G., and Buchner, J. (1994). Symmetric complexes of GroE chaperonins as part of the functional cycle. *Science* **265,** 656–659.

Staniforth, R. A., Burston, S. G., Atkinson, T., and Clarke, A. R. (1994). Affinity of chaperonin-60 for a protein substrate and its modulation by nucleotides and chaperonin-10. *Biochem. J.* **300,** 651–658.

Taguchi, H., Makino, Y., and Yoshida, M. (1994). Monomeric chaperonin-60 and its 50-kDa fragment possess the ability to interact with non-native proteins, to suppress aggregation and to promote protein folding. *J. Biol. Chem.* **269,** 8529–8534.

Todd, M. J., Viitanen, P. V., and Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: Implications for facilitated protein folding. *Science* **265,** 659–666.

van der Vies, S. M., Viitanen, P. V., Gatenby, A. A., Lorimer, G. H., and Jaenicke, R. (1992). Conformational states of ribulose bisphosphate carboxylase and their interaction with chaperonin 60. *Biochemistry* **31,** 3635–3644.

Weissman, J. S., Kashi, Y., Fenton, W. A., and Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* **78,** 693–702.

Yifrach, O., and Horovitz, A. (1994). Two lines of allosteric communication in the oligomeric chaperonin GroEL are revealed by the single mutation Arg196 → Ala. *J. Mol. Biol.* **243,** 397–401.

Zahn, R., Spitzfaden, C., Ottiger, M., Wüthrich, K., and Plückthun, A. (1994). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature* (*London*) **368,** 261–265.

10

Immunological Aspects of Chaperonins

ANTHONY R. M. COATES

Division of Molecular Microbiology
Department of Cellular and Molecular Sciences
St. George's Hospital Medical School
London SW17 ORE, United Kingdom

Chaperonins protect cellular structures and functions and so ensure cell survival (Langer and Neupert, 1991; Martin *et al.*, 1992). The role of chaperonins does not end here. In higher animals they probably protect the whole body by stimulation of the immune system. In fact chaperonins are one of the most potent stimulators of the immune system known (reviewed in Young, 1990; Kaufmann, 1994). The immune defenses in mammals are called "innate" and "adaptive." Those defense mechanisms that are already in place are innate; examples of these are phagocytes, natural killer cells, and complement. If these systems are

overwhelmed by an invading microorganism, adaptive immunity is activated in the form of antibody and T lymphocytes. Although the adaptive system takes some time to reach its peak, it improves the efficiency of the innate response and has specific memory of prior infection with the same microbe. This memory means that the second time a particular infectious agent invades an animal the adaptive immune response will respond much faster than after the first exposure.

Chaperonins act directly on the innate defense mechanisms, particularly on phagocytes. Chaperonins also stimulate a powerful adaptive immune response, namely antibody and T lymphocytes. A description of these responses is given in the following sections of this chapter and includes chaperonin (cpn) stimulation of subsets of T cells such as $\alpha\beta$ and $\gamma\delta$ T cells. A critical analysis of the significance of the immune response to chaperonins then follows, and this analysis is divided into, first, protection against infection and cancer after immunization with chaperonins and, second, a discussion of a possible role of chaperonins in autoimmune disease. It is concluded that current ideas such as molecular mimicry do not accommodate all the available evidence. It is proposed that chaperonins operate as multiplex antigens, which means that they trigger signals from many different types of cells simultaneously.

I. INNATE IMMUNITY: DIRECT EFFECT OF CHAPERONINS ON PHAGOCYTES

Surprisingly, not only are chaperonins important immunogens for the adaptive immune system, but they also act directly on innate phagocytes (see Fig. 1). For example, the *Mycobacterium leprae* cpn60.2 (also called 65-kDa antigen) induces cytokine secretion from a human monocyte line, according to the findings of Friedland *et al.* (1993). These workers carefully excluded bacterial lipopolysaccharide from the cpn60.2 preparation, which is important because lipopolysaccharide can activate monocytes. In another laboratory Peetermans *et al.* (1994) incubated mycobacterial cpn60.2 with human monocytes and monocyte-derived macrophages; this treatment also induces cytokine secretion. It seems that cpn60.2 is a much more powerful stimulant of cytokines than, for example, bacterial lipopolysaccharide. However, major histocompatibility complex (MHC) class II proteins are not increased on the surface of

these cells and there is no enhancement of reactive oxygen intermediates, so the monocytes are not classically "activated."

Do chaperonins contribute toward innate host defenses? Chaperonin-treated monocytes secrete cytokines, namely tumor necrosis factor (TNF) and interleukin-1 (Friedland *et al.*, 1993; Peetermans *et al.*, 1994), which are thought to be important for host defenses. Tumor necrosis factor is released from the cpn60.2-treated monocytes with kinetics similar to that of whole live mycobacteria (Friedland *et al.*, 1992). Tumor necrosis factor is thought to be involved in localization of mycobacteria by the immune system into granulomas (Kindler *et al.*, 1989) and has antimycobacterial action *in vitro* (Flesch and Kaufmann, 1990). Another cytokine, interleukin-1 β, which is released from cpn60.2-treated monocytes, is important for lymphocyte-triggering differentiation and function. These observations suggest that chaperonin-treated monocytes send signals simultaneously to lymphocytes and to macrophages. So the chaperonin–phagocyte interaction may trigger an important early warning system for the adaptive immune system, as well as inducing a line of defense in the innate system. It is quite possible that in certain situations, rather than exerting a protective effect, chaperonins may damage the host. For example, TNF, which is secreted by cpn60-stimulated phagocytes, can contribute to tissue damage in infectious and noninfectious lesions (Langermans *et al.*, 1992; Cerami and Beutler, 1988). Similarly, bacterial endotoxins and other functionally related molecules such as lipopolysaccharide of gram-negative bacteria cause cytokine-related disease (Mims *et al.*, 1993) and stimulate macrophages to secrete TNF and, in the case of lipopolysaccharide, interleukin-1 as well.

However, there are a number of problems that need to be addressed, before the chaperonin–phagocyte relationship is assigned the label of biological significance: (i) Chaperonins are usually regarded as intracellular molecules, so it could be argued that it is rather artificial to add them to the outside of phagocytes. Nevertheless *M. tuberculosis* cpn10 clearly can exist extracellularly, because it is recovered from the conditioned medium of bacterial cultures (Verbon *et al.*, 1990; Orme *et al.*, 1992). Presumably this cpn10 either leaks from dead cells, or is secreted by live bacteria. *Mycobacterium tuberculosis* is found extracellularly in infected tissue (Lucas, 1988), so it is likely that, in tuberculosis, bacterial cpn10 comes into contact with the surface of mammalian cells. Are human chaperonins found on the surface of cells or outside cells? There is evidence that cpn60-related molecules are present on the surface of bone marrow-derived macrophages (Wand-Wurttenberger *et al.*, 1991),

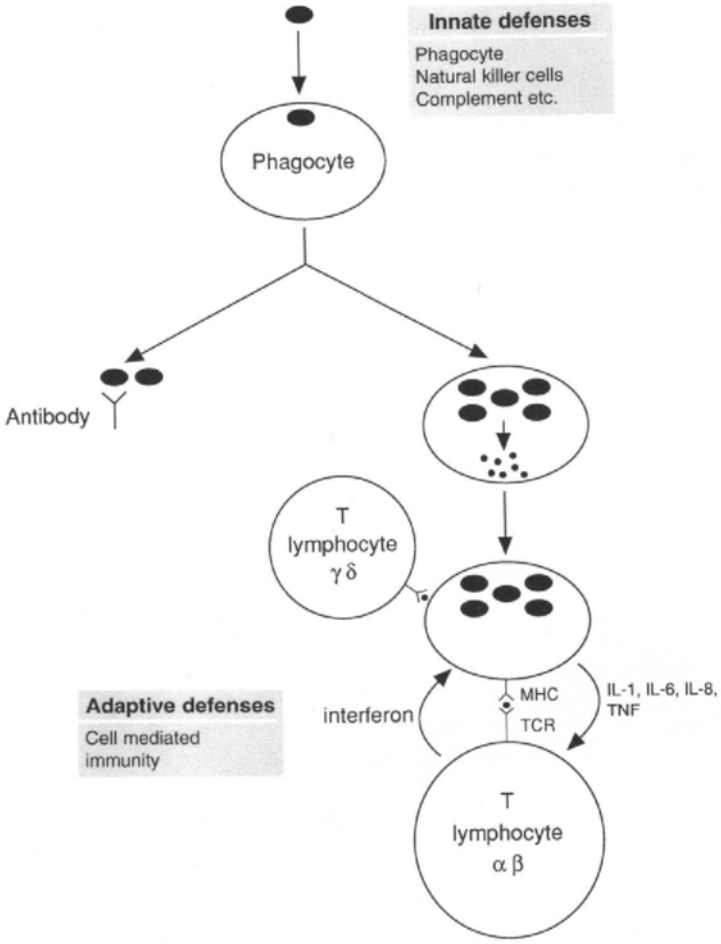


Fig. 1. A single bacterium shown at the top is engulfed by a phagocyte, which is one of the innate defenses of the immune system. If the organism survives, it proliferates locally, either intracellularly (shown here) or outside the cell. Extracellular growth of the organisms and/or release of toxins results in the production of specific antibodies by B lymphocytes. Intracellular proliferation leads to cell-mediated immunity. In most infections both antibody and cell-mediated immunity are stimulated. In the case of cell-mediated immunity, the microbe is digested within the phagocyte (in this case a macrophage), the chaperonins are fragmented into short peptides (●) and these are bound to the major histocompatibility complex (MHC) in the post-Golgi vesicle. The MHC–peptide complex moves toward the outside and decorates the surface of the cell. An $\alpha\beta$ T lymphocyte recognizes the peptide–MHC complex via its T cell receptor (TCR) and becomes activated. The activated T cell improves the defense against pathogens in a wide variety of ways

monocytic cells lines (Ferm *et al.*, 1992), and Daudi cells, but not in Raji or Epstein–Barr virus (EBV)-transformed B cells (Kaur *et al.*, 1993). Are these surface chaperonins recognized by other cells? There is no evidence for chaperonin receptors on the surface of monocytes, which may be required if extracellular chaperonins are to trigger phagocytes to release proinflammatory cytokines. On the other hand $\gamma\delta$ T lymphocytes do recognize cpn60 on the surface of human cells (Kaur *et al.*, 1993). The idea that chaperonins may have extracellular activity is important because it suggests a new biological role for these proteins. Chaperonin 10 has been located in human platelets but more importantly it may also be present in human maternal serum (Morton *et al.*, 1974; Cavanagh and Morton, 1994). If this report is confirmed in other laboratories, it is evidence for circulating chaperonin, which is clearly extracellular. Cavanagh and Morton (1994) claim that human cpn10 is identical with early pregnancy factor, which is involved in control over cell growth and development. This identification suggests that cpn10 may act rather like a hormone in stressful situations such as pregnancy. The idea that intracellular proteins may transfer between animal cells is not new. For instance, about 20% of radiolabeled glial cell proteins are transferred into the squid giant axon (Lasek *et al.*, 1974; Tytell, 1987). The relative molecular masses of these molecules range from 10,000 to more than 200,000. Intriguingly, at least two of these are heat shock proteins of 70 and 95 kDa (Tytell *et al.*, 1986; Tytell, 1987). Although chaperonins are also heat shock proteins, it is not yet known whether they are among these transferred glial proteins. It is thought that these transferred proteins are not taken into the axon by endocytosis or microphagocytosis, but are transferred by an unconventional process.

These data strongly suggest that bacterial chaperonins may come into contact with the surface of mammalian cells during an infection *in vivo,* but evidence for animal chaperonins on the surface of cells or in extracellular fluid or transferring between cells is less sure.

---

including secretion of interferons, which stimulate phagocytes to kill microbes, stimulate natural killer cells, and prevent viral replication. T lymphocytes also help cytotoxic T cells to inhibit replication of viruses and help B lymphocytes to produce antibody. Fragments of chaperonin also appear at the cell surface not bound to MHC and stimulate $\gamma\delta$ T cells. In addition, chaperonins stimulate the phagocyte to release cytokines (IL-1, IL-6, IL-8, and TNF), which stimulate the immune system by triggering activation, recruitment, and differentiation of T lymphocytes. The cytokines help cells to kill the bacterium but overstimulation in the process may result in damage to innocent tissues.

(ii) Most of the available evidence for phagocyte stimulation is centered on mycobacterial cpn60.2 molecules. Are both mycobacterial cpn60.2 and mycobacterial cpn10 involved? Work from the author's laboratory (F. Marelli, P. Mascagni, and A.R.M. Coates, unpublished observations, 1995) indicates that *M. tuberculosis* cpn10 stimulates interleukin-1 secretion from human synovial fibroblast-like cells. This observation suggests that mycobacterial cpn10 as well as cpn60.2 can stimulate nonlymphocyte cells directly (Table I). Clearly, if this is a general biological phenomenon, it should also apply to other bacterial chaperonins and even to mammalian chaperonins.

(iii) Is the chaperonin–phagocyte interaction a laboratory artifact due to contaminants? For example, might small amounts of antibody against chaperonin in the blood attach to the cell surface receptors for type 1 crystallizable fragment (Fc) of immunoglobulin and so activate the phagocyte by binding exogenous chaperonin? The phagocytes that were used by Friedland *et al.* (1993), and by Peetermans *et al.* (1994) were derived from the THP-1 human monocyte cell line and from heparinized peripheral venous blood of volunteers. Clearly, the monocytes from the volunteers might have come into contact with antichaperonin antibody, but it is much less likely that the THP-1 cell line was exposed to antichaperonin antibody, although it is possible that these type of antibodies might be present in the fetal calf serum that was used to grow the cells. Nevertheless, even if antichaperonin antibodies were absorbed to the Fc receptor, this would provide a mechanism for the action of chaperonins and would not contradict the proposed chaperonin–phagocyte interaction.

TABLE I

Cell Types Targeted by Chaperonins

Cell type	Refs.
Monocytes	Friedland *et al.* (1993); Peetermans *et al.* (1994)
Synovial fibroblast-like cells	Marelli, F., Mascagni, P., and Coates, A. R. M., unpublished data (1995)
$\alpha\beta$ T lymphocytes	Kaufmann *et al.* (1987); Mehra *et al.* (1992); Kaufmann (1994); Silva *et al.* (1993)
MHC restricted	
MHC unrestricted	
$\gamma\delta$ T lymphocytes	O'Brien *et al.* (1992); Born *et al.* (1990); Kaur *et al.* (1993)

An important and well-known contaminant of recombinant proteins is bacterial lipopolysaccharide, which probably derives from the commonly used bacterial expression systems. This contaminant can activate phagocytes. Friedland *et al.* (1993), Peetermans *et al.* (1994), and the author's group (F. Marelli, P. Mascagni, and A. R. M. Coates, unpublished data, 1995) took extraordinary steps to eliminate contaminating lipopolysaccharide from their chaperonin preparations. These steps included the removal of lipopolysaccharide by three passages through a Detoxigel column or polymyxin-coated beads. Measurement of lipopolysaccharide in these preparations reveals levels of the contaminant that are far too low to activate phagocytes. The level of lipopolysaccharide in the chaperonin preparations is less than 1 ng/ml, yet the chaperonin induces a similar TNF mRNA response as 10μg/ml of lipopolysaccharide. Heating the preparation to 65°C (which destroys chaperonin protein but not the more heat-stable lipopolysaccharide) removes TNF mRNA induction by chaperonin but not by lipopolysaccharide.

Another possible contaminant might be peptides from the bacterial expression systems which, perhaps bind to the chaperonins and act on cells in this form. This problem was addressed by the chemical synthesis of the cpn10 that had no contact with any recombinant expression cells and contains negligible levels of bacterial lipopolysaccharide, yet stimulates fibroblast-like synovial cells (F. Marelli, P. Mascagni, and A. R. M. Coates, unpublished work, 1995). These data indicate that contaminants are unlikely to be the cause of the chaperonin–phagocyte interactions. Rather, the cpn10 itself stimulates the phagocyte.

II. ADAPTIVE IMMUNITY: NATURE OF IMMUNE RESPONSE TO CHAPERONINS

A. Antibodies

If an experimental animal is immunized with bacteria, a powerful immune response against chaperonins occurs. This response is characterized by, first, anti-chaperonin antibodies produced from B lymphocytes and, second, the appearance of anti-chaperonin T lymphocytes. Anti-chaperonin antibodies were observed many years ago in sera of rabbits that were injected with *Escherichia coli* and *Pseudomonas aeruginosa,*

although at the time the workers called the cross-reactive protein "common antigen" (Kaiser, 1975; Hoiby, 1975; Sompolinsky *et al.*, 1980). Numerous monoclonal antibodies were produced against mycobacterial chaperonins by immunization of mice with whole bacteria, sonicates, or lysates (Engers *et al.*, 1985, 1986). The first set of these monoclonal antibodies contained the TB78 antibody (Coates *et al.*, 1981), which was used by Young and Davies (1983) to clone the well-known mycobacterial 65-kDa antigen (Shinnick, 1987), now called cpn60.2 (Coates *et al.*, 1993). Interestingly, some bacterial species contain two or more homologous cpn60 genes (Mazodier *et al.*, 1991; Rinke de Wit *et al.*, 1992; Kong *et al.*, 1993) and so, for convenience, these are called cpn60-1 or -2 and so on (Coates *et al.*, 1993). Anti-chaperonin 10 polyclonal antibodies and monoclonal antibody are also found in the serum of animals immunized with mycobacteria (Minden *et al.*, 1984). Humans who suffer from a wide variety of bacterial diseases also produce antibodies against bacterial cpn60s [reviewed by Young (1990)] and cpn10 (Coates *et al.*, 1989; Verbon *et al.*, 1990).

What is the significance of antibodies against chaperonins? Are they protective against invading microorganisms? Are they of no importance or are they harmful to the host? *In vivo*, antibodies are unable to bind effectively to antigens, such as chaperonins, which are located within the cell away from the cell surface. Nevertheless antichaperonin antibodies do not protect animals against infection with, for example, *E. coli* (Kaiser, 1975), or patients with lepromatous leprosy (Ilangumaran *et al.*, 1994). The most likely scenario is that antichaperonin antibodies are incidental to protective immune processes. As far as the potential host-damaging effect of antichaperonin antibodies are concerned, they are produced in a wide range of autoimmune disease (de Graeff-Meeder *et al.*, 1993; Xu *et al.*, 1993; Yokota *et al.*, 1993; Rambukkana *et al.*, 1993). These antibodies can bind to chaperonins that are located on the surface of mammalian cells (Wand-Wurttenberger *et al.*, 1991; Xu *et al.*, 1993), but they are not likely to be directly involved in harming the host because they are found in healthy people (Xu *et al.*, 1993; Coates *et al.*, 1989).

B. T Lymphocytes

A remarkable feature of the T lymphocyte response against chaperonins is its dominance over the response to other antigens. In bacterial

infection the immune system picks out chaperonins from hundreds of other proteins and mounts a major response to chaperonins but not to most of the other proteins. For example, in a leprosy patient one in three T lymphocytes that react to *M. leprae* respond to mycobacterial cpn10 (Mehra *et al.*, 1992). In *M. tuberculosis* immunized mice, one in five of the bacteria-reactive T lymphocytes respond to mycobacterial cpn60 (Kaufmann *et al.*, 1987). No doubt this observation explains why antichaperonin antibodies are also found with such regularity in infections, because helper T lymphocytes assist B lymphocytes to produce antibody. But why does the adaptive immune system choose chaperonins rather than other antigens? After all, bacterial chaperonins have a close sequence similarity to mammalian chaperonins, and so it might be dangerous for a host to invest a high proportion of its immune energy into sets of T cells that may cross-react with self-chaperonin. This response might lead to autoimmune disease, which could damage the host. Despite these dangers, the immune system operates a strong preference for chaperonins.

1. Multiplex Hypothesis

It is quite possible that immune preference for chaperonins reflects the influence of the innate immune system upon the adaptive immune system, in addition to a direct effect on T cells (see later in this chapter). This is the multiplex hypothesis, which focuses on the involvement of chaperonins in triggering simultaneous transmission of several messages (e.g., cytokines) along a channel of communication (the immune system). Chaperonins, in the absence of T lymphocytes, induce mononuclear cells and macrophages to secrete cytokines, in particular, interleukin-1 (Peetermans *et al.*, 1994). Interleukin-1, when administered simultaneously with antigen, will enhance the consequent immune response (Mims *et al.*, 1993); this is an adjuvant effect. The classical adjuvant is Freund's complete adjuvant, in which killed *M. tuberculosis* is emulsified with water in an oil vehicle. It is thought that adjuvants, such as Freund's complete adjuvant, act in part by inducing cytokines such as interleukin-1, which, in turn, trigger lymphocyte differentiation and function. Although chaperonins do not seem to have classical adjuvant activity, they can act as powerful carrier molecules, which means that they trigger B cells into a response, usually because they stimulate T helper cells. Lussow *et al.* (1991) demonstrated this stimulatory effect. *Mycobacterium bovis* cpn60 was linked to the repetitive malaria synthetic peptide

(NANP) 40 and injected into *M. bovis* BCG preimmunized mice. This treatment led to the production of anti-NANP immunoglobulin G (IgG) antibodies in the absence of adjuvant. So, chaperonins can act as carriers that can replace adjuvant *in vivo.*

The multiplex hypothesis is designed to embrace, as well as the well-known MHC-restricted chaperonin actions, all non-MHC-restricted effects of chaperonins that are beginning to emerge. These include (see Table I) MHC-unrestricted (Silva *et al.,* 1993) $\alpha\beta$ T lymphocyte targets, $\gamma\delta$ T cells (O'Brien *et al.,* 1992; Born *et al.,* 1990; Kaur *et al.,* 1993), monocytes (Friedland *et al.,* 1993; Peetermans *et al.,* 1994), and synovial fibroblast-like cells (F. Marelli, P. Mascagni, and A. R. M. Coates, unpublished data, 1995). In addition chaperonins may be involved in the growth and development of other cells if cpn10 is indeed early pregnancy factor (Cavanagh and Morton, 1994). For example, chaperonins may be involved in the induction of apoptosis, or programmed cell death. Work by Galli *et al.* (G. Galli, P. Ghezzi, P. Mascagni, F. Marcucci, and M. Fratelli, unpublished data, 1995) suggests that *M. tuberculosis* cpn10 induces apoptosis of human T lymphocytes. This interesting finding may herald other, so far unsuspected, roles for chaperonins.

2. Molecular Mimicry Hypothesis

There is another hypothesis to explain the immune preference for chaperonins. As far as this chapter is concerned, it is called the "molecular mimicry" hypothesis. It suggests that there is an inherent bias in the immune repertoire that favors recognition of conserved "self-like" proteins (Cohen and Young, 1992). This bias is introduced during development, and so is "hard-wired" into the immune system. If this were true, there should be a high frequency of responses to chaperonins in naive, unexposed lymphocyte populations. Fischer *et al.* (1992) have shown that there is indeed a high frequency of responses to mycobacterial cpn60 in human cord blood from newborn infants. Responses to other mycobacterial antigens are not elevated until later exposure. Similarly, O'Brien *et al.* (1992) demonstrated that a large subset of $\gamma\delta$ T lymphocyte cells that recognize cpn60 are present in the thymus. These reports support the molecular mimicry hypothesis. However, in inflammatory arthritis, T cells that react with mycobacterial cpn60 do not recognize human cpn60 (Life *et al.,* 1993). This observation is evidence, although not proof, against the molecular mimicry hypothesis. Chaperonins di-

rectly stimulate mononuclear cells that have no T cell receptor (Peetermans *et al.,* 1994) and this observation also does not fit the molecular mimicry hypothesis. Rather it suggests that chaperonins stimulate the immune system at multiple levels simultaneously, including monocytes, macrophages, fibroblast-like cells, perhaps other types of cells, and T cells.

3. Other Hypotheses

It is possible that the immune preference for chaperonins is due to exposure early in life to environmental bacteria that share epitopes with mycobacterial chaperonins. It is difficult to see how this can be true, because there is a high frequency of responses to mycobacterial cpn60 in the cord blood of newborn infants who have not been exposed to environmental bacteria. Another possibility is that bacteria synthesize large amounts of chaperonins, particularly when they are stressed, and this accounts for the dominance of chaperonins over other antigens. Although it is well known that stressed bacteria synthesize chaperonins in culture, the relative proportions of bacterial chaperonin versus other proteins produced during infection, for example, in active tuberculosis, still requires further investigation.

a. γδ Cells. T lymphocytes bear the T cell antigen receptor (TCR; see Fig. 1). The majority of T cells recognize antigen through TCR which consists of two chains called α and β. These cells require peptide fragments to be presented to them bound to the MHC, which resides on the surface of antigen-presenting cells such as macrophages. A further, smaller subset of T cells (1.5% of the total T cell population) recognizes antigen via a TCR that has γ and δ chains (Brenner *et al.,* 1988). Interestingly, 10–20% of $\gamma\delta$ T cells in the adult mouse react to cpn60 (O'Brien *et al.,* 1992) and to the 17 amino acid synthetic peptide corresponding to amino acids 180–196 of the *M. leprae* cpn60 sequence. A synthetic peptide representing the equivalent region of murine cpn60 also induces responses of $\gamma\delta$ T cell hybridomas, although most react weakly when compared to responses to the mycobacterial peptide (Born *et al.,* 1990). Furthermore, human peripheral $\gamma\delta$ T cells recognize cpn60 on the surface of human cells, and this recognition can be blocked by a monoclonal antibody against cpn60 (Kaur *et al.,* 1993). Do changes in the expression of self-cpn60 act as a signal for the detection and elimination of abnormal

cells such as those that are infected or malignant (Young and Elliott, 1989)? Although self-cpn60-related protein is constitutively expressed on the surface of human cells (Jarjour *et al.*, 1990), interferon-γ stimulates the expression of a surface-localized form of cpn60 on murine bone marrow macrophages (Kaufmann *et al.*, 1990). In areas of inflammation, such as multiple sclerosis brain lesions, there is colocalization of $\gamma\delta$ T cells and cpn60-expressing oligodendrocytes (Selmaj *et al.*, 1991). In addition, heat shock proteins and heat shock can regulate the expression of the murine class 1b molecules that may be the antigen-presenting molecules for $\gamma\delta$ T cells (Imani and Soloski, 1991). This evidence supports the idea that chaperonins are upregulated in stressful situations such as inflammation. The possibility that $\gamma\delta$ T cells eliminate cpn60-expressing cells is supported by the finding that cpn60-related surface protein is recognized by T cell clones that specifically lyse cpn60-expressing Daudi cells (Fisch *et al.*, 1990).

However, the role of $\gamma\delta$ T cells remains controversial. Despite the use of monoclonal antibodies that are highly specific for mammalian cpn60 (Kaufmann, 1994) and that immunoprecipitate surface molecules of the expected size, extensive microsequencing of these proteins is required to be certain that they are expressed as the full-length polypeptides. Further issues that need attention are the regulation of cpn60 expression in intact animals and whether it is the self or the microbial peptide that is recognized. On a wider front, the relative importance of chaperonins in relation to other molecules such as protease-resistant ligands, which also stimulate $\gamma\delta$ T cells (Pfeffer *et al.*, 1990), needs further study. Finally, there is the question of the mechanism of action of the cpn60 peptide that stimulates $\gamma\delta$ T cells. It is tempting to compare its action with that of superantigen on $\alpha\beta$ cells, but there are several differences between superantigen and cpn60. For instance, $\alpha\beta$ superantigen responses require MHC class II-presenting cells, whereas the cpn60 peptide does not require class II molecules for its action on $\gamma\delta$ T cells. It is also a short peptide of cpn60 that is active, whereas superantigen needs to be in its native, globular form. In addition, the cpn60 peptide requires Vγ and Vδ chains, but superantigen does not operate by binding to these chains.

b. αβ T Lymphocytes. This type of lymphocyte forms the majority of the T lymphocyte population. These lymphocytes recognize antigens bound to MHC on the surface of antigen-presenting cells (see Fig. 1).

There are two types of $\alpha\beta$ T cell, the helper T cell and the cytotoxic T lymphocyte. Helper T cell stimulation by antigen bound to class II MHC leads to activation of B lymphocytes to produce antibody, and priming of macrophages for microbicidal action. Cytotoxic T cells lyse cells that present antigen usually bound to class I MHC. Cytotoxic T cells are involved in the elimination of malignant or infected cells.

In animals that have been immunized with *Mycobacterium* spp. both cpn60 (Kaufmann *et al.,* 1987; Lamb *et al.,* 1988) and cpn10 (Minden *et al.,* 1984) are recognized by $\alpha\beta$ T cells. Healthy humans and patients with mycobacterial disease also react to cpn60 (Ilangumaran *et al.,* 1994; Emmrich *et al.,* 1986; Oftung *et al.,* 1987; Adams *et al.,* 1990) and to cpn10 (Mehra *et al.,* 1992) with antigen-specific helper and cytotoxic T cell responses. In the author's laboratory, healthy human peripheral blood mononuclear cells are found to respond, quite weakly, to *M. tuberculosis* cpn10 peptides (Fig. 2). These responses are MHC restricted (data not shown). This and previous work, cited above, suggest that there are low numbers of antichaperonin-specific lymphocytes circulating in the blood, even in healthy people.

Do human T cells recognize human chaperonins? The answer is unequivocally affirmative (reviewed in Kaufmann, 1994). For example, T cells can react with epitopes of chaperonin that are shared between self and non-self. Importantly, anti-self-chaperonin-specific responses occur in healthy adult individuals, as well as those with disease. This observation means that self-chaperonin-reactive T cells are not deleted in the thymus but are a normal component of the immune system. This rather uncomfortable idea raises the question of the function of these cells. Are they poised to defend the body against microbial invaders? There is some evidence to suggest that they are, and this evidence is discussed in Section III,A). Do they have an immune self-surveillance role whereby they recognize and lyse stressed host cells, for example, cells that are infected, malignant, or senescent? There is evidence to support the concept of immune surveillance. Murine cytotoxic T cells specific for *M. tuberculosis* cpn60 lyse stressed murine macrophages. Stress may be induced by interferon treatment or infection with cytomegalovirus (Koga *et al.,* 1989). Antisense oligonucleotide treatment of stressed macrophages interferes with chaperonin synthesis and prevents lysis by the chaperonin-specific T cells (Steinhoff *et al.,* 1994). These data show that host cpn60 is processed in stressed cells and that cpn60 epitopes are presented to T cells bound to MHC class I. Furthermore, the amino acid

AA	1	25	82	99
Seq.	AKVNIKPLEDKILVQANEAETTTAS ····· GEEYLILSARDVLAVVSK			
pept.	1	2	9	10
Donor				
AH			7.2 (0.1)	2.2 (10)
JR		3.0 (0.001)		
MO	14.0 (1)			2.2 (0.01)
FH	3.0 (10)		3.2 (1)	
TMH	7.9 (10)	2.7 (10)		

Fig. 2. Healthy human peripheral blood mononuclear cells respond to *M. tuberculosis* peptides (Marelli, F. M., Mascagni, P., Leoni, F., Pirovano, P., Al-Ghusein, H., Wilson, S. E., Gromo, G., Jolliffe, V. A., Blake, D. R., and Coates, A. R. M. (1994). Human T cell reactivity to *Mycobacterium tuberculosis* 10kDa antigen in healthy controls. Unpublished data). Peripheral blood mononuclear cells (200,000/well) were stimulated with the cpn10 peptides and cultured in 96-well round-bottom plates (Nunc, Roskilde, Denmark) at 37°C in humidified atmosphere with 5% CO_2. Culture medium consisted of RPMI 1640 (Gibco, Uxbridge, UK) supplemented with 2% human AB serum (South West Blood Transfusion Service, London, UK), 1% Nutri-Doma (Boeringher-Mannheim, Mannheim, DDR), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 m*M* glutamine. All cultures were carried out in five replicates. After 6 days, cells were pulsed with 1 μCi tritiated thymidine (Amersham International, Amersham, UK) for 6 h, and finally harvested onto glass fiber paper. Beta emission was assessed with a multidetector direct beta counter (Packard, Downer's Grove, IL). Results are expressed as stimulation index (cpm in presence of antigen : cpm in absence), which appears in the vertical columns with the peptide number at the top of each column. The proliferative responses were considered positive when the stimulation index was higher than 2. AA, amino acid; 1–99 (top line), cpn10 amino acid sequence number; seq. sequence of cpn10, which is shown in single-letter amino acid code; pept., peptides of cpn10 numbered 1 (AKV–VQA), 2 (KIL–TAS), 9 (GEE–LAV), 10 (RDV–VSK) (the overlapping sequences of the peptides are shown in the boxes); Donor, healthy donors (AH, JR, MO, FH, TMH). The optimal stimulating dose of peptide is shown in micrograms per milliliter in brackets. The stimulation index(ices) of each healthy donor is shown on the same horizontal line as the donor's initials.

sequence of these epitopes must be shared by host and bacterial cpn60s. This observation suggests that T lymphocytes recognize stressed cells by the MHC presentation of peptides that are derived from self-chaperonin.

What peptides reside in the MHC? Amino acid sequencing of peptides that are processed by MHC have revealed that a high proportion are derived from endogenous heat shock proteins (Rammensee *et al.*, 1993; Jardetzky *et al.*, 1991; Newcomb and Cresswell, 1993; Chicz *et al.*, 1993). Most of these peptides contain amino acid sequences that are located in highly conserved regions of heat shock proteins that are also present in microbial agents and in dietary protein (Kaufmann, 1994). Interestingly, peptides derived from hsp70, hsp83, and hsp90 were detected, but none from cpn60 or cpn10. Does this mean that chaperonins are not part of the immune surveillance mechanism or is their absence due to a technical anomaly? Most of the 200 or more sequenced peptides were derived from the MHCs of EBV-transformed B lymphocyte cells. It is conceivable that these B cell lines were not in a stressed state and so chaperonin was not upregulated or carried to the surface by MHC. However, it is rather odd that chaperonin sequences were not detected because if chaperonins have a major role in immune surveillance at least a small proportion of the peptides might be expected to be chaperonin fragments. Nevertheless, the general conclusion that can be drawn from these data is that conserved peptides either from self or from infectious agents or dietary protein are processed in normal healthy cells in MHC class I and class II pathways for presentation to T lymphocytes. This appears to be a very common, if not universal, phenomenon that ought to be regarded as physiological. What if the host is invaded by a microbial agent? Presumably, there is fast activation of the immune effector response and this is, perhaps, where chaperonins enter the picture. In other words infection stresses cells, which upregulate chaperonin expression, which is followed by chaperonin processing, presentation of chaperonin fragments by MHC, and lysis of these cells by antichaperonin T cells. Whether this increases the risk of the classical autoimmune diseases is unknown, but the likelihood must be very small because infectious disease is extremely common, whereas autoimmune disease is rather uncommon. Classical autoimmune diseases include rheumatoid arthritis, systemic lupus erythematosus, and Grave's disease. These diseases should be distinguished from the tissue destruction that can accompany persistent infection such as tuberculosis, which may have an autoimmune component and is commonplace. Conversely, it is possible that food-derived heat shock proteins, which enter the body through the gastrointestinal system, may cause tolerance.

III. SIGNIFICANCE OF IMMUNE RESPONSE

A. Protection against Invasion

A very important function of adaptive immunity is to improve resistance to infection. Doing so generates a specific memory of prior infection with the same microbe. Do chaperonins protect mammals against invasion with microorganisms? Although evidence is beginning to accumulate that immunization with chaperonins does induce protective adaptive immunity, the data are relatively limited and protection is not complete for every organism. For example, Noll *et al.* (1994) showed that cpn60 immunization protects mice against *Yersinia enterocolitica* infection, which causes food-associated gastroenteritis. These workers cloned and expressed the *Y. enterocolitica* cpn60. They isolated cpn60-specific $\alpha\beta$ T lymphocyte clones (CD_4^+) from both *Yersinia*-infected and *Yersinia* cpn60-immunized mice. After injection of these T cell clones into mice, significant protection was conferred against a lethal infection with *Y. enterocolitica.* Antibody does not protect against infection in organisms such as *Y. enterocolitica* that can survive inside cells, because the antibody cannot readily enter the cell.

Other attempts to induce protection *in vivo* with chaperonins have not been so successful. For example, recombinant vaccinia virus, although capable of stimulating the expression of *M. tuberculosis* chaperonin in infected cells, fails to protect mice against *M. tuberculosis* challenge (A. R. M. Coates and P. Chesters, unpublished data, 1993). Using a different strategy, Silva and Lowrie (1994) recorded protection against *M. tuberculosis* challenge. They inserted the mycobacterial cpn60.2 gene into a retroviral vector and showed cpn60.2 expression in murine monocyte-like cells. Then mice were injected with the cpn60.2-expressing macrophages; these mice were protected against subsequent *M. tuberculosis* challenge. These experiments demonstrate that, for mycobacterial infections, the method of presentation of the chaperonin is crucial if protection is to be achieved. In fact the presentation of chaperonin in the retroviral vector system is unusual; although the chaperonin recognition is T-cell receptor specific, it is, surprisingly, MHC nonrestricted (Silva *et al.,* 1993). The significance and mechanism of this interaction awaits further study.

In another intriguing set of experiments murine tumors were transfected with a retroviral vector expressing mycobacterial cpn60.2 (Lukacs

et al., 1993). The first, important finding was that the cpn60.2-expressing tumor cells lose their ability to form tumors in syngeneic mice; this effect is not T cell-dependent because no tumors are produced in mice that lack T cells. This observation suggests that cpn60.2 has a direct effect on the tumor cells, and the authors think that this might involve the protein p53 whose expression is raised in the transfected tumor cells (p53 is a protein encoded by the p53 tumor suppressor gene). The second, major finding, was that mice that had been immunized with the chaperonin-transfected cells are protected from tumors by a subsequent challenge with nontransfected tumor cells. The protective effect is mediated by cytotoxic T cells, which suggests that the presence of mycobacterial cpn60 enhances the immunological recognition of tumor antigens. Exactly how this works is not known, but it is possible that upregulation of chaperonins might play a role in enhancing antigen processing or presentation of other mitoantigens. Alternatively, peptides of other proteins may bind to chaperonins and it is these associated peptides, not chaperonins, that provide protection. It is interesting that vaccination of mice with three other heat shock proteins, gp96, hsp90, and hsp70, also protects the animals from the tumor from which the heat shock protein was obtained (Blachere *et al.*, 1993). If low-molecular-weight peptides are removed from the heat shock protein with ATP, the protein remains intact but no longer induces protection against tumor challenge (Udono and Srivastava, 1993). This effect suggests that the antigenicity of the heat shock proteins derives from the associated peptides, rather than from the heat shock protein itself. It is not known whether cpn60 protects against cancer in the same way as other heat shock proteins, but it is one possibility because chaperonins bind to other cellular proteins.

Work from another laboratory (Burns *et al.*, 1991) with the causative agent of whooping cough, *Bordetella pertussis,* shows that in this case cpn60 provides only limited protection. *Bordetella pertussis* cpn60 was purified and mice were vaccinated with the cpn60; this treatment conferred slight protection against an aerosol challenge with *B. pertussis.*

There is some evidence that $\gamma\delta$ T cells may also contribute toward protection against bacterial infection. In particular, $\gamma\delta$ T cells that react with mycobacterial cpn60.2 gather at the site of *Listeria* sp. infection in mice, and the bacteria multiply after depletion of $\gamma\delta$ T lymphocytes with monoclonal antibodies (Hiromatsu *et al.*, 1992).

The available evidence suggests that chaperonins induce protective adaptive immunity against bacterial infections and cancer. However, the

data are limited to only a few reports and it is too early to predict whether this protection applies to all infections and cancers. Furthermore, the relative importance of chaperonins in innate and adaptive protective immunity needs further study. In particular, the importance of the direct effects of chaperonins on cells other than via classical MHC-restricted presentation to T cells is a potentially interesting new area of research.

B. Autoimmune Disease

The role of chaperonins in autoimmune disease is controversial. At one end of the spectrum is the argument that although infection/immunization with chaperonin-containing organisms is universal, and healthy people have T cell responses to self-chaperonins, classical autoimmune disease is quite uncommon. So the presence of immune reactions to chaperonins may be incidental and unimportant. At the other end of the spectrum is the idea of molecular mimicry (Cohen and Young, 1991). The theory is based upon the observation that there is a high level of amino acid sequence conservation between chaperonins of microbial and mammalian origin. It proposes that during infection with a wide range of microbes chaperonin epitopes that are shared between microbes and mammals stimulate T lymphocytes. A high level of chaperonin presentation of shared chaperonin epitopes breaks tolerance to self-chaperonins and autoimmune disease develops (Kaufmann, 1990). A refinement of this idea (Cohen, 1991; Cohen and Young, 1991) suggests that lymphocyte recognition of molecular chaperones is "hard-wired" in the immune system early in its development. Chaperonins and a limited number of other self-antigens are controlled by anti-idiotype networks that constitute the hard wiring. These autoimmune lymphocyte networks are referred to as the immunological homunculus. This is a term that is borrowed from neurology and means an individual who is a dwarf but is perfectly proportioned; the term is used here with reference to the neural networks concerned with self that appear as a little person in somatotopical representation on the human cerebral cortex.

1. Animal Studies

The most convincing evidence for involvement of chaperonins in autoimmune disease is a series of experiments in animals, namely experimen-

tal insulin-dependent diabetes mellitus (IDDM) of nonobese diabetic (NOD) mice (Elias *et al.,* 1990), and adjuvant arthritis of rats (van Eden *et al.,* 1988). Chaperonin 60.2-reactive T cells either cause or suppress these model autoimmune diseases. The T cells recognize amino acids 180–188 of mycobacterial cpn60 in the case of adjuvant arthritis. Intraarticular injection of mycobacterial cpn60.2 itself induces joint inflammation in *M. tuberculosis*-sensitized rats (Winrow *et al.,* 1994b). Interestingly, immunization with mycobacterial cpn60.2 induces atherosclerosis in rabbits (Xu *et al.,* 1992) and this may turn out to be another important model of chaperonin-induced disease, although atherosclerosis is not a classical autoimmune disease.

Preimmunization of rats with mycobacterial cpn60.2 protects against adjuvant and other forms of arthritis. Immunization with cpn60.2 leads to T cell recognition of a single chaperonin epitope (256–265) that is shared by mycobacterial and rat cpn60, but this epitope is poorly recognized after immunization with adjuvant alone despite the heat-killed *M. tuberculosis* in the adjuvant (Anderton *et al.,* 1994). This epitope or others may be associated with the protection conferred by cpn60. Although a central pivotal role for cpn60 in adjuvant arthritis is possible, it is unlikely that a simple relationship between adjuvant arthritis and cpn60 exists. For example, although adjuvant arthritis in rats is induced by *M. tuberculosis* in mineral oil (the adjuvant), arthritis in DBA/1 mice can be triggered by intraperitoneal injection of the mineral oil pristane that contains no *M. tuberculosis.* Yet both types of arthritis can be modulated by preimmunization with recombinant mycobacterial cpn60 via a lymphocyte mechanism (Thompson *et al.,* 1991; Anderton *et al.,* 1994). This observation suggests that inflammation itself may be associated with an anti-cpn60 T cell response. This seems to be the case because injection of incomplete Freund's adjuvant (without *M. tuberculosis*) into the footpads of BALB/c mice induces acute inflammation and MHC class II-restricted mycobacterial cpn60 $\alpha\beta$ T cells (Anderton *et al.,* 1993). A high proportion of these T cells also responds to mouse and human cpn60. This result shows that T cells specific for microbial and mammalian cpn60s are activated during inflammatory responses, and are induced in the absence of exogenous cpn60. It seems that anti-self autoimmune lymphocyte responses are an integral part of the immune system.

2. *Human Studies*

a. T Cells with Specificity for Chaperonins. Antichaperonin T cells are found in a number of autoimmune diseases. In particular, T cells

that are specific for shared epitopes of bacterial and human chaperonins have been reported in reactive and juvenile arthritis (de Graeff-Meeder *et al.,* 1991; Hermann *et al.,* 1991). It is interesting that two CD_4^+ T cell clones from a patient with juvenile arthritis recognize cross-reactive epitopes in the 243–265 amino acid region, which is highly conserved between mycobacteria and mammals (Quayle *et al.,* 1992). This region is also recognized by T cells in adjuvant arthritis in rats (Anderton *et al.,* 1994). However, there is some doubt about whether T cells in the synovial fluid of juvenile arthritis cross-react with human cpn60. In this type of arthritis, bacterial cpn60-specific T cells do not respond to heat-shocked human cells that might be expected to express human cpn60 (Life *et al.,* 1993), although they do react to *E. coli* cpn60, which contaminates many recombinant cpn60 preparations. This observation provides evidence against the molecular mimicry hypothesis. Simple cross-reaction of T cells specific for human chaperonins and microbial chaperonins as an important cause of autoimmunity does not seem to apply to the autoimmune disease of the thyroid, Grave's disease, in which T cell clones are specific for either chaperonin or thyroid epithelial cells but not both (Trieb *et al.,* 1993). In rheumatoid arthritis, a disease characterized by destruction of the cartilage of the joint, the link between simple molecular mimicry at the T lymphocyte level and pathogenesis seems even less likely. Wilbrink *et al.,* (1993) cocultured human cartilage *in vitro* with mycobacterial cpn60-activated human synovial fluid mononuclear cells from patients with rheumatoid arthritis. Proteoglycan synthesis was measured as a marker of cartilage damage. They found that the mononuclear cells suppress cartilage proteoglycan synthesis, and that this effect is dependent on the production of interleukin-1 and TNF. They concluded that mycobacterial cpn60 can activate rheumatoid synovial mononuclear cells to suppress human cartilage synthesis and so, presumably, damage the cartilage. Recent work from the author's laboratory with rheumatoid arthritis patients (F. M. Marelli, P. Mascagni, and A. R. M. Coates, unpublished data, 1995) suggests that not only mycobacterial cpn60 activates human synovial membrane mononuclear cells, but cpn60s from other bacteria and even cpn10 also do so. Importantly, synovial fibroblast-like cells are induced to produce the proinflammatory cytokine interleukin-1 by chaperonin in the absence of synovial mononuclear cells. These data widen the cellular targets of chaperonins to include non-T cells in autoimmune disease.

b. Antibodies Specific for Chaperonins in Autoimmune Disease. These antibodies are found in a wide range of autoimmune diseases and

in healthy controls (reviewed in Winfield and Jarjour, 1991). Anti-human chaperonin antibodies have been described in children with juvenile chronic arthritis, systemic lupus erythematosus, and cystic fibrosis (de Graeff-Meeder *et al.*, 1993), in patients with Kawasaki disease (Yokota *et al.*, 1993), and in patients with atherosclerosis (Xu *et al.*, 1993), polymyositis, dermatomyositis, psoriatic arthritis, inflammatory bowel disease, epidermolysis bullosa aquisita, and bullous pemphigoid (Winfield and Jarjour, 1991). However, anti-human chaperonin autoantibodies were not found by other workers in the rheumatoid arthritis, Sjögren's syndrome, or Reiter's disease (Worthington *et al.*, 1993; Winfield and Jarjour, 1991). In addition, there are numerous reports of anti-bacterial chaperonin antibodies in many autoimmune diseases (reviewed in Winfield and Jarjour, 1991; Kaufmann, 1994).

There is no evidence from animal models that antibodies to chaperonins induce autoimmune disease, and in any case chaperonin-specific antibodies are found in healthy individuals. Furthermore, in some autoimmune diseases there are conflicting reports regarding these antibodies. Anti-chaperonin antibodies do not seem to be directly involved in the pathogenesis of these autoimmune diseases, but provide useful indications of the extent of antichaperonin activation in autoimmune disease.

c. Chaperonins Detected in Autoimmune Lesions. Increased levels of cpn60 have been recorded in rheumatoid arthritis (Winrow *et al.*, 1990; Karlsson-Parrs *et al.*, 1990), juvenile chronic arthritis (Boog *et al.*, 1992), atherosclerosis (Xu *et al.*, 1993), multiple sclerosis (Selmaj *et al.*, 1991), and ulcerative colitis (Winrow *et al.*, 1994a). However, Sharif *et al.* (1992) could find only minor differences in the distribution of cpn60 in rheumatoid arthritis synovia, and so was unable to confirm increased expression of cpn60 in rheumatoid arthritis. Interestingly, these latter workers recorded that cpn60 is expressed in normal synovium. Other reports have shown that there is no increase in cpn60 expression in systemic lupus erythematosus patients (Dhillon *et al.*, 1993) and those with Crohn's disease (Baca-Estrada *et al.*, 1994). Is increased cpn60 expression associated with colocalization of T cells? In the chronic plaques of multiple sclerosis (Selmaj *et al.*, 1991) and in chronic gastritis (Engstrand *et al.*, 1991), cells that express chaperonins are colocalized with $\gamma\delta$ T cells. These data suggest that chaperonins are present in normal tissues as well as in diseased tissue, but there is conflicting evidence about upregulation of chaperonin expression in disease. What if the opposite happens; in other words, there is too little heat shock protein? Agsteribbe *et al.* (1993) has reported an association between reduced

cpn60 in tissues (one-fifth of controls) and a fatal, systemic mitochondrial disease in a baby girl. This report suggests that insufficient cpn60 may have serious systemic consequences.

IV. CONCLUSION

The original hypothesis (Cohen, 1991; Cohen and Young, 1991) that the immune response to dominant self-antigens, such as chaperonins, is controlled by an idiotype network of lymphocytes as the immunological homunculus has resulted in a broad range of studies aimed at elucidating the role of molecular mimicry in autoimmune disease. Rather surprising results have emerged. (i) It seems that healthy people have antichaperonin lymphocyte immune responses. This observation suggests that self-reactive antichaperonin lymphocytes are physiological constituents of the immune system. Clearly, these lymphocytes, although autoreactive, do not cause autoimmune disease in healthy people. (ii) In inflammatory arthritis synovial T lymphocytes react with bacterial cpn60 but do not recognize human cpn60 (Life *et al.*, 1993). This is evidence against the molecular mimicry hypothesis. (iii) It has been assumed that chaperonin self-reactive T lymphocytes cause autoimmune disease. If this were the case, this population of anti-human chaperonin lymphocytes should be a major feature in a wide range of autoimmune disease, which is not observed. (iv) It is difficult to understand how a broad tissue distribution of chaperonins can lead to autoimmune tissue-specific disease. (v) Another puzzling finding is that out of over 200 peptides derived from the MHC cleft of antigen-presenting cells, none have been identified as cpns. This suggests that chaperonin presentation via the MHC cleft to $\alpha\beta$ T lymphocytes is an uncommon event, although chaperonin fragments could, perhaps, appear in the MHC cleft in stressed cells. (vi) The lymphocyte-based hypothesis of chaperonin action is likely to be too narrow, because chaperonins stimulate nonlymphocyte cells such as monocytes and fibroblast-like synovial cells. These observations suggest that chaperonins act at multiple levels in the immune system, namely on $\alpha\beta$ T cells, $\gamma\delta$ T cells, monocytes, macrophages, fibroblast-like cells, and probably others.

It is proposed here that chaperonins are multiplex antigens. Multiplex means a process that involves simultaneous transmission of several messages along a channel of communication (Fowler and Fowler, 1976). In

other words, chaperonins act simultaneously on lymphocytes, macrophages, fibroblasts, and other cells in response to stress, and this results in simultaneous transmission of immune messages such as cytokines to immune effector cells, which adapt to cope with the stress. This process is likely to be a very common physiological event, and may well be implicated, rather like steroids, in the maintenance of health. Whether any naturally occurring diseases are caused by the expression of too much or too little chaperonin is debatable. However, experimentally, it is possible to manipulate the chaperonin immune response to the benefit or disbenefit of the host, and so, again like steroids, it may be possible to use such an approach to benefit patients with disease.

ACKNOWLEDGMENTS

I am most grateful to M. J. Colston and A. Mitchison for helpful comments on this chapter. Also, I am very grateful to Rosemary Halms for typing the manuscript.

REFERENCES

Adams, E., Garsia, R. J., Hellquist, L., Holt, P., and Basten, A. (1990). T cell reactivity to the purified mycobacterial antigen p65 and p70 in leprosy patients and their household contacts. *Clin. Exp. Immunol.* **80,** 206–212.

Agsteribbe, E., Huckriede, A., Veenhuis, M., Ruiters, M. H. J., Niezen-Koning, K. E., Skjeldal, K., Gupta, R. S., Hallberg, R., van Diggelen, O. P., and Scholte, H. R. A fatal, systemic mitochondrial disease with decreased enzyme activities, abnormal ultrastructure of the mitochondria and deficiency of heat shock protein 60. (1993). *Biochem. Biophy. Res. Commun.* **193**(1), 146–154.

Anderton, S. W., van der Zee, R., and Goodacre, J. A. (1993). Inflammation activates self hsp60-specific T cells. *Eur. J. Immunol.* **23**(1), 33–38.

Anderton, S. M., van der Zee, R., Noordzij, A., and van Eden, W. (1994). Differential mycobacterial 65-kDa heat shock protein T cell epitope recognition after adjuvant arthritis-inducing or protective immunization protocols. *J. Immunol* **152**(7), 3656–3664.

Baca-Estrada, M. E., Gupta, R. S., Stead, R. H., and Croitoru, K. (1994). Intestinal expression and cellular immune responses to human heat-shock protein 60 in Crohn's disease. *Digestive Dis. Sci.* **39**(3), 498–506.

Blachere, N. E., Udono, H., Janetzki, S., Li, Z., Heike, M., and Srivastava, P. K. (1993). Heat shock protein vaccines against cancer. *J. Immunother.* **14,** 352–356.

Boog, C. J., De Graeff-Meeder, E. R., Lucassen, M. A., Van der Zee, R., Voorhorst-Ogink, M. M., van Kooten, P. J., Geuze, H. J., and van Eden, W. (1992). Two

monoclonal antibodies generated against human hsp60 show reactivity with synovial membranes of patients with juvenile chronic arthritis. *J. Exp. Med.* **175,** 1805–1810.

Born, W., Hall, L., Dallas, A., Boymel, J., Shinnick, T., Young, D., Brennan, P., and O'Brien, R. (1990). Recognition of a peptide by heat shock reactive γδ T lymphocytes. *Science* **249,** 67–69.

Brenner, M. B., Strominger, J. L., and Krangel, M. S. (1988). The γδ T cell receptor. *Adv. Immunol.* **43,** 133–192.

Burns, D. L., Gould-Kostka, J. L., Kessel, M., and Anciniega, J. L. (1991). Purification and immunological characterisation of GroEL-like protein from *Bordetella pertussis*. *Infect. Immun.* **59,** 1417–1422.

Cavanagh, A. C., and Morton, H. (1994). The purification of early-pregnancy factor to homogeneity from human platelets and identification as chaperonin 10. *Eur. J. Biochem.* **222,** 551–560.

Cerami, A., and Beutler, B. (1988). The role of cachectin/TNF in endotoxic shock and cachexia. *Immunol. Today* **9,** 28–31.

Chicz, R. M., Urban, R. G., Gorga, J. C., Vignali, D. A. A., Lane, W. S., and Strominger, J. L. (1993). Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J. Exp. Med.* **178,** 27–47.

Coates, A. R. M., Hewitt, J., Allen, B. W., Ivanyi, J., and Mitchison, D. A. (1981). Antigenic diversity of *Mycobacterium tuberculosis* and *Mycobacterium bovis* detected by means of monoclonal antibodies. *Lance* **ii,** 167–169.

Coates, A. R. M., Nicolai, H., Pallan, J. M., Guy, A., Chaparas, S. D., and Mitchison, D. A. (1989). The 45 kilodalton molecule of *M. tuberculosis* identified by immunoblotting and monoclonal antibodies as antigenic in patients with tuberculosis. *Brit. J. Exp. Pathol.* **70,** 215–225.

Coates, A. R. M., Shinnick, T. M., and Ellis, R. J. (1993). Chaperonin nomenclature. *Mol. Microbiol.* **8**(4), 787.

Cohen, I. R. (1991). Autoimmunity to chaperonins in the pathogenesis of arthritis and diabetes. *Annu. Rev. Immunol.* **9,** 567–589.

Cohen, I. R., and Young, D. B. (1991). Autoimmunity, microbial immunity and the immunological homunculus. *Immunol. Today* **12,** 105–110.

de Graeff-Meeder, E. R., Van der Zee, R., Rijkers, G. T., Schuurman, H. J., Kuis, W., Bijlsma, J. W. J., Zegers, B. J. M., and van Eden, W. (1991). Recognition of human 60kD heat shock protein by mononuclear cells from patients with juvenile chronic arthritis. *Lancet* **337,** 1368–1372.

de Graeff-Meeder, E. R., Rijkers, G. T., Voorhorst-Ogink, M. M., Kuis, W., Van der Zee, R., van Eden, W., and Zegers, B. J. (1993). Antibodies to human HSP60 in patients with juvenile chronic arthritis, diabetes mellitus, and cystic fibrosis. *Pediatr. Res.* **34**(4), 424–428.

Dhillon, V. B., McCallum, S., Norton, P., Twomey, B. M., Erkeller-Yuksel, F., Lydyard, P., Isenberg, D. A., and Latchman, D. S. (1993). Differential heat shock protein overexpression and its clinical relevance in systemic lupus erythematosus. *Ann. Rheumatic Dis.* **52**(6), 436–442.

Elias, D., Markovits, D., Reshef, T., Van der Zee, R., and Cohen, I. R. (1990). Induction and therapy of autoimmune diabetes in the non-obese diabetic (NOD/Lt) mouse by a 65kDa heat shock protein. *Proc. Natl. Acad. Sci. USA* **87,** 1576–1580.

Emmrich, F., Thole, J., van Embden, J., and Kaufmann, S. H. E. (1986). A recombinant 64kDa protein of *Mycobacterium bovis* BCG specifically stimulates human T4 clones reactive to mycobacterial antigens. *J. Exp. Med.* **163,** 1024–1029.

Engers, H. D., and Workshop Participants (1985). Results of a World Health Organization sponsored workshop on monoclonal antibodies to *Mycobacterium leprae*. *Infect. Immun.* **48,** 603–605.

Engers, H. D., and Workshop Participants (1986). Results of a WHO sponsored workshop to characterize antigens recognized by mycobacteria-specific monoclonal antibodies. *Infect. Immun.* **51,** 718–720.

Engstrand, L., Scheynius, A., and Pahlson, C. (1991). An increased number of γδ T cells and gastric epithelial cell expression of the groEL stress-protein homologue in *Helicobacter pylori*-associated chronic gastritis of the antrum. *Am. J. Gastroenterol.* **86,** 976–980.

Ferm, M. T., Söderström, K., Jindal, S., Grönberg, A., Ivanyi, J., Young, R., and Kiessling, R. (1992). Induction of human hsp60 expression in monocytic cell lines. *Int. Immunol.* **4,** 305–311.

Fisch, P., Malkovsky, M., Brankman, E., Sturm, E., Bolhuis, R. L. H., Priere, A., Sosman, J. A., Lam, V. A., and Sondolel, P. M. (1990). γδ T cell clones and natural killer cell clones mediate distinct patterns of non-major histocompatibility complex-restricted cytolysis. *J. Exp. Med.* **171,** 1567–1579.

Fischer, H. P., Sharrock, C. E. M., and Panayi, G. S. (1992). High frequency of cord blood lymphocytes against mycobacterial 65kDa heat-shock protein. *Eur. J. Immunol.* **22,** 1667–1669.

Flesch, I. E. A., and Kaufmann, S. H. E. (1990). Stimulation of antibacterial macrophage activities by B-cell stimulatory factor 2 (interleukin-6). *Infect. Immun.* **58,** 269–271.

Fowler, H. W., and Fowler, F. G. (1976). "The Concise Oxford Dictionary" (J. B. Sykes, ed.). Clarendon Press, Oxford.

Friedland, J. S., Remick, D. G., Shattock, R., and Griffin, G. E. (1992). Secretion of interleukin-8 following phagocytosis of *Mycobacterium tuberculosis* by human monocyte cell lines. *Eur. J. Immunol.* **22,** 1373–1378.

Friedland, J. S., Shattock, R., Remick, D. G., and Griffin, G. E. (1993). Mycobacterial 65-kD heat shock protein induces release of proinflammatory cytokines from human monocyte cells. *Clin. Exp. Immunol.* **91,** 58–62.

Galli, G., Ghezzi, P., Mascagni, P., Marcucci, F., and Fratelli, M. (1995). *Mycobacterium tuberculosis* 10kD heat shock protein increases both cell proliferation and cell death, depending on the state of the target cells. Submitted for publication.

Hermann, E., Lohse, A. W., Van der Zee, R., van Eden, W., Mayet, W. J., Probst, P., Poralla, T., Meyer zum Busehenfelde, K. H., and Fleischer, B. (1991). Synovial fluid-derived Yersinia-reactive T cells responding to human 65-kDa heat shock protein and heat-stressed antigen-presenting cells. *Eur. J. Immunol.* **21,** 2139–2143.

Hiromatsu, K., Yoshikai, Y., Matzuzaki, G., Ohga, S., Muramori, K., Matsumoto, K., Bluestone, J. A., and Nomoto, K. (1992). A protective role for γδ T cells in primary infection with *Listeria monocytogenes* in mice. *J. Exp. Med.* **175,** 49–56.

Hoiby, N. (1975). Cross-reactions between *Pseudomonas aeruginosa* and thirty-six other bacterial species. *Scand. J. Immunol.* **4**(Suppl. 2), 187–196.

Ilangumaran, S., Shanker Narayan, N. P., Ramu, G., and Muthukkaruppan, V. R. (1994). Cellular and humoral immune responses to recombinant 65-kD antigen of *Mycobacterium leprae* in leprosy patients and healthy controls. *Clin. Exp. Immunol.* **96,** 79–85.

Imani, F., and Soloski, M. J. (1991). Heat shock proteins can regulate expression of the T1a region-encoded class 1b molecule Qa-1. *Proc. Natl. Acad. Sci. USA* **88,** 10475–10479.

Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R., and Wiley, D. C. (1991). Identification of self peptides bound to purified HLA-B27. *Nature (London)* **353,** 326–329.

Jarjour, W., Mizzen, L. A., Welch, W. J., Denning, S., Shaw, M., Mimura, T., Haynes, B. F., and Winfield, J. B. (1990). Constitutive expression of a groEL-related protein on the surface of human cells. *J. Exp. Med.* **170,** 1857–1860.

Kaiser, B. (1975). Immunological studies of an antigen common to many Gram-negative bacteria with special reference to *E. coli. Int. Arch. Allergy Appl. Immunol.* **48,** 72–81.

Karlsson-Parrs, A., Söderström, K., Ferm, M., Ivanyi, J., Kiessling, R., and Klareskog, L. (1990). Presence of human 65kD heat shock protein (hsp) in inflamed joints and subcutaneous nodules of RA patients. *Scand. J. Immunol.* **31,** 283.

Kaufmann, S. H. E. (1990). Heat shock proteins and the immune response. *Immunol. Today* **11,** 129–136.

Kaufmann, S. H. E. (1994). Heat shock proteins and autoimmunity: A critical appraisal. *Int. Arch. Allergy Immunol.* **103,** 317–322.

Kaufmann, S. H. E., Vath, U., Thole, J. E. R., Van Embden, J. D. A., and Emmrich, F., (1987). Enumeration of T cells reactive with *Mycobacterium tuberculosis* organisms and specific for the 64-kDa protein. *Eur. J. Immunol.* **17,** 351–357.

Kaufmann, S. H. E., Schoel, B., Wand-Wurttenberger, A., Steinhoff, U., Munk, M. E., and Koga, T. (1990). T-cells, stress proteins, and pathogenesis of mycobacterial infections. *Curr. Top. Microbiol. Immunol.* **155,** 125–142.

Kaur, I., Voss, S. D., Gupta, R. S., Schell, K., Fisch, P., and Sondel, P. M. (1993). Human peripheral gamma delta T cells recognize hsp60 molecules on Daudi Burkitt's lymphoma cells. *J. Immunol.* **150**(5), 2046–2055.

Kindler, V., Sappino, A., Grau, G. E., Piguet, P., and Vassali, P. (1989). The inducing role of tumour necrosis factor in the development of bactericidal granulomas during BCG infection. *Cell* **56,** 731–740.

Koga, T., Wand-Wurttenberger, A., de Bruyn, J., Munk, M. E., Schoel, B., and Kaufmann, S. H. E. (1989). T cells against a bacterial heat shock protein recognize stressed macrophages. *Science* **241,** 1112–1115.

Kong, T. H., Coates, A. R. M., Butcher, P. D., Hickman, C. J., and Shinnick, T. M. (1993). *Mycobacterium tuberculosis* expresses two cpn60 homologs. *Proc. Natl. Acad. Sci. USA* **90,** 2608–2612.

Lamb, F. I., Kingston, A. E., Estrada, I., and Colston, M. J. (1988). Heterologous expression of the 65-kilodalton antigen of *Mycobacterium leprae* and murine T cell response to the gene product. *Infect. Immun.* **56,** 1237–1241.

Langer, T., and Neupert, W. (1991). Heat shock proteins hsp60 and hsp70: Their roles in folding, assembly and membrane translocation of proteins. *Curr. Top. Microbiol. Immunol.* **167,** 3–30.

Langermans, J. A. M., van der Hulst, M. E. B., Nibbering, P. H., Hiemstra, P. S., Fransen, L., and Van Furth, R. (1992). Interferon-γ-induced L-arginine-independent toxoplasmastatic activity in murine peritoneal macrophages is mediated by endogenous tumour necrosis factor α. *J. Immunol.* **148,** 568–574.

Lasek, R. J., Gainer, H., and Przybylski, R. J. (1974). Transfer of newly synthesized proteins from Schwann cells to the squid giant axon. *PNAS* **71,** 1188–1192.

Life, P., Hassell, A., Williams, K., Young, S., Bacon, P., Southwood, T., and Gaston, J. S. (1993). Responses to Gram negative enteric bacterial antigens by synovial T cells from patients with juvenile chronic arthritis: recognition of heat shock protein HSP60. *J. Rheumatol.* **20**(8), 1388–1396.

Lucas, S. B. (1988). Histopathology of leprosy and tuberculosis—an overview. *Brit. Med. Bull.* **44,** 584–599.

Lukacs, K. V., Lowrie, D. B., Stokes, R. W., and Colston, M. J. (1993). Tumour cells transfected with a bacterial heat shock gene lose tumorigenicity and induce protection against tumours. *J. Exp. Med.* **178,** 343–348.

Lussow, A. R., Barrios, C., van Embden, J., Van der Zee, R., Verdini, A. S., Pessi, A., Louis, J. A., Lambert, P. H., and Del Giudice, G. (1991). Mycobacterial heat-shock proteins as carrier molecules. *Eur. J. Immunol.* **21,** 2297–2302.

Martin, J., Horwich, A. L., and Hartl, F. U. (1992). Prevention of protein denaturation under heat stress by the chaperonin Hsp60. *Science* **258,** 995–998.

Mazodier, P., Guglielmi, G., Davies, J., and Thompson, C. J. (1991). Characterization of the groEL-like genes in *Streptomyces albus*. *J. Bacteriol.* **173,** 7382–7386.

Mehra, V., Bloom, B. R., Bajardi, A. C., Grisso, C. L., Sieling, P. A., Alland, D., Convit, J., Fan, X., Hunter, S. W., Brennan, P. J., Rea, T. H., and Modlin, R. L. (1992). A major T cell antigen of *Mycobacterium leprae* is a 10-kD heat-shock cognate protein. *J. Exp. Med.* **175,** 275–284.

Mims, C. A., Playfair, J. H. L., Roitt, I. M., Wakelin, D., and Williams, R. (1993). "Pathological Consequences of Infection." Medical Microbiology, Mosby, London.

Minden, P., Kelleher, P. J., Freed, J. H., Nielsen, L. D., Brennan, P. J., McPheron, L., and McClatchy, J. K. (1984). Immunological evaluation of a component isolated from *Mycobacterium bovis* BCG with a monoclonal antibody to *M. bovis* BCG. *Infect. Immun.* **46,** 519–525.

Morton, H., Hegh, V., and Clunie, G. J. A. (1974). Immunosuppression detected in pregnant mice by rosette inhibition test. *Nature (London)* **249,** 459–460.

Newcomb, J. R., and Cresswell, P. (1993). Characterization of endogenous peptides bound to purified HLA-DR molecules and their absence from invariant chain-associated α/β dimers. *J. Immunol.* **150,** 499–507.

Noll, A., Roggenkamp, A., Heesemann, J., and Autenrieth, I. B. (1994). Protective role for heat shock protein-reactive alpha beta T cells in murine yersiniosis. *Infect. Immun.* **62**(7), 2784–2791.

O'Brien, R. L., Fu, Y.-X., Cranfill, R., Dallas, A., Ellis, C., Reardon, C., Lang, J., Carding, S. R., Kubro, R., and Born, W. (1992). Hsp60-reactive $\gamma\delta$ cells: A large, diversified T lymphocyte subset with highly focused specificity. *Proc. Natl. Acad. Sci. USA* **89,** 4348–4352.

Oftung, F., Mustafa, A. S., Husson, R., Young, R. A., and Godal, T. (1987). Human T-cell clones recognize two abundant *M. tuberculosis* protein antigens expressed in *E. coli*. *J. Immunol.* **138,** 927–931.

Orme, I. M., Miller, E. S., Roberts, A. D., Furney, S. K., Griffin, J. P., Dobos, K. M., Chi, D., Rivoire, B., and Brennan, P. J. (1992). T lymphocytes mediating protection and cellular cytolysis during the course of *Mycobacterium tuberculosis* infection. *J. Immunol.* **148,** 189–196.

Peetermans, W. E., Raats, C. J., Langermans, J. A., and van Furth, R. (1994). Mycobacterial heat-shock protein 65 induces proinflammatory cytokines but does not activate human mononuclear phagocytes. *Scand. J. Immunol.* **39**(6), 613–617.

Pfeffer, K., Schoel, B., Gulle, H., Kaufmann, S. H. E., and Wagner, H. (1990). Primary responses of human T cells to mycobacteria: A frequent set of gamma/delta T cells are stimulated by protease-resistant ligands. *Eur. J. Immunol.* **20,** 1175–1179.

Quayle, A. J., Wilson, K. B., Li, S. G., Kjeldsen-Kragh, J., Oftung, F., Shinnick, T., Sioud, M., Forre, O., Capra, J. D., and Natvig, J. B. (1992). Peptide recognition, T cell receptor usage and HLA restriction elements of human heat-shock protein (hsp) 60 and mycobacterial 65 kDa hsp-reactive T cell clones from rheumatoid synovial fluid. *Eur. J. Immunol.* **22,** 1315–1322.

Rambukkana, A., Das, P. K., Witkamp, L., Yong, S., Meinardi, M. M., and Bos, J. D. (1993). Antibodies to mycobacterial 65-kDa heat shock protein and other immunodominant antigens in patients with psoriasis. *J. Invest. Dermatol.* **100**(1), 87–92.

Rammensee, H. G., Falk, K., and Rutzschke, O. (1993). Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.* **11,** 213–244.

Rinke de Wit, T. F., Bekelie, S., Osland, A., Miko, T. L., Hermans, P. W. M., van Soolmgen, D., Drijfhout, J.-W., Schoningh, R., Janson, A. A. M., and Thole, J. E. R. (1992). Mycobacteria contain two groEL genes: The second *Mycobacterium leprae* groEL gene is arranged in an operon with groES. *Mol. Microbiol.* **6,** 1995–2007.

Selmaj, K., Brosman, C. F., and Raine, C. S. (1991). Colocalisation of lymphocytes bearing $\gamma\delta$ T cell receptor and heat shock protein 65kDa oligodendryocytes in multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **88,** 6452–6456.

Sharif, M., Worrall, J. G., Singh, B., Gupta, R. S., Lydyard, P. M., Lambert, C., McCulloch, J., and Rook, G. A. (1992). The development of monoclonal antibodies to the human mitochondrial 60-kDa heat-shock protein, and their use in studying the expression of the protein in rheumatoid arthritis. *Arthritis Rheumatism* **35**(12), 1427–1433.

Shinnick, T. M. (1987). The 65-kilodalton antigen of *Mycobacterium tuberculosis. J. Bacteriol.* **169,** 1080–1088.

Silva, C. L., and Lowrie, D. B. (1994). A single mycobacterial protein (HSP 65) expressed by a transgenic antigen-presenting cell vaccinates mice against tuberculosis. *Immunology* **82,** 244–248.

Silva, C. L., Lukacs, K., and Lowrie, D. B. (1993). Major histocompatibility complex non-restricted presentation to CD4+ T lymphocytes of *Mycobacterium leprae* heat-shock protein 65 antigen by macrophages transfected with the mycobacterial gene. *Immunology* **78**(1), 35–42.

Sompolinsky, D., Hertz, J. B., Hoiby, N., Jensen, K., Mansa, B., Pedersen, V. B., and Samra, Z. (1980). An antigen common to a wide range of bacteria. 2. A biochemical study of a 'common antigen' from *Pseudomonas aeruginosa. Acta Pathol. Microbiol. Scand. B* **88,** 253–260.

Steinhoff, U., Zügel, U., Wand-Wurttenberger, A., Hengel, H., Rosch, R., Munk, M. E., and Kaufmann, S. H. E. (1994). Prevention of autoimmune lysis by T cells with specificity for a heat-shock protein by anti-sense oligonucleotide treatment. *Proc. Natl. Acad. Sci. USA* **91**(11), 5085–5088.

Thompson, S. J., Hitsumoto, Y., Ghoraishian, M., van der Zee, R., and Elson, C. J., (1991). Cellular and humoral reactivity pattern to the mycobacterial heat shock protein hsp65 in pristane induced arthritis susceptible and hsp65 protected DBA/1 mice. *Autoimmunity* **11**(2), 89–95.

Trieb, K., Sztankay, A., Hermann, M., Gratzl, R., Szabo, J., Jindal, S., and Grubeck-Loebenstein, B. (1993). Do heat shock proteins play a role in Graves' disease? Heat

shock protein-specific T-cells from Graves' disease thyroids do not recognize thyroid epithelial cells. *J. Clin. Endocrinol. Metab.* **77**(2), 528–535.

Tytell, M. (1987). Characterization of glial proteins transferred into the squid giant axon. *In* "Glial–Neuronal Communication in Development and Regeneration" (H. H. Althaus and W. Seifert, eds), pp. 249–261. Springer-Verlag, Berlin.

Tytell, M., Greenberg, S. G., and Lasek, R. J. (1986). Heat shock-like protein is transferred from glia to axon. *Brain Res.* **363,** 161–164.

Udono, H., and Srivastava, P. K, (1993). Heat shock protein 70-associated peptides elicit specific cancer immunity. *J. Exp. Med.* **178,** 1391–1396.

van Eden, W., Thole, J. E., Van der Zee, R., Noordzij, A., van Embden, J. D., Hensen, E. J., and Cohen, I. R. (1988). Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis. *Nature (London)* **331,** 171–173.

Verbon, A., Kuijper, S., Jansen, H. M., Speelman, P., and Kolk, A. H. J. (1990). Antigens in culture supernatant of *M. tuberculosis:* Epitopes defined by monoclonal and human antibodies. *J. Gen. Microbiol.* **136,** 955–964.

Wand-Wurttenberger, A., Schoel, B., Ivanyi, J., and Kaufmann, S. H. (1991). Surface expression by mononuclear phagocytes of an epitope shared with mycobacterial heat shock protein 60. *Eur. J. Immunol.* **21**(4), 1089–1092.

Wilbrink, B., Holewijn, M., Bijlsma, J. W., van Roy, J. L., den Otter, W., and van Eden, W. (1993). Suppression of human cartilage proteoglycan synthesis by rheumatoid synovial fluid mononuclear cells activated with mycobacterial 60-kd heat-shock protein. *Arthrits Rheumatism* **36**(4), 514–518.

Winfield, J., and Jarjour, W. (1991). Do stress proteins play a role in arthritis and autoimmunity? *Immunol. Rev.* **121,** 193–220.

Winrow, V. R., Mojdehi, G., Mapp, P. I., Rampton, D. S., and Blake, D. R. (1990). Immunohistological localisation of stress proteins in inflammatory tissue. *In* "Stress Protein in Inflammation" (R. Burdon, C. Rice-Evans, D. Blake, and V. Winrow, eds.), pp. 237–251. Richelieu Press, London.

Winrow, V. R., Mojdehi, G. W., Ryder, S. D., Rhodes, J. M., Blake, D. R., and Rampton, D. S. (1994a). Stress proteins in colorectal mucosa. Enhanced expression in ulcerative colitis. *Digest. Dis. Sci.* **38**(11), 1994–2000.

Winrow, V. R., Ragno, S., Morris, C. J., Colston, M. J., Mascagni, P., Leoni, F., Gromo, G., Coates, A. R. M., and Blake, D. R. (1994b). Arthritogenic potential of the 65kDa stress protein—an experimental model. *Ann. Rheum. Dis.* **53,** 197–201.

Worthington, J., Rigby, A. S., MacGregor, A. M., Silman, A. J., Carthy, D., and Ollier, W. E. (1993). Lack of association of increased antibody levels to mycobacterial hsp65 with rheumatoid arthritis: Results from a study of disease discordant twin pairs. *Ann. Rheum. Dis.* **52**(7), 542–544.

Xu, Q., Dietrich, H., Steiner, H. J., Gown, A. M., Schoel, B., Mikuz, G., Kaufmann, S. H. E., and Wick, G. (1992). Induction of arteriosclerosis in normocholesterolemic rabbits by immunisation with heat shock protein 65. *Arterio. Thromb.* **12,** 789–799.

Xu, Q., Leuf, G., Weimann, S., Gupta, R. S., Wolf, H., and Wick, G. (1993). Staining of endothelial cells and macrophages in atherosclerotic lesions with human heat-shock protein-reactive antisera. *Arterio. Thromb.* **13**(12), 1763–1769.

Yokota, S., Tsubaki, K., Kuriyama, T., Shimizu, H., Ibe, M., Mitsuda, T., Aihara, Y., Kosuge, K., and Nomaguchi, H. (1993). Presence in Kawasaki disease of antibodies to mycobacterial heat-shock protein HSP65 and autoantibodies to epitopes of human HPS65 cognate antigen. *Clin. Immunol. Immunopathol.* **67**(2), 163–170.

Young, D. B. (1990). Chaperonins and the immune response. *Semin. Cell. Biol.* **1,** 27–35.
Young, R. A., and Davies, R. W. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
Young, R. A., and Elliott, T. J. (1989). Stress proteins, infection and immune surveillance. *Cell* **59,** 5–8.

Index

P

T