1993

Specificity of Chaperonin GroEL Binding to the Precursor of the Small Subunit of Ribulose-1,5-Bisphosphate Carboxylase.

Carmen Wheelock Dessauer

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation


https://digitalcommons.lsu.edu/gradschool_disstheses/5494

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Specificity of chaperonin GroEL binding to the precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase

Dessauer, Carmen Wheelock, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1993
SPECIFICITY OF CHAPERONIN GroEL BINDING TO THE PRECURSOR OF THE SMALL SUBUNIT OF RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The Department of Biochemistry

by Carmen Wheelock Dessauer B. S., Louisiana State University, 1988 May 1993
ACKNOWLEDGEMENTS

I am grateful to C. P. Georgopoulos for providing the groE clone, A. R. Means for providing the calmodulin clone, and R. Broglie and N.-H. Chua for providing the wheat pS cDNA clone.

I thank the faculty members who served on my advisory committee, Drs. Simon Chang, Alan Klotz, Jim Moroney, and Jeff Nelson. I also want to thank Dr. Ezzat Younathan for patiently answering all my many questions.

The members of Dr. Bartlett's laboratory: Michael Bracey, Beth Floyd, Hee Jin Kim, and Pilar Tovar create a stimulating environment for the discussion of a wide range of scientific subjects. I am also thankful for the advice given to me by Drs. Sam Landry, Steve Pomarico, and Tim Fawcett. We are all indebted to our mentor, Dr. Sue Bartlett, for her friendship, contagious optimism, and insightful guidance.

Finally, I thank my mother and father for always encouraging me to reach for the highest goals; in their minds everything was possible for their little girl. I also must thank the many other members of my family, by birth and by marriage, who have patiently dealt with my demanding schedule and who rarely complained about forgotten birthdays or missed family gatherings. But most of all I am forever indebted to my husband, Patrick, whose support, encouragement, and cooking sustained me throughout my graduate career.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>LITERATURE REVIEW</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 1: Identification of a Region in a Chloroplast Precursor Protein Required for Binding the <em>E. coli</em> Chaperonin GroEL</td>
<td>39</td>
</tr>
<tr>
<td>CHAPTER 2: Identification of a Protein-Protein Contact Site Between GroEL and a Chloroplast Precursor Protein</td>
<td>73</td>
</tr>
<tr>
<td>CHAPTER 3: Specificity of Chaperonin GroEL Binding to a Chloroplast Precursor Protein</td>
<td>86</td>
</tr>
<tr>
<td>CONCLUSIONS AND FUTURE PROSPECTS</td>
<td>105</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>114</td>
</tr>
<tr>
<td>VITA</td>
<td>125</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Cellular Roles of Molecular Chaperones</td>
<td>5</td>
</tr>
<tr>
<td>II. Sequence of Site Specific Mutations Created in the Truncation PAxpS90T</td>
<td>59</td>
</tr>
<tr>
<td>III. Association Constants for Derivatives of PAxpS</td>
<td>98</td>
</tr>
<tr>
<td>IV. Association Constants for Mutations Created in PAxpS90T</td>
<td>99</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Removal of bound GroEL from PAxpS</td>
<td>52</td>
</tr>
<tr>
<td>1.2</td>
<td><em>In vivo</em> and <em>in vitro</em> binding of GroEL to the precursor and mature form of PAxpS</td>
<td>53</td>
</tr>
<tr>
<td>1.3</td>
<td>C-terminal truncations of PAxpS</td>
<td>55</td>
</tr>
<tr>
<td>1.4</td>
<td><em>In vivo</em> binding of GroEL to C-terminal truncations of PAxpS</td>
<td>56</td>
</tr>
<tr>
<td>1.5</td>
<td>Replacement of hydrophobic amino acids with charged residues (HCH) in the truncation PAxpS90T</td>
<td>60</td>
</tr>
<tr>
<td>1.6</td>
<td>Destabilization of a helix by substitution of amino acids with proline in the truncation PAxpS90T</td>
<td>61</td>
</tr>
<tr>
<td>1.7</td>
<td>Identification of a binding domain by substitution of amino acids with glycine in the truncation PAxpS90T</td>
<td>62</td>
</tr>
<tr>
<td>1.8</td>
<td>Destruction of the hydrophobic moment by substitution of amino acids with serine and glutamine in the truncation PAxpS90T</td>
<td>64</td>
</tr>
<tr>
<td>1.9</td>
<td>Transit sequence deletions in the 90 amino acid truncation of PAxpS</td>
<td>65</td>
</tr>
<tr>
<td>2.1</td>
<td>Two dimensional peptide maps</td>
<td>82</td>
</tr>
<tr>
<td>2.2</td>
<td>Sequence of iodinated peptides from protein A and pS</td>
<td>84</td>
</tr>
<tr>
<td>3.1</td>
<td>Scatchard plots of GroEL binding to PAxpS90T and PAxpS90THCH3</td>
<td>95</td>
</tr>
<tr>
<td>4.1</td>
<td>Model for the interaction of PAxpS and GroEL</td>
<td>111</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

1,8-ANS = 1-anilino-napthalene-8-sulfonate
bisANS = 1,1'-bi(4-anilino)naphthalene-5,5'-disulfonic acid
DHFR = dihydrofolate reductase
DTT = dithiothreitol

E. coli = Escherichia coli
ER = endoplasmic reticulum
IgG = immunoglobulin G
Ka = association constant
Kd = dissociation constant
L = large subunit of Rubisco
PA = staphylococcus protein A
PAxpS = fusion of pS to the C-terminus of PA separated by a Factor Xa protease site
pS = precursor form of the small subunit of Rubisco
RT = room temperature
Rubisco = Ribulose-1,5-bisphosphate carboxylase
S = small subunit of Rubisco
SDS-PAGE = sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TCA = trichloroacetic acid
ABSTRACT

Although chaperonin-assisted protein folding has been studied *in vitro* by a number of investigators, the features of an unfolded or partially folded polypeptide that are recognized and bound by chaperonins are not known. I addressed this question using the precursor of the small subunit (pS) of ribulose-1,5-bisphosphate carboxylase as a model substrate for GroEL, the bacterial chaperonin. The precursor protein was expressed in *E. coli* as a C-terminal fusion to protein A. Protein A-pS (and any associated cellular proteins) was isolated by affinity chromatography. GroEL could be eluted from the fusion protein by ATP and either GroES or casein, consistent with results of *in vitro* folding assays.

Using deletions from the C-terminus of pS I defined the smallest truncation of pS, PAxpS90T, that binds GroEL with high avidity. A series of site-specific mutations targeting the C-terminal 15 - 20 amino acids of PAxpS90T was constructed and analyzed for the ability to bind GroEL. Two of these mutations bound significantly less GroEL than PAxpS90T, suggesting that this region is required for avid GroEL binding. I demonstrated a physical interaction between GroEL and this region of pS with a novel assay that utilizes the protection of tyrosine residues from iodination upon formation of specific protein-protein complexes. Finally, I further showed
that at least half of the transit sequence of pS is also required for avid binding to GroEL.

The association constants for the interaction of GroEL with PAxpS, PAxpS90T, or its mutated derivatives, were determined and fell within the range $3.7 \times 10^7$ to $2.7 \times 10^6$ M$^{-1}$. Analysis of the affinity constants for PAxpS90T mutants allowed us to define a possible recognition motif for GroEL's interaction with pS. This motif includes the recognition of both hydrophobic and positively charged amino acids. The motif need not be helical, but structural rigidity may be a requirement for recognition (and binding) by GroEL.
LITERATURE REVIEW
Christian Anfinsen defined the first rules of protein folding in the thermodynamic hypothesis or self-assembly principle (Anfinsen, 1973). This hypothesis states that the native structure of a protein when in its normal physiological environment is determined by the lowest Gibbs free energy of all possible interatomic interactions of the system. Therefore, this spontaneous process is determined solely by the linear sequence of amino acids and requires no energy or factors extrinsic to the polypeptide itself. This basic tenet of molecular biology was developed from classic experiments on the in vitro refolding of bovine pancreatic ribonuclease (Haber and Anfinsen, 1962). Haber and Anfinsen showed that upon dilution from denaturants and reducing agents, ribonuclease spontaneously refolded into a catalytically active form, indistinguishable from the original native protein. However, even during these early folding experiments, investigators were disturbed by the kinetics of this process. The in vitro experiments required hours to produce a native product, while the rate of synthesis of protein chains in vivo required minutes. Goldberger et al. (1966) soon discovered that an enzyme (protein disulfide isomerase) found in the endoplasmic reticulum of cells catalyzes the rapid formation of disulfide bonds, producing a native ribonuclease in less than 2 minutes. This may have been the first hint of the existence of a class of proteins that would revolutionize the ideas of protein folding in the cell.
Molecular Chaperones

The same year that Anfinsen accepted the Nobel Prize for his work that led to the thermodynamic hypothesis, mutants of *Escherichia coli* called *groE* that appeared to block lambda phage head assembly were discovered by Georgopoulos *et al.* (1972). Although it would be several years before the global importance of this gene and its homologs would be recognized, it marked one of the key discoveries in the cellular protein folding problem. The *groE* gene product and protein disulfide isomerase were among the first of many classes of proteins to be discovered that are required for proper folding and/or assembly of proteins *in vivo*. Only after a great deal of work and several years were these classes of proteins finally named and defined by John Ellis in 1987 as "molecular chaperones". This term defines a family of proteins that mediates the correct folding of polypeptides and/or their assembly into oligomeric complexes, but are not themselves part of the final native structure. (Note: The erroneous English spelling of "chaperone" has been adopted throughout this field of study.)

The discovery of molecular chaperones does not negate the *in vitro* folding experiments. On the contrary, these experiments were, and continue to be, invaluable in deciphering the early stages of refolding. Two current models address the protein folding process (reviewed in Baldwin, 1989 and Ptitsyn, 1987). In the framework model, protein
folding consists of three stages: (1) the formation of fluctuating secondary structure in an unfolded chain, (2) the merging and thus stabilization of these secondary structures to form an intermediate globular structure, and (3) the adjustment of this intermediate by squeezing out water from the hydrophobic core as the side chains are fitted together. In the molten globule model, folding begins with a hydrophobic collapse to form the "molten globule state" which consists of a fluid, yet compact hydrophobic core with secondary structure present but few long range interactions. This core then reorganizes itself into the correct secondary and tertiary structures. Both models agree that many of the intermediates formed are in rapid equilibrium with the unfolded state and are only slowly converted to a native conformation. Although in vitro experiments continue to provide important details of the folding process, many of the conditions necessary for in vitro folding differ drastically from physiological conditions. In general, the solubility of "unfolded" or partially folded chains in aqueous solvents is poor (Fischer and Schmid, 1990). Folding reactions therefore require very dilute protein solutions to avoid premature aggregation of proteins prior to the final slow-folding steps. Furthermore, most in vitro folding reactions are performed at low temperatures. In contrast, a cell contains a very high concentration of proteins and the proteins may begin folding before the entire polypeptide chain is released from the ribosome. In addition, oligomeric complexes cannot assemble
until constituent polypeptides reach a specific location. Finally, most organisms grow at temperatures well above those used in *in vitro* folding reactions. A pool of unprotected, and incompletely assembled polypeptides could easily aggregate or misfold. Thus, it is easy to understand the necessity for a family of proteins that provides a safe environment for folding, translocation or assembly, processes which often involve the transient exposure of interactive protein surfaces to the environment.

Chaperones are found in all organisms. They vary widely in size, proposed function and location in the cell, and so a number of classes have been designated. These chaperone classes are outlined in the following table.

<table>
<thead>
<tr>
<th>Name of protein family</th>
<th>Proposed roles</th>
</tr>
</thead>
</table>
| Nucleoplasmins         | Nucleosome assembly  
                        | Possibly transcription and ribosome assembly. |
| Chaperonins            | Protein folding and assembly  
                        | Protein transport  
                        | Stress Protection  
                        | DNA replication and mRNA turnover? (E. coli) |

Table I

Cellular roles of Molecular Chaperones
<table>
<thead>
<tr>
<th>Table I, continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heat shock protein 70</strong> (Hsp70/ Hsc70/ Bip/ DnaK)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Heat shock protein 90</strong> (Hsp90)</td>
</tr>
<tr>
<td><strong>Peptidyl prolyl-cis-trans-isomerase (PPI/ CsA/ FKBP)</strong></td>
</tr>
<tr>
<td><strong>Protein disulfide isomerase (PDI)</strong></td>
</tr>
<tr>
<td><strong>Signal recognition particle (SRP)</strong></td>
</tr>
<tr>
<td><strong>TRAP or p28</strong></td>
</tr>
<tr>
<td><strong>p88</strong></td>
</tr>
<tr>
<td><strong>Trigger factor</strong></td>
</tr>
<tr>
<td><strong>Sec B protein</strong></td>
</tr>
<tr>
<td><strong>Pap D protein</strong></td>
</tr>
<tr>
<td><strong>Prosequence of:</strong></td>
</tr>
<tr>
<td>subtilisin</td>
</tr>
<tr>
<td>(\alpha)-lytic protease</td>
</tr>
<tr>
<td>bovine pancreatic trypsin inhibitor</td>
</tr>
</tbody>
</table>
Nucleoplasmin

Nucleoplasmin was the first protein to be described as a chaperone (Laskey et al., 1978). It is an acidic protein that binds histones and thus shields their positive charges. This interaction prevents non-specific binding of DNA to the histones and promotes the assembly of nucleosome core particles which are the basic repeating element of chromatin. The absence of nucleoplasmin in the reconstitution of these core particles leads to aggregation, although nucleoplasmin itself is not part of the final complex. Nucleoplasmin also interacts with ribonucleoprotein particles, suggesting that it may play a role in the assembly of these particles or in post-transcriptional RNA processing or transport (Dingwall and Laskey, 1990).

Heat Shock Protein 70

The heat shock protein 70 (Hsp70) class of chaperones is the largest, with possibly the most diverse functions. It is ubiquitous in nature and is found in a large number of compartments within the eukaryotic cell. The *E. coli* homolog (DnaK) promotes export of proteins when overexpressed (Phillips and Silhavy, 1990), can protect RNA polymerase from heat inactivation, and can resurrect heat-inactivated aggregates of RNA polymerase (Skowyra et al., 1990). Furthermore, the synthesis of DnaK is increased upon heat shock and the accumulation of denatured proteins. DnaK serves to regulate the heat shock response
through its interaction with $\sigma^{32}$ (Liberek et al., 1992) and possibly via its autophosphorylating activity (McCarty and Walker, 1991). DnaK also functions in conjunction with two other proteins, DnaJ and GrpE, to initiate replication complexes (Zylicz et al., 1989). The ATPase activity of DnaK, considered to be crucial in its activities, is stimulated 50-fold in the presence of these proteins (Liberek et al., 1991).

Multiple homologs of Hsp70 are found in eukaryotic cells. Many Hsp70s are expressed constitutively, but their synthesis increases upon heat shock. The function of these homologs are reviewed in Gething and Sambrook (1992) and are summarized as follows. The cytosolic forms of Hsp70 interact with newly synthesized proteins, presumably to aid in folding or to maintain them in a transport-competent state (i.e. unfolded or partially unfolded) for translocation to the ER, mitochondria, or chloroplast. In addition, Hsp70 promotes the disassembly of clathrin cages and targets proteins to lysosomes for degradation. Cytosolic Hsp70 also travels to the nucleus and nucleolus upon heat shock and associates with insoluble polypeptides and partially assembled proribosomes. The ER, mitochondrial, and chloroplast homologs of Hsp70 all play a role in receiving unfolded polypeptides as they are transported across the respective membranes. The ER protein (BiP) also associates with malfolded polypeptides that cannot exit the ER and possibly interacts with an ER homolog of the E. coli protein DnaJ. In all cases mentioned,
substrates of Hsp70 can be released upon addition of ATP but not with nonhydrolyzable analogs of ATP. All of these functions point to Hsp70's role in binding and stabilizing proteins that have not yet folded or are denatured upon heat shock and stress.

**Heat Shock Protein 90**

The Hsp90 class of proteins is present in all prokaryotic and eukaryotic organisms examined and can associate with a variety of proteins such as steroid hormone receptors, protein kinases, actin, tubulin, and retroviral transforming proteins (reviewed in Craig, 1988). Hsp90 presumably stabilizes these proteins in an inactive or unassembled state until they reach their proper destinations. Hsp90 actually masks the DNA-binding activity of steroid hormone receptors, releasing them only after the receptor binds a hormone molecule (Picard *et al.*, 1988, Picard *et al.*, 1990, and Dalman *et al.*, 1991). Another protein, p59, which contains a region of homology to the chaperone peptidyl prolyl-cis-trans-isomerase, also is present in this complex (Lebeau *et al.*, 1992). Hsp90 binds ATP and can undergo autophosphorylation (Csermely and Kahn, 1991), just as the Hsp70s can, but it is not clear if cleavage of ATP is necessary for its activity. In an *in vitro* folding reaction with denatured citrate synthase, Hsp90 increased the yield of native protein in the absence of ATP (Wiech *et al.*, 1992)
Peptidyl Prolyl-cis-trans-Isomerase

The peptidyl prolyl-cis-trans-isomerase (PPI) class of proteins are also ubiquitous in nature. Cyclophilin and FK506 Binding Protein, although unrelated to one another, can catalyze the interconversion of the cis and trans proline rotamers (Fischer et al., 1989, Siekierka et al., 1989, and Harding et al., 1989). These proteins are distinguished by their ability to bind the immunosuppressive agents cyclosporin A and FK506, respectively. The role of these proteins in vivo is still unclear, but they are believed to catalyze slow folding steps which occur late in protein folding pathways and are often associated with proline isomerization. Prolyl isomerases can accelerate the folding of several proteins in vitro (Fischer and Bang, 1985, and Lin et al., 1988), while the cyclophilin-type protein, NinaA, is required for the folding and/or stability of rhodopsins 1 and 2 in the ER (Stamnes et al., 1991). In addition to its role in proline isomerization, PPI can also act at earlier stages in protein folding, decreasing aggregation of intermediates and thus increasing the yield of native protein (Freskgård et al., 1992).

Protein Disulfide Isomerase

Protein disulfide isomerase (PDI) catalyzes disulfide bond formation in vitro. The efficiency of disulfide bond formation is improved for at least Ribonuclease T1 with the simultaneous use of PPI in these reactions (Schönbrunner and Schmid, 1992). PDI is thought to be required for
folding of proteins in the ER (Bulleid and Freedman, 1988) and deletion of the PDI gene in yeast is lethal (Lamantia et al., 1991), although the exact cause of death is unknown. The mammalian PDI is homologous to the *E. coli* protein thioredoxin, the beta subunit of prolyl-4-hydroxylase, and small domains of several ER resident proteins (reviewed in Noiva and Lennarz, 1992). Thioredoxin is not as efficient as the ER form of PDI in catalyzing the refolding of RNase and may not be limited to disulfide bond formation in its *in vivo* functions. Thioredoxin is required for assembly of filamentous phages in *E. coli*, however its redox activity does not play a role in this function (Gething and Sambrook, 1992). In chloroplasts, proteins which have limited homology to *E. coli* thioredoxin serve as electron donors and as regulators of carbon metabolism (Cséke and Buchanan, 1986 and Crawford et al., 1985). Although results of *in vitro* studies and yeast genetic studies suggest that PDI does participate in chaperone activities, its actual role *in vivo* must still be determined.

**Signal Recognition Particle**

The signal recognition particle (SRP) helps proteins destined for the ER to either cross or become incorporated into the lipid bilayer (Walter and Blobel, 1980). These proteins are synthesized with a hydrophobic signal sequence which is recognized by the SRP as it emerges from the ribosome during translation (Walter and Blobel, 1981a). The binding of this sequence halts translocation until the entire complex binds to a ER
membrane receptor (Walter and Blobel, 1981a,b). The nascent polypeptide is then transferred to the "translocation machinery", the SRP is released, and cotranslational transport continues (Meyer et al., 1982 and Gilmore et al., 1982). The SRP is a complex of 6 proteins (Walter and Blobel, 1980) and an RNA molecule (7SL) (Walter and Blobel, 1982). *E. coli* uses a similar particle in the export of at least some proteins (Phillips and Silhavy, 1992). Homologs of the signal sequence binding subunit and the 7SL RNA form a ribonucleoprotein particle in *E. coli* that interacts specifically with the signal sequence of nascent secretory proteins (Luirink et al., 1992).

Two other *E. coli* proteins (Trigger factor and SecB) facilitate export of proteins by maintaining them in a loosely folded state for translocation (Crooke and Wickner, 1987 and Kumamoto, 1989). Once in a complex with a precursor protein, SecB will bind to a peripheral membrane protein, SecA, which interacts with part of the translocation machinery, SecY and SecE.

Several other proteins or protein units act as chaperones for a single protein or complex. TRAP (p28) associates with CD3 chains until they assemble with other subunits of the T-cell receptor (Pettey et al., 1987). P88 stabilizes newly synthesized major histocompatibility complex class I heavy chains before assembly (Degen and Williams, 1991) and the *E. coli* protein Pap D functions in the assembly of the attachment appendages of bacteria or P pili (Sambrook and Gething, 1989 and Holmgren and
Branden, 1989). Finally, several proteins require their prosequences in order to properly fold, including insulin (Steiner and Clark, 1968), carboxypeptidase Y (Winther and Sørensen, 1991), subtilisin (Zhu et al., 1989), α-lytic protease (Silen and Agard, 1989), and bovine pancreatic trypsin inhibitor (BPTI, also known as aprotinin) (Weissman and Kim, 1992). These sequences may facilitate folding via a variety of mechanisms. The prosequence of BPTI substantially increases the rate of formation and the yield of native protein by providing a single cysteine that acts as an intramolecular thiol-disulfide reagent (Weissman and Kim, 1992). The pro-sequence of α-lytic protease can even be separated from the mature protein (Silen and Agard) and still accelerate the rate-limiting step on the folding pathway (Baker et al., 1992), promoting formation of an active protease. Other pro-regions may provide a scaffold onto which the mature protein can fold. These pro-sequences either are later cleaved by a protease to obtain a functional protein or undergo autocatalytic cleavage (Silen et al., 1989 and Ikemura and Inouye, 1988). The obvious advantage of prosequences is the covalent linkage of the chaperone to its target protein.

The *E. coli* Chaperonin GroEL

The last class of chaperones is the chaperonin (Cpn) family which will be the focus of the remainder of this dissertation. As mentioned earlier, Georgopoulos and others originally identified the *E. coli* member of
this family (GroEL) by screening for bacterial genes essential for bacteriophage development (Georgopoulos et al., 1973 and Sternberg, 1973). The groE mutants could be suppressed by mutation of the lambda phage gene E (hence the name groE) which encodes the major structural subunit of the phage head. Subsequently, Georgopoulos and coworkers (1973) discovered that the groE locus is necessary for the assembly of a dodecameric protein structure called the head-tail connector of lambda and for the assembly of T4 phage (Georgopoulos et al., 1972). Mutations in groE lead to aggregation of these lambda proteins (Georgopoulos et al., 1973). In addition to the block in phage assembly, the groE mutants also exhibited effects on normal bacterial functions (Georgopoulos and Ang, 1990). Some of these phenotypes included a reduction in overall rates of DNA and RNA synthesis at nonpermissive temperatures, a block in cell division leading to formation of long filaments without septa, and a reduction in generalized protease activity. Further characterization of groE mutants revealed that the reduction in RNA synthesis is due to the impaired activity of RNA polymerase in the absence of a functional GroEL-GroES complex (Wada et al., 1987). In addition, GroEL interacts with an RNase E-like activity which affects RNA processing but not cleavage upon mutation of GroEL (Sohlberg et al., 1993 and Chanada et al., 1985). Since these early mutants were isolated on the basis of a block in bacteriophage growth, the spectrum of mutations and phenotypes observed for normal
bacterial functions was limited. Mutations at the \textit{groE} locus also cause defective export of \(\beta\)-lactamase (Kusukawa \textit{et al.}, 1989) and a reduction in SOS repair of UV-irradiated bacteriophage (Liu and Tessman, 1990).

The \textit{groE} locus maps to 94 min on the \textit{E. coli} genetic map (Georgopoulos and Eisen, 1974) and is actually an operon encoding two protein products, GroES and GroEL, with molecular masses of 10,368 and 57,259, respectively (Hemmingsen \textit{et al.}, 1988). These genes are constitutively expressed as a single transcript of approximately 2,100 nucleotides from a promoter under the control of \(E\sigma^{70}\) and accounts for approximately 1\% of the total soluble protein in \textit{E. coli} grown at 37°C. Synthesis increases as much as 10 fold (up to 10\% of the soluble \textit{E. coli} proteins) under heat shock or other stress conditions (reviewed in Georgopoulos and Ang, 1990). The bulk of transcription during heat shock is carried out by the \(E\sigma^{32}\) RNA polymerase holoenzyme which specifically recognizes heat shock promoters. This dual level of control is comforting for the \textit{E. coli} cell since both gene products are required for \textit{E. coli} viability at all temperatures (Fayet \textit{et al.}, 1989).

Both GroES and GroEL are acidic proteins with predicted pI values of 5.92 and 5.63, respectively (Hemmingsen \textit{et al.}, 1988). GroEL is composed of 14 identical subunits arranged in two cylindrical stacks of 7 subunits each and the holoenzyme is 125-130 Å in diameter and 100-114 Å in height (Hendrix, 1979). GroES is a symmetrical ring-like structure of
7 subunits and cosediments with GroEL in the presence, but not in the absence, of ATP (Chandrasekhar et al., 1986). In addition, GroES can specifically be retained on a GroEL-affinity column (Chandrasekhar et al., 1986) and has approximately 50% homology to GroEL (Welch, 1991). GroEL has a very weak ATPase activity (0.8 moles of ATP hydrolyzed per second per mole of GroEL particles) (Hendrix, 1979) which is partially inhibited in the presence of GroES (Chandrasekhar et al., 1986). Genetic evidence also indicates that these two proteins interact *in vivo*. Mutations in *groEL* can be isolated that compensate for a temperature-sensitive phenotype of certain *groES* mutations in an allele specific manner (Tilly and Georgopoulos, 1982).

Genetic evidence has also provided clues as to the function of the GroE proteins *in vivo*. Mutations in the *rpoA* gene, coding for the α-subunit of RNA polymerase, were identified as extragenic suppressors for the temperature-sensitive phenotype of certain *groES* mutations (Wada et al., 1987). These results, in conjunction with the identification of GroEL in RNA polymerase preparations (Paetkau and Coy, 1972), suggest that the GroE proteins may play a role in the assembly of the RNA polymerase core enzyme (Georgopoulos and Ang, 1990).

The expression of a specific mutation of GroEL (*groEL411*) was required to suppress the temperature-sensitive phenotype of a mutation in the single-stranded DNA binding protein, *ssb*, required for DNA replication
Overexpression of the wild type GroEL or GroES could not substitute for the \textit{groEL411} mutant suggesting that GroEL411 had obtained a new function. The functional structure of SSB is a tetramer. However, the SSB1 mutant protein tends to dissociate to monomers at physiological concentrations (Williams \textit{et al.}, 1984). SSB1 can be boiled and later regain full activity upon cooling (Meyer \textit{et al.}, 1980), therefore the inactive monomers are not irreversibly denatured at the non-permissive temperature. It appears that GroEL411 may stabilize tetramer formation or actively reorganize SSB1 monomers back into tetramers.

The overexpression of \textit{groEL} and \textit{groES} can compensate for mutations in a variety of genes such as acetolactate synthase II and transaminase B of the \textit{ilv} operon, several multimeric members of the histidine (\textit{his}) operon of \textit{Salmonella typhimurium}, and the secretion dependent genes, \textit{secA} and \textit{secY}, of \textit{E. coli} (Van Dyk \textit{et al.}, 1989).

Overproduction of GroEL and GroES can overcome lethal jamming of the bacterial export machinery (Phillips and Silhavy, 1990) and can suppress the temperature-sensitive phenotype of certain \textit{dnaA} mutants (Fayet \textit{et al.}, 1986 and Jenkins \textit{et al.}, 1986). (The \textit{dnaA} gene product is required for initiation of DNA replication at the oriC site.) In all cases studied except \textit{rpoH} (see below), overexpression of the GroE proteins does not bypass the need for the mutant polypeptide. Fayet \textit{et al.} (1986) suggested that GroEL
and GroES must stabilize the mutant proteins at the non-permissive temperatures. The \textit{rpoH} (\textit{htpR}) gene codes for $\sigma_{32}$ which controls the bulk of transcription at high temperatures and can be deleted in \textit{E. coli} only at temperatures below 20°C (Zhou \textit{et al.}, 1988). Overexpression of GroEL and GroES can partially compensate for this deletion, allowing colony formation at temperatures as high as 39°C (Kusukawa and Yura, 1988). In this case, the GroE proteins do not take on the role of $\sigma_{32}$, rather they are believed to bypass the deletion by stabilizing cellular proteins at these higher temperatures. Therefore, overexpression of the GroE proteins actually bypasses the need for all the other heat shock proteins normally present at elevated temperatures.

Yet another example of the folding/assembly function of the GroE proteins is reflected in their requirement for the assembly of oligomeric proteins. Assembly of the dimeric Rubisco (ribulose-1, 5-bisphosphate carboxylase/oxygenase) from \textit{Rhodospirillum rubrum} (Goloubinoff \textit{et al.}, 1989a) and of the nitrogenase MoFe protein (Govezensky \textit{et al.}, 1991) in \textit{E. coli} requires the GroE proteins. Mutations in either \textit{groE} gene resulted in no Rubisco activity or assembled MoFe protein. Bochkareva \textit{et al.} (1988) demonstrated early on that GroEL could associate with newly synthesized proteins other than the lambda structural proteins. Pre-$\beta$-lactamase and chloramphenicol acetyltransferase were cross-linked to GroEL when translated \textit{in vitro} in an S30 extract from \textit{E. coli}. Either apomyoglobin or
heat-denatured apomyoglobin but not native myoglobin, could compete with the newly-synthesized proteins for GroEL. All of these examples point to the ability of GroEL, in conjunction with GroES, to facilitate folding, assembly, and/or secretion of newly synthesized proteins.

Rubisco Binding Protein

The concept of molecular chaperones actually resulted from studies of the chloroplast chaperonin, also known as rubisco binding protein (Barraclough and Ellis, 1980). Rubisco is a soluble chloroplast protein composed of eight large (55 kDa) and eight small (14 kDa) polypeptide subunits (Chapman et al., 1988). The large subunit is encoded and synthesized within the chloroplast (McIntosh et al., 1980). The small subunit is encoded in the nuclear DNA and is synthesized as a precursor protein containing a 5 kDa NH$_2$-terminal transit sequence that directs its transport into the chloroplast (Chua and Schmidt, 1978). Once inside the chloroplast, the transit sequence is cleaved and the mature protein assembles with the large subunit of Rubisco to form the holoenzyme. When radioactive label is added to intact chloroplasts during light-driven protein synthesis, the large subunit of Rubisco is essentially the only soluble labelled product. Barraclough and Ellis (1980) discovered that these newly synthesized large subunits were not incorporated into the Rubisco holoenzyme, but were bound by a large molecular weight protein which was composed of 60 kDa subunits. This novel protein was called
the rubisco binding protein and was predicted to be involved in an intermediate step in Rubisco assembly. After prolonged incubations in the light under chase conditions, radioactive large subunits were released from rubisco binding protein and co-migrated with the active Rubisco holoenzyme. Incorporation of newly-synthesized radioactive subunits into the holoenzyme was not seen in the dark, when intact chloroplasts are no longer producing large amounts of ATP. However, addition of ATP could substitute for light in promoting the assembly of large subunits into Rubisco (Bloom et al., 1983).

Accumulation of rubisco binding protein shows tissue-specific regulation and its abundance is correlated with that of Rubisco (Hemmingsen, 1990), but even plastids in non-photosynthetic tissues that contain no Rubisco express low levels of the binding protein (Hemmingsen et al., 1988). Levels of rubisco binding protein increase as pea seedlings are shifted from dark- to light-grown conditions and oscillate in a circadian manner (Hemmingsen, 1990). The regulation of rubisco binding protein is also under heat shock control; however, unlike many other heat shock proteins, the levels of this protein are only slightly increased (Hemmingsen, 1990).

Rubisco binding protein is actually composed of approximately an equal amount of two distinct 60 kDa polypeptides, distinguished as alpha and beta (Hemmingsen and Ellis, 1986). Electron microscopy indicates
that the structure is a stack of two rings with 7-fold symmetry (Pushkin et al., 1982), identical to that described for the *E. coli* chaperonin, GroEL. In addition, like GroEL, rubisco binding protein also has a weak ATPase activity (Pushkin et al., 1982). Rubisco binding protein reversibly dissociates slightly in the presence of millimolar amounts of ATP and this dissociation is increased at decreased temperature (Bloom et al., 1983). The tetradecameric complex, however, seems to predominate *in vivo* (Roy et al., 1988b) and is stable in the presence of ATP at high concentrations of the rubisco binding protein *in vitro* (Roy et al., 1988a). The purification of rubisco binding protein from pea eventually led to the isolation of the nuclear-encoded genes for the alpha subunit from *Ricinus communis* (castor bean) and *Triticum aestivum* (wheat) (Hemmingsen et al., 1988). These sequences, published in conjunction with the sequence for the *groE* operon, showed clearly that rubisco binding protein is the chloroplast homolog of GroEL. Both the alpha and beta subunits have now been sequenced for *Brassica napus* and show 49% identity between each other and 47% and 52% identity to GroEL, respectively (Hemmingsen, 1990). The chloroplast chaperonin is unique in its composition of alpha and beta subunits and the relevance of both subunits to the function of this chaperonin is unknown.
The Mitochondrial Chaperonin

In searching for heat-inducible, ribosome-associated proteins responsible for regulation of translation during heat shock, McMullin and Hallberg (1987) generated antibodies against a 58 kDa protein from *Tetrahymena thermophila*. Although this protein had nothing to do with translation *per se* (Hallberg, 1990), the hsp58 antibodies cross-reacted with similarly sized proteins from *E. coli, Saccharomyces cerevisiae, Xenopus laevis, Zea mays*, and human cells (McMullin and Hallberg, 1988). The *E. coli* and the *Saccharomyces* proteins shared several characteristics with the *Tetrahymena* mitochondrial protein including heat inducibility, sedimentation coefficients and virtually identical morphologies as seen with the electron microscope. These results suggested that hsp58 from *Tetrahymena* mitochondria is a homolog of GroEL. Subsequently, several mitochondrial homologs of GroEL were isolated and characterized. The *Neurospora crassa* mitochondrial chaperonin was purified and shown to possess the same 7-fold symmetry composed of 14 subunits (Hutchinson *et al.*, 1989). The *Saccharomyces* mitochondrial cpn60 shares 54% identity with *E. coli* GroEL and 45% identity with the wheat chloroplast cpn60 (Reading *et al.*, 1989). Disruption of the *Saccharomyces* cpn60 gene results in cells that are not viable at any temperatures (Reading *et al.*, 1989). Chaperonin-60 (cpn60) family members have since been found in all bacteria and mitochondria examined and in all plastids,
including chloroplasts, chromoplasts, and leucoplasts (reviewed in Ellis, 1990b). They are abundant, constitutive proteins that are essential for growth and increase to varying extents upon stress (Ellis, 1990b). All members of this family are composed of 14, 60 kDa subunits with one exception, the mammalian mitochondrial chaperonin (see below, Viitanen et al., 1992b).

The designation chaperonin-60 (cpn60) is used to differentiate the 60 kDa, 14 subunit complex from the other member of the chaperonin family, GroES. GroES, or chaperonin-10 (cpn10), works in conjunction with GroEL in E. coli, as mentioned previously. One would thus expect the mitochondrial and chloroplast chaperonin-60s to also function in conjunction with their respective chaperonin-10 homologs. A mammalian mitochondrial cpn10 has been isolated and shown to form a complex with GroEL in the presence of ATP, inhibit the ATPase activity of GroEL, and participate in the release of substrates from GroEL (Lubben et al., 1990). Further studies of the mammalian mitochondrial cpn10 reveal that it has 41% identity to the bacterial cpn10 and exists as a 65 kDa complex (Hartman et al., 1992). A chaperonin-10 homolog has also been identified in chloroplasts (P. Viitanen, personal communication) and a large number of bacteria ranging from Thermophilic bacterium to Synechococcus (reviewed in Hartman et al., 1992).
Chaperonin Dependent *In Vitro* Refolding Reactions

Goloubinoff *et al.* (1989a) showed that *groEL* and *groES* were required *in vivo* for the assembly of *Rhodospirillum* dimeric Rubisco in *E. coli*. Subsequently, using purified GroEL and GroES, Goloubinoff *et al.* (1989b) demonstrated that the *in vitro* reconstitution of active dimeric Rubisco from a completely unfolded state depends on the presence of cpn60, cpn10 and MgATP. The absence of cpn60 led to unproductive aggregates, whereas the absence of cpn10 or MgATP led to a stable complex between cpn60 and the unfolded Rubisco. The later addition of MgATP to the complex resulted in the partial dissociation of the cpn60 14-mer and the release of Rubisco from cpn60, but active dimeric Rubisco was not formed. Active Rubisco was only formed from a cpn60-Rubisco complex in the presence of both cpn10 and MgATP. Nonhydrolyzable analogues of ATP or cpn10 alone had no effect on the release of Rubisco from cpn60. Later studies revealed that both the chaperonin-dependent reconstitution of dimeric Rubisco and the uncoupled ATPase activity of cpn60 require low concentrations of ionic K⁺, NH₄⁺, or Rb⁺ (Viitanen *et al.*, 1990). The formation of a complex between cpn60 and cpn10 does require ATP but not the presence of K⁺ which suggests that ATP hydrolysis is not necessary for complex formation (Viitanen *et al.*, 1990).

Viitanen *et al.* (1990) developed conditions that support Rubisco folding and assembly in absence of chaperonins. At temperatures below
15°C, spontaneous reconstitution will occur, although the reaction is still stimulated 10-fold in the presence of chaperonins and MgATP at this temperature. In the absence of K⁺ and MgATP, spontaneous folding was inhibited by the presence of a molar excess of cpn60 over Rubisco. The complex formed between cpn60 and Rubisco under these conditions led to active Rubisco up to 16 hours after complex formation upon addition of K⁺ and MgATP in the presence of cpn10. The role of cpn10 may be to couple the K⁺-dependent hydrolysis of ATP to the release of the folded substrate from cpn60. A working hypothesis for this reaction is as follows. Cpn60 can bind to unfolded or partially folded polypeptides in the absence of any other cofactors. The polypeptide is therefore stabilized and prevented from aggregating. Upon ATP hydrolysis in the presence of cpn10, the substrate is released and obtains a native conformation. ATP hydrolysis alone results in the release of Rubisco but in an inactive, aggregated state. In the presence of ATP, cpn60 and cpn10 exist as a complex, thereby preventing the wasteful hydrolysis of ATP (Chandrasekhar et al., 1986).

Many steps in this process are still ill-defined, but it is believed that the chaperonins facilitate folding by preventing off pathway events which typically lead to aggregation. In this model the unfolded protein is sequestered and protected, allowing folding to occur while immobilized on the chaperonin. Chaperonins do not change the folding pathway of a
protein as evidenced by the same pH dependence of the folding reaction and the appearance of similar intermediates in the presence and absence of chaperonins (Martin et al., 1991 and Zahn and Plückthun, 1992). Results of folding studies with dimeric Rubisco suggest that cpn60 recognizes an intermediate along the folding pathway which contains significant alpha helical content, rather than the unfolded protein (van der Vies et al., 1992). The products of the chaperonin-dependent reconstitution reaction are folded monomers which spontaneously form the dimeric protein. One caveat of these studies is the existence of equilibria between all species along the folding pathway. Therefore, the possibility that cpn60 reacts with poorly populated conformational states, such as the unfolded state, that are in equilibrium with the folding intermediate cannot be ruled out.

In addition to the bacterial cpn60 (GroEL), the yeast mitochondrial cpn60 and the chloroplast cpn60 (Rubisco binding protein) also support reconstitution of dimeric Rubisco, but only in the presence of MgATP and the bacterial cpn10 (GroES) (Goloubinoff et al., 1989b). The efficiency of this reaction is only 10-25% that of the homologous system from E. coli, but these proteins have probably evolved to work with their own respective cpn10 counterparts. This is certainly true for the mammalian mitochondrial cpn60 (Viitanen et al., 1992b). As mentioned earlier, the mammalian mitochondrial cpn60 is the only cpn60 characterized so far
that is not composed of 14 subunits upon purification. This cpn60 homologue is composed of a single ring of 7 subunits and can form a stable complex with the mammalian mitochondrial cpn10, but not the bacterial GroES. The mammalian mitochondrial cpn60 facilitates the reconstitution of dimeric Rubisco in the presence of MgATP, K⁺, and the mitochondrial cpn10, but not the bacterial cpn10.

The chaperonin-assisted refolding of rhodanese is probably the most studied of the refolding systems. Rhodanese is a monomeric, mitochondrial matrix protein that does not contain a cleavable transit sequence. The refolding of rhodanese in the absence of chaperonins requires nondenaturing detergents to minimize aggregation and is reversible (Mendoza et al., 1991b). The detergents function by interacting with hydrophobic surfaces that are exposed and which would otherwise interact with one another leading to aggregates. This mechanism may be similar to chaperonin-assisted refolding which requires both cpn60 and cpn10, in addition to MgATP and K⁺ (Mendoza et al., 1991a). The bacterial cpn60 does possess exposed hydrophobic surfaces, as monitored by the fluorescent probe bisANS, that may interact with substrates (Mendoza et al., 1991a). One cpn60 particle binds one rhodanese molecule (Mendoza et al., 1991a) and fluorescence studies suggest that a molten globule intermediate is bound by cpn60 (Martin et al., 1991). The addition of large amounts of α₁-casein could compete with rhodanese in
binding to cpn60. In the absence of cpn10, only small amounts of rhodanese were released from cpn60 after addition of MgATP and K⁺. The addition of casein resulted in the complete release of rhodanese and its subsequent aggregation. ATP hydrolysis may result in the transient release of rhodanese but in the absence of any competitive substrate or GroES, rhodanese quickly rebinds to cpn60. One model for rhodanese refolding states that folding on the surface of cpn60 requires initial binding of two or more segments of the protein substrate and a sequential release of them on ATP hydrolysis (Martin et al., 1991). However to date, only limited data support this model. In addition to refolding, cpn60 protects rhodanese (Mendoza et al., 1992) and α-glucosidase (Neugebauer and Rudolph, 1991) against heat inactivation, supporting its role as a cellular "protector" during heat shock.

Refolding of citrate synthase also requires both cpn60 and cpn10 (Zhi et al., 1992) for folding in vitro in the absence of detergents or other folding assistants. However, one group of proteins does not require cpn10 in chaperonin-assisted refolding. In the absence of cpn10, hydrolysis of ATP by cpn60 promotes the release of substrate which is partially folded but catalytically inactive (Martin et al, 1991). Whether the released substrate simply aggregates or continues along the proper folding pathway generating a native protein depends on the nature of the individual substrate and its propensity to spontaneously fold under the given
conditions. In the presence of cpn10, ATP hydrolysis is coupled to folding (Viitanen et al., 1990), which may proceed through intermediate conformations that are progressively released (Martin et al., 1991), apparently committed to completion of the folding process. The proteins that require cpn10 in chaperonin-dependent folding all have a strong tendency to aggregate, whereas those proteins that do not require cpn10 generally can refold quite efficiently in the absence of any chaperonin. This is certainly true for dihydrofolate reductase (DHFR) whose spontaneous refolding can be arrested by cpn60, resulting in a stable complex of cpn60 and a folding intermediate of DHFR (Viitanen et al., 1991). This intermediate is released upon addition of ATP to form the native protein, but neither hydrolysis of ATP nor the presence of cpn10 is required. However, cpn10 does potentiate the maximum effect of ATP. The overall effect of cpn60, ATP, and/or cpn10 is to actually slow the folding of DHFR (Martin et al., 1991). The physiological relevance of such a system is uncertain. Cpn60 can increase the yield of native glutamine synthetase (Fisher, 1992), pre-β-lactamase (Zahn and Plückthun, 1992), and tryptophanase (Mizobata et al., 1992) in the presence of nonhydrolyzable ATP (or even ADP in the case of tryptophanase), but in the absence of cpn10. In these cases, cpn60 may act as a detergent, reducing aggregation and thus increasing the yield of native protein. The native form of pre-β-lactamase, but not that of the mature β-lactamase,
will undergo a net unfolding in the presence of cpn60 which can be reversed with ATP (Laminet et al., 1990). The authors concluded that cpn60 recognizes a non-native conformation in equilibrium with the native state, thereby trapping an intermediate form of the protein. The slight destabilization of the mature protein conferred by the precursor sequence (Liu et al., 1988) may be all that is necessary for recognition and binding by chaperones and thus maintenance of a transport-competent form of the precursor protein.

Are chaperonins catalysts?

The answer to whether chaperonins function as enzymes may depend partly on the protein substrate. The Oxford English Dictionary defines a catalyst as a substance which, when present in small amounts, increases the rate of a chemical reaction or process but which is chemically unchanged by the reaction. Chaperonins can facilitate folding at concentrations similar to that of their substrate and can recycle and be used in subsequent reactions (Langer et al., 1992). In every case studied, chaperonins increase the yield of native protein, but their effect on the rate of refolding is substrate specific. The rate of refolding of dimeric Rubisco is accelerated 10-fold in the presence of cpn60 and cpn10 (Goloubinoff et al., 1989b), whereas refolding rates of citrate synthase (Buchner et al., 1991), α-glucosidase (Neugebauer and Rudolph, 1991), and pre-β-lactamase (Laminet et al., 1990) are unchanged, and rates of DHFR
(Martin et al., 1991) and rhodanese (Mendoza et al., 1991a) are decreased. If chaperonins function to stabilize folding intermediates and thus prevent off-pathway events leading often to aggregation, folding rates could be accelerated simply by the presence of a higher concentration of folding intermediate available for folding. Many enzymes are able to catalyze reactions by stabilizing the transition state of the reaction. In the folding process, the transition state is again substrate specific in that one of several intermediates may play the key role in progression to the native state. Cpn60 recognizes various intermediates depending on the substrate. Cpn60 binds the molten globule form of DHFR and rhodanese (Martin et al., 1991) and intermediates of pre-β-lactamase and Rubisco that have characteristics of molten globules (Zahn and Plückthun, 1992 and van der Vies et al., 1992). However, cpn60 recognizes only the unfolded form or weakly the first folding intermediate of lactate dehydrogenase (Badcoe et al., 1991). Even the number of substrates bound to cpn60 ranges from one to two polypeptides per cpn60 particle depending on the substrate itself. Regardless of how they function, chaperonins are catalysts for only a subset of proteins with which they interact.

**Chaperonins in Organellar Transport**

Most of the *in vitro* refolding reactions described above have used the bacterial cpn60 (GroEL) due to its ease of purification. However, the
chloroplast and mitochondrial cpn60s also participate in folding
(Goloubinoff et al., 1989b). One would assume that this role includes the
folding of proteins that are imported into the organelles. The chloroplast
cpn60 forms complexes with a large number of proteins upon import,
including the small subunit of Rubisco, the β-subunit of ATP synthase,
glutamine synthetase, and the light-harvesting chlorophyll a/b binding
protein (Lubben et al., 1989). Similarly, in the absence of ATP, newly
imported proteins form stable complexes with the mitochondrial cpn60
which are released upon readdition of ATP (Ostermann et al., 1989).
Mutants of the yeast mitochondrial cpn60 gene, such as mif4, have
allowed investigators to further characterize cpn60’s role in transport. The
mif4 mutant can undergo protein transport but is defective in the folding
and/or assembly of newly imported proteins (Cheng et al., 1989).
Translocation into the matrix is dependent on the mitochondrial Hsp70.
Proteins remaining in the matrix are transferred from Hsp70 to cpn60 in
order to facilitate folding and assembly (Cheng et al., 1989 and Ostermann
et al., 1989). Not only is cpn60 required for the assembly of other
proteins, but it is also required for the folding and assembly of itself upon
import into mitochondria (Cheng et al., 1990), which could raise the old
argument about the chicken and the egg. The same requirement is not
true for cpn10 (Mascagni et al., 1991).
Sorting to the intermembrane space also requires both Hsp70 and cpn60 (Cheng et al., 1989) in cases where precursor proteins contain bipartite N-terminal targeting sequences. A typical positively charged presequence directs the protein into the matrix, followed by an export targeting sequence similar to the bacterial signal sequences which contain a positively charged N-terminus, a hydrophobic core, and a polar carboxyl terminus (Randall and Hardy, 1989). The presence of the export sequence prolongs the interaction of the protein with cpn60 after import (Koll et al., 1992). The hydrolysis of ATP is then required to release the protein for export to the mitochondrial intermembrane space. Cpn60 may therefore have a dual function in translocation: mediating the folding of proteins entering the matrix and preventing folding or aggregation of proteins destined to undergo further translocation (Koll et al., 1992). The protection of proteins destined for export to the intermembrane space may be necessary since import and export are not tightly coupled and a pool of proteins may transiently build up while awaiting further translocation (Cheng et al., 1989).

**Coupled Action of Hsp70 and Chaperonins**

Genetic and biochemical data have suggested a cooperation between Hsp70 and cpn60 in the transport and folding of proteins entering the mitochondria. Langer et al. (1992a) have further characterized the successive action of Hsp70 and cpn60 by developing an *in vitro* refolding
system that may mimic the folding pathway of newly synthesized or imported proteins. Model proteins were used to reconstitute this system. They included the *E. coli* Hsp70 (DnaK); the *E. coli* protein DnaJ, which normally works in conjunction with DnaK; GrpE, another heat shock protein which also interacts with DnaK and DnaJ; the *E. coli* chaperonins GroEL and GroES; and denatured rhodanese which acted as a substrate for the system. Denatured rhodanese aggregates when diluted from denaturant into buffer alone. High amounts of either DnaK or DnaJ suppress this aggregation, whereas combined they act synergistically to completely suppress aggregation. ATP hydrolysis normally results in the release of substrates from Hsp70 (Munro and Pelham, 1986), however, unfolded rhodanese (rho) forms a complex of approximately 230 kDa of unknown stoichiometry with DnaK and DnaJ that is stable even in the presence of ATP. Rhodanese stabilized by DnaK/DnaJ or DnaK alone did not fold to an active conformation. GrpE strongly stimulates the ATPase activity of DnaK when DnaJ is also present (Liberek *et al.*, 1991). Its addition to the rho-DnaK-DnaJ complex allowed folding to occur at a very slow rate in the presence of ATP, reaching a maximum of 30% after 8 hours. Although rhodanese is recognized by cpn60, the addition of cpn60 and cpn10 to the complex did not result in reactivation in the absence of GrpE. Addition of GrpE resulted in a rapid folding of rhodanese to an active conformation. These results were extrapolated into a model for
protein import *in vivo*. DnaK binds to extended sequences emerging from the ribosome or from the inner mitochondrial membrane. The nascent chain then adopts a more collapsed conformation, but its release from DnaK is limited by the addition of DnaJ to the complex. Transfer of the polypeptide to cpn60 requires MgATP and GrpE as a coupling factor. The final folding is then mediated by cpn60 in an ATP-dependent reaction requiring cpn10. In order for this model to be applicable for mitochondria or chloroplasts, homologs of DnaJ and GrpE must be present in the organelles. Eukaryotic homologs of DnaJ have been discovered (Blumberg and Silver, 1991), but GrpE homologs have not yet been identified.

The need for two different chaperones and associated proteins in this system may not be obvious since cpn60 and cpn10 facilitate the folding of rhodanese in the absence of any other chaperone. However, Hsp70 and cpn60 differ in their abilities to recognize substrates. Hsp70 binds to small peptides (Flynn *et al.*, 1989) or proteins (Langer *et al.*, 1992a) in an extended conformation (Landry *et al.*, 1992) which may be necessary for the recognition of nascent chains emerging from the ribosome. In contrast, cpn60 often recognizes a molten globule conformation (Martin *et al.*, 1991) and binds peptides, albeit very weakly, in a helical conformation (Landry *et al.*, 1992). The combined efforts of these chaperones may allow many different proteins and conformations to be recognized.
Cytosolic Chaperonin

The bacterial cpn60 is clearly required for folding and assembly of proteins within the bacterial cell, as are the mitochondrial and chloroplast cpn60s within their respective organelles. Purified bacterial cpn60 can form stable complexes with over half of the soluble proteins in *E. coli* after their denaturation with 5 M guanidinium hydrochloride (Viitanen *et al.*, 1992a). This promiscuous binding of unfolded proteins leads one to ask whether cpn60 interacts with all newly synthesized proteins and, if so, what fulfills this role within the cytosol of eukaryotic cells. Although the extent of cpn60’s role in folding of newly synthesized proteins is unknown, a cytosolic homolog of cpn60 has been identified. Two different groups noted a weak, but significant, similarity between 11 cpn60 proteins and a protein called the t-complex polypeptide1 (TCP-1) (Ellis, 1990b and Gupta, 1990). This polypeptide is a product of a gene carried in the t locus of chromosome 17 of mice and is associated with changes in spermatogenesis (North, 1991). TCP-1 is present in the cytosol of all cells and homologs exist in other mammals, yeast, *Drosophila* and pea. A monoclonal antibody against TCP-1 recognized a cytosolic protein in pea that is composed of 62 kDa subunits (Ellis, 1990b). Antibodies against the *E. coli* chaperonin GroEL also recognized a cytosolic protein in oats (Grimm *et al.*, 1991).
The first real evidence for a cytosolic chaperonin was a report by Trent et al. (1991). This group found an abundant heat shock protein of the thermophilic bacterium *Sulfolobus shibatae* (TF55) to be very similar in sequence to TCP-1. Thermophilic factor 55 (TF55) has an ATPase activity and binds unfolded proteins *in vitro*. Structurally, TF55 also resembles TCP-1 (Lewis et al., 1992), and is reminiscent of cpn60 in that it consists of two stacked rings of 8-9 subunits. Further characterization of TCP-1 revealed that newly synthesized chains of *α*- and *β*-tubulin become transiently bound to a 900 kDa particle in rabbit reticulocyte lysates in a protease-sensitive conformation (Yaffe et al., 1992). Hydrolysis of ATP releases the tubulin subunits as functional *α-β*-tubulin heterodimers. The major component of this 900 kDa particle is the 58 kDa protein that cross-reacts with a monoclonal antibody against TCP-1. Purification of the TCP-1 particle from mouse cell homogenates revealed eight polypeptides ranging from 72 - 53 kDa in addition to TCP-1 (Lewis et al., 1992 and Gao et al., 1992). At least three of the polypeptides in the 50 kDa range that make up the hetero-oligomeric ring have homology to the major 58 kDa protein (Frydman et al., 1992). Two additional polypeptides cross-reacted with anti-Hsp70 antibodies (Lewis et al., 1992). This suggests that the cytosolic chaperonin, TCP-1, may participate in the folding and assembly of proteins in conjunction with members of the Hsp70 family. Frydman et al. (1992) demonstrated the ability of the purified hetero-oligomeric
complex to promote the \textit{in vitro} ATP-dependent refolding of luciferase, a task that the GroEL and GroES system cannot accomplish.

**Future Directions**

Although research in the chaperonin field has progressed at an extremely fast rate, many more questions must be addressed. The question of how cpn60 interacts with cpn10 and substrates is beginning to be addressed (Langer \textit{et al.}, 1992b). It appears that the GroES ring binds asymmetrically to either end-surface of the GroEL cylinder which causes marked structural changes at the opposite end of the cylinder. Substrates of GroEL seem to be accommodated within the central cavity of the cpn60 cylinder, but can be crosslinked to both GroEL and GroES in the GroEL-GroES-substrate complex formed in the presence of non-hydrolyzable ATP analogues (Bochkareva and Girshovich, 1992). But how does GroES exerts its allosteric effects on GroEL and what effect does the conformational change have on binding to a substrate? Is cpn10’s role \textit{in vivo} also substrate-specific? The list of questions are seemingly endless. In the research described in this dissertation, I have investigated which specific features of a substrate are recognized by GroEL. The results presented here will help us to determine what universal features may be presented in unfolded proteins that targets them for binding to GroEL.
CHAPTER 1

Identification of a Region in a Chloroplast Precursor Protein Required for Binding the *E. coli* Chaperonin GroEL
A large array of cellular machinery that interacts with unfolded and newly synthesized proteins has recently been discovered. Various components of this machinery can assist in the immediate folding of these proteins, deliver them to their appropriate compartments, and even promote proline isomerization or the formation of disulfide bonds. One class of proteins that performs several of these functions is the chaperonin family. Members of this family include the bacterial protein GroEL (Hemmingsen et al., 1988), the chloroplast Rubisco binding protein (Hemmingsen et al., 1988), and the hsp60 protein from mitochondria (McMullin and Hallberg, 1988 and Reading et al., 1989). These proteins consist of subunits with a mass of approximately 60 kDa and thus have been renamed chaperonin-60 (cpn60). The E. coli, chloroplast and fungal mitochondrial cpn 60s are composed of 14 subunits (Hendrix, 1979, Pushkin et al., 1982, McMullin and Hallberg, 1988), while the mammalian mitochondrial cpn60 is functional with only 7 subunits (Viitanen et al., 1992b). GroEL interacts in an ATP-dependent manner with another E. coli protein GroES (chaperonin-10) (Chandrasekhar et al., 1986). Both GroEL and GroES are encoded by the groE operon and were first discovered due to their requirement in bacteriophage head assembly (Georgopoulos et al., 1973). Mitochondria also contain a homologue of chaperonin-10 (Lubben et al., 1990). The chaperonins appear to stabilize unfolded or partially folded protein structures and prevent the formation of aggregates or other
aberrant structures which are off the folding pathway. In organelles, these molecules also play a role in protein transport, presumably binding to proteins as they enter the organelle. Although cpn60 binds its substrate in the absence of other proteins, (Lecker et al., 1989), the proper folding and/or release of the substrate requires ATP hydrolysis and in most cases the presence of chaperonin-10 (Goloubinoff et al., 1989b, Martin et al., 1991). One chaperonin molecule (or 14 subunits) binds only one or two substrate molecules (Martin et al., 1991, Laminet et al., 1990). This implies that either the subunits of cpn60 interact to make up one binding site or that 14 binding sites exist but binding to multiple substrates is limited by steric hindrance.

Little is known about the exact features of a substrate which are recognized by chaperonins. Using lactate dehydrogenase as a substrate, Badcoe et al. obtained evidence that the E. coli chaperonin binds the completely unfolded or earliest folding intermediate, but not to the molten globule form of this protein (Badcoe et al., 1991). On the other hand, Martin et al. (1991) suggested that this same chaperonin binds rhodanese and DHFR in their molten globule conformations. The earliest folding intermediate of dimeric Rubisco also binds GroEL (van der Vies et al., 1992). This species has considerable alpha helical content and has been referred to as a molten globule. The authors point out, however, that
unfolded states of any given protein are likely to be in rapid equilibrium with the molten globule state.

A vast array of proteins are substrates for cpn60. Substrates known to bind chaperonins in vivo include the NifA transcriptional activator and the protein products of the NifH and NifDK genes (Govezensky et al., 1991), whereas pre-β-lactamase, chloramphenicol acetyltransferase (Bochkareva et al., 1988), and over 50% of the soluble E. coli proteins unfolded by treatment with guanidinium hydrochloride (Viitanen et al., 1992a) will bind to GroEL in vitro. The alpha subunit of F1-ATPase, the beta subunit of the thylakoid ATP synthase, glutamine synthetase and several other proteins have been shown to associate with chaperonins upon transport into mitochondria (Prasad et al., 1990) or chloroplasts (Lubben et al., 1989). In addition, a requirement for the presence of cpn60 (± cpn10) for refolding in vitro has been established for rhodanese (Martin et al., 1991), citrate synthase (Buchenr et al., 1991), DHFR (Viitanen et al., 1991) and dimeric Rubisco (Goloubinoff et al., 1989). Clearly this vast array of cpn60 substrates must contain some motif that differentiates them from their folded counterparts.

The precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase (pS) is yet another substrate for chaperonins. After transport into the chloroplast pS (or S) transiently associates with the chloroplast cpn60 before assembly with the large subunit into the holoenzyme (Ellis
and van der Vies, 1988, Lubben et al., 1989). In addition, pS binds the bacterial cpn60 (GroEL) with high avidity *in vivo* when expressed as a C-terminal fusion to protein A (Landry and Bartlett, 1989). In the studies reported here, we used truncations of the protein A/pS fusion to delimit the smallest portion of pS that binds GroEL with high avidity, and then constructed a number of site-specific mutations in this truncation. Most of the mutants retained the ability to bind GroEL with high avidity, but two of them had a greatly diminished ability to bind the chaperonin.

**Materials and Methods**

*Construction of pS truncations* -- All manipulations of DNA were performed as described by (Landry and Bartlett, 1989). The construction of plasmid PAxpS has been previously described (see Fig. 1.3 and Landry and Bartlett, 1989). The truncations of the pS gene were made using the vector PAxpSCAM, which expresses a tripartite fusion consisting of protein A, pS and calmodulin, so that the Xba site in the 3' non-coding portion of the CAM sequence (Putkey et al., 1983) could be utilized for truncating pS.

The truncations PAxpS51T and PAxpS105T were made as follows. PAxpSCAM was digested with Ball or Smal, respectively, and ligated to a Xba adapter, 5'TAGGATAGT 3' and 3'ATCCTATCAGATC 5'. The DNA was then digested with Xbal and religated. Since pRIT2T (Pharmacia), the parent plasmid of PAxpS, contains the restriction sites utilized for
constructing PAxpS71T and PAxpS122T, these truncations were constructed in a derivated of pSP65, pSCAM, containing the pS cDNA sequence from pW9 (Broglie et al., 1983) fused to the cDNA sequence of calmodulin. The plasmid was digested with either EcoO109 or ApaLI, treated with mung bean nuclease, ligated to the Xba adaptor, and then digested with Ncol. In the case of PAxpS122T, the fragment was ligated into the Ncol-Xbal cut PAxpSCAM vector. The fragment for PAxpS71T was first ligated to Ncol-Xbal cut pSCAM65 and subsequently moved into PAxpSCAM with Ncol and PstI. PAxpS78T and PAxpS90T were made using the Xho/Xba adapter, 5’TCGACTAGT 3’ and 3’GATCAGATC 5’.

pSCAM65 was digested with either Xhol and Xbal or Sall and Xbal. These vectors were then ligated with the Xho/Xba adapter. The truncations were then moved into PAxpS with Ncol and Pstl.

The plasmid PAxpS-X/N was constructed in order to make intramolecular religations possible when creating truncations. The RIT2T plasmid (Pharmacia) was digested with EcoRV, ligated with the Xbal stop codon linker (NEB), digested with Xbal, and religated. In a similar manner a NotI site was introduced at the Pvull site of pRIT2T. The pS gene and Factor Xa site were cloned into pRIT2TX/N with the EcoRI-Pstl fragment from PAxpS. In order to construct PAxpS86T, the Ncol-Pstl fragment of pS65 was first isolated before digesting with NlaIV. It was then ligated with the Xbal stop codon linker, and digested with Xbal and Sphl. The
fragment was ligated with PAxpS-X/N which had been digested with Sphl and XbaI. All subsequent truncations of other constructs or site-specific mutants were made in PAxpS-X/N utilizing the Xho/Xba adapter and intramolecular religations.

**Other constructs** -- The plasmid PAxS+52-72 was made as follows. PAxS (Landry and Bartlett, 1989) was digested with Eco0109, treated with mung bean nuclease, and then digested with EcoRI. This fragment was cloned into PAxpS-X/N which had been digested with Mscl and EcoRI. To create PAxpSΔ1-28, PAxpS-X/N was digested with Ncol, treated with mung bean nuclease, digested with Pvull and religated. PAxpSΔ29-47 was constructed as follows. PAxpS-X/N was digested with Pvull and ligated with the MET-containing Sphl linker (NEB). The DNA was then digested with Sphl and religated.

**Site-specific mutants** -- The HCH class of mutations which replaced hydrophobic amino acids with charged amino acids were made using two complementary oligonucleotides which span the region between the Eco0109 and XhoI sites in the pS sequence

\[(5'\text{GCC(T/G)AC(T/G)TAAGCAGG(T/A)CGACTACC} \right.\]
\[(\text{T/G)TA(T/G)ACGCTCAAAGTGGG(T/A)TCCCTGCC} 3' \right.\] and

\[3'\text{G(A/C)TG(A/C)ATTCGTCC(A/T)GCTGATGG(A/C)AT(A/C)} \left.\text{TGCGAGGTTCCACC(A/T)AGGGACGAGCA} 5'\).\] At six separate sites an equal amount of the wild type and mutant base were included to
allow generation of a total of 64 different mutants. This set of oligonucleotides was cloned into the EcoO109 and Xhol sites of pS65. The transformation was not plated, but grown overnight in 3 ml of LB. The DNA was isolated and retransformed into *E. coli* DH5α. DNA was isolated from single colonies and analyzed by restriction analysis. At each possible mutation site the presence or absence of a restriction enzyme site indicated whether the site contained the wild type or mutant base.

All other site specific mutations were made using the Altered Sites Mutagenesis System (Promega). The pS gene was cloned into the pSelect phagemid vector by digesting PAxpS with EcoRI and PstI and moving this fragment into pSelect which had been digested with the same enzymes. The mutagenic oligonucleotides used to create the β-turn mutants were as follows:

T 1: 5’ CTTGATCCCGCAAGTGGGTCCCCTGCCTC 3’
T 2: 5’CTTGATCCCGGCAAGTGGGTG 3’
T 3: 5’ CTTGATCCCGCCCACTGGGTCCCTGCCTC 3’.

The mutagenic oligonucleotides used to create the proline mutants were:

P 2 and 3: 5’ CGGAGGCCCCCTTGAAGCCGGTCCCCTACTTGAT3’
P 3’: 5’ GGTCGACTACTTGATACCCTCCAAGTGGG 3’.

PAxpSP2 and P3, although made with the same mutagenic oligonucleotide, differ in the first mutagenic position which was wild type in the case of PAxpSP2. PAxpSP3 was used in the starting vector in
order to make PAxpSP3’. The oligonucleotide used to create PAxpSP3 was used to make PAxpSP4 utilizing PAxpSP3’ in the starting vector.

The Glycine mutants were made by first utilizing a complementary adapter that spanned the Eco0109 and HincII sites of pS. It was cloned into Sn5-N/S (constructed as described for PAxS (24) in the vector pS-pSP65) which lacks the transit sequence. This mutant contained 3 Glycine changes and destroyed the HincII site. The PAxpSG4 and G5 mutants were made by the altered sites mutagenesis method using PAxpSG3 as the starting vector. The mutagenic oligonucleotides that were utilized are:

G4: 5’ GGGGCAGGGCGGTACTTGATCCGCTCC 3’, and
G5: 5’ GGGCAGGGCGACGGCGGGATCCGCTCCAAG 3’.

The serine and glutamine mutants were also constructed utilizing the altered sites mutagenesis method and the following mutagenic oligonucleotides.

S1:5’ CGGAGGCCCTCTCGTCGCCAGTCCGACTCTCGATCCGCTCC AAG 3’ ; Q4: 5’ CCTCTTGAAGCAGCAGGACCAGCAGCAGCGCTC CAAGTGGG 3’ ; Q4’: 5’ CCTCTCCACGGAGCAGCAGCAGAAGCAG CAGGACTACTTGATCC 3’ ; Q5: 5’ GACTACTTGATTCAGTCCAAG TGGGTG 3’.
The sequence of each mutant construct was confirmed using double-stranded templates and the dideoxy chain termination method (Fawcett and Bartlett, 1990).

**Expression of Protein A fusions** -- Protein A or fusion proteins isolated from *E. coli* cells expressing the appropriate plasmids were bound to IgG agarose essentially as described (Landry and Bartlett, 1989) except that the buffer (Buffer A) in which the cells were resuspended and with which the column matrices were washed consisted of 50 mM Tris-HCl, 50mM NaCl, pH 7.4 plus 0.05mg/ml leupeptin. A low salt buffer was used to reduce nonspecific binding due to hydrophobic interactions. We found that elution of the fusions from the agarose with acid as described previously resulted in retention of some GroEL on the column, possibly due to precipitation. In order to elute all of the GroEL bound to the fusions, the elution was carried out using 2 washes of 7M urea followed by 3 washes of 0.5M HAc adjusted to pH 3.4 with ammonium acetate. One set of fusion proteins eluted in this manner was retained in order to assess the amount of GroEL bound *in vivo*. The second set of fusion proteins bound to IgG-agarose were incubated with 50mM Tris-HCl, pH 7.4, 50mM NaCl, 10mM MgCl₂, 5mM KCl, 10mM ATP and either GroES or casein in a 10-fold molar excess. The mixture was rocked for 30 min and then the matrix was washed batchwise two times with 15 ml of buffer A. An approximate 2-fold molar excess of purified GroEL was then added and
incubated 30 min while rocking. The column matrices were again washed 2 times with buffer A and the fusion and GroEL eluted as above. The eluted proteins were precipitated with 10% TCA, washed with ice-cold acetone and redissolved in 0.1M carbonate-DTT for analysis by SDS-PAGE (Landry and Bartlett, 1989).

Protein purification -- GroEL was purified from DH5a E. coli cells transformed with pOF39 which expresses the entire groE operon (Fayet et al., 1986). The cells were grown to late stationary phase at 37°C, centrifuged and resuspended in 0.1 volume of buffer A. The cells were lysed on ice by sonication and the cell debris was removed by centrifugation at 10,000xg for 10 min. The 41-56% saturated ammonium sulfate cut was resuspended in 50mM Tris, pH 7.4, 50mM NaCl and heated to 50°C for 30 min. The denatured protein was removed by centrifugation at 10,000xg for 15 min and the soluble protein was fractionated on a Sephacryl S-200 column. The GroEL eluted in the void volume and was subsequently applied to a Sepharose S-400 column. The GroEL peak was then applied to a Cibacron Blue agarose column (Sigma). GroEL fractions were pooled and stored in 10% glycerol at -20°C.

GroES was purified from cells harboring the plasmid pOF39ΔKpn. (This deletion results in overexpression of GroES without concomitant overexpression of GroEL and was constructed by digesting pOF39 with KpnI and religating the vector). The cells were grown to late stationary at
37°C, centrifuged and resuspended in 0.1 volume of buffer A. The cells were lysed by sonication, and the lysate was subjected to centrifugation to remove cell debris. The supernatant was fractionated with ammonium sulfate and the 35-50% saturated ammonium sulfate cut was dialyzed against 50mM Tris, pH 7.4, 150 mM NaCl. The samples were heated to 50°C for 20 min. The denatured proteins were removed by centrifugation and the supernatant was heated to 90°C for 20 min in 1ml aliquots, cooled, and centrifuged again. GroES was pooled and stored in 10% glycerol at -20°C.

Results

GroEL can be specifically removed from and bound to PAxpS in vitro

Protein A-pS (PAxpS) fusions were isolated by incubating IgG-agarose with lysates from *E. coli* cells expressing the appropriate plasmids. The column matrix was then extensively washed with buffer. Protein which was removed from the columns with urea at this stage gave us an estimate of the amount of GroEL bound *in vivo*. This has been shown previously (Landry and Bartlett, 1989) and is repeated in lanes 1 and 3 of Figure 1.1 for protein A and PAxpS. In the case of protein A, the only other protein which is bound to the column is DnaK. The IgG-agarose column matrix alone bound neither DnaK nor GroEL, therefore DnaK must be specifically binding to protein A. PAxpS binds both DnaK and GroEL (Landry and Bartlett, 1989). GroEL was partially eluted from PAxpS with
an excess of MgATP and K\(^+\) as shown in lane 4. However, removal of almost all of the GroEL required a 10-fold excess of GroES or casein (lanes 5 - 6). The large excess of casein or GroES required for complete removal of GroEL likely reflects the fact that the pS moiety does not require prior denaturation to bind GroEL and does not obtain a state that inhibits GroEL binding. Therefore, GroEL can continually rebind to PAxpS. Once the GroEL that co-purified with the PAxpS was removed, purified GroEL could bind the fusion in the absence of K\(^+\), MgATP, or GroES (lane 7). These components are not necessary for binding of GroEL to an unfolded substrate but are required for the subsequent refolding of that substrate (Viitanen et al., 1990). GroEL did not bind to protein A alone, even when added in excess. The PAxpS fusion bound GroEL in vitro in a ratio of roughly 14 GroEL subunits per molecule of fusion. GroEL is assembled into a double doughnut of 14 identical subunits, therefore approximately one mole of GroEL was bound per mole of fusion. Similarly, rhodanese and \(\beta\)-lactamase require a 1:1 molar ratio of GroEL:substrate to prevent aggregation or folding, respectively (Martin et al., 1991, Laminet et al., 1990) and GroEL binds 1 - 2 molecules of DHFR (Martin et al., 1991).

Deletion of the transit sequence results in a decrease in bound GroEL

Deletion of the pS transit sequence results in increased expression of the PAxS chimeric protein (Landry and Bartlett, 1989) and a GroEL:fusion
Figure 1.1 Removal of bound GroEL from PAxpS. The proteins Protein A (lanes 1 and 2) and PAxpS (lanes 3-7) were affinity purified from *E. coli* lysates. The GroEL and DnaK that copurified with these proteins is shown in lanes 1 and 3. Protein A was incubated with 0.3 nmoles of purified GroEL in lane 2. Lanes 4-6 represent the amount of GroEL remaining bound to PAxpS after incubation with: Buffer A + 5 mM K⁺ + 10 mM MgATP (lane 4); Buffer A + K⁺, MgATP, and a 10 fold excess of either GroES (lane 5) or casein (lane 6). Once GroEL was removed from the fusion, 0.3 nmoles of purified GroEL was allowed to bind, followed by extensive washing (lane 7). The proteins were separated on SDS-PAGE and stained with Coomassie Blue. Molecular weight markers (BRL) are shown on the left and correspond to the sizes shown in figure 6.
Figure 1.2. \textit{In vivo} and \textit{in vitro} binding of GroEL to the precursor and mature form of PAxpS. PAxpS (lanes 1 and 2) and PAxS (lanes 3 and 4) were expressed in \textit{E. coli} and affinity-purified. The GroEL that copurified with these fusion proteins is shown in lanes 1 and 3 and represents the amount of GroEL bound \textit{in vivo}. In lanes 2 and 4, bound GroEL was removed from the fusion proteins with MgATP and casein. They were then incubated with 1 n mole of purified GroEL for 30 min and washed extensively. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers (BRL) are shown on the left.
protein ratio that is much less than that observed with PAxpS. However, the reason for this decreased level of binding in vivo was not clear (Landry and Bartlett, 1989). To determine whether pS and S have similar avidities for GroEL, or whether the transit sequence contributes to the increase in GroEL binding to the precursor protein, we compared the abilities of PAxpS and PAxS to bind GroEL in vitro. Even when a large excess of GroEL was incubated with the fusion proteins, PAxS bound much less GroEL than PAxpS (Figure 1.2, lanes 2 and 4). These results clearly show that the transit sequence plays an important role in the binding of GroEL to PAxpS.

A 14 amino acid region is required for binding with high avidity

In order to delineate the region of pS that is important for GroEL binding, a series of C-terminal truncations of pS were constructed and expressed in E. coli. Each truncation is numbered for the position of the C-terminal amino acid of the protein (Figure 1.3). Amino acids are numbered starting from the first residue of the transit sequence. Figure 1.4 shows the in vivo binding for each of these truncations. As seen before, DnaK bound each fusion protein to the same extent. The other proteins seen in Figure 4 correspond to IgG, which over time will bleed off the column matrix, and to breakdown products which vary according to the stability of each fusion. These fusions are normally degraded to the size of protein A alone and therefore do not interfere with the measurements of GroEL bound to the Protein A-pS fusions. As the
Figure 1.3  C-terminal truncations of PAxpS. A. Plasmid map of PAxpS. B. The pS sequence is numbered starting from the first amino acid of the transit sequence. All the pS truncations occur as fusions to the C-terminus of protein A. Restrictions sites are abbreviated as follows: N, Ncol; P, Pvull; Sp, Sphl; B, Ball; E, Eco0109; S, Sall; V, NlaIV; X, Xhol; Sm, Smal; Ap, ApaLI. Those proteins that bind GroEL with high avidity are designated by (+), while those which bind with low avidity are designated by (-).
Figure 1.4  *In vivo* binding of GroEL to C-terminal truncations of PAXpS. C-terminal truncations of PAXpS were expressed in *E. coli* and affinity-purified. The amount of GroEL that copurified with PAXpS51T (lane 1), PAXpS71T (lane 2), PAXpS78T (lane 3), PAXpS90T (lane 4), PAXpS105T (lane 5), PAXpS122T (lane 6), and PAXpS (lane 7) is shown. The sizes of the truncations range from 35 to 48.6 kDa (full length). The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers are shown on the left.
truncated proteins became smaller, there was a sharp drop in the amount of GroEL bound to each fusion. PAxpS90T bound GroEL in a ratio similar to the full length pS whereas PAxpS78T, which is 12 amino acids shorter, bound GroEL in a ratio of only 1:1. We have made another truncation to further examine this region. PAxpS86T is 8 amino acids longer than PAxpS78T yet it bound GroEL in a stoichiometry similar to wild type pS (data not shown). With the exception of PAxpS51T and PaxpS122T, all truncations were also tested in *in vitro* rebinding studies, and showed the same pattern of GroEL binding (data not shown). Since PAxpS78T may truncate pS in the middle of a putative GroEL binding site, we expanded the region of interest to include 15 residues from amino acid 72 to 86. PAxpS86T is the shortest truncation that bound with the same avidity as full length pS but it expresses very poorly in *E. coli*. All site-specific mutations in this region therefore were analyzed as truncations to amino acid 90.

*Site-specific mutagenesis of the mature portion of the precursor protein*

Site-specific mutagenesis was used to determine what features of residues 72-86 of the precursor protein are required for GroEL binding. The first set of mutations replaced the most hydrophobic amino acids (leucine, valine, and isoleucine) with charged amino acids (Table II, HCH 2-5). These mutant polypeptides bound GroEL to the same extent as the wild-type PAxpS90T (Figure 1.5, lanes 3-6).
We next examined the ability of GroEL to bind substrates with altered secondary structures. The only such features that normally occur in full-length S in this region are an alpha helix and a beta turn (Knight et al., 1989). In order to destroy the $\beta$-turn, the middle two amino acids of the turn, serine and lysine, were replaced with amino acids which strongly disfavor turns such as alanine and histidine. These mutations had no effect on the ability of the pS fusion to bind GroEL (data not shown).

In order to destroy the alpha helix we made two sets of mutations which used proline or glycine as helix breakers. Proline and glycine residues effectively destabilize alpha helices, especially when multiple residues are present (Strehlow et al., 1991, MacArthur and Thornton, 1991, Pingchiang et al., 1990, O'Neil and DeGrado, 1990, Chakrabartty et al., 1991). The mutations containing proline substitutions were constructed so that proline replaced the amino acids that strongly promote helix formation singly and in combination. Additionally, the prolines were distributed uniformly throughout the region (Table II, Pro 2-4). Even the mutant protein that contained four prolines bound GroEL \textit{in vitro} to the same extent as wild type pS (Figure 1.6). Thus, an alpha helix (or a nascent alpha helix) is not required for substrate binding by GroEL. However, since prolines impart some rigidity to the amino acid sequence, we constructed a similar set of mutations substituting glycine for the amino acids that favor helix formation (Table II, Gly 3-5). The mutant
Table II

Sequence of Site-Specific Mutations Created in the Truncation PAxpS90T

<table>
<thead>
<tr>
<th>Sequence of pS amino acid 67 to 86</th>
<th>Hydrophobic Moment $\mu$(105)/9</th>
<th>Hydrophobicity (KD) b</th>
<th>Net Charge</th>
<th>High Avidity GroEL Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>W.T. PLSTEALLKQVDYLIRSKMV</td>
<td>2.50</td>
<td>0.137</td>
<td>+1</td>
<td>Yes</td>
</tr>
<tr>
<td>Gly5 ---------------GG-G-GG---------</td>
<td>0.98</td>
<td>-0.221</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>Gly4 ---------------GG-GG------------</td>
<td>1.32</td>
<td>0.097</td>
<td>+1</td>
<td>Yes</td>
</tr>
<tr>
<td>Gly3 ---------------GG-GG------------</td>
<td>1.36</td>
<td>-0.085</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>Ser5 ---------------SS-S-S-----------</td>
<td>0.99</td>
<td>-0.421</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>Gln4 ---------------Q-QQQ-----------</td>
<td>0.85</td>
<td>-1.19</td>
<td>+1</td>
<td>Yes</td>
</tr>
<tr>
<td>Gln4* ---------------QQQ--Q----------</td>
<td>1.37</td>
<td>-1.32</td>
<td>+1</td>
<td>Yes</td>
</tr>
<tr>
<td>Gln5 ---------------QQQ--Q--Q-------</td>
<td>1.27</td>
<td>-1.26</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>Pro2 -----------------P--P------------</td>
<td>2.39</td>
<td>0.336</td>
<td>+2</td>
<td>Yes</td>
</tr>
<tr>
<td>Pro3 -----------------P--P--P---------</td>
<td>2.39</td>
<td>0.053</td>
<td>+2</td>
<td>Yes</td>
</tr>
<tr>
<td>Pro3* ----------------P--P--P--P------</td>
<td>2.13</td>
<td>0.105</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>Pro4 -----------------P--P--P--P------</td>
<td>2.09</td>
<td>0.205</td>
<td>+1</td>
<td>Yes</td>
</tr>
<tr>
<td>HCH2 -----------------R--D--RR-------</td>
<td>0.92</td>
<td>-2.02</td>
<td>+2</td>
<td>Yes</td>
</tr>
<tr>
<td>HCH3 -----------------RR--R----------</td>
<td>1.93</td>
<td>-1.21</td>
<td>+4</td>
<td>Yes</td>
</tr>
<tr>
<td>HCH4 -----------------RR--RR---------</td>
<td>2.41</td>
<td>-0.737</td>
<td>+3</td>
<td>Yes</td>
</tr>
<tr>
<td>HCH5 -----------------RR--D--R--------</td>
<td>0.92</td>
<td>-1.58</td>
<td>+3</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Calculated according to Eisenberg et al. (84) using the hydrophobicity scale of Kyte and Doolittle (82) and averaged over 9 amino acids. These are indicated by a bar and correspond to the region with the highest hydrophobic moment.

Calculated according to Kyte and Doolittle (82). The number reported is the average over all amino acids shown.
Figure 1.5 Replacement of hydrophobic amino acids with charged residues (HCH) in the truncation PApS90T. Site-specific mutations (shown in Table II, HCH 2-5) of the putative binding region of pS were made in the truncation PApS90T. PApS (lane 1), PApS90T (lane 2), PApS90THCH2 (lane 3), PApS90THCH3 (lane 4), PApS90THCH4 (lane 5), and PApS90THCH5 (lane 6) were expressed in E. coli and affinity-purified. Bound GroEL and DnaK was removed from the fusion proteins with MgATP and casein. They were then incubated with purified GroEL for 30 min and washed with buffer. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers are shown on the left. Arrows indicate the position of GroEL and PApS90T and its derivatives.
Figure 1.6 Destabilization of a helix by substitution of amino acids with proline in the truncation PAxpS90T. Site-specific mutations (shown in Table II, Pro 2-4) were constructed in the 90 amino acid truncation of PAxpS. As in Fig. 4, PAxpS90TP2 (lane 1), PAxpS90TP3 (lane 2), PAxpS90TP3* (lane 3), and PAxpS90TP4 (lane 4) were expressed in *E. coli* and affinity-purified. The bound GroEL was removed and the fusion proteins were incubated with purified GroEL. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. The single arrow represents the position of PAxpS90T derivatives whereas the double arrow corresponds to the position of protein A breakdown products.
Figure 1.7 Identification of a binding domain by substitution of amino acids with glycine in the truncation PAxpS90T. The wild-type protein, PAxpS90T, (lane 1) and the mutant proteins PAxpS90TG3 (lane 2), PAxpS90TG4 (lane 3), and PAxpS90TG5 (lane 4) were expressed in *E. coli* and affinity purified (see Table II, Gly 3-5). As in Fig. 4, the bound GroEL was removed and purified GroEL was allowed to bind. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers are shown on the left.
proteins PAxpS90TG4 (Gly4) and PAxpS90TG5 (Gly5) cluster 4 or 5 glycines in the middle of the region that contains the greatest number of amino acids favoring alpha helix formation. Both PAxpS90TG3 (Gly3) and PAxpS90TG4 bind GroEL in a ratio of approximately 1 pS derivative to 14 GroEL polypeptides, but the ability of PAxpS90TG5 to bind GroEL is reduced dramatically (Figure 1.7).

The only discernable difference between PAxpS90TG5 and the numerous other mutant proteins described above is the absence of a strong helical hydrophobic moment in the region of residues 72-86 regardless of whether Eisenberg et al. (1984) or Kyte and Doolittle’s (1982) hydropathy scale, which differ mainly in the values assigned to amino acids with very hydrophilic side chains, was used.

In light of the failure of the PAxpS90TG5 substrate to bind GroEL as avidly as the wild-type protein or any of the other mutant proteins, we constructed four additional mutations that also alter the hydrophobic moment of the protein in this region. The first of these replaces the same amino acids that were mutated in the PAxpS90TG5 with serine. Like proline and glycine, serine occurs infrequently in helices but serine is intermediate between these two amino acids in the restriction of its backbone movements (Ramachandran and Sasisekharan, 1968). In the other three mutations glutamine, which is frequently found in alpha helices (Levitt, 1978), replaced specific hydrophobic amino acids thereby
Figure 1.8 Destruction of the hydrophobic moment by substitution of amino acids with serine and glutamine in the truncation PAxpS90T. The wild-type protein, PAxpS90T, (lane 1) and the mutant proteins PAxpS90TS5 (lane 2), PAxpS90TQ4 (lane 3), PAxpS90TQ4* (lane 4), and PAxpS90TQ5 (lane 5) were expressed in E. coli and affinity-purified (see Table II, Ser5 and Gln4-5). As in Fig. 4, the bound GroEL was removed and purified GroEL was allowed to bind. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers are shown on the left.
Figure 1.9 Transit sequence deletions in the 90 amino acid truncation of PAXpS. A, structure of the deletions made in the transit sequence of PAXpS90T. The hatched area in PAXs(+52-72)90T is repeated as a spacer of 20 amino acids. Restriction sites are labelled as in Fig. 3. Those proteins which bind GroEL with high affinity are designated by (+), while those binding with low affinity are designated by (-).
Figure 1.9 Transit sequence deletions in the 90 amino acid truncation of PAxpS. B, PAxpS90T (lane 1), PAxpS(Δ1-28)90T (lane 2), PAxpS(+52-72)90T (lane 3), and PAxpS(+52-72)90TG5 (which contains the Gly 5 mutation) (lane 4) were expressed in E. coli and affinity-purified. As in Fig. 4, the bound GroEL was removed and purified GroEL was added. The proteins were separated by SDS-PAGE and stained with Coomassie Blue. Molecular weight markers are shown on the left.
destroying the amphipathic nature of the helix. The sequences for all four mutations are found in Table II. However, when expressed in *E. coli*, affinity purified, and examined for the ability to bind GroEL *in vitro*, only PAxPS90TS5 bound significantly less GroEL than the wild type protein (Figure 1.8). This argues against the idea that a hydrophobic moment is essential to GroEL binding.

*Deletions in the transit sequence of pS*

We further examined the effect of the transit sequence on the high avidity binding of GroEL to pS by constructing two deletion mutants. One contained only the first 28 amino acids of the transit sequence (PAxPS\(\Delta\)29-47) and the other contained only the last 20 amino acids of the transit sequence (PAxPS\(\Delta\)1-28) as shown in Figure 1.9A. Each of these mutant proteins retained the ability to bind GroEL *in vitro* with high avidity (Figure 1.9B, lane 2 and data not shown). To rule out the trivial explanation that the transit sequence simply serves as a spacer between the mature protein and protein A which allows avid binding of GroEL, a mutant was constructed that lacks the entire transit sequence and repeats the first 20 amino acids of the mature protein (PAxS(+52-72)90T). This mutant contains the same number of amino acids separating the mature protein from protein A as one of the mutants with half of the transit peptide deleted, but its ability to bind GroEL is drastically reduced (Figure 1.9B, lane 3). Furthermore, when the Gly5 mutation is placed in the
context of PAxS(+52-72)90T, the resulting fusion protein (PAxS(+52-72)90TG5) binds no GroEL in vitro (Figure 1.9B, lane 4). These results suggest that the transit sequence contains two functionally redundant regions in terms of its role in GroEL binding and does not serve simply as a spacer to separate the mature S protein from the protein A moiety. Since chloroplast transit sequences contain many of the same features (high serine and threonine content along with a net positive charge) throughout the length of their sequence (Keegstra et al., 1989), it may not be surprising that either half of the transit sequence can facilitate binding to the chaperonin.

Discussion

In the studies described here we have shown that the PAxpS fusion protein exhibits many of the same binding characteristics as substrates used in GroEL refolding studies in vitro. One molecule of PAxpS binds one GroEL molecule (or 14 subunits). GroEL is removed from the fusion in an ATP-dependent manner. GroES is also required for efficient release but casein can replace GroES in this reaction, presumably acting as an alternate substrate for GroEL (Martin et al., 1991). Finally, purified GroEL binds very tightly to the fusion protein in vitro. The one limitation of the in vitro studies we describe here is a strong bias in favor of binding GroEL. We observed differences between some of the mutant proteins in their abilities to bind GroEL in vivo but these differences were reduced when
analyzed *in vitro*. This may reflect the different abilities of the mutants to effectively compete for GroEL with cellular substrates. Despite this slight bias, we can draw several conclusions from these studies.

The transit sequence, or a portion of it, is necessary but not sufficient for binding GroEL with high avidity (1 molecule of chaperonin per molecule of substrate) to the PAxpS fusion protein. The main difference between our binding assay and other refolding assays is that our substrate requires no prior denaturation in order to effectively bind to GroEL. This is due to the presence of the transit sequence and the nature of our substrate. Even in the presence of an excess of the chaperonin, PAxS bound less GroEL than PAxpS. When the transit sequence is deleted and the first 20 amino acids of S are duplicated in the fusion, much less GroEL is bound. Thus, the transit sequence does not act simply as a spacer between protein A and S. Bacterial signal sequences slow, but do not prevent, folding of the mature portion of the precursor protein to a native-like structure (Liu *et al.*, 1988). The transit sequence could have a similar effect on the folding of the mature portion of pS, thereby acting to destabilize structure and promote GroEL binding. However, since S cannot assemble into the holoenzyme under conditions used in our study, sites recognized by GroEL may still be exposed even if S is folded. Alternatively S may, in the absence of large subunits, be unstable and unfold to some extent.
Analysis of GroEL binding to the C-terminal deletions of S indicated that amino acids 72-86 of pS, which correspond to the first alpha helix in S assembled in the holoenzyme (Knight et al., 1989), are required for binding GroEL with high avidity. We constructed a number of site-specific mutations of PAxpS90T designed to alter secondary structure, charge or overall hydrophobicity of this region and assayed the mutant proteins for the ability to bind GroEL. Most of the mutant proteins retained the ability to bind GroEL with high avidity, but two of them, Gly5 (PAxpS90TG5) and Ser5 (PAxpS90TS5), bound much less of the chaperonin than the wild type truncation. In the absence of the transit peptide, the Gly5 protein bound no GroEL, reinforcing the notion that the transit sequence and the region of pS between amino acids 72 and 86 are both required for binding GroEL with high avidity. Results of folding studies using pre-β-lactamase might also suggest that GroEL interacts with both the signal sequence and the mature part of this bacterial precursor protein (Zahn and Plückthun, 1992).

What features of Gly5 and Ser5 distinguish them from the other mutant proteins and the wild type truncation? Examination of the hydrophobic moment, overall hydrophobicity, helical tendency and net charge of the region of pS and the various mutants between amino acids 72 and 86 revealed that Gly5 and Ser5 are not unique from the other mutations or wild type in any single parameter, but these are the only
constructs in which the probability of alpha helix formation (or structural rigidity), net positive charge and hydrophobicity are all lower than the other proteins.

The proline mutants show that, although some rigidity may be required, a helical structure is not an absolute requirement for GroEL binding. A helical hydrophobic moment in at least the context of the proline mutants cannot be the major contributor to GroEL binding. In addition, since GroEL binds to a protein containing only β-sheets (Schmidt and Buchner, 1992), multiple structural motifs may be recognized. GroEL contains several hydrophobic patches near its surface as determined by its ability to bind the hydrophobic fluorescent reporter, 1,1′-bi(4-anilino)naphthalene-S,5′-disulfonic acid (Mendoza et al., 1991a). Furthermore, a non-ionic detergent can replace GroEL in refolding rhodanese in vitro, suggesting that the chaperonin may provide a hydrophobic environment in which folding of proteins can occur (Tandon and Horowitz, 1986). It is therefore no surprise that hydrophobic residues may be required for binding. Further, since GroEL is very acidic with a pI of 5.63 (Hemmingsen et al., 1988) its affinity for positively charged sequences is also not completely unexpected. Finally, recent NMR data indicates that GroEL can promote helix formation in small peptides (Landry and Gierasch, 1991). GroEL may have some preference for a structural element, even if
this is in the form of structural rigidity induced by the presence of several proline residues.

We have used GroEL as a model chaperonin in this study to identify a binding site in a chloroplast precursor. Both the chloroplastic and mitochondrial chaperonins associate with newly imported proteins into their respective organelles. In addition, the mitochondrial and chloroplastic chaperonins can substitute for GroEL in the refolding of dimeric Rubisco (Goloubinoff et al., 1989) It is therefore very likely that GroEL recognizes regions of pS similar to those recognized by the chloroplastic chaperonin upon import of the precursor. The GroEL binding site identified is located in the mature region of PAxpS and is not dependent upon a specific amino acid sequence. The chaperonin may recognize a combination of overall hydrophobicity, net positive charge and structural rigidity or helical tendency. This kind of broad specificity is necessary since in vitro experiments have shown that over 50% of the soluble proteins from E. coli can interact with GroEL in their unfolded states (Viitanen et al., 1992a). Further studies will be required to determine the relative contribution of each of these features to binding of a chaperonin by PAxpS.
CHAPTER 2

Identification of a Protein-Protein Contact Site Between GroEL and a Chloroplast Precursor Protein
Molecular chaperones are characterized as a class of proteins which recognize and stabilize unfolded or partially folded protein intermediates during protein folding and/or translocation into organelles (Ellis, 1990). What each member of this class of proteins recognizes as unfolded or misfolded is still, however, unclear. The *E. coli* chaperonin, GroEL, can bind a large number of proteins as assayed by *in vitro* refolding reactions (Goloubinoff *et al.*, 1989; Martin *et al.*, 1991; Viitanen *et al.*, 1991) and other binding assays (Viitanen *et al.*, 1992), although a particular domain important for this interaction has not been identified for any of these proteins with the exception of the chloroplast precursor form of the small subunit of Rubisco (pS) (Chapter 1). We recently identified by site-directed mutagenesis a 15 amino acid region in the mature part of pS that is required for binding GroEL with high avidity. In order to determine what features of this region are necessary for this protein-protein interaction, we must first show that GroEL physically contacts this region of pS in a specific manner.

Protein-protein contact sites can be identified and examined by a number of methods. Landry *et al.* (1991) used NMR to identify interactions between GroEL and the side chains of small peptides. Although this technique allows for excellent detailed information on the structural characteristics of these interactions, it can only be used for the very weakest GroEL binding peptides (Kd ≈ 10-1000 µM). To identify
strong GroEL binding sites within larger proteins, other methods must be used. UV crosslinking was used to identify binding sites between the 54 kD subunit of the signal recognition particle (SRP) and the signal sequence of the preprolactin precursor (Zopf et al., 1990). However, for a large complex such as GroEL, with 14 identical subunits, such a task is tedious at best. Therefore we developed a novel protection assay to examine the nature of the GroEL-pS complex.

Materials and Methods

Expression of protein A fusions -- The construction of PAxpS90T and PAxpS78T has been previously described (Chapter 1). PAxpS(Δ62-72)90T was made in the following manner. The wheat pS cDNA was cloned into pSP65 (Promega) which contained a deletion of the two Drall sites. The corresponding DNA (pSn5) was digested with Drall and trimmed with Bal31 nuclease. The DNA was then digested with HindIII (at a site located at the 3' end of pS), and the insert that was removed was replaced with the corresponding fragment from pSn5 digested with Drall and HindIII. This deletion of 30 base pairs, 5' to the Drall site, was shown to be in frame by analysis of in vitro translation products and subsequently moved into the protein A vector with the restriction enzymes Ncol and PstI.

The protein A fusion proteins were purified from DH5α E. coli lysates of cells expressing the appropriate plasmids as described previously.
(Chapter 1). Briefly, the cells were centrifuged, resuspended in Buffer A (100 mM Tris, 50 mM NaCl, pH 7.4) plus 0.05 mg/ml leupeptin, and lysed by sonication. The clarified lysate was incubated batchwise for 30 min with 400 ul of IgG-agarose column matrix that had been previously iodinated with cold KI (see below). The excess lysate was removed and the column matrix was washed batchwise with buffer A. The DnaK and GroEL that copurifies with several of the fusion proteins was removed with an excess of casein in the presence of 10 mM MgATP and 5 mM KCl. The columns were again washed batchwise with buffer A. To form a complex with GroEL, the column was incubated for 30 minutes with a 3 molar excess of purified GroEL. The excess GroEL was removed by repeated washings and the fusion protein was ready for iodination.

Iodination conditions -- The IgG-agarose column matrix that was used to bind the fusion proteins was first iodinated with cold KI. This does not affect the ability of the matrix to bind the protein A fusions. Three IODO-BEADS were first washed with Buffer A and dried on Whatman paper. Each bead was incubated for 5 min at RT with 120 ul of buffer A plus 10 ul of 0.046 mM KI in 0.1 N NaOH. The IgG-agarose was then added to the beads and incubated for 10 minutes at RT while rocking. The columns were washed several times and then incubated with the clarified E. coli lysates.
Due to the sensitivity of GroEL to either the highly oxidative environment or the iodination of its tyrosines, the conditions used for iodination of the fusion proteins were a modification of the IODO-BEADS protocol suggested by the manufacturer (Pierce Chemical Co.). One IODO-BEAD was used for each reaction. The beads were first washed briefly in 1 ml of Buffer A, dried on Whatman paper, and then incubated with 195 ul of buffer + 0.5 mCi of Na\(^{125}\)I for 5 min at RT. The non-absorbed label was then removed and the column material was added to the IODO-BEAD in a total of 600ul of buffer. The mixture was incubated for 2 min at RT while rocking. The column material was then removed and placed in a disposable Econo-column (Bio-Rad) and immediately washed several times with buffer to remove free iodine. The fusions were eluted from the column with 2 washes of 7M urea followed by 3 washes of 0.5M HAc adjusted to pH 3.4 with ammonium acetate. The proteins were TCA precipitated and separated by SDS-PAGE. After staining the gel with Coomassie brilliant blue and drying, each fusion protein was excised from the PAGE gel and electroeluted from the gel slice in a buffer containing 50mM Tris, 0.384 M Glycine, and 0.1% SDS. Complete elution of the protein required 2.5 hours using a constant power of 3 Watts per gel slice. The eluted protein was then dialyzed overnight against 1 liter of 50mM ammonium bicarbonate, 50 mM β-mercaptoethanol, and 0.1% SDS.
Sample preparation for peptide mapping -- This protocol is essentially as described by Boyle et al. (1991) and the C.B.S. manufacturer’s manual for the HTLE-7000 system. The eluted proteins were precipitated from the ammonium bicarbonate buffer with 15% TCA using 20 ug of RNaseA as a protein carrier. The protein was pelleted, washed with ice-cold acetone, and then incubated with performic acid for 60 min at 0°C in order to oxidize all methionines and cysteines. Performic acid was prepared by incubating 900 ul of formic acid with 100 ul of hydrogen peroxide (33%) for 30 min at RT. The performic acid/protein mixture was rapidly diluted with cold water, frozen in a dry ice/EtOH bath and lyophilized using a Savant "Speed-Vac". The pellet was then resuspended in 50 ul of 50 mM ammonium bicarbonate, pH 8.0-8.3, and digested with 10 ug of TPCK-treated trypsin overnight at 37°C. The sample was vortexed for 1 min and another aliquot of trypsin was added and incubated further for 5 hours. When the digestion was complete, the samples were diluted with 400 ul of water, frozen and lyophilized. The pellet was washed once with 400 ul of water and again lyophilized to remove any traces of ammonium bicarbonate. The protein pellet was resuspended in pH 1.9 buffer (0.58 M formic acid, 1.36 M acetic acid) and vortexed for a full 2 min. Insoluble particulate matter was pelleted by centrifugation for 2 min in a microcentrifuge at RT. The supernatant was transferred to a new tube avoiding any debris. The sample was again lyophilized, resuspended in 10
ul of pH 1.9 buffer, and spotted on a thin layer cellulose plate (Merck).
The first separation dimension was high voltage electrophoresis. The plate
was prewet in pH 1.9 buffer to concentrate the sample spot and
electrophoresed for 40 min at 1000 Volts using the HTLE-7000 system.
The plate was dried for 30 min and transferred to a chromatography tank
containing 4.3 M n-butanol, 3.75 M pyridine, and 1.07 M acetic acid.
After 7 hours the plates were dried overnight and placed with film for 1
week.

Results and Discussion

The precursor of the small subunit of RuBisCO (pS) binds to GroEL
both in vitro and in vivo when expressed as a C-terminal fusion to protein
A (Landry and Bartlett, 1989; Chapter 1). Furthermore, a site in the
mature portion of pS was identified by C-terminal truncations and site-
specific mutagenesis to be required for this interaction (Chapter 1). The
smallest truncation of PAxpS that bound to GroEL with the same avidity as
the full length precursor was PAxpS90T, which contains the entire transit
sequence plus 43 amino acids of the mature protein. This truncated
precursor was used to determine whether the site identified by
mutagenesis is the site of GroEL contact. In order to do this, we devised a
protection assay which takes advantage of a tyrosine in the middle of the
binding site identified by site-specific mutagenesis. This tyrosine should
be protected from iodination upon formation of a complex between
PAxpS90T and GroEL. There are also 5 additional tyrosines in PAxpS90T that can serve as controls for the specificity of complex formation (Broglie et al., 1983; Uhlen et al., 1984). One of these also resides in the mature part of pS, 15 amino acids to the N-terminal side of the tyrosine in the putative binding site. The other 4 tyrosines are all present in protein A. In order to identify the location of each tyrosine-containing peptide in pS on the TLC plates, we used two additional constructs (PAxpS(Δ62-72)90T and PAxpS78T) in which the first and second tyrosines, respectively, of pS are deleted.

Protein A and each of the fusion proteins were purified and iodinated using a modification of Pierce’s IODO-BEADS protocol (see Materials and Methods). In addition, the complex between GroEL and PAxpS90T was also iodinated. The proteins were separated by SDS-PAGE, electroeluted, and digested with trypsin. The peptide fragments that were generated were separated by 2D-peptide mapping, using high voltage electrophoresis in the first dimension and chromatography in the second. The peptide map of protein A reveals four main peptides that are ¹²⁵I-labelled, two of which are overlapping (Figure 2.1A). The separation of these main peptides is quite good considering that the sequences of these peptides differ by only 1 to 3 amino acids (Figure 2.2). Also visible are several peptides which are due to partial proteolysis by the trypsin or that were not completely
soluble in one dimension or the other. Partial proteolysis products were visible on all the autoradiograms in varying amounts.

The peptide map of PAXpS90T (Figure 2.1B) is almost identical to that of protein A except for the appearance of two additional $^{125}$I-labelled peptides that are well-resolved in both dimensions. These were generated from the pS portion of the fusion and represent the sequences of P5 and P6 (see Figure 2.2). The lighter labelling of these peptides is due either to the limited solubility of these peptides or the very short iodination reaction times. Iodination conditions similar to Pierce’s recommended protocol reveal an even labelling of all peptides suggesting that the modified protocol favors iodination of tyrosines present in protein A.

The peptide maps of PAXpS(Δ62-72)90T and PAXpS78T allowed us to identify the sequence of the labelled peptides generated from pS. These constructs delete the first or second tyrosine and surrounding sequences present in pS, respectively. Both peptide maps have a single well-resolved labelled peptide missing as compared to PAXpS90T. Based on these maps, the peptide labelled with the closed arrow corresponds to P5, whereas the open arrow corresponds to P6. This identification agrees fairly well to the predicted mobilities of these peptides based on electrophoretic charge, relative hydrophobicity and molecular weight (Boyle et al., 1991).
Figure 2.1. **Two dimensional peptide maps.** Trypsin digests of iodinated protein A (A), PAxpS90T (B), PAxpS78T (C), PAxpS(Δ62-72)90T (D) were spotted on thin layer chromatography plates in pH 1.9 buffer. High voltage electrophoresis was run in the first dimension followed by chromatography in the second dimension. The origin is in the lower left-hand corner. The peptide map for the trypsin digest of PAxpS90T, iodinated as a complex with GroEL, is shown in (E). The closed and open arrows point to the position of the P5 and P6 peptides, respectively. The peptide map of PAxpS(Δ62-72)90T was run separately from the others and hence the extent of partial proteolysis and the position of the electrophoresis and chromatography fronts are slightly different.
Protein A

P1. DQQSAFYEILNMPNLNEAQR
P2. EQQNAFYEILNMPNLNEEQR
P3. EQQNAFYEILHLPLNNEEQR
P4. EQQNAFYEILHLPLTTEEQR

pS

P5. (59) FETLSYLPPPLSTEALLK (75)
P6. (76) QVDYLIR (82)

Figure 2.2. Sequence of iodinated peptides from protein A and pS. Sequence of potentially iodinated peptides generated from trypsin digestion of PAxpS90T. The amino acids in the pS peptides are numbered from the N-terminus of pS. The proteins PAxpS(Δ62-72)90T and PAxpS78T delete the first and second tyrosines in pS corresponding to peptides P5 and P6, respectively.
The peptide map of PAxpS90T, iodinated in a complex with GroEL, revealed that the only tyrosine protected from $^{125}$I-labelling was present in the P6 peptide. This peptide represents the same region of the mature protein that was identified by site-specific mutagenesis as required for a high avidity interaction with GroEL. However, a tyrosine present only 15 amino acids away is not protected by formation of the PAxpS90T-GroEL complex. This suggests that GroEL binds to a very specific region of pS or that any contact with other portions of the protein is highly solvent exposed to allow iodination to occur. Unfortunately chloroplast transit sequences rarely contain tyrosines and pS is no exception (von Heijne et al., 1991), therefore, we are not able to determine if other interactions may occur in this region.

In conclusion, we have developed a novel method to examine the contact sites of protein-protein interactions. Using this method we showed that GroEL interacts with the precursor of the small subunit of Rubisco at a site previously identified by site-specific mutagenesis. In addition, GroEL does not interact with other sites within pS or protein A that can be examined using this method, implying that this interaction is highly specific.
CHAPTER 3

Specificity of Chaperonin GroEL Binding to a
Chloroplast Precursor Protein
Ellis defined molecular chaperones as "a family of cellular proteins which mediate the correct folding of other polypeptides, and in some cases their assembly into oligomeric structures, but which are not components of the final functional structures" (Ellis, 1990a). This definition can be expanded to include proteins which bind to nascent polypeptides and prevent the formation of incorrect structures. These chaperones, although they never directly facilitate folding, are required to keep precursors in a transport-competent state until their final destination is reached.

Molecular chaperones are postulated to have a role in protein folding/translocation events in prokaryotic and eukaryotic cells (reviewed in Gething and Sambrook, 1992). As proteins emerge from ribosomes, they may transiently associate with a member of the chaperonin family (cpn60) (at least in *E. coli* and organelles), an Hsp70 homolog, or the signal recognition particle (SRP) if destined for the ER. In prokaryotes, the *E. coli* chaperone SecB stabilizes precursors destined for secretion from bacteria, while peptidyl prolyl *cis-trans* isomerase (PPI) awaits these proteins in the periplasm to aid in slow folding steps. In eukaryotes, proteins entering the ER are greeted by a resident ER Hsp70, along with protein disulfide isomerase (PDI) and PPI. Those precursor proteins headed for the mitochondria or chloroplasts are escorted by a cytosolic Hsp70 homolog. Even after these precursors enter the mitochondria or chloroplasts they are transiently bound by yet another set of Hsp70 and chaperonin homologs.
Both the *E. coli* chaperonin (GroEL) and the ER Hsp70 (BiP) bind proteins during heat shock or other stress conditions, while the *E. coli* Hsp70 (DnaK) can resurrect certain protein aggregates. With such assorted functions and the ability to interact with many different nascent chains, it is no wonder that recognition motifs for the various molecular chaperones are still a mystery.

Hsp70 recognition domains were identified for the immunoglobin heavy chain and the hemagglutinin of influenza virus (Gething and Sambrook, 1990). Based on these two proteins, it appears that Hsp70 may bind to sequences that form subunit interfaces. Flynn *et al.* (1991) used small peptides to further define the binding specificity of Hsp70. The peptide binding site on Hsp70 was filled by a stretch of only 7 amino acids. There was a preference for small aliphatic amino acids in the middle positions of the peptides and a slight preference for positively charged residues at position 7. Negatively charged and hydrophilic amino acids were excluded at all positions.

It is clear from NMR studies of Hsp70 and the *E. coli* chaperonin GroEL that the mode of binding differs between these two chaperones. Landry *et al.* (1992) showed that the vsv-C peptide bound to the *E. coli* Hsp70 is in an extended conformation but it is helical when bound to GroEL. These studies also indicate that Hsp70 interacts with the
backbone of the peptide while GroEL interacts with the amino acid side chains. Two other peptides bind GroEL with a similarly weak affinity (Kd ≈ 10-1000 uM) (Landry and Gierasch, 1991), but it is difficult to define a consensus motif from just these three peptides.

In order to better define a GroEL recognition motif, we measured the affinity constants for a number of polypeptides that bind to GroEL. We previously identified a 15 amino acid region of the precursor form of the small subunit (pS) of ribulose-1,5-bisphosphate carboxylase that binds GroEL with high avidity (Chapter 1). In addition, we developed a protection assay to show a physical interaction between GroEL and this region of pS (Chapter 2). Site-specific mutations made in the GroEL binding region of pS bound GroEL to varying extents. We therefore used Scatchard analysis to measure the affinity of each of these proteins toward GroEL and will propose one possible substrate recognition motif for GroEL.

Materials and Methods

PAxpS constructs and mutants were previously described (Chapter 1). The site-specific mutations were constructed in the 90 amino acid truncation of PAxpS and are named according to the substituting amino acid. GroEL was purified to homogeneity according to published methods (Chapter 1). The concentration of GroEL stock solutions was determined
by quantitative amino acid analysis with reference to the published amino acid sequence (Hemmingsen et al., 1988).

_Determination of binding constants_ -- The various PAxpS derivatives were synthesized in _E. coli_ and purified on IgG-agarose as described previously (Chapter 1) except the buffer used throughout was 40 mM Hepes, 50 mM NaCl, pH 7.4. Any GroEL and DnaK that copurified with the fusion proteins was removed using casein and MgATP, leaving only a purified fusion protein bound to an IgG-agarose matrix. This matrix was diluted in buffer and then divided equally into several Eppendorf tubes with 120 ul of matrix per tube in a total volume of 300 ul. In addition, a second set of tubes was created as a negative control which contained exactly the same amount of IgG-agarose matrix in 300 ul, but with no protein bound. An equal volume of various concentrations of GroEL mixes were added to each set to give a final concentration of 40 mM Hepes, pH 7.4, 50 mM NaCl, 2.5% glycerol, and 0-400 nM GroEL (14-mer). For those constructs that we anticipated would have higher dissociation constants we increased the concentration of GroEL to 800 nM to reach saturation. For PAxpS78T, however, this was never obtained due to the high concentrations of GroEL required. Therefore, the Kd for this construct is only an approximation. These reactions were incubated for 90 minutes while rocking at RT. The time necessary for establishing equilibrium was determined experimentally. After incubation, the column
matrix was centrifuged for 2 min and the top 250 ul of the supernatant was removed. The GroEL concentration for each supernatant was determined using the BCA (Pierce) protein assay with known concentrations of GroEL as the standard curve. The pelleted column matrix was transferred to Econo-columns (Biorad), washed quickly with 0.5 mls of buffer (40 mM Hepes, 50 mM NaCl, pH 7.4) and eluted with 7 M urea followed by 0.5 M acetic acid, pH 3.4. The eluted fusion protein and bound GroEL were TCA precipitated and separated by SDS-PAGE. The amount of fusion protein present was estimated by excising the gel segment containing the polypeptide, grinding it in liquid N₂, and eluting the polypeptide in 50mM Hepes, 200mM β-mercaptoethanol, and 0.2% SDS. The samples were boiled for 5 minutes and incubated overnight at 37°C before pelleting the gel pieces by centrifugation. The amount of Coomassie stain bound to the protein was measured by absorbance at 587 nm and compared to a standard curve prepared under identical conditions.

Scatchard analysis -- For Scatchard and Hill plots it is necessary to determine the amount of free and bound ligand which, in this case, is GroEL. The free ligand concentration (Sf) was simply the concentration of GroEL measured in the supernatant after pelleting the column matrix and bound GroEL in the reaction mixtures. The amount of bound GroEL (SB) was taken as the difference between the total concentration of GroEL, as
measured with no fusion present on the IgG-agarose matrix, and the free
concentration of GroEL with fusion present
\( (\text{IgG(Sf)} - \text{PAXpS (Sf)} = \text{PAXpS (Sb)} ) \). The dissociation constants were
calculated from the best fitting regression lines generated through an
iterative least squares analysis of the data. This analysis was performed
using the Ligand analysis program of Munson and Rodbard (1980). The
Hill constants were generated using the Equilibrium Binding Data Analysis
program by McPherson (1983).

*Estimation of GroEL's cellular concentration -- E. coli* lysate containing
the plasmid PAXpS was separated by SDS-PAGE. The amount of GroEL
present in the lysate was determined by western blotting using known
concentrations of GroEL as standards. Cell density was determined by
plating serial dilutions of the cells before lysis.

**Results**

*Determinaton of the dissociation constant for the GroEL-pS complex*

We determined the affinity constant for the interaction of the 800 kD
GroEL complex and the precursor of the small subunit (pS) of Rubisco. We
expressed pS as a C-terminal fusion to protein A and previously showed
that GroEL does not bind to protein A nor the IgG-agarose matrix used to
purify the protein fusions (Landry and Bartlett, 1989; Chapter 1). Our
method for measuring the dissociation constants took advantage of the
fact that any complex formed between GroEL and the pS moiety can be
easily removed from solution by rapid centrifugation of the IgG-agarose matrix. Therefore, the concentration of free GroEL could be measured in the remaining supernatant. The precursor form of the small subunit of Rubisco has a distinct advantage over other substrates in that it does not appreciably obtain a conformational state no longer recognized by GroEL. As a fusion to protein A, pS remains soluble and can establish an equilibrium between the bound and free state with GroEL. This "recognized conformation" is most likely maintained due to the presence of the transit sequence and the absence of the large subunits of Rubisco with which small subunits assemble to form the holoenzyme. In Figure 3.1A, the Scatchard plot for the interaction of PAxpS90T and GroEL is shown. PAxpS90T is a truncation of the small subunit at amino acid 90 and is the smallest truncation of pS that binds GroEL to the same extent as the full-length protein (Chapter 1, Table III). All mutations of pS will be in the context of this truncation. The Scatchard plot yields an association constant of $1.88 \times 10^7$ M\(^{-1}\) and a dissociation constant of 53 nM. The Scatchard plot is completely linear throughout the data points which is also reflected in a Hill coefficient of 1.0. The measured dissociation constant (Kd) is two orders of magnitude lower than that predicted by Viitanen et al. (1992a) for other substrates. Assuming the volume of the E. coli cell to be $10^{-12}$ ml, we estimate that the cellular concentration of GroEL is approximately 1.4 uM. Therefore, most of the fusion should be
bound by GroEL in the cell. A significant portion does remain in a complex with GroEL upon resuspension of the cells, lysis, and subsequent dilution of the soluble proteins (Chapter 1).

All the mutants of PApS90T or other constructs have linear Scatchard plots and thus Hill coefficients of 1.0. By estimating the amount of fusion from SDS-PAGE, we can calculate the number of binding sites to fall within the range of 0.5 to 1.0. Therefore, approximately one GroEL molecule (14-mer) binds 1-2 molecules of fusion. Examples are shown in Figure 3.1A and B. Bochkareva et al. (1992) showed that GroEL also binds 1-2 molecules of rhodanese but in a cooperative manner. However, since denatured rhodanese self-aggregates neither the dissociation constant nor the Hill coefficient for this interaction could be measured.

The 90 amino acid truncation (PAxPS90T) has the same affinity for GroEL as does the full length pS (Table III), strongly suggesting that the first 90 amino acids of pS contains the strongest, and possibly the only, binding site for GroEL. Using mutagenesis and peptide mapping, we previously showed that the site of GroEL binding is located within amino acids 72-86 of truncated pS. However, we also showed that the presence of the transit sequence (first 47 amino acids) increased the amount of GroEL bound to the mature protein in vitro. The dissociation constant for PAxS, which lacks the transit sequence, is 235 nM, at least four times that for PAxPS. A further C-terminal truncation of pS, which deletes half of the
Figure 3.1 Scatchard plots of GroEL binding to PÃexpS90T and PÃexpS90THCH3. Scatchard plots of GroEL binding to PÃexpS90T (A) and PÃexpS90THCH3 (B). The insets represent the concentration dependence of binding GroEL to these proteins. Approximately 312 pmol of PÃexpS90T (A) and 75 pmol of PÃexpS90THCH3 (B) bound to an IgG-agarose matrix were incubated batchwise for 90 min at 25°C while rocking in 600 ul of buffer (50 mM Hepes, 50 mM NaCl, pH 7.4) containing 0 to 400 nM GroEL. The matrix was centrifuged for 2 min to separate GroEL bound to the fusion protein from free GroEL in solution. The concentration of free GroEL was measured by the BCA protein assay. Identical conditions were used with IgG-agarose matrix containing no bound fusion protein to measure free GroEL concentrations in the absence of fusion protein. The fraction of bound GroEL represents the concentration of bound GroEL divided by the concentration of fusion protein present in the reaction.
binding site within the mature part of pS (PAxpS78T) but leaves the transit sequence intact, has a dissociation constant of well over 1000 nM. If there is a GroEL binding site in the transit sequence it is very weak compared to the site within the mature protein, but the transit sequence may help to stabilize the overall GroEL-pS complex or may interact with the mature portion of S in such a way that motifs recognized by GroEL are displayed.

We also determined association constants for a number of PAxpS90T derivatives containing mutations in the GroEL binding site in the mature portion of pS. The sequences of these mutations in PAxpS90T are shown in Table IV. The range of dissociation constants for mutations in this limited area is 27 to 370 nM, almost a 14-fold difference in magnitude. It is important to note that not all the mutations resulted in decreased association constants. Several mutant proteins have significantly increased affinities towards GroEL. The mutant protein with the greatest affinity for GroEL contains a large number of positive charges throughout the binding region. This theme is continued with several other PAxpS90T derivatives. For example, replacement of a single arginine at position 82 with either glutamine or proline results in a significant decrease in the association constant (Ka) for the interaction of these proteins with GroEL (compare Gln4* with Gln5 and Pro3 with Pro4). Conversely, the
### Table III

Association Constants for Derivatives of PAXpS

<table>
<thead>
<tr>
<th>Construct</th>
<th>Association Constant (K_a)</th>
<th>Kd (nM)</th>
<th>(1/ K_a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAXpS</td>
<td>(1.87 ± 0.37) x 10^7</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>PAXS</td>
<td>(4.24 ± 0.45) x 10^6</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td>PAXpS78T</td>
<td>(7.6 ± 1.9) x 10^5</td>
<td>1300</td>
<td></td>
</tr>
<tr>
<td>PAXpS90T</td>
<td>(1.88 ± 0.19) x 10^7</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

*a Association Constants ± the standard error of the mean were calculated using the Ligand analysis program by Munson and Rodbard (1980)*
Table IV

Association Constants for Mutations Created in PAnpS90T

<table>
<thead>
<tr>
<th>Sequence of pS amino acid 67 to 86</th>
<th>Ka (x 10⁻⁶)*</th>
<th>(1/Ka) Kd (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCH3</td>
<td>37.4 ± 1.7</td>
<td>27</td>
</tr>
<tr>
<td>Gly3</td>
<td>27.5 ± 3.2</td>
<td>36</td>
</tr>
<tr>
<td>Pro2</td>
<td>26.9 ± 4.4</td>
<td>37</td>
</tr>
<tr>
<td>HCH2</td>
<td>19.8 ± 3.5</td>
<td>50</td>
</tr>
<tr>
<td>W.T.</td>
<td>18.8 ± 1.9</td>
<td>53</td>
</tr>
<tr>
<td>HCH4</td>
<td>16.0 ± 1.1</td>
<td>62</td>
</tr>
<tr>
<td>Pro3</td>
<td>16.1 ± 2.2</td>
<td>62</td>
</tr>
<tr>
<td>HCH5</td>
<td>16.1 ± 2.0</td>
<td>62</td>
</tr>
<tr>
<td>Pro3*</td>
<td>13.8 ± 1.2</td>
<td>72</td>
</tr>
<tr>
<td>Gln4*</td>
<td>10.4 ± 1.1</td>
<td>96</td>
</tr>
<tr>
<td>Gly4</td>
<td>10.1 ± 0.8</td>
<td>100</td>
</tr>
<tr>
<td>Pro4</td>
<td>9.1 ± 0.4</td>
<td>109</td>
</tr>
<tr>
<td>Gln4</td>
<td>7.3 ± 0.9</td>
<td>136</td>
</tr>
<tr>
<td>Gln5</td>
<td>5.6 ± 0.7</td>
<td>180</td>
</tr>
<tr>
<td>Ser5</td>
<td>5.1 ± 0.4</td>
<td>196</td>
</tr>
<tr>
<td>Gly5</td>
<td>2.7 ± 0.3</td>
<td>370</td>
</tr>
</tbody>
</table>

* Association Constants ± the standard error of the mean were calculated using the Ligand analysis program by Munson and Rodbard (1980).
substitution of arginine for isoleucine at position 81 results in more than a two-fold increase in the Ka (compare HCH3 and HCH4).

Another feature of this region that is important for GroEL binding is the presence of a hydrophobic amino acid(s) near position 72-74. The replacement of a leucine at position 73 with a proline decreases the Ka (compare Pro2 and Pro3). In addition, replacement of the hydrophobic residues of the wild type protein at positions 72-74 and 77 with glutamine decreases the association constant almost two-fold. However, a positively charged amino acid appears to partially compensate for the loss of hydrophobic amino acids in this part of the binding site (HCH4).

Two other general trends are notable. Three of the four mutations with the lowest affinity towards GroEL have tyrosine 79 and the adjacent leucine 80 replaced with more polar, noncharged amino acids. The second observation is that a positive charge may be required at position 75. Almost all of the PApS90T derivatives with the highest affinity for GroEL contain this positive charge. The general motif for GroEL binding of PApS may therefore consist of two closely spaced regions containing several amino acids with hydrophobic side chains, each of which is followed by an amino acid with a positively charged side chain. Replacement of at least some of the hydrophobic amino acids with positively charged ones does not diminish (and may enhance) GroEL binding.
NMR analysis of small peptides bound to GroEL revealed that the chaperonin promotes alpha helix formation (Landry and Gierasch, 1991). On the other hand, GroEL can bind to an all beta-sheet protein (Schmidt et al. 1992). Several of the PAxpS90T derivatives contain a number of prolines and glycines (Table IV) and thus are predicted to lack secondary structure between amino acids 67 and 86. Yet these derivatives bind GroEL avidly. Although secondary structure appears not to be a prerequisite for GroEL binding, GroEL possibly promotes formation of secondary structures in those substrates which form them. However, we cannot rule out the possibility that structural rigidity is a prerequisite for GroEL binding (see below).

Results of NMR studies also show that GroEL significantly reduces the mobility of side chains and not the backbone of peptides (Landry et al., 1992). The Gly5 mutant has the least affinity for GroEL of any PAxpS derivatives. The reduction in the available side chains of the Gly5 mutant could point to GroEL’s preference for binding to the side-chains of proteins, regardless of the secondary structure. Ser5 contains serines in the same positions as the glycines in Gly5, yet the GroEL:Ser5 complex has a Ka almost twice that of the GroEL:Gly5 complex. The reduction in available side chains of the Gly5 and the corresponding increase in Ka of the GroEL:Gly5 complex could point to GroEL’s preference for binding to the side chains of proteins, regardless of secondary structure. However,
this difference could also be due to the increase in flexibility and rotational freedom afforded by the glycine residues.

Conclusions

In summary, we determined the association constant for the interaction of GroEL with the precursor form of the small subunit of Rubisco. The dissociation constant for this interaction is 53 nM (Ka = 1.88 x 10^7 M\(^{-1}\)). We showed that mutations in the GroEL binding site of pS identified previously can alter the affinity of pS for GroEL. In addition, we identified several features of this binding site that contribute to GroEL binding. In essence, this site consists of two patches, each of which contains two or more hydrophobic amino acids immediately followed by a positive charge. Although these elements do seem to be important for recognition of this one region of pS by GroEL, our results clearly demonstrate GroEL’s ability to avidly bind a wide range of sequences. For example, replacement of two hydrophobic amino acids in either patch with positively charged ones has little effect on GroEL binding. GroEL must present a surface that is a patchwork of many hydrophobic and negatively charged amino acids. Such a notion is consistent with GroEL’s acidic pI, its ability to bind the hydrophobic probe bis-ANS (Mendoza et al., 1991a) and its ability to bind a variety of different substrates. In all four of the PAxpS90T truncations with the least affinity for GroEL, at least one of the hydrophobic patches has been disrupted. Furthermore, in three of the four
mutants a positively charged amino acid has been replaced with an uncharged one, resulting in a net charge of zero for the region between amino acids 67 and 86 of pS.

Results of studies of the import of pre-cytochrome b$_2$ fused to the mouse dihydrofolate reductase show that deletion of 19 amino acids of the cytochrome b$_2$ export sequence results in drastically reduced binding of the chimeric protein to the mitochondrial chaperonin HSP60 (Koll et al., 1992). The deleted region contains three positive charges followed by three hydrophilic residues, four hydrophobic amino acids and another positive charge. Similarly, each of the three peptides used in NMR binding studies with GroEL contains two positively charged amino acids separated by several hydrophobic residues (Landry and Gierasch, 1991; Landry et al., 1992). Thus, our results as well as those described above suggest that a substrate for GroEL contains several adjacent hydrophobic residues and an overall positive charge.

Interestingly, SecB binds both positively charged peptides and the hydrophobic fluorescent probe 1,8-ANS. Randall (1992) proposed that SecB interacts with substrates by binding to a positively charged region on the substrate which induces a conformational change in SecB, further exposing a hydrophobic patch on SecB to which the substrate can also bind. Thus it appears that at least two molecular chaperones, SecB and GroEL, can recognize similar features in substrates. On the other hand,
while hydrophobic side chains and a positive charge clearly are important for avid binding of GroEL to pS and its derivatives, other substrates presenting a different array of features at their surfaces may also bind the chaperonin with avidity.
CONCLUSIONS AND FUTURE PROSPECTS
What features do chaperonins recognize that are present in the unfolded or partially folded form of a linear sequence of amino acids but which are not available for recognition when this sequence is folded into its native conformation? To address this question we utilized the precursor form of the small subunit of Rubisco as a model substrate for the \textit{E. coli} chaperonin GroEL. This precursor protein transiently binds the chloroplast chaperonin upon import before assembly with the large subunit of Rubisco to form the holoenzyme. Landry and Bartlett (1989) showed that the \textit{E. coli} chaperonin binds specifically with pS expressed as a C-terminal fusion to protein A (PAxpS). This binding did not require prior denaturation of PAxpS. We further characterized this binding system and showed that some GroEL can be released from the fusion protein with MgATP and K$^+$. However, complete removal of GroEL required either chaperonin-10 (GroES) or casein as well as MgATP and K$^+$. Purified GroEL binds to the protein A-pS fusion protein \textit{in vitro} with a dissociation constant of 53 nM (Ka = $1.88 \times 10^7$ M$^{-1}$). This dissociation constant is much lower than the cellular concentration of GroEL (1.4 uM) and thus most of the fusion should be bound by GroEL \textit{in vivo}.

Using C-terminal truncations of PAxpS and site-specific mutagenesis, we located a stretch of 15 amino acids (positions 72-86) in the mature portion of pS that is required for avid binding of GroEL. In addition, a tyrosine residue present in this putative binding site is protected
by GroEL from iodination upon complex formation, whereas other tyrosine residues present both in pS and protein A are not protected from iodination. This suggests that GroEL actually contacts or binds the fusion protein at or near the site identified by site-directed mutagenesis.

If one were to predict the location of a chaperonin binding site in pS, logical choices would include residues normally hidden in the folded protein, such as those in the hydrophobic core, or at contact sites with other small subunits or with the large subunits of Rubisco. Chaperonins might be required to sequester these potentially "sticky" sites from the cellular environment before assembly. However, the GroEL binding site identified in the mature region of pS does not correspond to any of the subunit interactive sites, either with small or large chains, that were determined from the X-ray crystal structure of the holoenzyme (Knight et al., 1989). The GroEL binding site also does not correspond with the putative assembly domain identified by Wasmann et al. (1989). The assembly domain is a conserved 16 amino acid sequence present in higher plant small subunits but which is absent in the small subunits of the cyanobacterial Rubisco. It was proposed to facilitate assembly of the eukaryotic holoenzyme and corresponds to amino acids 96 - 111 of wheat pS, a region absent in our avid binding PAxpS derivative, PAxpS90T.

Residues present at the hydrophobic core would normally be inaccessible in the folded protein but might be exposed in the unfolded or
partially folded form of pS. However, only 2 (Leu 73 and Val 77) of the 11 core residues of S are present in the GroEL binding site. Mutations at these sites had little effect on GroEL binding. Therefore exposed core residues cannot be the major contributors to GroEL recognition of pS, although they may be part of a more complex recognition scheme.

The transit sequence of pS might also be considered a potential binding site for chaperonins. Upon import into chloroplasts, the transit sequence is the first part of the protein that emerges from the membrane. Results of mitochondrial import studies suggest that Hsp70 homologs recognize emerging polypeptides and help to reel them into the organelle. These nascent proteins are then passed on to the mitochondrial chaperonin (Cheng et al., 1989). Certainly the presence of the transit sequence strengthens the interaction of the small subunit with GroEL. The affinity constant for the interaction of GroEL with the full length precursor as a fusion to protein A (PAxpS) is $1.88 \times 10^7 \text{ M}^{-1}$ versus $4.24 \times 10^6 \text{ M}^{-1}$ for the mature small subunit (PAxS). GroEL weakly binds a polypeptide consisting of the transit sequence and the first 31 amino acids of the mature protein which includes 6 amino acids of the avid binding site ($K_a = 7.6 \times 10^5 \text{ M}^{-1}$). Thus residues 78-90 clearly are responsible for avid binding of GroEL, although the presence of the transit sequence increases GroEL's affinity for the precursor protein. Deletion of either the first two thirds or the last third of the transit sequence still supports an increase in
GroEL binding as compared to the mature protein alone. This is not completely unexpected since transit sequences are generally similar in nature throughout the length of their sequence, containing a large number of hydroxylated amino acids and a net positive charge (Keegstra et al., 1989). It is still unclear whether GroEL recognizes a site in the transit sequence or whether this region simply stabilizes the GroEL-pS complex, possibly by presenting a few nonspecific contacts or by promoting a conformation in pS that makes the binding site more accessible to GroEL.

Since the GroEL binding site in pS contains few amino acids normally sequestered upon folding and is located in the mature protein there must be other features that define a chaperonin recognition motif. By measuring the affinity constants for the interaction of GroEL with a series of site-specific mutations created in the 90 amino acid truncation of PAXpS, we noted several key elements important for binding GroEL. The GroEL binding site in pS consists of a duplicated hydrophobic patch immediately followed by a positive charge. These two elements are separated by 2 additional amino acids and a negative charge. Most of the mutant derivatives of PAXpS90T had affinities for GroEL similar to it. Neither PAXpS90T or the mutant derivatives share sequence homology with other polypeptides that binds GroEL. Thus GroEL can bind a wide range of sequences. In order to accommodate such a variety of features, GroEL must present a surface that is a patchwork of hydrophobic and
negatively charged residues. Previous studies have defined a hydrophobic site on GroEL (Mendoza et al., 1991a) and its pI of 5.63 (Hemmingsen et al., 1988) indicates the presence of a number of acidic residues. A model for the interaction of PAxpS and GroEL is shown in Figure 4.1.

Although amino acids 69 to 81 in the wild-type sequence of pS form an amphipathic alpha helix in the Rubisco holoenzyme (Knight et al., 1989), it is still unclear what, if any, secondary structure is required for binding GroEL. Our results as well as results of previous folding studies might suggest that GroEL can bind structures ranging from completely unfolded (Badcoe et al., 1991) to nearly native (Mendoza et al., 1992b). A patchwork surface on a protein of GroEL’s magnitude could accommodate this range of structure since it has the potential to be all things to all (or nearly all) substrates. A protein substrate simply has to bind a complementary pattern on GroEL’s surface. This may explain why the estimate of the number of substrate molecules bound by a single GroEL molecule range from 1-2. In some cases, the orientation of a substrate on the surface of GroEL could prevent binding of a second substrate molecule. The number of rhodanese molecules bound to GroEL changes from 1-2 upon addition of laurylmaltoside. The presence of the detergent may reduce the steric hindrance of binding two rhodanese molecules by interacting with a site(s) normally bound by GroEL.
Figure 4.1 Model for the interaction of PAxpS and GroEL. The symbol $\phi$ represents hydrophobic residues, whereas, the + and - symbols represent basic and acidic amino acids, respectively.
The characterization of the GroEL binding region of pS provides the first clues towards understanding GroEL's recognition of a substrate, however, it is far from complete. The features identified are present in small peptides that bind GroEL (Landry et al., 1991) and in other protein substrates of the chaperonin (Koll et al., 1992) but probably do not define all the elements recognized by GroEL. The tandem nature of the motif defined in pS seems to be an important feature. PAxpS78T, which contains only the first hydrophobic stretch and positive charge of the avid binding site binds GroEL very weakly (Ka = 7.6 x 10^5 M^-1 compared to 1.88 x 10^7 M^-1 for PAxpS90T). The nature of the spacer region and the distance between these repeats could easily be examined using this system. We have also expressed several other chloroplast precursor proteins (such as acetolactate synthase, glutamine synthetase, and phosphoribulokinase) as C-terminal fusions to protein A. All of these proteins bound GroEL with high avidity. The identification of GroEL binding domains in these proteins would lead to additional clues about potentially other recognition motifs of GroEL.

The PAxpS protein could provide a binding assay to screen for functional GroEL or alternative chaperonin molecules. The ams gene encodes an RNase E-like activity which functionally interacts with GroEL (Sohlberg et al., 1993). A groEL fragment coding for amino acids 307 to 548 of the GroEL protein was found to complement a mutation of the ams
gene (Chanada et al., 1985). Expression and purification of this partial GroEL molecule may define the minimal sequence required for function. Its binding to PAxpS, elution with MgATP, interaction with GroES, and ability to compete with the wild-type GroEL would easily be assayed using the techniques described in chapters 1 and 3.

In order to answer many of the remaining questions concerning the GroEL-GroES proteins, methods other than those described here will be required. Several groups are attempting to determine the crystal structures of GroEL and GroES. While investigators impatiently wait for those structures, the mechanism of GroES action needs to be determined. One proposal suggests that GroES functions to potentiate the maximum effect of ATP and to couple its hydrolysis to the release of a folded protein. How many ATP molecules are hydrolyzed during the folding of a single substrate, and is this number substrate dependent? What role does GroEL play in the folding of those proteins that do not require GroES or even MgATP? To what extent do chaperonins coordinate their actions with other chaperones in the cell? All of these questions must be answered before we can truly grasp the nature of protein folding in the cell.
LITERATURE CITED


Ellis, R. J. (1990a) *Seminars in Cell Biol.* 1, 1-9

Ellis, R. J. (1990b) *Science* 250, 954-959


Carmen Wheelock Dessauer was born on April 29, 1966 in Marrero, Louisiana. She attended Mandeville High School and received several academic scholarships, including the L. S. U. Chancellor’s Scholarship and the A. R. Choppin Scholarship, to attend Louisiana State University in Baton Rouge. Her major was Biochemistry and she graduated with highest honors receiving a B. S. degree in May of 1988. Her research experience as an undergraduate included two summers at the United States Department of Agriculture under the direction of Dr. Barbara Triplett and a year in Dr. Sue Bartlett’s laboratory at Louisiana State University. She pursued a graduate degree in Biochemistry under the direction of Dr. Bartlett and upon completion of her degree requirements will receive a Ph.D. degree. She has accepted a postdoctoral position in Dr. Alfred G. Gilman’s laboratory at the University of Texas, Southwestern Medical Center in Dallas.
Candidate: Carmen Wheelock Dessauer

Major Field: Biochemistry

Title of Dissertation: Specificity of Chaperonin GroEL Binding to the Precursor of the Small Subunit of Ribulose-1,5-Bisphosphate Carboxylase

Approved:

[Signatures]

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination: April 1, 1993