Article No. jmbi.1999.3040 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 292, 163–172

JMB

GroEL Recognises Sequential and Non-sequential Linear Structural Motifs Compatible with Extended β-Strands and α-Helices

Jean Chatellier, Ashley M. Buckle and Alan R. Fersht*

Cambridge University Chemical Laboratory and Cambridge Centre for Protein Engineering and MRC Centre, Hills Road Cambridge, CB2 2QH UK

The chaperonin GroEL binds a variety of polypeptides that share no obvious sequence similarity. The precise structural, chemical and dynamic features that are recognised remain largely unknown. Structural models of the complex between GroEL and its co-chaperonin GroES, and of the isolated apical domain of GroEL (minichaperone; residues 191-376) with a 17 residue N-terminal tag show that a linear sequential sequence (extended β-strand) can be bound. We have analysed characteristics of the motifs that bind to GroEL by using affinity panning of immobilised GroEL minichaperones for a library of bacteriophages that display the fungal cellulose-binding domain of the enzyme cellobiohydrolase I. This protein has seven non-sequential residues in its binding site that form a linear binding motif with similar dimensions and characteristics to the peptide tag that was bound to the minichaperone GroEL(191-376). The seven residues thus form a constrained scaffold. We find that GroEL does bind suitable mutants of these seven residues. The side-chains recognised do not have to be totally hydrophobic, but polar and positively charged chains can be accommodated. Further, the spatial distribution of the side-chains is also compatible with those in an α-helix. This implies that GroEL can bind a wide range of structures, from extended β-strands and α-helices to folded states, with exposed side-chains. The binding site can accommodate substrates of approximately 18 residues when in a helical or seven when in an extended conformation. The data support two activities of GroEL: the ability to act as a temporary parking spot for sticky intermediates by binding many motifs; and an unfolding activity of GroEL by binding an extended sequential conformation of the substrate.

Keywords: phage display; protein-ligand interactions; GroES; GroEL minichaperone; cellulose-binding domain

**Corresponding author*

Introduction

The GroE chaperonin of *Escherichia coli* is essential for the folding and assembly of some newly synthesised polypeptides and for the recovery of stress-denatured proteins (Goloubinoff *et al.*, 1989). This complex molecular machine consists of 14 identical 57.5 kDa subunits of GroEL, which form a cylinder containing a central cavity ~45 Å wide (Braig *et al.*, 1994; Saibil *et al.*, 1993), and seven or 14 10 kDa GroES subunits, and is fuelled by ATP hydrolysis. Each 57.5 kDa subunit consists of three domains. The ATP-binding equatorial domain (residues 6-133 and 409-523) connects the two heptameric rings stacked back-to-back that form the cylinder. The apical domain (residues 191-376) forms the flexible opening of the cylinder and contains the polypeptide-binding and GroES-binding sites, which line the inner wall of the cavity. The intermediate domain (residues 134-190 and 377-408) makes intersubunit contacts within a ring and transmits ATP and GroES-mediated allosteric effects.

Abbreviations used: *A*, absorbance; BSA, bovine serum albumin; CBD, cellulose-binding domain; ELISA, enzyme-linked immunosorbant assay; IPTG, isopropyl-β-D-thiogalactoside; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; sht, short histidine tail; t.u., transducing units.

E-mail address of the corresponding author: arf10@cam.ac.uk

0022-2836/99/360163–10 $30.00/0

The mechanism of the GroE chaperonin is still under debate (Fenton & Horwich, 1997; Hartl, 1996; Lorimer, 1996). Certainly, there is a spectrum of mechanisms for GroE-mediated protein folding, depending on the nature of the substrate (Peres Ben-Zvi *et al*., 1998). The basic function of GroEL in facilitating folding is to provide a hydrophobic binding site (Buckle *et al*., 1997; Fenton *et al*., 1994; Zahn *et al*., 1996a), which prevents aggregation of folding intermediates and can lead to the unfolding of compact states (Zahn *et al*., 1996b). At one extreme, proteins can fold in a ''cage'' of GroEL subunits.

The recently determined crystal structure of the GroEL.GroES.$(ADP)_7$ complex at 3 Å resolution shows that GroES caps the end of the GroEL ring and each of its seven subunits binds the apical, polypeptide-binding domains of GroEL *via* the extended segment of its modile loop (Figure 1(a)). ATP-driven movements of apical and intermediate domains of GroEL allow GroES to bind to and change the character of the polypeptide binding site from being hydrophobic to hydrophilic, facilitating the release of bound polypeptide (Xu *et al*., 1997). The bacteriophage T4 co-chaperonin Gp31 has a similar binding mode to that of GroES (Hunt *et al*., 1997). But, crystallographic studies on intact GroEL have not resolved the structures of any bound peptide substrate. The structural nature of binding of non-native proteins by GroEL is unresolved (for a review, see Horovitz, 1998).

Functionally active monomeric minichaperones (Chatellier *et al*., 1998; Zahn *et al*., 1996a), which contain only a part of the polypeptide binding domain of GroEL (Fenton *et al*., 1994), are more amenable to detailed peptide recognition studies than is the intact tetradecamer. The crystal structure at 1.7 Å resolution of GroEL(191-376) with a 17 residue N-terminal tag reveals the interaction of the N-terminal tag of one molecule with the active site of a neighbouring molecule (Figure 1(b)). This shows the detail of an individual binding site and suggests a model for the binding of long peptides around the rim of seven subunits in the cavity of the tetradecamer (Buckle *et al*., 1997). GroEL minichaperones, therefore, constitute a paradigm for understanding GroEL-substrate interactions at the atomic level.

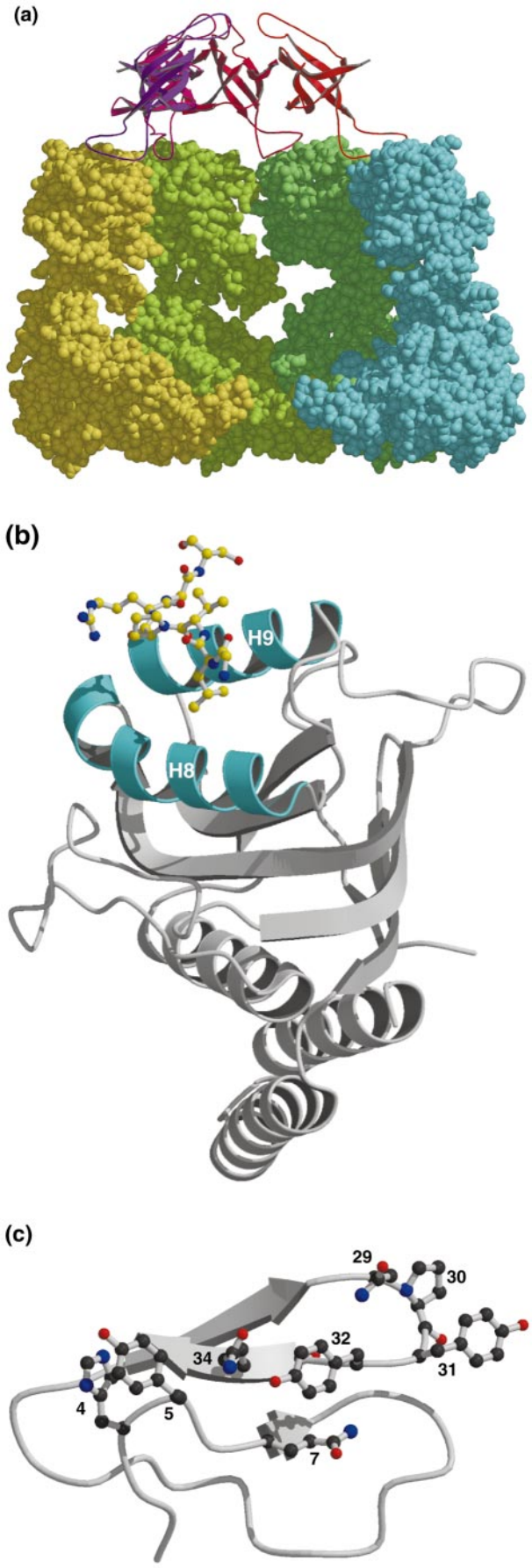
Further experiments are required to answer the questions: What structural and chemical features of polypeptide substrates are recognised by GroEL? What conformations do substrates adopt whilst bound to GroEL? And how does GroEL discriminate between native proteins and its misfolded protein substrates? Random mutagenesis of simple peptides followed by selection by binding to immobilised GroEL (affinity panning) can be used to determine the nature of the side-chains that bind. But this procedure has the disadvantage that the three-dimensional structure of such peptides is not generally constrained. To overcome this problem, we have used affinity panning of GroEL minichaperones for a library of bacteriophages that display constrained peptides. As a peptide model, we chose the C-terminal cellulose-binding domain (CBD) of the enzyme cellobiohydrolase I from the fungus *Trichoderma reesei* (Figure 1(c); Kraulis *et al*., 1989). This domain belongs to the knottin family, proteins that bind with a high degree of specificity to their target (Le Nguyen *et al*., 1990). Knottins share a common scaffold comprising a small triple-stranded antiparallel β-sheet and disulphide bond framework (Figure 1(c)). Cellulose binding is mediated by a 400 $Å^2$ patch of seven functional residues (Linder *et al*., 1995) that form a hydrophilic, flat face of the domain (Figure 1(c)). In the primary sequence, these residues are located in two separate regions: one close to the N terminus (residues 4, 5 and 7) and the other near the C terminus (residues 29, 31, 32 and 34). In the library, the seven functional residues were randomised (Smith *et al*., 1998). The side-chains of those seven residues have a similar stereochemistry to those in the proposed model for a substrate bound to GroEL (Buckle *et al*., 1997). The cellulose-binding domain thus offers a stable scaffold from which to identify the nature, size and shape of surface sites that bind GroEL. We interpret our findings in the light of the structural models of the GroEL-GroES and minichaperone GroEL(191-376)-peptide tag complexes.

Results

Selection of cellulose-binding domain (CBD) proteins

Techniques for defining substrate specificity of GroEL are readily suggested by analogy of GroEL minichaperones with monomeric Hsp70 family members, although the folding activity of the latter is ATP-dependent (Bukau & Horwich, 1998). The substrate-specificity of chaperones has been defined using the panning of large libraries of phage-displayed peptides (Smith, 1985) against the immobilised, purified chaperones (Blond-Elguindi *et al*., 1993; Gragerov *et al*., 1994; Hightower *et al*., 1994).

To identify the nature, size and shape of surface sites that bind GroEL, we used affinity panning of immobilised minichaperone short histidine tail (sht)-GroEL(191-345) for a large repertoire (5.5×10^8) of bacteriophages that display CBD randomly mutated at the seven discontinuous positions on the domain flat face that binds cellulose (Smith *et al*., 1998). The immobilisation of minichaperone sht-GroEL(191-345) was determined by ELISA; immobilised minichaperones were functional in refolding cyclophilin A as previously described (data not shown; Altamirano *et al*., 1997). The results of the selection are summarised in Figure 2. The number of colonies increased substantially at each round of selection (Figure 2(a)). The titre of eluted phages (t.u. total) increased with subsequent rounds (data not shown). The ELISA signal corresponding to the binding of these phage populations also increased with later rounds of selection (Figure 2(b)).



Characterisation of GroEL minichaperone-binding CBD proteins

A potential drawback for this approach, however, is the non-specific adsorption of phages to the solid phase during panning (Adey *et al.*, 1995). In the study of Hsp70 chaperones, specific elution of phages by the addition of ATP may have alleviated this problem. Therefore, we assayed single phage clones for binding activities by ELISA. The results obtained for three representative phages obtained after rounds 3 and 4 are depicted in Figure 3. The binding to immobilised minichaperone sht-GroEL(191-345) of the selected CBD-phage was highly specific. The binding background in the absence of phage or of wild-type CBD-phage was very low. A weak binding was also observed either against the non-substrate bovine serum albumin (BSA) (Figure 3) or the solid phase (data not shown). In addition, similar results were obtained using minichaperones devoid of the short histidine tail, indicating that the N-terminal tag did not interfere with the selection (data not shown). The binding activity was dramatically diminished after treatment of CBD-displayed phages with β-mercaptoethanol indicating the presence of disulfide bonds responsible for the proper fold of the domains on the phage tip (Smith *et al.*, 1998; data not shown). The mutation Y203E abolished the *in vitro* and *in vivo* activities of intact GroEL (Fenton *et al.*, 1994) and minichaperones (Chatellier *et al.*, 1998). The mutation Y203E drastically decreased the binding of minichaperone GroEL to the selected CBD-phages (Figure 3). This suggests that the recognition of denatured polypeptide substrates and CBD-phages by minichaperones involves the same residues.

Sequences of selected CBD proteins

A total of 24 different CBD sequences were identified from 55 clones with specific binding activity (Table 1). The diversity of sequences that bind minichaperone is consistent with the ability of GroEL to recognise a wide variety of unrelated

Figure 1. (a) Cross-section of the GroEL-GroES interface as described by Xu *et al.* (1997) (PDB code 1aon). Four subunits of GroES (orange to magenta) and GroEL (yellow to blue) are shown. (b) Three-dimensional structure of minichaperone GroEL(191-376), showing seven residues of the 17 residue N-terminal tag bound in the polypeptide binding site (PDB code 1kid; Buckle *et al.*, 1997). Helices H8 and H9 are coloured cyan. The bound peptide is drawn as ball and stick. (c) Structure of the cellulose binding domain (CBD) from *T. reesei* in solution (PDB code 1cbh; Kraulis *et al.*, 1989). The seven residues randomised in the library used (Smith *et al.*, 1998), forming the discontinuous cellulose-binding patch on the flat protein surface (Linder *et al.*, 1995), are shown. Secondary structure representation is drawn with MolScript (Kraulis, 1991), and rendered with Raster3D (Merritt & Bacon, 1997).

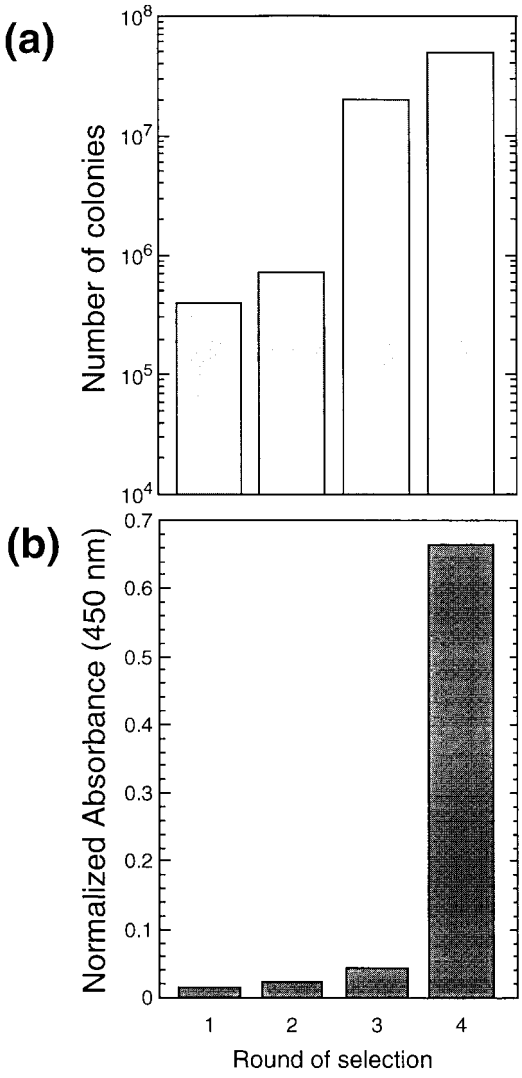


Figure 2. Selection of CBD-phage library against minichaperone sht-GroEL(191-345). (a) Number of colonies obtained after each round of selection. (b) Polyclonal phage (7×10^9 t.u.) from each round of selection were assayed for binding to minichaperone sht-GroEL(191-345) by ELISA.

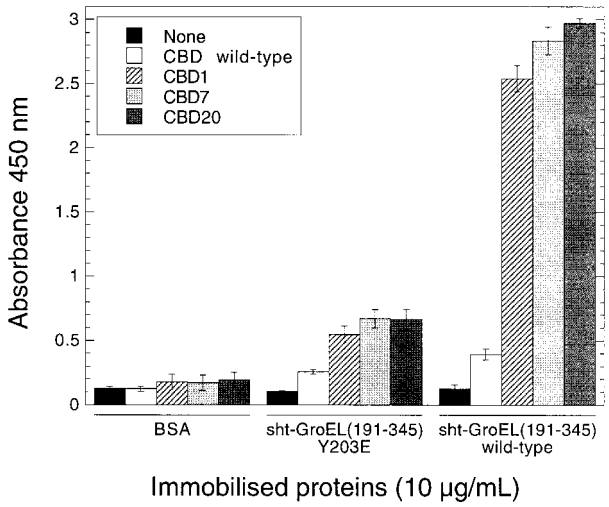


Figure 3. Binding specificity of CBD-selected phages to minichaperone sht-GroEL(191-345) wild-type or mutant Y203E determined by ELISA. See Table 1 for the sequences.

polypeptides. The frequency of amino acid type (hydrophobic, polar or positively charged) at each position of the discontinuous patches in the domains generated by selection is represented in Figure 4. The hydrophobic nature of the sequences is also consistent with the probability that GroEL binds to motifs preferentially buried inside properly folded protein. But the side-chains recognised do not have to be all hydrophobic; indeed polar and positively charged chains can be accommodated (see Table 1).

Similarities to the GroES mobile loop

The formation of the GroEL-GroES complex is mediated through a central hydrophobic tripeptide (Ile25-Val26-Leu27) of the GroES mobile loop (Landry *et al.*, 1993; Xu *et al.*, 1997). The ''consensus'' sequence, corresponding to the more frequent residues at each position of the discontinuous patch and some other, adjacent to the mutated residues, shares a significant degree of identity with the mobile loop of GroES, although the tripeptide is not fully conserved (Table 2).

To asses these similarities, the ability of a synthetic peptide corresponding to residues 16 to 32 of GroES mobile loop (see Table 2 for the sequence) to displace bound selected CBD-phages from GroEL minichaperone was tested by competition ELISA (Figure 5). The synthetic GroES mobile loop peptide does inhibit the binding of a similar amount (about 10^9 t.u.) of selected CBD-phages (Figure 5). The concentration of free peptide required for 50 % inhibition of binding (IC_{50}) calculated from the competition curve ranges from 10×10^{-6} to 30×10^{-6} M (Figure 5). The affinity of the clone selected at the fourth round of panning, CBD20, is about twice that of clones CBD1 or 7 from the third round of selection. The selected CBD-phages also inhibit to some extent the binding of GroEL (2.5 µM) to GroES (2.5 µM) in the presence or absence of 1 mM ADP (data not shown). Selected CBD peptides that bind tightly to minichaperone are good candidates for structural studies of GroEL substrate complexes.

Taken together, these results indicate that selected CBD-phages bind within the polypeptide-binding site of GroEL minichaperone as the synthetic GroES mobile loop peptide.

Table 1. Sequences of discontinuous patches in the wild-type cellulose binding domain from *T. reesei* and proteins generated by selection

Name	Residues number[a]										Frequency[b]	
	4	5	6	7	29	30	31	32	33	34	r3	r4
Wild-type												
CBD	H	Y	G	Q	N	P	Y	Y	S	Q	1	0
CBD1	S	S	G	L	P	P	H	G	S	L	1	0
CBD2	N	R	G	A	P	P	I	I	S	A	1	0
CBD3	G	E	G	C	I	P	A	P	S	F	1	0
CBD4	S	I	G	Y	R	P	S	S	S	I	1	0
CBD5	R	Q	G	Y	V	P	R	I	S	L	1	0
CBD6	H	Y	G	R	M	P	F	R	S	G	1	0
CBD7	R	S	G	S	L	P	C	D	S	M	1	0
CBD8	P	I	G	H	L	P	K	I	S	I	1	0
CBD9	Q	K	G	W	L	P	G	I	S	Q	1	0
CBD10	G	I	G	S	S	P	L	Y	S	L	1	0
CBD11	N	R	G	A	V	P	I	S	S	R	0	1
CBD12	S	S	G	Q	P	P	H	G	S	L	0	1
CBD13	R	L	G	P	V	P	G	I	S	L	0	1
CBD14	F	H	G	L	S	P	A	S	S	L	0	1
CBD15	L	C	G	R	V	P	V	S	S	R	0	1
CBD16	L	G	G	S	H	P	G	L	S	I	0	1
CBD17	R	S	G	Q	L	P	S	M	S	P	0	1
CBD18	L	G	G	S	H	P	G	L	S	L	0	1
CBD19	Q	C	G	H	P	P	V	R	S	T	0	1
CBD20 & 22	R	Q	G	T	L	P	A	R	S	P	0	2
CBD21	Q	L	G	L	H	P	L	S	S	S	0	1
CBD23	Q	P	G	R	S	P	Q	T	S	H	0	1
CBD24	Q	V	G	A	S	P	C	S	S	L	0	1

[a] Residues not randomized in the CBD library (Smith *et al.*, 1998) are underlined.
[b] Number of indidual sequences found from 24 clones: r, round of selection.

Discussion

To characterise the nature, size and shape of surface sites that bind GroEL, we used affinity panning of immobilised functional minichaperone sht-GroEL(191-345) for a large library of bacteriophages that display the fungal cellulose-binding domain of the enzyme cellobiohydrolase I randomly mutated at seven discontinuous positions on the domain hydrophilic, flat face that binds cellulose (Figures 2 and 3). CBD was chosen, since the side-chains of those seven positions are constrained and have a similar stereochemistry to those in a previously proposed model for a substrate bound to GroEL (Figure 1; Buckle *et al.*, 1997). Wild-type CBD does not bind detectably to sht-GroEL(191-345), but mutations at those seven positions (Table 1) lead to tight binding (Figure 5). Thus, we know which residues bind to GroEL and what is their overall stereochemistry.

Table 2. Sequence alignment of ''consensus'' CBD protein and mobile loop of GroES

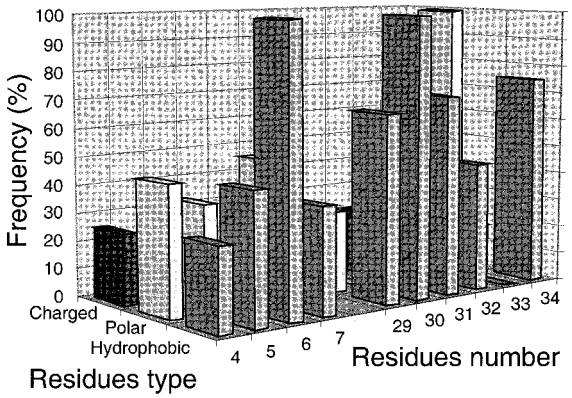
Name	Sequence
Consensus CBD	**t**q**s**RS**G**S**L**L**P****GS**SL
GroES (residues 16 to 32)	EVE**T**K**S**AG**G**IV**L**T**GS**AA

The consensus sequence corresponds to the more frequent residues at each position of the discontinuous patch (residues 4-7 and 29-34) plus residues 1-3 and 28. Residues not randomised in the CBD library (Smith *et al.*, 1998) are underlined. Residues adjacent to the mutated residues are in lowercase. Identical residues are in bold.

Nature of the GroEL-bound polypeptides

The nature of GroEL-polypeptide interactions is promiscuous (Landry & Gierasch, 1991a). The GroEL-polypeptide complexes are predominantly stabilised by hydrophobic interactions (Fenton *et al.*, 1994), although not always (Aoki *et al.*, 1997). Hydrophobic and positively charged side-chains tend to interact favorably with GroEL whereas negatively charged side-chains tend to repel (Itzhaki *et al.*, 1995). Overall electrostatic interactions also play a role in favoring the formation of GroEL-substrate complex (Perrett *et al.*, 1997). The diversity of sequences found from panning that bind minichaperone is consistent with the ability of GroEL to recognise a wide variety of unrelated polypeptides (Table 1). We find also that the recognised side-chains do not have to be totally hydrophobic, but polar and positively charged chains are accommodated (see Table 1).

Thus, GroEL recognises and binds a variety of polypeptides that share no obvious sequence similarity but is still able to discriminate between non-native and native proteins (Martin *et al.*, 1991; Mendoza *et al.*, 1992; Schmidt & Buchner, 1992). Therefore, the information required for recognition is not supplied only by the specific primary sequence. A more ''global'' mechanism of recog-



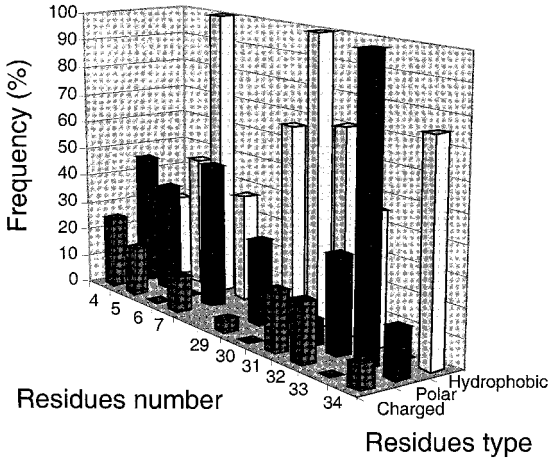


Figure 4. Frequency of amino acid type at each of the positions in the discontinuous patches in the domains generated by selection. The amino acids are divided into three major groups: those that are hydrophobic, those that are polar, and those that carry a positive charge (only one residue selected is negatively charged; see Table 1). The positions that were not randomised in the CBD library (Smith *et al.*, 1998) are included.

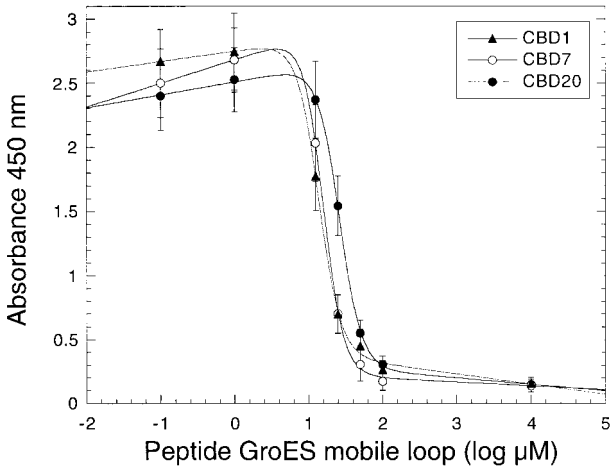


Figure 5. Inhibition of CBD-selected phages binding to minichaperone sht-GroEL(191-345) by varying concentrations of free synthetic peptide corresponding to residues 16 to 32 of GroES mobile loop determined by ELISA. See Tables 1 and 2 for the sequences.

nition should occur. The presence of solvent-exposed backbone peptide bonds and irregular (i.e. that comprises distinct regions of the polypeptide chain) hydrophobic surfaces is one common characteristic of non-native proteins. These surfaces could provide the necessary binding motif for recognition by the GroEL hydrophobic binding sites.

Conformation of GroEL-bound polypeptides

Extended β-strand

The selected constrained peptides exhibit a non-sequential linear binding motif with similar dimensions and chemical characteristics to the GroEL-bound mobile loop of GroES and the minichaperone GroEL(191-376)-bound peptide tag (Figure 1). The mobile loop of GroES that binds to the apical domains of GroEL adopts a characteristic β-hairpin turn (Xu *et al.*, 1997). NMR studies of the synthetic peptide representing the mobile loop of either GroES or Gp31, which differ in sequence, also adopt the same bulged hairpin conformation (Landry *et al.*, 1996). The crystal structure of minichaperone GroEL(191-376) shows the N-terminal peptide tag (sequence GLVPRGS) bound within the polypeptide-binding site (Buckle *et al.*, 1997) (Figure 1(b)). The seven residue (GLVPRGS) peptide binds to the same site as the mobile loop of GroES, and their conformations are strikingly similar (Figure 6(a)). Both adopt a roughly extended conformation that runs across the surface of two helices, H8 and H9, that form the binding site.

It could be argued that the binding of GLVPRGS in the crystal (Buckle *et al.*, 1997) is artefactual, and that fortuitously the peptide is binding to the GroES binding site and the real binding site for substrates is elsewhere in the protein. But, in the accompanying paper, it is shown that the binding site of minichaperones in solution for real denatured states spans the same region of structure as in the crystal (Tanaka & Fersht, 1999). In addition, we have recently demonstrated a genuine interaction between this peptide and GroEL minichaperone in solution (J.C. *et al.*, unpublished results) and our results indicate that selected CBD-phages bind, as does the synthetic GroES mobile loop peptide, within the polypeptide-binding site of GroEL minichaperone. Our findings suggest that the binding mode and/or site of GroES to GroEL are likely to mimic substrate binding.

It is clear that a preferred conformation of GroEL-bound substrates is extended. Both polar and non-polar interactions are a feature of GroES and peptide binding (Buckle *et al.*, 1997; Xu *et al.*,

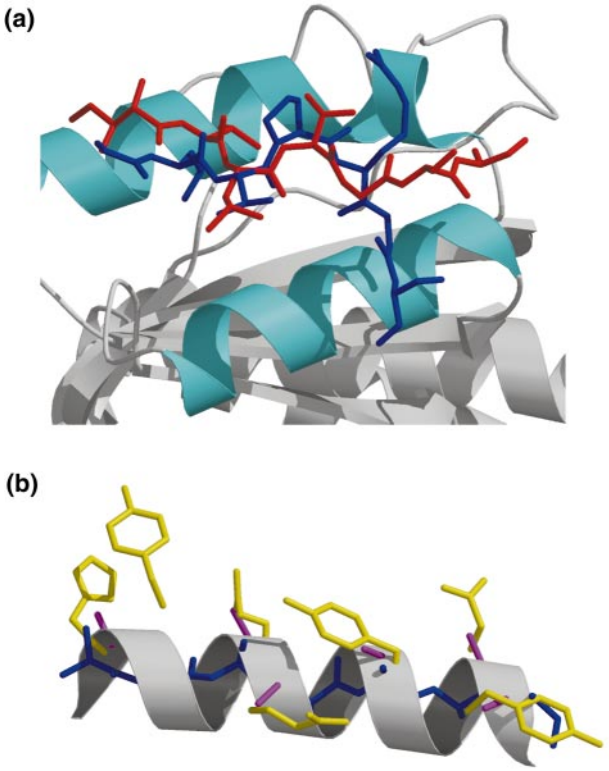


Figure 6. (a) Structural comparison of minichaperone-peptide complex (Buckle *et al.*, 1997) and the mobile loop of GroES bound to GroEL (Xu *et al.*, 1997). Residues 191 to 376 of the apical domain in the GroEL-GroES complex were least-squares fitted onto residues 191 to 376 of the minichaperone-peptide complex using LSQKAB (CCP4, 1994). The GroES mobile loop is in red, the minichaperone peptide is in blue. Helices H8 and H9 that make up the polypeptide binding site are coloured cyan. (b) Superposition of α-helix and β-strand and CBD side-chains. The C^{β} atoms of side-chains of CBD (yellow bonds), randomised in the bacteriophage library (Smith *et al.*, 1998), were manually fitted onto C^{β} atoms of helix H12 side-chains (residues 358 to 375) of the apical domain of GroEL (C^{β} atoms only, bonds drawn in magenta) and C^{β} atoms of a regular β-strand (blue bonds). The Figures are drawn to scale. Secondary structure representation is drawn with MolScript (Kraulis, 1991), and rendered with Raster3D (Merritt & Bacon, 1997).

1997). This is also consistent with preferential binding of an extended conformation: the hydrogen-bonding potential of the substrate backbone can be satisfied by polar residues in the active site (Buckle *et al.*, 1997).

α-Helical conformation

Not only is the spatial distribution of the seven randomised side-chains of CBD virtually superimposable with those of an extended β-strand, but they are also virtually superimposable with those in an α-helix (Figure 6(b)). Amphipathic α-helical structures are recognized by GroEL (Brazil *et al.*, 1997; Landry & Gierasch, 1991b). In addition, we show in the accompanying paper that the N-terminal fragment of rhodanese bound to minichaperone GroEL (193-335) adopts a helical conformation (Kobayashi *et al.*, 1999). Thus, the binding site of GroEL is compatible with binding exposed hydrophobic and positively charged side-chains in extended β-strands and α-helices. Our model shows that the polypeptide binding site of GroEL can accommodate an helix of approximately 18 residues (three to four turns) or an extended structure of approximately seven residues (Figure 6).

Structural and mechanistic implications

Specificity of binding of substrates

GroEL binds linear stretches of non-polar (mainly β-sheet formers; Ile, Phe, Val, Leu, and Trp), but also, polar and positively charged amino acids (including strong α-helix formers; Gln and Arg) that belong then to any type of secondary structure, or are even discontinuous in the polypeptide chain. For example, the side-chains of asparagine and glutamine residues could mimic a peptide backbone, fulfilling a dual role in the active site: aliphatic portions of their side-chains allow binding to hydrophobic side-chains whereas their polar groups can participate in hydrogen bonding with the substrate backbone. The presence of residues of varied shape, size and character at the binding site, which is of high flexibility, creates a very irregular and pitted surface that is highly adaptable to the binding of a multitude of substrates. Our data imply that GroEL can bind a wide range of structures, from extended β-strands and α-helices to folded states, with exposed side-chains. As suggested before (Richarme *et al.*, 1994), GroEL could act as an amphiphilic organiser of protein folding. Denatured states have the highest tendency to bind to GroEL because they have the exposed hydrophobic groups in extended strands and amphipathic helices that have the highest inherent binding affinity.

Passive mechanisms for unfolding

The binding of an extended conformation of the substrate not only implies a mechanism whereby GroEL discriminates between non-native and native proteins, but is also consistent with an unfolding activity of GroEL (Shtilerman *et al.*, 1999; Zahn *et al.*, 1996b) in the two following passive ways. First, specificity for binding an extended β-strand causes tension at the active site of GroEL or minichaperones to stretch a bound denatured hydrophobic patch into an extended β-strand conformation (Buckle *et al.*, 1997). Second, GroEL binds to α-helices in folding intermediates and misfolded proteins (Landry & Gierasch, 1991a; Kobayashi *et al.*, 1999). Since a likely feature of misfolded proteins would be incorrectly packed amphipathic helices that expose non-polar residues on their surface, then binding of the hydrophobic

faces of such helices would lead to a tension to disrupt such complexes. The ''tension'' in both cases is just a consequence of equilibrium binding.

In vivo, it would be unfavourable for GroEL to bind its natural protein substrates in their native state. Since many native proteins have exposed non-polar patches of random shape and size on the surfaces, a linear-shaped binding site that binds incorrectly packed helices and applies binding pressure towards an extended structure would effectively discriminate against native structures.

Active rack mechanism

The peptide and GroES binding sites form a ring around the inside top rim of the GroEL cylinder (Buckle *et al.*, 1997). A ring of contiguous binding sites has a great avidity for substrates that occupy multiple subsites. The mobile loop of GroES has evolved to bind to the polypeptide-binding site, effectively mimicking a bound peptide substrate. GroEL-bound peptides are displaced by GroES *via* a combination of competitive binding and movements of the apical domain (Xu *et al.*, 1997). On the binding of ATP, or ADP and GroES, this ring becomes expanded into a crown-like structure (Roseman *et al.*, 1996). The thrusting apart of subsites *via* the ATP-driven twisting and translation of the apical and intermediate domains (Xu *et al.*, 1997) weakens further the binding of substrates (Corrales & Fersht, 1996). We have proposed that such movements may forcibly unfold the protein by mechanical forces (Buckle *et al.*, 1997) as recently demonstrated (Shtilerman *et al.*, 1999). It is also possible that GroEL cannot completely unfold misfolded structures by binding pressure alone, but once more than one apical domain has locally unfolded segments of structure firmly bound, mechanical unfolding could further unfold the protein. This would be possible only if binding is tight, since the forces needed to disrupt the structures would be relatively large.

Conclusions

GroEL has two main activities: (i) the ability to allow the folding of aggregation-prone intermediates (such as molten globules), bound weakly by their exposed hydrophobic surfaces, in isolation within the central cavity (the ''folding cage'') (Weissman *et al.*, 1996); and (ii) the propensity to unfold kinetically trapped misfolded proteins that then have further attempts at productive folding (''iterative annealing'') (Todd *et al.*, 1994; Weissman *et al.*, 1994). These activities constitute a folding and annealing cage (Corrales & Fersht, 1996). GroEL binds linear stretches of non-polar, polar and positively charged side-chains that belong to any type of secondary structure, or are even discontinuous in the polypeptide chain. This widely applicable mode of binding allows the widest range of features of denatured states to be bound by GroEL.

Materials and Methods

Bacterial strains

The *E. coli* strains C41(DE3), a mutant of BL21(DE3) capable of expressing toxic genes (Miroux & Walker, 1996), and TG1 (Gibson, 1984) were used.

Proteins expression and purification

The histidine-tagged (short histidine tail; sht)-minichaperones GroEL(191-345) wild-type and the mutant Y203E (Chatellier *et al.*, 1998) were expressed by inducing the phage T7 promoter of pRSETA vectors with isopropyl-β-D-thiogalactoside (IPTG) in *E. coli* C41(DE3) cells and purified essentially as described (Buckle *et al.*, 1997; Zahn *et al.*, 1996a). GroE proteins, ~57.5 kDa GroEL and ~10 kDa GroES, were expressed and purified as previously described (Chatellier *et al.*, 1998).

Peptide library

The cellulose-binding domain (CBD) library has been described (Smith *et al.*, 1998). The seven residues forming the discontinuous cellulose-binding patch (Linder *et al.*, 1995) on the flat protein surface of the C-terminal CBD of cellobiohydrolase I from the fungus *T. reesei* (Kraulis *et al.*, 1989) were randomised (Smith *et al.*, 1998). The phagemid particles, displaying a large repertoire (5.5×10^8) of mutant CBDs fused to the g3 protein of the filamentous phage fd (Smith *et al.*, 1998) produced by rescuing the library with VCS-M13 helper phage (Stratagene), were screened.

Selection

Purified sht-GroEL(191-345) minichaperone was coated on immunotubes (Nunc, Maxisorb) overnight at 4 °C at a concentration of 10 μg/ml in bicarbonate buffer (50 mM $NaHCO_3$, pH 9.6). The efficiency of the coating was verified by ELISA (enzyme-linked immunosorbant assay) using horseradish peroxidase conjugated anti-GroEL antibody (Sigma) with 3′,3′,5′,5′-tetramethylbenzidine (TMB, Boehringer Mannheim) as substrate. The refolding capacity of minichaperone sht-GroEL(191-345) coated on immunotube for cyclophilin A renaturation was verified essentially as described (Altamirano *et al.*, 1997).

The phage repertoire was selected by panning on coated immunotubes essentially as described (Smith *et al.*, 1998). Acid was used to elute bound bacteriophages: 5.1×10^{12} transducing units (t.u.) phage particles were used for the first round of selection and approximately 1×10^{12} t.u. for the subsequent rounds.

Screening of phage-displayed peptides by ELISA

Single ampicillin resistant colonies resulting from infection of *E. coli* TG1 by phages liberated after selection were screened to identify those producing minichaperone-binding phage by ELISA essentially as described (Clackson *et al.*, 1991; Smith *et al.*, 1998). Binding was detected using horseradish peroxidase conjugated anti-M13 antibody (Pharmacia), diluted in phosphate-buffered saline (PBS; 25 mM NaH_2PO_4, 125 mM NaCl, pH 7.0), with 3′,3′,5′,5′-tetramethylbenzidine (TMB, Boeh-

ringer Mannheim) as a substrate. Reactions were stopped with 50 µl of 1 M H_2SO_4 after ten minutes and readings taken by subtracting A_{650} from the A_{450}.

Sequencing of selected peptides

Individual clones were amplified by PCR (polymerase chain reaction) using *Taq* DNA polymerase (Promega) and oligonucleotides pelBB.lmb3 and g3fwd.lmb2 (Smith *et al*., 1998). PCR cycle sequencing reactions with fluorescent dideoxy chain terminators (Applied Biosystems) were carried out according to the manufacturer's instructions, using oligonucleotide lmb3 (Smith *et al*., 1998). Sequencing reactions were analysed on an Applied Biosystems 373A Automated DNA Sequencer and sequence analysis was performed using SeqEd (Applied Biosystems) and MacVector (IBI Kodak, New Haven, CT).

Peptide synthesis and competition assays

A peptide corresponding to the mobile loop of GroES (residues 16 to 32, numbered as described by Hemmingsen *et al*., 1988) was synthesised on a Synergy Personal Peptide Synthesizer (Applied Biosystems) which performs solid-phase synthesis using 9-fluoromethylcarbonyl (Fmoc) protection with free termini. Cleavage from the resin was performed with trifluoroacetic acid (TFA)/phenol/thioanisol/ethanethiodiol/water mixture and the peptide purified by HPLC (high pressure liquid chromatography) and characterised by mass spectrometry. The purified peptide was lyophilized and stored at −20 °C.

Phages were purified using PEG (polyethylene glycol) as described (McCafferty *et al*., 1990). To verify the structural integrity of peptide displayed, phages were treated with 1% (v/v) β-mercaptoethanol in PBS for one hour prior to minichaperone binding by ELISA. The inhibition of the binding of the phage displayed CBD by the free peptide was analysed by phage-ELISA as described above, adding different concentrations (between 10,000 and 0.01 µM) of free peptide dissolved in 0.1% (v/v) TFA solution to the phage supernatants.

Acknowledgements

We thank Dr G. Smith for the availability of the peptide library prior to publication. Postdoctoral fellowships from FEBS (until 6/97) and the European Union (from 7/1997) to J.C. are gratefully acknowledged.

References

Adey, N. B., Mataragnon, A. H., Rider, J. E., Carter, J. M. & Kay, B. K. (1995). Characterization of phage that bind plastic from phage-displayed random peptide libraries. *Gene,* **156**, 27-31.

Altamirano, M. M., Golbik, R., Zahn, R., Buckle, A. M. & Fersht, A. R. (1997). Refolding chromatography with immobilized mini-chaperones. *Proc. Natl Acad. Sci. USA,* **94**, 3576-3578.

Aoki, K., Taguchi, H., Shindo, Y., Yoshida, M., Ogasahara, K., Yutani, K. & Tanaka, N. (1997). Calorimetric observation of a GroEL-protein binding reaction with little contribution of hydrophobic interaction. *J. Biol. Chem.* **272**, 32158-32162.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. & Gething, M.-J. H. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell,* **75**, 717-728.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2. 8 Å. *Nature,* **371**, 578-586.

Brazil, B. T., Cleland, J. L., McDowell, R. S., Skelton, N. J., Paris, K. & Horowitz, P. M. (1997). Model peptide studies demonstrate that amphipathic secondary structures can be recognized by the chaperonin GroEL (cpn60). *J. Biol. Chem.* **272**, 5105-5111.

Buckle, A. M., Zahn, R. & Fersht, A. R. (1997). A structural model for GroEL-polypeptide recognition. *Proc. Natl Acad. Sci. USA,* **94**, 3571-3575.

Bukau, B. & Horwich, A. L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell,* **92**, 351-366.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Chatellier, J., Hill, F., Lund, P. & Fersht, A. R. (1998). *In vivo* activities of GroEL minichaperones. *Proc. Natl Acad. Sci. USA,* **95**, 9861-9866.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature,* **352**, 624-628.

Corrales, F. J. & Fersht, A. R. (1996). Toward a mechanism for GroEL-GroES chaperone activity: an ATPase-gated and -pulsed folding and annealing cage. *Proc. Natl Acad. Sci. USA,* **93**, 4509-4512.

Fenton, W. A. & Horwich, A. L. (1997). GroEL-mediated protein folding. *Protein Sci.* **6**, 743-760.

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

Gibson, T. J. (1984). Studies on the Epstein-Barr virus genome. PhD thesis, University of Cambridge.

Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. (1989). GroE heat shock proteins promote assembly of foreign prokaryotic ribalose *bis* phosphate carboxylate oligomers in *Escherichia coli*. *Nature,* **337**, 44-47.

Gragerov, A., Zeng, L., Zhao, X., Burkholder, W. & Gottesman, M. E. (1994). Specificity of DnaK-peptide binding. *J. Mol. Biol.* **235**, 848-854.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature,* **381**, 571-580.

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

Hightower, L. E., Sadis, S. E. & Takenaka, I. M. (1994). In *The Biology of Heat Shock Proteins and Molecular Chaperones*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Horovitz, A. (1998). Structural aspects of GroEL function. *Curr. Opin. Struct. Biol.* **8**, 93-100.

Hunt, J. F., van der Vies, S., Henry, L. & Deisenhofer, J. (1997). Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell,* **90**, 361-371.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). Nature and consequences of GroEL-protein interactions. *Biochemistry,* **34**, 14581-14587.

Kobayashi, N., Freund, S. M. V., Chatellia, J., Zahn, R. & Fersht, A. R. (1999). NMR analysis of the binding

of a rhodanere peptide to a minichaperone in solution. *J. Mol. Biol.* **292**, 181-190.

Kraulis, P. (1991). MolScript, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989). Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry,* **28**, 7241-7257.

Landry, S. J. & Gierasch, L. M. (1991a). Recognition of nascent polypeptides for targeting and folding. *Trends Biochem. Sci.* **16**, 159-164.

Landry, S. J. & Gierasch, L. M. (1991b). The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry,* **30**, 7359-7362.

Landry, S. J., Zeilstra, R. J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. (1993). Characterization of a functionally important mobile domain of GroES. *Nature,* **364**, 255-558.

Landry, S. J., Taher, A., Georgopoulos, C. & van der Vies, S. (1996). Interplay of structure and disorder in cochaperonin mobile loops. *Proc. Natl Acad. Sci. USA,* **93**, 11622-11627.

Le Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R. A., Favel, A. & Coletti-Previero, M. A. (1990). Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie,* **72**, 431-435.

Linder, M., Mattinen, M.-L., Kontelli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G. & Annila, A. (1995). Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci.* **4**, 1056-1064.

Lorimer, G. H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding *in vivo*. *FASEB J.* **10**, 5-9.

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L. & Hartl, F. U. (1991). Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature,* **352**, 36-42.

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature,* **348**, 552-554.

Mendoza, J. A., Butler, M. C. & Horowitz, P. M. (1992). Characterization of a stable, reactivatable complex between chaperonin 60 and mitochondrial rhodanese. *J. Biol. Chem.* **267**, 24648-24654.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289-298.

Peres Ben-Zvi, A., Chatellier, J., Fersht, A. R. & Goloubinoff, P. (1998). Minimal and optimal mechanisms for GroE-mediated protein folding. *Proc. Natl Acad. Sci. USA,* **95**, 15275-15280.

Perrett, S., Zahn, R., Stenberg, G. & Fersht, A. R. (1997). Importance of electrostatic interactions in the rapid binding of polypeptides to GroEL. *J. Mol. Biol.* **269**, 892-901.

Richarme, G., el Yaagoubi, A., de Crouy-Chanel, A. & Kohiyama, M. (1994). Specific interaction of the *Escherichia coli* chaperone GroEL (60-kDa heat shock protein) with the liganded form of the galactose binding protein. *Biochem. Mol. Biol. Intern.* **34**, 955-961.

Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. (1996). The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell,* **87**, 241-251.

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

Schmidt, M. & Buchner, J. (1992). Interaction of GroE with an all-beta-protein. *J. Biol. Chem.* **267**, 16829-16833.

Shtilerman, M., Lorimer, G. H. & Englander, S. W. (1999). Chaperonin function: folding by forced unfolding. *Science,* **284**, 822-825.

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* **228**, 1315-1317.

Smith, G., Patel, S. U., Windass, J. D., Thornton, J. M., Winter, G. & Griffiths, A. D. (1998). Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J. Mol. Biol.* **277**, 317-332.

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

Tanaka, N. & Fersht, A. R. (1999). Identification of substrate binding site of GroEL minichaperone in solution. *J. Mol. Biol.* **292**, 173-180.

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

Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. & Horwich, A. L. (1996). Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell,* **84**, 481-490.

Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. *Nature,* **388**, 741-750.

Zahn, R., Buckle, A. M., Perrett, S., Johnson, C. M. J., Corrales, F. J., Golbik, R. & Fersht, A. R. (1996a). Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proc. Natl Acad. Sci. USA,* **93**, 15024-15029.

Zahn, R., Perrett, S., Stenberg, G. & Fersht, A. R. (1996b). Catalysis of amide proton exchange by the molecular chaperones GroEL and SecB. *Science,* **271**, 642-645.

Edited by J. Karn

(Received 20 May 1999; received in revised form 16 July 1999; accepted 18 July 1999)