ELSEVIER

Biophysical Chemistry 100 (2003) 453–467

Biophysical Chemistry

www.elsevier.com/locate/bpc

Annealing function of GroEL: structural and bioinformatic analysis

George Stan[a], D. Thirumalai[b,*], George H. Lorimer[c], Bernard R. Brooks[a]

[a]*Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA*

[b]*Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

[c]*Center for Biomolecular Structure and Organization, Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*

Received 5 April 2002; accepted 12 June 2002

Abstract

The *Escherichia coli* chaperonin system, GroEL–GroES, facilitates folding of substrate proteins (SPs) that are otherwise destined to aggregate. The iterative annealing mechanism suggests that the allostery-driven GroEL transitions leading to changes in the microenvironment of the SP constitutes the annealing action of chaperonins. To describe the molecular basis for the changes in the nature of SP–GroEL interactions we use the crystal structures of GroEL (T state), GroEL–ATP (R state) and the GroEL–GroES–$(ADP)_7$ (R″ state) complex to determine the residue-specific changes in the accessible surface area and the number of tertiary contacts as a result of the $T \rightarrow R \rightarrow R''$ transitions. We find large changes in the accessible area in many residues in the apical domain, but relatively smaller changes are associated with residues in the equatorial domain. In the course of the $T \rightarrow R$ transition the microenvironment of the SP changes which suggests that GroEL is an annealing machine even without GroES. This is reflected in the exposure of Glu386 which loses six contacts in the $T \rightarrow R$ transition. We also evaluate the conservation of residues that participate in the various chaperonin functions. Multiple sequence alignments and chemical sequence entropy calculations reveal that, to a large extent, *only the chemical identities and not the residues themselves* important for the nominal functions (peptide binding, nucleotide binding, GroES and substrate protein release) are strongly conserved. Using chemical sequence entropy, which is computed by classifying aminoacids into four types (hydrophobic, polar, positively charged and negatively charged) we make several new predictions that are relevant for peptide binding and annealing function of GroEL. We identify a number of conserved peptide binding sites in the apical domain which coincide with those found in the 1.7 Å crystal structure of 'mini-chaperone' complexed with the N-terminal tag. Correlated mutations in the HSP60 family, that might control allostery in GroEL, are also strongly conserved. Most importantly, we find that charged solvent-exposed residues in the T state (Lys 226, Glu 252 and Asp 253) are strongly conserved. This leads to the prediction that mutating these residues, that control the annealing function of the SP, can decrease the efficacy of the chaperonin function.

Keywords: GroEL; *Escherichia coli*; Annealing; Chaperonin function

*Corresponding author.
E-mail address: thirum@glue.umd.edu (D. Thirumalai).

0301-4622/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(02)00298-3

1. Introduction

The chaperonin machinery, GroEL and GroES, helps folding of newly synthesized proteins in *Escherichia coli* and facilitates in vitro refolding of a number of structurally unrelated proteins under conditions (non-permissive) that promote aggregation or trapping in misfolded states [1]. Developments in protein folding theory [2–4], in vitro biophysical studies of GroEL-assisted folding [5–7], and the determination of the structures of GroEL [8], GroES [9], GroEL–ATP complex [10], and the GroEL–GroES–$(ADP)_7$ complex [11] have enhanced our understanding of this class of nanomachines. The iterative annealing mechanism (IAM), which incorporates these developments [12] dissects the workings of this ATP-driven nanomachine into three distinct parts: (i) The ATP- and GroES-induced changes in the architecture of GroEL that take place in the course of the chaperonin cycle [13]. (ii) The fate of the substrate protein (SP) as a result of capture and encapsulation by the nanomachine. (iii) The coupling between the allosteric transitions of the GroEL system and the folding kinetics of the SP [5,14]. An exploration of the annealing action, at the molecular level, must incorporate these three major events that occur in the chaperonin cycle.

The class I chaperonins, belonging to the HSP60 family, consist of two heptameric rings stacked back-to-back that communicate with one another and operate in a coordinated manner. The presence of a large central cavity in GroEL prompted the suggestion that the sole function of GroEL is to act as a passive Anfinsen cage, i.e. to sequester the SP and allow folding to occur in isolation. The IAM invokes an active role of GroEL. Annealing of the SP occurs as a result of the *changes in the microenvironment felt by the SPs*. The nature of interactions between SP and GroEL changes because of the spectacular allosteric transitions that GroEL undergoes in the ATP-driven cycle. Thus, the fundamental annealing action is intimately linked to the large conformational changes in GroEL. Iterative annealing is often required to get sufficient yield of the folded SP, i.e. n, the number of binding and release events, is usually greater than one [15]. It has recently been noted that the folding rate of Rubisco, which is recognized to be a stringent SP for GroEL, increases by a factor of four in a single ring mutant of GroEL [16]—one from which ADP and GroES are not released. In this experiment [16] there is a change in the microenvironment of the SP because ATP and GroES binding have triggered the $T \rightarrow R \rightarrow R''$ transitions. Hence, the model proposed in [16] *does not correspond to the original Anfinsen cage picture*, but is a special case of IAM with $n=1$.

Although the active role played by GroEL is not in dispute, the molecular basis for the annealing action (the way the SP is folded by interactions with GroEL and GroES) has not been fully clarified. Much attention has focused on the 'nominal' functions of GroEL. These are SP binding, ATP hydrolysis, and binding and release of GroES. Residues that control these functions have been identified using mutagenesis experiments [17] and structural analysis [18,19]. However, the most crucial function of GroEL is to assist folding of a number of structurally unrelated proteins that are kinetically trapped. Computational studies [14] show that alternation of the cavity lining from hydrophobic to hydrophilic is sufficient to 'smooth out' the energy landscape of the SP. In this process the SP undergoes unfolding [18,20] (locally or globally) by a force transmitted to it by GroEL domain movements [21]. Whatever the mechanisms of producing a favorable SP free energy landscape may be, the crystal structures of GroEL [8] (referred to as T-state from now on), the Cryo-EM picture of the GroEL–ATP complex (R state) [10], and GroEL–GroES–$(ADP)_7$ (R″ state) [11] can be used to analyze residue specific movements that provide the requisite changes in the character of inner walls of GroEL.

Of particular importance in the chaperonin cycle is the chemical character of the residues lining the surface of the GroEL central cavity. In the T state, hydrophobic residues that line the mouth of the cavity facilitate the capture and sequestration of the SP, while in the R″ state hydrophilic residues line the cavity walls. A less spectacular, but significant, change in the wall character of GroEL occurs as a result of $T \rightarrow R$ transition. These structural changes are accompanied by dramatic domain shifts in the GroEL *cis* ring, especially

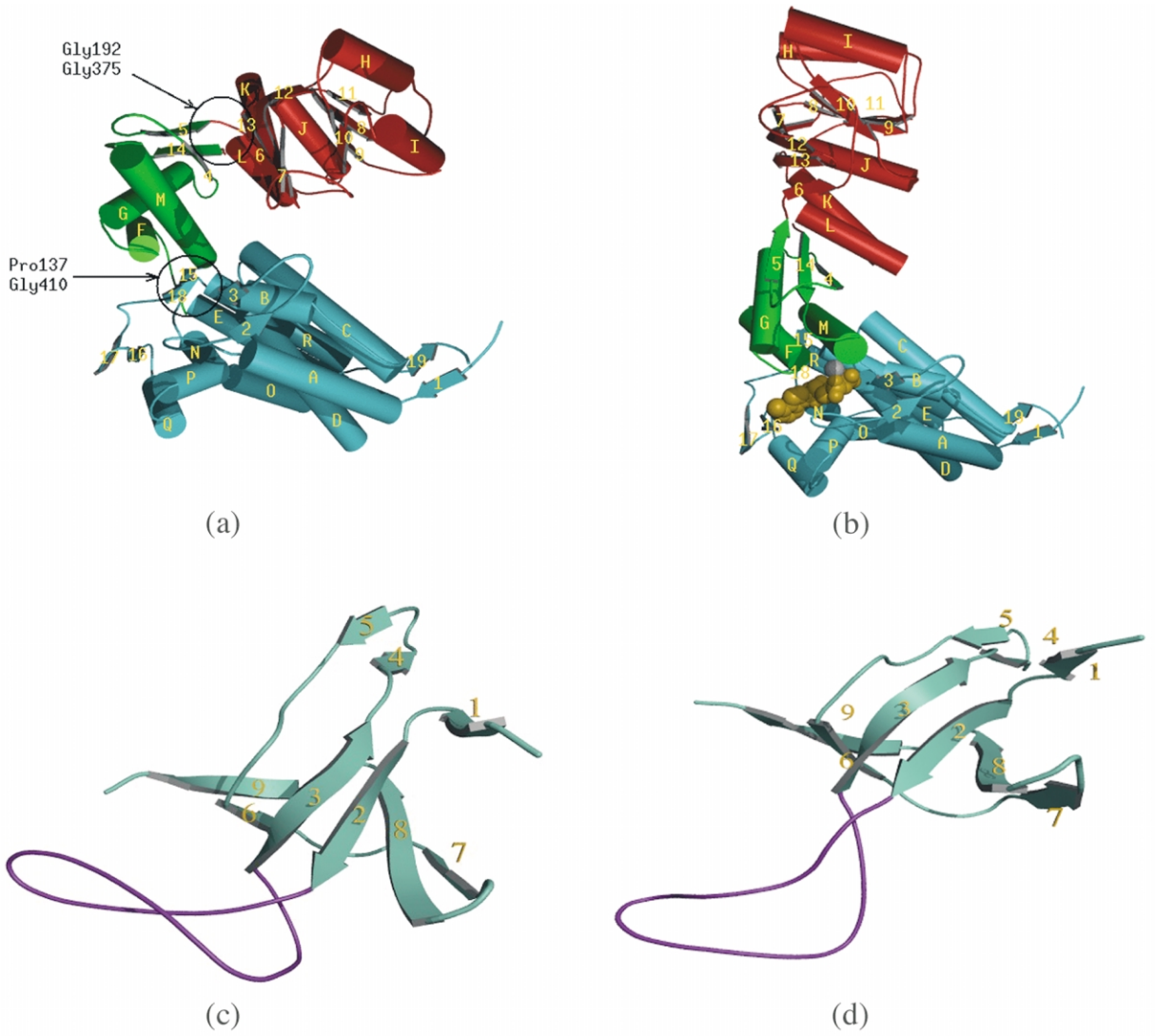

Fig. 1. (a,b) Schematic drawing of one subunit in the GroEL *cis* ring in the free (a) and liganded (b) state. Colors represent the equatorial domain (blue), the intermediate domain (green) and the apical domain (red). Cylinders represent α-helices, arrows β-strands, and continuous lines extended strands. Pivot points for domain movement are indicated by circles and arrows. (c,d) One subunit of the GroES molecule: (c) unbound [28]; (d) bound to GroEL [11]. All secondary structure elements are represented in green, except for the 'mobile loop' (purple). Figures (c) and (d) were produced using Molscript [33], Povscript [34], Rasmol [35,36] and PovRay [37]. Figures (a), (b), and (d) were adapted from [11].

those residues associated with the SP-binding sites in the apical (A) domain, the nucleotide binding sites in the equatorial (E) domain (Fig. 1a,b), and the residues that control the annealing function of GroEL. The structure of the GroEL *trans* ring is only slightly modified between these states. The co-chaperonin GroES undergoes significant structural transformations between its unbound and bound state, mainly in the 'mobile loop' responsible for the formation of the GroEL–GroES interface (Fig. 1c,d). The purpose of this note is to analyze the role of these residues in the annealing

action of GroEL, and to correlate these with the degree of evolutionary conservations in related HSP60 sequences. We show that several previously unidentified residues, which are dispersed throughout the GroEL structure, are crucial to the overall function of chaperonins. We predict that mutating the chemical identity of residues that provide a hydrophilic lining of the GroEL cavity in the R″ state can be deleterious to its function.

2. GroEL residues that control the annealing function of the SP

The annealing action, that stochastically alters the free energy landscape of the SP, is due to switches in the nature of residues that line the inner cavity of the GroEL particle. Residue-specific structural movements can be characterized through the changes in the area per residue accessible to solvent molecules. The algorithm used for the calculation of the accessible area per residue, due to Lee and Richards [22], employs a computer-generated space-filling model of the protein located on a three-dimensional grid. The surface area of a residue exposed to solvent is determined as the locus of the center of a solvent molecule which makes maximum contact with the residue atoms. To compare the accessible area for different types of amino acids, we used

$$r_s = a_s / a_0 \tag{1}$$

where a_s is the accessible surface area of the amino acid *X* in the polypeptide chain, and a_0 is the reference accessible surface area of the amino acid in an extended Gly-*X*-Gly tripeptide [23]. A small value of r_s implies that the residue is deeply buried in the polypeptide chain structure, while a value of the order 1 is typical of residues on the protein surface. We compared the accessible area of amino acids in the liganded and unliganded states of both GroEL and GroES. We define the relative change of accessible surface area between the R″ and the T states as

$$\Delta = \frac{r_s(\mathrm{R}'') - r_s(\mathrm{T})}{\mathrm{R}_s(\mathrm{T})} \times 100 \tag{2}$$

where $\mathrm{R}_s(\mathrm{T}) = \max\{r_s(\mathrm{T})\}$, is the largest accessibility ratio for residues in the T state. In the case of GroEL, $\mathrm{R}_s(\mathrm{T}) = 1.08$, corresponding to Ser 43 ($a_s(\mathrm{T}) = 131$ Å^2), while for GroES, $\mathrm{R}_s(\mathrm{T}) = 1.09$, corresponding to Ala 22 ($a_s(\mathrm{T}) = 123$ Å^2). Significant changes in the accessible surface area are associated with $|\Delta| \geq 20\%$.

To further analyze the structural changes in the $\mathrm{T} \rightarrow \mathrm{R} \rightarrow \mathrm{R}''$ transitions we also computed the residue specific changes in the number of contacts

$$\Delta Z_i = \frac{Z_i(\alpha_{j+1}) - Z_i(\alpha_j)}{Z_i(\alpha_j)} \tag{3}$$

where α_1, α_2 and α_3 correspond to T, R and R″ states, respectively. The number of contacts is

$$Z_i(\alpha_j) = \sum_{k \neq i} \Theta(\mathrm{R}_{\mathrm{cut}} - r_{ik}) \tag{4}$$

where r_{ik} is the minimum distance between heavy atoms in the side chains of residues *i* and *k*, $\mathrm{R}_{\mathrm{cut}} =$ 5.2 Å is the cutoff distance for contact, and $\Theta(x)$ [$\Theta(x) = 1$ if $x \geq 0$ and is 0 otherwise] is the step function. Quantitative assessment of the crucial structural changes accompanying the concerted allosteric transitions can be made using ΔZ_i and Δ.

To identify the residues responsible for the changes in the inner surface of GroEL, we determined Δ per residue in each domain of a GroEL *cis* ring following the $\mathrm{T} \rightarrow \mathrm{R}''$ transition using the T and R″ crystal structures (Fig. 1). The residues found on the external surface of the GroEL dome in the R″ state were ignored, as this surface is thought not to play an active role in the chaperonin cycle. For the present discussion, the secondary structural elements of interest also include residues at neighboring sequence positions.

2.1. Apical domain (*residues 191–376*)

A number of residues in the apical domain has a reduced exposure to the solvent ($\Delta < 0$), partly due to the bias imposed against the external surface residues and partly due to the formation of the GroEL–GroES interface (Fig. 2a). The GroES binding sites [17], i.e. residues in helices H and I, are substantially less exposed to the solvent in the R″ state. Among the established binding sites Leu234 undergoes the largest change in Δ. Our

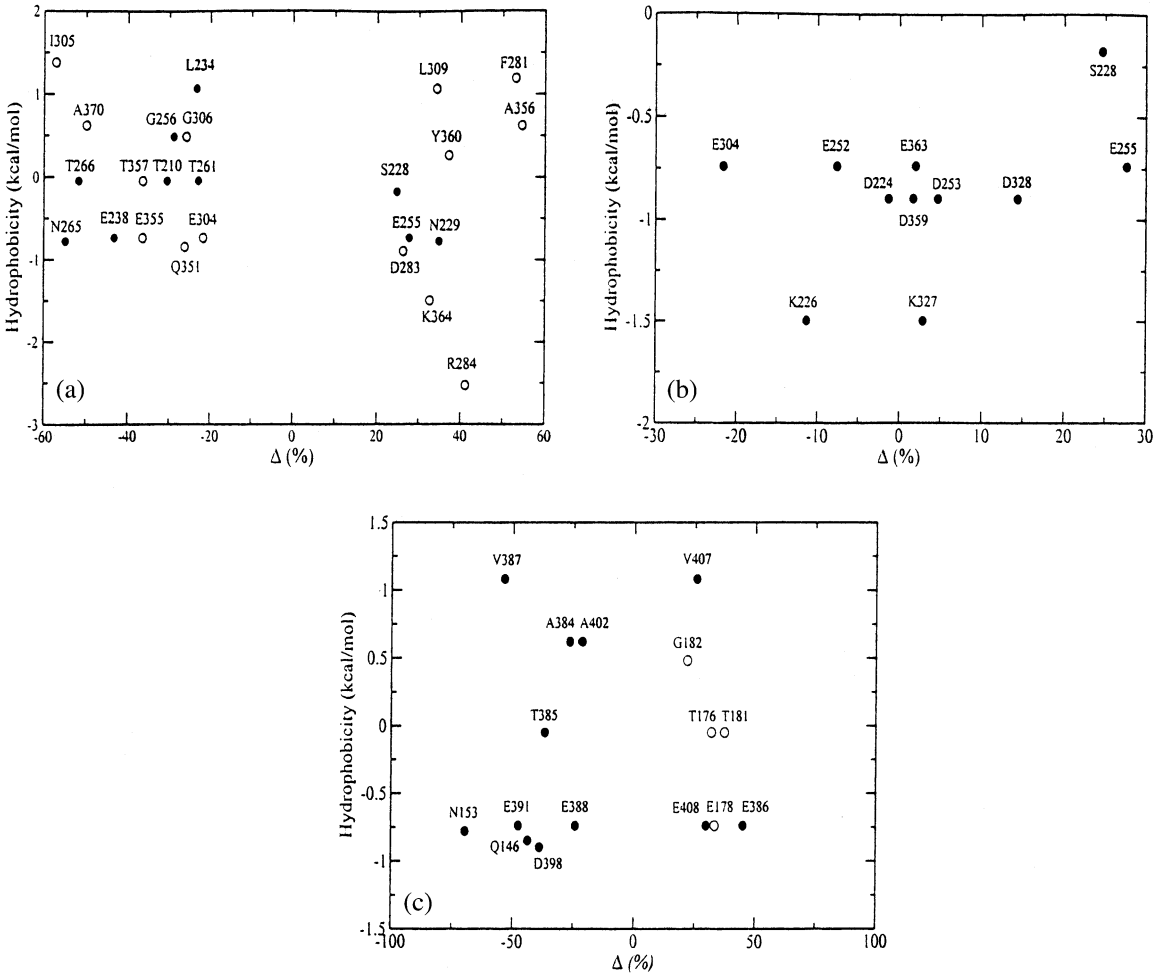

Fig. 2. Residues with significant changes (except for part b) in accessible area in the liganded vs. unliganded GroEL: (a) the apical domain. Filled circles indicate residues in the long loop between strands 6 and 7 and helices H and I. Other residues are shown with open circles. (b) A set of hydrophilic residues found on the cavity lining in the R″ state. (c) The intermediate domain. Filled symbols represent residues on or near helices F and M. Normalized hydrophobicity values were taken from Eisenberg et al. [38].

analysis also suggests other residues in helices H and I, namely, Gly256, Thr261, Asn265 and Thr266 also have large negative values of Δ. Recently, Buckle et al. [19], have proposed Thr261 and Asn265, both of which are in helix I, as SP binding sites. Other residues that are SP binding sites (Leu237, Leu259, Val263 and Val264) move considerably less in the course of the T→R″ transition. Chen and Siegler [18] have speculated, based on crystal structure of peptides bound to the apical domain, that the segment 301–311 may indirectly contribute to peptide binding. We find that Glu304, Ile305, Gly306. ($\Delta<0$) and Leu309 ($\Delta>0$) undergo substantial fluctuations in the course of the T→R″ transition.

We have also focused on a set of hydrophilic residues found in the cavity lining in the R″ state [11]: Asp 224, Lys 226, Ser 228, Glu 252, Asp 253, Glu 255, Glu 304, Lys 327, Asp 328, Asp 359 and Glu 363. In this case, no threshold values were imposed for the minimum surface area per residue in the T state, nor regarding the change in

accessible area between the T and R″ state. With the exception of Ser228, Glu257 and Glu304 all other (Fig. 2b) have relatively small changes in accessible area. The movement in most of the hydrophilic residues results in the SP experiencing a hydrophilic environment in R″. The mechanism of switching the chemical character of the cavity lining consists of the withdrawal of the hydrophobic residues into the GroEL–GroES interface and the small displacement of hydrophilic residues to the cavity surface. Because the hydrophilic residues contribute to changes in the wall character we predict that they are important in the annealing function of GroEL, i.e. drastic changes in these residues might reduce the efficiency of chaperonins. We have previously suggested [12] that, because the cavity lining is likely to be less hydrophobic in the R (GroEL–ATP complex) than in the T state, GroEL can serve as an annealing nanomachine for some SPs even without GroES. Recently, Ranson et al. [10] have obtained a 10 Å map of the R state using cryo-EM. Assuming that domains execute rigid body movements they fitted the low-resolution map to atomic structures to get an estimate of the coordinates of the R state. We find that even in the R state a sufficient number of hydrophilic residues line the cavity so that the interaction with the SP is reduced. This observation is reflected in the number of contacts some of the hydrophilic residues make in the two allosteric states (Fig. 3a). For example, Asp224 and Lys225 have two more contacts in the R state than in the T state. Similarly, Val263 and Val264 have lesser contacts in the R state than in the T state. This suggests that these hydrophobic residues have moved away from the inner lining and are replaced by hydrophilic residues. We should emphasize, however, that the major changes in the wall character take place only upon GroES binding [either R′ (GroEL–GroES–$(ATP)_7$] or R″ state. This is reflected in ΔZ_i for several residues. For example, Glu257 and Glu304 gain contacts while hydrophobic residues Leu259, Phe204 (loses six contacts), and Ile205 lose contacts. Thus, for stringent SPs the full machinery is required.

A spectacular change in the number of contacts occurs for Glu386 which forms an inter-subunit salt bridge with Arg197 of the adjacent apical domain. The number of contacts that Glu386 makes *decreases from seven in the T state to just 1 in the R state*. This creates local instability in the apical domain which facilitates the observed counterclockwise rotation in the T→R transition. Transition to the R state also weakens the SP-GroEL interaction in the R state. Glu386 gains a contact in the R→R″ transition.

2.2. *Intermediate domain* (*residues 134–190 and 377–408*)

Helices F (residues 141–151) and M (residues 386–409) are responsible for closing the nucleotide-binding site by clamping onto the equatorial domain upon GroES binding [11]. Two charged residues, with $\Delta > 0$ Glu 386 and Glu408 are found in the cavity lining in the R″ state. However, many residues in helix M exhibit a large loss of accessible area whereas $\Delta < 0$ only for two residues in helix F (Fig. 2c). A large negative Δ value is found for Asp398, a residue that is required for ATP hydrolysis. Because Asp398 gains contacts only after R→R″ (or R′) transition we conclude that coordination with Mg^{2+} occurs only upon GroES binding. Residues Thr181 and Gly182 (more exposed), are associated with elements close to the hinge between the apical and intermediate domain, namely helix G and strands 4 and 5. The domain movements in the T→R transition also lead to loss in contacts in a number of residues; Thr181 loses four contacts whereas Gly182 loses two. Cys138, which is correlated with Cys519 [24], has four fewer contacts in the R state than in the T state. The number of contacts that I-domain residues make does not change significantly in the R→R″ transition. The exception is Asn153 which gains four contacts. This increase is also reflected in the large negative value of Δ. It has been shown that mutations of Ile150, Ser151 and Ala152 compromise ATPase activity. Our analysis suggests that the neighboring Asn153 might also be an interesting candidate for future mutagenesis experiments.

2.3. *Equatorial domain* (*residues 6–133 and 409–548*) [*25*]

The equatorial domain moves very little during the chaperonin cycle. However, important changes

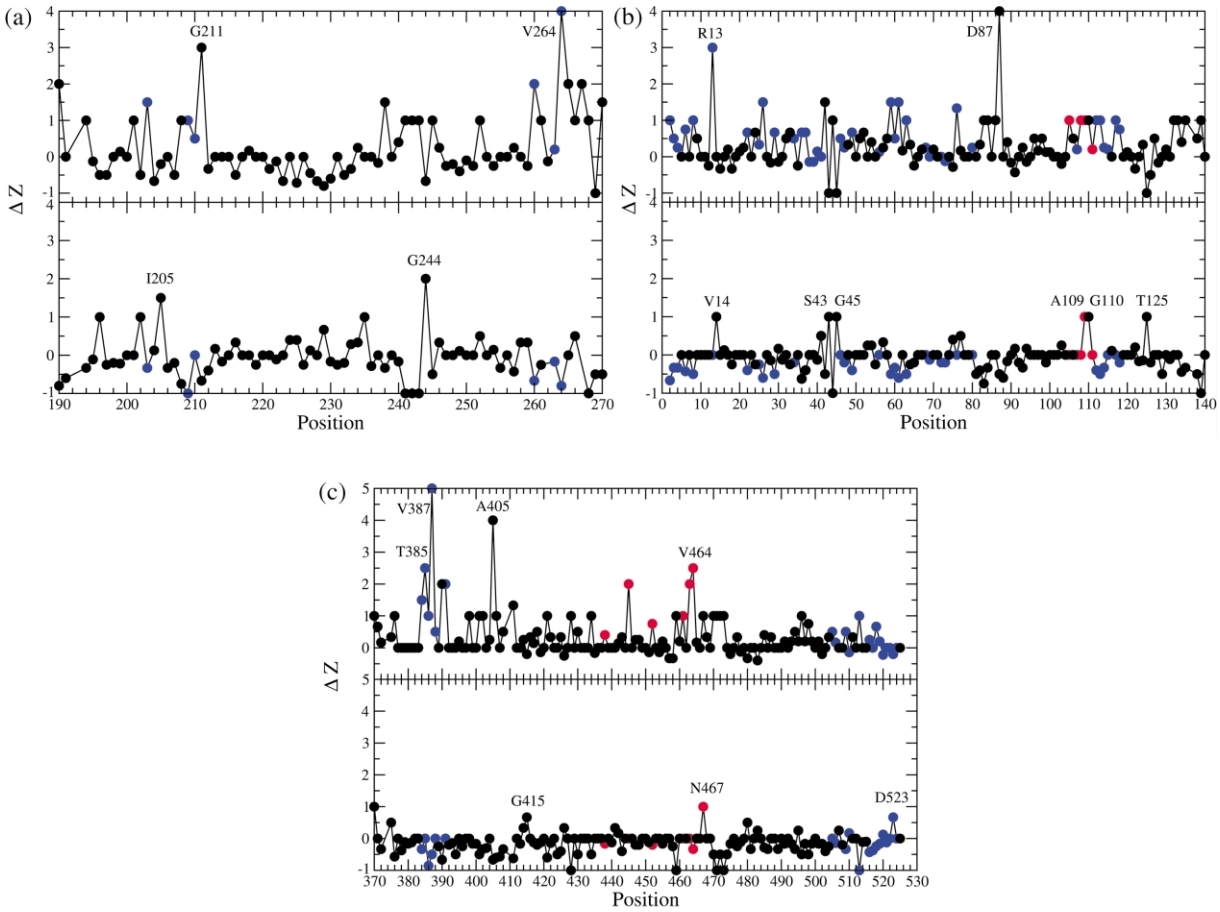

Fig. 3. Relative change in the number of contacts for residues in the GroEL *cis* ring following the T→R transition (lower panel) and the R→R″ transition (upper panel): (a) Apical domain; (b) and (c) Equatorial domain. Red (Blue) dots represent residues with inter-ring (inter-subunit) contacts in the R″ state.

in the accessible area occur due to the motion of the intermediate domain. Two equatorial domain structural elements are closely involved in nucleotide-binding, the stem loop (residues 34–52) and helix C (residues 65–85). Analysis of equatorial domain residues with a large Δ shows almost total absence of accessible area gains ($\Delta > 0$) in this domain. All of the residues in the stem loop and helix C become significantly less exposed. Among the remaining residues, two categories can be distinguished: residues near the nucleotide-binding site (Val54, Gly86, Asp87, Pro113, Leu494, Tyr506 and Leu513), and residues participating in the inter–ring interface (Gly9, Lys15, Val107, Val438, Glu461, Ser463 and Val464). The large change of accessible area of residues involved in the inter–ring interface may be an artifact of the two structures considered in this study: the unbound GroEL structure has a single heptameric ring, while the GroEL–GroES–$(ADP)_7$ complex contains both GroEL rings. Nevertheless, this difference allows us to examine the inter–ring interface. Three of the residues found through our approach are among those with a major role in stabilizing this interface: Glu461, Ser463 and Val464 [26]. In addition, the mutation E461K, which locally alters the balance of electrostatic interactions, affects GroES binding and polypep-

tide release, without affecting the ATP activity [17].

In accord with the calculations of Δ we find that, with the exception of Asp87, there are no dramatic changes in the number of contacts in the equatorial domain in the $T \rightarrow R \rightarrow R''$ transitions. The coordination number of Asp87 increases from 1 to 5 as the $R \rightarrow R''$ transition occurs. This residue, which is strongly conserved (see below) coordinates with Mg^{2+}.

2.4. Subunit–subunit interactions

The integrity of the GroEL oligomer is achieved by attractive subunit–subunit interactions. We have identified the residues that provide stability to GroEL multimer in the T and R″ states. The residues that are in contact between two adjacent A domains in the T and R″ states are completely different (Fig. 3). After large movement of the A domains, the interactions that emerge between the A domains in the R″ state involve residues Glu304, Ile305, Gly306, Gln348 and Gln351 in one subunit, and residues Tyr203, Glu209, Thr210, Ala260, Val263 and Val264 in the adjacent subunit. Thus, the subunit–subunit interactions between A domains in the T state are replaced by an altogether new set of residues in the R″ state. The inter–subunit interactions in the T state are largely between charged residues, whereas weaker interactions are present in the R″ state [27]. The changes in the coordination numbers (Fig. 3) reveal that the intermediate domain forms a new set of interactions in the R″ state that is absent in T or R states. For example, the number of contacts for Val387 and Ala405 increases by five and four, respectively.

A different picture emerges when considering the subunit–subunit interactions involving the E domain. A large subset of interactions are preserved in both the T and R″ states, which is consistent with the relatively small displacement of residues in the $T \rightarrow R''$ transition. Because in the $T \rightarrow R''$ transition, which is driven by ATP and GroES binding, the adjacent subunits move apart in a counter-clockwise manner, the E domains experience a torque. The small displacement of the residues in the E domain due to this torque implies a large moment of inertia associated with this domain. The dense inertial mass of this domain has a functional role in imparting a stretching force to the SP [12].

2.5. GroES

The largest changes in accessible area for GroES (Fig. 1c,d) residues, shown in Fig. 4, were computed using the structures of the free state [28] and the one complexed with GroES [11]. As in the case of the GroEL apical domain, most hydrophobic residues participating in the GroEL–GroES interface exhibit a significant loss of accessible area. These are clustered in the mobile loop (residues 16–33), which is the most flexible element of the GroES structure. Exceptions are Arg14, which appears in the cavity lining, Ile25 (a residue that participates in the interface between GroEL and GroES) faces the exterior of the R″ complex, and Ala32. The neighboring Ala33 has $\Delta < 0$. The residues required for the GroEL–GroES interface, Ile25, Val26 and Leu27, are shown in bold labels. Among them Ile25 and Leu27, are located on the external surface in the R″ state. Residues Leu49 $\Delta > 0$ and Arg47 $\Delta < 0$, located at the rigid β-hairpin loop that blocks the access to the complex cavity have large values of Δ. The functional significance of the exposure of a hydrophobic residue and burial of a positively charged residue upon binding to GroEL is not clear. Because these residues do not directly interact with GroEL we surmise that they must play a role in the release of GroES. Mutations of these GroES residues might shed additional light on their roles. There is considerable movement in the highly conserved Gly 72 that is located near the GroEL–GroES interface.

3. Sequence conservation in the HSP60 family and functional implication

The integration of the structural and sequence analysis allows us to infer the role of the most evolutionary conserved residues in relation to the various functions of chaperonins. We expect strong sequence conservation at positions that are associated with the nominal functions (ATPase activity,

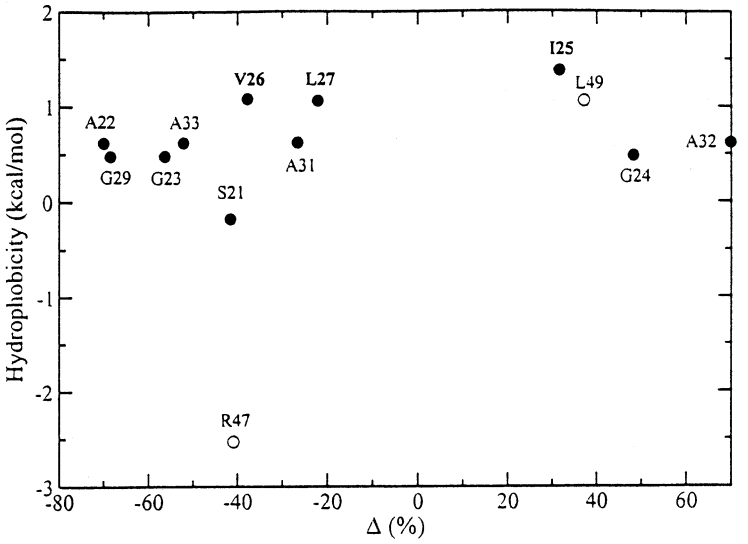

Fig. 4. Relative change in accessible area for residues in bound vs. unbound GroES structure.

SP binding, and interactions with GroES) of GroEL. These functions *do not* impact directly on the most crucial function of the chaperonin machinery, namely, to enable kinetically trapped SPs to reach their native state. The preceding structural analysis has isolated some of the residues that are involved in the alterations in the nature of the SP–GroEL interactions. In this section we use multiple sequence alignments together with sequence entropy to address the extent to which these residues are conserved in the HSP60 family. Brocchieri and Karlin [27] performed multiple sequence alignment using 43 HSP60 sequences together with a site-specific conservation index [CI(i)] to pinpoint many of the crucial residues in the ATP/ADP and Mg^{2+} binding sites and substrate binding sites. The present analysis leads to several new predictions.

Sequence alignment was performed using the PSI-BLAST program [29]. The sequences are aligned against the *E. coli* chaperonins. In all, there are about 450 sequences for GroEL and about 250 for GroES. To assess the extent of conservation we use the sequence entropy (SE)

$$S(i) = -\sum_{j=1}^{m} p_j(i)\ln p_j(i) \qquad (5)$$

where $p_j(i)$ is the frequency of occurrence of an amino acid of type j at position i and $m=20$ is the total number of amino acid types. Gaps in the sequence alignment are not included in the determination of $p_j(i)$. The reference SE, in which all $p_j(i)$ are the same, is $S(i)=\ln 20\simeq 3$, corresponds to a situation in which all amino acids can occur with equal probability. If residue i is perfectly conserved, $S(i)=0$. The sequence entropy is approximately related to $\mathrm{CI}(i)\approx 1-S(i)$. For the purposes of classifying the degree of conservation, we assume that strongly conserved residues have $0<S(i)\leqslant 0.25$ and weakly conserved residues have $0.25<S(i)\leqslant 1.0$. In this section the numbers in parentheses following the indicated amino acids represent either the sequence entropy or the chemical sequence entropy (see below).

Mutational analysis of Fenton et al. [17] has already identified many (not all) residues that are important in the nominal functions (ATPase activity, polypeptide binding and its release) of GroEL. Because SP folding is intimately related to the sequence of concerted transitions $\mathrm{T}\rightarrow\mathrm{R}\rightarrow\mathrm{R}''$ we are interested in the extent to which the residues that provide differing microenvironment to the SP as GroEL undergoes $\mathrm{T}\rightarrow\mathrm{R}\rightarrow\mathrm{R}''$ transitions are conserved. Although structures alone can be used

to identify the residues that provide a hydrophilic environment to the SP sequence analysis can be used to get an evolutionary perspective on the conservation of such residues in the HSP60 family. Through a combined analysis testable predictions can be made.

3.1. Apical domain

Fenton et al. [17] suggested that the putative SP binding sites are Tyr199 (0.22), Tyr203 (0.23), Phe204 (0.43), and residues located in helices H and I, namely, Leu234 (0.78), Leu237 (0.09), Leu259 (0.16), Val263 (0.49) and Val264 (0.1). Of the eight residues only five are strongly conserved. Thus, sequence entropy for the A domain shows (Fig. 5a,b) that relatively few residues involved in the binding of the SPs and GroES are strongly conserved. Given the importance of the role of helices H and I in the capture of SP, it is somewhat surprising that these residues do not show stronger conservation. A better assessment of conservation is obtained using the chemical sequence entropy (CSE), $S_T(i)$.

$$S_T(i) = -\sum_{j=1}^{4} p_j^T(i) \ln p_j^T(i) \qquad (6)$$

where the summation is over the type of residues. The four types are: hydrophobic (Cys, Phe, Ile, Leu, Trp, Val, Met, Tyr and Ala), polar (Gly, Pro, Asn, Thr, Ser, Gln and His), positively charged (Arg and Lys), and negatively charged (Asp and Glu). The values of CSE for several of the SP binding sites of Fenton et al. are small (Fig. 5c,d). Except for Tyr203 (0.15) the CSE for the other seven residues are close to zero. This implies that in the SP binding sites only substitutions by other hydrophobic residues are tolerated. The CSE calculations also show that, two previously unidentified residues Ile205 ($S_T = 0.15$) and Asn206 ($S_T = 0.07$), which are in the loop connecting β6 and β7, could also influence peptide binding.

Buckle et al. [18] suggested, based on a 1.7 Å crystal structure of the apical domain with an N-terminal tag, that Ile230, Glu238, Ala241, Glu257, Ala260, Thr261, Asn265, Arg268, Thr270 and Val271 are also potential SP-binding sites. Previous calculations [27] and sequence entropy computations (Fig. 5a) do not show strong conservation at these positions. However, using the CSE values we infer that Ile230 (0.12), Glu238 (0.23), Glu257 (0.06), Asn265 (0.02), Thr261 (0.23), Arg268 (0.18) and Val271 (0.04) are strongly conserved. This lends support to the findings of Buckle et al. [18]. We also predict Leu262 ($S_T = 0$), whose chemical identity is perfectly preserved, should be considered a putative SP binding site. Chen and Siegler [19] have suggested that the segment 301–311 may also indirectly influence peptide binding. The chemical entropies of Glu304 and Leu309 are 0.06 and 0.07, respectively. This highlights the possible importance of these residues in influencing SP binding [19].

The pair of Gly at 192 and 375 is absolutely conserved [$S(i) \equiv 0$] (Fig. 5a,b). The *en bloc* large amplitude rigid body movement of the A domain occurs around this pivot point. The present study reveals other strongly conserved residues, namely Leu221 (strand 8), Lys226 (loop between strand 8 and helix H), and residues Pro246–Asp253 (strand 9). We find that $S_T(i) \approx 0$ for these residues which shows that in the HSP60 family their chemical character is evolutionarily conserved. Charged residues Lys226, Glu252 and Asp253, which point into the GroEL cavity upon nucleotide and GroES binding, also have relatively small $S(i)$ values. The importance of these residues has not been emphasized in previous studies that focused almost exclusively on the nominal functions of GroEL. *If the annealing action of GroEL is linked to the allostery-driven alterations in the wall character of the cavity* [12]*, then we predict that mutations of the charged residues* (*for example Lys226, Glu252 and Asp253*) *can greatly compromise the efficiency of assisted folding*.

3.2. Intermediate domain

Upon ATP binding the I domain exercises a 25° movement towards the E domain around the hinge Pro137 and the perfectly conserved Gly410. Among the four residues comprising the two hinges in GroEL only Pro137 is not absolutely conserved [$S(137) \approx 1.5$]. The most common

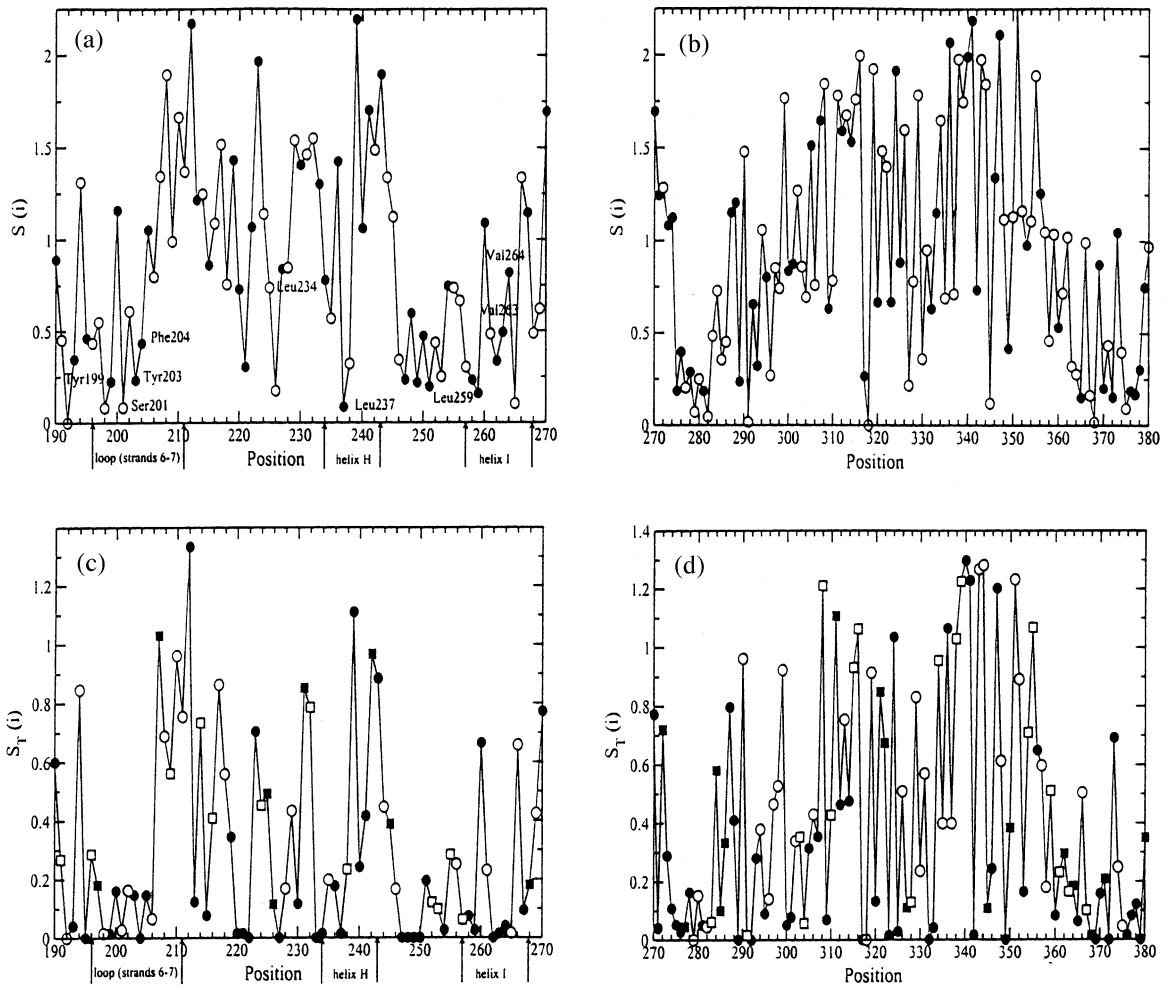

Fig. 5. (a,b) The sequence entropy [Eq. (5)] of the GroEL apical domain as a function of the sequence position. (a,b) Conservation of amino acid types. Hydrophobic residues are shown with filled symbols and hydrophilic residues with empty symbols. The nine labeled residues have been identified as required for binding non-native polypeptides [17]. The positions of the corresponding secondary structure elements (long loop between strand 6 and 7, helices H and I) are indicated with arrows. (c,d) Sequence entropy calculated using Eq. (6): hydrophobic (filled circles), polar (empty circles), positively charged (filled squares), and negatively charged (empty squares). Continuous lines serve as a guide to the eye.

substitution at this site, among the 481 consensus HSP60 sequences, is Lys. Neither the chemical character, nor size seems to be important at this position. Strongly conserved residues, with large values of Δ, are Val169, Gly173 [$S(173) \approx S_T(173) \approx 0$], Lys393 [$S_T(393) = 0.07$], and Asp398 [$S(398) \approx S_T(398) = 0.03$]. Besides these residues, others (Ser151, Ile150, Ile175, Glu386, Glu388 and Glu391) are also highly conserved with an average value of $S_T = 0.10$. Some of these and Asp398 make contact with either ADP in the asymmetric structure (1*aon*) or ATP in the symmetric structure (1*der*) [30].

Kass and Horovitz [31] have recently argued that correlated mutations, which couple non-contacting pairs, triplets of residues, in GroEL and GroES might serve as signaling networks for allosteric communications. They have suggested,

using sequence analysis, that only 31 pairs out of a possible 87 751 pairs are coupled. Out of the two residues in the I domain that are correlated with other residues Val158 is linked with Arg58 in the E domain. Interestingly, the chemical entropy values for Arg58 and Val158 are 0.07 and 0.03, respectively, implying very strong conservation at these positions. Other methods of searching for sequence correlation do not reveal the importance of these coupled sites. Our analysis further validates the suggestion by Kass and Horovitz.

3.3. Equatorial domain

The strongly conserved residues are Asp52, Gly86, Asp87, all of which are located on the stem loop or on helix C. Asp87, which is essential for ATP hydrolysis and polypeptide release [17], has $S(i)=0.05$. Relatively conserved residues are Thr50, Val54 and Ser509. Positions within or around the stem loop are strongly conserved, while little conservation occurs for the helix C. Residues 86–91, which are located at or near the nucleotide-binding site, are almost perfectly conserved. The analysis of Brocchieri and Karlin, done by a very different method, also points to these residues as having a value of $CI \approx 1$.

There are a few coupled sites in the E domain that are strongly conserved according to the CSE values. For example, we find that Cys138 and Cys519 are highly conserved. This is particularly interesting in light of a study by Horovitz et al. [24] who found that addition of ADP led to the dissociation of C519S but not the double mutant C138S/C519S.

As an example of network of interactions between residues in the I and E domains Kass and Horovitz [31] single out Asp58, Ala81 and Asp328. Such network of residues, which are not in contact in the available structures, may mediate allosteric transitions [31]. Molecular dynamics simulations [32] had also suggested that Arg58, Asp83 and Lys327 may interact with each other in the R state. The CSE for all the residues except Ala81 is less than 0.15 which implies very strong correlation. We also find that Asp83 and Lys327 lose contacts ($\Delta Z<0$) in the T→R transition which further lends support to the importance of this network of residues.

3.4. GroES

The formation of a well-defined interface with helices H and I of GroEL is the crucial role of GroES in the chaperonin cycle. The interface is formed by interaction of the mobile loop residues (Glu16–Ala33) with the hydrophobic residues of helices H and I. Although it is not indicated in the crystal structure the three charged residues (Glu16, Glu18 and Lys20) in the mobile loop could form salt bridges with the oppositely charged residues in H and I. From these considerations we would expect strong sequence conservation in at least a few of the mobile loop residues. Sequence entropy using 250 aligned GroES sequences shows (Fig. 6a) little conservation even for residues that undergo large changes in the accessible surface area (Fig. 4) to form the interface with GroEL. With the exception of Gly24 [$S(24)=0.19$] none of the residues in the mobile loop is conserved.

Sequence entropy alone fails to identify crystallographically relevant GroES residues (Ile24, Val26 and Leu27) that interact with the hydrophobic residues from helices H and I. On the other hand, CSE shows (Fig. 6b) that the nature of several mobile loop residues are nearly perfectly conserved ($S_T(i)\approx 0$). The values of $S_T(i)$ for Ile25, Val26 and Leu27 are 0, 0.08 and 0.07, respectively. This finding, which was not realized in previous sequence analysis, shows that at these positions only substitutions by other hydrophobic residues are tolerated. We also find that Thr28 [$S_T(28)=0.23$], which can make hydrogen bonds with residues in GroEL, is strongly conserved. The hydrophobic patch Val65 ($S_T=0$), Ile66 ($S_T=0.15$) and Phe67 ($S_T=0.03$) are highly conserved. There are very few studies examining the effect of GroES mutations on the function of GroEL. The present analysis suggest possible candidates for such studies.

4. Conclusions and predictions

The analysis of the crystal structures of the unliganded T, R and the R″ states reveals that

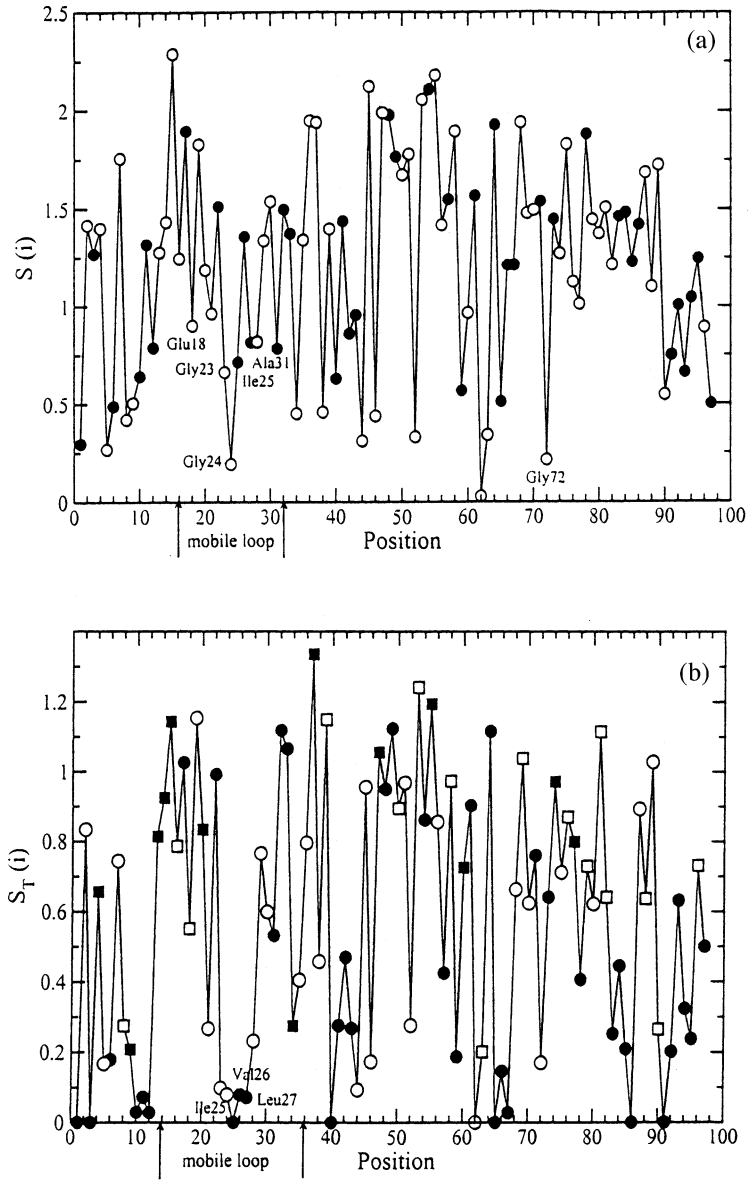

Fig. 6. GroES sequence entropy. The position of the mobile loop is indicated between arrows. Labeled residues are among those exhibiting significant changes of accessible area. Symbols used are the same as in Fig. 1.

many residues have large loss or gain in the accessible surface area and the number of contacts. Such large displacements appear to facilitate the important changes in the nature of the interaction between the SP and the cavity. The key function of the chaperonin nanomachine is to rescue SPs that are likely to aggregate. Prior to the occurrence of a few crucial events in the chaperonin cycle (ATP hydrolysis, binding of ATP to the trans ring, etc.) the annealing of the encapsulated SP commences. According to IAM, annealing becomes possible because of the change in the character of

the cavity as the $T \rightarrow R \rightarrow R''$ transitions take place. Residues (e.g. Lys226, Glu251 and Asp252) that provide a hydrophilic environment to the SP in the R″ state *are not* perfectly conserved. Nevertheless, IAM suggests that preserving the chemical identity, which are solvent-exposed in the T state, is critical. In accord with this observation the entropy $[S_T(i)]$, that takes into account only the chemical identity, calculations show that these residues are *strongly conserved*. This study leads to an experimentally testable prediction that mutating these seemingly non-conserved residues can compromise the efficiency of the chaperonin machinery. In general, it is seen that with the exception of a relatively small number of sites only the chemical identities of residues are evolutionarily conserved in the HSP60 family. Thus, a great degree of plasticity, at the sequence level, is observed in the evolution of this class of proteins.

Acknowledgments

We are pleased to dedicate this paper to Dr John Edsall. We thank Helen R. Saibil for bringing Ref. [10] to our attention. We are grateful to Amnon Horovitz for providing a copy of Ref. [31] prior to publication. We would like to thank Ruxandra Dima for useful discussions. This work was supported in part by a grant from NSF (to D.T.) through grant number NSF CHE02-09340.

References

[1] W.A. Fenton, A.L. Horwich, GroEL-mediated protein folding, Prot. Sci. 6 (2000) 743–760.
[2] K.A. Dill, H.S. Chan, From Levinthal to pathways to funnels, Nat. Struct. Biol. 4 (1997) 10–19.
[3] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Theory of protein folding: the energy landscape perspective, Annu. Rev. Phys. Chem. 48 (1997) 545–600.
[4] D. Thirumalai, S.A. Woodson, Kinetics of folding of proteins and RNA, Acc. Chem. Res. 29 (1996) 433–439.
[5] O. Yifrach, A. Horovitz, Coupling between protein folding and allostery in the GroE chaperonin system, Proc. Natl. Acad. Sci. USA 97 (2000) 1521–1524.
[6] M.J. Todd, P.V. Viitanen, G.H. Lorimer, Dynamics of the chaperonin ATPase cycle: implication for facilitated protein folding, Science 256 (1994) 659–666.
[7] H.S. Rye, A.M. Roseman, S. Chen, K. Funak, W.A. Fenton, H.R. Saibil, A.L. Horwich, GroEL–GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings, Cell 97 (1999) 325–338.
[8] K. Braig, Z. Otwinowski, R. Hedge, D.C. Boisvert, A. Joachimiak, A.L. Horwich, P.B. Sigler, The crystal-structure of the bacterial chaperonin at 2.8 Å, Nature 371 (1994) 578–586.
[9] J.F. Hunt, A.J. Weaver, S.J. Landry, L. Gierasch, J. Diesenhofer, The crystal structure of the GroES co-chaperonin at 2.8 Å resolution, Nature 379 (1996) 37–45.
[10] N.A. Ranson, G.W. Farr, A.M. Roseman, B. Gowen, A.L. Horwich, H.R. Saibil, ATP-bound states of GroEL captured by Cryo-Electron Microscopy, Cell 107 (2001) 869–879.
[11] Z. Xu, A.L. Horwich, P.B. Sigler, The crystal structure of the asymmetric GroEL–GroES–$(ADP)_7$ chaperonin complex, Nature 388 (1997) 741–750.
[12] D. Thirumalai, G.H. Lorimer, Chaperonin-mediated protein folding, Annu. Rev. Biophys. Biomol. Struct. 30 (2001) 245–269.
[13] A. Horovitz, Y. Fridmann, G. Kafri, O. Yifrach, Allostery in chaperonins, J. Struct. Biol. 135 (2001) 104–114.
[14] M.R. Betancourt, D. Thirumalai, Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity, J. Mol. Biol. 287 (1999) 627–644.
[15] M.J. Todd, G.H. Lorimer, D. Thirumalai, Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism, Proc. Natl. Acad. Sci. USA 93 (1996) 4030–4035.
[16] A. Brinker, G. Pfeifer, M.J. Kerner, D.J. Naylor, F.U. Hartl, M. Hayer-Hartl, Dual function of protein confinement in chaperonin-assisted protein folding, Cell 107 (2001) 223–233.
[17] W.A. Fenton, Y. Kashi, K. Furtak, A.L. Horwich, Residues in chaperonin GroEL required for polypeptide binding and release, Nature 371 (1994) 614–619.
[18] L.L. Chen, P.B. Sigler, The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity, Cell 99 (1999) 757–768.
[19] A.M. Buckle, R. Zahn, A.R. Fersht, A structural model for GroEL-polypeptide recognition, Proc. Natl. Acad. Sci. USA 94 (1997) 3571–3575.
[20] R. Zahn, S. Perrett, G. Stenberg, A.R. Fersht, Catalysis of amide proton exchange by the molecular chaperones GroEL and SecB, Science 271 (1996) 642–645.
[21] M. Shtilerman, G.H. Lorimer, S.W. Englander, Chaperonin function: folding by forced unfolding, Science 284 (1999) 822–825.
[22] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400.
[23] S. Miller, J. Janin, A.M. Lesk, C. Chothia, Interior and surface of monomeric proteins, J. Mol. Biol. 196 (1987) 641–656.
[24] A. Horovitz, E.S. Bochkareva, O. Yifrach, A.S. Girshovich, Prediction of an inter-residue in the chaperonin

GroEL from multiple sequence alignment is confirmed by double-mutant cycle analysis, J. Mol. Biol. 238 (1994) 133–138.
[25] Structural models of GroEL do not include 23 carboxy-terminal residues, which are not visible in the available crystal structures [8, 11, 30]
[26] J.S. Weissman, H.S. Rye, W.A. Fenton, J.M. Beechem, A.L. Horwich, Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES, Cell 83 (1995) 577–587.
[27] L. Brocchieri, S. Karlin, Conservation among HSP60 sequences in relation to structure, function, and evolution, Prot. Sci. 9 (2000) 476–486.
[28] J.F. Hunt, private communication.
[29] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[30] D.C. Boisvert, J. Wang, Z. Otwinowski, A.L. Horwich, P.B. Sigler, The 2.4 Å crystal structure of the bacterial chaperonin GroEL complexed with ATP-γS, Nat. Struct. Biol. 3 (1996) 170–177.
[31] I. Kass, A. Horovitz, Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations', Prot.: Struct. Funct. Genet. 48 (2002) 611–617.
[32] J.P. Ma, P.B. Siegler, Z. Xu, M. Karplus, A dynamic model for the allosteric mechanism of GroEL, J. Mol. Biol. 302 (2000) 303–313.
[33] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Cryst. 24 (1991) 946–950.
[34] Povscript was written by Dan Peisach of Brandeis University.
[35] R. Sayle, E.J. Milner-White, RasMol: biomolecular graphics for all, Trends Biochem. Sci. 20 (1995) 374.
[36] H.J. Bernstein, Recent changes to RasMol, recombining the variants, Trends Biochem. Sci. 25 (2000) 453–455.
[37] POV-ray is Copyright 1991, 1997 by the POV-ray Team®.
[38] D. Eisenberg, E. Schwartz, M. Komaromy, R. Wall, Analysis of membrane and surface protein sequences with the hydrophobic moment plot, J. Mol. Biol. 179 (1984) 125–142.