doi:10.1016/j.jmb.2008.12.032 J. Mol. Biol. (2009) 387, 390–406

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Allostery Wiring Diagrams in the Transitions that Drive the GroEL Reaction Cycle

Riina Tehver[1], Jie Chen[1] and D. Thirumalai[1,2]*

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

[2]*Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*

Received 28 September 2008; received in revised form 5 December 2008; accepted 8 December 2008
Available online
24 December 2008

Determining the network of residues that transmit allosteric signals is crucial to understanding the function of biological nanomachines. During the course of a reaction cycle, biological machines in general, and *Escherichia coli* chaperonin GroEL in particular, undergo large-scale conformational changes in response to ligand binding. Normal mode analyses, based on structure-based coarse-grained models where each residue is represented by an α carbon atom, have been widely used to describe the motions encoded in the structures of proteins. Here, we propose a new C^{α}–side chain elastic network model of proteins that includes information about the physical identity of each residue and accurately accounts for the side-chain topology and packing within the structure. Using the C^{α}–side chain elastic network model and the structural perturbation method, which probes the response of a local perturbation at a given site at all other sites in the structure, we determine the network of key residues (allostery wiring diagram) responsible for the T→R and R″→T transitions in GroEL. A number of residues, both within a subunit and at the interface of two adjacent subunits, are found to be at the origin of the positive cooperativity in the ATP-driven T→R transition. Of particular note are residues G244, R58, D83, E209, and K327. Of these, R38, D83, and K327 are highly conserved. G244 is located in the apical domain at the interface between two subunits; E209 and K327 are located in the apical domain, toward the center of a subunit; R58 and D83 are equatorial domain residues. The allostery wiring diagram shows that the network of residues are interspersed throughout the structure. Residues D83, V174, E191, and D359 play a critical role in the R″→T transition, which implies that mutations of these residues would compromise the ATPase activity. D83 and E191 are also highly conserved; D359 is moderately conserved. The negative cooperativity between the rings in the R″→T transition is orchestrated through several interface residues within a single ring, including N10, E434, D435, and E451. Signal from the trans ring that is transmitted across the interface between the equatorial domains is responsible for the R″→T transition. The cochaperonin GroES plays a passive role in the R″→T transition. Remarkably, the binding affinity of GroES for GroEL is allosterically linked to GroEL residues 350–365 that span helices K and L. The movements of helices K and L alter the polarity of the cavity throughout the GroEL functional cycle and undergo large-scale motions that are anticorrelated with the other apical domain residues. The allostery wiring diagrams for the T→R and R″→T transitions of GroEL provide a microscopic foundation for the cooperativity (anticooperativity) within (between) the ring (rings). Using statistical coupling analysis, we extract evolutionarily linked clusters of residues in GroEL and GroES. We find that several substrate protein binding residues as well as sites related to ATPase activity belong to a single functional

**Corresponding author.* Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA. E-mail address: thirum@glue.umd.edu.

Abbreviations used: ENM, elastic network model; SPM, structural perturbation method; SCA, statistical coupling analysis.

network in GroEL. For GroES, the mobile loop residues and GroES/GroES interface residues are linked.

Keywords: allostery wiring diagram; chaperonin GroEL; structural perturbation method; functional residues; evolutionary imprints

Edited by D. Case

Introduction

Repeated cycling between distinct allosteric states is required for the functions of numerous biological nanomachines. During a typical reaction cycle, biological machines (for example, *Escherichia coli* chaperonin GroEL,[1] ATP synthase,[2] and molecular motors[3,4]) visit a number of states triggered by ligand binding. The transitions between the various structures are often powered by hydrolysis of ATP, binding, or release of ligands. Determining the structural elements and specific residues that are responsible for transmitting the allosteric signal is the key to understanding how molecular machines work. More generally, it is suspected that a network of residues that are interspersed throughout the structure conspire to bring about the large-scale conformational changes between distinct allosteric states.[5] The network of residues, which may be thought of as transmitting signals along the most probable dynamical route for facilitating the functionally relevant structural transitions, is referred to as the allosteric wiring diagram. Structure-based methods are generally needed to determine the allostery wiring diagram.[6] We also use a sequence-analysis-based method to extract functionally related networks.[7,8]

The large conformational changes in the allosteric transitions make it currently impractical to use standard molecular dynamics simulations to reliably obtain the dynamical pathways connecting two states in a reaction cycle. A practical method to study the dominant conformational changes between two allosteric states is to use a normal mode analysis.[9–17] The theoretical rationale for the method is based on the assumption that a protein is normally at the free-energy minimum in its native state with small deviations from the minimum. Based on this assumption, one can build a harmonic elastic network using the structures of the states that are visited during the reaction cycle of a nanomachine or an enzyme. Significant conformational changes are often associated with only a few low-frequency normal modes. This is in accord with the expectation that protein structures have evolved to accommodate biologically necessary transitions as efficiently as possible. In accord with this expectation, examination of the dynamical trajectories connecting the various states of GroEL shows considerable plasticity in the structural transitions.[18]

Normal modes are most easily calculated using structure-based coarse-grained models. In the usual elastic network model (ENM), a given protein structure is represented using the coordinates of the C^{α} atoms and the associated contact maps.[19] A contact between two C^{α} atoms i and j exists if the distance between them is less than a cutoff value. With this definition, a two-dimensional contact map can be constructed. However, detailed analyses of protein structures that have been used to obtain statistical potentials show that besides direct interaction between side chains, there are substantial numbers of interactions between side chains and the backbone atoms.[20] It is likely that C^{α}–side chain interactions add substantially to the dense packing of proteins. In order to take such interactions into account, we propose a new and more detailed, but still computationally tractable, ENM that includes information about the physical identity of each residue in addition to the system topology. We describe each amino acid using two interaction centers, one representing the C^{α} atom and the other the center of mass of the side chain. Such C^{α}–side chain models have been profitably used in simulations of protein folding.[21,22] In the context of the normal mode analysis, the inclusion of the side chains gives a more accurate description of the topology of the protein and the packing within the structure. We also incorporate information about the chemical and physical characteristics of the amino acids by including residue-specific spring constants.

We use the Betancourt–Thirumalai potential[23] to scale the strength of interactions in the C^{α}–side chain ENM. The van der Waals radii of the side chains are used as the length scale for contact distances. (For more details, see the Methods section).

We demonstrate the efficacy of the C^{α}–side chain ENM by analyzing the allosteric transitions that GroEL undergoes during its reaction cycle. The associated allostery wiring diagrams of GroEL $T \rightarrow R$ and $R'' \rightarrow T$ transitions are constructed based on the C^{α}–side chain ENM using the structural perturbation method (SPM) approach. A network of residues interspersed throughout the subunit and at the interface between the adjacent subunits accounts for the positive cooperativity in the ATP-driven $T \rightarrow R$ transition. Similarly, the origin of negative cooperativity is linked to residues that lie at the interface between the equatorial domains across the two rings as well as those that are at the interface between the equatorial domains within a ring.

We also study the potential functional relationships between different residues within GroEL sequence using evolutionary data. Statistical coupling analysis (SCA)[7,8] yields a functional wiring diagram that incorporates several ATPase rate related and substrate protein or GroES binding residues for GroEL. For GroES, we find a linkage between sev-

eral GroEL binding residues and GroES/GroES interface residues.

Results and Discussion

Allosteric states in GroEL

The chaperonin GroEL,[24–27] an ATP-consuming machine that rescues substrate proteins that are otherwise destined to aggregate, has been thoroughly studied using both C^{α}[28,29] and all-atom[30] models. GroEL is a molecular chaperonin whose spectacular (driven by the binding of ATP, cochaperonin GroES, and the substrate protein) allosteric transitions facilitate the folding of cytosolic proteins.[1,25] The intact GroEL particle is made up of two heptameric rings stacked back to back. Each subunit of the GroEL molecule is a 548-residue chain that is usually divided into the apical (residues 191–376), equatorial (1–133, 409–548), and intermediate (134–190, 377–408) domains. Figures 1a and 2a illustrate the three domains in red, green, and blue, respectively. Assembly of the machine into two heptameric rings creates two large inner cavities where the substrate proteins are sequestered transiently (for a few seconds) during the reaction cycle.[31] Each of the two GroEL rings undergoes the same, but out of phase, complex allosteric cycle consisting of a series of conformational changes between T, R, R′, and R″ states (Fig. 3). High-resolution crystal structures have been determined for GroEL in the T and R″ states.[1,32] An approximate structure for the R state, based on cryoelectron microscopy measurements, is also known.[33] In the T state, a GroEL ring can bind various substrate proteins. ATP binding triggers a transition to the R state. Subsequent cochaperonin GroES binding and ATP hydrolysis drive GroEL, via the R′ state, to the R″ state. In response to a signal from the opposite trans ring, GroES, ADP, and the substrate protein are released and the GroEL ring returns from the R″ to the T state. All the conformational changes are associated with large structural rearrangements or domain movements that also enable structural changes in the substrate protein and are thus essential for the proper function of the chaperonin. In particular, the change in the microenvironment that the substrate protein feels as the reaction cycle proceeds is intimately related to the annealing function of GroEL.[25,34] Thus, allostery and rescue of the substrate protein are intimately coupled.[25] We use the C^{α}–side chain ENM-based models in conjunction with SPM to map the allostery wiring diagrams for the various allosteric transitions.

Single-subunit T→R transition is dominated by the lowest-energy mode

The GroEL functional cycle begins with the transition from the T to the R state that is triggered by ATP binding. To study the T→R transition and determine the key residues that support it, we performed the normal mode analysis of the transition using the C^{α}–side chain ENM (see Methods) representations of the Protein Data Bank (PDB) structures 1AON (chain H in the T state) and 2C7E (chain A in the R state). In Fig. 4a, we plot the overlap [Eq. (2) in Methods] of each of the lowest 50 normal-mode eigenvectors associated with the transition between the T and R states. The lowest-energy (nonzero) eigenmode, mode 7, has an overlap of 0.83, and thus describes the transition between the structures remarkably well. The second highest overlap mode is mode 9 with a significantly lower overlap of 0.26. The overlaps of all other modes are relatively insignificant (Fig. 4a). Thus, the T→R transition is well described by a single low-energy mode.

Mode 7 corresponds to a counterclockwise twist and a slight upward tilt of the apical domain with respect to the equatorial domain with the intermediate domain residues acting as hinges. Figure 5a shows the amplitudes of vibration [see Eq. (3) in Methods] of all the residues in mode 7. The hinge residues in the intermediate domain show significantly smaller fluctuations than do the residues in other domains. The highest-amplitude region corresponds to apical domain residues 340–370 or helices K (339–353) and L (361–371) and the loop connecting them.[28,35] The dramatic movement of helices K and L during the R→R″ transition, which was first identified in a previous study,[13] results in the change in the polarity of residues that line the inner cavity. The change in the microenvironment of the cavity as the reaction cycle proceeds is the fundamental reason that GroEL is an annealing machine.

We see that in the dominant mode, the helices undergo significant movement even prior to the R→R″ transition. The correlations between different residues in mode 7 are shown in Fig. 6a. The orientational cross-correlations between the fluctuations of residues are calculated as $\mathbf{e}_i^M \cdot \mathbf{e}_i^M$ where $\mathbf{e}_i^M$ is the unit vector in the direction of the displacement of the ith residue in a given mode M. While the correlations within each domain are generally high, there are a few exceptions. In particular, the residues 334–371, which include helices K and L, move together in a correlated fashion but are anticorrelated ($\mathbf{e}_i^M \cdot \mathbf{e}_i^M \approx -1$) with most of the rest of the apical domain (see Fig. 6a).

Allostery wiring diagram for the single-subunit T→R transition

The network that transmits allostery in proteins or the allostery wiring diagram can be predicted using the SPM (see Methods). To the best of our knowledge, the wiring diagram that is specifically responsible for the GroEL T→R transition has not been calculated before. According to SPM, mutations of the residues that are most critical to an allosteric transition ("hot-spot" residues) lead to the largest change in significant normal modes. The response of GroEL T-state mode 7 to local perturbations (mutations) in Fig. 5b shows that many of the residues that show significant response to local

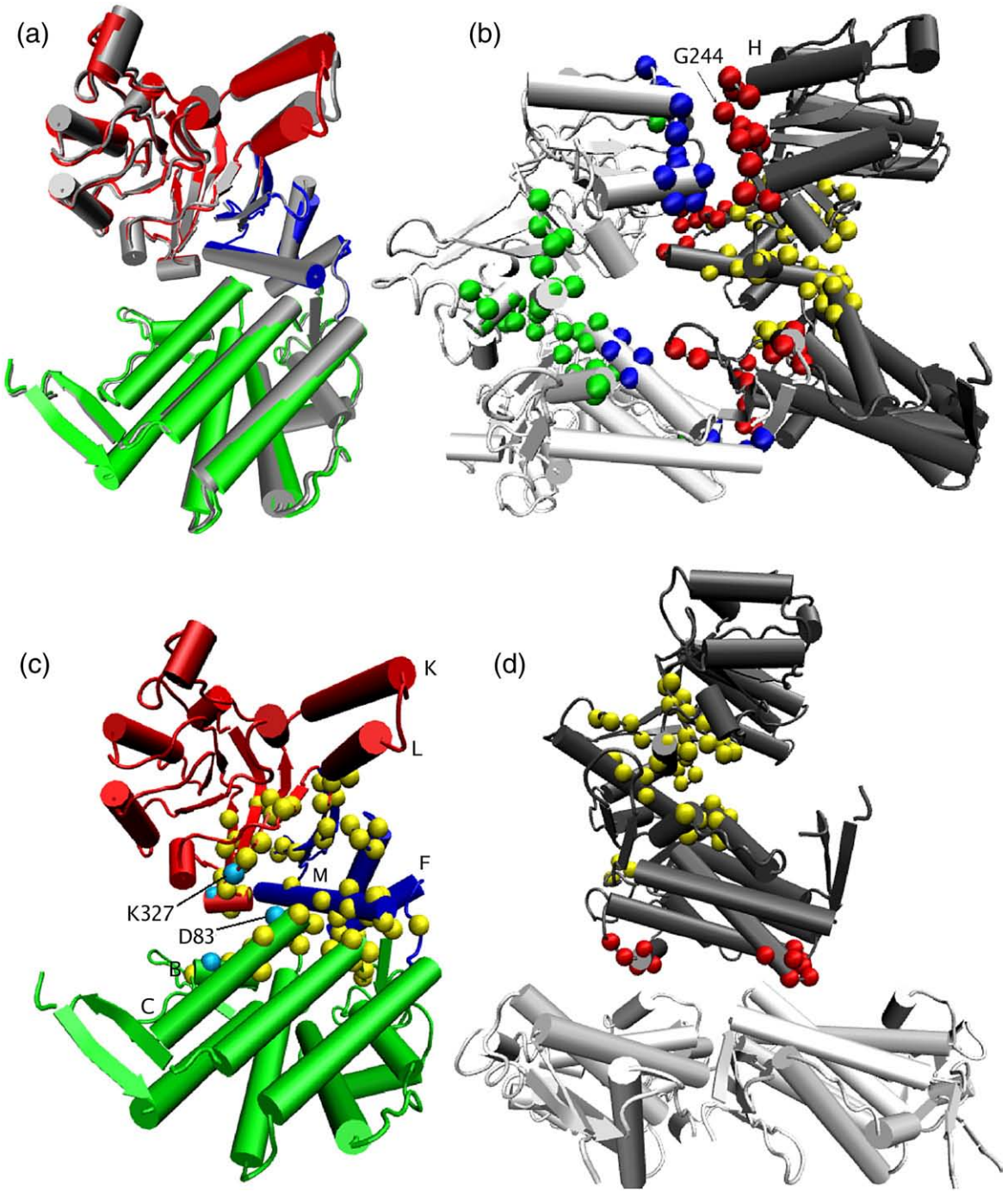

Fig. 1. Illustration of the GroEL T→R transition and the associated allostery wiring diagram. (a) Single-subunit structure in the T state. The equatorial, intermediate, and apical domains are shown in green, blue, and red, respectively. The motions of the structural elements due to the dominant mode are shown in gray. (b) Structure of two adjacent subunits of GroEL (the chains are shown in dark and light gray) in the T state with residues in the allostery wiring diagram highlighted in color. The critical interface residues are in red and blue, and the other hot-spot residues are in yellow and green. The interface residue G244 (see the text) is explicitly labeled. (c) The residues that determine the allostery wiring diagram for the T→R transition are shown in yellow. Residues that respond most to perturbation (see Fig. 5b) are highlighted in light blue. Helices K, L, F, and M (see the text) are also labeled. The domains are colored as in (a). (d) GroEL T-state structure (dark gray) with adjacent trans equatorial domain (light gray). The interface hot-spot residues are shown in red; the rest are shown in yellow.

perturbations are in the intermediate domain. This finding accords well with earlier dynamical simulations, which showed that the initial events in the T→R transition involve movements of helices F (residues 141–151) and M (386–409) (Fig. 1c) toward the equatorial domain.[18] However, the largest δω residues (explicitly labeled in Fig. 5b) are located in the apical and equatorial domains. Using the

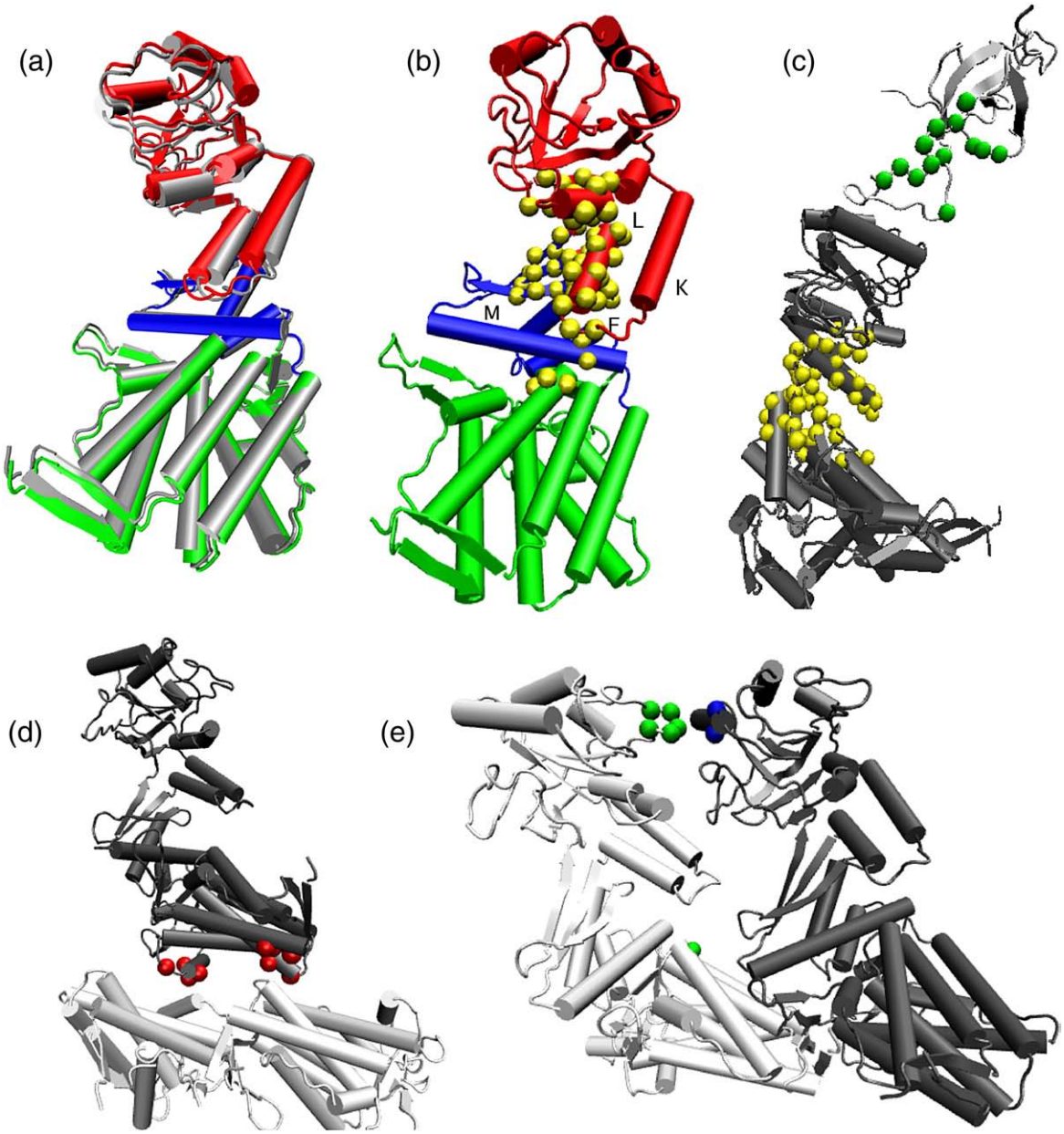

Fig. 2. Illustration of the GroEL R″→T transition and the associated allostery wiring diagram. (a) Single-subunit structure in the R″ state. The equatorial, intermediate, and apical domains are shown in green, blue, and red, respectively. The motions of the structural elements due to the dominant mode are shown in gray. (b) The significant residues that determine the allostery wiring diagram for the transition from the R″→T state are shown in yellow. Helices K, L, F, and M (see the text) are labeled. The domains are colored as in (a). (c) GroEL (dark gray)–GroES (light gray) model. The allostery wiring diagram residues are explicitly shown in yellow (GroEL) and green (GroES). (d) GroEL R″-state structure (dark gray) with adjacent trans right equatorial domain (light gray). The interface hot-spot residues are in red. (e) Structure of two adjacent subunits, in dark and light gray, in the R″ state with the interface residues that belong to the allostery wiring diagram are shown in green and blue.

criterion $\delta\omega > 2\langle\delta\omega\rangle$ [see Eq. (5) in Methods], we extract 67 significant or hot-spot residues. The residues, listed in Table 1, are mapped onto the GroEL structure in Fig. 1c. Most of the hot-spot residues are located in the intermediate domain, highlighting the importance of the hinge residues in facilitating the T→R transition. The importance of a few hot-spot residues that are located in the apical and equatorial domains are discussed below.

Because almost all of the ENM calculations have been done using only C^{α} representations of the structures, it is important to compare the present results with those obtained using the standard approach. To this end, we also performed the normal mode analysis and SPM calculation of the standard C^{α} model of the GroEL T→R transition for comparison[29] (data not shown). The normal mode analysis of the C^{α} model also shows that a single

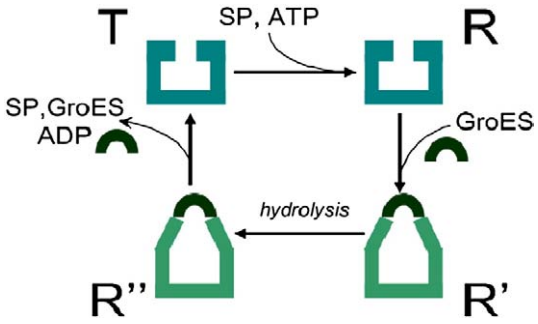

Fig. 3. A sketch of the GroEL reaction hemicycle. Substrate protein and nucleotide binding trigger an allosteric transition T→R. Subsequent GroES binding and ATP hydrolysis result in the transitions R→R′ and R′→R″. The cycle is complete when, due to a signal from the trans ring (not shown), ligands are released and GroEL undergoes an allosteric transition R″→T. We describe the first (T→R) and the last (R″→T) step in the reaction cycle using available structures.

dominant mode, with an overlap of 0.63, describes the structural changes during the T→R transition well. There are, however, significant differences between the SPM results using the C^{α}–side chain ENM and the C^{α}-only model. Compared to the C^{α} model, the C^{α}–side chain ENM predicts an additional set of residues in the allostery wiring diagram: 54–59, 78–84, 207–215, and 326–332. The residues within this cluster are directly linked in the C^{α}–side chain ENM due to the orientation of their side chains. Specifically, residues 54–59 belong to the equatorial domain helix B; residues 78–84 are part of helix C, adjacent to helix B. Residues 207 to 215 are part of the apical domain but they are in close proximity to helix B in the T state. In particular, the distance between the closest side-chain atoms (residue 55 atom OG and residue 209 atom OE) is only 4.9 Å. (The corresponding backbone C^{α} distance is outside of the interaction cutoff and, thus, considering only the backbone–backbone interactions does not create a direct contact between these residues.) Finally, residues 321–332 are part of a β-sheet that is parallel to residues 211–216.

The prediction that the residues (54–59, 78–84, 207–215 and 326–332) are dynamically important in the T→R transition is consistent with earlier findings using totally different methods. In a previous study that probed the dynamics of the allosteric transitions in GroEL using a self-organized polymer model,[18] the rupture of the salt bridge D83–K327 (residues that have the highest δω values according to SPM in Fig. 5b) was shown to be a significant early event in the T to R transition. Our allostery wiring diagram also explains why the introduction of a disulfide bond between residues D83 and K327 arrests the GroEL allosteric cycle.[36] Several of these residues (R58, A81, N328) belong to the allosteric networks identified by Kass and Horovitz based on sequence analysis.[37] In addition, residues 54–59 belong to a conserved word (52-DGVRVAKEI-60), and K207 and E214 are moderately conserved.[38]

The interface residues between subunits in the allostery wiring diagram as foundation for cooperativity in the T→R transition

It is known that the T→R transition is cooperative and concerted in the sense that the transitions in the seven subunits are synchronous. In order to describe the cooperative nature of the GroEL allosteric transitions, it is important to determine the interface residues in the wiring diagram that transmit the allosteric signals. To identify the most significant interface residues, we constructed two subunits of GroEL in the T state (Fig. 1b). The T→R transition of a GroEL model, with two adjacent subunits, is best described by modes 7 (overlap 0.49) and 13 (overlap 0.35) as shown in Fig. 4a. The

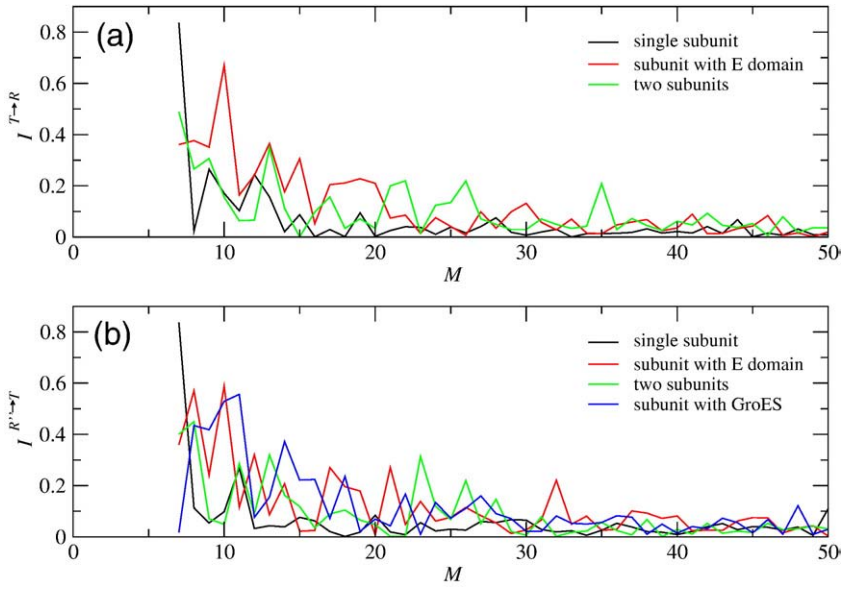

Fig. 4. (a) Overlap, computed using Eq. (2) in Methods, for the lowest 50 normal modes (*M*) associated with the T→R transition. The single-subunit, two-subunit, and a subunit with two trans equatorial domains are plotted in black, green, and red as shown in the legend. (b) Overlap for the R″→T transition. The single-subunit, two-subunit, and a subunit with two trans equatorial domains are plotted in black, green, and red as shown in the legend; the results from the GroEL–GroES complex are shown in blue. See the text for the discussion of the highest overlap modes.

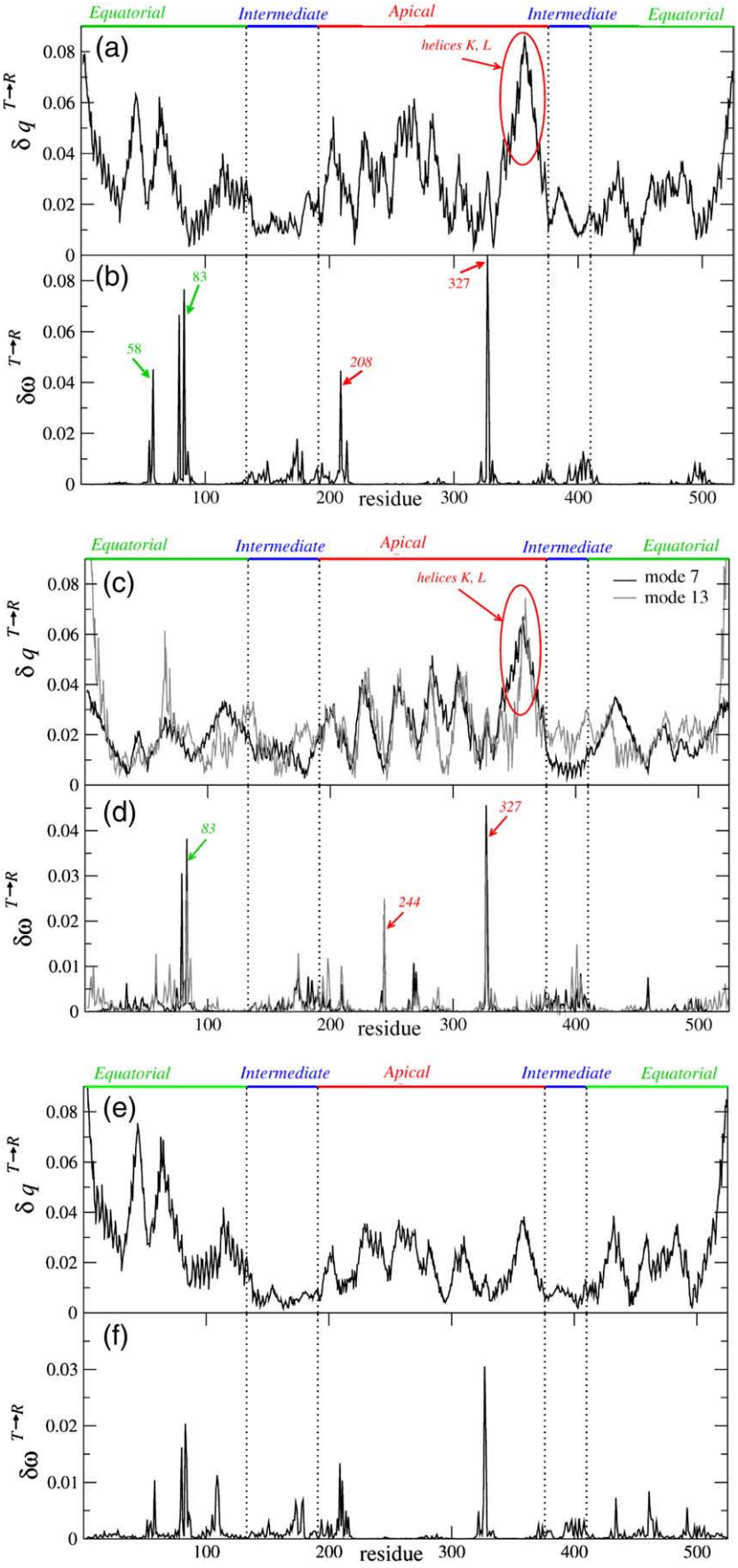

Fig. 5 (*legend on next page*)

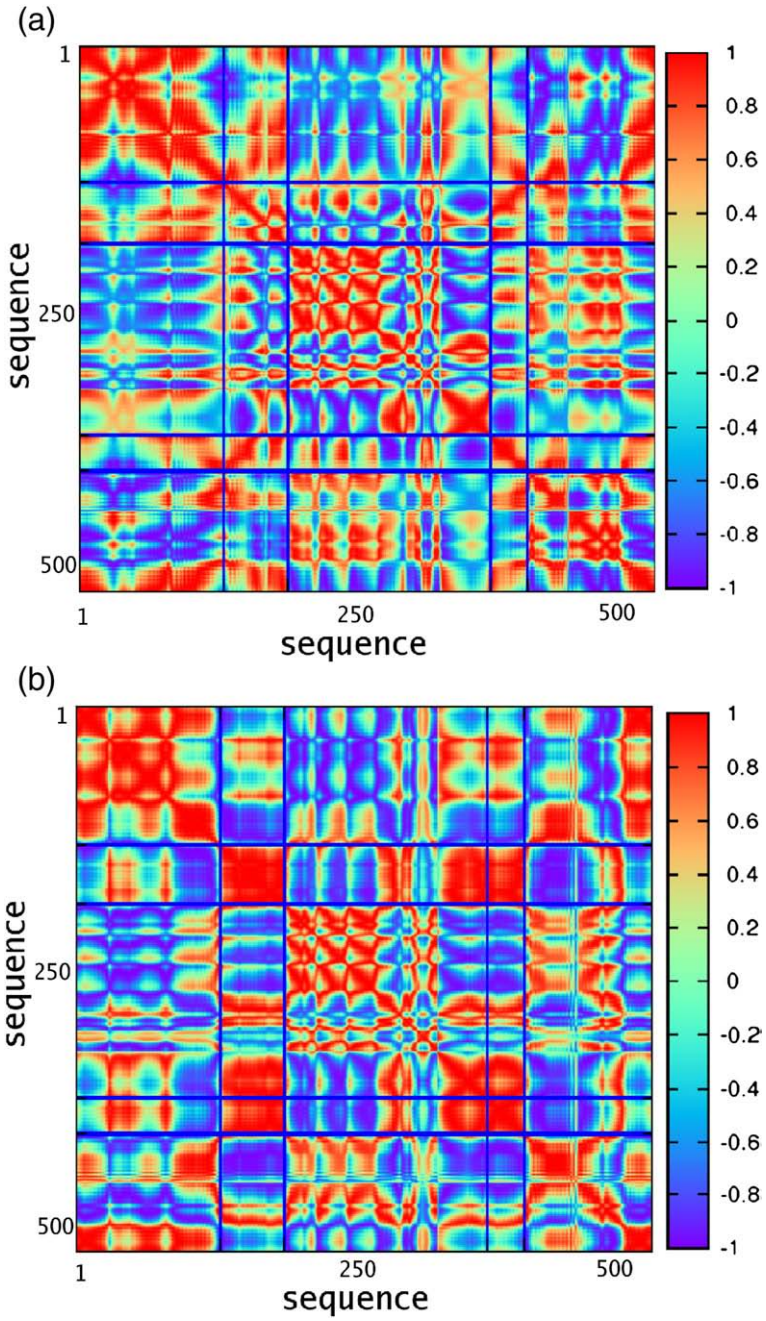

Fig. 6. (a) The correlations between all residue pairs in the C^{α}–side chain ENM of a single subunit of GroEL in the T state in mode 7. Red indicates high (positive) level of correlation between the motions of two residues, blue denotes negative correlation, and green indicates lack of correlation. The color scheme for the numerical values of the correlation function is shown on the right. The domain boundaries are drawn in dark blue. (b) The same as (a) for a single subunit of GroEL in the R″ state.

amplitudes of vibration for the modes are shown in Fig. 5c. The results are similar to those obtained using a single subunit in that there is a noticeable reduction in the amplitudes of motion of the intermediate domain residues. In addition, helices K and L (residues 339–371) show the largest amplitude among the apical domain residues. The SPM result for the modes is shown in Fig. 5d. The one- and two-subunit structures yield rather similar results with many hot-spot residues in the intermediate domain. Residues D83 and K327 have the largest $\delta\omega$ values. A new residue with a large $\delta\omega$ value is residue G244 (Fig. 5d). Based on our calculations, the cooperativity of the T→R transition

Fig. 5. Analysis of the dominant modes of the GroEL T→R transition. (a and b) Results based on C^{α}–side chain ENM for a single subunit in the T state. (a) The amplitudes of motion in the dominant normal mode. The region with the highest amplitude corresponds to helices K and L. (b) Response $\delta\omega$ as a function of residue number *i* for the dominant mode. The residues with the largest $\delta\omega$ values are labeled. (c and d) Results for the GroEL model with two adjacent subunits. (c) The amplitudes of motion in the dominant normal modes. The region with the highest amplitude corresponds to helices K and L. (d) $\delta\omega$ for each residue for the dominant modes. The residues with the largest $\delta\omega$ values are labeled. (e and f) Results for a GroEL subunit with adjacent trans equatorial domain. (e) The amplitudes of motion in the dominant normal mode. (f) $\delta\omega$ for each residue for the dominant mode.

Table 1. Hot-spot residues using SPM for the highest overlap mode for each transition

Transition	Chain	Hot-spot residues
T→R, single subunit	H	54–59, 78–80, 82–87, 137, 149–151, 170–175, 177–179, 189, 190, 194, 195, 207–211, 213–215, 322, 326–329, 331, 332, 374–376, 393, 401–410, 493–495, 497–499
T→R, two adjacent subunits	H	33–35, 46, 47, 53–60, 78–86, 146, 170–189, 194, 195, 208–212, 241–246, 267–271, 326–329, 374, 375, 383, 385, 391–394, 403–405, 407–409, 458–460, 491–493
	I	7–9, 57–59, 75–80, 82, 83, 85–88, 113–116, 206–211, 213–215, 225, 230–232, 255–261, 283–285, 302–304, 307–309, 311, 393–402
T→R, single subunit with *trans* E plate	H	57-59, 78-87, 105-111, 170-175, 177-180, 208–215, 322, 323, 325–329, 393, 394, 397, 398, 433–435, 460–464, 491–493
R″→T, single subunit	A[a]	82–84, 165–177, 187–196, 283–285, 287–289, 330–332, 358–379, 403–405
R″→T, single subunit+GroES	A	83–85, 164–177, 186–197, 287, 288, 330, 357–370, 372–379, 403–405, 407–409
R″→T, two adjacent subunits	A	82–85, 154–157, 163–177, 187–193, 255–257, 283–285, 287–291, 333, 334, 344–349, 351–369, 371–377, 400, 401, 403–405
	G	82–84, 165–177, 188–196, 282–285, 287–289, 291, 303–307, 341–347, 358–360, 363–378
R″→T, single subunit with *trans* E plate	A	9–15, 82–85, 104, 105, 109, 165–177, 187–195, 330–332, 357–360, 364–379,403–405, 434, 435, 460, 463–465

[a] The chain labeling is according to PDB file 1AON.

should be particularly sensitive to mutations of residue G244. A previous study[39] showed that the number of contacts involving G244, located at the end of helix H and at the interface between the two subunits in the apical domain in T the state, changes significantly in the R→R″ transition. The residue is also located next to a set of highly conserved residues (246–253).[38]

By mapping the hot-spot residues (listed in Table 1 for mode 7) onto their structures, we find that 33 of the 85 hot-spot residues of chain H (per the chain labeling in the PDB structure 1AON) and 24 of the 62 hot-spot residues of chain I belong to the inter-subunit interface. We define interface residues as residues that make at least one contact with a residue in the adjacent subunit. The interface hot-spot residues, highlighted in green and blue in Fig. 1b, show that the large number of interface residues in the allostery wiring diagram is the possible foundation for the strong intraring positive cooperativity.

A calculation of the normal modes of a GroEL T-state structure with additional two trans ring equatorial domains (Fig. 1d) leads a single dominant normal mode with a significant overlap with the T to R transition—mode 10 has an overlap of 0.67 (Fig. 4). The mode is shown in Fig. 5e. Perturbations of the residues in this mode (Fig. 5f) reveal that the residues with the highest δω values are D83, E209, and K327 and that 12 hot-spot residues (listed in Table 1) are in the interring interface as shown in Fig. 1d. Based on the allostery wiring diagram and the locations of the hot-spot residues, the interring interface interactions seem to play a less critical role in the T→R transition than the intraring interactions. These observations show that the cooperativity in the R→T transition arises largely from interactions within a single ring.

R″→T transition is dominated by the lowest-energy eigenmode

The normal mode analysis for a single GroEL subunit in the R″→T transition using PDB structures 1AON (chain A in the R″ state) and 1GR5 (chain A in the T state) shows that the global transition is well described by mode 7. The overlap (Methods) of mode 7 with the R″ to T transition is 0.84. The overlaps of the GroEL R″→T transition with the lowest 50 normal modes are shown in Fig. 4b.

Mode 7, which encompasses the collective motions of the residues of all three domains, describes a downward and slightly sideways tilt of most of the apical domain and an opposite direction sideways movement of helices K and L. The mode is illustrated in Fig. 2a. The intermediate domain acts as a hinge supporting the shift of the apical domain with respect to the equatorial domain. (See dotted line in Fig. 7e). As we already noted in the discussion of the T→R transition, the intermediate-domain residues are significantly less mobile than equatorial- or apical-domain residues. Among the apical-domain residues, the helices K and L stand out due to their higher amplitudes of vibration. Because of their involvement in the reaction cycle of GroEL, we surmise that the large-scale movement of helices K and L[18] and the change in the surface polarity of the cavity are linked. Thus, the flexibility in this region is crucial for the annealing function of GroEL. The

Fig. 7. Analysis of the dominant modes of the GroEL R″→T transition. (a, b) Results for the GroEL model with two adjacent subunits. (a) The amplitudes of motion in the two dominant normal modes. (b) δω for each residue for the dominant modes. (c, d) Results for a GroEL subunit in the R″ state with adjacent trans equatorial domain. (c) The amplitudes of motion in the dominant normal modes. (d) δω for each residue for the same modes. (e and f) Results based on C^{α}–side chain ENM for a single GroEL subunit in the R″ state with a neighboring GroES subunit. (e) The amplitudes of motion in the dominant normal modes (shown in green and blue). There is a significant reduction in the amplitudes of motion of the GroES binding residues as well as of helices K and L as compared to the amplitudes calculated based on a single subunit of GroEL alone (black dashed line). (f) Response δω as a function of residue number *i* for the dominant modes.

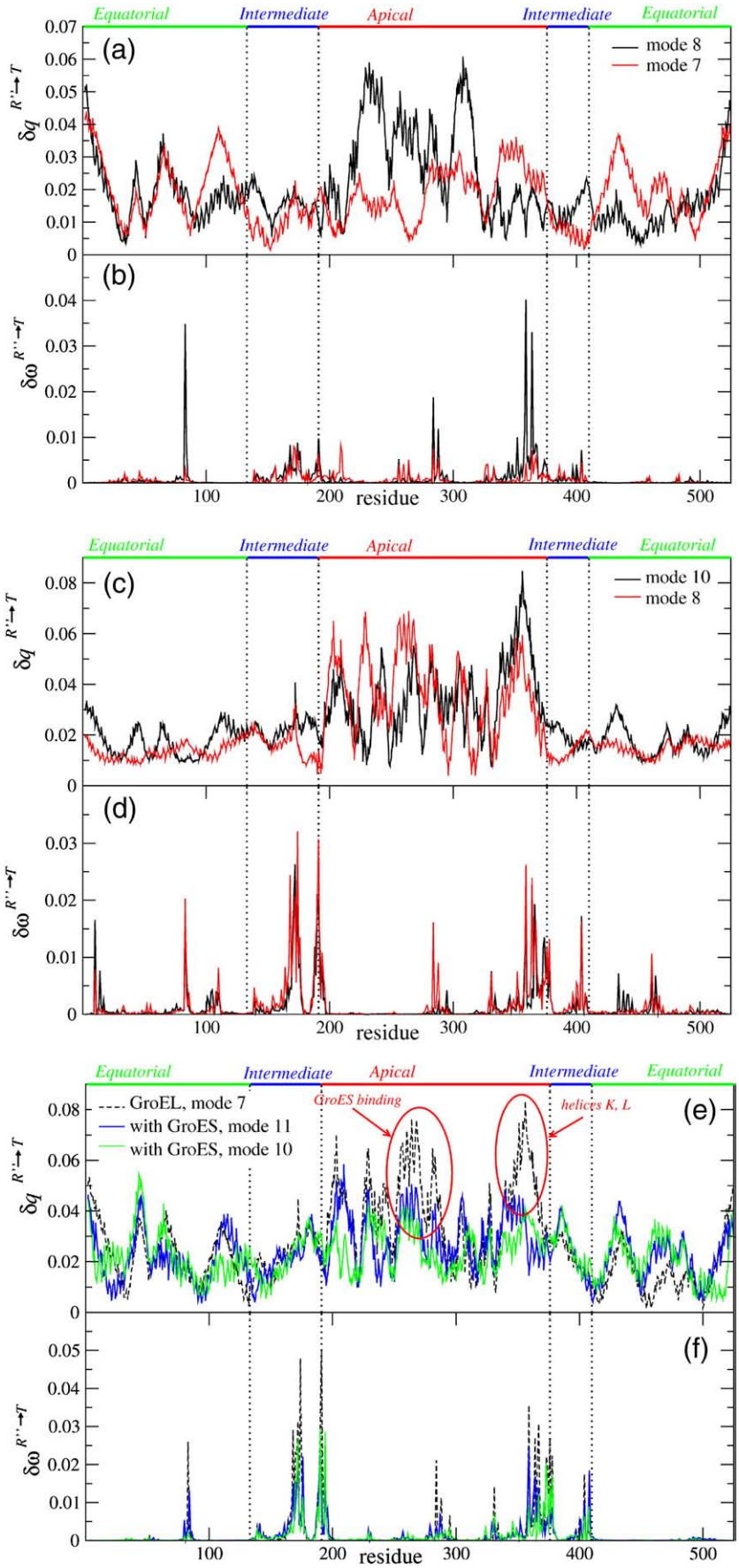

Fig. 7 (*legend on previous page*)

correlations between the motions of residues in mode 7 are shown in Fig. 6b. The amount of correlation within each domain is striking. The notable exceptions are helices K and L that are correlated with the adjacent intermediate-domain residues rather than the other apical-domain residues.

The response of mode 7 to perturbations (dotted line in Fig. 7f) identifies 61 residues as part of the allostery wiring diagram. Most of the hot-spot residues are located in the intermediate domain. A few of the residues are located in the equatorial or apical domains, especially in helix L, which, as noted earlier, moves opposite to the direction of most of the apical domain. The locations of the hot-spot residues, listed in Table 1, are also shown on the GroEL R″-state structure in Fig. 2b. The residues with the highest δω values are D83, V174, E191, and D359, and we therefore expect that the R″→T transition is most sensitive to mutations of these residues. Of those residues, D359 is moderately conserved, D83 is highly conserved, and E191 is at the beginning of a conserved word.[38] The relevance of E191 is supported by earlier experiments that showed that the mutation E191G inhibits GroEL functionality.[40]

GroES plays a passive role in the R″→T transition

In the R″ state, the GroEL apical domain is capped by GroES. To determine how the presence of GroES affects the normal modes and influences the allostery, we performed a normal mode analysis of the GroEL–GroES complex, illustrated in Fig. 2c. In the presence of GroES, the R″→T transition is no longer described by a single normal mode. Instead, mode 11 gives an overlap of 0.56 and mode 10 an overlap of 0.53 with the transition (Fig. 4b). Both modes describe a tilt of the apical domain toward the equatorial domain, as discussed in the previous section. The amplitudes of motion of all the GroEL residues in both of the modes (Fig. 7e) reveal two regions where the presence of GroES suppresses GroEL residue fluctuations. First, many apical domain residues between 225 and 275 are part of the GroES binding region and the reduction in their vibration amplitudes is not unexpected. However, the second region (residues 350–365) spans the region associated with helices K and L and the loop connecting them, and is not in direct contact with GroES. We had already noted that many of the residues in this region were hot-spot residues that are linked to the annealing function of GroEL. We find the plausible signaling pathway linking GroES binding residues and helices K and L remarkable. Experiments have shown that the GroEL mutant D361K lacks the ability to bind GroES.[41] Our calculations suggest that the residues 350–365 are also allosterically linked to GroES binding residues, and mutations of several of those residues could change the affinity of GroEL for GroES.

From the SPM analysis of the GroEL–GroES structure (Fig. 7f), the residues with large δω or hot-spot residues were identified (see Table 1). The residues with the highest δω values are either at or near positions D83, V174, E191, and D359. Remarkably, there are no hot-spot residues among the GroES binding residues, highlighting the passive role of GroES in the R″→T transition. We predict that mutations of the GroES binding residues should have a minimal effect on the GroEL R″→T transition, which involves release of GroES and ADP.

In addition, we also found the wiring diagram of GroES that corresponds to the normal mode 11 of the GroEL–GroES complex. The locations of GroES hot-spot residues are shown in Fig. 2c. Kass and Horovitz,[37] using correlated mutations in GroEL and GroES, found networks where mutations of cochaperonin residues were statistically linked to mutations of chaperonin residues. They identified GroES residues S21, L27, S35, and Y71 as being relevant (Table 2 of Ref. 37). We find that GroES residues S21 and S35 are indeed hot-spot residues. Also, residues 68–70, which neighbor residue Y71, are hot-spot residues. Surprisingly, none of the GroES hot-spot residues belong to the GroEL binding motif,[42] which is perhaps related to the functional proximity of the requirement that the GroEL binding should be promiscuous. Bioinformatic analysis[39] has also determined that, with the exception of G24, none of the GroES mobile loop (binding) residues were conserved.

Allostery wiring diagram that transmits the signal from the trans ring

The GroEL cis ring transition from R″ to T state is triggered by a signal from the trans ring.[25] We therefore calculated the normal modes of the GroEL R″-state structure together with two trans ring equatorial domains as illustrated in Fig. 2d. For this system, modes 10 (overlap 0.59) and 8 (overlap 0.57) describe the R″→T transition most accurately. Inclusion of the trans ring equatorial domain significantly reduced the amplitudes of vibration of the adjacent cis equatorial domain (Fig. 7c). In addition, the exact motion of residues of helices K and L (in mode 10) and GroES binding region (mode 8) are also reduced. These findings again highlight the allosteric connections between these residues.

The inclusion of the trans ring has allowed us to find several new hot-spot residues in the equatorial domain that trigger intraring communication. The 20 residues, obtained using SPM at the interring interface, are 9–15, 104, 105, 109–111, 434, 435, and 460–465. The relatively large number of critical interring interface residues highlights the importance of the interface between the rings, and hence the role of the trans ring on the R″→T transition. This is in sharp contrast to the lack of GroES interface residues among hot-spot residues for the R″→T transition or the smaller number of trans ring interface residues in the T→R transition. These observations show that the signal from the trans ring, resulting in negative cooperativity, arises almost exclusively from interactions across equatorial

domains between the rings. Many of the hot-spot interring interface residues such as residues A109, G110, and 463–465 were predicted previously using the C^{α}-based ENM.[28] However, residues 9–15 and E434 and D435 are also critical to this transition. In Ref. 41, mutant R13G was experimentally found to have reduced ATPase activity. The effects of R13G–A126V were also analyzed by Aharoni and Horovitz,[43] who found that the mutation perturbed the negative cooperativity between the rings but not the positive cooperativity within each ring, consistent with our findings. We predict that mutations of many residues among 9–15 as well as E434 and D435 would alter the GroEL R″→T transition and associated ATPase activity.

Finally, we also performed a normal mode analysis of a double-subunit fragment of GroEL (Fig. 2e) to investigate the role and significance of the interface residues between two subunits within the cis ring. Modes 8 and 7 have the highest overlap with the transition; the overlap values are 0.44 and 0.4, respectively. The modes are rather similar to the ones we described for the single-chain R″ to T transition. The SPM yields a familiar list of significant residues, including only very few (2+6) intersubunit hot-spot residues (G256, E257, chain A and A84, 303–307, chain G). Of those, a mutation, E257A, was experimentally found to alter GroEL ATPase rate.[44] Our list of hot-spot residues does not include R197[30,45] (it includes the neighboring D196). This could be due to the fact that in the crystal structure we used (1AON), R197 was modeled as Alanine. We thus found a significantly smaller number of interface hot-spot residues within a ring compared to interring interface hot-spot residues. This may be a reflection that it is the signal from the trans ring that drives the R″→T transition in the cis ring. The extent of negative cooperativity between rings is, to a large extent, associated with movements across the equatorial domains.

GroEL signaling wiring diagrams were also recently obtained by Chennubhotla *et al.*[46,47] using a novel Markov signal propagation model. Their predicted signal transduction pathways include some of the same residues that we have noted as critical to GroEL allostery. For instance, residues E409, E408, A370, K371, and M193 from one of the pathways[46] are all hot-spot residues in the R″→T transition (see Table 1). Many of the interface residues within K105–A109 and R452–E461 are also part of the allostery wiring diagram, involved in the interring communication, according to our calculations. The agreement between the predictions made using two vastly different techniques suggests that the common residues must play a critical role in the reaction cycle of GroEL.

Functionally linked residues using a sequence-based method

The SPM is suited to obtain a network of residues that are most relevant in driving the large-scale domain movements in the reaction cycle of GroEL. Residues that are responsible for binding of substrate protein and ATP that are localized in certain regions of GroEL are not predicted by SPM, as they do not pertain to mechanical movements. In this context, methods that analyze the evolutionary imprints might be more useful. To tease out the functionally relevant residues, we have used our formulation[8] of the SCA introduced by Lockless and Ranganathan[7] and Suel *et al.*[48] The assumption of SCA is that statistically significantly coincident mutations of a group of residues stem from biological or functional linkage. It should be stressed that the SCA cannot identify residues that are absolutely conserved across a given family.

GroEL

Using a set of 499 aligned sequences for GroEL (Methods), we found a cluster of 21 co-evolving residues. The residues are listed in Table 2 and depicted on GroEL structure in Fig. 8. It is interesting to note that a large fraction of the residues are in the apical domain, located particularly at or near the putative substrate protein and GroES binding sites. This set includes residues N206, D224, I227, I230, R231, A241, K242, I270, E304, and K311.[41,49,50] Many of the remaining equatorial- and intermediate-domain residues can be linked to GroEL interaction with ATP. For example, S154, L494, and T500 are near the ATP binding site. A383 has been experimentally linked to GroEL ATPase activity,[41] while M389 and G180 are near A383. C138 neighbors P137 that in turn affects GroEL ATPase rate.[51] M389 also belongs to helix M that runs close to the ATP binding pocket. Thus, evolutionary analysis of GroEL sequences has produced a cluster of mainly ligand-binding related residues. The set of residues predicted to be relevant by SCA is in sharp contrast with the structure-based wiring diagrams obtained from SPM that consist largely of residues that drive domain movements.

GroES

Using 487 GroES-related sequences (Methods), the SCA predicts 17 co-evolving residues (Table 2 and Fig. 8). Of those, 4 belong to the mobile loop (S21, L27, T28, and G29) and 2 (R14 and S35) are adjacent to it. Several of the other residues (R4, H7, I66, N68, K74, and L92) belong to the GroES–GroES intersubunit interface and thus play a role in the GroES heptamer formation. Three of the residues (S21, L27,

Table 2. SCA-predicted residue clusters

Protein	SCA-predicted residues
GroEL	C138 K142 S154 G180 N206 D224 I227 I230 R231 A241 K242 I270 E304 K311 A383 M389 G472 G474 L494 T500 M514
GroES	R4 H7 R14 S21 L27 T28 G29 S35 I66 N68 D69 G70 K74 S75 V83 M86 L92

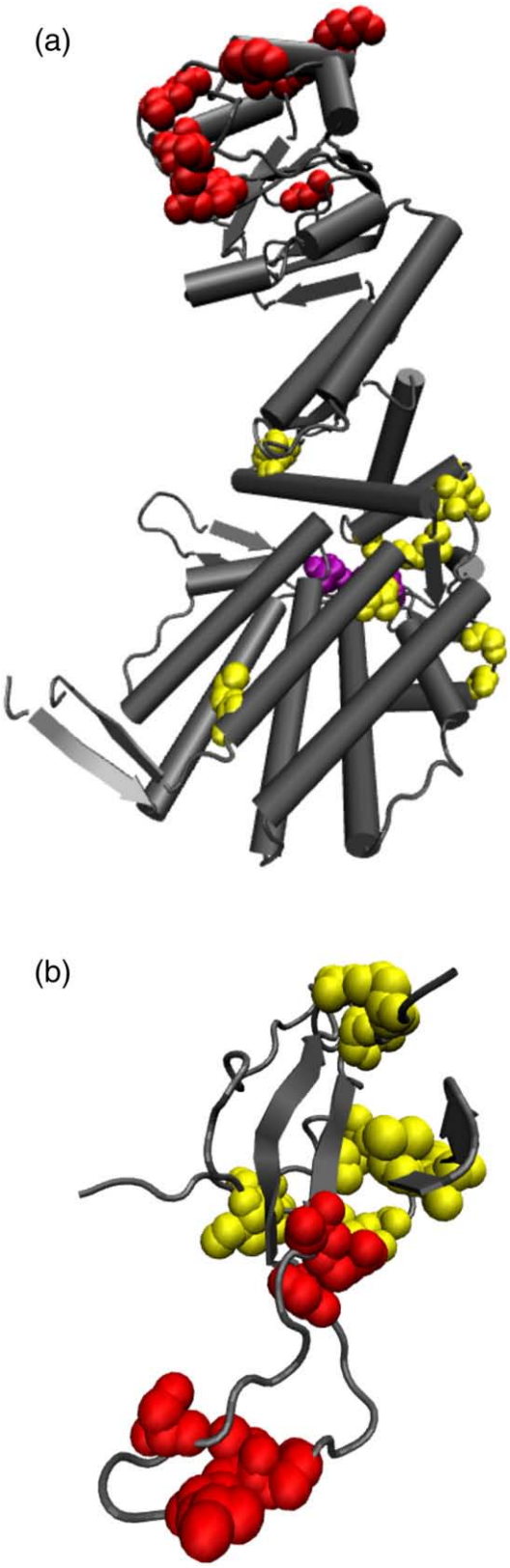

Fig. 8. Set of residues predicted using SCA. (a) The most significant GroEL residues in the apical domain are shown in red; the intermediate and equatorial domain residues are shown in yellow. Nucleotide is shown in purple. (b) The most significant GroES mobile loop and adjacent SCA residues are shown in red; the others are shown in yellow.

and S35) have been previously identified by Kass and Horovitz using a sequence-based approach.[37] Just as in GroEL, the network of residues that are predicted to be relevant in the allosteric transitions of GroEL–GroES is markedly different (Figs. 8 and 2c).

To understand the differences between the predictions based on SPM and SCA, we consider the foundations of both methods. In the structure-based SPM, the networks of residues that respond to local perturbations are probed. As such, they provide an allostery wiring diagram that encodes for the conformational dynamics without regard to residues that have evolved to bind ligands such as ATP. Consequently, we expect SPM to produce a set of residues (localized in the hinges or interfaces) that are linked to the allosteric motions of a protein. In contrast, SCA is based on the covariations of evolutionarily related sequences and does not take structural changes into account. As a result, the SCA can locate accurately residues that are functionally relevant. In the context of GroEL, these functions pertain to ATP binding and hydrolysis and recognition sites for substrate proteins. Clearly, the overall annealing function requires a linkage between allosteric transitions (related to mechanics) and that are driven by functional recognition of ATP and substrate protein.

SPM and SCA are vastly different approaches for analyzing the information encoded in protein sequences and structures. We postulate that the SPM can be a useful tool to extract mechanically critical residues that are related to protein dynamics in the native state, whereas SCA might produce functionally significant residues that are related to protein–protein and protein–nucleotide interactions. The current anecdotal evidence suggests that both sequence and structure-based methods might be necessary to identify networks of residues that drive reaction cycles in molecular machines.

Conclusions

We have determined the network of residues that transmit allosteric signals in the key transitions that occur in the reaction cycle of GroEL using a new C^{α}–side chain representation of proteins and SPM. The explicit inclusion of side chains in the model gives a more accurate description of the topology of the protein and the interactions within the structure than C^{α} models. The C^{α}–side chain ENM also incorporates information about the chemical and physical characteristics of the amino acids by including residue-specific spring constants. Using the model in conjunction with SPM,[6] we have mapped the allostery wiring diagrams for the T→R and the R″→T transitions of GroEL. We identify previously unidentified residues that transmit the allosteric signals both within a single subunit and at the subunit interfaces (see below). We also find that the spectacular movements of helices K and L, throughout the GroEL allosteric cycle, are needed for the

alterations in the polarity of the cavity. During the T→R and R″→T transitions, the K and L helices move in a concerted manner, but are always anti-correlated to the other apical domain residues. The coordinated movement of helices K and L, which was previously linked to the switch of the polarity of the GroEL cavity during the R→R″ transition,[18] is crucial for the annealing function of GroEL.

The allostery wiring diagram for the T→R transition is predominantly localized in the intermediate domain hinge. In addition, a few apical and equatorial domain residues, including R58, D83, E209, and K327, also play a crucial role. The charged K327 is particularly critical to the T→R transition. It follows that mutations of any the residues in the wiring diagram or creation of disulfide bonds between them (including D83–K327)[36] should measurably change the GroEL T→R transition.

The origin of intraring cooperativity is dictated by residues at the interface of two adjacent subunits. We find a large number of intraring interface residues in the wiring diagram, providing a microscopic foundation for the origin of the intraring positive cooperativity. In particular, G244 is critical to the concerted transition. Relatively few interring interface residues are important to the T→R transition, in sharp contrast to the allostery wiring diagram of the R″→T transition (see below), which highlights that intraring interactions, and not interring interactions, determine the positive cooperativity within a ring.

The R″→T transition is supported by an allostery wiring diagram that also consists of a large number of intermediate domain hinge residues and a few residues in the apical and equatorial domains (Fig. 2b and Table 1). We find the R″→T transition is particularly sensitive to mutations of residues D83, V174, E191, and D359, and hence mutations of these residues should compromise the transition and the associated ATPase rate significantly. Dynamics involving the multisubunit R″-state structures reveal a new picture for the allosteric communication between the subunits in the R″→T transition. The wiring diagram includes interring interface residues that provide a communication pathway for the signal from the trans ring. The large number of significant interring interface residues furnishes us with a detailed structural foundation for the importance of the negative cooperativity between the rings. The negative cooperativity between the rings is sensitive to mutations of the interface residues N10 and E461 and the a few adjacent residues. Interface residues E434 and D435 also are critical in determining communication between the rings. There are relatively few intraring interface residues that are critical to the R″→T transition and, thus, based on the allostery wiring diagram, we conjecture that it is the communication between the rings that is critical to the R″→T transition. The relative motions of the interface between the rings determine the extent of negative cooperativity.

Remarkably, in the R″→T transition, the wiring diagram of GroEL does not include any GroES binding residues. Neither does the wiring diagram of GroES include any GroEL binding residues, leading us to conclude that the role of GroES is purely passive in the R″→T transition. We do find, however, an allosteric connection between the GroES binding sites and GroEL helices K and L. We thus predict that mutations of the residues of the helices will affect the affinity of GroES for GroEL, even though these residues are not part of the GroES binding site. The GroES residues that are important to the chaperonin allosteric cycle include residues S21 and S35 and residues that are adjacent to them as well as residues N68–G70. It is also likely that mutations in K and L helices could affect the release of GroES and, hence, the efficiency of the intact machine. Our results clearly show that positive cooperativity involves a network of residues within a single ring, while the origin of negative cooperativity is linked to interactions between the interfaces of interring equatorial domains. It is worth emphasizing that these subtle effects are difficult to realize in the SR1 construct that completely arrests the R″→T transition.

We have also analyzed the co-evolution of mutations of GroEL and GroES residues. Using SCA, we have mapped a markedly different, yet functionally relevant, wiring diagrams for GroEL and GroES. We find that several substrate protein and GroES binding residues (N206, D224, I227, I230, R231, A241, K242, I270, E304, K311) and ATPase-rate- or ATP-binding-related residues (S154, L494, T500, A383) belong to the same evolutionarily linked cluster in GroEL. For GroES, we find an association between several mobile loop residues (S21, L27, T28, G29) and GroES/GroES interface residues (R4, H7, I66, N68, K74, L92).

The two different methods (SPM and SCA) used to analyze GroEL and GroES yield rather different allostery wiring diagrams. There are only a few residues common to both the networks (GroEL residues G180 and R231, GroES residue S21), but the coincidences are not statistically significant. The predicted differences can be understood based on the foundations of the methods. The determination of the allostery wiring diagram using SCA is based on evolutionary covariation of residues. The sequence-based prediction could be linked to the functional relationship or interactions between a protein and its ligands. Because in SCA only one conserved site (j) in a subalignment of a multiple sequence alignment is treated as a perturbation, only covariation with respect to j can be extracted. It is clear that multiple variations in sequence positions can affect function. The inability to probe sequence covariation at multiple positions is an inherent weakness of SCA. The SPM is based on protein structures and thus produces sets of transition-specific residues that are structurally or mechanically important. In contrast to SCA, the SPM automatically accounts for collective behavior and hence is physically a more appropriate method for determining the allostery wiring diagram.

Methods

Elastic network models

We introduce a C^{α}–side chain ENM where each amino acid (except Gly) is represented by the coordinates of the C^{α} atom and that of the center of mass of the side-chain heavy atoms. For Gly, only the C^{α} atom is used. We impose a harmonic potential between all interaction sites (C^{α} and side chain) that are within a cutoff radius R_c in the folded structure. In the structure-based elastic network representation of the protein, the potential energy is

$$E = \frac{1}{2}\sum_{i,j:d_{ij}^0<R_c} \kappa_{ij}\left(d_{ij} - d_{ij}^0\right)^2 \quad (1)$$

where d_{ij} is the distance between interaction sites i and j, d_{ij}^0 is the corresponding distance in the native structure, and κ_{ij} is the spring constant (see below). The sum is over all pairs of sites that are in contact in the native conformation. Sites i and j are in contact if d_{ij} is less than a cutoff distance, R_c, whose value is chosen so that the B-factors using the energy function in Eq. (1) and the measured values are as close as possible.[28] The R_c values for different allosteric transitions in GroEL are given in Table 3.

The spring constants, κ_{ij}, are chosen based on the physical and chemical properties of the residues. We use $\kappa_{ij} = \varepsilon_{ij}/(\sigma_i/2 + \sigma_j/2)^2$ where ε_{ij} is the Betancourt–Thirumalai potential[23] and σ_i is the van der Waals diameter of the residue i. The Betancourt–Thirumalai potential, a statistical effective pair potential between amino acid residues, is constructed based on empirical knowledge of the contact frequencies of amino acids in known protein structures.

Normal mode analysis

We perform a normal mode analysis using the energy function in Eq. (1) and obtain a spectrum of frequencies for the normal modes along with the corresponding eigenvectors. Since only a few of the lowest-frequency normal modes often describe the allosteric transitions, we compute the overlap between the conformational changes based on the crystal structures of the two allosteric states (α and β) and the eigenvector a_M of the Mth normal mode (calculated based on the state α)[52] using

$$I_M^{\alpha\to\beta} = \frac{\sum \mathbf{a}_{iM}\,\Delta\mathbf{r}_i}{\sqrt{\sum \mathbf{a}_{iM}^2 \sum \Delta\mathbf{r}_i^2}} \quad (2)$$

where $\Delta\mathbf{r}_i$ is the change in the position of the ith site between α and β. It follows from Eq. (2) that $0 \le I_M \le 1$. The closer the value is to 1, the better a given allosteric transition is represented by mode M.

Table 3. Cutoff values in the C^{α}–side chain ENM calculations

Transition and model	R_c (Å)
T→R, single subunit	10
T→R, two adjacent subunits	8
T→R, single subunit with trans equatorial (e) plate	10
R″→T, single subunit	10
R″→T, single subunit+GroES	12
R″→T, two adjacent subunits	10
R″→T, single subunit with trans E plate	10

R_c, the cutoff value for native contacts, is obtained by best fit to experimentally determined B-factors.[28]

The relative amplitude of a site i in mode M is calculated from the normalized eigenvectors a_{iM} using:

$$\delta q_{iM}^{\alpha\to\beta} = \sqrt{a_{i_xM}^2 + a_{i_yM}^2 + a_{i_zM}^2} \quad (3)$$

where a_{i_γ} denotes the displacement of the site i in the γ direction.

Structural perturbation method

The SPM is best explained in terms of propagation of response of a local perturbation in an ordered system. Normal modes are the elementary excitations that transmit the local vibrations in ordered solids. In solids, local perturbation propagates throughout the sample coherently indicative of a long-range order. Similarly, the extent to which a residue at a given site in a structure responds to a perturbation far away can be used to assess allosteric coupling. The SPM allows us to quantify how strong the coupling is to a mutation at a particular site. The greater the response (higher $\delta\omega_{iM}$, see below), the more significant a specific residue is to a given mode. The high $\delta\omega_{iM}$ residues map a network or an allosteric wiring diagram that provides the energetic foundation for the allosterically significant modes. In general, it is found that the residues in the allostery wiring diagram are also strongly conserved.[6]

In practice, SPM probes the response of a normal mode M to a mutation of a residue i. In the context of normal mode analysis, perturbation of the spring constant around a site mimics the effect of a mutation. The response to such a perturbation is calculated using

$$\delta\omega_{iM}^{\alpha\to\beta} = \frac{1}{2}\sum_{j:d_{ij}^0<R_c} \delta\kappa_{ij}\left(d_{ij,M} - d_{ij}^0\right)^2 \quad (4)$$

where $\delta\kappa_{ij}$ is the perturbed spring constant and $(d_{ij,M} - d_{ij}^0)$ is the change in the distance between residues i and j in a mode M. It should be stressed that the SPM probes global response to short-range perturbations [Eq. (4)], which is the hallmark of allostery. The residues that are associated with high $\delta\omega_{iM}$ values are considered most critical to the significant modes and thus can be identified as the hot-spot residues for a given transition. The higher the $\delta\omega$ value of a residue, the more important it is to the transition. In practice, we use the criterion $\delta\omega_{iM} > 2\langle\delta\omega_M\rangle$ to determine the hot-spot residues, where $\langle\delta\omega_M\rangle$ is given by

$$\langle\delta\omega_M\rangle = \sum_{i=1}^{N} \delta\omega_{iM}/N \quad (5)$$

with N being the total number of residues. To obtain the specific residues listed in Table 1, we used a local average (average over the side chain and C^{α} of the site i and its nearest neighbors) value in the highest overlap mode for $\delta\omega_i$.

Statistical coupling analysis

In order to identify the residues that are evolutionarily linked, we used the sequence-based SCA.[7,48] In the reformulation of the SCA,[8] the statistical "free energy" at position i in a multiple sequence alignment is defined as,

$$\frac{\Delta G_i}{kT^*} = \sqrt{\frac{1}{C_i}\sum_{x=1}^{20}\left[p_i^x \ln\left(\frac{p_i^x}{p_x}\right)\right]^2} \quad (6)$$

where kT^* is an arbitrary energy unit, C_i is the number of types of amino acids that appear at position i, p_x is the mean frequency of amino acid of type x in the alignment. $p_i^x = \frac{n_i^x}{N_i}$, where n_i^x is the number of times an amino acid of type x appears at a position i, and $N_i = \sum_{x=1}^{20} n_i^x$. This definition of free energy resembles the sequence entropy $S_i = -\sum_{x=1}^{20} p_i^x \ln(p_i^x)/\ln(20)$.

The basic hypothesis of the SCA is that a covariation between two positions i and j may be inferred by comparing the statistical properties of the multiple sequence alignment and a subalignment of sequences (derived from the original multiple sequence alignment) in which a given amino acid is conserved ($S_j=0$) at j (restricted subalignments). The restriction, $S_j=0$, in the subalignment is referred to as a sequence perturbation. The effect of the perturbation is assessed from the statistical free-energy change,

$$\frac{\Delta\Delta G_{ij}}{kT^*} = \sqrt{\frac{1}{C_i}\sum_{x=1}^{20}\left[p_{i,R}^x \ln\left(\frac{p_{i,R}^x}{p_x}\right) - p_i^x \ln\left(\frac{p_i^x}{p_x}\right)\right]^2} \qquad (7)$$

where $p_{i,R}^x = n_{i,R}^x/N_{i,R}$ and $n^x{}_{i,r}$ is the number of restricted sequences in the subalignment in which amino acid x appears in the ith position, and $N_{i,R} = \sum_{x=1}^{20} n_{i,R}^x$.

In practice, we performed a standard PSI-BLAST query for GroEL (based on 1AON.pdb chain A) and GroES (1AON.pdb, chain O). The 500 best matches were obtained and realigned using Clustalw package.[53] We manually deleted the sequences that were too long or too short in order to minimize long gaps in the alignment. We were left with 499 aligned sequences, each 587 residues long, including gaps, for GroEL, and 487 sequences of 139 residues, including gaps, for GroES. In subsequent clustering of the restricted subalignments, we only considered the ones that contained at least 55% of the full multiple sequence alignment. This criterion was based on the requirement that all subalignments obey the central limit theorem.[8]

We performed a hierarchical clustering of $\Delta\Delta G_{ij}$ by using the Euclidean distance as a measure of the similarity of two vectors.[8] We obtained robust clusters of positions i and perturbations j and considered a residue significant when it was present in both the clustered positions and clustered perturbations. Our final clusters contained 21 significant residues for GroEL and 17 for GroES.

Acknowledgements

R.T. is a Ruth A. Kirschstein postdoctoral fellow and her work was supported by the National Institute of General Medical Sciences (1F32GM082009). D.T. is grateful to the National Institutes of Health (1R01GM067851-01) and the Air Force Office of Scientific Research (FA9550-07-1-0098) for support.

References

1. Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroES(ADP)7 chaperonin complex. *Nature*, **388**, 741–750.
2. Boyer, P. D. (1997). The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749.
3. Vale, R. D. & Milligan, R. A. (2000). The way things move: looking under the hood of molecular motors. *Science*, **288**, 88–95.
4. Houdusse, A. & Sweeney, H. L. (2001). Myosin motors: missing structures and hidden springs. *Curr. Opin. Struct. Biol.* **11**, 182–194.
5. Zheng, W., Brooks, B. R. & Thirumalai, D. (2006). Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc. Natl Acad. Sci. USA*, **103**, 7664–7669.
6. Zheng, W., Brooks, B. R., Doniach, S. & Thirumalai, D. (2005). Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure*, **13**, 565–577.
7. Lockless, S. W. & Ranganathan, R. (1999). Evolutionary conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295–299.
8. Dima, R. & Thirumalai, D. (2006). Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci.* **15**, 258–268.
9. Levitt, M., Sander, C. & Stern, P. S. (1983). Normal-mode dynamics of a protein: bovine pancreatic trypsin inhibitor. *Int. J. Quant. Chem.* **10**, 181–199.
10. Go, N. (1983). Theoretical-studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
11. Ma, J. (2003). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, **13**, 373–380.
12. Bahar, I. & Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **388**, 586–592.
13. Maragakis, P. & Karplus, M. (2005). Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J. Mol. Biol.* **352**, 807–822.
14. Rader, A. J., Chennubhotla, C., Yang, L. W. & Bahar, I. (2006). The Gaussian network model: theory and applications. In *Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems* (Cui, Q. & Bahar, I., eds), pp. 41–64, Chapman & Hall CRC Press, Chicago, IL.
15. Sanejouand, Y. H. (2006). Functional information from slow mode shapes. In *Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems* (Cui, Q. & Bahar, I., eds), pp. 91–109, Chapman & Hall CRC Press, Chicago, IL.
16. Tama, F. & Brooks, C. L., III (2006). Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 115–133.
17. Schröder, G. F., Brunger, A. T. & Levitt, M. (2007). Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure*, **15**, 1630–1641.
18. Hyeon, C., Lorimer, G. H. & Thirumalai, D. (2006). Dynamics of allosteric transitions in GroEL. *Proc. Natl Acad. Sci. USA*, **103**, 18939–18944.
19. Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908.
20. Buchete, N. V., Straub, J. E. & Thirumalai, D. (2004). Development of novel statistical potentials for protein fold recognition. *Curr. Opin. Struct. Biol.* **14**, 225–232.
21. Cheung, M. S., Klimov, D. & Thirumalai, D. (2005). Molecular crowding enhances native state stability and refolding rates. *Proc. Natl Acad. Sci. USA*, **102**, 4753–4758.
22. Klimov, D. K. & Thirumalai, D. (2000). Mechanisms

and kinetics of β-hairpin formation. *Proc. Natl Acad. Sci. USA*, **97**, 2544–2549.
23. Betancourt, M. R. & Thirumalai, D. (1999). Pair potentials for protein folding: choice of reference states and sensitivity of predicted motive states to variations in the interaction schemes. *Protein Sci.* **8**, 361–389.
24. Xu, Z. & Sigler, P. B. (1998). GroEL/GroES: structure and function of a two-stroke folding machine. *J. Struct. Biol.* **124**, 129–141.
25. Thirumalai, D. & Lorimer, G. H. (2001). Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269.
26. Horwich, A. L., Farr, G. W. & Fenton, W. A. (2006). GroEL–GroES-mediated protein folding. *Chem. Rev.* **106**, 1917–1930.
27. Lin, Z., Madan, D. & Rye, H. S. (2008). GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **15**, 303–311.
28. Zheng, W., Brooks, B. R. & Thirumalai, D. (2007). Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. *Biophys. J.* **93**, 2289–2299.
29. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. (2002). Molecular mechanisms of chaperonin GroEL–GroES function. *Biochemistry*, **42**, 491–501.
30. Ma, J. & Karplus, M. (1998). The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc. Natl Acad. Sci. USA*, **95**, 8502–8507.
31. Grason, J., Gresham, J. & Lorimer, G. (2008). Setting the chaperonin timer: a two-stroke, two-speed, protein machine. *Proc. Natl Acad. Sci. USA*, **105**, 17339–17344.
32. 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.
33. Ranson, N. A., Farr, G. W., Roseman, A. M., Gowen, B., Fenton, W. A., Horwich, A. L. & Saibil, H. R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, **107**, 869–879.
34. Betancourt, M. R. & Thirumalai, D. (1999). Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* **287**, 627–644.
35. Yang, Y. (2006). Site-directed mutagenesis of GroEL: developing a system for monitoring allosteric movements by flourescence resonance energy transfer. Master's thesis, University of Maryland, College Park, MD.
36. Murai, N., Makino, Y. & Yoshida, M. (1996). GroEL locked in a closed conformation by an interdomain cross-link can bind ATP and polypeptide but cannot process further reaction steps. *J. Biol. Chem.* **271**, 28229–28234.
37. Kass, I. & Horovitz, A. L. (2002). Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins*, **48**, 611–617.
38. Brocchieri, L. & Karlin, S. (2000). Conservation among HSP60 sequences in relation to structure, function, and evolution. *Protein Sci.* **9**, 476–486.
39. Stan, G., Thirumalai, D., Lorimer, G. H. & Brooks, B. R. (2003). Annealing function of GroEL: structural and bioinformatic analysis. *Biophys. Chem.* **100**, 453–467.
40. Klein, G. & Georgopoulos, C. (2001). Identification of important amino acid residues that modulate binding of *Escherichia coli* GroEL to its various cochaperones. *Genetics*, **158**, 507–517.
41. 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.
42. Stan, G., Brooks, B. R., Lorimer, G. H. & Thirumalai, D. (2005). Identifying natural substrates for chaperonins using a sequence-based approach. *Protein Sci.* **14**, 193–201.
43. Aharoni, A. & Horovitz, A. (1996). Inter-ring communication is disrupted in the GroEL mutant Arg13 Gly: Ala126 Val with known crystal structure. *J. Mol. Biol.* **258**, 732–735.
44. Danziger, O., Shimon, L. & Horovitz, A. (2006). Glu257 in GroEL is a sensor involved in coupling polypeptide substrate binding to stimulation of ATP hydrolysis. *Protein Sci.* **15**, 1270–1276.
45. Yifrach, O. & Horovitz, A. (1994). Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry*, **34**, 5303–5308.
46. Chennubhotla, C., Yang, Z. & Bahar, I. (2008). Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol. BioSyst.* **4**, 287–292.
47. Chennubhotla, C. & Bahar, I. (2006). Markov propagation of allosteric effects in biomolecular systems: application to GroEL–GroES. *Mol. Syst. Biol.* **2**, 36.
48. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. I. (2003). Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59–68.
49. Buckle, A. M., Zahn, R. & Fersht, A. R. (1997). A structural model for GroEL–polypeptide recognition. *Proc. Natl Acad. Sci. USA*, **94**, 3571–3575.
50. Chen, L. L. & Sigler, P. B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell*, **99**, 757–768.
51. Bochkareva, E. S., Horovitz, A. & Girshovich, A. S. (1994). Direct demonstration that ATP is in contact with Cys-137 in chaperonin GroEL. *J. Biol. Chem.* **269**, 44–46.
52. Zheng, W. & Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl Acad. Sci. USA*, **100**, 13253–13258.
53. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.