Solution structures of GroEL and its complex with rhodanese from small-angle neutron scattering

P Thiyagarajan[1], SJ Henderson[2] and A Joachimiak[1]*

Background: Molecular chaperonins 60 are cylindrical oligomeric complexes which bind to unfolded proteins and assist in their folding. Studies to identify the location of the protein substrate have produced contradictory results: some suggest that the substrate-binding site is buried within the interior of the complex, whereas others indicate an external (polar) location.

Results: Small-angle neutron scattering (SANS) measurements were made on GroEL chaperonin and on a complex of GroEL with rhodanese. The radius of gyration and the molecular weight determined from SANS measurements of GroEL agree well with those from its crystal structure. The positions of residues which were unresolved in the crystal structure have been confirmed. In addition, through model fitting of the SANS data, conformational changes in solution have been assessed and the location of bound rhodanese has been determined.

Conclusions: The overall structure of GroEL in solution is similar to the crystal structure. In GroEL the N-terminal and C-terminal residues are organized compactly near the equator of the cylinder and the apical domains are flared by about 5°. The best fit of SANS data suggests the existence of an equilibrium between the complex and single rings and monomers. SANS data for the GroEL–rhodanese complex are consistent with a model wherein one rhodanese molecule binds across the opening to the chaperonin cavity, rather than within it.

Addresses: [1]Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA and [2]Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

*Corresponding author.

Key words: chaperonin, conformational changes, GroEL solution structure, rhodanese binding, SANS

Received: **15 Sep 1995**
Revisions requested: **13 Oct 1995**
Revisions received: **30 Oct 1995**
Accepted: **1 Nov 1995**

Structure 15 January 1996, **4**:79–88

Introduction

The relationship between protein structure in solution and in crystal form has been studied extensively [1–4]. Due to the difficulties encountered with the crystallization of large multisubunit macromoleculer assemblies, solving their structures by X-ray crystallography is highly challenging. Furthermore, the conditions of crystallization and packing in the crystal may alter local three-dimensional structures [5], bringing into question the functional validity of macromolecular crystal structures [6,7]. Small-angle scattering (SAS), using either X-rays (SAXS) or neutrons (SANS), can provide important low-resolution information on the structure of such assemblies and their interactions in solution that may have more functional relevance than crystal-derived data. These methods yield information on the size, morphology, and composition of macromolecular complexes and can detect conformational changes and substrate binding. When the crystal structure coordinates of the individual components of macromolecular complexes are available, more reliable information on the solution structure of complexes can be delineated by using that information in the model fitting of the SAS data.

SAXS and SANS methods, when applied to biomacromolecules in solution, yield complementary information due to the difference in the nature of the interactions of X-rays and neutrons with matter. The X-ray scattering from an atom results from the electron clouds around its nucleus, whereas neutrons are scattered by the nucleus. The contrast for the X-ray scattering for proteins in aqueous solutions is rather low because of the small difference in the scattering cross section of the solvent and the protein. The contrast can be varied, to some degree, by the addition of high concentrations of sucrose; however, the applicability of such an approach is limited, because that may alter the thermodynamics of the solution. Because the energy of X-rays useful for SAXS falls around 8–10 keV, radiation damage in the form of denaturation and aggregation of proteins is often a concern. In the case of SANS, the scattering power of atomic nuclei varies in an irregular manner from nucleus to nucleus. The most important aspect of neutron scattering that is useful for the study of biological macromolecules is the large difference in the neutron scattering cross section of hydrogen and deuterium. Therefore, the large contrast available for proteins in D_2O solutions makes SANS very effective even at the low-flux neutron sources. Contrast variation can be accomplished by changing the deuteration level in the solvent or solute, and this change may not significantly alter the thermodynamics of the system. Radiation damage is much lower in the case of SANS because the energies of the neutrons fall in the range of 10–80 meV, even though obtaining

SANS data requires more exposure time. The SANS method can therefore only provide time- and space-averaged low-resolution data and cannot probe the dynamic nature of the interaction. Notwithstanding these limitations, SANS is still a very useful tool for the determination of the size, shape, conformational change, and the molecular weight of macromolecular systems. Hence, we employed SANS (without contrast variation) rather than SAXS for the present study.

Molecular chaperonins 60 (cpn60) have been identified as essential constituents of cells [8]. Chaperonins protect cellular proteins against stress conditions that cause protein denaturation, such as heat shock, and poisoning with alcohols, free radicals and metals. Furthermore, chaperonins help the newly synthesized polypeptides to fold [8–10]. The folding process appears to be quite dynamic and may involve many rounds of protein binding and release [10]. The 2.8 Å crystal structure of GroEL, a cpn60 from *Escherichia coli*, has shown that the 57.4 kDa subunits are assembled into two seven-membered rings that stack back-to-back [11]. Each subunit is composed of three domains that are loosely connected by antiparallel strands. In the crystal structure, about 5% of the amino acid residues comprising N- and C-terminal sequences have not been resolved, presumably due to their disorder. The seven-membered ring has an outer diameter of 137 Å and combines with a second ring to form a cylindrical particle with a length of 146 Å. Each ring has a cavity (volume ~125×10^3 Å^3) with a circular mouth 45 Å in diameter [11]. The presence of this cavity was suggested by electron micrograph (EM) reconstructions [12–14] and was later confirmed by X-ray crystallography. The cavity was implicated in protein binding and folding [13–16]; however, the cavity is not big enough to accommodate large multidomain proteins. Calculation of the volumes of the cavity suggests that only proteins with molecular weights up to 72 kDa (assuming densities found typically in protein crystals) can be accommodated inside a single-ring cavity [11], but it is not understood how the unfolded proteins are directed into the cavity.

The SANS measurements were performed to determine the solution structures of GroEL chaperonin and the GroEL–rhodanese complex in D_2O. Rhodanese was selected as a substrate because its folding intermediate is known to bind to GroEL [17,18]. The GroEL–rhodanese complex is well characterized and is stable for extended periods of time at low temperature [17–19]. Rhodanese can be released from GroEL only in the presence of ATP and GroES [19]. Thus, the rhodanese should remain bound to GroEL during SANS data collection. These data were systematically analyzed and modeled by using the crystal coordinates of GroEL, and the structural features of both the native GroEL and its complex with rhodanese in solution were determined.

Results and discussion

Chaperonin structure in solution

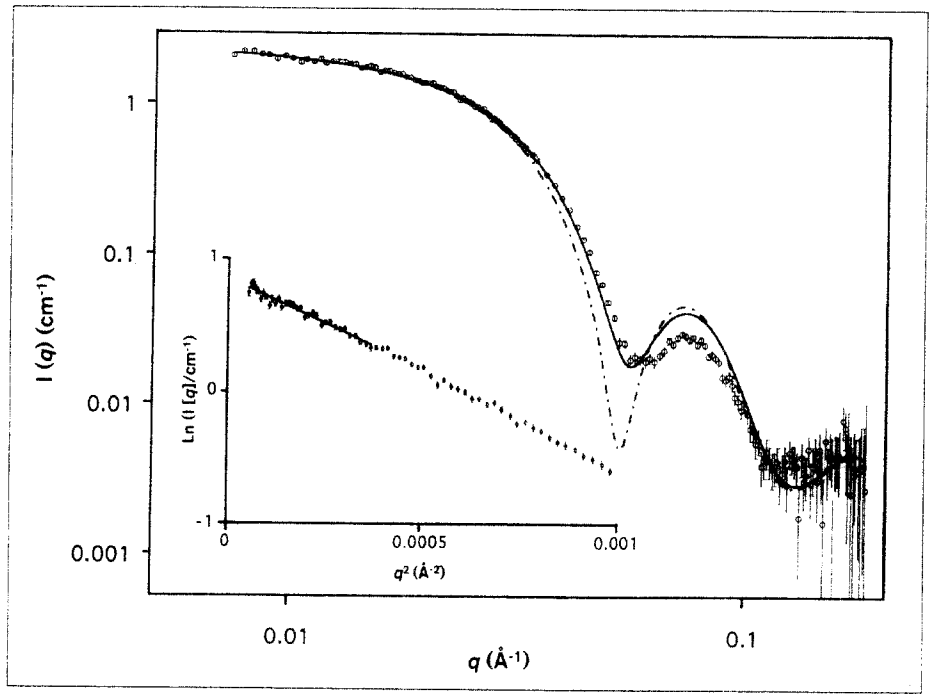
The SANS experiments on GroEL chaperonin at two concentrations (in the dilute regime) in buffered 98.5% D_2O solutions were measured at the 30 m SANS instrument at the High Flux Isotope Reactor at Oak Ridge National Laboratory (see the Materials and methods section). Because the radius of gyration (R_g) remained the same at both concentrations, and because the differential scattering cross section at momentum transfer ($q\rightarrow0$), $d\Sigma(0)/d\Omega$, scaled linearly with the concentration within the experimental error, interparticle effects are insignificant in this concentration range. Hence, the SANS data are presented only for the sample with the higher GroEL concentration (4.57 mg ml^{-1}; 5.5 μM) in Figure 1. In order to cover a wide q range ($4\pi\sin\theta/\lambda$, where θ is half of the Bragg angle, and λ [the neutron wavelength]=4.75 Å) measurements were made at two sample-to-detector distances (12 m and 3 m). The modeling of the SANS data of GroEL employed the available crystal coordinates at 2.8 Å resolution. The Guinier analysis of the SANS data for the 4.57 mg ml^{-1} GroEL (Fig. 1 insert) yielded an R_g of 63.2±0.8 Å and $d\Sigma(0)/d\Omega$ of 2.73±0.02 cm^{-1}. The R_g value agrees well with the calculated R_g of 63.63 Å from the crystal coordinates [11]. The extrapolated $d\Sigma(0)/d\Omega$ value corresponds to a molecular weight of 663.3 kDa (as calculated using equation 1), which is about 82.6% of the true molecular weight of the tetradecamer (803.3 kDa). (The significance of the lower value of the molecular weight from SANS data of the GroEL complex will be discussed later.) The molecular weight (M) of the scattering molecule is given by:

$$M = 1000 \times \frac{d\Sigma(0)}{d\Omega} \times \frac{N_A}{(\upsilon^2 C \Delta\rho^2)} \qquad (1)$$

where N_A is the Avogadro's number and using the $d\Sigma(0)/d\Omega$ value from the Guinier analysis, the protein concentration (C) of 4.57 mg ml^{-1}, a protein contrast ($\Delta\rho$) of 3.225×10^{10} cm^{-2} (determined from the amino acid composition by assuming that 90% of the labile hydrogen atoms exchanged with deuterium), and a partial specific volume (υ) of 0.722 cm^3 g^{-1} (calculated for the amino acid composition of GroEL). The important result here is that the molecular weight (weight-averaged) determined from the SANS data is close to the true molecular weight of GroEL.

Modeling of the SANS data of GroEL was done by employing the Monte Carlo method (described in the Materials and methods section) on the crystal coordinates for GroEL at 2.8 Å resolution. Note that the low q region is sensitive to the overall size of the particle, R_g; and the high q region is more sensitive to the local structures describing the relative position of the protein domains and the bound protein substrate, as well as the presence of equilibria between the complex and its single ring and

Figure 1



Comparison of experimental and calculated SANS curves for GroEL chaperonin. Key: (o) measured SANS data for 5.5 μM *E. coli* GroEL; (– · –) calculated SANS curve from the crystal data of GroEL; (–) calculated curve after smearing with a Gaussian resolution function with $\Delta q/q=1.77\times10^{-3}$ and $\Delta q/q=5.61\times10^{-3}$ for the sample-to-detector distances of 12 m and 3 m respectively. The insert is the Guinier plot of the measured data.

monomeric subunits. To evaluate the individual models, the integral discrepancy factor (IDF), defined as

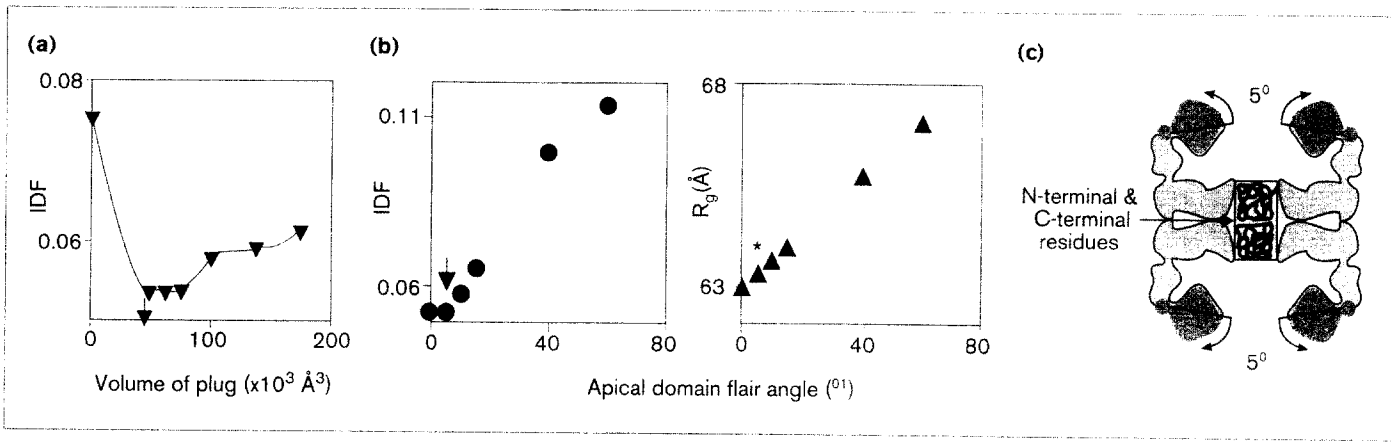
$$\mathrm{IDF}=\Sigma\,|\,I_e(q)-s\,I_m(q)\,|\,q^2\times\frac{dq}{I_e(q)\,q^2dq} \qquad (2)$$

was calculated [20]. Equation 2 uses the invariant of the scattering signal. This function weights the intensities by q^2 which becomes important in assessing the discrepancies in the high q region (describing the local structures), where the scattering signal decays exponentially. I_e and I_m are the experimental and calculated scattering intensities respectively and s is a scale factor proportional to the GroEL concentration. The model yielding the smallest IDF value, as well as an R_g value consistent with that from SANS data, was chosen to be the best. An important feature used in the modeling of the SANS data of GroEL is a peak at q=0.074 Å^{-1} (Fig. 1). Other factors that can affect the structure need to be carefully considered to account for this difference. Simultaneous fitting of both the low q region and the secondary peak places enormous constraints on the possible structural models for GroEL. Although the R_g value and the q value of the peak agree well for both the experimental and calculated differential scattering cross section, the amplitude of the peak does not (Fig. 1). Modeling of differential scattering cross section using the crystal data shows that the peak amplitude is sensitive to several factors including the location of the missing amino acid residues in the GroEL crystal, the tertiary conformation of GroEL chaperonin, and, to a smaller extent, the equilibrium governing the oligomeric state of the chaperonin.

Location of missing N- and C-terminal residues

Of the 547 amino acid residues in each monomer subunit, 5 N-terminal and 26 C-terminal residues of each subunit are not resolved in the crystal structure of GroEL, and they appear to be disordered [11]. These 434 missing residues in the crystal structure of GroEL account for 40.6 kDa (~5.1%) of the 803.3 kDa for the whole GroEL chaperonin. Cryo-EM studies have suggested that these missing residues may fill the cavity near the equator [13]. In the GroEL X-ray structure also, crystallographically disordered N- and C-terminal residues appear to project from equatorial domains into the central channel near the equator of the double-ring structure. In order to determine the location and packing density of the N- and C-terminal residues of GroEL in solution, a protein mass of 40.6 kDa (in the form of a cylinder or 'plug') was included at the equator of the hollow cylinder from the crystal structure. Modeling studies were then performed by varying the distribution of the missing amino acid residues. As described in the Materials and methods section, we kept the total number of points (based on the calculated partial specific volume of 0.700 cm^3 g^{-1} for these amino acid residues) the same when changing the volume of the plug. By so doing, we conserve the scattering cross section of these residues. The values of IDF and the volume of the plug from the calculated models for the plug configuration are shown in Figure 2a. The best agreement between the calculated and the experimental data (smallest IDF) was obtained when the missing amino acid residues were localized on both sides of the equator of the chaperonin double-ring structure. They were modeled in the form of a solid

Figure 2



Modeling of the GroEL structure in solution. **(a)** The missing mass at the N- and C-terminal regions of GroEL was added to the crystal data of GroEL as a cylinder with r=20 Å. The integral discrepancy factor (IDF) is shown as a function of the volume occupied by the cylinder. The zero volume point is for the GroEL with no plug, and the arrow points to the volume determined from the amino acid partial specific volume. For further modeling, a cylinder with r=20 Å and h=40 Å located at the equator was selected. **(b)** All 14 apical domains were rotated outward for each subunit by using a center of rotation defined by h=58 Å from the equator and r=65 Å. The IDF and radius of gyration (R_g) values are given as a function of the flair angle. The arrow points to the IDF local minimum and the asterisk to the R_g value (63.2±0.8 Å) determined from SANS data. The best fit corresponds to a flair angle of 5°. **(c)** Cross section of GroEL indicating the plug and apical domains where modifications to the structure (indicated by arrows) were introduced.

cylinder filling the available space in the cavity with a volume very close to that expected from the partial specific volume consideration. Changing either the shape or the packing density of this region resulted in a poorer fit to the experimental data (increasing IDF values in Fig. 2a). As expected, these changes affect R_g only marginally (data not shown). Thus, the N- and C-terminal residues of all 14 subunits of GroEL seem to condense near the equator, presumably preventing the exchange of protein substrates between the individual rings.

The N-terminal residues appear to play an important role in the stability of the chaperonin (residues Ala2 and Lys3 are highly conserved in the GroEL family) and therefore are likely to be located near the equatorial plane of the chaperonin complex [21,22]. In contrast, the sequence homology of the C-terminal region of chaperonins is low, and its sequence suggests neither α nor β structure. It is likely that the role of these residues in chaperonin function may be of limited importance due to the fact that removal of 16 C-terminal residues does not affect the function of GroEL [23,24].

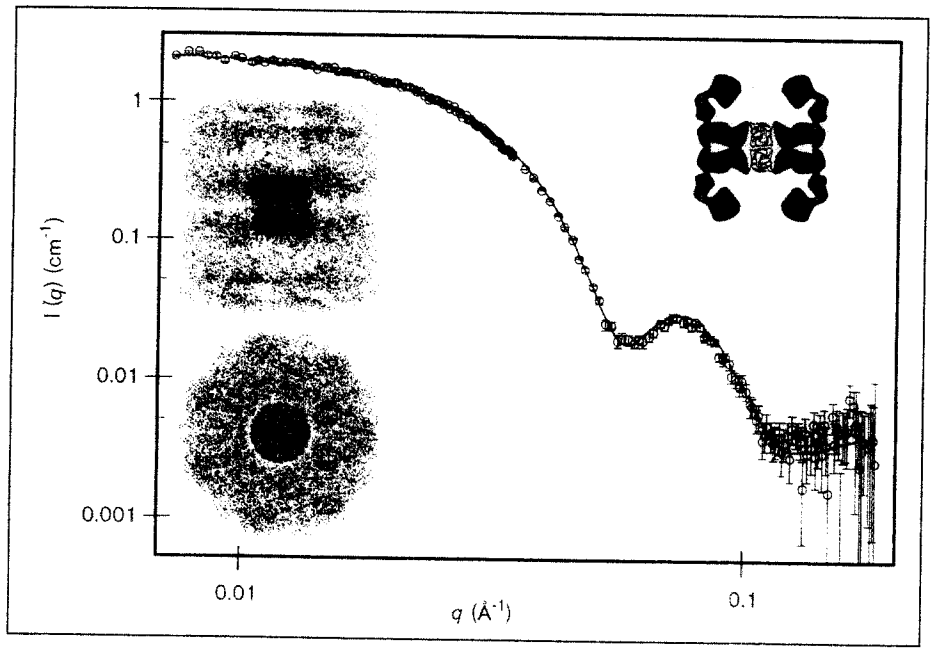
Orientation of apical domains

It has been reported that GroEL shows significant flexibility and plasticity in its structure [25,26]. Also, the X-ray structure suggested that the apical domains might exhibit two kinds of flexibility: *en bloc* movement generated by a hinge-like motion at its junction with the intermediate domain; and a local or segmental flexibility within the apical domain [11]. In addition, recent steady-state fluorescence polarization studies have shown that the two motion components in GroEL, namely, the slow and the fast, can account for the fluorescence depolarization [26]. The authors concluded that the presence of the fast motion suggests that cpn60 is not a rigid protein. The movement of a large part of the protein, *en bloc*, can alter the R_g and this alteration can easily be detected by SANS. We have systematically analyzed the possibility of closing and opening the mouth to the cavity by altering the orientation of the apical domains (0–60°) that are located near the poles of chaperonin [11] (Fig. 2b). Changes in R_g and IDF were used to validate the models. Figure 2b shows that when the apical domains were moved angularly from 0–10°, IDF shows a shallow minimum around 5° and increases sharply for higher flair angles, while R_g increases monotonically with increasing angular movement of the apical domains. The point at which the calculated R_g value is consistent with that from SANS data and the lowest IDF value occurs (0.0530), corresponds to the conformation in which the apical domains are moved outward by 5° (Fig. 2b). Because SANS data are time-averaged, it is likely that the apical domains may be displaced outward by 5° with respect to their positions in the crystal. This angular displacement of the apical domains results in an increase of about 2.7 Å in the diameter at the cavity opening (Fig. 2c), and this increase may have implications for binding of larger proteins.

Excellent agreement between the experimental and calculated data (IDF=0.050) is observed when the molar ratio of 0.9:0.025:0.075 for the equilibria between the double rings, single rings and monomers, respectively, is introduced (Fig. 3). This modeling was done by using the

Figure 3

The best fit of the measured SANS data for GroEL, giving due consideration to the plug location and volume, flairing of apical domains, and the existence of equilibria between the chaperonin and its subunits. Inserts on the left are side and top views of GroEL indicating the location and the shape of the plug. The icon on the right represents a cross section of GroEL with the plug.



calculated scattering intensities for the double ring, single ring, and a single subunit. Because we know the number of GroEL molecules from its concentration, the SANS data for the final state can be calculated by first multiplying the scattering intensities of the above individual systems by their number and then summing them on the basis of their molar ratios. The contribution of single rings and subunits to scattering at low q is very small. Only the high q region of the calculated scattering curve is sensitive to the presence of the three systems considered. The best estimate of the molar ratios was reached by looking for agreement between the calculated and experimental scattering curves, using IDF as an indicator. Furthermore, the R_g and the molecular weight obtained from the experimentally determined $d\Sigma(0)/d\Omega$ are also consistent with the previous equilibria. Given that only 90% of the GroEL is in the double-ring configuration (concentration=4.11 mg ml^{-1}), recalculation of the molecular weight from the SANS data yields a value of 738 kDa; and this value is in close agreement with the expected molecular weight of 803 kDa. Moreover, the calculated $d\Sigma(0)/d\Omega$ value for the systems in equilibrium is 3.04 cm^{-1}, which is about 10% larger than the experimental $d\Sigma(0)/d\Omega$ value of 2.73 cm^{-1}. The lower than expected values of $d\Sigma(0)/d\Omega$ and molecular weight for GroEL may be due to the uncertainty in the absolute calibration of the SANS data (~10%), coupled with hydration effects of the protein by the D_2O solvent — an effect that is very difficult to quantify, but is supposed to decrease the protein contrast. Observed equilibria are consistent with the recent reports on the presence of single rings and monomers in GroEL solutions, as well as in preparations of related bacterial chaperonins [27–31] and an archaeosome (a cpn60 from archaea [32]).

The solution structure of *E. coli* GroEL from synchrotron SAXS has been reported recently [33]. The major conclusions from this study are that the GroEL exists as a cylinder with an R_g of 66.2 Å (radius=68 Å and length=150.7 Å) and has no solvent cavities in solution. These results contradict the observations from crystallographic and cryo-EM studies [11,13]. The crystal structure of GroEL shows that it is assembled as a hollow cylinder (radius=68.5 Å and length=146 Å) with two large solvent cavities. The present study agrees well with the crystal structure data. The discrepancies between the SAXS study [33] and the present SANS study may be explained in the following way. First, the determination of R_g from the SAXS data was done in a q region of the SAXS data that is above the Guinier region (q R_g ~1) for the measured size (R_g=66.2 Å). If the same q region as the SAXS study [33] is used for the Guinier analysis, our SANS data on GroEL give a similarly large R_g value. The volume of GroEL (solid cylinder) from the SAXS study [33] is 219×10^3 $Å^3$, which is more than twice the size expected (96.4×10^3 $Å^3$) from the calculated partial specific volume of GroEL (v=0.722 cm^3 g^{-1}). Thus, the solvent-excluded volume predicted by SAXS for the solid cylindrical model is quite large, and the only way to obtain a reasonable value (based on partial specific volume) is by having large cavities inside the cylindrical volume. The modeling of our SANS data shows that filling the cavities of the double-ring GroEL complex with protein-like density leads to poor agreement with the experimental data. The present study

clearly demonstrates the presence of two independent solvent cavities in the double-ring structure of GroEL chaperonin [11,13] and these two cavities are separated by a plug, preventing any exchange between the cavities.

Structure of the GroEL–rhodanese complex

The SANS data are sensitive to the conformation of chaperonin (as shown previously) and the location of bound substrate [2]. Bovine rhodanese is a mitochondrial protein of 33.8 kDa (calculated v=0.719 cm^3 g^{-1}). Rhodanese becomes insoluble when denatured and the aggregates can be solubilized in the presence of GroEL [17–19]. We chose this protein as a substrate because refolding of rhodanese, in the absence of chaperonin 10 and ATP, is very slow at 15°C [17,18]; and, hence, the folding intermediate will remain bound to GroEL during the SANS experiment. Fluorescence polarization studies have shown that GroEL is very flexible in solution. However, binding of the rhodanese folding intermediate causes freezing of the local motion in GroEL. This observation suggests that the structure of cpn60 can be stabilized in binary complexes with protein substrates [26]. The models for GroEL–rhodanese complexes were chosen on the basis of previous suggestions by a number of research groups [13–15,17,34].

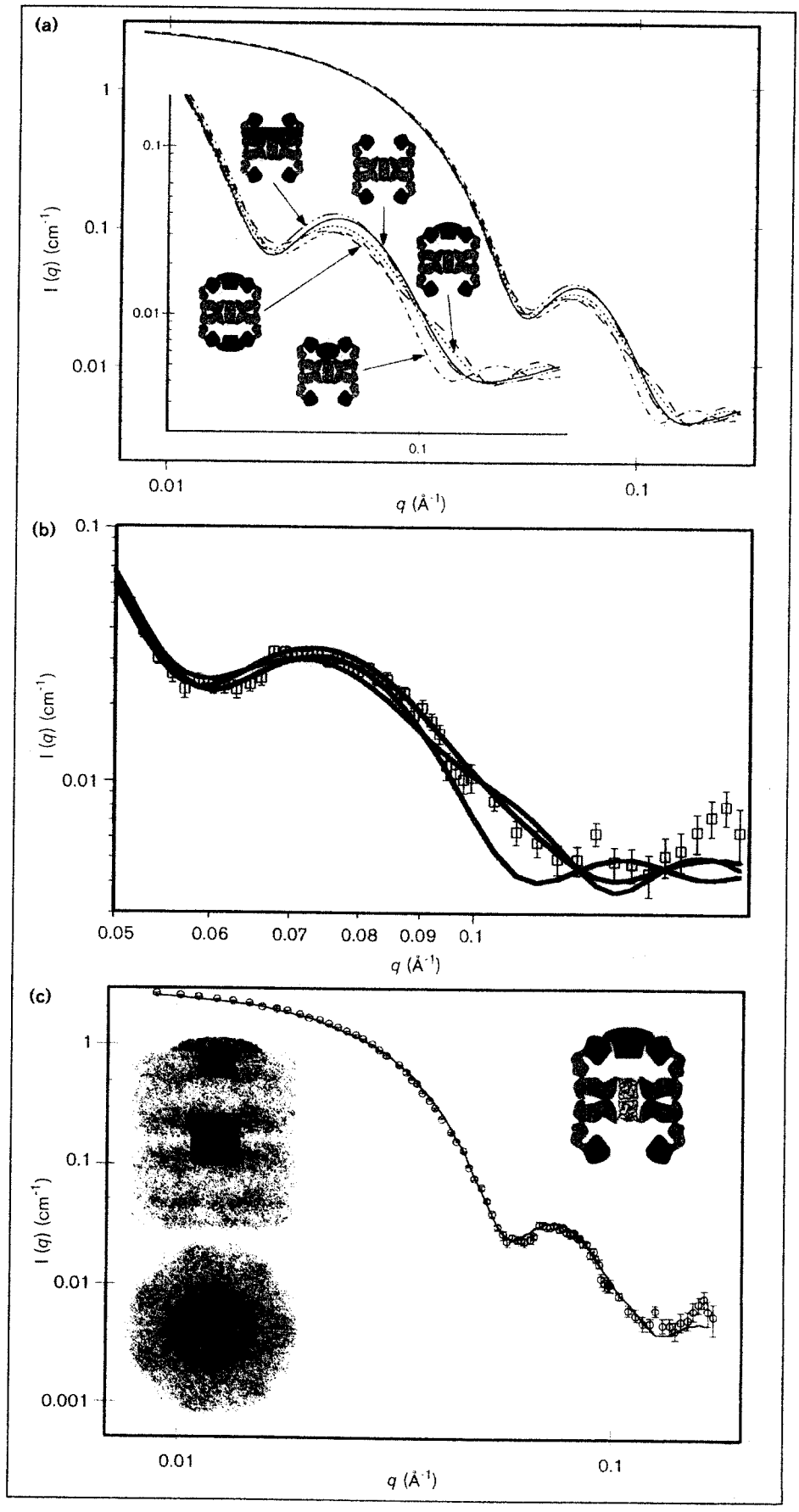
The SANS data measured for the GroEL–rhodanese complex (GroEL concentration of 5.5 μM, saturated with rhodanese) were compared with those for GroEL and with GroEL–rhodanese complex models (Fig. 4a). The difference between the measured values of R_g for GroEL (63.2±0.8 Å) and for the GroEL–rhodanese complex (64.3±0.5 Å) is small. The $d\Sigma(0)/d\Omega$ from the SANS data for the complex is 2.875±0.03 cm^{-1}, which corresponds to a molecular weight of 764 kDa (91.3% of that expected for the GroEL–rhodanese complex); and this is larger than that for the free GroEL by ~26 kDa. It is important to point out that parameters such as the concentration, partial specific volume, and the uncertainty in the bound D_2O reduce the accuracy of the molecular weight determination from the SANS experiments. However, when the best estimates of these parameters are used, SANS can yield information on the molecular weight within an uncertainty of about 10%. The interesting observation here is that the molecular weight of the GroEL–rhodanese complex is greater than that for GroEL, and the difference is very close to the molecular weight of one rhodanese molecule. The binding of a small protein to a large chaperonin has increased the R_g value of the complex by only ~1 Å, which suggests that the substrate binding did not cause any large conformational change in GroEL. On the basis of this small change in the R_g value, we assumed that the GroEL in the complexed state has an equilibrium identical to that for free GroEL.

Comparing the models of the GroEL–rhodanese complex

In the present study, we evaluated four different models suggested previously for the GroEL–protein complexes: first, protein substrate bound within the ring cavity; second, protein substrate bound to the outside surface of chaperonin; third, protein substrate bound at the top opening of chaperonin; and fourth, two protein substrates bound to chaperonin. In our modeling, rhodanese was allowed to assume conformations such as ellipsoid, ring, and 'champagne cork' (cylinder+semi-ellipsoid). Because of the dominance in size of GroEL over the rhodanese folding intermediate, the scattering curves for the chosen models depend much less on the shape of rhodanese and more on its average position with respect to GroEL. The important conclusion from these modeling experiments is that the larger differences between the SANS data for GroEL and the GroEL–rhodanese complex are observed only in the high q region, and the low q region is insensitive to the location of bound rhodanese (Fig. 4a). For each model, the R_g and the IDF [20] with the SANS data in the whole q region were computed.

The small but significant increases in R_g and $d\Sigma(0)/d\Omega$ for the GroEL–rhodanese complex, when compared with those for GroEL, indicate that the GroEL–rhodanese complexes should be present in solution. In fact, GroEL alone fits poorly to the GroEL–rhodanese data (IDF=0.070). The small increase in R_g also suggests that the rhodanese is bound within the confines of GroEL but not outside. This eliminates the model wherein the rhodanese is bound to the outside surface of chaperonin (IDF=0.071). The $d\Sigma(0)/d\Omega$ for the GroEL–rhodanese complex yields an increase in molecular weight of 26 kDa over that for free GroEL. This small increase indicates that probably only one rhodanese molecule is bound to GroEL. Therefore, the two-molecule protein substrate model can be eliminated (IDF=0.057). Among the remaining three models, the champagne cork and the ellipsoid models for the rhodanese bound across one end of GroEL are indistinguishable in the measured q region. They both agree with the measured SANS data equally well (IDF=0.051). On the other hand, a model in which rhodanese is placed as an ellipsoid inside the cavity, agrees poorly with the SANS data (IDF=0.073) (Fig. 4a,b). In fact, this model fits the SANS data the worst of all the tested models. A more detailed comparison of the ellipsoidal and champagne cork models with our SANS data on the GroEL–rhodanese complex in the q region 0.05–0.15 $Å^{-1}$ (Figs. 4b,c) indicates that the protein substrate is probably bound at the mouth (IDF=0.051) rather than inside the cavity of GroEL (IDF=0.073). Our observations show that a small difference in the IDF values is still sufficient for discriminating between the models. An inspection of Figure 4b shows that the deviation from the experimental data at q=0.06–0.12 $Å^{-1}$ is larger for the two-protein substrate model (red line) (IDF=0.057) than that for the one-substrate model (green line) (IDF=0.051). In the case of the two-substrate model, it is seen that the minimum in the intensity at q=0.06 $Å^{-1}$ is slightly larger than that for the

Figure 4



SANS curves of the GroEL–rhodanese complex. **(a)** SANS curves calculated for free GroEL (continuous line) (IDF=0.070) and several theoretical models of the GroEL–rhodanese complex are compared with the experimental data for the GroEL–rhodanese complex. (The insert shows an enlargement of the high q region where the larger differences are observed.) The key is as follows: (i) an ellipsoid inside the cavity (–·–) (IDF=0.073); (ii) a ring located outside GroEL (–···–) (IDF=0.071); (iii) a 'champagne cork' bound at the mouth of GroEL (···) (IDF=0.051); (iv) two 'champagne corks' bound to the ends of GroEL (– – –) (IDF=0.057; (v) an ellipsoid bound at the mouth of GroEL (IDF=0.051) (data not shown because the scattering curve is virtually identical to the single 'champagne cork' model). The champagne cork was created by combining a cylinder with r=20 Å and h=20 Å, and a semi-ellipsoid with semi-axes 40×40×5 Å, making up a total volume of 41.9×10^3 Å^3. **(b)** Comparison of the experimental data in the high q region with the best fit 'champagne cork' bound at the mouth of GroEL (green line), two 'champagne corks' bound to the ends of GroEL (red line) and an ellipsoid inside the cavity (blue line). **(c)** Fit of the experimental SANS data with the rhodanese substrate in the form of a 'champagne cork' model. Inserts on the left are side and top views of the GroEL–rhodanese complex showing the location (at z=60 Å from the equator) and the shape of the bound rhodanese. The icon on the right represents a cross section of the complex.

one-substrate model and the secondary peak is more asymmetric (with a shoulder at $q>0.1$ Å^{-1}) for the two-substrate model than that for the one-substrate model.

Thus, only one rhodanese molecule appears to bind to the chaperonin, presumably allowing the cpn10 co-chaperonin to bind on the other side of the cylinder. From our data, it is not clear why the GroEL double-ring structure binds only a single protein substrate (even with a large molar excess of substrate), but our data are consistent with the earlier observations of Mendoza *et al.* [19], that only one rhodanese molecule binds to GroEL. Perhaps binding of the folding intermediate introduces subtle changes in the conformation of chaperonin that are below the detection limits of SANS but are sufficient to prevent binding of a second rhodanese molecule to GroEL.

Of the two models that best fit the experimental data of the GroEL–rhodanese complex, the champagne cork model appears to agree better with the mutational [16] and EM data [13]. Figure 4c shows the SANS data for the GroEL–rhodanese complex in solution, along with the model that best fits the data. In this model, rhodanese appears to be extensively and exclusively bound to the apical domains of GroEL. These domains have been implicated in protein binding and folding by mutational analysis [16]. The rhodanese location and the high-affinity interaction suggest an extensive network of contacts between the rhodanese and the apical domains of GroEL at the opening to the cavity. In the GroEL–rhodanese complex, the apical domains appear to remain flared by 5° (the IDF increases when apical domains are moved in or out of this position). It is likely that binding of larger proteins with GroEL may involve more distortion in this region. Modeling studies also revealed that the N- and C-terminal regions of GroEL remain condensed near the equator in the presence of bound protein substrate and both cavities remain filled with solvent. The rhodanese folding intermediate seems to spread across the GroEL opening and assumes a globular shape (champagne cork or ellipsoid-like) that may have 'molten globule'-like properties, as suggested for protein folding intermediates [17,19,35]. Thus, GroEL appears to provide a large active surface for the binding of unfolded proteins. Our model imposes constraints on the mechanism of chaperonin-mediated protein folding and assembly of multisubunit complexes. The model also implies that the rhodanese folding intermediate is significantly exposed to the bulk solvent and multiple rounds of protein binding and release are easily envisaged. Our data are also consistent with the report that the folding intermediate of 3-isopropylmalate dehydrogenase binds to one axial end of the chaperonin from *Thermus thermophilus* [36]. This enzyme folding intermediate was bound to one end of holo-chaperonin and was readily accessible to a specific antibody. Earlier reports indicated that the stable folding intermediates (including rhodanese) are sensitive to proteolysis when bound to chaperonins, suggesting surface rather than cavity binding [17,19,35]. Our results are consistent with the report that rhodanese in the GroEL-bound state remains accessible to proteases [17], but our data differ with the same study's suggestion that such rhodanese is bound inside the cavity.

In summary, our data clearly indicate that the protein substrate is much more exposed to the solvent than previously thought. Our data are highly consistent with the EM reconstructions of the GroEL–malate dehydrogenase complex [13] and GroEL mutational studies [16], and do not support those models in which the protein substrate is either bound inside the cavity [15,17] or outside the GroEL cylinder [34]. Furthermore, it is unlikely that the protein substrate can cross the equatorial plane and exchange freely between chaperonin cavities.

Biological implications

Chaperonins 60 (cpn60) of which GroEL from *Escherichia coli* is an example, help to fold proteins *in vivo* and *in vitro* into their functional three-dimensional structures. These large oligomeric complexes first recognize and bind to the unfolded protein. Transformation of the protein folding intermediate into its native conformation requires the cooperation of cpn60 and cpn10 (e.g. GroES) chaperonins as well as ATP hydrolysis. It has been suggested that the cpn60 chaperonin's cavity provides an 'infinite dilution' environment for the unfolded protein, thus preventing any non-productive interaction between the different unfolded polypeptides during their folding. Studies to date have produced conflicting results with respect to the location of the protein substrate: some show the substrate bound inside the cavity of the chaperonin and others suggest that it is bound at the pole of the chaperonin.

In order to understand the functional role of chaperonins in protein folding it is essential to know their structure. X-ray crystallography has provided the detailed atomic structure of *E. coli* GroEL and cryo-electron microscopy determined the lower-resolution structure of GroEL and its complexes with ATP and proteins. Our objective was to determine the structure of GroEL and its complex with a protein folding intermediate in solution. We carried out small-angle neutron scattering (SANS) studies on GroEL and a GroEL–rhodanese complex. The SANS data of GroEL were compared with the scattering data calculated from the coordinates of the GroEL crystal structure. The present study reveals that the solution structure of GroEL from SANS is similar, but not identical to that in the crystal. The differences between the crystal and solution structures allow the extraction of additional structural information. The

location and packing density of the residues at the N and C termini, which appear disordered and are therefore missing in the crystal structure, have been determined. These N- and C-terminal regions seem to be condensed near the equator of GroEL. While N-terminal residues may confer specificity to form oligomeric chaperonin, the C-terminal residues may play a structural role in blocking the transfer of unfolded proteins between the cavities of GroEL. This study also shows that the apical domains in the solution structure are projected outward slightly, as compared with the crystal. The flexibility of the apical domains and the larger opening of the mouth seen in the solution structure may enable GroEL to provide the necessary surface for the binding of denatured proteins of different sizes.

The intriguing question that carries important mechanistic implications is the location of the binding site for the unfolded protein in the chaperonin. If the unfolded protein is bound inside the chaperonin cavity, then the folding process could proceed in an infinite dilution environment. However, if, as this study demonstrates, the folding intermediate binds across the opening of the cpn60 cavity (i.e. on the surface), the roles of the solvent and other cofactors also have to be considered in determining the mechanism of chaperone-mediated protein folding. With the folding intermediate bound on the surface of the chaperonin, assembly of oligomeric proteins can also be explained by the interaction of multiple chaperonin–protein complexes. However, the detailed explanation of the chaperonin–protein interaction must await high-resolution structural studies of chaperonin–protein complexes.

Materials and methods

Materials

E. coli GroEL was obtained from the overproducing strain provided by Dr A Horwich (Yale University). The cells were grown in 2×TY medium 16 g Trypton [Difco], 10 g yeast extract [Difco], 5 g NaCl per liter) and upon induction with 1 mM isopropylthiogalactoside (IPTG) large quantities of GroEL and GroES were produced. GroEL was purified in three steps. GroEL was chromatographed on a Q Sepharose Fast Flow column (Pharmacia Biotech, Inc., Piscataway, NJ) and eluted with a linear (0–500 mM) NaCl gradient. Chaperonin fractions were concentrated by using Filtron 300 (Filtron Technology Corp., Northborough, MA) and were purified by gel permeation chromatography on a Sephacryl S-300 column (Pharmacia). This step was followed by separation on a high-resolution MonoQ (16/10) column (Pharmacia), and GroEL was eluted with a linear (100–350 mM) NaCl gradient. GroEL was more than 99% pure, as determined from overloaded polyacrylamide gel electrophoresis under denaturing conditions and silver staining. For short periods, 0.25 mM GroEL was stored at +4°C in 20 mM Tris-HCl (pH 7.5), 1 mM DTT, and 100 mM NaCl. Alternatively, 0.25 mM GroEL was stored with 30% glycerol at −25°C. Prior to data collection, the chaperonin complex was purified rapidly (8 min) on a Superdex 200 column (Pharmacia) by using FPLC and was concentrated on a Centricon 100 membrane (Amicon, Inc., Beverly, MA) in D_2O buffered with 40 mM Na/K phosphate (pH 7.0) and 1 mM DTT. The GroEL concentration was determined spectrophotometrically from its molar extinction coefficient. The GroEL–rhodanese complex was prepared in the following way: 6 M guanidinium/HCl-denatured rhodanese was added in 7.5-fold molar excess to GroEL chaperonin. Insoluble rhodanese was removed by centrifugation, and the complex was purified rapidly by gel permeation chromatography on a Superdex 200 column by using FPLC, in a similar way to that described by Holdan *et al.* [17]. The complex was concentrated in D_2O as described previously. Denaturing polyacrylamide gels confirmed that the purified preparation contained the GroEL–rhodanese complex.

SANS measurements and data processing

SANS data were measured at 15°C for 5.5 μM (0.47%) and 2.9 μM (0.23%) solutions of GroEL in 98.5% D_2O buffered with 40 mM Na/K phosphate (pH 7.0) and 1 mM DTT, at the WC Koehler 30 m SANS facility at the High Flux Isotope Reactor at Oak Ridge National Laboratory. The effects of interparticle interaction in this concentration regime were insignificant as the values of R_g were the same at both of these concentrations, and the $d\Sigma(0)/d\Omega$ scaled linearly with the concentration. Because the concentrations considered here are in a high dilution regime for SANS, we show the data only for the higher concentration (5.5 μM). The neutron wavelength was 4.75 Å ($\Delta\lambda/\lambda$ ~5%), and measurements were made at two sample-to-detector distances, 12 m (4 h) and 3 m (1.5 h) to obtain sufficient statistical precision. For the 12 m sample-to-detector distance, the sizes of the source and sample slits (irises) were 3 cm and 1 cm, respectively; and they were separated by a distance of 7.5 m. For the 3 m sample-to-detector distance, a source slit of 4.5 cm was used; and the distance between the source and sample slits was 3.5 m. Samples were contained in quartz cells with a path length of 0.5 cm. The data were corrected for the sample transmission, detector efficiency on a cell-by-cell basis, and for the backgrounds from the instrument, the quartz cell, and the solvent, prior to radial averaging. Net intensities were converted to an absolute differential scattering cross section per unit sample volume ($d\Sigma(q)/d\Omega$ in units of cm^{-1}) by comparison with precalibrated secondary standards [37]. The incoherent background for each sample was determined on the basis of the number density of hydrogens and was subtracted as a flat background prior to modeling.

Modeling

A Monte Carlo method was used to model the proposed structures with the measured SANS data (S Henderson, unpublished data). The crystallographic coordinates provide the location in space and the scattering power of the elements of GroEL. The solvent-excluded volumes of the amino acid residues are calculated from the van der Waals radii of the atoms and their bond lengths. Random points in space are chosen around the center of the mass of each amino acid, where the number of such points for each amino acid is proportional to the product of the amino acid volume and its scattering length density difference with respect to the solvent. Calculation of all pairwise distances between such points and binning these as a function of distance generates the pair distribution function P(r) as a function of r, and this function is then readily transformed to $d\Sigma(q)/d\Omega$ *versus* q. The instrumental smearing function for the SANS instrument (due to neutron wavelength spread, beam size, and detector pixel size) is then applied to the calculated $d\Sigma(q)/d\Omega$ to produce the smeared model intensities, and compared with the measured SANS intensities per molecule. This Monte Carlo approach to evaluate model intensities from protein crystal coordinate files allows easy extension to the study of bound proteins that have crystal coordinates. Modeling is still possible even if such crystal coordinates are not available for proteins, such as for the rhodanese folding intermediate. This is due to the fact that the partial specific volume for proteins varies between 0.70 and 0.74 $cm^3\ g^{-1}$ and can be calculated from the amino acid composition. Hence, the volume occupied by a protein is proportional to its molecular weight. For the rhodanese folding intermediate, we calculated the partial specific volume from its amino acid composition (0.719 $cm^3\ g^{-1}$), and this calculation leads to its volume (V) by using the relationship: $V\ (Å^3)=0.719\times10^{24}\times Mwt\ (Da)/6.02\times10^{23}$. A shape,

orientation, and position of this volume can be selected with respect to the GroEL structure. This volume is then filled with k random points such that the density of points (k/V) is the same as the average random point density (PD) in the GroEL structure. While modeling with different shapes for the bound protein, the point density of the bound protein is adjusted such that the total number of points in the volume is always conserved. The previously described procedure of calculating the P(r) function, followed by conversion to $d\Sigma(q)/d\Omega$, is then repeated. Similarly, the distribution of N- and C-terminal residues within GroEL was modeled. The volume occupied by these amino acid residues is determined by using the calculated partial specific volume of 0.700 $cm^3\ g^{-1}$ (based on the amino acid composition) and their total molecular weight. In the modeling, this volume (Vp) was randomly filled with the total number of points ($Np=Vp\times PD$) and was allowed to change its volume and point density such that Np is always conserved.

Acknowledgments

We thank Arthur Horwich of Yale University for providing the *E. coli* strain that overexpresses GroEL. We also thank Jonathan Trent of Argonne National Laboratory for enticing discussions prior to the beginning of this project. This project was supported by the US Department of Energy Office of Health and Environment Research and BES–Materials Science, under Contract No. W-31-109-Eng-38 with the University of Chicago, as well as by the Director's R&D Fund, Oak Ridge National Laboratory under Contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

References

1. Clore, G.M. & Gronenborn, A.M. (1994). Structure of larger proteins, protein–ligand and protein–DNA complexes by multidimensional heteronuclear NMR. *Protein Sci.* **3**, 372–390.
2. Knott, R., Hansen, S. & Henderson, S.J. (1994). A small angle X-ray scattering study of the binding of cyclosporin-A to calmodulin. *J. Struct. Biol.* **112**, 192–198.
3. Schiffer, M., Stevens, F.J., Westholm, F.A., Kim, S.S. & Carlson, R.D. (1982). Small angle neutron scattering study of Bence–Jones protein Mcg: comparison of structures in solution and in crystal. *Biochemistry* **21**, 2872–2878.
4. Svergun, D.I., Koch, M.H.J., Pedersen, J.S. & Serdyuk, I.N. (1994). Structural model of the 50S subunit of *Escherichia coli* ribosomes from solution scattering. *J. Mol. Biol.* **240**, 78–86.
5. Lawson, C.L., Zhang, R.-g., Schevitz, R.W., Otwinowski, Z., Joachimiak, A. & Sigler, P.B. (1988). Flexibility of the DNA-binding domains of *trp* repressor. *Proteins* **3**, 18–31.
6. Staacke, D. Walter, B. Kisters-Woike, B. Wilcken-Bergmann, V.v. & Muller-Hill, B. (1990). How Trp repressor binds to its operator. *EMBO J.* **9**, 1963–1967.
7. Haran, T.E., Joachimiak, A. & Sigler, P.B. (1992). The DNA target of the *trp* repressor. *EMBO J.* **11**, 3021–3030.
8. Fayet, O., Ziegelhoffer, T. & Georgopoulos, C. (1989). The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171**, 1379–1385.
9. Hendrick, J.P. & Hartl, F.-U. (1993). Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.* **62**, 349–384.
10. Weissman, J.S. Kashi, Y., Fenton, W.A. & Horwich, A.L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release by nonnative forms. *Cell* **78**, 693–702.
11. Braig, K., *et. al.*, & Sigler, P.B. (1994). The crystal structure of the bacterial GroEL at 2.8 Å. *Nature* **371**, 578–586.
12. Schmidt, M., *et. al.*, & Buchner, J. (1994). Symmetric complexes of GroEL chaperonins as part of the functional cycle. *Science* **265**, 656–659.
13. Chen, S., *et al.*, & Saibil, H. (1994). Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature* **371**, 261–264.
14. Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.-U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO. J.* **11**, 4757–4765.
15. Braig, K., Simon, M., Hainfield, J., Furuya, F. & Horwich, A.L. (1993). A polypeptide bound by the chaperonin GroEL is located within a central cavity. *Proc. Natl. Acad. Sci. USA* **90**, 3978–3982.
16. 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.
17. Holdan, R., Tempst, P. & Hartl, U. (1995). Binding of defined regions of a polypeptide to GroEL and its implications for chaperonin-mediated protein folding. *Nat. Struct. Biol.* **2**, 587–595.
18. Mendoza, J.A., Lorimer, G.H. & Horowitz, P.M. (1991). Intermediates in the chaperonin-assisted refolding of rhodanese are trapped at low temperature and show a small stoichiometry. *J. Biol. Chem.* **266**, 16973–16976.
19. Mendoza, J.A., Butler, M.C. & Horowitz, P.M. (1992). Characterization of a stable, reactivatable complex between chaperonin 60 and mitochondrial rhodanese. *J. Biol. Chem.* **267**, 24648–24654.
20. Feigin, L.A. & Svergun, P.I. (1987). *Structure Analysis by Small Angle X-ray and neutron scattering.* (Taylor, J.W., ed), p. 94, Plenum Press, NY.
21. Horovitz, A., Bochkareva, E.S. & Girshovich, A.S. (1993). The N terminus of the molecular chaperonin GroEL is a crucial structural element for its assembly. *J. Biol. Chem.* **268**, 9957–9959.
22. Horovitz, A., Bochkareva, E.S., Kovalenko, O. & Girshovich, A.S. (1993). Mutation Ala2→Ser destabilizes intersubunit interactions in the molecular chaperone GroEL. *J. Mol. Biol.* **231**, 58–64.
23. McLennan, N.F., Girshovich, A.S., Lissin, N.M., Charters, Y. & Masters, M. (1993). The strongly conserved carboxy-terminus glycine-methionine motif of the *Escherichia coli* GroEL chaperonin is dispensable. *Mol. Microbiol.* **7**, 49–58.
24. Burnett, P.B., Horwich, A.L. & Low, K.B. (1994). A carboxy-terminal deletion impairs the assembly of GroEL and confers a pleiotropic phenotype in *Escherichia coli* K-12. *J. Bacteriol.* **176**, 6980–6985.
25. Zahn, J., Harris, R., Pfeifer, G., Pluckthun, A. & Baumeister, W. (1993). Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *J. Mol. Biol.* **229**, 579–584.
26. Gorovits, B.M. & Horowitz, P.M. (1995). The molecular chaperonin cpn60 displays local flexibility that is reduced after binding with an unfolded protein. *J. Biol. Chem.* **270**, 13057–13062.
27. Taguchi, H., Makino, Y. & Yoshida, M. (1994). Monomeric chaperonin-60 and its 50-kDa fragment possess the ability to interact with non-native proteins, to suppress aggregation, and promote protein folding. *J. Biol. Chem.* **269**, 8529–8534.
28. Mendoza, J.A., Demler, B. & Horowitz, P.M. (1994). Alteration of the quaternary structure of cpn60 modulates chaperonin-assisted folding. *J. Biol. Chem.* **269**, 2447–2451.
29. Viitanen, P.V., *et al.*, & Cowan, N.J. (1992). Mammalian mitochondrial chaperonin 60 functions as a single toroidal ring. *J. Biol. Chem.* **267**, 695–698.
30. Lissin, N.M. (1995). *In vitro* dissociation and self-assembly of three chaperon 60s: the role of ATP. *FEBS Lett.* **361**, 55–60.
31. Ishii, N., Taguchi, H., Sasabe, H. & Yoshida, M. (1995). Equatorial split of holo-chaperonin from *Thermus thermophilus* by ATP and K^+. *FEBS Lett.* **362**, 121–125.
32. Quaite-Randall, E., Trent, J.D., Josephs, R. & Joachimiak, A. (1995). Conformational cycle of the archaeosome, a TCP1-like chaperonin from *Sulfolobus shibatae*. *J. Biol. Chem.* **270**, 1–6.
33. Igarashi, Y., *et al.*, & Kihara, H. (1995). Solution X-ray scattering study on the chaperonin GroEL from *Escherichia coli*. *Biophys. Chem.* **53**, 259–266.
34. Weiss, C. & Goloubinoff, P. (1995). A mutant at position 87 of the GroEL chaperonin is affected in protein binding and ATP hydrolysis. *J. Biol. Chem.* **270**, 13956–13960.
35. Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. & Hartl, F.-U. (1991). Chaperonin mediated protein folding at the surface of GroEL through a "molten-globule"-like intermediate. *Nature* **352**, 36–42.
36. Ishii, N., Taguchi, H., Sasabe, H. & Yoshida, M. (1994). Folding intermediate binds to the bottom of bullet-shaped holo-chaperonin and is readily accessible to antibody. *J. Mol. Biol.* **236**, 691–696.
37. Wignall, G.D. & Bates, F.S. (1986). Absolute calibration of small-angle neutron scattering data. *J. Appl. Cryst.* **20**, 28–40.