86547

5/2/95

ANL/CMB/PP--86547

Solution Structure of GroEL and Its Complex with Rhodanese from Small Angle Neutron Scattering

P. Thiyagarajan, Stephen J. Henderson* and Andrzej Joachimiak#

Argonne National Laboratory, 9700 S. Cass Ave, Argonne, IL 60439

* Oak Ridge National Laboratory, Oak Ridge, TN 37831

To whom correspondence should be addressed

ABSTRACT

Small angle neutron scattering measurements were made on *Echerichia coli* GroEL chaperonin and its complex with rhodanese in D_2O. GroEL contains 14 equivalent subunits arranged as two stacked seven-member rings. The radius of gyration from the experimental data agrees well with that obtained from the calculated form factor by using the coordinates of the x-ray structure of *E. coli* GroEL. The x-ray crystal structure does not include about 5% of the terminal amino acid residues so an equivalent protein volume was added inside the GroEL cavity and its location was determined. In addition, the best fit of the solution scattering data requires five degrees of flaring of the apical domains and the inclusion of small contributions from seven-subunit rings and monomers. The modeling of the SANS data for the GroEL/rhodanese complex supported a "champagne cork" model suggesting that rhodanese is bound across one opening of GroEL rather than inside its cavity.

INTRODUCTION

The relationship between the functional protein structure in solution and that determined by x-ray crystallography has been studied extensively (1-3). Large multisubunit assemblies offer special challenges (4). Small angle neutron scattering (SANS) can provide important lower resolution information on the structure of such assemblies and their interactions in solution. SANS can detect conformational change and substrate binding, and can provide information on the size, morphology and composition of macromolecular complexes. More detailed information on the structure of complexes can be obtained when the crystal coordinates are available for the major component.

Molecular chaperonins of hsp60 class have been identified as an essential constituent of cells (5). Chaperonins protect cellular proteins against stress conditions that cause denaturation. Furthermore, chaperonins help newly synthesized polypeptides to fold. In *E. coli,* the GroEL chaperonin exists as a tetradecamer. The 2.8-Å crystal structure of GroEL has shown that the 57.4-kD subunits are assembled into two 7-member rings that stack back to back (6). Each subunit is composed of three domains that are connected loosely by anti-parallel 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 7-member ring has an outer diameter of 137 Å, and it combines with a second ring to form a cylindrical particle 146 Å long. Each ring has a cavity (~125 $kÅ^3$) with a circular mouth 45 Å wide (6). The presence of this cavity was suggested from electron micrograph reconstructions (7-9) and was confirmed by x-ray crystallography.

The central cavity was implicated in protein binding (8-11). However, it may not be big enough to accommodate large multidomain proteins. Calculation of the cavity volume suggests that only proteins up to about 72 kD (assuming densities found typically in protein crystals) can be accommodated inside a single-ring cavity (6). Furthermore, it is not clear how proteins can be directed into the cavity.

SANS measurements in D_2O were made to determine the solution structure of GroEL chaperonin and the GroEL/rhodanese complex. These data were modeled using the crystal structure of GroEL to produce models for both native GroEL and its complex with rhodanese in solution.

Chaperonin Structure in Solution. The SANS data of GroEL chaperonin in buffered 98.5% D_2O solution were measured at the 30 m SANS instrument at the High Flux Isotope Reactor at Oak Ridge National Laboratory, and are shown in Figure 1. In order to cover sufficient q range ($4\pi \sin\theta/\lambda$, where θ is half the scattering angle and $\lambda = 4.75$ Å, the neutron wavelength) measurements were made at two different sample-to-detector distances (12 m and 3 m). The SANS data from GroEL is consistent with a hollow cylindrical structure. The radius of gyration (R_g) determined from the experimental data (Fig. 1a insert) is 63.2±0.8 Å, which agrees well with R_g=63.3 Å calculated from the crystal structure. An important feature for modeling in the SANS data of GroEL is a peak at q = 0.074 $Å^{-1}$. Although the R_g values and the position of the peak agree well for both, the experimental and calculated form factors, the amplitudes of the peak do not, without careful consideration of other factors. Modeling $d\Sigma/d\Omega(q)$ from the crystal data shows that the peak amplitude is sensitive to the location of the disorder in the crystal GroEL residues, conformation of the chaperonin, and the equilibria between chaperonin and its subunits (12).

In the crystal structure of GroEL, 5 N-terminal and 26 C-terminal residues of each 548 amino acid polypeptide are not seen in electron density and appear disordered (6). These residues seem to project into the solvent cavity near the equatorial plane of double toroid and account for 40.6 kD of mass (about 5.1 %) out of 803.3 kD for the whole GroEL chaperonin. In order to determine the distribution and the location of N- and C-terminal residues in GroEL, a number of models were tested (12). The best agreement between the calculated and the experimental data was obtained when the missing amino acid residues were localized on both sides of the equator of the chaperonin double-ring structure in the form of a solid cylinder filling the available cavity space with a specific volume common to globular proteins (1.2 A^3/Dalton) (Fig. 1b). Cryo-electron microscopic studies

have suggested the presence of such mass near the equator (8). Changing the shape or the extent (density) of this region resulted in a poorer fit to the experimental data. Thus, the N- and C-terminal sequences of GroEL appear to condense near the equator, presumably preventing the exchange of protein substrates between individual rings. The N-terminal residues appear to play an important role in the stability of the chaperonin complex (14-15), and residues A2 and K3 are highly conserved among the GroEL family. In contrast the sequence homology of the C-terminal region of chaperonins is low, its sequence suggest neither α nor β structure, and 16 C-terminal residues can be entirely removed without affecting the GroEL function (13), thus the role of these residues in chaperonin function may be of limited importance.

It has been reported that GroEL shows significant flexibility and plasticity in its structure (16). We have systematically analyzed the possibility of closing and opening the mouth of the cavity by altering the orientation of the apical domains (0 - 60 degrees) that are located near the poles of chaperonin (6) (Fig. 1b). The best fit resulted when the apical domains were moved outward by five degrees. Since SANS data are time-averaged, this suggests that on an average the apical domains are displaced five degrees from the crystal coordinate positions. This angular displacement corresponds to a translation of up to 3.9 Å and the mouth diameter increases by up to 2.7 Å (Fig. 1b).

Excellent agreement between the experimental and calculated data was seen when the molar ratios of 0.9/0.025/0.075 for the equilibria between the double rings/single rings/subunits respectively were introduced (Fig. 1c). The presence of single rings and monomers of GroEL and related chaperonins were reported previously (17-20). The chaperonin dissociation to monomers and self-assembly has been described recently for *E. coli* GroEL and two other homologous chaperonins (20), and for TCP-1-like chaperonin from *Sulfolobus shibatae* (21).

Structure of the Chaperonin/Rhodanese Complex. Bovine rhodanese is a 33.8-kD protein. Rhodanese becomes insoluble when denatured and the aggregates can be solubilized in the presence of GroEL (22-23). We chose this protein as a substrate because it will remain bound to chaperonin during the SANS experiments, since the refolding of rhodanese in the absence of GroES and ATP is very slow at 15 ^{o}C (22). The aim of our experiment was to distinguish between various protein-substrate binding modes that have been suggested for GroEL/protein complexes: i) protein-substrate bound within the ring cavity, ii) protein-substrate bound to the outside surface of chaperonin, iii) protein-

substrate bound on top of chaperonin, iv) two protein-substrates bound to chaperonin. SANS data are sensitive to the conformation of chaperonin (as shown above) and the location of bound substrate (2).

The SANS data measured for the GroEL/rhodanese complex were compared with those for GroEL and with GroEL/rhodanese complex models (Fig. 2). The difference between the measured R_g values of GroEL (63.2±0.8 Å) and the GroEL/rhodanese complex (64.3±0.5 Å) is small, and the R_g for the model that best fits the experimental data for GroEL/rhodanese complex is 64.2 Å (Fig. 2). The binding of a small protein to a large chaperonin has increased the R_g value of the complex by about 1 Å, which suggests that the substrate binding did not cause any large conformational change in GroEL, and it should be bound within GroEL. More detailed modeling of GroEL/rhodanese complexes showed differences in the secondary and tertiary peaks in the high q region (representative examples are shown in Fig. 2a). The models for GroEL/rhodanese complexes were chosen on the basis of previous suggestions by a number of research groups (8-10). Rhodanese was allowed to assume many conformation in the models; ellipsoid, ring, and "champagne cork" (cylinder + semi-ellipsoid) shapes were used. Because of the size dominance of GroEL over rhodanese, the scattering curve for models depend much less on the shape of rhodanese, and more on its average position with respect to GroEL. For each model the integral discrepancy factor (R) was computed over entire q range as a measure of disagreement between the model and the experimental data (Fig. 2a) (24).

The important conclusion from these modeling experiments is that the differences between the SANS data for GroEL and the GroEL/rhodanese complexes are observed only in the high q region, and the low q region is not very sensitive to the location of bound rhodanese (Fig. 2a). Nevertheless, the small increase in R_g for the GroEL/rhodanese complex appears meaningful. Our modeling showed that the best agreement between the model and the experimental data can be obtained with a single protein substrate in the form of an ellipsoid or a "champagne cork" bound across the mouth of GroEL (R=0.051) (Fig 2b). Thus, only one rhodanese molecule appears to bind with high affinity to the chaperonin, presumably allowing the GroES co-chaperonin to bind on the other side of the cylinder. It is not clear from our data why the GroEL double ring structure binds only a single protein substrate (even with large molar excess of substrate), but our data are consistent with earlier observations of Mendoza *et al.* (17), that only one rhodanese molecule binds to GroEL. Perhaps protein binding introduces subtle changes in the conformation of chaperonin that are below the resolution limit for SANS, but sufficient to prevent binding of second rhodanese molecule to GroEL.

Out of two models which best fit the experimental data of GroEL/rhodanese complex, a "champagne cork" model appears to agree better with mutational (11) and electron microscopic data (8). In this model rhodanese appears to be extensively and exclusively bound to the apical domains of GroEL. These domains were implicated in protein binding and folding by mutational analysis (11). This high affinity interaction must involve extensive network of contacts between rhodanese and chaperonin and presumably involves large surface of apical domains of GroEL at the opening to the cavity. In the GroEL/rhodanese complex, the apical domains are flared five degrees. Accommodation of larger proteins may require 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 substrate. Rhodanese seems to spread across the GroEL opening and appears to assume a globular shape (a "champagne cork" or ellipsoid like). Thus the main role of GroEL is to provide a large active surface for the unfolded proteins to bind. Our model imposes constraints on the mechanism of chaperonin-mediated protein folding and assembly of multisubunit complexes.

Our data suggest that overall structure of GroEL in solution is similar to x-ray structure determined in the crystal. The differences involve a small flaring of apical domains, addition of the N- and C-terminal regions compactly at the equator, and accounting for the existing equilibria between chaperonin and its subunits (double rings, single rings and monomers). One molecule of denatured protein binds across the opening to the chaperonin cavity and appears to interact extensively with apical domains. Our data are highly consistent with the electron microscopic reconstructions of the GroEL/malate dehydrogenase complex (8), and do not support those models where the protein substrate is bound inside the cavity or outside the GroEL cylinder. It is also unlikely that the protein substrate can exchange freely between chaperonin cavities.

Acknowledgments

We thank Dr. Arthur Horwich of Yale University for providing strain the *E. coli* that overexpresses GroEL. We also thank Dr. Jonathan Trent of ANL for enticing discussions prior to beginning this project. This project was supported by the U.S. Department of Energy Office of Health and Environment Research and BES, under Contract No. W-31-109-Eng-38, and by the Directors 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. *Protein Science* **3**, 372-390 (1994)

2. Knott, R., Hansen, S. & Henderson, S.J. *J. Struct. Biol.* **112**, 192-198 (1994)

3. Schiffer, M., Stevens, F.J., Westholm, F.A., Kim, S.S. & Carlson, R.D. *Biochemistry* **21**, 2872-2878 (1982)

4. Svergun, D.I., Koch, M.H.J., Pedersen, J.S. and Serdyuk, I.N. *J. Mol. Biol.* **240**, 78-86 (1994)

5. Hendrick, J.P. & Hartl, F.-U. *A. Rev. Biochem.* **62**, 349-384 (1993).

6. Braig, K. *et. al.*, *Nature* **371**, 578 (1994).

7. Schmidt, M., Rutkat, K., Rachel, R., Pfeifer, G., Jaenicke, R., Viitanen, P., Lorimer, G., and Buchner, J. *Science* **265**, 656-659 (1994)

8. Chen, S. *et al. Nature* **371**, 261-264 (1994)

9. Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.U. *EMBO. J.* **11**, 4757-4765, (1992).

10. Braig, K., Simon, M., Hainfield, J., Furuya, F., & Horwich, A.L. *Proc. Natl. Acad. Sci. USA* **90**, 3978-3982 (1993).

11. Fenton, W.A., Kashi, Y., Furtak, K. & Horwich, A.L. *Nature* **371**, 614-619 (1994).

12. The modeling method is described in detail elsewhere (S.J. Henderson, submitted, *J. Appl. Cryst.* (1995). Briefly, a Monte Carlo method is used to generate the pair distribution function of the native enzyme or enzyme/substrate complex, where the scattering power of the protein distribution is modeled by a similar distribution of random coordinates, where

the local density of such random coordinates is proportional to the local scattering power of the protein.

13. McLennan, N.F., Girshovich, A.S., Lissin, N.M., Charters, Y & Masters, M. *Mol. Microbiol.* **7**, 49-58 (1993)

14. Horovitz, A., Bochkareva, E.S. & Girshovich, A.S. *J. Biol. Chem.* **268**, 9957-9959 (1993)

15. Horovitz, A., Bochkareva, E.S., Kovalenko, O. & Girshovich, A.S. *J. Mol. Biol.* **231**, 58-64 (1993)

16. Zahn, J., Harris, R., Pfeifer, G., Pluckthun, A. & Baumeister, W. *J. Mol. Biol.* **229**, 579-84 (1993).

17. Taguchi, H., Makino, Y. & Yoshida, M. *J. Biol. Chem.* **269**, 8529-8534 (1994)

18. Mendoza, J.A., Demeler, B. & Horowitz, P.M. *J. Biol. Chem.* **269**, 2447-2451 (1994)

19. Viitanen, P.V., Lorimer, G.H., Seetharam, R., Gupta, R.S., Oppenheeim, J., Thomas, J.O. & Cowan, N.J. *J. Biol. Chem.* **267**, 695-698 (1992)

20. Lissin, N.M. *FEBS Lett.* **361**, 55-60 (1995)

21. Quaite-Randall, E., Trent, J.D., Josephs, R. & Joachimiak, A. (unpublished results)

22. Mendoza, J.A., Lorimer, G.H. & Horowitz, P.M. *J. Biol. Chem.* **266**, 16973-16976 (1991)

23. Mendoza, J.A., Demler, B. & Horowitz, P.M. *J. Biol. Chem.* **269**, 2447-2451 (1994)

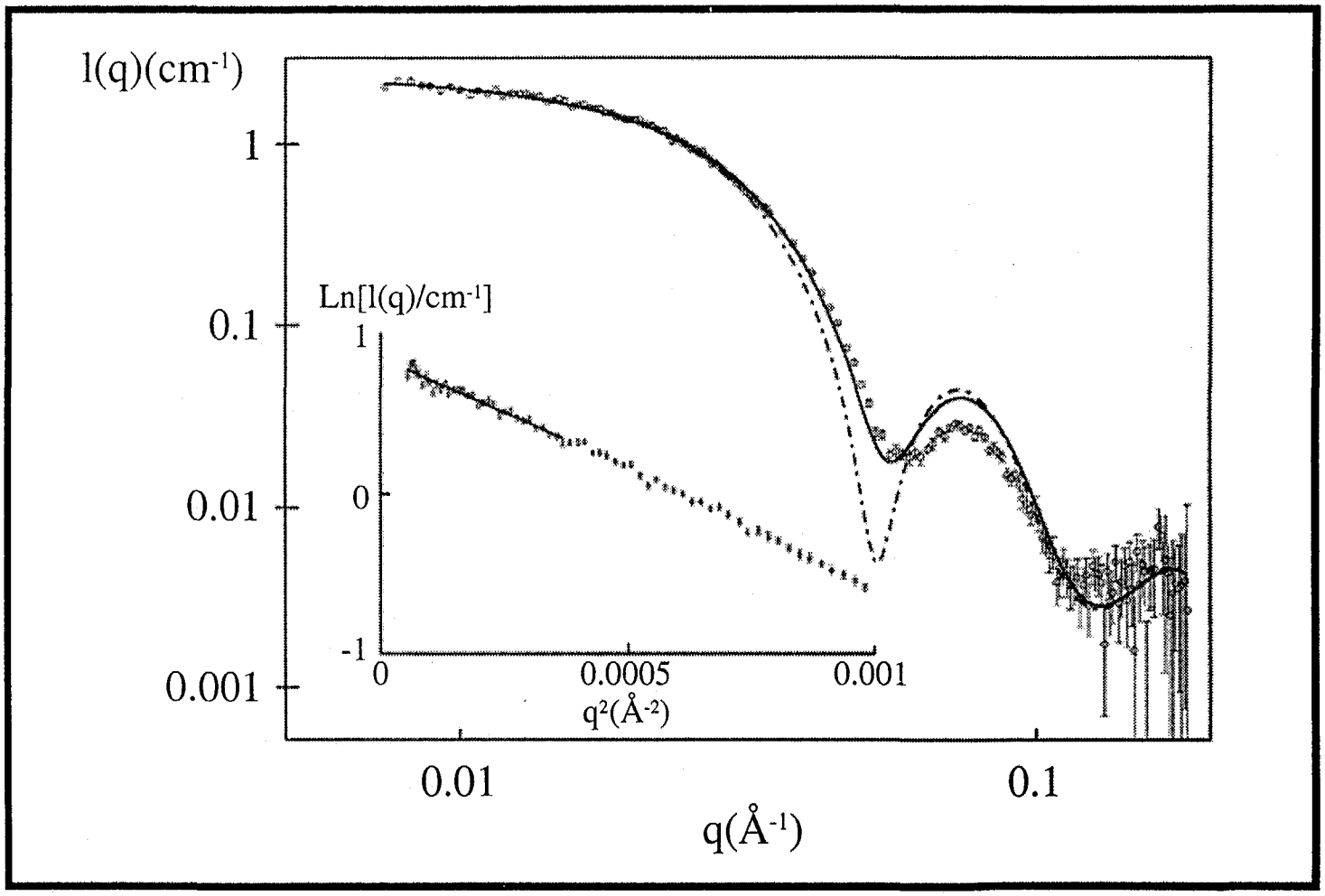
24. Feigin, L.A. & Svergun, P.I. Structure analysis by small angle x-ray and neutron scattering, Ed. Taylor, J.W. Plenum Press, NY p94 (1987)

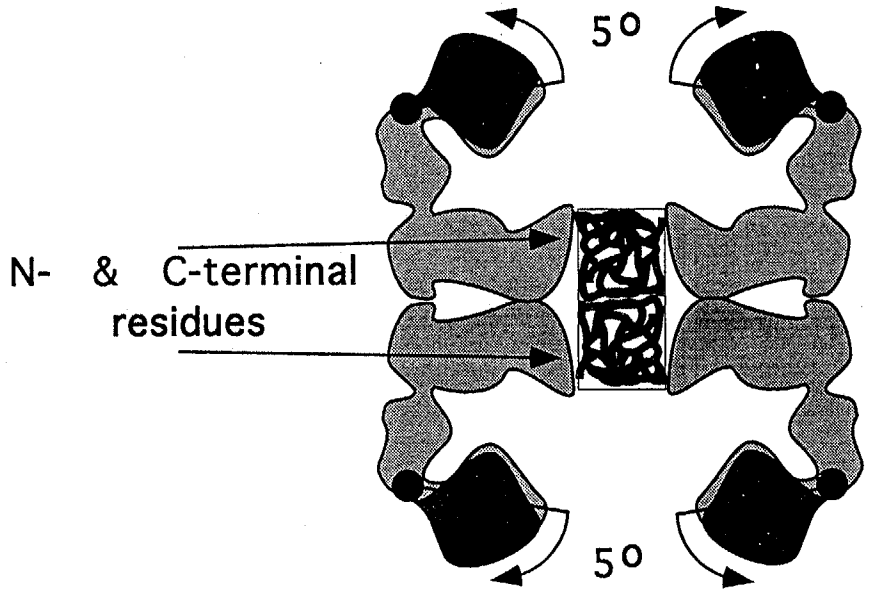
FIGURE LEGENDS

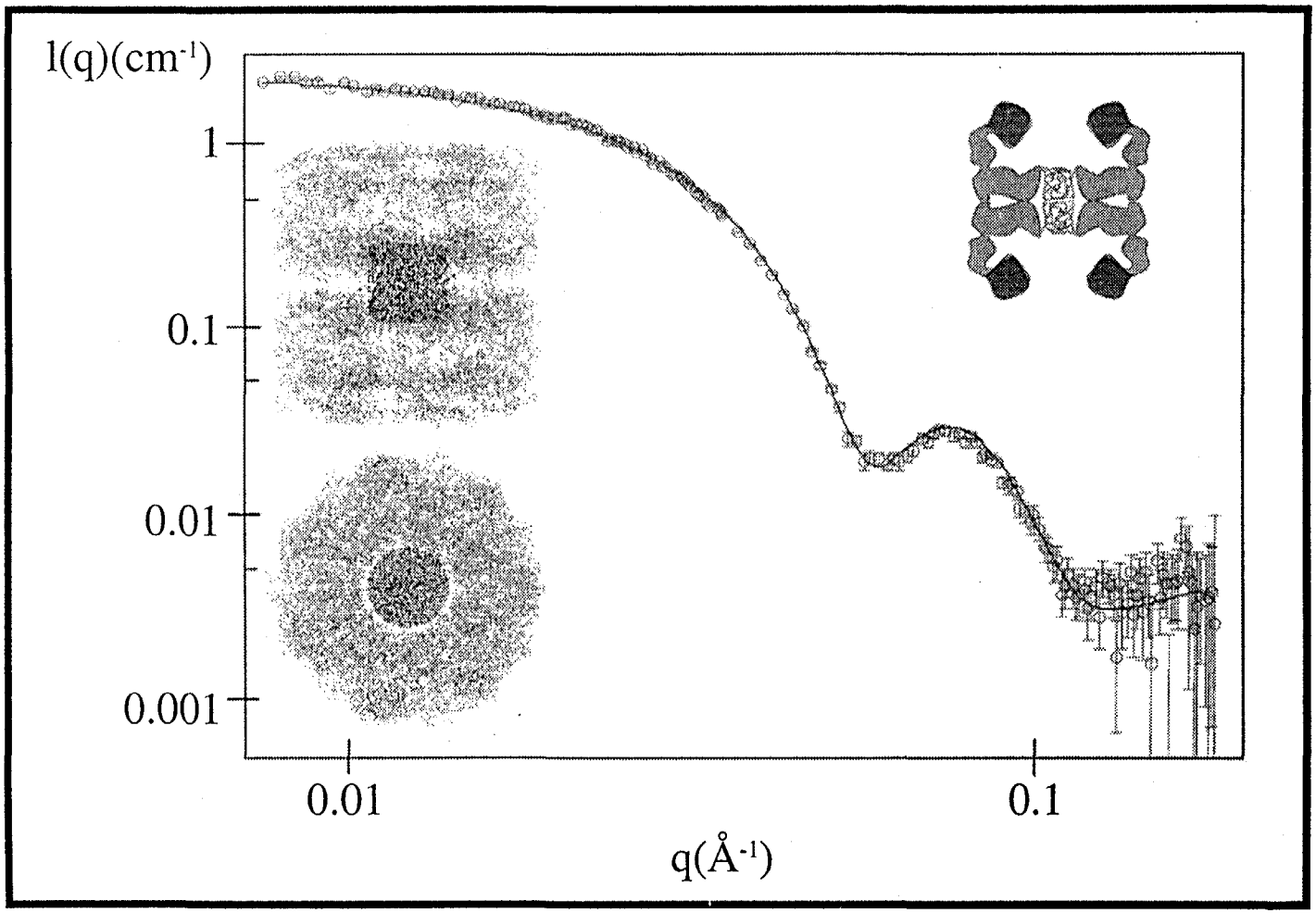
Fig. 1. SANS data of *E. coli* GroEL chaperonin.
a) SANS data were measured at 15°C for a 5.5 μM solution of GroEL in 98.5 % D_2O at the 30-m SANS instrument at the High Flux Isotope Reactor at Oak Ridge National Laboratory. The neutron wavelength was 4.75 Å ($\Delta\lambda/\lambda$ ~5%) and the sample-detector distances were 12 m and 3 m. Samples were contained in quartz cells with a path length of 0.5 cm. The data were corrected for the detector efficiency and instrument and solvent backgrounds prior to azimuthal averaging. Net intensities were converted to an absolute differential cross section per unit sample volume [$d\Sigma/d\Omega(q)$ in units cm^{-1}] by comparison with precalibrated secondary standards. (o), measured SANS data for GroEL; (--), calculated SANS curve from the crystal data of GroEL; (-), and the same calculated SANS after smearing with a Gaussian resolution function with $\Delta q/q$ =1.77E-3 for 12 m and $\Delta q/q$ =5.61E-3 for 3 m. The insert is the Guinier plot yielding the R_g of GroEL.
b) Crystal-structure-based modeling and fitting of GroEL was done as described (12). The icon shows cross section of GroEL in which the following modifications to the structure were introduced: i) missing mass at the N- and C-terminal fragments of the protein were added to the crystal data as a cylinder with h = 40 Å and r = 20 Å and located inside the cavity near the equator of the double ring complex as indicated by arrows; specific density for this protein region was assumed to be typical for globular proteins; ii) apical domains were rotated 5° outward using for each subunit center of rotation (•) defined by h=58 Å, r=65 Å.
c) Comparison of SANS curve of the GroEL model shown in (b) (-) with the experimental data (o) (insert shows distribution of scattering atoms used in the modeling with highlighted missing mass and the icon shows cross section of the GroEL). *E. coli* GroEL was obtained from the overproducing strain provided by Dr. A. Horwich from Yale University. The cells were grown overnight in 2 x TY medium and upon IPTG induction large quantities of GroEL and GroES were produced. GroEL was purified in three steps using FastQ, Sephacryl S-300 and MonoQ columns and was more then 99% pure as determined from SDS polyacrylamide gel electrophoresis under denaturing conditions and silver staining. Prior to data collection the chaperonin complex was purified rapidly on Superdex 200 column and concentrated with buffered D_2O on a Centricon 100 membrane in 40 mM Na/Kphosphate buffer, pH = 7.0, and 1 mM DTT.

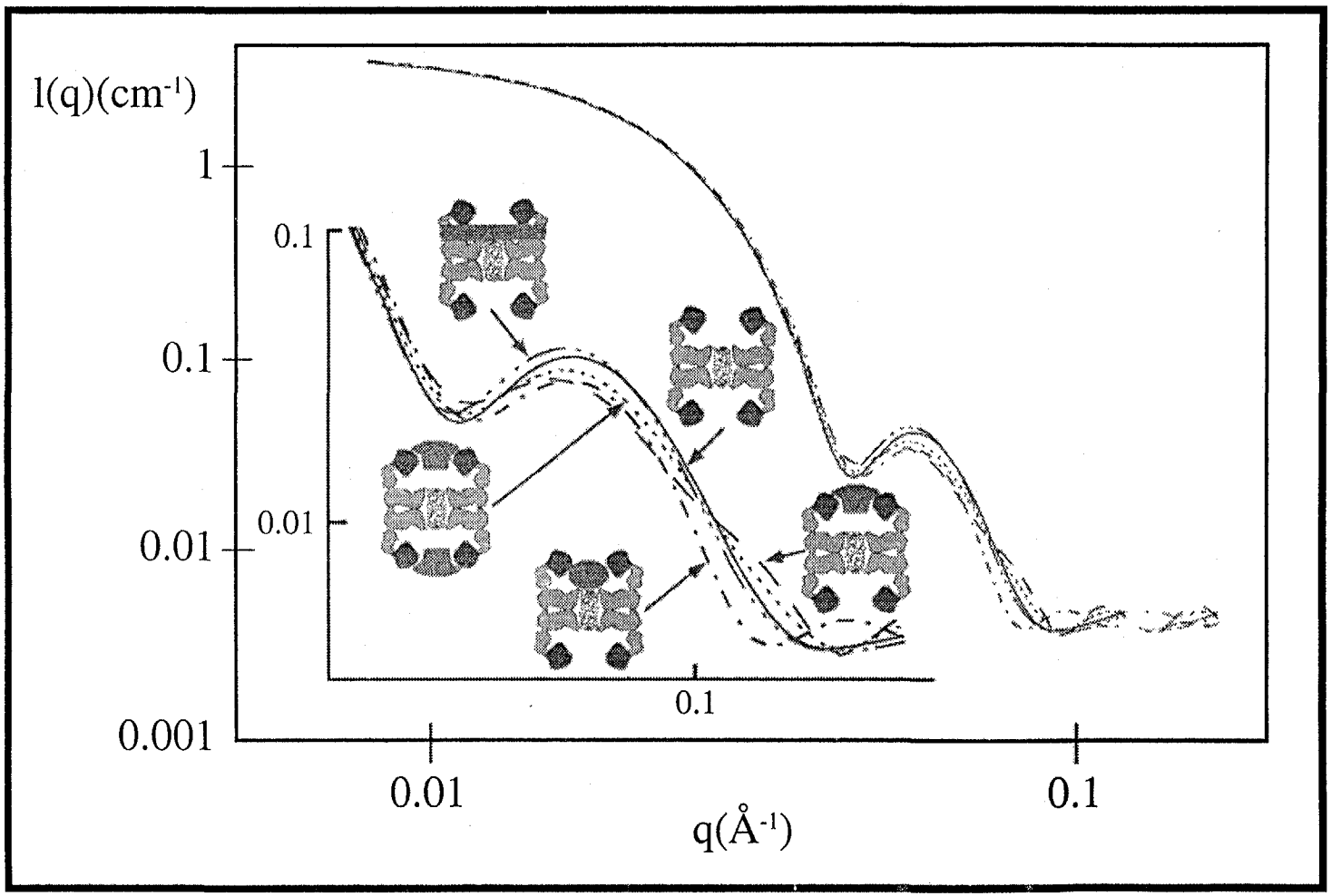
Fig. 2. SANS curves of *E. coli* GroEL/rhodanese complex. a) SANS curves calculated for GroEL and several models of GroEL/rhodanese complex were compared. To evaluate

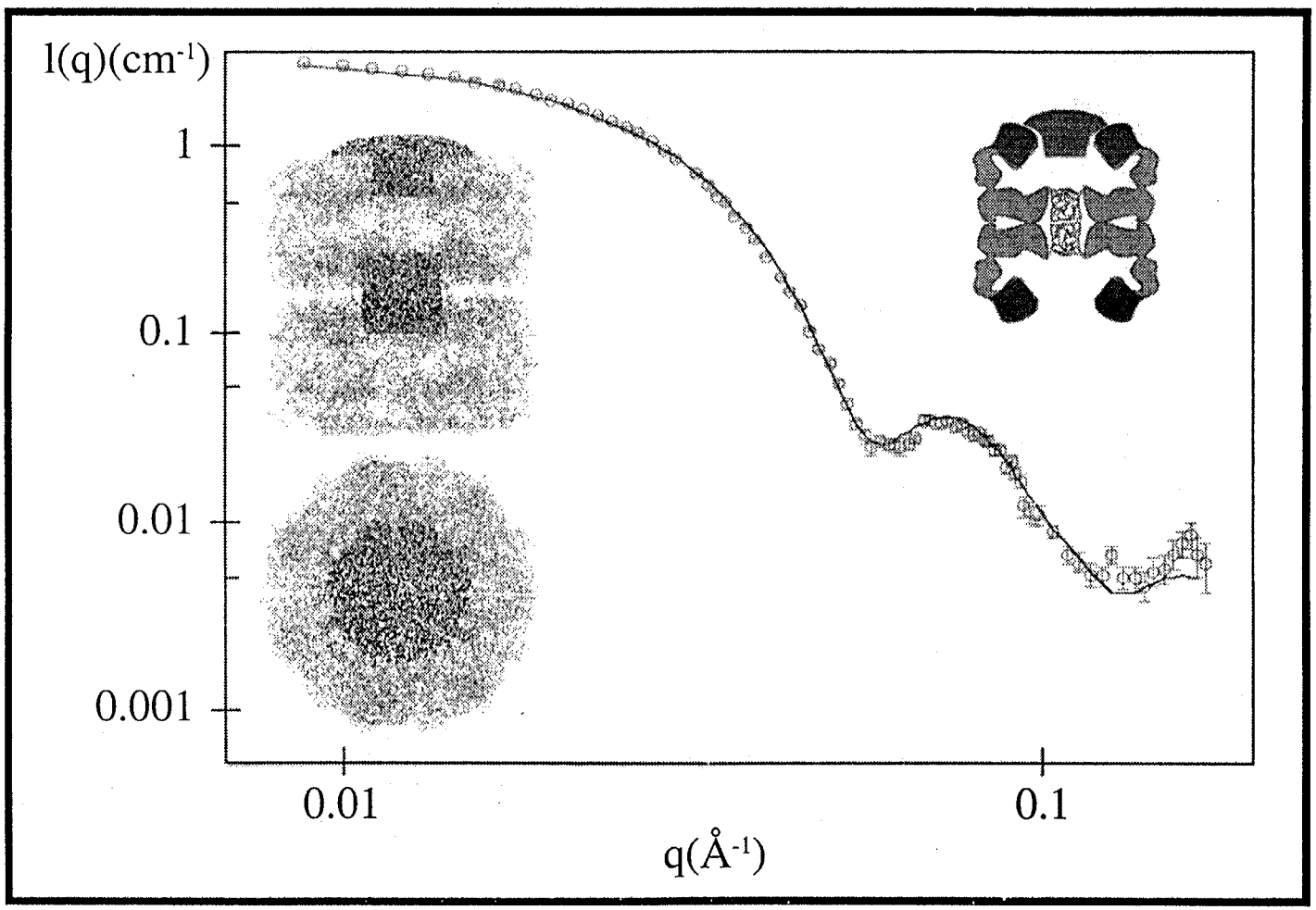
individual models, the integral discrepancy factor, defined as $R=\int |I_e(q) - s\, I_m(q)|\, q^2\, dq / \int I_e(q)\, q^2 dq$, was calculated (24). We compared experimental data obtained for GroEL/rhodanese complex with: i) free GroEL (-) (R=0.070) and the following theoretical models for the rhodanese bound to GroEL and approximated as: ii) an ellipsoid inside the cavity (-··-) (R=0.073), iii) a ring located outside GroEL (-····-) (R=0.071), iv) a "champagne cork" bound at the mouth of GroEL (···) (R=0.051), v) as two "champagne corks" bound to ends of GroEL (---) (R=0.054), vi) an ellipsoid bound at the mouth of GroEL (R=0.051) (data not shown since the scattering curve is virtually identical to iv). The champagne cork was created by combining a cylinder with r = 20 Å and h = 20 Å, and a semi-ellipsoid with semi-axes of 40 x 40 x 5 Å, making up a total volume of 41.9 kÅ^3; b) Fit of the experimental SANS data with the rhodanese substrate in the form of a "champagne cork" model (-) R=0.051. Inserts (side and top view) and the icon representing a cross section of the GroEL/rhodanese complex show the location (at z=60 Å from the equator) and suggested shape of the bound rhodanese. The GroEL/rhodanese complex was prepared in the following way: 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 Superdex 200 column using FPLC. The complex was concentrated in D_2O as described in Fig. 1. Polyacrylamide gels under denaturing conditions confirmed that the purified preparation contained GroEL/rhodanese complexes (data not shown).


I(q)(cm-1)
1
0.1
0.01
0.001
Ln[I(q)/cm-1]
1
0
-1
0
0.0005
0.001
q2(Å-2)
0.01
0.1
q(Å-1)


5°
N- & C-terminal residues
5°


I(q)(cm-1)
1
0.1
0.01
0.001
0.01
0.1
q(Å-1)


I(q)(cm-1)
1
0.1
0.01
0.001
0.1
0.01
0.1
0.01
0.1
q(Å-1)


I(q)(cm-1)
1
0.1
0.01
0.001
0.01
0.1
q(Å-1)