LETTERS

Chaperonin complex with a newly folded protein encapsulated in the folding chamber

D. K. Clare[1], P. J. Bakkes[2]†, H. van Heerikhuizen[2], S. M. van der Vies[2] & H. R. Saibil[1]

A subset of essential cellular proteins requires the assistance of chaperonins (in *Escherichia coli*, GroEL and GroES), double-ring complexes in which the two rings act alternately to bind, encapsulate and fold a wide range of nascent or stress-denatured proteins[1–5]. This process starts by the trapping of a substrate protein on hydrophobic surfaces in the central cavity of a GroEL ring[6–10]. Then, binding of ATP and co-chaperonin GroES to that ring ejects the non-native protein from its binding sites, through forced unfolding or other major conformational changes, and encloses it in a hydrophilic chamber for folding[11–15]. ATP hydrolysis and subsequent ATP binding to the opposite ring trigger dissociation of the chamber and release of the substrate protein[3]. The bacteriophage T4 requires its own version of GroES, gp31, which forms a taller folding chamber, to fold the major viral capsid protein gp23 (refs 16–20). Polypeptides are known to fold inside the chaperonin complex, but the conformation of an encapsulated protein has not previously been visualized. Here we present structures of gp23–chaperonin complexes, showing both the initial captured state and the final, close-to-native state with gp23 encapsulated in the folding chamber. Although the chamber is expanded, it is still barely large enough to contain the elongated gp23 monomer, explaining why the GroEL–GroES complex is not able to fold gp23 and showing how the chaperonin structure distorts to enclose a large, physiological substrate protein.

Chaperonin–substrate (binary) complexes were formed by rapidly mixing urea-denatured gp23 with GroEL. Ternary complexes were generated by adding gp31 (bacteriophage T4 GroES homologue) and the ATP transition state analogue ADP·AlF_3. Cryo-electron microscopy (cryo-EM) data sets of 30,000–35,000 particles were collected of both preparations, and initial three-dimensional maps were obtained by treating each data set as a single structure with seven-fold symmetry. The resulting maps showed GroEL and GroEL–gp31 complexes with some additional densities in the binding cavities (Supplementary Fig. 1). As in our earlier study on malate dehydrogenase (MDH) folding, we expected the non-native substrate to form heterogeneous and asymmetric complexes with the chaperonins. Therefore we used a combination of multivariate statistical analysis (MSA) and competitive projection matching to sort the images into more homogeneous classes and determine their structures, without imposing any symmetry[10,21].

The binary complexes were resolved into five classes (Supplementary Fig. 2, Supplementary Table 1). In agreement with our mass spectrometry results[22,23], we observed empty GroEL (20% of the images) and GroEL with extra density in either one (40%) or both rings (40%). The classes displaying the largest fraction of the substrate density are shown in Fig. 1, with the GroEL subunit domains fitted into the maps. The GroEL rings deviate little from seven-fold symmetry (Fig. 1c, d, g, h and rotational correlation analysis, not shown). This is unlike GroEL–MDH complexes in which the GroEL apical domains are bunched together on the side of the ring where MDH binds[10]. The large gp23 densities contact at least five of the seven GroEL apical domains in the ring, and the density is located deep inside the cavity, mainly around helix I and the underlying hydrophobic segment[7] (Fig. 1b, c, f–h). Whether substrate is bound or not, the open rings are very similar to one another and to apo GroEL.

Most of the ternary complexes fell into three well-defined structural classes (Fig. 2, Supplementary Table 1): 'empty' (no apparent substrate density, 33% of the images), substrate bound in the *trans* (open) ring (32%), and substrate bound in both *cis* and *trans* rings (35%). Consistent with the presence of the ATP analogue, the *trans* ring of all three complexes displayed the same intra-ring β-sheet contact between neighbouring equatorial domains as in the GroEL–GroES–ATP complex[20,24], although there is some variation in orientation of the apical domains. Less substrate density is resolved in the *trans* ring than in the binary complexes but it is found in a very similar location, that is, deep in the cavity and interacting with most of the apical domains (Fig. 2e, f, h and i). In the *cis* chamber, the substrate density fills most of the space, without strong contacts to either GroEL or gp31 (Fig. 2h). At lower density thresholds, the *cis* substrate density extends to contact several regions of the gp31 lid and the GroEL apical domains, potentially exerting force on the chamber (not shown). The atomic structure of gp23 is not known, but sequence alignment and genetic analysis have shown that it is closely related to the T4 vertex protein, gp24, for which the structure has been determined[25]. The encapsulated gp23 density is remarkably similar in shape to the major domain of gp24 (Fig. 3c).

Unexpectedly, we found that the *cis* chamber of the *trans*-occupied complex is slightly compressed, estimated as a 7% decrease in volume, and is significantly expanded in the *cis/trans* occupied complex, with a 12% increase in volume, relative to the *cis* chamber in the empty complex. This can be seen by the differences in density of the mobile loops of gp31, which interact with the apical domains of GroEL (arrows, Fig. 3a–c), and in the cross-sections through the *cis* apical domains which are almost losing contact at some positions (Fig. 3d–f). The expansion suggests that the folding substrate is exerting pressure on the folding chamber. No density is visible for the small, mobile insertion domain of gp23, but there is space at the base of chamber to accommodate it in a disordered form (Fig. 3c). The compression of the folding chamber in the *trans*-bound complex reveals an allosteric effect of the *trans* substrate, and suggests a mechanism for newly bound substrates to prime dissociation of the folding chamber. The nature of this allosteric effect was surprising, because the connections between GroEL and gp31 are stronger, rather than weaker as expected in preparation for the release of gp31.

Extracting the gp23 densities from the maps revealed remarkably well-defined structures for both the initially captured, non-native

[1]Department of Crystallography and Institute for Structural and Molecular Biology, Birkbeck College, Malet Street, London WC1E 7HX, UK. [2]Department of Pathology, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands. †Present address: Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands.

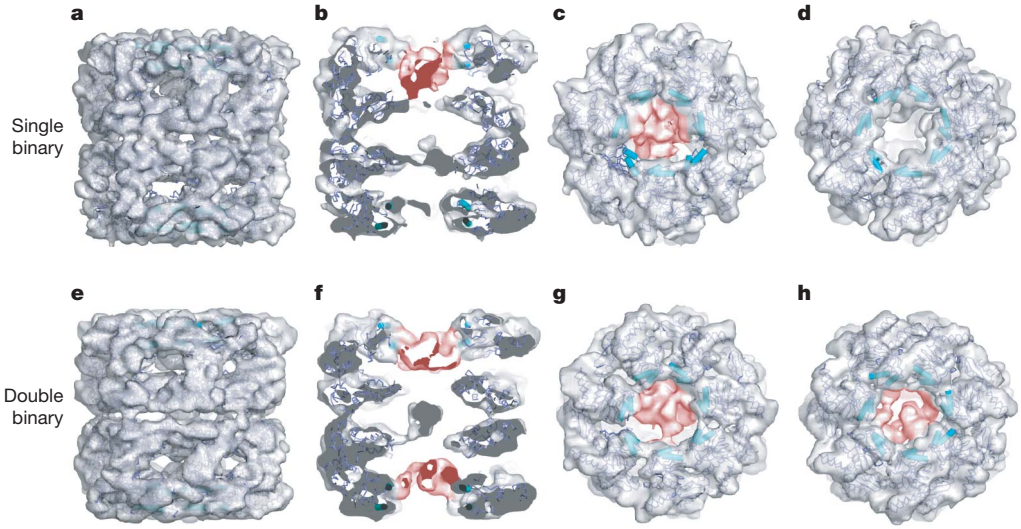

Figure 1 | Asymmetric reconstructions of GroEL with non-native gp23 in one or both rings. **a**–**d**, Binary complex with gp23 in one ring (red density), with the crystal structures of GroEL domains fitted into the maps, shown from the side (**a**), as a central section (**b**), from the top (**c**) and from the bottom (**d**). **e**–**h**, Same views of the binary complex with gp23 in both rings (red density). The EM density maps were sharpened between 20 and 10 Å. Automated docking of the atomic coordinates of the 42 GroEL domains as rigid bodies into each complex gave excellent fits to the maps, with hinge residues of neighbouring domains in proximity to each other, except for a few regions such as the intermediate domains of some subunits. Helices H and I are shown as cyan cylinders. The carboxy termini of some subunits are visible and either contact the substrate (**b**, upper ring) or bend away from it (**f**, lower ring). Interaction of the flexible GroEL C termini with substrates is consistent with earlier reports[29,30]. The diameter of GroEL complexes is around 140 Å and the resolution of the maps is around 11 Å at 0.5 Fourier shell correlation.

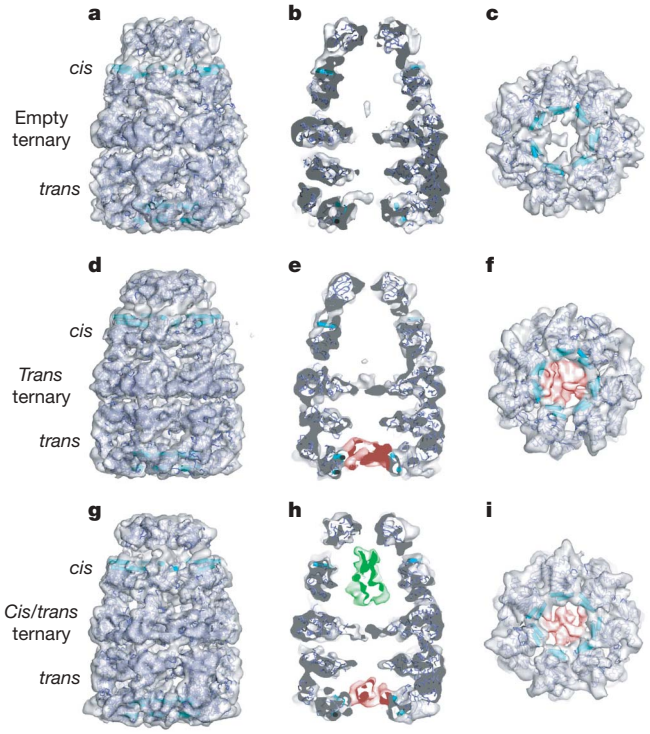

Figure 2 | Asymmetric reconstructions of GroEL–gp31 without visible substrate, with gp23 in the open ring and with gp23 in both rings. **a**–**c**, Empty ternary complex, with the crystal structures of GroEL and gp31 fitted as 49 individual domains, shown from the side (**a**), as a central section (**b**) and from the bottom (**c**). **d**–**f**, Same views of the *trans*-bound ternary complex. **g**–**i**, Same views of the *cis*/*trans* bound ternary complex. Density of gp23 in the *trans* rings is shown in red and in the *cis* ring in green. The resolution of the maps is around 10 Å at 0.5 Fourier shell correlation, and they were therefore sharpened between 20 and 10 Å. The GroEL and gp31 domain coordinates fit very well into the density maps, with only a few minor mismatches.

form and the almost fully folded state in the folding chamber of the GroEL–gp31 complex (Fig. 4). The cryo-EM density for gp23 bound to the top rings of the different binary complexes represents over half of the native molecular volume (Fig. 4a, b), showing that the non-native protein is constrained in position and shape when bound to the chaperonin complex (disordered regions are not seen in these averaged structures). The substrate densities in the second ring of doubly occupied complexes (bottom rings) are less well defined, possibly because one substrate dominates the alignment and classification, and the second substrate is not necessarily in a fixed position relative to the first. In the *trans* occupied ternary complex, the observation of less substrate density suggests that it is less well ordered in this class (Fig. 4c). In the *cis* complex, the encapsulated density is clearly recognizable as a low resolution version of the major domain in the gp24 crystal structure (Fig. 4d, e).

A recent FRET (fluorescence resonance energy transfer) study shows that another large, stringent GroEL substrate, Rubisco, occupies a more compact conformation in the *trans* ring of a ternary complex than in the open ring of a binary complex[14]. Our findings do not show a more compact conformation for gp23 in the *trans* ring of the ternary complex. Moreover, our mass spectrometry data shows that Rubisco binds to GroEL with strong negative cooperativity between the rings, but that gp23 binds to both rings[23,26]. Therefore, these two large substrates may have distinct modes of interaction with GroEL. A diversity of GroEL–substrate interaction modes can be anticipated, depending on the folding pathways and intermediates of different substrate proteins.

It is estimated that all of the GroEL in *E. coli* would be required to fold the large number of gp23 subunits produced during T4 infection[27]. It has been proposed that gp23 monopolizes the cell's GroEL through specific amino-terminal regions of gp23 that pause its translation and target the nascent chain to GroEL[27]. Our observation of well-defined electron density for the large domain of gp23 bound to GroEL, in a relatively small number of structural classes, is consistent with the notion that there is a specific binding region on gp23 that targets it to GroEL, leading to a preferred mode of binding.

In our previous study of GroEL–MDH complexes, most of the MDH density was observed in a similar region of the GroEL surface

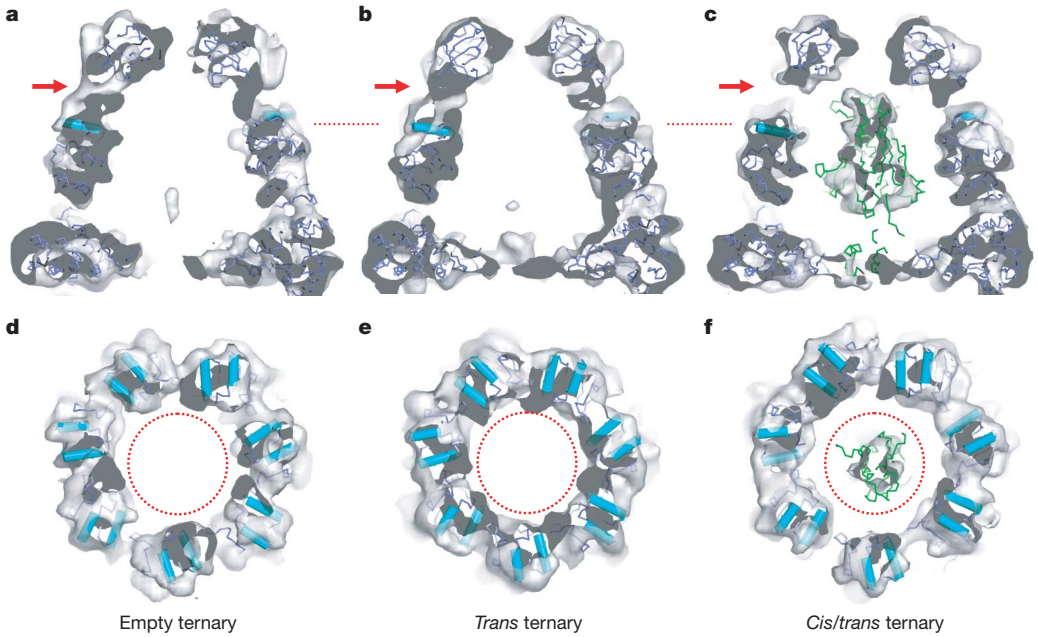

Figure 3 | Folding chambers of the GroEL–gp23–gp31 complexes. **a**–**c**, Side view sections and **d**–**f**, cross-sections through the apical domains of the folding chambers (at the position of the red dotted lines in **a**–**c**). Shown are views of the empty complex (**a** and **d**), the *trans*-occupied complex (**b** and **e**) and the *cis/trans* occupied complex (**c** and **f**). The red arrows in **a**–**c** show the loss of density at the contact between gp31 and GroEL. Correspondingly, the dotted red circles in **d**–**f** are all the same size (45 Å in diameter) and highlight the expansion of the apical domain ring in the *cis/trans* complex and the contraction of the *trans*-occupied apical domain ring. The major domain of the gp24 structure fits very well into the *cis* substrate density (**c**, **f**), unlike the mobile insertion domain of gp24 in the extended conformation seen in the crystal structure, where it makes an inter-subunit contact. In this position it clashes with the GroEL C termini, but there is clearly space available for it closer to the major domain.

to what we report here for gp23, namely the lower part of helix I and the underlying segment[10]. But MDH (33 kDa) only interacts with three of the seven apical domains, whereas gp23 (56 kDa) contacts at least five of the seven sites and occupies more of the binding cavity than MDH. The more asymmetric interaction of MDH with the cavity appears to perturb the GroEL symmetry in both rings[10]. The internal location of the substrate around helix I and the underlying segment leaves the helix H/I groove accessible to the mobile loop of the co-chaperonin even with a large substrate protein bound in the open cavity. However, it was reported that glutamine synthetase (51 kDa) occupies a more external binding site[28]. It is possible that the external binding site preferentially binds more folded states, similar to that suggested for one class of MDH complexes[10].

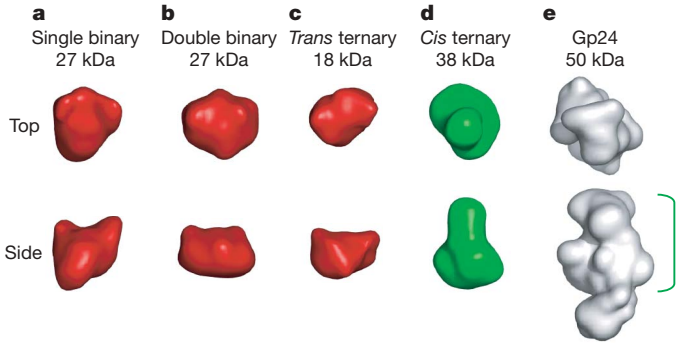

Figure 4 | Substrate densities isolated from the binary and ternary complexes. Shown are top and side views of substrate densities isolated from the GroEL–gp23 binary complex (**a**), the GroEL–$(gp23)_2$ binary complex (**b**), the *trans*-only GroEL–gp23–gp31 ternary complex (**c**) and the *cis* ternary complex (**d**) compared to the low resolution filtered density of the gp24 crystal structure (**e**). The isolated substrate densities were low-pass filtered at 15 Å and their approximate molecular masses were determined at a density threshold of 1σ of the complete complex. A larger observed mass in the class average reflects a more homogeneous class and therefore a more consistent structure for that sub-population. The green line in **e** indicates the major domain of gp24.

Once gp23 is bound to GroEL, gp31 (but not GroES) can encapsulate and fold the capsid protein[19]. The inability of GroES to encapsulate gp23 can be explained by the elongated shape adopted by newly folded gp23 inside the GroEL–gp31 chamber, and implies that the folding intermediate formed upon ATP and co-chaperonin binding is too bulky or too extended for GroES to bind. In this respect, it should be noted that gp31 has longer mobile loops and a larger internal space, which we show here to be stretched to the limit in order to encapsulate gp23.

The present work gives the first visualization of a newly folded substrate protein trapped in a largely native conformation inside the folding chamber of GroEL. The well-defined gp23 density in the GroEL–gp31 *cis* chamber unexpectedly shows that this physiological substrate protein is trapped in a unique position and orientation in the chaperonin chamber. The combination of size and shape of the *cis* chamber and gp23 results in the gp23 major domain being wedged into the chamber in a restricted position. The small insertion domain, which makes inter-subunit contacts in the viral capsid[25], is likely to be mobile and disordered until assembly of the capsid hexamers, which form as soon as the folded gp23 is released from GroEL–gp31 (ref. 22). The encapsulated substrate causes a significant expansion of the folding chamber in the region of the apical domains, and is exerting pressure on the connections between GroEL and gp31 and the inter-subunit contacts between apical domains of the *cis* ring. Thus, even though the folding chamber formed by GroEL–gp31 is larger than that formed by GroEL–GroES, an enclosed gp23 monomer still exerts pressure on it. Previous studies of GroEL–GroES and GroEL–gp31 complexes in various nucleotide states did not reveal any significant changes in the conformation of the folding chamber[20,24] (see also Supplementary Results and Supplementary Discussion). The expansion and compression observed here show that folding substrates directly and indirectly affect the conformation of the chamber. In conclusion, this study reveals a remarkable view of the chaperonin folding chamber strained to encapsulate a physiological substrate protein.

METHODS SUMMARY

The GroEL, gp23 and gp31 were expressed and purified as described[19]. The binary GroEL–gp23 complexes were prepared as described[23], yielding final concentrations of 1 μM GroEL oligomer, 2.5 μM gp23 monomer and 100 mM urea. GroEL–gp23–gp31 ternary complexes were formed by adding a twofold molar excess of gp31 heptamer over GroEL tetradecamer to preformed binary complexes, along with ADP to give a final concentration of 2.5 mM. After incubation at 24 °C, KF and $KAl(SO_4)_2$ were added, giving final concentrations of 20 mM KF and 2 mM $KAl(SO_4)_2$, to form the ATP transition state analogue ADP·AlF_3 to generate folding-active complexes.

Images were recorded on a 200 kV FEG microscope on photographic film and processed at 2.8 Å per pixel, with final data sets of 30,000 and 35,000 side views of the binary and ternary complexes, respectively. A starting model for the binary complex was obtained by angular reconstitution, and our previously determined GroEL–ADP–gp31 structure[20] was used as a starting model for the ternary complexes. The data sets were sorted into classes showing different substrate features by a combination of MSA and competitive projection matching[10], and the atomic structures of the GroEL subunit domains, gp31 and gp24 subunits were docked into the final, asymmetric maps as separate rigid bodies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 May; accepted 12 September 2008.

1. Houry, W. A. *et al.* Identification of *in vivo* substrates of the chaperonin GroEL. *Nature* **402,** 147–154 (1999).
2. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122,** 209–220 (2005).
3. Rye, H. S. *et al.* GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97,** 325–338 (1999).
4. Sigler, P. B. *et al.* Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67,** 581–608 (1998).
5. Horwich, A. L., Fenton, W. A., Chapman, E. & Farr, G. W. Two families of chaperonin: Physiology and mechanism. *Annu. Rev. Cell Dev. Biol.* **23,** 115–145 (2007).
6. Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371,** 578–586 (1994).
7. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371,** 614–619 (1994).
8. Lin, Z., Schwartz, F. P. & Eisenstein, E. The hydrophobic nature of GroEL-substrate binding. *J. Biol. Chem.* **270,** 1011–1014 (1995).
9. Farr, G. W. *et al.* Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* **100,** 561–573 (2000).
10. Elad, N. *et al.* Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* **26,** 415–426 (2007).
11. Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–$(ADP)_7$ chaperonin complex. *Nature* **388,** 741–750 (1997).
12. Motojima, F. *et al.* Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl Acad. Sci. USA* **101,** 15005–15012 (2004).
13. Lin, Z. & Rye, H. S. Expansion and compression of a protein folding intermediate by GroEL. *Mol. Cell* **16,** 23–34 (2004).
14. Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nature Struct. Mol. Biol.* **15,** 303–311 (2008).
15. Mayhew, M. *et al.* Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature* **379,** 420–426 (1996).
16. Laemmli, U. K., Beguin, F. & Gujer-Kellenberger, G. A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.* **47,** 69–85 (1970).
17. van der Vies, S. M., Gatenby, A. A. & Georgopoulos, C. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature* **368,** 654–656 (1994).
18. Hunt, J. F., van der Vies, S. M., Henry, L. & Deisenhofer, J. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell* **90,** 361–371 (1997).
19. Bakkes, P. J., Faber, B. W., van Heerikhuizen, H. & van der Vies, S. M. The T4-encoded cochaperonin, gp31, has unique properties that explain its requirement for the folding of the T4 major capsid protein. *Proc. Natl Acad. Sci. USA* **102,** 8144–8149 (2005).
20. Clare, D. K. *et al.* An expanded protein folding cage in the GroEL-gp31 complex. *J. Mol. Biol.* **358,** 905–911 (2006).
21. Elad, N., Clare, D. K., Saibil, H. R. & Orlova, E. V. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *J. Struct. Biol.* **162,** 108–120 (2008).
22. van Duijn, E. *et al.* Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nature Methods* **2,** 371–376 (2005).
23. van Duijn, E., Heck, A. J. & van der Vies, S. M. Inter-ring communication allows the GroEL chaperonin complex to distinguish between different substrates. *Protein Sci.* **16,** 956–965 (2007).
24. Ranson, N. A. *et al.* Allosteric signaling of ATP hydrolysis in GroEL-GroES complexes. *Nature Struct. Mol. Biol.* **13,** 147–152 (2006).
25. Fokine, A. *et al.* Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc. Natl Acad. Sci. USA* **102,** 7163–7168 (2005).
26. van Duijn, E. *et al.* Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **128,** 4694–4702 (2006).
27. Snyder, L. & Tarkowski, H. J. The N terminus of the head protein of T4 bacteriophage directs proteins to the GroEL chaperonin. *J. Mol. Biol.* **345,** 375–386 (2005).
28. Falke, S. *et al.* The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.* **348,** 219–230 (2005).
29. Tang, Y. C. *et al.* Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125,** 903–914 (2006).
30. Machida, K. *et al.* Hydrophilic residues 526KNDAAD531 in the flexible C-terminal region of the chaperonin GroEL are critical for substrate protein folding within the central cavity. *J. Biol. Chem.* **283,** 6886–6896 (2008).
31. Chaudhry, C. *et al.* Role of the gamma-phosphate of ATP in triggering protein folding by GroEL-GroES: Function, structure and energetics. *EMBO J.* **22,** 4877–4887 (2003).
32. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142,** 334–347 (2003).
33. Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J. Struct. Biol.* **116,** 9–16 (1996).
34. Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
35. van Heel, M. *et al.* A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
36. Navaza, J. *et al.* On the fitting of model electron densities into EM reconstructions: A reciprocal-space formulation. *Acta Crystallogr. D* **58,** 1820–1825 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Westlake, D. Houldershaw and L. Wang for computing and EM support. This work was carried out at the School of Crystallography, Birkbeck College, and was supported by a Wellcome Trust programme grant, EU 3D EM Network of Excellence and 3D Repertoire grants.

Author Information The three-dimensional reconstructions for the two binary complexes and the three ternary complexes are deposited in the EBI-MSD EMD database with accession codes EMD-1544 through to EMD-1548. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.R.S. (h.saibil@mail.cryst.bbk.ac.uk) and S.M.v.d.V. (vdvies@vumc.nl).

METHODS

Sample preparation. Binary complexes. GroEL was diluted to a concentration of 1.2 μM (oligomer) in 40 mM Tris-HCl pH 7.5, 10 mM KCl, 10 mM $MgCl_2$. 25 μM gp23 monomer was denatured in 6 M urea for 2 h at 24 °C and 300 r.p.m. on a bench top shaker. The denatured gp23 was then added to 90 μl of GroEL solution in ten 1-μl additions, with vortexing and centrifugation after each addition, giving a final concentration of 1 μM GroEL, 2.5 μM gp23 and 600 mM urea. To maximize the occupancy of the gp23–GroEL bound complex, the denatured gp23 was added in a 2.5 molar excess over GroEL oligomer. The urea was diluted to 100 mM by addition of 500 μl of buffer to the GroEL–gp23 complex. This mixture was then concentrated in a Vivaspin with a 5 kDa cut-off, to a final concentration of 1 μM GroEL oligomer, 2.5 μM gp23 monomer and 100 mM urea.

Ternary complexes. The non-hydrolysable ATP analogue ADP·AlF_3 has been shown to produce folding active ternary complexes for several GroEL substrates[31]. We therefore used it to produce folding-active GroEL–gp23–gp31:ADP·AlF_3 complexes. 0.75 μl of 140 μM (heptamer) gp31 was added to 50 μl of binary complex prepared as above, followed by 0.3 μl of 400 mM ADP. The mixture was incubated for 10 min at 24 °C. Then 1 μl of 1 M KF was added and mixed, followed by 1 μl of 100 mM $KAl(SO_4)_2$, mixing and 5 min incubation. To maximize the number of ternary complexes, a twofold molar excess of gp31 over GroEL was added to the preformed binary complexes.

Data collection. Samples (3.5 μl) were loaded on to freshly glow discharged (45 s) c2/2 holey carbon C-flat grids (Protochips Inc.), blotted and plunged into liquid nitrogen cooled liquid ethane. Low dose images were recorded on an F20 microscope (FEI) at 50,000× on Kodak SO-163 film with a defocus range of 1–3 μm.

The images showed an even distribution of side and end views very similar to that seen for GroEL alone. The number of side views is crucial to the study of GroEL complexes as they contain all the information required to reconstruct the molecule in three dimensions. In contrast, for the GroEL–MDH binary complex, a modified GroEL had to be used to increase the number of side views[10].

Image processing. The images were digitized at a step size of 7 μM on a Zeiss photodensitometer (Intergraph), giving a sampling of 1.4 Å per pixel. The defocus was determined with CTFFIND3[32]. Particles were manually picked with XIMDISP and extracted into 512 × 512 pixel boxes with LABEL[33]. The cut-out particles were corrected for the phase reversals of the contrast transfer function (CTF) in SPIDER[34], cropped to 320 × 320 pixels and 2 × 2 averaged to give a final sampling of 2.8 Å per pixel. The boxed particles were filtered between 200 and 6 Å for the binary complexes and between 285 and 6 Å for the ternary complexes and then normalized. The images were centred in SPIDER using side and end view projections of the crystal structures of GroEL and GroEL–GroES that were low pass filtered to 40 Å. The aligned images were then classified using MSA in IMAGIC into classes each containing on average five images[35]. Using these classes, end views, bad images and GroEL lacking gp31 in the ternary complex sample were removed to produce two data sets containing 30,000 and 35,000 side view images of the binary and ternary complexes, respectively. The binary data set, including an end view class, was aligned and initially reconstructed by angular reconstitution in IMAGIC with imposed seven-fold symmetry, treating it as a single structure.

Binary complexes. The initial map was used as a starting model for projection matching in SPIDER, with the number of references increasing to 155 (90–105° in β and 0–51.4° in γ) which gave a γ spacing of 2.1°. The resulting reconstruction was masked to generate two reference structures, one without substrate density and the other containing one-third of the sevenfold averaged substrate density. At this stage the symmetry was relaxed from *C*7 to *C*1 and the reconstructions were used for competitive projection matching, in which the images are sorted by cross-correlation value to projections of both reference structures. The images that aligned to the substrate-bound reference were then subjected to MSA, and two eigenimages that represented substrate variance were used to separate the images into two populations of substrate-bound structures, GroEL–gp23 and GroEL–$(gp23)_2$, which, along with the empty structure, were used for further competitive projection matching.

For the asymmetric refinement, the number of reference images was increased from 155 to 528 per reconstruction for further competitive alignment with an angular spacing of 3° from 78–90° in β and 3.1° from 0–360°in γ and with the mirror option turned on in the SPIDER operation AP SH. After another three rounds of alignment, when the image alignments were stable for each of the three species, they were subjected to MSA. This led to a further separation to give one empty, two GroEL–gp23 and two GroEL–$(gp23)_2$ species. The separation was refined again by projection matching. When the alignment had stabilized, 95% of the images aligned to the same references in consecutive alignments. MSA was carried out on each of the subsets to check that the substrate variations had been minimized (Supplementary Fig. 3a–c). The total number of images in the data set limited the number of different complexes that could be discriminated and therefore no more separations were carried out. The resolutions of the five final reconstructions, determined by the 0.5 criterion of the Fourier shell correlation, were 11–12 Å.

Ternary complexes. The initial model used to align the ternary complexes was the previously published GroEL–gp31–ADP structure[20], re-projected at an angular spacing of 2.1° (90–110° in β and 0–51.4° in γ). The resulting seven-fold symmetric reconstruction contained density for both *cis* and *trans* bound substrate and was used for further alignment and MSA, eventually yielding five classes. Classes 1–3 together represented 70% of the data and contained well aligned images, whereas classes 4 and 5 contained poorly aligned images. After this separation we continued the alignment procedure with classes 1, 2 and 3 as independent data sets. The alignment of these three classes stabilized quickly with 95% of the images aligning to the same reference in consecutive rounds, and the calculated eigenimages showing that the variation due to substrate had been removed (Supplementary Fig. 3d–f).

Once the alignment had stabilized for each of the complexes the symmetry was relaxed from *C*7 to *C*1 and the number of references increased to give a final angular spacing of 2° (90–110° in β and 0–360° in γ and with the mirror option turned on in the SPIDER operation AP SH). The refinement was stopped when more than 95% of the images aligned to same reference in consecutive rounds of alignment. The images of the *cis*/*trans* bound complex were also aligned to the *trans* only bound references and the resulting reconstruction had almost identical *cis* substrate density, confirming the reliability of the *cis* substrate density. The resolutions of the three maps were all around 10 Å by the 0.5 Fourier shell correlation criterion.

Docking of atomic coordinates. As there was no symmetry applied to the reconstructions, all 14 subunits of GroEL (1GRL and 1AON) and 7 subunits of gp31 (1G31) were fitted individually. Each of the GroEL subunits was further split into its three separate domains, apical 192–373, intermediate 137–191 and 373–409, and equatorial 2–136 and 410–525, which were docked as rigid bodies. The mobile loops of gp31 (residues 15–36) were omitted since they were not in the GroEL-bound conformation in the crystal structure. The initial docking was done manually in Pymol (www.pymol.org) and then refined by the automatic docking program URO[36]. However the fitting of the intermediate domains was not accurate since the density for these domains in most of the maps was not as well defined as that of the other domains. Therefore, the intermediate domains had to be checked and manually corrected to maintain the connectivity between the equatorial and intermediate domains. After adjustment of the intermediate domains all the connections between it and the equatorial and apical domains were within 7 Å (α-carbon distance), with the majority less than that.