research papers

Docking structures of domains into maps from cryo-electron microscopy using local correlation

Alan M. Roseman

Medical Research Council, Laboratory of
Molecular Biology, Hills Road, Cambridge
CB2 2QH, England

Correspondence e-mail:
roseman@mrc-lmb.cam.ac.uk

Accurate maps of large macromolecular complexes can be calculated from cryo-electron micrographs of non-crystalline specimens to resolutions of about 10 Å. A method to dock the atomic structures of domains solved by X-ray crystallography or nuclear magnetic resonance into cryo-EM maps is presented. Domains can be docked independently into large complexes without prior definition of the boundaries. No special symmetry is assumed and the procedure is suitable for general application to almost any system where a cryo-EM map (at 15 Å resolution or better) of a complex has been obtained and the atomic structures of the components are available. This is achieved through use of a real-space density-matching procedure based on local correlation. A complete asymmetric unit search correlating a density object derived from the atomic coordinates and the density of the EM map is performed. The correlation coefficient is calculated locally in real space using only values of the search object and corresponding samples extracted from the EM map which are under the 'footprint' of the positioned search object. The procedure has been demonstrated by docking the domains of GroEL from the crystal structure into a cryo-EM map Fourier filtered to 12 or 15 Å resolution. The correct positions were found without applying any additional constraints. A model of the oligomer built from the docked domains compared favourably with the known crystal structure, confirming the validity of the approach. The procedure is designed to facilitate the incorporation of additional constraints on the docking solutions, which could help to dock using lower resolution maps.

1. Introduction

1.1. Cryo-EM of preparations of single particles

Any sufficiently large macromolecular complex that can be purified and prepared in solution at high enough concentration (a few milligrams per millilitre) is potentially a good specimen for single-particle cryo-electron microscopy (cryo-EM) analysis. Images of many particles representing views at different orientations must be combined in order to calculate the three-dimensional (3D) structure of the object. Maps of 20–10 Å resolution are now regularly produced. Details of the methods used are described in the book *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Frank, 1996) and references therein.

The information provided by the single-particle technique is complementary to X-ray crystallography. It is suited to analysis of large structures, greater than a few hundred kilodaltons. Advantages of the technique are that the specimens are prepared directly from solution, preserved in a frozen

hydrated state (Dubochet *et al.*, 1988) and relatively little material is required (of the order of micrograms). It is possible to trap transient structures by freezing (Berriman & Unwin, 1994) and to deal with heterogeneous populations by classification and separation of the images (van Heel & Frank, 1981; Frank, 1990). This can be performed automatically as part of the image-processing refinement (Gabashvili *et al.*, 1999; Rye *et al.*, 1999; http://www.cell.com/cgi/content/full/97/3/325/DC1; Schoehn *et al.*, 2000). In contrast, smaller less flexible objects are more tractable by X-ray crystallography and atomic models are generated.

Cryo-EM images and electron-diffraction patterns of two-dimensional arrays of proteins have been used to calculate maps at 3.4–3.7 Å resolution into which atomic models could be built [bacteriorhodopsin (Henderson *et al.*, 1990), light-harvesting complex (Kühlbrandt *et al.*, 1994), tubulin (Nogales *et al.*, 1998)]. The success with electron crystallography suggests that the resolution of maps calculated by the single-particle method can be improved. It was possible to directly trace the α-helical fold of the protein subunit in the 7.4 Å map of hepatitis B core shells (Böttcher *et al.*, 1997). This discovery of a new fold has been confirmed by the 3.5 Å X-ray structure of the same shells (Wynne *et al.*, 1999) and is currently the only example of a new fold solved by cryo-EM of single particles. It is more usual that detailed models of structures have to be obtained by docking atomic models into the EM maps and the domain-docking procedure method presented here is an important contribution towards this goal.

1.2. Review of docking of X-ray structures into EM maps

1.2.1. Manual docking. The first EM maps into which atomic structures were docked were of complexes of icosahedral viruses with their receptors (Olson *et al.*, 1993; Bella *et al.*, 1998) or fragments of antibodies (FABs; Smith *et al.*, 1993; Liu *et al.*, 1994; Stewart *et al.*, 1997) or of actomyosin filaments (Rayment *et al.*, 1993; Schröder *et al.*, 1993). The components of these systems (the viral shells, FABs, receptors, myosin) were large molecules with distinctive shapes at the resolution of the maps (20–30 Å), which allowed the docking to be performed with confidence and reasonable accuracy. When the complex of human rhinovirus and the FAB was crystallized and solved, it was found that the model from docking was accurate to within 4 Å (Smith *et al.*, 1996).

The head domain of myosin (myosin subfragment 1 or S1) has been docked manually into EM maps of complexes with actin. Myosin and actin were either docked directly into the map of the complex (Schröder *et al.*, 1993) or myosin was docked after an F-actin model had been positioned in the map of the actomyosin complex (Rayment *et al.*, 1993). In the latter the F-actin model was docked into an EM map of the actin filament that had been aligned with the EM map of the actomyosin complex.

GroEL apical domains have a distinct triangular shape at low resolution and were also amenable to manual docking (White *et al.*, 1997). The intermediate domains and equatorial domains could not be unambiguously docked into the ∼25 Å maps because interfaces between the apical, intermediate and equatorial domains of one subunit and the interface between intra-ring equatorial domain packing were not clearly resolved.

The motor proteins kinesin and nonclaret disjunctional protein (ncd) have been manually docked into EM maps of their complexes with microtubules (Hirose *et al.*, 1999). Labelling of the motor proteins with gold markers adds distinct features to the maps which has helped to define their orientation (Rice *et al.*, 1999). This and use of other biophysical data has helped to reduce the disparity in proposed models of this system.

Ribosome complexes are examples of systems with no symmetry where manual docking has provided insight (Stark *et al.*, 1997; Agrawal *et al.*, 1998).

1.2.2. Quantitative docking. Stewart *et al.* (1993) optimized the docking of a hexon subassembly of the icosahedral adenovirus shell by cross correlation. Programs such as *X-PLOR* (Brünger, 1992) have been used for quantitative refinement of the manual placements in the viral antibody–receptor complexes (Wikoff *et al.*, 1994; Cheng *et al.*, 1995; Hewat *et al.*, 1997; Che *et al.*, 1998; Kolatkar *et al.*, 1999), operating in reciprocal space. Recently, more sophisticated programs have been developed for quantitative docking with more thorough searching of the icosahedral asymmetric unit in real space (Cheng *et al.*, 1995; Grimes *et al.*, 1997). These procedures were used on cryo-EM maps at ∼25 Å resolution which have little internal structure and the scoring methods were not set up to make use of such features. Docking in the icosahedral context has been concisely reviewed in Baker *et al.* (1999). The *trans* apical domains of GroEL–GroES complexes have a close packing of apical domains within a heptameric ring. *X-PLOR* was used to refine the apical domain position and packing, imposing strict sevenfold non-crystallographic symmetry (NCS) and van der Waals terms to prevent interpenetration of the subunits (Rye *et al.*, 1999; http://www.cell.com/cgi/content/full/97/3/325/DC1). Mendelson & Morris (1994) implemented a reciprocal-space refinement of a model of F-actin incorporating helical constraints. They also refined a model of the complex of myosin and actin (Mendelson & Morris, 1997). The refined model of the actomyosin complex was significantly different from the manual fits and the *R* factor had improved.

In complexes where ligands are bound externally, for example to viral shells or to actomyosin fibres, the additional components are located mainly in an outer radial zone. Another way of isolating a component is to determine a difference density between a map of the complex and a map of a sub-complex that excludes the component to be docked. The program *SITUS* (Wriggers *et al.*, 1999) is suited mainly to these types of application and has been used to dock the motor protein ncd into a map of an ncd–microtubule complex. It was also used to find a docking position for kinesin in a kinesin–microtubule complex, which was then refined to optimize the correlation coefficient (Kikkawa *et al.*, 2000). An initial manual docking of terminal domains of clathrin allowed the local region of the cryo-EM map around the terminal domains

to be extracted. The position and orientation were then refined using exhaustive cross correlation between a search object derived from the coordinates and the region of the map extracted (Musacchio *et al.*, 1999). Global correlations between a search object and the EM map are used by the program *COAN* (Hanein *et al.*, 1998; Volkmann & Hanein, 1999) to define the measure-of-fit. It was applied to docking S1 and fimbrin into maps of their complexes with actin. The global correlation coefficients are less reliable if the search object represents only a portion of the EM map. The usefulness of incorporating additional biochemical data to discriminate potential solutions was demonstrated. Hewat & Blaas (1996) implemented a cross-correlation procedure to dock an FAB model into density isolated from a complex with human rhinovirus.

Thus, quantitative docking has until now been performed using only highly constrained symmetrical systems or where the component to be docked is somehow isolated from the map of the complex. The program presented here is capable of docking domains independently into large complexes without prior definition of the boundaries. No special symmetry is assumed and the procedure is suitable for general application to almost any system where a cryo-EM map of a complex has been obtained and the atomic structures of the components are available. The procedure has been demonstrated by docking all the domains of GroEL into a 15 Å cryo-EM map without applying any additional constraints.

2. The docking procedure

The core of the docking procedure is a 3D density-correlation algorithm. An atomic model of the domain to be docked, derived by X-ray crystallography or nuclear magnetic resonance, is used to calculate a density-search object compatible with the EM map. This is exhaustively correlated with the density of the EM map. All possible positions and orientations of the search object within the asymmetric unit of the EM map are searched in order to find the optimal match, which is defined by the maximum correlation coefficient. The correlation coefficient is calculated locally in real space using only values of the search object above a defined threshold and corresponding samples extracted from the EM map that are under the 'footprint' of the positioned search object. The molecular envelope of the search object and the density variations within it are correlated with the EM map, which is also a 3D map containing internal structure. Such internal features become significant at about 15 Å resolution.

The advantage of determining the correlation coefficient locally at each position searched is that the two regions of density maps to be compared are scaled together. It is not necessary to fit or derive an absolute scale or background level of density representing the solvent surrounding the specimen in order to match the search object and the EM map because solvent is not included the density-search object calculated from the atomic coordinates. Global cross correlations can be calculated more rapidly by making use of the convolution theorem and the fast Fourier transform algorithm. However,

in that case the scaling is not performed locally and the score will be sensitive to the scaling of the search object to the EM map, which may not be straightforward to achieve. As a consequence of this, false-positive solutions arise. The structure in the EM map that does not correspond to the domain being docked effectively acts as 'noise' and distorts the scaling of the search object and EM map.

Since with the local correlation procedure only the map density under the footprint of the search object is considered at each point in the search, no density from adjacent domains contributes when the correlation at the correct position is being evaluated. Therefore, the boundary between this domain and the rest of the complex does not need to be known *a priori* in the EM map. Indeed, the docking problem is partly solved once this boundary is defined. Each domain or fragment for which an atomic model is available can be docked into the EM reconstruction of the larger complex individually. If coordinates are available for only some parts of the structure, then these can be docked in without reference to the other parts. It is assumed that the domain that is being docked does indeed exist within the complex in that conformation.

Packing and other constraints are not taken into account during the search, but can be applied by filtering the list of top-scoring solutions at a later stage. For example, clashes of symmetry-related or adjacent domains indicate solutions that can be discarded. The approach is to dock each domain independently and then to look for compatible solutions. If the map has been symmetrized then the symmetry information will have been implicitly used to improve the signal-to-noise ratio in the map and in the definition of the asymmetric unit. However, the docking of domains is not constrained by the search to be compatible with the symmetry.

2.1. Generating the density-search object

The program *O* (Jones *et al.*, 1991) is used as the visual interface. After the EM map and the atomic coordinates of the domain are read into *O*, an initial placement of the coordinates into the asymmetric unit of the EM map is made. The coordinates of the domain at this new position are written out and the search-object density is calculated from them. The origin of the search is defined by the centre of mass of the search object.

The search object is calculated from the coordinates of atoms in the domain by interpolating the atomic *Z* number into a grid with the same sampling as the EM map, as described in Stewart *et al.* (1993) or implemented as CP FROM PDB in *SPIDER* (Frank *et al.*, 1996). This creates a map proportional to the Coulomb potential of the molecule, which is what the electron beam in the electron microscope 'sees'. The EM map and search object are low-pass Fourier filtered to the same resolution, the resolution at which the EM map is believed to be valid. *SPIDER* (Frank *et al.*, 1996) was used to apply the Fourier filters. A high-pass filter was also applied to remove low-resolution terms beyond the lowest spatial frequencies present in the complex. This flattens the background in the EM map because frequencies describing

density changes over distances larger than the dimension of the molecule are filtered out. An additional step at this stage could be to manipulate the search object, *e.g.* to apply reciprocal-space scaling to match the power spectrum of the search object to the EM map in order to compensate for the effect of the contrast-transfer function of the electron microscope on the map. A density threshold defining those parts of the search object to be used is assigned. The threshold was chosen so that the isosurface boundary defined around the search object contained approximately the predicted volume of the domain.

2.2. 3D density-matching algorithm

The density-search object is converted into a list of coordinates of voxels and their sample values, S_i, where i runs from 1 to n, the number of samples in the search object. Values lower than the cutoff are not considered to be part of the search object. A set of transformations describing the search region (usually a translationally limited region of the map defining the asymmetric unit, but with angular sampling representing all orientations) is applied to the coordinates. This is a six-dimensional (three translation parameters plus three orientation parameters) search. At each position, the density values for equivalent samples, E_i, are interpolated from the EM map and the standard correlation coefficient is calculated.

(a)

(b)

Figure 2
GroEL apical and equatorial domains, atomic and density representations. Close-up of an all-atom representation of the (a) apical and (b) equatorial domains. The blue envelope surrounding each is an isosurface of the density (15 Å Fourier cutoff) derived from each. These low-resolution density maps are the search objects that are correlated with the cryo-EM map in order to find the positions of the domains in the EM map of the complete oligomer. It is apparent that at 15 Å sharp features are smoothed and side chains extending out from the surface are rounded off. The atomic domains and maps were displayed in *O*.
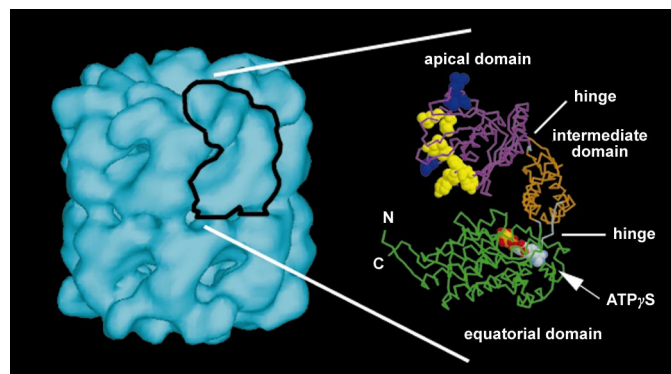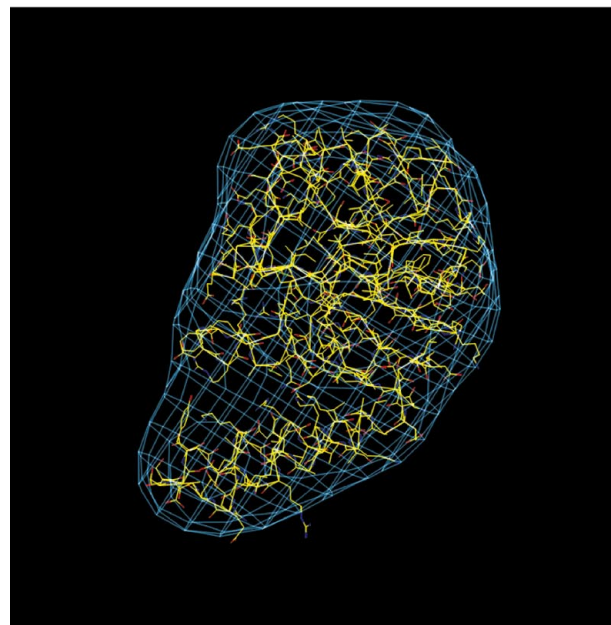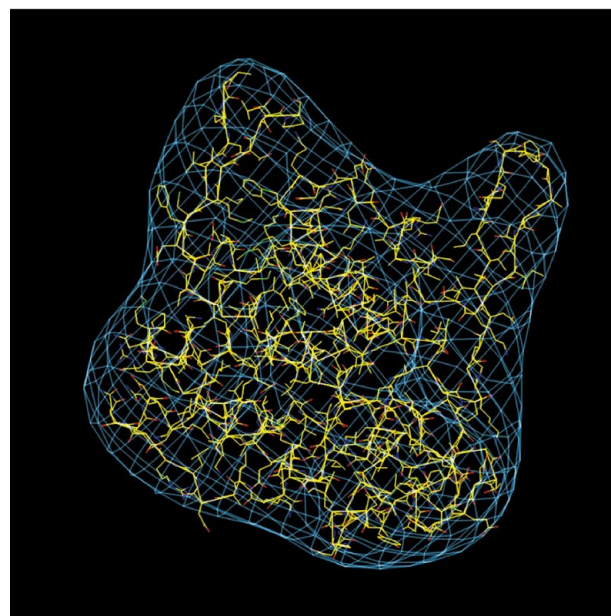

Figure 1
Structure of the GroEL subunit in relation to the oligomer. One GroEL subunit has been outlined in a 15 Å representation of the 14-mer (left) with an expanded view of the subunit from X-ray crystallography shown on the right. Three domains are shown colour coded: apical (purple), intermediate (orange) and equatorial (green). They are separated by two hinge regions, which are rich in conserved glycines (Fenton *et al.*, 1994). The domains are colour coded as defined for the docking trial. Regions of grey between them correspond to short loops between the domains that were not defined as part of any domain. The equatorial domain contains the ATP-binding site (shown here containing the non-hydrolysable ATP analogue ATPγS drawn in space-filling representation) and forms inter- and intra-ring contacts. The intermediate domain joins the equatorial and apical domains and contains residues involved in ATP hydrolysis. The apical domain contains residues required for substrate–protein and GroES binding, shown in space-filling representation in yellow and blue, respectively. The equatorial domain contains the amino (N) and carboxy (C) termini of the protein chain. The diameter of the oligomer is ∼140 Å. AVS was used to generate the isosurface of a 15 Å map calculated from the coordinates of the GroEL 14-mer (PDB accession number 1oel; Braig *et al.*, 1995). The subunit was drawn with *RASMOL*, using a subunit extracted from the coordinates of the GroEL–ATPγS complex (PDB accession number 1der; Boisvert *et al.*, 1996). Adapted from Roseman *et al.* (1996).

$$\text{Correlation} = \frac{1}{n} \sum_{i=1,n} \frac{(S_i - \overline{S})(E_i - \overline{E})}{\sigma_S \sigma_E},$$

where σ_S and σ_E are the standard deviations of the sets of samples S_i and E_i, respectively.

Maximizing the correlation coefficient is equivalent to minimizing the residual, calculated as the sum of squared differences between the two sets of samples, after the two sets of values have been scaled to the same mean and standard deviation,

$$\text{residual} = \sum_{i=1,n} (S_i - E_i)^2.$$

The program *DOCKEM* calculates the correlations systematically over the search region; the parameters for the 10 000 search positions with the highest correlation are saved to a file. This file is read by another program *DOCKXSOLN*, which generates the atomic models for the top-scoring positions. They are visualized in *O* and checked for compatibility with available constraints.

2.3. Stepwise summary of the procedure

(i) Read in the EM map and the atomic coordinates for the domains into *O*. Position the domains. Save the coordinates.

(ii) Convert the atomic coordinates into a simulated EM density to make the search object. Fourier filter to the chosen resolution.

(iii) Search for the best match of the search-object density (derived from the X-ray model) and the EM map density in the six-parameter search space using the program *DOCKEM*.

(iv) Generate the atomic coordinates of the positioned domains for the top scores. The program *DOCKXSOLN* performs this using the transformations corresponding to the highest correlations, which are output by the program *DOCKEM*.

(v) Examine the atomic domain positions overlaid with the EM map in *O*.

(vi) If relevant, generate symmetry-related domains and examine in *O*.
*CCP*4 programs are used to manipulate the coordinates.

The solutions generated at stage (iv) could be automatically filtered to match criteria, such as the requirement that symmetry-related subunits do not clash or do not clash with previously positioned domains. The top scores consistent with the constraints could then be presented for visual inspection.

3. Trial runs of docking domains into a cryo-EM map of GroEL

A cryo-EM map (Roseman, Ranson, Gowen & Saibil, in preparation) and atomic coordinates of the unliganded form of GroEL (Braig *et al.*, 1995) were used to test the procedure and programs. GroEL is a large oligomer composed of 14 60 kDa subunits arranged in two back-to-back heptameric rings (Fig. 1). Each subunit has three domains joined by short flexible loops containing conserved glycine residues (Fenton *et al.*, 1994). GroEL and its co-chaperone GroES are involved in helping other proteins to fold (Ellis, 1996). This involves cycles of binding and release of the substrate proteins from an internal cavity which are driven by ATP binding and hydrolysis. GroEL adopts a number of different conformations during the ATPase cycle (Roseman *et al.*, 1996). The changes in conformation are very likely to occur as rigid-body movements of the domains about the hinge regions, which is the case for the two different forms observed as crystal structures, apo GroEL (Braig *et al.*, 1994, 1995) and the GroEL–GroES–ADP complex (Xu *et al.*, 1997).

GroEL was selected because a crystal structure and a cryo-EM map of this large multi-domain complex were available, which allowed extensive tests of the docking method. The GroEL molecules crystallized were packed into the *C*222 space group, imposing a twofold symmetry axis between the two rings. In the final refinement, sevenfold NCS was not imposed and each subunit refined with a slightly different conformation. The most variable part was the apical domains, which exhibited a rigid-body motion about the hinge region. The root-mean-square (r.m.s.) deviation of the apical domain atoms from the average conformation was ∼1.5 Å per atom. The molecules crystallized were a double mutant deficient in negative cooperativity (Aharoni & Horovitz, 1996), whereas for electron microscopy wild-type GroEL was vitrified from solution. In the EM reconstruction sevenfold symmetry was strictly imposed but each heptameric ring was independently determined. No twofold symmetry between the rings was imposed.

The three domains of a GroEL subunit were extracted from the coordinates deposited by Braig *et al.* (1995) in the PDB (accession number 1oel). They were defined as the apical (residues 191–373), equatorial (residues 2–134 and 411–525) and the intermediate (residues 140–190 and 376–409) domains. A few residues in short loops connecting the domains were omitted from the models (Fig. 1).

The EM map was displayed in the program *O* and the coordinates of the domains imported. Since the EM map has sevenfold symmetry strictly imposed, the asymmetric unit is a 1/7 wedge of a hypothetical cylinder encompassing the molecule. For practical reasons, docking was performed separately for each heptameric ring of the EM map. Each of the three domains was roughly positioned in the centre of the asymmetric unit of the map and the coordinates were saved. The asymmetric unit must be defined to contain a continuous subunit density. No 'wrap-around' of density within the asymmetric unit occurs because symmetry is not built into the search. Six density-search objects were then generated from the coordinates, since there are three different domains to be docked into each of the two heptameric rings. Fig. 2 shows the calculated 15 Å density-search objects overlaid on the coordinates of the apical and equatorial domains.

The program *DOCKEM* was run to find the best match between each model domain and the EM map. The density matching was performed with the maps low-pass filtered at 15

or 12 Å. The high-pass filter was set at 300 Å. The translation search covered the complete asymmetric unit with a step size of 3.62 Å. At each position a complete search of angular space was performed, sampling with a step size of 4°. It was found that when the docking trials were repeated from slightly different start positions in the asymmetric unit, the results for the smaller intermediate domain were sensitive to the initial placement. Reducing the translational sampling to 1.81 Å steps enabled the correct position to be indicated by the highest correlation coefficient in every case. This finer search was performed reducing the translation-search region by 30% in each dimension in order to save computing time. In a *de novo* application, such a restriction of the search region for a smaller domain would be justified if the larger domains were positioned first. In GroEL the positioning of the apical and equatorial domains defines the position of the smaller intermediate domain quite accurately. An alternative way to reduce computation time would be to perform fine searches in local regions around top-scoring positions from a coarse search. This is a consideration because the run time for one of these trials at 3.62 Å sampling covering the complete asymmetric unit is 1–2 d on a DEC alpha computer with an EV6 processor. Later test runs were also performed at 20, 25 or 30 Å. The transformations output by *DOCKEM* were applied to the coordinates and were viewed in *O*.

A model of a subunit in each ring was made by concatenating the coordinates for each of the three docked domains. Sevenfold symmetry-related copies were generated in order to create a model of the heptameric ring using the *CCP*4 (Collaborative Computational Project, Number 4, 1994) program *PDBSET*. This model of the heptameric ring and the crystal structure were aligned, minimizing the r.m.s. difference
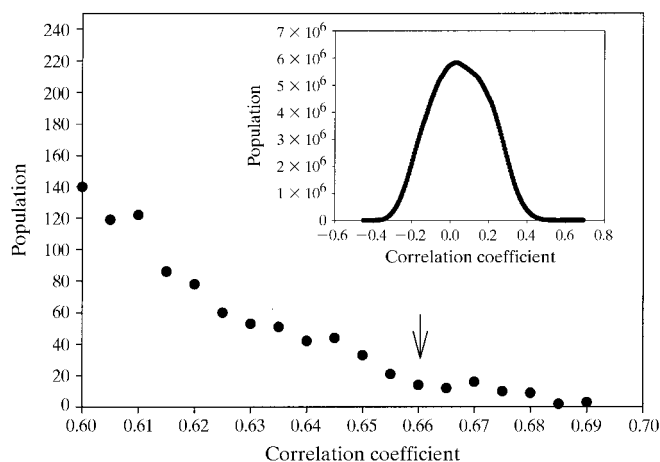
Table 1
Correlation coefficients (%) and their significance (as standard deviations from the mean of all correlation coefficients calculated in the asymmetric unit search) for the six domains docked.

	15 Å	12 Å
Apical domain (~20 kDa)		
Open ring	68.9 (4.2)	54.9 (6.3)
Closed ring	64.3 (3.8)	52.2 (5.7)
Intermediate domain (~9 kDa)		
Open ring	72.2 (3.5)	70.0 (4.8)
Closed ring	72.1 (3.4)	64.7 (4.7)
Equatorial domain (~26 kDa)		
Open ring	68.4 (5.1)	53.8 (8.7)
Closed ring	67.7 (5.7)	50.6 (7.9)

of C^α-atom positions, using the program *LSQMAN* (Kleywegt, 1996). The domainwise r.m.s. difference of C^α positions between the crystal structure and the model of the subunit made by docking was then calculated, also with *LSQMAN*.

4. Results and discussion

The models of the heptameric rings of complete subunits generated from the docked domain positions were not very different from the crystal structure, indicating that the domains were correctly docked into the EM map. This was the case for docking at 15 and 12 Å. In each case the highest correlation indicated the correct position. No other constraints were required. At 15 Å the highest correlations were between 3.4 and 5.7 standard deviations from the mean of all correlations evaluated. The correlations are more significant when using 12 Å densities, with the highest correlations between 4.7 and 8.7 standard deviations above the mean, though the actual correlation values are lower than for 15 Å (Table 1). This may be because the EM map is noisier at 12 Å than 15 Å or because an inaccuracy in the magnification assumed for the EM map has greater effect at higher resolution.

A histogram of all the correlations for the docking of the apical domain at 15 Å is bell shaped, with a small extension at the high correlation end (Fig. 3, inset). The family of highest correlations all represent very similar solutions. The top correlation coefficient is 0.69. The highest correlation that corresponds to a position significantly different from the top score is 0.66, 0.2 standard deviations below the maximum (Fig. 3). The docked positions are shown for all six domains in Fig. 4. There is no clash between the different docked domains or between symmetry-related copies of any domain and any other domain. Fig. 5 shows an overlay of the C^α representation of the crystal structure and the docked domain positions.

The r.m.s. difference of C^α atoms between the model created by docking and the crystal structure was calculated for each domain (Table 2). The domains docked into the ring most similar to the crystal structure had r.m.s. differences in position of ≤ 4 Å. These figures are good considering that the coarseness of the search was 3.6 Å and 4°. The movement of the apical domain in the other ring is significantly greater than
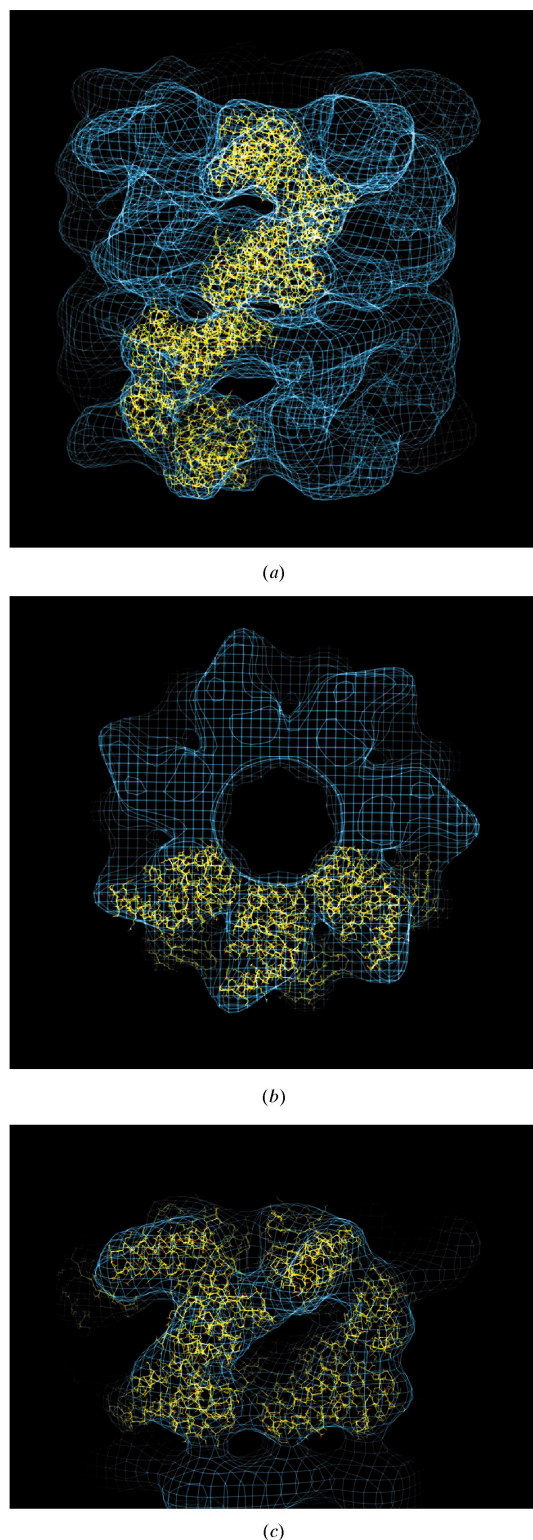

Figure 3
Histogram of all correlation coefficients computed in the asymmetric unit for apical domain docking at 15 Å (inset) and enlarged view of the highest correlations. The family of highest correlations all represent the correct position. The arrow marks the highest correlation that is significantly different (>3.6 Å r.m.s. difference in atom positions and tilted by more than 10°) from the top-scoring position. This solution is still part of the shoulder of the top peak of solutions. It is tilted by 14° from the top-scoring solution and the r.m.s. difference between corresponding C^α atoms is 6.2 Å.

(a)

(b)

(c)

Figure 4
Three views of the crystallographic domains docked into the EM map (15 Å Fourier cutoff) are shown. (*a*) View from the side of the six independently docked domains (the two rings of the map were independently calculated and are not identical). (*b*) End view looking along the sevenfold axis at the fit of the apical domain and two adjacent symmetry-related domains. They do not clash into one another and there is a reasonable interface between them. (*c*) Close-up view of two adjacent complete subunits. There are no clashes between different domains or their symmetry-generated counterparts. The maps and coordinates were displayed in *O*.

the sampling and a small conformation change has been detected. They have opened outwards compared with the crystal structure. The differences in the other domain positions may not be significant at this level of sampling in the docking search.

The differences between the model of the oligomer made by docking and the crystal structure are quite small and the comparison shows that there is no gross error in the docking positions assigned. If the crystal structure of the oligomer was not known, then other criteria could have been used to establish confidence that a sensible conformation of the subunit had been modelled. The fact that the domains do not clash and that the hinge regions can be appropriately connected are good indicators.

Docking was only partially successful when it was attempted at lower resolutions. The significance of the highest correlation for the domain docking is plotted for maps at 12, 15, 20 or 25 Å for cases where this indicated the correct position of the domain (Fig. 6). Though the highest correlation indicated the correct placement for all the domains at 12 and 15 Å, only some were located correctly at 20 or 25 Å. None of the highest correlations indicated the correct position at 30 Å. The significance of correlation coefficients is greater for larger domains and for density matching at higher resolutions.

The apical domains could be docked at 20 Å using only the local correlation coefficient. The docking of the equatorial domains at 20 Å was not so successful because the packing between the equatorial domains is tighter and the boundary between them is less well defined. The incorrect solution given by the highest correlation could be rejected because it is incompatible with the packing into a heptameric ring. Therefore, it may be possible to achieve docking at lower resolutions than 15 Å if suitable information is available that can be applied as constraints in order to filter the possible
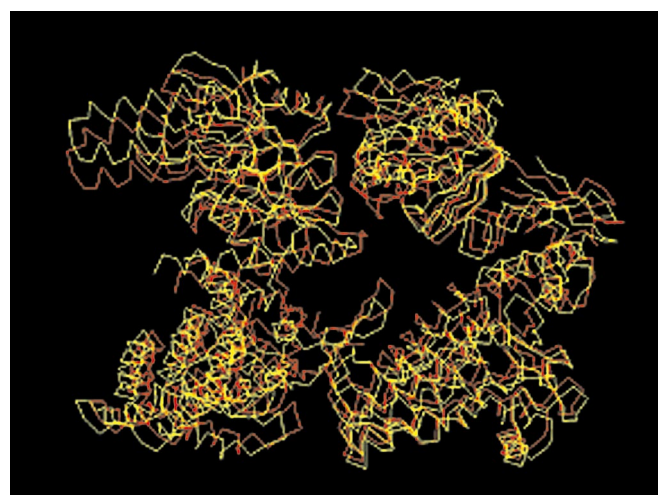

Figure 5
Overlay of the model oligomer of GroEL subunits generated from the docked domain positions (yellow) and the crystal structure (red), showing two neighbouring subunits from the heptameric ring. Both are shown as C^{α} traces generated in *O*. The docking program has produced a model of the heptameric ring very similar to the crystal structure conformation. The model shown is of the more closed ring.

Table 2
R.m.s. deviation of C$^\alpha$ atoms (Å) between the heptameric ring model from the docked domain positions and crystal structure, per domain.
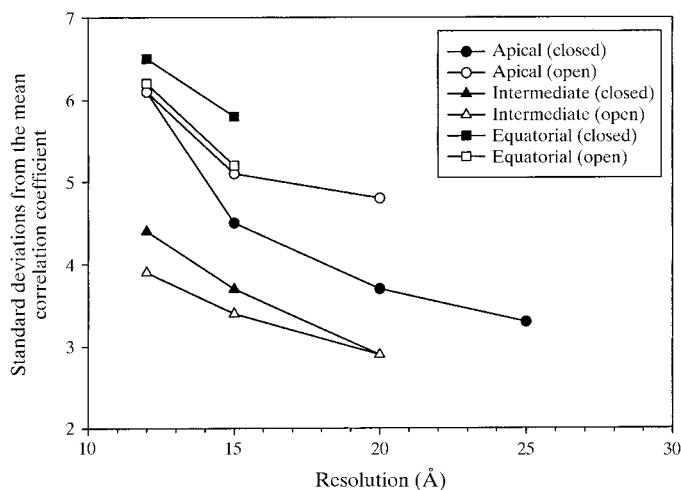
	Closed ring	Open ring
Apical domain	2.6	5.5
Intermediate domain	3.6	3.7
Equatorial domain	4.0	1.5

solution set. At 15 Å there were enough features for the correlation to indicate the correct solution.

Now that the method has been shown to work using a coarse sampling, an improvement would be to interpolate peaks more accurately or run the search locally around best solutions with a finer step. The small movements could then be interpreted with more confidence. Another improvement could be to include magnification of the electron microscope as a parameter searched. The program may also have application in docking structures into low-resolution maps of large complexes obtained by X-ray crystallography.

5. Conclusions

The programs and algorithm have been demonstrated to work using real cryo-EM data, docking the atomic coordinates of domains of three different shapes and sizes into the cryo-EM map of GroEL. The equatorial (26 kDa), apical (20 kDa) and intermediate (9 kDa) domains were all placed in the correct position. The significance of the correlation coefficient was higher for the larger domains. A resolution of 15 Å or better was required in the EM map for the correct position of all the domains to be identified by the highest correlation. The highest correlation for some of the docking trials performed at

lower resolutions corresponded to false solutions that could be rejected using the sevenfold symmetry (symmetry-generated domains clashed). Therefore, other constraints could help to identify correct solutions if only a lower resolution EM map was available.

These trials of the procedure on the GroEL map are very encouraging. The docked domains did not clash and were compatible with connectivity criteria, producing a self-consistent solution. Once a model is built it should be tested against all available constraints, such as symmetry, connectivity, amino- or carboxy-terminus orientation and position of known epitopes.

Favourable results have been obtained using a trial structure where the solution was known. The real value will be demonstrated when the program is used on unknown structures, but then the validity of the resulting model will have to be tested by mutational or other experiments. These will offer indirect proof and confirm that correct hypotheses have been derived from the model generated. Ultimately, the larger complex may be solved by EM at high enough resolution to follow the fold of the protein or by X-ray crystallography. This absolute verification is likely to take longer and the advantage of the docking approach is that an atomic model can be made sooner and may be useful for suggesting other kinds of experiments. The program is available from the author on request: roseman@mrc-lmb.cam.ac.uk.

Figure 6
The significance of the highest correlation coefficient for the domain docking is plotted for docking at 12, 15, 20 or 25 Å resolution for cases where this indicated the correct position of the domain. For these calculations, the translation search was restricted to a cube of dimension 25 Å approximately centred around the correct position of the domain. The step sizes were 3.62 Å and 4°. All orientations were searched. In general, the significance improves with docking of larger domains and at higher resolution.

References

Agrawal, K., Penczek, P., Grassucci, R. A. & Frank, J. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 6134–6138.
Aharoni, A. & Horovitz, A. (1996). *J. Mol. Biol.* **258**, 732–735.
Baker, T. S., Olson, N. H. & Fuller, S. D. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 862–922.
Bella, J., Kolatkar, P. R., Marlor, C. W., Greve, J. M. & Rossmann, M. G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4140–4145.
Berriman, J. & Unwin, P. N. T. (1994). *Ultramicroscopy*, **56**, 241–252.
Boisvert, D. C., Wang, J., Otwinowski, Z., Horwich, A. L. & Sigler, P. B. (1996). *Nature Struct. Biol.* **3**, 170–177.
Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.
Braig, K., Adams, P. D. & Brünger, A. T. (1995). *Nature Struct. Biol.* **2**, 1083–1094.
Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). *Nature (London)*, **371**, 578–586.
Brünger, A. T. (1992). *X-PLOR. Version* 3.1. *A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.
Che, Z., Olson, N. H., Leippe, D., Lee, W., Mosser, A. G., Rueckert, R. R., Baker, T. S. & Smith, T. J. (1998). *J. Virol.* **72**, 4610–4622.

Cheng, R. H., Kuhn, J. K., Olson, N. H., Rossmann, M. G., Hok-Kin, C., Smith, T. J. & Baker, T. S. (1995). *Cell*, **80**, 621–630.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A. W. & Schultz, P. (1988). *Quart. Rev. Biophys.* **21**, 129–228.

Ellis, R. J. (1996). Editor. *The Chaperonins.* San Diego: Academic Press.

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

Frank, J. (1990). *Quart. Rev. Biophys.* **23**, 281–329.

Frank, J. (1996). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies.* San Diego: Academic Press.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). *J. Struct. Biol.* **116**, 190–199.

Gabashvili, I. S., Agrawal, R. K., Grassucci, R. & Frank, J. (1999). *J. Mol. Biol.* **286**, 1285–1291.

Grimes, J. M., Jakana, J., Ghosh, B., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I. & Prasad, B. V. V. (1997). *Structure*, **5**, 885–893.

Hanein, D., Volkmann, N., Goldsmith, S., Michon, A. M., Lehman, W., Craig, R., DeRosier, D., Almo, S. & Matsudaira, P. (1998). *Nature Struct. Biol.* **5**, 787–792.

Heel, M. van & Frank, J. (1981). *Ultramicroscopy*, **6**, 187–194.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. (1990). *J. Mol. Biol.* **213**, 899–929.

Hewat, E. A. & Blaas, D. (1996). *EMBO J.* **15**, 1515–1523.

Hewat, E. A., Verdaguer, N., Fita, I., Blakemore, W., Brookes, S., King, A., Newman, J., Domingo, E., Mateau, M. G. & Stuart, D. I. (1997). *EMBO J.* **16**, 1492–1500.

Hirose, K., Lowe, J., Alonso, M., Cross, R. A. & Amos, L. A. (1999). *Mol. Biol. Cell*, **10**, 2063–2074.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kikkawa, M., Okada, Y. & Hirokawa, N. (2000). *Cell*, **100**, 241–252.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Kolatkar, P. R., Bella, J., Olson, N. H., Bator, C. M., Baker, T. S. & Rossmann, M. G. (1999). *EMBO J.* **18**, 6249–6259.

Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). *Nature (London)*, **367**, 614–621.

Liu, H., Smith, T. J., Lee, W. M., Mosser, A., Rueckert, R. R, Olson, N. H., Cheng, R. H. & Baker, T. S. (1994). *J. Mol. Biol.* **240**, 127–137.

Mendelson, R. & Morris, E. P. (1994). *J. Mol. Biol.* **240**, 138–154.

Mendelson, R. & Morris, E. P. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 8533–8538.

Musacchio, A., Smith, C. J., Roseman, A. M., Harrison, S. C., Kirchhausen, T. & Pearse, B. M. F. (1999). *Mol. Cell*, **3**, 761–770.

Nogales, E., Wolf, S. G. & Downing, K. H. (1998). *Nature (London)*, **391**, 199–203.

Olson, N. H., Kolatkar, P. R., Oliveira, M. A., Cheng, R. H., Greve, J. M., McClelland, A., Baker, T. S. & Rossmann, M. G. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 507–511.

Rayment, I., Holden, H. M, Whittaker, M., Yohn, C. B., Lorenz, M., Holmes, K. C. & Milligan, R. A. (1993). *Science*, **261**, 58–65.

Rice, S., Lin, A. W., Safer, D., Hart, C. L., Naber, N., Carragher, B. O., Cain, S. M., Pechatnikova, E., Wilson-Kubalek, E. M., Whittaker, M., Pate, E., Cooke, R., Taylor, E. W. & Milligan, R. A. (1999). *Nature (London)*, **402**, 778–784.

Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. (1996). *Cell*, **87**, 241–251.

Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. & Horwich, A. L. (1999). *Cell*, **97**, 325–338.

Schoehn, G., Quaite-Randall, E., Jiménez, J. L., Joachimiak, A. & Saibil, H. R. (2000). *J. Mol. Biol.* **296**, 813–819.

Schröder, R. R., Manstein, D. J., Jahn, W., Holden, H., Rayment, I., Holmes, K. C. & Spudich, J. A. (1993). *Nature (London)*, **364**, 171–174.

Smith, T. J., Chase, E. S., Schmidt, T. J., Olson, N. H. & Baker, T. S. (1996). *Nature (London)*, **383**, 350–354.

Smith, T. J., Olson, N. H., Cheng, R. H., Liu, H., Chase, E., Lee, W. M., Leippe, D. M., Mosser, A. G., Ruekert, R. R. & Baker, T. S. (1993). *J. Virol.* **67**, 1148–1158.

Stark, H., Rodnina, M. V., Rinke-Appel, J., Brimacombe, R., Wintermeyer, W. & van Heel, M. (1997). *Nature (London)*, **389**, 403–406.

Stewart, P. L., Chiu, C. Y., Huang, S., Muir, T., Zhao, Y., Chait, B., Mathias, P. & Nemerow, G. R. (1997). *EMBO J.* **16**, 1189–1198.

Stewart, P. L., Fuller, S. D. & Burnett, R. M. (1993). *EMBO J.* **12**, 2589–2599.

Volkmann, N. & Hanein, D. (1999). *J. Struct. Biol.* **125**, 176–184.

White, H., Chen, S., Roseman, A., Yifrach, O., Horovitz, A. & Saibil, H. (1997). *Nature Struct. Biol.* **4**, 690–694.

Wikoff, W. R., Wang, G., Parrish, C. R., Cheng, R. H., Strassheim, M. L., Baker, T. S. & Rossmann, M. G. (1994). *Structure*, **2**, 595–607.

Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). *J. Struct. Biol.* **125**, 185–195.

Wynne, S. A., Crowther, R. A. & Leslie, A. G. W. (1999). *Mol. Cell*, **3**, 771–780.

Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). *Nature (London)*, **388**, 741–750.