research papers

Combining X-ray and electron-microscopy data to solve crystal structures

Jorge Navaza

Laboratoire de Microscopie Electronique
Structurale, Institut de Biologie Structurale
Jean-Pierre Ebel, 41 Rue Jules Horowitz,
F-38027 Grenoble, France

Correspondence e-mail: jorge.navaza@ibs.fr

Low-resolution electron-microscopy reconstructions can be used as search models in molecular replacement or may be combined with existing monomeric structures in order to produce multimeric models suitable for molecular replacement. The technique is described in the case of viral and subviral particles as well as in the case of oligomeric proteins.

1. Introduction

Transmission electron microscopy (TEM) and X-ray crystallography are two complementary techniques used in structural biology. TEM fascinates by its apparent simplicity in the visualization of isolated biological systems in almost *in vitro* conditions. The observed systems are in general large assemblies of macromolecular structures, for example viruses and viral particles. The main limitation of the technique is the rather low resolution of the images produced, which is usually in the nanometre range.

On the other hand, X-ray crystallography routinely determines at atomic resolution the structures of individual proteins or complexes involving a limited number of proteins. The crystallographic phase problem is solved either experimentally (isomorphous replacement and related techniques) or by numerical methods in which information from previously determined molecular structures is efficiently used (molecular replacement). However, in the case of the very large complexes that are now crystallized, phasing by isomorphous replacement is often difficult and frequently atomic models do not exist. In these cases, EM and X-ray data may be combined to start the process of crystal structure determination. Indeed, an initial crystal structure may be obtained by molecular replacement (MR) based on a low-resolution EM reconstruction (Dodson, 2001). Phases are then extended by density-modification techniques, in particular solvent flattening and noncrystallographic symmetry averaging. This has been performed in the case of large icosahedral particles (viruses), where the high symmetry is a guarantee of success in the phase-extension process. Moreover, the model is usually obtained by cryo-TEM, a technique that already provides a good representation of the isolated particles that constitute the crystal.

More recently, we have applied the same technique to oligomeric proteins, a classic problem in protein crystallography. The particle sizes were too small to use cryo-TEM, so the three-dimensional reconstructions were performed using negatively stained samples. The main problems to solve were linked to the solvent contribution to the observed structure

factors at low resolution, which is the range of resolution at which EM and X-ray data overlap.

However, EM and X-ray data may also be combined in another way: very often X-ray crystallography determines the structures of the individual proteins that constitute the assemblies whose low-resolution reconstructions have been determined by TEM. It is then possible to interpret the EM map in terms of atomic models, which provides considerable complementary information to molecular biologists. This is achieved by docking individual molecules into the EM reconstruction, a technique related to molecular replacement, although a simpler one as the role of observed structure factors is played by the complex Fourier coefficients of the EM map, so that phases are now at our disposal. In crystallography the docking technique is most useful in the case of oligomeric structures or protein assemblies for which only approximate models for the individual monomers are available, as the composite models thus obtained are usually good search models for MR.

The combined use of EM and X-ray data to solve crystal structures will be illustrated with some applications to viral and subviral particles and some multimeric proteins.

2. Use of cryo-TEM reconstructions to solve crystal structures

Customarily, X-ray diffraction data is collected at 'high' resolution, usually starting between 15 and 10 Å, as the low-resolution part of the spectrum contains the solvent contribution, which is only qualitatively explained by models and is not essential to recover the sought macromolecular atomic structures. On the other hand, the resolution of EM reconstructions is typically within the same 15–10 Å range. Also, in cryo-TEM what is reconstructed is the perturbation in scattering density arising from the particle relative to the background vitreous-ice scattering density. In the EM case this background plays the role of the solvent in the X-ray case. Therefore, in order to use EM maps as search models in molecular replacement, we must, in principle, measure low-resolution reflections to increase the overlap of resolution, subtract the bulk-solvent contribution from the X-ray spectrum and subtract the background contribution from the EM map.

At low resolution, we may invoke Babinet's principle to estimate the solvent contribution to the structure factors. In this approximation the contribution is taken into account by two parameters: a Gaussian resolution dampening factor, B_{sol}, and the ratio between the average solvent density and the average protein density, K_{sol}, entering into the expression

$$F_{obs}(\mathbf{h}) = F_{particle}(\mathbf{h})[1 - K_{sol} \exp(-B_{sol}|\mathbf{h}|^2)]. \quad (1)$$

However, when optimized against experimental data, the physical meaning of the bulk-solvent correction parameters is often lost (Glykos & Kokkinidis, 2000). A better approximation, along the same line of ideas, is discussed in Fokine & Urzhumtsev (2002) for refinement procedures.

By analogy to the TEM case, we may consider the X-ray structure factors as corresponding to a density where the average value of the bulk-solvent contribution is zero. This amounts to subtracting a constant value from the crystal electron density which affects only the $F_{obs}(0)$ coefficient. Thus, disregarding the Gaussian resolution dampening factor, we must simply subtract a 'solvent minus background' constant from the X-ray intensities, which we expect to be a small correction in a first approximation.

The use of maps instead of coordinates in molecular replacement is not a problem. Indeed, the *TABLING* program of *AMoRe* (Navaza, 1994) will read the map and compute its finely sampled Fourier coefficient. These will be the molecular scattering factors from which crystal structure factors will be calculated by interpolation for current values of the positional parameters of the molecules within the crystal. The only precaution that must be taken is to set the background value of the map to zero.

We applied the technique to the $T = 1$ subviral particle ($\sim$250 Å diameter) of IBDV, constituted of VP2, the only component of the virus icosahedral capsid (Coulibaly *et al.*, 2005). The hexagonal $P6_3$ unit cell has parameters $a = 258.950$, $b = 258.950$, $c = 347.265$ Å, $\gamma = 120°$. If the particles are accommodated in a close-packed configuration they would occupy $\pi/18^{1/2}$ of the available volume. With one particle per unit cell the particle diameter would be 305.5 Å, with two particles 242.5 Å and with three particles 211.8 Å; the unit cell thus contains two viral particles. Moreover, since there are six space-group symmetry operations, only one-third of the viral particle is crystallographically independent. Therefore, one of the particles, or more precisely its centre of mass, must lie on one of the two threefold-site symmetry positions, (1/3, 2/3, z) or (0, 0, z), in fractional coordinates (*International Tables For*
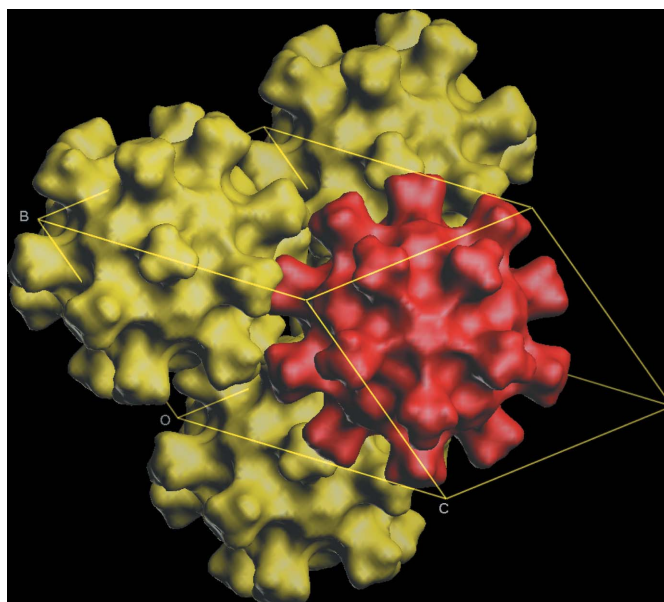

Figure 1
Crystal packing of the subviral particle of IBDV constituted of VP2 trimers, obtained by molecular replacement using the cryo-TEM map as search model.

Crystallography, 2006). The last position can be ruled out because there is no room for two particles along the *c* axis. Although this information was not used in the molecular-replacement procedure, it allowed us to quickly assess the putative solutions.

The cryo-TEM reconstruction was obtained with the *RIco* suite of programs which uses icosahedral symmetry-adapted functions to represent the scattering density (Navaza, 2003). This enhances the signal-to-noise ratio and permits model-independent reconstructions. The TEM micrographs were produced by Jean Lepault on a CM120 electron microscope. The estimated EM map resolution was about 20 Å. As the crystallographic data were complete starting at 50 Å, reflections in the 50–20 Å resolution range were used for MR. The correct solution was well contrasted, with a correlation co-efficient of 53% compared with 37% for the first incorrect position. The oriented and translated model is presented in Fig. 1 together with symmetry-related particles to show the crystal packing. This molecular-replacement solution can be used to start the phase-extension process by noncrystallo-graphic symmetry averaging. The icosahedral subviral particle thus determined contains 20 of the trimers shown in Fig. 2.

3. Fitting atomic models into EM reconstructions

The problem of fitting an atomic model into an EM map may be addressed using the ideas of molecular replacement. Some important differences exist between the two problems which impose modifications of the conventional molecular-replacement programs. The most important differences are:

(i) phase information is available,

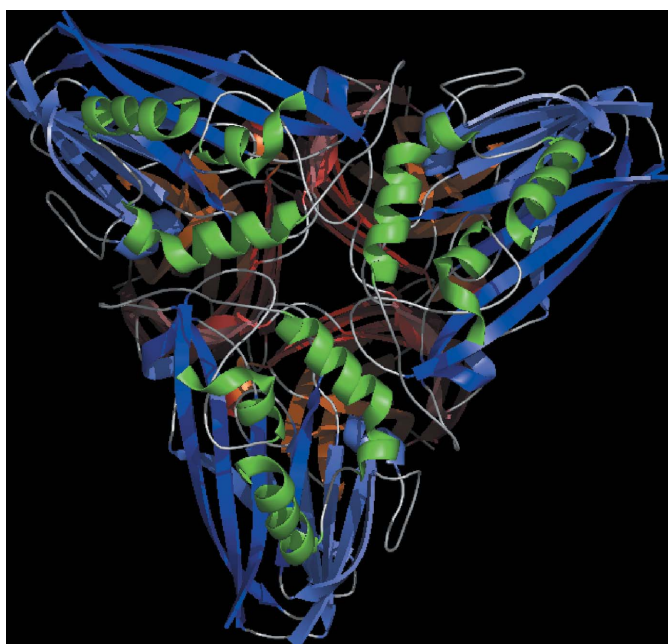(ii) the symmetry of the EM reconstructions is not in general the point group of a crystallographic symmetry,

(iii) EM images suffer from lack of resolution (typically below 15 Å) and low signal-to-noise ratio.

A consequence of the low resolution of the EM maps is that the boundaries of individual molecules cannot be easily defined, so that the extraction of volumes containing single molecules is inevitably biased. It is then necessary to consider large volumes containing several copies of the independent molecules without making any assumption concerning their shapes. These copies are related by the symmetry imposed during the EM reconstruction (icosahedral, helical, point-group symmetry), so that equivalent molecules will, in general, be sampled at non-equivalent grid points. Indeed, most of the imposed symmetries are not compatible with equally spaced Cartesian grids, the standard way in which EM maps are presented. Therefore, even if the information content of a whole map is the same as that of its asymmetric part, the form in which data are available determines the procedures to use the information in an efficient way.

The image of the biological assembly can be used to guess initial positions of the search models. Hence, the fitting problem can be reduced to the application of a rigid-body refinement protocol starting from putative locations of the model molecules, instead of performing exhaustive six-dimensional searches or separate rotational and translational searches as in the standard MR procedure. Indeed, phase information is straightforwardly derived from the EM recon-struction; its presence dramatically increases the radius of convergence of the refinement procedures compared with the standard phaseless MR case. The rigid-body refinement program *FITING* from the *AMoRe* suite of programs has been adapted to take into account noncrystallographic symmetry and phase information. The procedure for fitting molecular
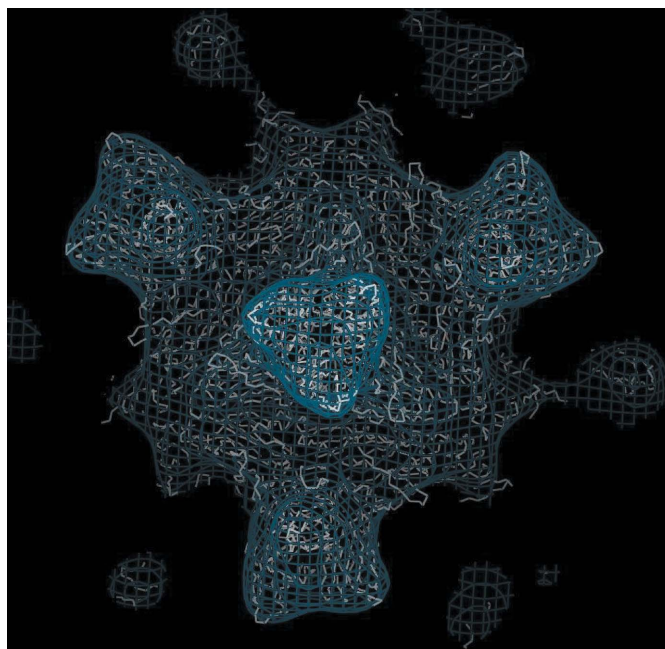

Figure 2
Top view of the VP2 trimer.

Figure 3
Fit of VP2 trimers into the cryo-TEM reconstruction of the *T* = 1 subviral particle. View along a threefold axis.

models into EM reconstructions has been implemented in a package called *URO* (Navaza *et al.*, 2002).

In Fig. 3, we show the results of applying *URO* to the $T = 1$ subviral particle of IBDV. The independent molecular model was taken as the VP2 trimer. 20-fold symmetry was imposed during the fitting procedure. The resulting correlation between the complex Fourier coefficients of the EM and the model based maps is 77.9% when using data in the 400–20 Å resolution range.

4. Use of pseudo-atomic models derived from EM reconstructions to solve crystal structures

The intact $T = 13$ IBDV (700 Å diameter) crystallizes in space group $P2_1$, with unit-cell parameters $a = 854.0$, $b = 692.2$, $c = 792.4$ Å (Coulibaly *et al.*, 2005). Its cryo-TEM reconstruction at about 20 Å resolution was obtained with the *RIco* program based on micrographs produced by Jean Lepault. The crystal structure was solved by MR, but instead of the EM map we used the pseudo-atomic model resulting from the fit of the VP2 trimers into the EM map. The result of this fit is shown in Fig. 4; the correlation coefficient of this fit was 91.0% when using Fourier coefficients within the 400–35 Å resolution range. The virion capsid contains 260 VP2 trimers, of which only four and a third belong to the icosahedral asymmetric unit.

AMoRe gave a clear MR solution using data in the 50–15 Å resolution shell. At the output of the translation function there were 30 positions, with correlations ranging between 42.1 and 36.4%, which eventually refined to the correct solution. The first incorrect solution had a correlation of 28.1%. Using the *FITING* program, the 260 VP2 trimers were refined as rigid
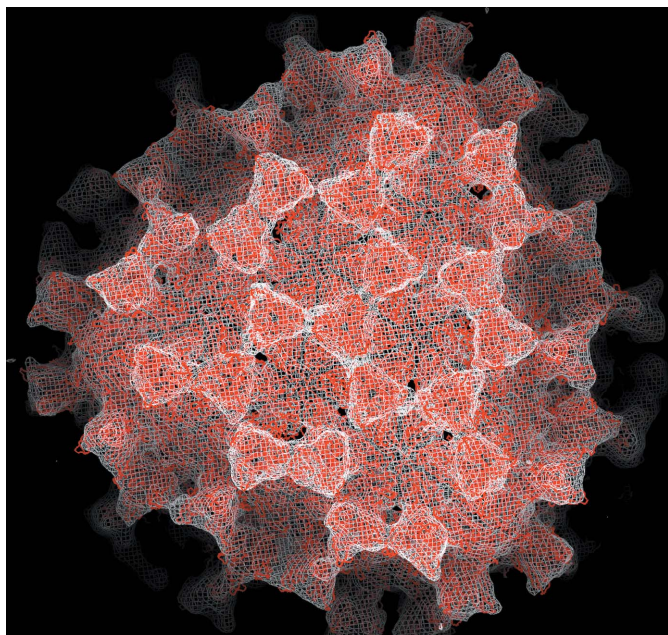
bodies against data up to 7 Å resolution to correct possible magnification errors in the EM reconstruction. 60-fold noncrystallographic symmetry averaging was then performed for density modification and model-bias removal. The refined atomic model is shown in Fig. 5; its correlation with the EM map is 82.1% within the 400–35 Å shell.
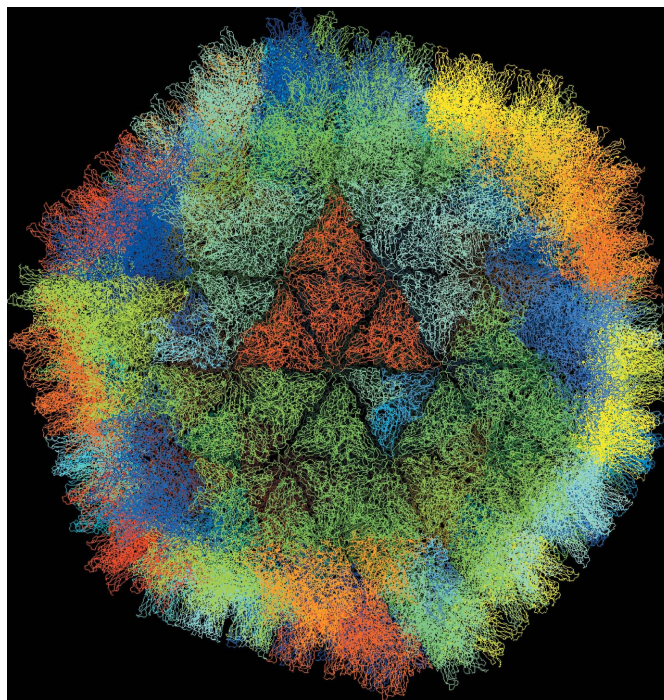

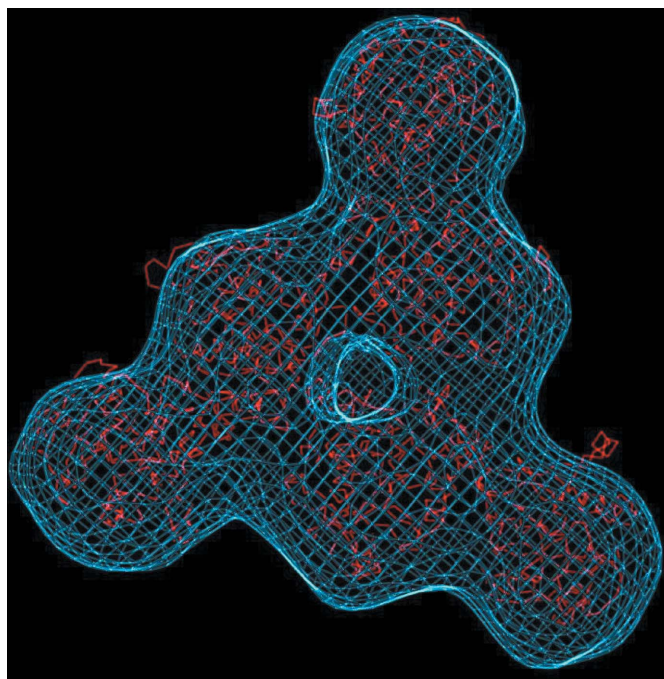**Figure 5**
Atomic model of IBDV based on 260 VP2 trimers.

Figure 4
Fit of the VP2 trimers into the cryo-TEM reconstruction of the $T = 13$ viral particle.

Figure 6
Fit of *Agrobacterium tumefaciens* glycogen synthase monomers into the negative-stain EM reconstruction of PaGS.

5. Negative-stained multimeric proteins as search models for MR

When the size of the particles is small, cryo-TEM images have such a small contrast that particles cannot be differentiated from the background. In such cases the technique of choice is negative staining (NS), which is usually performed with heavy-metal salts. Besides the contrast, the advantage of NS is the low radiation damage. The main drawbacks are the particle distortion as part of the staining process and the limited resolution of the images (~20 Å or lower).

NS can be used with particles greater than about 100 kDa. Some proteins, in particular multimeric proteins, are larger than this so that in principle NS-TEM reconstructions could be used as search models for molecular replacement. We have already studied the feasibility of the technique with *Pyrococcus abyssi* glycogen synthase (PaGS; Trapani *et al.*, 2006). In this case, owing to the lack of low-resolution data, a homologous monomer was fitted into the NS-TEM map and the resulting trimer (Fig. 6) could be used successfully in MR.

A more interesting case was 3-dehydroquinate dehydratase from *Candida albicans* (~170 kDa), a tetrahedral dodecameric protein which crystallizes in the orthorhombic $P2_12_12$ space group, with unit-cell parameters a = 159.1, b = 308.1, c = 97.1 Å (GenBank gi:68482603). The molecular-replacement solution was difficult to obtain. As there are two independent molecules, the orientation of the second molecule was forced to satisfy the rotational NCS, which allowed us to test many putative first-molecule positions. The NS-TEM reconstruction was calculated by Guy Schoehn; its resolution was estimated to be about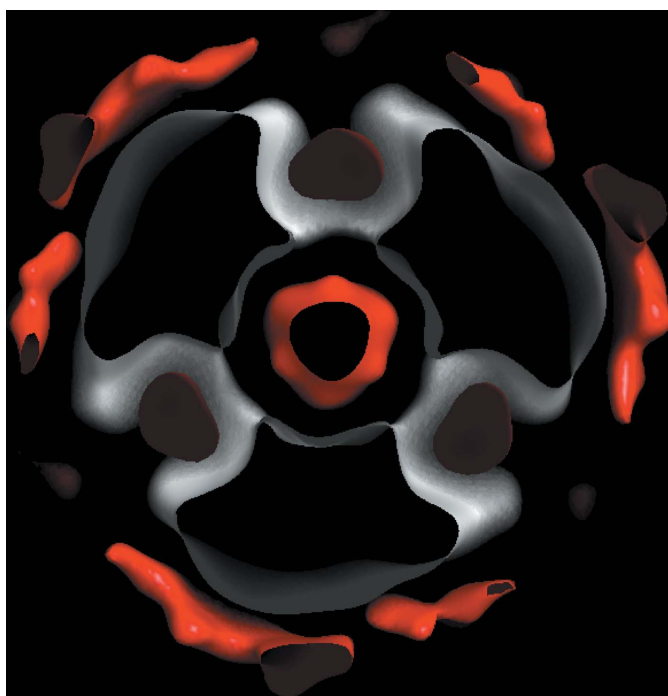 15 Å, a rather high resolution for the technique, achieved by carefully selecting a few thousand images. Usually, NS-TEM maps have important negative regions at the surface of the particle, where the heavy-metal salts are deposited, as shown in Fig. 7. The values of the scattering density function in these regions were set to the average background value. Although we had rather complete diffraction data starting at 40 Å, a contrasted solution was eventually found by working with data in the 25–15 Å resolution shell and by enhancing the high-resolution part of the model's spectrum by dividing by a temperature-like Gaussian factor. The correlation coefficient of the solution was 40.6% and that of the first wrong position was 36.4%. The low-resolution crystal structure thus obtained (Fig. 8) allowed us to obtain phases up to 2.8 Å without any atomic model intervention (Trapani *et al.*, 2008).

6. Conclusions

We have discussed and illustrated with examples how EM and X-ray data may be combined in order to build a search model suitable for molecular-replacement calculations. Our results
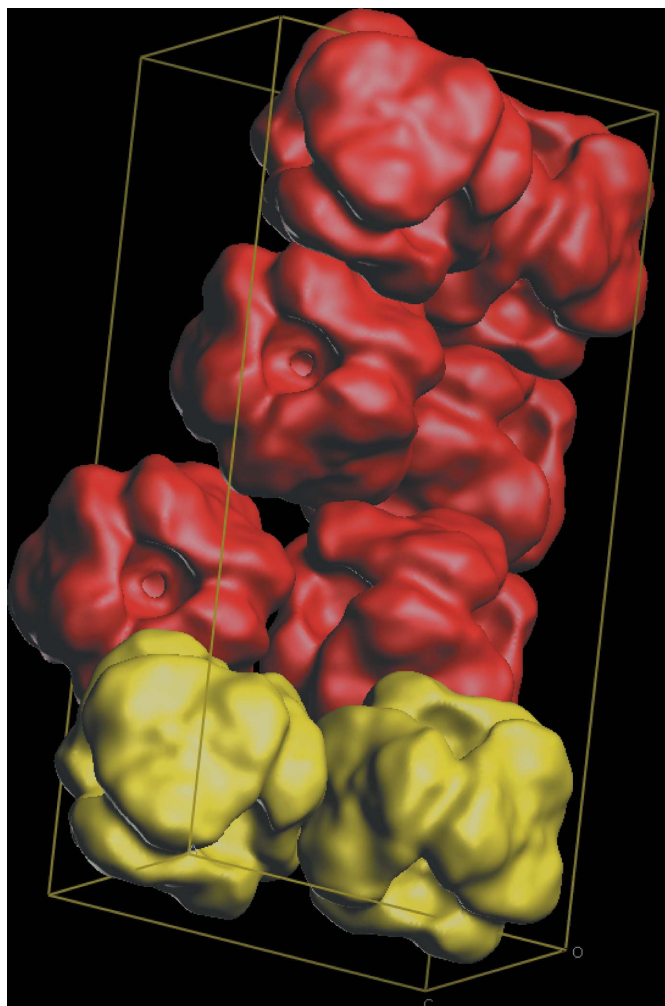

Figure 7
Sections of the negative-stain EM reconstruction of the dehydroquinase at 1σ (white) and -1σ (red).

Figure 8
Crystal structure of the dehydroquinase based on the NS-TEM reconstruction.

also suggest that EM reconstructions obtained by negative staining, a technique of wider applicability than cryo-EM, are suitable models for medium-size proteins (greater than ~100 kDa), provided that complete low-resolution diffraction data are available. Very low resolution spectra seem difficult to use. Although it is possible to measure data at quite low resolution, the standard models for bulk-solvent correction need to be improved in order to render them exploitable.

References

Coulibaly, F., Chevalier, C., Gutsche, I., Pous, J., Navaza, J., Bressanelli, S., Delmas, B. & Rey, F. A. (2005). *Cell*, **120**, 761–772.
Dodson, E. J. (2001). *Acta Cryst.* D**57**, 1405–1409.
Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D**58**, 1387–1392.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 1070–1072.
International Tables for Crystallography (2006). Vol. A, *Space-group Symmetry*, 1st online ed., edited by Th. Hahn. Chester: International Union of Crystallography.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. (2003). *J. Struct. Biol.* **104**, 13–23.
Navaza, J., Lepault, J., Rey, F. A., Álvarez-Rúa, C. & Borge, J. (2002). *Acta Cryst.* D**58**, 1820–1825.
Trapani, S., Abergel, C., Gutsche, I., Horcajada, C., Fita, I. & Navaza, J. (2006). *Acta Cryst.* D**62**, 467–475.
Trapani, S., Schoehn, G., Monchois, V., Navaza, J. & Abergel, C. (2008). In preparation.