

# Supplanting crystallography or supplementing microscopy? A combined approach to the study of an enveloped virus

Erika J. Mancini<sup>a</sup> and Stephen D. Fuller<sup>a,b\*</sup>

<sup>a</sup>The Structural Biology Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, and <sup>b</sup>Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of Oxford, Headington, Oxford OX3 7BN, England

Correspondence e-mail: steve@strubi.ox.ac.uk

Received 8 June 2000  
Accepted 1 August 2000

The recent advances in the resolution obtained by single-particle reconstructions from cryo-electron microscopy (cryo-EM) have led to an increase in studies that combine X-ray crystallographic results with those of electron microscopy (EM). Here, such a combination is described in the determination of the structure of an enveloped animal virus, Semliki Forest virus, at 9 Å resolution. The issues of model bias in determination of the structure, the definition of resolution in a single-particle reconstruction, the effect of the correction of the contrast-transfer function on the structure determined and the use of a high-resolution structure of a subunit in the interpretation of the structure of the complex are addressed.

## 1. Introduction

The inclusion of so many electron-microscopy talks in a CCP4 workshop is a sign of the maturing of cryo-EM and image reconstruction of single particles. It has passed from a 'hobby' practiced by a small devoted community into a structural tool of general interest. Two other complementary events signal this maturation.

The first development is the increasing demand that the results of EM be deposited in public databases in the same way as has become routine for the deposition of the results of X-ray crystallography in the Protein Data Bank (PDB). The first cryo-EM structures to be deposited were those generated by electron crystallography and were deposited in the PDB (Table 1). It was relatively straightforward to incorporate these high-resolution structures because they resulted in atomic models and could be deposited as coordinates. Further, the criteria for the resolution of the electron crystallographic structures are familiar ones to the X-ray crystallographic community. The deposition of other EM structures has been less straightforward. The PDB was focused on coordinates and did not accommodate density maps, which are the most natural form in which low-resolution structures are produced. Structure factors, which can be used to recalculate the density map, can be deposited along with the coordinates; however, they are not naturally produced by single-particle reconstruction methods. Carazo and coworkers (Carazo & Stelzer, 1999; Carazo *et al.*, 1999) initiated a pioneering effort to establish a multidimensional database, BioImage, which served as a repository for a number of cryo-EM maps without the requirement for atomic models. The structures deposited in the BioImage database include those derived from cryo-EM by single-particle reconstruction methods as well as multi-dimensional data sets from light microscopy, atomic force microscopy and other methods. The past year has seen the

**Table 1**

Cryo-EM structures deposited in the MSD (May 2000).

The table shows the structures derived from cryo-EM and image reconstruction that had been deposited in the Macromolecular Structure Database (MSD; Keller *et al.*, 1998) by May 2000. All of the entries include atomic models. Four include the structure factors that could be used to calculate the corresponding maps.

| Identifier | Descriptor                                    | Technique                                       | Resolution (Å) | Citation                        | Structure factors |
|------------|---|---|----------------|---------------------------------|-------------------|
| 1qgc       | Foot-and-mouth disease virus-antibody complex | Icosahedral reconstruction                      | 30             | Hewat <i>et al.</i> (1997)      | No                |
| 1d3e       | Rhinovirus 16-receptor complex                | Icosahedral reconstruction                      | 28             | Kolatkhar <i>et al.</i> (1999)  | Yes               |
| 1kvp       | Spiroplasma virus Spv4                        | Icosahedral reconstruction                      | 27             | Chipman <i>et al.</i> (1998)    | No                |
| 1d3i       | Rhinovirus 16-receptor complex                | Icosahedral reconstruction                      | 26             | Kolatkhar <i>et al.</i> (1999)  | Yes               |
| 1dgi       | Poliovirus-receptor complex                   | Icosahedral reconstruction                      | 22             | He <i>et al.</i> (2000)         | No                |
| 1ego       | 70S ribosome                                  | Single-particle reconstruction without symmetry | 11.5           | Gabashvili <i>et al.</i> (2000) | No                |
| 1dyl       | Semliki Forest virus                          | Icosahedral reconstruction                      | 9              | Mancini <i>et al.</i> (2000)    | Yes               |
| 1c2x       | Ribosomal complex                             | Single-particle reconstruction without symmetry | 7.5            | Mueller <i>et al.</i> (2000)    | No                |
| 487d       | Ribosomal complex                             | Single-particle reconstruction without symmetry | 7.5            | Mueller <i>et al.</i> (2000)    | No                |
| 1tub       | Tubulin                                       | Electron crystallography                        | 3.7            | Nogales <i>et al.</i> (1998a,b) | No                |
| 1brd       | Bacteriorhodopsin                             | Electron crystallography                        | 3.5            | Henderson <i>et al.</i> (1990)  | Yes               |
| 1at9       | Bacteriorhodopsin                             | Electron crystallography                        | 3.0            | Kimura <i>et al.</i> (1997)     | No                |
| 2at9       | Bacteriorhodopsin                             | Electron crystallography                        | 3.0            | Mitsuoka <i>et al.</i> (1999)   | No                |

logical next step in which several structures derived by single-particle reconstruction have been deposited in the Macromolecular Structure Database (MSD; Keller *et al.*, 1998), a successor to the PDB. These entries are primarily viral structures and are all accompanied by atomic models (Table 1). Today, cryo-EM structures comprise less than 0.1% of the more than 12 000 entries in the MSD. Three of the deposited cryo-EM structures are accompanied by structure factors so that the map, as well as the derived model, is accessible. There is increasing support for insisting that EM structures be deposited as a requirement for publication, in the same way as currently required for X-ray structures.

The second development is the dawning of a movement in the EM community toward standardizing data formats and image-processing packages. Several groups have modified their image-processing packages to utilize the CCP4 map format (Collaborative Computational Project, Number 4, 1994) for both images and maps. While this format is not ideal for image processing of electron-microscopic data, it does have the attraction of easing the interface to X-ray crystallographic data and programs. Increased standardization will ease the task of depositing structures. Indeed, without an increased level of standardization, it is difficult to envision a

coherent approach to the validation of the data determined by single-particle reconstruction. Understandably, few in the EM community would like to make results available without generally agreed measures for their reliability.

This contribution will address some of the issues that arise in the attempt to combine structures determined by cryo-EM and by X-ray crystallography. We will follow the example of the determination and interpretation of the 9 Å resolution structure of the enveloped Semliki Forest virus (SFV; Mancini *et al.*, 2000), in which the complementarity of X-ray and cryo-EM was particularly useful. The reader should see Mancini *et al.* (2000) for a discussion of the results of the study.

## 2. Semliki Forest virus (SFV)

Several features of the SFV system make the combination of cryo-EM and X-ray particularly useful in its structure determination. SFV comprises 240 copies of a capsid protein (C)

assembled into an icosahedral nucleocapsid that encloses the 42S positive-strand RNA genome. The nucleocapsid is enveloped by a membrane containing 240 copies of each of two single-spanning transmembrane proteins, E1 and E2, and of an extrinsic envelope protein, E3. The C protein and the envelope protein complexes are arranged with  $T = 4$  symmetry and are linked by the transmembrane tails of E1 and E2. Although the capsid and envelope share the same  $T = 4$  triangulation number, the units are arranged differently: the capsid protein is clustered as hexamers and pentamers, while the envelope contains trimer clustered spikes (E1E2E3)<sub>3</sub> so that the two arrangements are complementary (Fuller, 1987; Fuller *et al.*, 1995).

SFV is a particularly advantageous case for the determination of an enveloped virus structure. The virus was the prototype used for the description of pH-triggered fusion after endocytosis and hence a comprehensive literature exists concerning the biochemistry and cell biology of this virus (Kielian & Helenius, 1986). The preparation of homogeneous virus in relatively large quantities is well established. Indeed, it is possible to grow crystals of the virus very easily, although these crystals have yet to yield diffraction beyond 30 Å resolution. X-ray crystallography has yielded structures of the

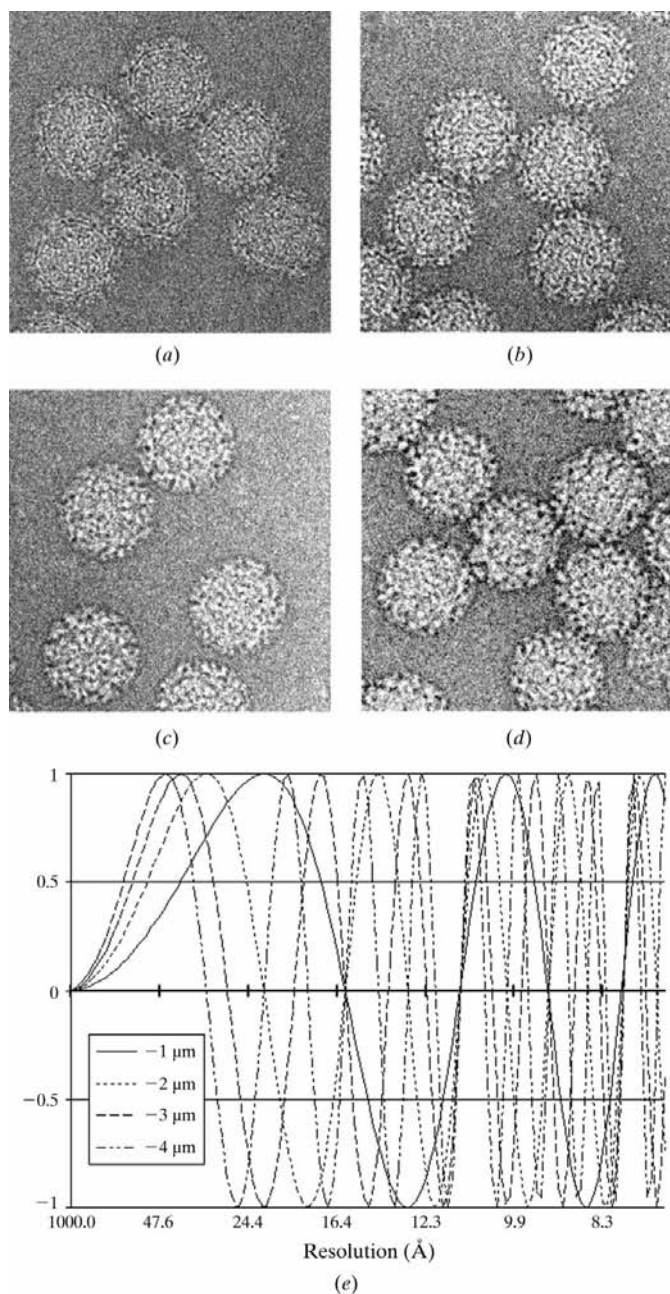
capsid protein of SFV (Choi *et al.*, 1997) to accompany those of other members of the alphavirus family (Choi *et al.*, 1991). Interestingly, all of the alphavirus capsid proteins crystallize as dimers (Choi *et al.*, 1991, 1997). In some of the crystal forms a carboxy-terminal tail can be seen to form a dimer linkage which is believed to mimic the interaction of the E2 tail with the capsid (Lee *et al.*, 1996; Skoging *et al.*, 1996). Finally, cryo-EM had provided structures of SFV (Fuller *et al.*, 1995) and of other alphaviruses (Cheng *et al.*, 1995; Paredes *et al.*, 1998) to better than 25 Å resolution that could serve as starting points for higher resolution structure determination. Cryo-EM had also revealed the structures of several variants and complexes that should assist in the functional interpretation of a higher resolution structure (Ferlenghi *et al.*, 1998; Smith *et al.*, 1995). A final spur to obtaining higher resolution structure from cryo-EM is the work on crystals of a soluble domain of E1 (Wengler & Rey, 1999) that will yield a high-resolution structure for the fusion protein of the virus.

### 3. Reconstruction

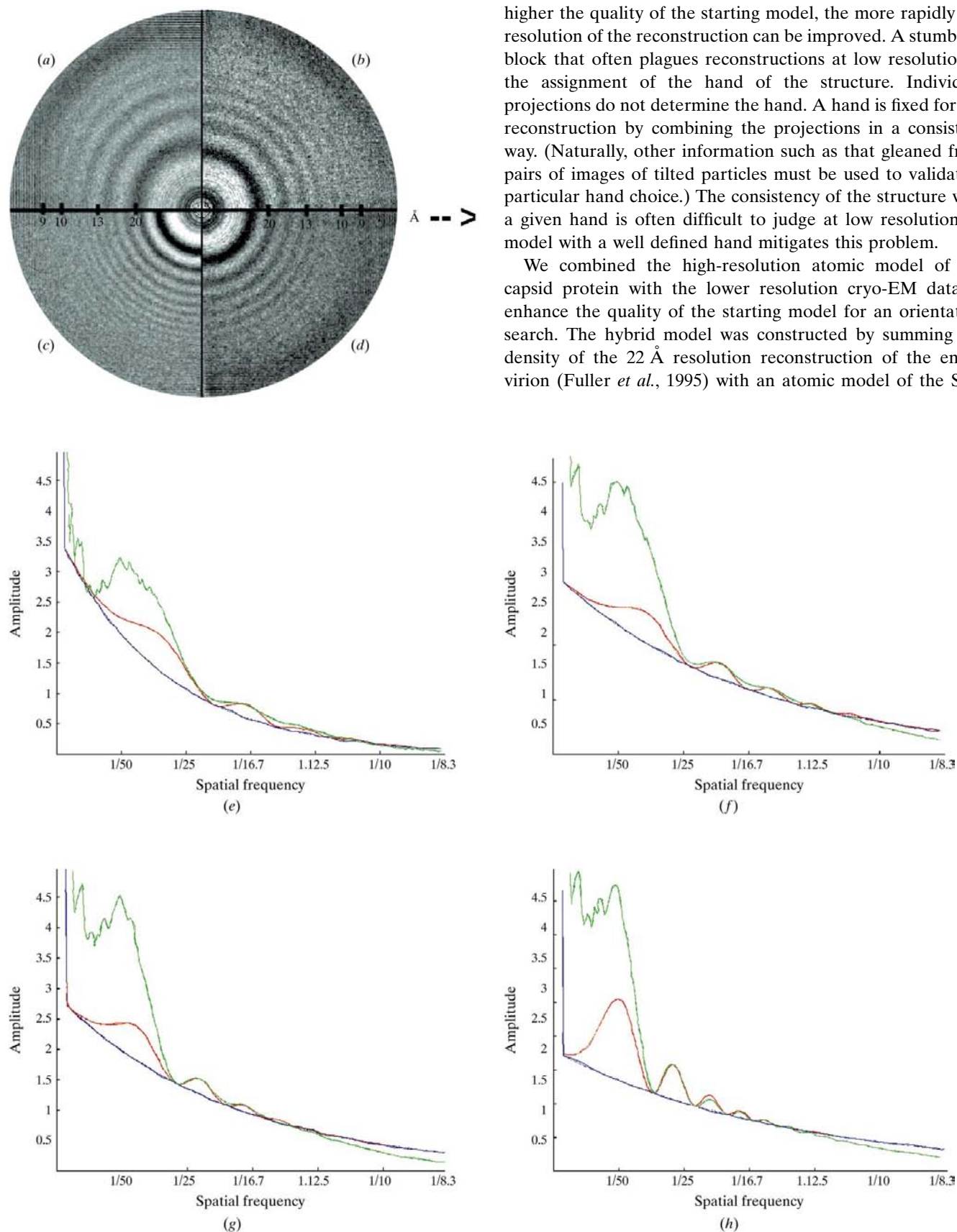
The quality of the data used in a single-particle reconstruction can only be assessed by the resolution achieved in the reconstruction (see below). Nevertheless, the data collection could be optimized for achieving higher resolution. The use of a very homogeneous sample (Fig. 1) allowed the collection of micrographs from which hundreds of usable particle images could be extracted. The final reconstruction contains 5276 out of 9950 particle images initially selected from 48 micrographs (Mancini *et al.*, 2000). The micrographs were recorded at a magnification of 50 000 $\times$  on a Philips CM200-FEG electron microscope equipped with a GATAN 626 cryo-transfer holder and operated at 200 kV. Uniform coverage of the resolution range to 9 Å was achieved by selecting micrographs which covered a broad defocus range (−975 to −7628 nm). A portion of the carbon film was included in each field to mitigate charging effects and to ease the visualization of the transfer function. Fig. 1 shows typical fields of SFV particles at different defoci. The use of very coherent field-emission gun (FEG) illumination also aids the data collection by increasing the strength of high-resolution information in the images. Cryo-EM provides images of an unstained sample in which defocus phase contrast is the major component the image contrast. The phase contrast transfer function (CTF) for the defoci in Figs. 1(a)–1(d) is seen in Fig. 1(e). Structural features with different characteristic dimensions become clearly visible at different defocus values. This can be seen in Fig. 1 by the variation in image features with the change in the defocus of the images  $\delta f = -1 \mu\text{m}$  to  $\delta f = -4 \mu\text{m}$ . The lipid bilayer is clearly visible in the closer-to-focus images while the coarser features of the virus are enhanced in Fig. 1(d). These phase-contrast effects are more prominent with the use of FEG data which also eases the determination of the CTF for the images so that it can be more precisely corrected.

Single-particle reconstruction requires the determination of the relative orientations of the projections of the particle seen in the images and their combination in an unbiased way to

determine the structure. Practically, this involves comparing the observed projections (*i.e.* images) with those expected for a model of the structure (Fig. 3; Baker & Cheng, 1996; Baker *et al.*, 1999; Crowther *et al.*, 1994; Fuller *et al.*, 1996; Mancini *et al.*, 1997). As the quality of the reconstruction improves during reconstruction cycles, so does the quality of the model and consequently the accuracy of the orientations. This leads to an improved structure in the subsequent cycle. Obviously, the



**Figure 1** Cryo-electron microscopy of Semliki Forest virus. Selected regions from different cryo-electron micrographs of SFV particles which were used for image processing and three-dimensional reconstruction. The effect of defocusing on the appearance of viral features is shown for  $\delta f = 1$  (a), 2 (b), 3 (c) and 4  $\mu\text{m}$  (d). The theoretical phase contrast transfer functions for these defocus values are shown in the plot in (e). Note that further defocus values would be necessary to give complete coverage to 7 Å resolution.



**Figure 2**

Average power spectra, CTF curves and fit to the CTF curves. The logarithm of the power spectrum averaged over particles is shown for four micrographs, at nominal defoci of  $\delta f = -1$  (a),  $-2$  (b),  $-3$  (d) and  $-4 \mu\text{m}$  (c). Rotationally averaged profiles of these spectra (green line), the calculated CTF (red line) and the Gaussian baseline (blue line) are shown in (e), (f), (h) and (g), respectively.

higher the quality of the starting model, the more rapidly the resolution of the reconstruction can be improved. A stumbling block that often plagues reconstructions at low resolution is the assignment of the hand of the structure. Individual projections do not determine the hand. A hand is fixed for the reconstruction by combining the projections in a consistent way. (Naturally, other information such as that gleaned from pairs of images of tilted particles must be used to validate a particular hand choice.) The consistency of the structure with a given hand is often difficult to judge at low resolution. A model with a well defined hand mitigates this problem.

We combined the high-resolution atomic model of the capsid protein with the lower resolution cryo-EM data to enhance the quality of the starting model for an orientation search. The hybrid model was constructed by summing the density of the 22 Å resolution reconstruction of the entire virion (Fuller *et al.*, 1995) with an atomic model of the SFV

capsid derived from the structure of a monomer of the SFV C protein crystal structure (Choi *et al.*, 1997). The model density was normalized so that it had the same average value as the map at the radii of the capsid. The initial model of the capsid was based on an arrangement of C protein which had exact quasi-equivalence and corresponded to that first proposed for the Ross River virus (Cheng *et al.*, 1995)

This capsid atomic model provides a very high contrast handed component of the model for the orientation determination. The atomic model was added to the current reconstruction to produce a new hybrid model for the next cycle of the orientation determination until the resolution of  $\sim 14$  Å was reached. Past this resolution, the departure of the capsid protein arrangement from strict quasi-equivalence became apparent and the initial atomic model was no longer included.

#### 4. CTF correction

Always remember, your focus determines your reality.  
Qui-Gon in G. Lucas (1999), *The Phantom Menace*.

The projected image of the unstained object recorded in cryo-EM is related to the projection of the object's Coulomb potential by convolution with the CTF of the microscope. CTF accentuates certain resolution ranges at the expense of others, allowing the visualization of the unstained object. This however also means that the amplitudes and phases need to be corrected to achieve a faithful representation of the structure. This correction or deconvolution is one of the key steps in extending the resolution of cryo-EM procedures to higher resolution. Deconvolving the CTF in the images involves a series of intrinsically unstable operations that run the risk of inadvertently amplifying noise or degrading signal. An important point, which is occasionally finessed in the rush to calculate a structure, is that CTF correction alone can never restore the missing information near the nodes when only a single defocus is available. Only combining a range of complementing defocuses can restore the information needed to characterize the specimen over the complete range of resolutions.

Correction of the CTF requires determining the set of parameters that allow its calculation. These are elucidated from the power spectrum of the image and then checked by observing the positions of phase reversals in the higher resolution portion of the transform. A quasi-empirical approach is commonly used which has proved sufficient for resolutions to 7 Å for thin specimens. This procedure was made more reliable by the use of FEG data, because multiple zeros could be seen in the power spectrum of the micrographs, as seen in Fig. 2. The CTF includes both phase- and amplitude-contrast contributions (Erickson & Klug, 1971; Toyoshima & Unwin, 1988) and is a function of the resolution (or scattering vector,  $\nu$ ) which depends on the electron wavelength ( $\lambda$ ), the defocus ( $\delta f$ ), the spherical aberration of the lens ( $C_s$ ) and the fraction of amplitude contrast ( $F_{\text{amp}}$ ).

$$\text{CTF}(\nu) = (1 - F_{\text{amp}}) \sin[\chi(\nu)] + F_{\text{amp}} \cos[\chi(\nu)],$$

where

$$\chi(\nu) = \pi\lambda\nu^2(\delta f - 0.5C_s\lambda^2\nu^2).$$

The curves shown in Fig. 1(e) display the phase-contrast component of the CTF ( $F_{\text{amp}} = 0$ ) and reveal how the CTF reverses phases and alters intensity as a function of resolution.

The correspondence between the theoretical power spectrum and that of a real image must also include the effect of an exponential decay in amplitude with resolution. This decay reflects the effect of the limited temporal coherence of the electron beam, ice thickness, movement and other factors which degrade the high-resolution information in the image. Consequently, an observed power spectrum must be described by the CTF parameters and others describing the baseline and this exponential decay.

The CTF must be evaluated for each micrograph. We used the incoherent average of the particle images' power spectra (de Haas *et al.*, 1999) to elucidate the CTF parameters. An estimate of the defocus value is obtained from the position of the function's first zero and a table of first CTF zeros, calculated using  $C_s = 2$  mm (for the Philips CM200 FEG) and  $F_{\text{amp}} = 0.1$  (typical value for vitrified samples; Toyoshima & Unwin, 1988) using rotationally averaged power spectra (Fig. 2). Next, the estimate of the defocus was used as a starting value to optimize the agreement of the predicted power spectrum with the observed power spectrum. The two curves were plotted and the defocus was adjusted until the best agreement between the predicted and the observed power spectrum was achieved. The baseline was determined from the values of the rotationally averaged power spectrum at the first and third zero of the CTF, based on the CTF for a given defocus.

Once the parameters for the imaging conditions for each micrograph were established, they were used to compensate for the CTF effects in the individual images and in the final reconstruction. The final reconstruction  $\mathcal{R}_{\text{corrected}}$  was obtained by summing Fourier transforms  $\mathcal{F}_i$  of the individual reconstructions  $\mathcal{R}_i$  each multiplied by the corresponding CTF ( $\text{CTF}_i$ ), dividing by the sum of the squares of the CTFs and then back-transforming. The parameter  $\delta$  reflects the signal to noise and determines the weight of the low-amplitude information near the CTF nodes in the corrected image. The addition of  $\delta$  avoids amplification of the noise near the CTF nodes. Typically, the values of  $\delta$  were between 0.1 and 0.5.

$$\mathcal{R}_{\text{corrected}} = \mathcal{F}^{-1} \left\{ \frac{\sum_i [\text{CTF}_i \times \mathcal{F}(\mathcal{R}_i)]}{\sum_i (\text{CTF}_i)^2 + \delta} \right\}.$$

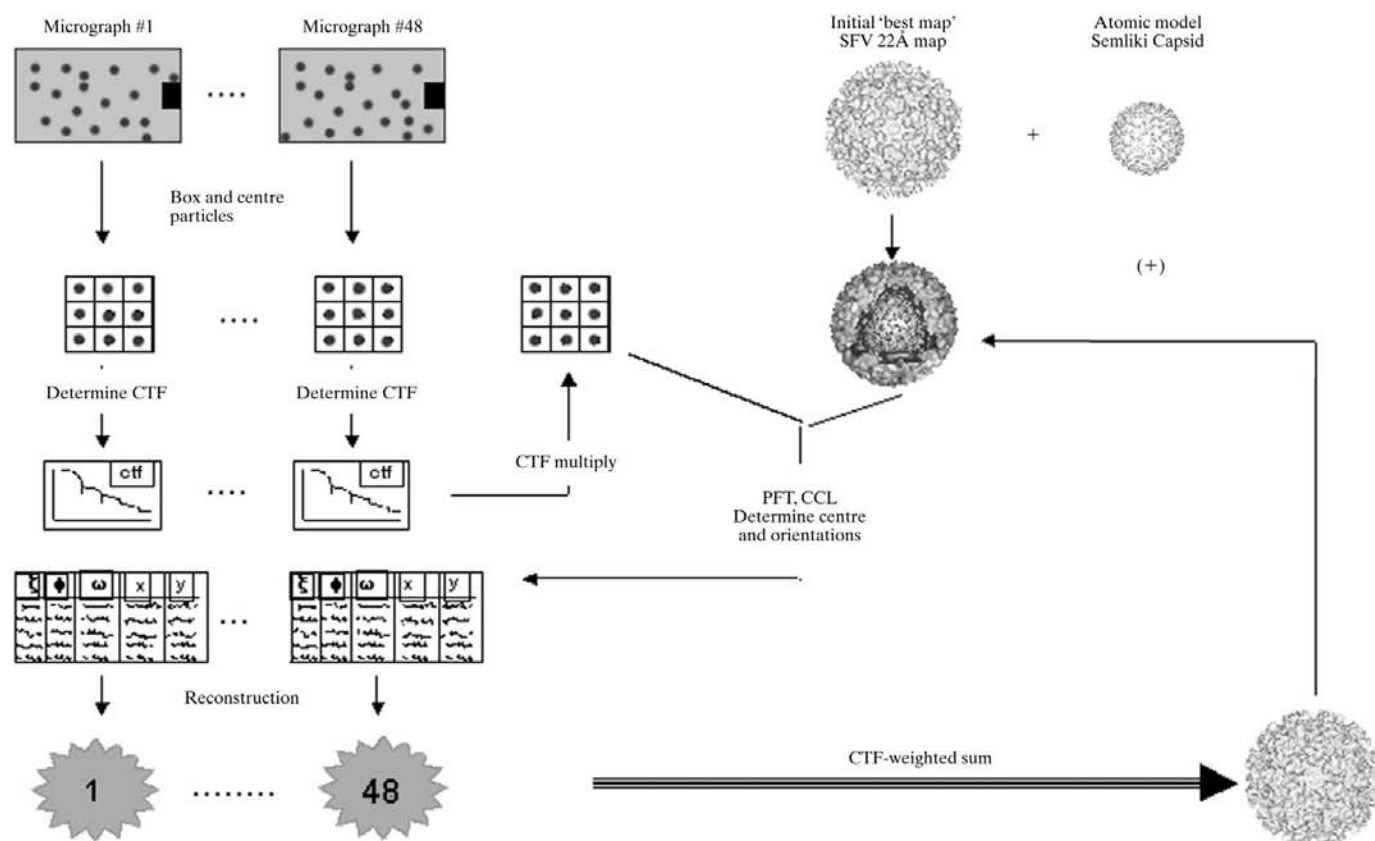
Compensation of the CTF was also important for orientation determination. Individual particle images were convolved with the CTF to improve the reliability of model-based orientation searches *versus* the CTF-corrected models. Multiplication of the image transform by the appropriate CTF decreases amplitudes of the less reliable portions of the image transform (*i.e.* near the CTF nodes). Although this treatment degrades the appearance of the image of the particle, it accentuates the

reliable portion of the information for determining the orientation. The effect of CTF correction on an individual image (Figs. 4*a* and 4*b*) and on an individual reconstruction (Figs. 4*c* and 4*d*) is shown in Fig. 4. Fig. 4 also shows the effect of the CTF correction on the amplitudes of their power spectra.

The combination of the individual reconstructions from images at a range of complementary defocusses, each weighted by its CTF, corrects for the data missing at the nodes. The weighted average reconstruction contains reliable phases because it combines complementary defocusses. The amplitudes are also affected by the imaging process. Unfortunately, the effects on amplitudes are less well understood than those on the phases and are more difficult to model. The exponential envelope, mentioned above, decreases the amplitudes of the higher resolution portion of the transform. The averaging of images compensates for this loss of intensity so that the combined phases are reliable; however, the amplitudes still need to be adjusted to correspond to the true transform of the object. The low-resolution portion of the transform is also affected by the imaging process and appears artificially strong. Several factors have been suggested for the enhancement of

the low-resolution portion of the transform, including inelastic scattering and the thickness of the water layer; however, useful models for its correction are not available. Comparison of low-angle X-ray results with image reconstructions from cryo-EM confirm that the low-resolution components are amplified and the higher resolution components attenuated relative to those of the object (Thuman-Commike *et al.*, 1999). Correction of these effects requires attenuating the low-resolution portion of the transform and boosting the high-resolution portion. Boosting is usually accomplished by applying an inverse temperature factor to sharpen the reconstruction. This is a relatively standard practice in high-resolution cryo-EM (Böttcher *et al.*, 1997; Conway *et al.*, 1997) and is routinely used in electron crystallography (Henderson *et al.*, 1990; Kimura *et al.*, 1997; Mitsuoaka *et al.*, 1999; Nogales *et al.*, 1998*a,b*). Attenuation of the low-resolution components is achieved by applying an envelope function to downweight them relative to the high-resolution ones.

The SFV case is fortunate in that there is low-resolution X-ray data for the closely related Sindbis virus (Harrison *et al.*, 1971) which provides a guide to an appropriate envelope. We combined such an envelope [ $\exp(-6.75 \text{ \AA}^2/d^2)$ , where  $d$  is the



**Figure 3**

Scheme of the icosahedral reconstruction method for the 9 Å resolution SFV structure. Schematic diagram of the three-dimensional reconstruction method used for the 9 Å resolution SFV structure, illustrating the important steps in the process. Several hundred particle images were selected from 48 micrographs with a range of defocusses. The orientation of each particle was determined by comparison with a hybrid model formed by summing the atomic model of the SFV capsid with the current reconstruction of the entire virus. Particle images were convolved by the corresponding CTF to correct the phases and to decrease the amplitudes near the CTF nodes. A separate reconstruction was calculated from each image. The reconstructions were multiplied by the CTF and combined as the CTF-normalized sum. An envelope applied to compensate for the effect of imaging on the amplitudes by down-weighting the low-resolution components and boosting the high-resolution ones.



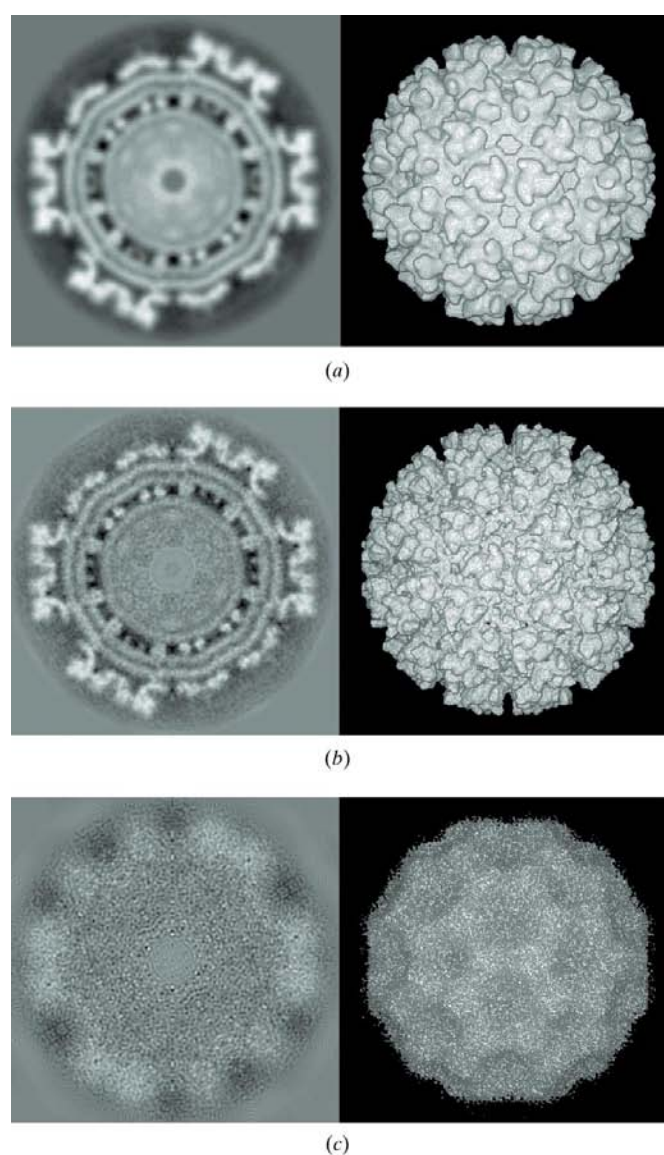
resolution in Å] for the low-resolution components and a relatively conservative boosting [ $\exp(15 \text{ \AA}^2/d^2)$ ] for the high-resolution components. The effects of this choice of amplitude weighting and of other choices are seen in Fig. 4.

### 5. How good is it? Assessing the resolution

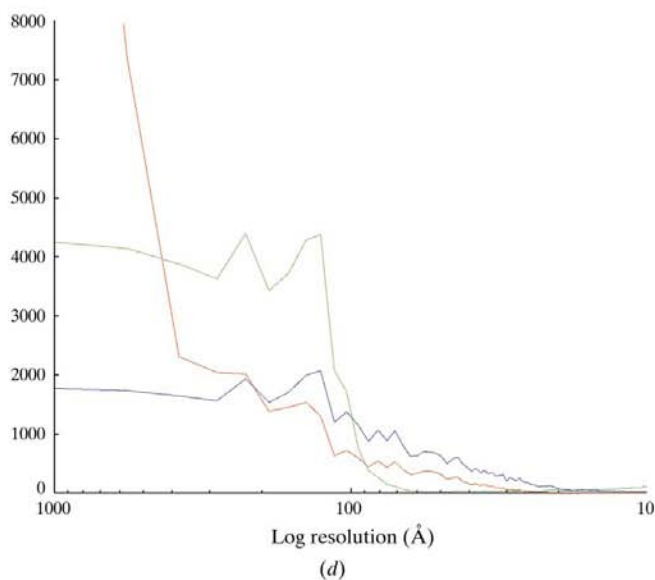
Single-particle reconstruction differs from crystallography in that the resolution attainable in the result must be assessed from the result itself. Examination of the power spectrum of the image (Fig. 2) reveals the resolution to which the modulation of the transfer function can be observed. This provides an upper limit to the information which can be extracted from the image. There is no guarantee that the reconstruction will

reach this resolution, since reconstruction is also affected by such factors as the accuracy of orientation determination and the variation between particles. Variation is particularly important for large complexes and may correspond to a more significant limit to the resolution achieved with some specimens than the quality of the microscope or the processing.

There are two fundamentally different methods for determining the resolution in a single-particle reconstruction. The first is the most general and involves dividing the data into two halves and then calculating two independent reconstructions which are then compared by Fourier shell correlation (FSC) or difference phase residual (DPR; Baker *et al.*, 1999). The resolution to which these measures satisfy a quality limit is taken as the resolution of the reconstruction. There is significant discussion in the literature over the choice of the appropriate quality limit (Gabashvili *et al.*, 2000; Mueller *et al.*, 2000). Grigorieff (2000) has pointed out that these measures are often applied in a way that gives an artificially optimistic resolution estimate. Nevertheless, this approach remains the most common in the field and represents the simplest way to compare the quality of reconstructions performed by different groups. Often, an FSC limit of 0.5 is used as a conservative measure of resolution. In well behaved reconstructions, the precise choice of cutoff value is relatively unimportant since the FSC falls sharply near the resolution limit. A second less general method is the use of internal criteria within the image processing itself. In the case of icosahedral reconstruction, for example, this could be the average phase residual for the individual images relative to the average (Stewart *et al.*, 1991). This criterion is usually extremely conservative since the higher resolution information is relatively noisy and no



**Figure 4** Restoring the amplitudes. (a) Central section and isosurface representation of the final reconstruction before amplitude restoration; (b) after amplitude restoration with the  $\exp(-6.75 \text{ \AA}^2/d^2)$  and  $\exp(15 \text{ \AA}^2/d^2)$  envelope (where  $d$  is the resolution in Å); (c) after incorrect amplitude restoration which overemphasizes the low-resolution region but boosts the high resolution to the same extent. (d) Profiles of the reconstruction power spectra calculated before amplitude restoration (red line), after proper amplitude restoration (blue line) and after incorrect amplitude restoration (green line). The profiles are plotted on a log scale.



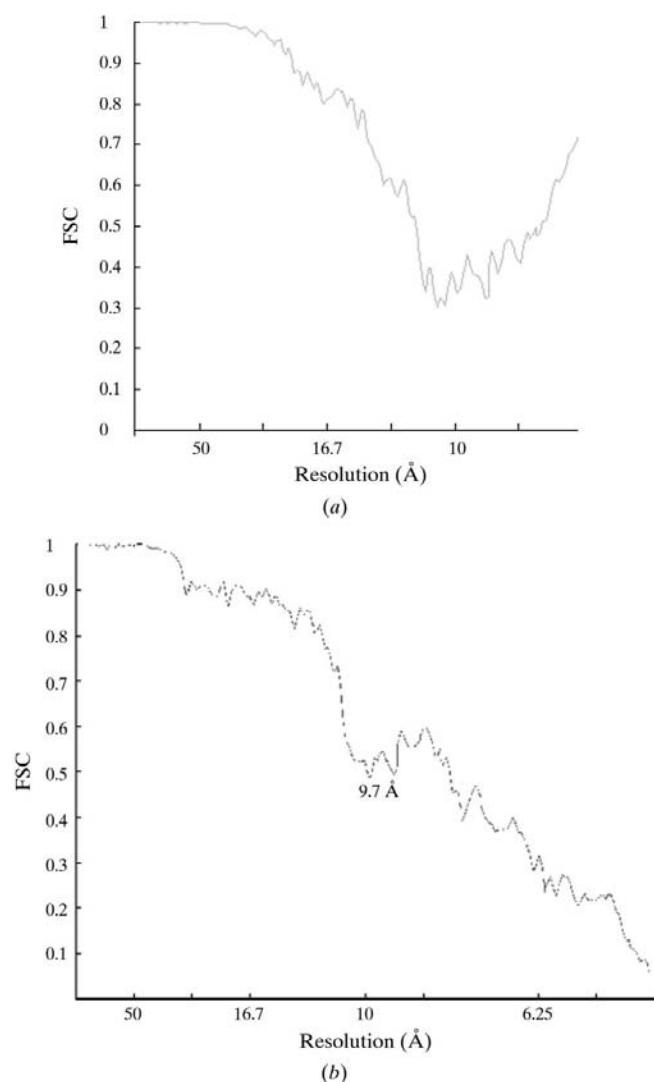
increase in signal to noise similar to that obtained in crystallographic averaging occurs for the individual images.

SFV provides an example in which a single resolution measure is too coarse a description of the quality of the structure (Mancini *et al.*, 2000). FSC of independent reconstructions of the whole virion reaches 0.5 at 13.6 Å resolution. FSC of successive 35 Å thick shells revealed that the resolution varied with the position in the structure. The RNA was unsurprisingly the region of the map with the lowest resolution. The nucleocapsid (NC) and the spikes were the regions that yielded the highest resolution. A monotonic fall-off of resolution with increasing radius could suggest a problem in determining orientation; however, the variation in resolution observed here matches the positions of structural features. It is easiest to interpret as a reflection of variable order in different

portions of the structure (Mancini *et al.*, 2000). The FSC for the well ordered NC region is shown in Fig. 5.

A different way of assessing the resolution is more relevant to the combination of EM and X-ray data being discussed here. The high-resolution structure of the capsid protein is known from X-ray crystallographic studies of the crystals of the dimer (Choi *et al.*, 1997). We can use the similarity of the EM density to that of the electron density corresponding to the atomic structure as a measure of the resolution of the EM map. This requires determining the positions of the atomic structure that provides the best fit to the observed density. The fit allows us to use the comparison of the high-resolution structure determined by X-ray crystallography with the cryo-EM map to assess the resolution of the reconstruction. The results (Fig. 5) support a better resolution than that suggested by the FSC analysis.

This procedure may seem somewhat circular. An atomic model of the capsid based on the high-resolution structure was used to help in the finding of orientations to generate the map. Does the initial use of this model rule out comparison with the atomic structure as a measure of the quality of the reconstruction? We know that we have avoided the problem of model bias by the observation that the model was only helpful for finding orientations until the 14 Å resolution. Past this



**Figure 5**  
Assessing the resolution of the reconstructed nucleocapsid. (a) The Fourier shell correlation (FSC) calculated between two independent reconstructions of the capsid density (each of which contained half of the total data). (b) The FSC calculated between the reconstruction of the capsid density (which contained all of the data) and the atomic model of the capsid.



**Figure 6**  
Fit of the C protein atomic structure to the nucleocapsid density. The fit of the atomic structure of the C protein (Choi *et al.*, 1997) to the observed capsid density. A real-space fit was performed for each of the four independent positions of the asymmetric unit using the program *EMFIT* (Cheng *et al.*, 1995). The C protein crystal structure is represented as a ribbon diagram drawn with the program *BOBSCRIPT*. The ribbon diagram is coloured by the density of the map as described in the text. A contour of the map corresponding to  $1.5\sigma$  is shown.



resolution, the model hindered the finding of orientations and was discarded. Reassuringly, the reconstruction process converged to a structure that differed from the model, although the fit shows that the departure was a slight rearrangement of the capsid subunits.

### 6. Fitting and interpretation

The optimal fit between the atomic model of the capsid and the present reconstruction was found by using a modification of the program *EMFIT*. This program was originally authored by Professor M. Rossmann (Purdue University) and first used for the fitting of the Sindbis virus capsid protein to the Ross River virus structure at 25 Å resolution (Cheng *et al.*, 1995). *EMFIT* identified the best fit in real space between the atomic model of the capsid protein and each of the four independent monomer densities in a shell between the inner leaflet of the bilayer and the RNA (Mancini *et al.*, 2000).

*EMFIT* works in two steps. First, it uses a real-space rigid-body procedure in which the atomic coordinates are rotated through all possible orientations and translations in the icosahedral asymmetric unit. The quality of the fit is defined by an *R* factor and by the overlap between density and atoms. The positions of the best 20 fits are then refined by a local search algorithm. The reasonableness of the fit is evaluated by the number of overlapping residues between symmetry-related or quasi-symmetry-related molecules. This objective fitting also allowed refinement of the scale of the map (yielding 2.52 Å pixel<sup>-1</sup> versus the nominal 2.64 Å pixel<sup>-1</sup>). Real-space refinement was chosen because it seemed less sensitive to the presence of other non-modelled density (the inner leaflet of the membrane, the tail of the spike, the RNA *etc.*) in the region of the map which was being fitted. The results with *EMFIT* appeared relatively insensitive to small changes in the envelope which were used to correct the amplitudes (Thouvenin & Hewat, 2000).

The fit is shown in Fig. 6. The SFV capsid protein does not have prominent secondary structural features, such as  $\alpha$ -helices, that should be visible at 9 Å resolution. Helical features are indeed seen in the transmembrane regions of the reconstruction (Mancini *et al.*, 2000). The capsid protein does contain clusters of secondary structural elements that give a characteristic variation in the internal density of the capsid protein. The positions of these clusters match those of variations in the reconstructed density (Fig. 6) and provide a reassuring confirmation of the quality of the map (Mancini *et al.*, 2000). Similar variations are not seen in reconstructions at lower resolution which typically reveal protein subunits as shells of relatively uniform density (Baker *et al.*, 1999; Fuller *et al.*, 1995; Grimes *et al.*, 1997).

Comparison of the original capsid model (Cheng *et al.*, 1995) with the one refined from the high-resolution map shows a very small departure. The original model imposed strict quasi-equivalence in which the hexamers of the *T* = 4 capsid had exact local sixfold symmetry. The higher resolution structure shows that the hexamer is slightly bent so that this sixfold symmetry is perturbed. The density shows an alteration

in the structure of the capsid protein relative to the high-resolution structure of the dimeric protein found in the crystal. A portion of the carboxy-terminal arm of the protein appears to have rearranged to form a connection between adjacent subunits. The density corresponding to this connection is sufficiently consistent that it is enhanced by real-space averaging using the body of the capsid protein in the program *GAP* (Grimes *et al.*, 1995, 1997; Mancini *et al.*, 2000). The position of the carboxy-terminus of the E2 protein relative to the fitted capsid is precisely as predicted for the interaction suggested previously (Lee *et al.*, 1996; Skoging *et al.*, 1996).

The fit of the capsid protein into the map and structure factors calculated from the map have been deposited in the MSD as entry 1dyl. A pseudo-temperature factor (following the suggestion of Professor M. G. Rossmann, Purdue University) has been included in the entry to give a relative measure of the correspondence of the map and the fitted model. The temperature factor is inversely proportional to the density of the map at the positions of the atoms in the model. Hence, portions of the model that lie in lower density regions would be visualized as red in a temperature-factor colouring provided by such programs such as *RASMOL* (Sayle & Milner-White, 1995) or *BOBSCRIPT* (Esnouf, 1997, 1999) (Fig. 6).

### 7. Conclusions

The combination of the X-ray and cryo-EM provided enormous advantages in the determination and interpretation of the SFV structure. X-ray information aided the finding of orientations to generate the reconstruction and assisted in the correction of amplitude effects to obtain an interpretable map. The availability of a high-resolution structure allowed the interpretation to pass from a description of features of the density to an atomic model that suggested interactions and rearrangements. We hope that our experience with the SFV system will motivate others to continue the combination of X-ray crystallographic and cryo-EM approaches which have formed the focus of this workshop and are becoming more common in structural biology (Baker *et al.*, 1999).

### References

- Baker, T. S. & Cheng, R. H. (1996). *J. Struct. Biol.* **116**, 120–130.
- Baker, T. S., Olson, N. H. & Fuller, S. D. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 862–922.
- Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.
- Carazo, J. M. & Stelzer, E. H. (1999). *J. Struct. Biol.* **125**, 97–102.
- Carazo, J. M., Stelzer, E. H., Engel, A., Fita, I., Henn, C., Machtynger, J., McNeil, P., Shotton, D. M., Chagoyen, M., de Alarcon, P. A., Fritsch, R., Heymann, J. B., Kalko, S., Pittet, J. J., Rodriguez-Tome, P. & Boudier, T. (1999). *Nucleic Acids Res.* **27**, 280–283.
- Cheng, R., Kuhn, R., Olson, N., Rossmann, M. & Baker, T. (1995). *Cell*, **80**, 621–630.
- Chipman, P. R., Agbandje-McKenna, M., Renaudin, J., Baker, T. S. & McKenna, R. (1998). *Structure*, **6**, 135–145.
- Choi, H.-K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. (1997). *Proteins*, **27**, 345–359.

- Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). *Nature (London)*, **354**, 37–43.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D* **50**, 760–763.
- Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). *Nature (London)*, **386**, 91–94.
- Crowther, R. A., Kiselev, N. A., Böttcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). *Cell*, **77**, 943–950.
- Erickson, H. P. & Klug, A. (1971). *Philos. Trans. R. Soc. London Ser. B*, **261**, 105–118.
- Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134, 112–113.
- Esnouf, R. M. (1999). *Acta Cryst. D* **55**, 938–940.
- Ferlenghi, I., Gowen, B., de Haas, F., Mancini, E. J., Garoff, H., Sjöberg, M. & Fuller, S. D. (1998). *J. Mol. Biol.* **283**, 71–81.
- Fuller, S. D. (1987). *Cell*, **48**, 923–934.
- Fuller, S. D., Berriman, J. A., Butcher, S. J. & Gowen, B. E. (1995). *Cell*, **81**, 715–725.
- Fuller, S. D., Butcher, S. J., Cheng, R. H. & Baker, T. S. (1996). *J. Struct. Biol.* **116**, 48–55.
- Gabashvili, I. S., Agrawal, R. K., Spahn, C. M., Grassucci, R. A., Svergun, D. I., Frank, J. & Penczek, P. (2000). *Cell*, **100**, 537–549.
- Grigorieff, N. (2000). *Acta Cryst. D* **56**, 1270–1277.
- Grimes, J., Basak, A. K., Roy, P. & Stuart, D. (1995). *Nature (London)*, **373**, 167–170.
- Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I. & Prasad, B. V. V. (1997). *Structure*, **5**, 885–893.
- Haas, F. de, Paatero, A. O., Mindich, L., Bamford, D. H. & Fuller, S. D. (1999). *J. Mol. Biol.* **294**, 357–372.
- Harrison, S. C., David, A., Jumbblatt, J. & Darnell, J. E. (1971). *J. Mol. Biol.* **60**, 533–538.
- He, Y., Bowman, V. D., Mueller, S., Bator, C. M., Bella, J., Peng, X., Baker, T. S., Wimmer, E., Kuhn, R. J. & Rossmann, M. G. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 79–84.
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). *J. Mol. Biol.* **213**, 899–929.
- Hewat, E. A., Verdaguier, N., Fita, I., Blakemore, W., Brookes, S., King, A., Newman, J., Domingo, E., Mateu, M. G. & Stuart, D. I. (1997). *EMBO J.* **16**, 1492–1500.
- Keller, P. A., Henrick, K., McNeil, P., Moodie, S. & Barton, G. J. (1998). *Acta Cryst. D* **54**, 1105–1108.
- Kielian, M. & Helenius, A. (1986). *The Togaviridae and Flaviviridae*, edited by S. Schlesinger & M. J. Schlesinger, pp. 91–119. New York: Plenum.
- Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). *Nature (London)*, **389**, 206–211.
- Kolatkhar, P. R., Bella, J., Olson, N. H., Bator, C. M., Baker, T. S. & Rossmann, M. G. (1999). *EMBO J.* **18**, 6249–6259.
- Lee, S., Owen, K. E., Choi, H. K., Lee, H., Lu, G., Wengler, G., Brown, D. T., Rossmann, M. G. & Kuhn, R. J. (1996). *Structure*, **4**, 531–541.
- Mancini, E. J., Clarke, M., Gowen, B., Rutten, T. & Fuller, S. D. (2000). *Mol. Cell*, **5**, 255–266.
- Mancini, E. J., de Haas, F. & Fuller, S. D. (1997). *Structure*, **5**, 741–750.
- Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y. & Fujiyoshi, Y. (1999). *J. Mol. Biol.* **286**, 861–882.
- Mueller, F., Sommer, I., Baranov, P., Matadeen, R., Stoldt, N., Woehnert, J., Goerlach, M., van Heel, M. & Brimacombe, R. (2000). *J. Mol. Biol.* **298**, 35.
- Nogales, E., Wolf, S. G. & Downing, K. H. (1998a). *Nature (London)*, **391**, 199–203.
- Nogales, E., Wolf, S. G. & Downing, K. H. (1998b). *Nature (London)*, **393**, 191.
- Paredes, A. M., Heidner, H., Thuman-Commike, P., Prasad, B. V., Johnson, R. E. & Chiu, W. (1998). *J. Virol.* **72**, 1534–1541.
- Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374.
- Skoging, U., Vihinen, M., Nilsson, L. & Liljestrom, P. (1996). *Structure*, **4**, 519–529.
- Smith, T. J., Cheng, R. H., Olson, N. H., Peterson, P., Chase, E., Kuhn, R. J. & Baker, T. S. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10648–10652.
- Stewart, P. L., Burnett, R. M., Cyrklaff, M. & Fuller, S. D. (1991). *Cell*, **67**, 145–154.
- Thouvenin, E. & Hewat, E. (2000). *Acta Cryst. D* **56**, 1350–1357.
- Thuman-Commike, P. A., Tsuruta, H., Greene, B., Prevelige, P. E. Jr, King, J. & Chiu, W. (1999). *Biophys. J.* **76**, 2249–2261.
- Toyoshima, C. & Unwin, P. N. T. (1988). *Ultramicroscopy*, **25**, 279–291.
- Wengler, G. & Rey, F. A. (1999). *Virology*, **257**, 472–82.