Single-particle electron cryo-microscopy: towards atomic resolution

Marin van Heel[1]*, Brent Gowen[1], Rishi Matadeen[1], Elena V. Orlova[1],
Robert Finn[1], Tillmann Pape[1], Dana Cohen[1], Holger Stark[1]†, Ralf Schmidt[2],
Michael Schatz[1,2] and Ardan Patwardhan[1]

[1] Imperial College of Science, Medicine and Technology, Department of Biochemistry, London SW7 2AY,
UK
[2] Image Science Software GmbH, Heilbronnerstraße 10, D-10711 Berlin, Germany

  * Author to whom correspondence should be addressed.
  † Present address: MPI für Biophysikalische Chemie, Am Fassberg 11, 37077 Göttingen, Germany.

1. Introduction

In the past few years, electron microscopy (EM) has established itself as an important – still upcoming – technique for studying the structures of large biological macromolecules. EM is a very direct method of structure determination that complements the well-established techniques of X-ray crystallography and NMR spectroscopy. Electron micrographs record images of the object and not just their diffraction patterns and thus the classical 'phase' problem of X-ray crystallography does not exist in EM. Modern microscopes may reach resolution levels better than ~ 1.5 Å, which is more than sufficient to elucidate the polypeptide backbone in proteins directly. X-ray structures at such resolution levels are considered 'excellent'. The fundamental problem in biological EM is not so much the instrumental resolution of the microscopes, but rather the radiation sensitivity of the biological material one wants to investigate. Information about the specimen is collected in the photographic emulsion with the arrival of individual electrons that have (elastically) interacted with the specimen. However, many electrons will damage the specimen by non-elastic interactions. By the time enough electrons have passed through the object to produce a single good signal-to-noise (SNR) image, the biological sample will have been reduced to ashes. In contrast, stable inorganic specimens in material science often show interpretable details down to the highest possible instrumental resolution.

Electron cryo-microscopy ('cryo-EM') – electron microscopy of hydrated specimens at liquid nitrogen or liquid helium temperatures – has proven indispensable for producing reliable images of intact biological structures. Cooling the specimen prevents the evaporation

of bound water from the molecules in the microscope's vacuum and therefore keeps the molecules in good shape ('structure preservation'). This approach has resulted in a quantum-leap improvement in specimen preservation compared to the earlier negative-stain preparation of dry specimens. The cooling also keeps the radiation damage localised, and therefore limited, which allows one to use a higher electron dose for the same level of radiation damage (Stark *et al.* 1996). Most of the current cryo-EM work is based on the vitreous-ice specimen-preparation technique in which a solution of molecules is rapidly frozen by plunging the EM grid into liquid ethane and then transferring to the electron microscope while cooled to liquid nitrogen temperatures (~ 100 K). After initial experiments by other groups (Taylor & Glaser, 1976; Chanzy *et al.* 1976), the practical problems of freeze-plunging and cryo-imaging techniques, namely how to freeze to form vitreous rather than crystalline ice, how to transfer without contamination, and how to prevent rapid ice build-up within the microscope vacuum, were investigated and solved by Jacques Dubochet and his colleagues at the EMBL in Heidelberg (Adrian *et al.* 1984; Dubochet *et al.* 1988) in the 1980s.

The cooling of the specimen leads to 'cryo-protection' factors of 10–30 (Stark *et al.* 1996) allowing a corresponding increase in the exposure dose applied. The higher dose directly leads to proportionally higher SNRs in the electron micrographs. However, for high-resolution imaging the tolerable levels of electron exposure at liquid helium temperatures amount to only 5–10 electrons/Å^2 and such exposure levels still lead to extremely noisy images. The noise variance in the high-frequency details of the image exceeds the amount of signal variance in that spatial-frequency range. Thus, the high-frequency SNR in the micrographs is much smaller than unity. To overcome this inherent noise problem in electron microscopy of biological macromolecules one is forced to image a large number of copies of the molecule, under the 'low-dose' conditions, and then to apply some form of averaging to improve the SNR of the data. The first natural choices of samples for such averaging procedures were regular aggregates of molecules like two-dimensional (2D) crystals (Unwin & Henderson, 1975; Henderson *et al.* 1990), helical tubes (Miyazawa *et al.* 1999) (which can be seen as 2D crystals folded back upon themselves); or icosahedral viruses (Crowther, 1971).

Cryo-EM of individual non-crystallised macromolecular assemblies ('single particles') has made substantial progress in the past years. This success includes the study of icosahedral structures. Purists may protest against the classification of icosahedral structures as 'single particles' because their high degree of symmetry considerably facilitates the reconstruction process and makes the process reminiscent of the analysis of helical fibres or even of 2D crystals. Icosahedral particles are simply particles with the highest possibly pointgroup symmetry (pointgroup: '532'). Icosahedral viruses were typically studied at resolution levels of ~ 35 Å (for an example of early work along these lines see Baker *et al.* (1988)). In 1997, a sequence of three papers on icosahedral structures (Trus *et al.* 1997; Conway *et al.* 1997; Böttcher *et al.* 1997) brought the resolutions achievable with the technique to less than 10 Å. Currently, the highest-resolution cryo-EM reconstruction exploiting icosahedral symmetry is the nominal 5·9 Å structure of Tomato Bushy Stunt Virus (TBSV) reviewed here, but the nominal 7·4 Å resolution of hepatitis B core structure (Böttcher *et al.* 1997) was based on the stringent 0·5 threshold criterion discussed below. Single-particle three-dimensional (3D) reconstructions of molecular assemblies with symmetry levels lower than these icosahedral ones have improved considerably since 1994. In that year, the first results of the zero-tilt 3D reconstructions appeared (Orlova & van Heel, 1994) performed with the 'angular reconstitution' approach proposed some years before (van Heel, 1987). The approach was

particularly successful in elucidating the entirely asymmetric structure of the bacterial ribosome.

Single-particle techniques can be separated into techniques requiring tilts of the specimen holder in the EM and 'zero-tilt' techniques. The classical tomography approach requires many pictures of the same particles tilted into different orientations. The technique, which traces back to – among others – the group of Professor Walther Hoppe (Hoppe *et al.* 1974; Oettl *et al.* 1983), has been rejuvenated by the introduction of automatic tomography where the computer controls the data collection in the cryo-EM (Dierksen *et al.* 1992). Another tilting approach requiring only two exposures of the sample is the Random Conical Tilt ('RCT') approach pioneered by Radermacher and co-workers (Radermacher *et al.* 1987; Radermacher, 1988). The (macroscopic) tilts required for these approaches intrinsically limit the achievable resolution level: the huge differences in defocus existing within the tilted image are very difficult to tackle computationally. The RCT approach includes explicit tilting of the specimen and may thus provide the absolute handedness of an object in cases where that information is not available. The issue of handedness determination is also discussed in Section 7.

Zero-tilt reconstructions exploit the random orientations that particles assume in the vitrified solvent to extract 3D information. Because no explicit macroscopic tilting of the sample holder is required, zero-tilt single-particle approaches can intrinsically reach very high resolution levels. The two main zero-tilt techniques are 'projection matching' and 'angular reconstitution' (for more detailed comparisons between single-particle approaches see van Heel *et al.* (1997) and Schatz *et al.* (1997)).

Projection matching is based on finding the projection directions of the individual molecular images by correlating the images to 're-projections' of an earlier 3D reconstruction (van Heel, 1984b; Harauz & Ottensmeyer, 1984; Harauz & van Heel, 1986a). Harauz & Ottensmeyer (1984) used projection matching to orient individual nucleosomes for 3D reconstruction relative to an earlier model structure. It was shown (van Heel, 1984b; Harauz & van Heel, 1986a) that it is even feasible to first assign Euler angles using a random number generator and then to iteratively refine the Euler-angles assignments by projection matching. Various forms of projection matching have been applied to icosahedral structures (Baker & Cheng, 1996) and to asymmetric structures (Frank *et al.* 1995). More general and more powerful is the angular reconstitution approach, which is based on finding common line projections between 2D projection images of the 3D structure (van Heel, 1987; Orlova & van Heel, 1994). Noise reduction by multivariate statistical classification of the raw EM projections is an integral part of the approach. Angular reconstitution has already yielded structures at resolution levels of 5–9 Å resolution of a number of different specimens, including two of the examples used below.

The structure analysis of single particles by cryo-EM involves a large number of techniques and algorithms that are all essential for achieving quasi-atomic levels of resolution. In this review, we give a complete overview of single-particle cryo-EM by our angular reconstitution, including various (unpublished) critical details of our latest developments. It is already feasible to obtain structures at 5·9 Å resolution with the approach, sufficient to resolve β-sheets in an entirely β-sheet structured virus. We are, however, by no means fundamentally limited to this resolution level. We discuss the developments required for achieving resolution levels better than $\sim$ 4 Å, that is, for achieving resolution levels allowing interpretation of the macromolecular maps in terms of atomic co-ordinates.

2. Electron microscopy

2.1 Specimen preparation

One of the most critical – yet not always predictable – aspects of modern cryomicroscopy is the specimen preparation. It is here that the art and the science of cryo-EM meet. Many projects have been stalled or shelved completely owing to a lack of good images. When lack of material or the insufficient purity of the specimens is to blame, this may be beyond the control of the microscopist. The full creativity and devotion of the microscopist can be challenged, however, where it concerns control of the behaviour of single molecules in their interaction with the EM grid. What good are samples that fully clump together on the edges of the holes in the carbon foil? Purists, including some of the authors of this review, insist that vitrified water cryo-EM is the *only* acceptable form of cryo-microscopy. Others, however, believe that anything goes that keeps the samples happy. Water may be used directly or may be substituted or complemented by polar replacements such as glucose (Unwin & Henderson, 1975), trehalose (Hebert *et al.* 1997; Walz *et al.* 1997), tannin, etc. Indeed, most atomic resolution structures determined by electron crystallography have been elucidated using such carbohydrates as embedding media. For single-particle methodologies vitreous-ice embedding remains the prevailing specimen preparation technique, although glucose negative stain mixtures (Orlova *et al.* 1997), and vitreous-ice negative stain solutions (Boettcher *et al.* 1996, 1999; Adrian *et al.* 1998) have also been used successfully. The famous last words on these issues have not yet been spoken. Single-particle methodologies have only matured during the past few years and only now we can start objectively assessing the various embedding media in terms of their 3D results.

The hydrophobicity properties of the grids and the support films, and the humidity of the environment are important parameters for successful sample preparation, yet these parameters are not that easy to control. Controlled environment chambers (Bellare *et al.* 1988) may be used to regulate the local atmospheric environment during freezing. A recent, very positive development is the use of semiconductor fabrication technology to create holey films with regular, predetermined hole geometries (Ermantraut *et al.* 1998). Apart from increasing the reproducibility of specimen preparation, one may expect that such grids will play an important role in computer-controlled EM data collection (Kisseberth *et al.* 1997). Still, the number of parameters in the production of good samples is large and not entirely controllable. For optimising the specimen preparation conditions it is essential to have direct access to an entry-level cryo-EM. Once good samples have been obtained, the frozen vitreous-ice samples can be stored 'forever' in liquid nitrogen. There then is enough time to find the best possible microscope for the final data collection.

2.2 The electron microscope

The typical microscope with which much of the current cryo-EM work is performed is a 200 kV instrument equipped with a highly coherent field emission electron gun (FEG) and a side-entry, liquid nitrogen-cooled specimen holder. An FEG substantially improves the transfer of information through the microscope – especially in the high-resolution domain – when compared to a conventional hairpin tungsten filament gun. There are now probably more than 30 200 kV FEG microscopes from Philips, JEOL or Hitachi in use for this type

of high-resolution cryo-EM. Microscopes equipped with a liquid-helium-cooled stage are available from JEOL and Philips. Until recently, the only two liquid-helium microscopes in the world were the JEOL 3000SFF or Dr Yoshinori Fujiyoshi in Kyoto (Fujiyoshi *et al*. 1991) and the Philips CM20 'Sophie' of Dr Fritz Zemlin in Berlin (Zemlin *et al*. 1996). The first commercial Philips CM300 FEG with a liquid-helium stage has been installed in our London laboratory. JEOL has installed a number of liquid-helium-cooled stage 300 kV FEG microscopes. Apart from the benefit derived from improved cryo-protection actor at liquid helium temperatures, this new generation of top-entry or detached-tip stages has an improved stability over the liquid-nitrogen cooled side-entry holders. Note that the modern cryo-EM is often equipped with a 1024 × 1024 or 2048 × 2048 cooled CCD camera under computer control. In automated tomography studies (Dierksen *et al*. 1992) the CCD serves directly as the data collection device. For high-resolution molecular structure analysis, however, the device serves primarily for optimising the electron microscopical imaging conditions. CCD cameras may become important in the context of remote control of the expensive (shared) instrumentation (Kisseberth *et al*. 1997).

2.3 Acceleration voltage, defocus, and the electron gun

An important aspect of choosing the correct electron microscope for data collection is the choice of acceleration voltage. The general rule-of-thumb is that the higher the voltage, the higher the achievable resolution levels. However, the price tag of the instrument is proportional to (at least) the square of the acceleration voltage of the instrument. Thus, one will typically use a 100–200 kV microscope for optimising the specimen preparation conditions. For achieving resolution levels better than ~ 10 Å, one will then move to an expensive 200 kV or even a 300 kV instrument for data collection. The contrast transfer function (CTF) of the instrument reaches higher resolution levels with increasing voltage (Fig. 1(*a*) and (*b*)).

Close to Scherzer focus, the spherical aberration ('Cs') and the defocus level combine to create a broad band of uninterrupted information transfer up to the resolution where the CTF crosses zero for the first time (Fig. 1(*a*) and (*b*)). This very high resolution is not necessarily relevant for cryo-EM of single molecules. The highest instrumental resolution is reached close to 'Scherzer focus', and at such defocus values the transfer of the low-frequency information is extremely poor. The low frequency information is, however, essential for locating the particles in the micrographs. Thus, even though high-frequency information is present in the images taken close to focus, the noisy particles are no longer discernible against the background due to the missing low-frequency information and thus they are simply lost. Therefore, one will generally work at relatively large defocus values (up to 3 μm). At large defocus values, the CTF oscillates many times in the high-frequency domain (Fig. 1(*c*) and (*d*)).

When a conventional hairpin-filament electron gun is used to illuminate the specimen, the coherence of the illumination is substantially lower than when a Field Emission Gun (FEG) is applied. When the ('lateral') coherence is low, the CTF is averaged over an area of the size of the source in the back focal plane of the objective lens (van Heel, 1978). In other words, the partially coherent transfer function is the 'convolution' of the coherent CTF with the shape of the illuminating source. As a consequence, the image information in the frequency

ranges where the CTF oscillates rapidly, will be smeared out. When a FEG source is used, the image information in frequency ranges where the CTF oscillates can still be recorded in the micrographs. The oscillations need to be computationally corrected for after the image has been recorded ('CTF correction'). Thus, the use of a FEG (Zemlin *et al.* 1996) is indispensable for high-resolution imaging (better than $\sim$ 10 Å) of individual macromolecules at these large defocus values. These rapid CTF oscillations at high spatial frequencies have an interesting side effect: they cause the information in the image to be spread over a very extended point spread function (PSF). This issue (illustrated in Fig. 2) will be discussed in detail in Section 8.

As was mentioned above, the higher 'Scherzer focus' resolution of the cryo-EM images achieved at 200–300 kV as opposed to 100 kV is currently not of large practical importance for biological cryo-EM. What is very important is the effect of the thickness of the specimen on the effective transfer of high-resolution information. The differences in defocus over the height of the specimen (the size of actual particle, or the full thickness of the vitreous-ice layer in which it is embedded) will cause smearing out of the high-frequency image information, which is difficult to compensate for. Without any form of compensation, the net effect is that of an uncertainty in the exact defocus value for the CTF correction, which leads to a 'defocus envelope'. The effect of this envelope damping function is shown in Fig. 1(*e*) and (*f*) for the 100 kV and the 300 kV situation, and assuming a defocus uncertainty of $\sim$ 400 Å. For a 300 kV microscope (Fig. 3), the specimen is effectively much 'thinner' than for a 100 kV microscope and the damping is thus much reduced at 300 kV. This effect is also known as 'curvature of the Ewald sphere' and we here see the consequences of the fact that the radius of the Ewald sphere is much smaller at 100 kV than at 300 kV (DeRosier, 2000).

Another much-discussed issue in this context is that of the contrast achieved in the images as function of the voltage. At higher voltages – owing to the reduced interaction with the specimen – the contrast in the images is lower. (The contrast in an image is defined as the standard deviation of the modulations in the image due to the presence of the object, divided by the average background exposure level.) From this observation some workers prefer to work at reduced voltages, say 120 kV, even when the instrument allows one to work at 200 kV (Conway *et al.* 1997). However, contrast is only one part of the story. At higher voltages the radiation damage associated with a given exposure dose is also reduced and the 'survival dose' may be simply proportional to the acceleration voltage as was measured for solid nitrogen crystals at liquid-helium temperatures (Zemlin *et al.* 1999). Thus, to compensate for the lower image contrast, one may increase the exposure level, which, in turn will lead to an increased SNR in the images. As a net effect, the amount of information one can collect from a single particle does not necessarily change with increased acceleration voltage. Note that most 'atomic-resolution' reconstructions of 2D crystals have been performed on 2D crystalline object that are thin (less than $\sim$ 60 Å thick), on 200–300 kV FEG cryo-EMs mostly operating at liquid-helium temperatures (Kimura *et al.* 1997; Nogales *et al.* 1998).

## 2.4 Magnification and data collection

The best magnification for data collection is – in principle – dictated by the radiation sensitivity of the biological material. Assume, for example, that a sample can tolerate a maximum exposure dose of 10 e$^-$/Å^2 for preserving the 3 Å information one wants. It is important to collect the micrographs at approximately this exposure level, since that is the

(a)

(b)

(c)

(d)

(e)

(f)
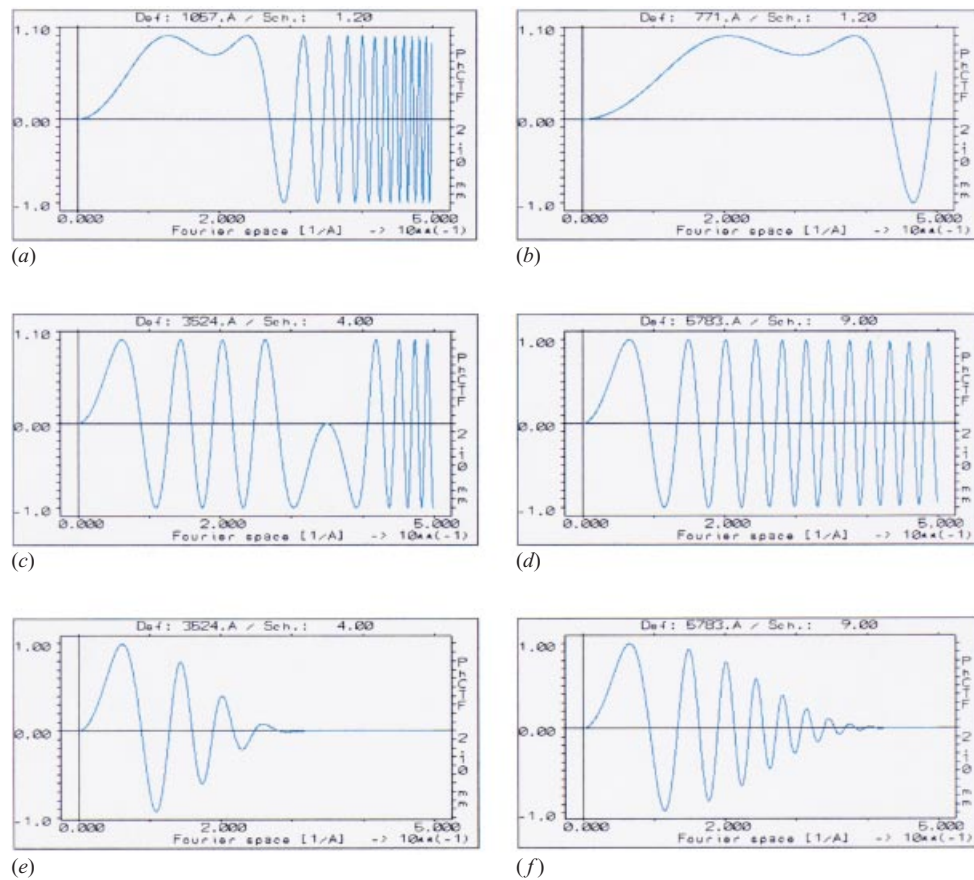
Fig. 1. Influence of voltage and 'defocus envelope'. These idealised PhCTF curves are calculated for field emission (FEG) systems and exclude the influence of partial coherent illumination, chromatic aberration, etc.: they serve exclusively for illustrating the limiting effect of specimen thickness. Since no 100 kV FEG systems are commercially available, the 100 kV curves would apply to the available 200 kV or 300 kV systems used at a lower voltage. At 1·2 Scherzer defocus (a, b)(, the CTFs reflect the difference in 'genuine' instrumental resolution between the two instruments. The two 'working' defocus values chosen (c, d) lead to a comparable contrast in the low-frequency components of the images. The CTF 'working' curves are chosen as close as possible to Scherzer focus, at an approximate limit of detectability of large single particles (small viruses, ribosomes) embedded in vitreous ice. The highest frequency (the 'Nyquist' frequency) in all curves corresponds to 1 Å/pixel, or a spatial frequency of 1/2 Å. (a) The phase contrast transfer function ('PhCTF' or, for short, 'CTF') for a hypothetical 100 kV microscope equipped with a field emission gun ('FEG'). (b) the CTF of the Philips CM300 FEG (Fig. 2) at the same Scherzer defocus parameters (1·2 Scherzer) as the 100 kV instrument depicted in (a). (c) CTF of a standard EM at 100 kV (defocus: 6 Scherzer or ∼ 2600 Å underfocus; infinitely thin object). (d) CTF at 300 kV with comparable parameters (defocus: 10 Scherzer, or ∼ 6500 Å underfocus; infinitely thin object). (e) CTF of the 'standard' EM at 100 kV assuming a 400 Å defocus uncertainty (defocus: 5 Scherzer or ∼ 2600 Å underfocus). (f) CTF at 300 kV with comparable parameters (defocus: 10 Scherzer, or ∼ 6500 Å underfocus; 400 Å defocus uncertainty).

level at which the SNR of the raw input data is maximised. We need to maximise the information content in the molecular images. Note that the preferred definition of the SNR is that of the variance of the signal over the variance of the noise. Some researchers – unfortunately – define the SNR as the standard deviation of the signal over the standard deviation of the noise. The later definition leads to the somewhat awkward conclusion that
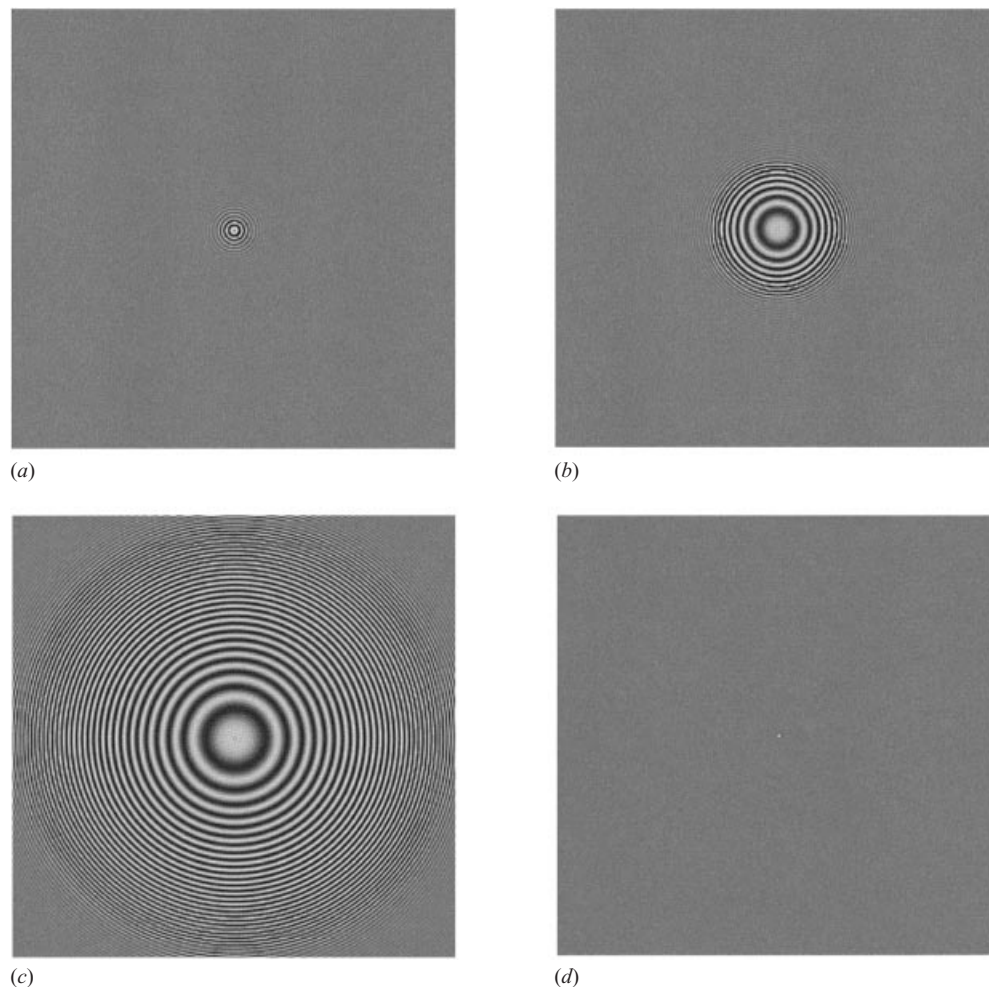
Fig. 2. Extent of the point spread function (PSF) in real space. The PSF is the Fourier transform of the CTF and represents an alternative way of describing the same concept. The PSF, at defocus values that are typically used in cryo-EM, lead to a spread of the information of a single 1 Å 'pixel' of the image (256 × 256 pixels), over an area as large as an *E. coli* ribosome. The defocus values used here, in part correspond to the CM300 defocus values used in Fig. 1. The calculations were performed with the 'Impose CTF' option of the IMAGIC TRANSFER program (see main text). (*a*) PSF of the CM300 at 1·1 Scherzer (∼ 800 Å underfocus). (*b*) PSF CM300 at 9 Scherzer (6000 Å underfocus). (*c*) PSF CM300 at ∼ 23 Scherzer (1·5 μm underfocus). (*d*) PSF CM300 at 9 Scherzer (6000 Å underfocus) after CTF correction by phase flipping.

the SNR increases with the square root of the exposure dose whereas with the first definition the SNR increases proportionally to the exposure. Although both definitions are possible, the variance-based SNR definition reflects better the aspect of 'information harvesting' in noise-limited data. We need to maximise the 'information' collected from each molecule since this will minimise the number of molecular images required to achieve the required resolution and will facilitate all further data processing. Information – as originally defined by Shannon in 1948 (Shannon, 1948; van Heel, 2000b) – is defined as:

$$I = {}^2\log(1 + \text{SNR}). \tag{1}$$
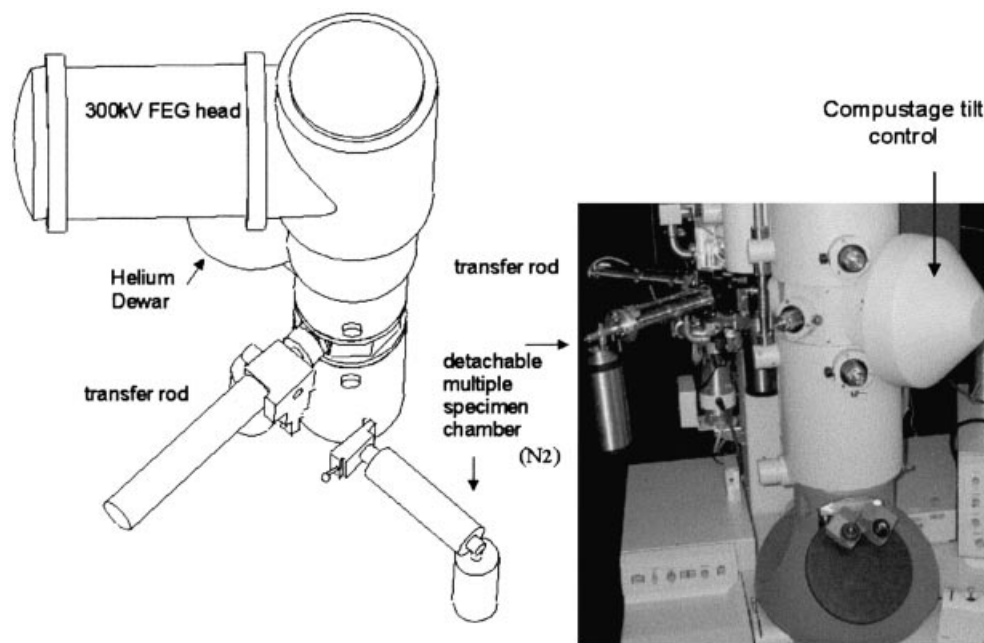
**Fig. 3.** The Philips CM300 liquid helium microscope. This is the first Philips 300 kV FEG electron microscope equipped with a liquid-helium cooled specimen stage. A unique property of this stage is that it can also operate at liquid-nitrogen temperatures. One may switch between the modes of operation without changing the specimen.

For noise-limited images (or for noise-limited spatial frequency ranges of the images), the SNR is small compared to unity, and the information collection becomes proportional to the (variance-based) SNR and to the exposure level. The closer one operates to maximum tolerable exposure, the more information is collected per individual molecular image.

Photographic emulsions – still the best means of collecting the electron microscopic data for our purposes – will operate well at range of exposures leading to optical densities of $\sim$ 0·2–1·5 OD. The density of photographic film reacts linearly to electron exposure (Valentine, 1965), and this range of optical densities may correspond to exposure levels of, say, 0·1–2·0 electrons/μm^2, depending on the photographic material and the development specifics. Within certain limits, the film leaves flexibility to match the data collection to the properties of the densitometer one uses. Photographic film is a very good electron image transducer in that the granularity of the image is largely determined by the electron counting statistics, and not so much by the granularity of the film (Valentine, 1965). More modern research has also indicated that the properties of the photographic film are still hard to beat. In an eye-opening study on this issue it was shown that images from a CM300 FEG, taken at a magnification of only 10 500, contained useful information up to the 2·04 Å resolution level (Cheng & Taylor, 1998). That study was performed using an optical diffractometer (OD) to evaluate the micrographs.

It is good practice to check the quality of each micrograph in the OD rather than to digitise all micrographs collected because many micrographs are ruined by drift, vibrations, etc., and thus do not achieve the expected theoretical limits. As an example, for cryo-EM data

(*a*)   (*b*)
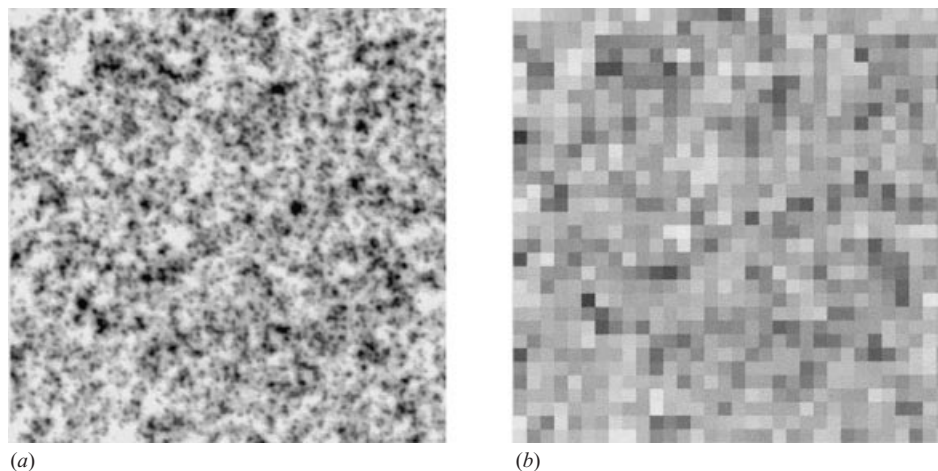
Fig. 4. Photographic emulsion exposed to electrons. (*a*) A small (256×256 pixels) area of a Kodak SO163 emulsion exposed to electrons without a sample in the microscope, and digitised here at $1 \cdot 1$ µm per pixel (equivalent to 23 100 dpi or 'dots-per-inch'). The emulsion has been exposed to an estimated $\sim 0 \cdot 5$ e$^-$/pixel. At an EM magnification of 50 000 $\times$, each pixel would have corresponded to $0 \cdot 22$ Å, and the total exposure to ~ 10 e$^-$/Å². The images of grains are somewhat blurred due to the limited numerical aperture of the microscope lens ($0 \cdot 19$) used for recording this image and due to the thickness of the emulsion, making it impossible that grains positioned at different heights in the emulsion are simultaneously in focus. (*b*) For comparison, the same area is shown as it would be seen by a densitometer with a $\sim 9 \times 9$ µm² sampling aperture (~ 3200 dpi). At an EM magnification of 50 000 $\times$, each pixel would have corresponded to $\sim 1 \cdot 8$ Å/pixel. A prerequisite for achieving 'atomic' resolutions is the digitisation of the micrographs at ~ 1 Å/pixel. Although only few high-speed densitometers can reliably reach 9 µm/pixel, the electron emulsions can, albeit at low contrast, resolve periodic information requiring $1 \cdot 0$ µm per pixel digitisation (see main text).

collection on a CM300 FEG at liquid-helium temperature, in a project aimed at achieving resolution levels below 4 Å, we expose the sample at ~ 10 e$^-$/Å². Using a magnification of $\sim 50 000$, this leads to micrographs (Kodak SO163, developed in D19 for 12 min) with an estimated average optical density of $\sim 0 \cdot 6$. These micrographs are then digitised using a pixel size of 5 µm $\times$ 5 µm corresponding to ~ 1 Å $\times$ 1 Å on the specimen scale. To familiarise the reader with the typical appearance of an electron micrograph at very high resolution, we have digitised a small area of a photographic emulsion (Kodak SO163) exposed to electrons (without a sample in the microscope) at an estimated density of $\sim 0 \cdot 2$ e$^-$/µm² (Fig. 4(*a*)). This image was digitised using the Image Science patchwork densitometer in combination with a $6 \times$ microscope lens with a numerical aperture of $0 \cdot 18$. The image consists of a 256×256 pixel area with a pixel size of $1 \cdot 1$ µm. By binning 8×8 pixels, we emulate what the image would look like when digitised using a more standard ~ 9 µm $\times$ 9 µm pixel size (Fig. 4(*b*)).

3. Digitisation and CTF correction

Micrographs taken under optimal imaging conditions can contain massive amounts of information. Indeed, the experiment by Cheng & Taylor (1998) indicates that it can make sense to digitise a micrograph using a pixel size smaller than 1 µm $\times$ 1 µm (~ 1 Å $\times$ 1 Å at

10 000 magnification). At 1 μm × 1 μm/per pixel and 8 bits/pixel, one electron micrograph could produce up to 80 000 × 100 000 bytes ≈ 8 gbyte of information. The Cheng & Taylor (1998) experiment was aimed at finding the absolute limits of the technology. In reality, one would probably never use a scanning step size below 2 μm × 2 μm so as to not introduce serious limitations due to the size of the silver-halide crystals in the emulsion (Fig. 4(*a*)) which makes the modulation transfer function (MTF) of the film drop to too low levels. In our typical scanning situations a single 8 cm × 10 cm micrograph recorded at 50 000 × magnification and using a 5 μm × 5 μm (or 3 μm × 3 μm) sampling aperture, yields ∼ 320 000 000 (800 000 000) pixels, or around a Gigabyte of information per micrograph.

3.1 The patchwork densitometer

Densitometry at a few μm/pixel has been a problem in cryo-EM. The best densitometers traditionally were the flatbed densitometers such as the Perkin–Elmer PDS 1010. These densitometers, however, require one mechanical movement per digitised pixel and can therefore only digitise on the order of ∼ 1000 pixels/s meaning that it could take up to 250 h (10 days!) to scan a single micrograph. Data sets needed for single particle analyses can easily consist of a few hundred micrographs. For our purposes, a densitometer must be capable of a sustained data collected rate that is at least two orders of magnitude higher than that of the previous generation, i.e. a rate in excess of ∼ 100 000 pixels/s. Rotating drum densitometers had a poor reputation 25 years ago, but they appear to have made a revival. According to their specifications, some are now capable of operating in the 5–8 μm/pixel high-resolution high-speed realms, in spite of belonging to the category of instruments requiring one mechanical movement per pixel.

Most of today's scanners are hybrid scanners in the sense that for one mechanical movement they capture many pixels using an image sensor that contains many sensitive elements, say, a linear CCD array. Why the need for hybrid scanners? There are currently no image sensors that can match the information content of a whole micrograph. Moreover, there are also fundamental limits to the amount of information that can be transferred through a light-optical system. High-resolution densitometers for electron microscopy thus typically implement some form of mechanical scanning of a multi-element optical sensor. The Zeiss SCAI Scanner, for example, scans a linear CCD array in front of the negative. This top-of-the-range scanner is capable of high-speed scanning of a micrograph at 7 μm × 7 μm resolution. A disadvantage of linear CCD scanners is that when the number of pixels required in the direction of the linear CCD chip exceeds the length of the linear array, different stripes must be mounted together to form a continuous image. This is mechanically a difficult task requiring a high level of positioning accuracy. In our experience, we found that in some low-cost scanners, the mechanical positioning required for moving the linear CCD chip (or the negative) from one line to the neighbouring one was not sufficiently precise.

We have constructed a new type of hybrid 'checkerboard' densitometer (Fig. 5; Patwardhan *et al.* 2000a) consisting of a computer-controlled *x–y* table, a cooled 2D CCD-camera, and a stable light source illuminating only that area of the micrograph that is seen by the camera. The densitometer digitises the micrograph checkerboard-wise as (overlapping) single patches, the high-precision *x–y* stage being used for moving the micrograph from one patch to the next. Image information from overlapping regions can be used to estimate the
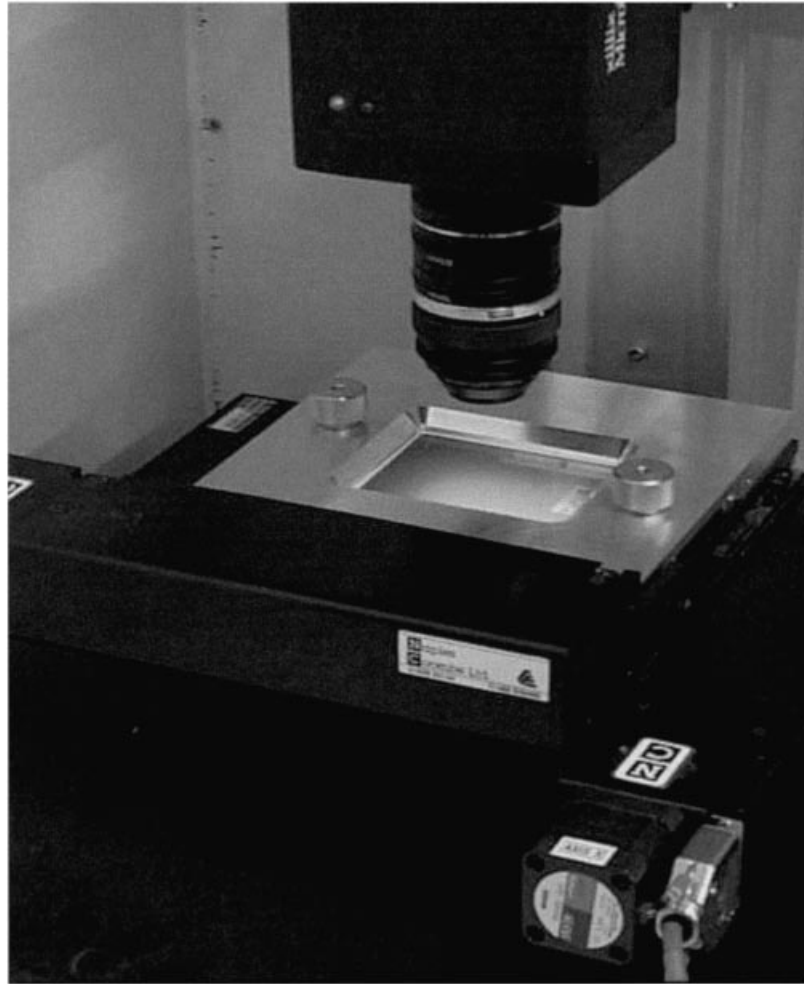
Fig. 5. The ImSc chequerboard densitometer. This high-speed high-resolution densitometer collects individual patches of $\sim 1536 \times 1024$ pixels (on a cooled CCD camera) from the full required area of the micrograph. The digitised overall image is created in the computer by integrating all patches into a contiguous digitised micrograph. For details see main text.

relative positions of neighbouring patches by means of cross correlation functions. The measured refined positions of the patches can then be used to 'stitch' the individual patches into a huge single image covering the entire micrograph. Alternatively, depending on the required resolution, only may simply exploit the reproducibility precision of the x–y table to direct the micrograph to the next field. The software required to control the system is implemented under the IMAGIC image processing system (van Heel *et al.* 1996).

The no-compromise design of the patchwork densitometer gives a number of advantages over other currently available systems. In the first place, the resolution levels can be chosen from a continuous range of, say, 2 µm × 2 µm up to 50 µm × 50 µm, to fit the specific EM experiment. This range of sampling choices can be achieved by using an appropriate lens. If we assume a pixel size of 7 µm × 7 µm on the CCD chip, then a ×4 microscope objective or macro-lens will lead to a sampling step of 1·75 µm on the micrograph. A lens optimised for

imaging at $\sim 1:1$ could cover continuous sampling ranges from $4\,\mu m$ to $10\,\mu m$, etc. A second very important aspect is the speed of data collection. A cryo-EM data set may consist of 100 micrographs, each of which must be digitised at, say, $5\,\mu m \times 5\,\mu m$. The patchwork densitometer can sustain very high data-collection rates because only one mechanical positioning (~ 1 s) is required to collect all pixels covered by the CCD chip. Let us assume the (cooled) CCD camera supports 1024×1024 pixels at 12 bits per pixel, and that it takes ~ 1 s to read the contents of the CCD and 1 s to store the collected information (~ 2 Mb) onto the hard disk. With these parameters, a continuous data-collection rate of ~ 300 Kpixels/s can be sustained. A full micrograph at $16\,000 \times 18\,000$ pixels ($5\,\mu m \times 5\,\mu m$) can thus be scanned in ~ 1000 s (15 min). At the lower resolution level of $7\,\mu m \times 7\,\mu m$ the digitisation takes only half as long; and again (and half as long as that if one were to collect only 8-bit data rather than essentially 16-bit data). With such a high-speed densitometer, the processing of 100 micrographs scanned at $5\,\mu m \times 5\,\mu m$ would thus take ~ 25 densitometry hours (~ 12 h at $7\,\mu m \times 7\,\mu m$) and is doable. Such data-collection operations would not have been conceivable with the earlier generation of 'single-pixel' flatbed densitometers. Note that we are here speaking of raw data sets (before processing) of around 60 Gigabyte. The handling of such massive amounts of data requires – as is discussed below – a correspondingly large computational infrastructure.

3.2 Particle selection

After digitisation, the individual particle images must be pinpointed in the micrographs and this task is critical as was indicated above. The problem here is – again – that the closer to Scherzer focus an image is taken, the better the high-resolution data components are preserved but the weaker the low-frequency data components that determine the visibility of the individual particles. Interactive particle selection still prevails although automatic particle picking has been possible for a long time (van Heel, 1982; Frank & Wagenknecht, 1984; Lata *et al*. 1995). The simplest objects to find in a cryo-EM image are spherical viruses: cross-correlation with a rotationally symmetric average shape of the virus is reliable in finding virtually all viruses in the image (Conway & Steven, 1999). For more asymmetric shapes, other, more elaborate approaches may be required and the algorithms are still rough. Automatic particle selection results thus need to be verified visually and may not necessarily lead to significant savings in time. It is obvious that more development must be done on this aspect of processing since the amount of information collected is directly proportional to the number of particles used.

An approach that has been applied to select barely visible icosahedral viruses from close-to-focus images is to take a second high-defocus image after taking one close-to-focus. Particle picking can then be performed on the second image (and the first alignments and orientation searches, see below) and can then be applied to the close-to-focus images. The relative orientations of the two micrographs must, of course, be known precisely in order to be able to apply the co-ordinates found for one image to the other. One can, in principle, also merge the information from two or more micrographs, taken at different defocus values to create a single image in which both the low-frequency components from the second exposure and the high-frequency components from the first exposure are present. With such a merged micrograph all subsequent alignment/reconstruction procedures can be applied to same basis data set.
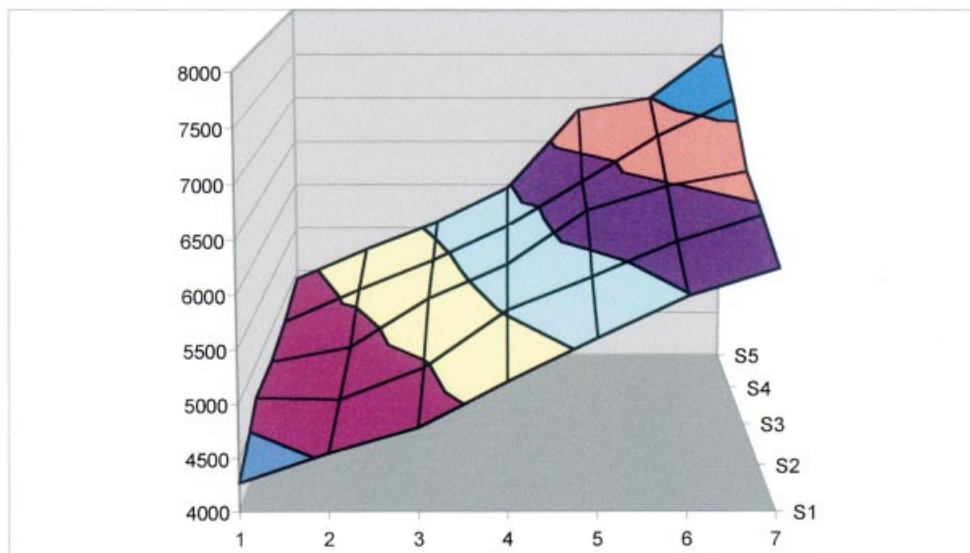
3.3 Position dependent CTF correction

The determination and the correction of the CTF is a multiphase procedure and, given the crucial importance of the correction in achieving high-resolution results, is the subject of continuing research and development. It is most common at present to correct the CTF at the end of the 3D reconstruction procedures for each micrograph, or for sets of micrographs in a given defocus range separately (Böttcher *et al.* 1997; Conway *et al.* 1997; Orlova *et al.* 1997). We now prefer to perform the primary CTF correction directly after the digitisation of the micrographs and prior even to the actual particle selection. We first calculate a defocus 'map' characterising the defocus variations over each micrograph. The full micrograph is subdivided into, say, 2048×2048 pixel patches for which the defocus and astigmatism parameters are determined. The 2048×2048 fields can be either equally spaced over the micrograph or spaced at arbitrary positions over the micrograph. Based on these measurements, a 'map' is generated covering the full micrograph that characterises the defocus and astigmatism values over the image. The assumption that is often made in cryo-EM is that all particles in one micrograph have the same defocus and thus must be corrected with the same CTF. However, if we assume that the grid is tilted by $\sim 2°$, then, at $\times 50\,000$ magnification, we can easily have defocus differences of over 500 Å within the micrograph. Such defocus uncertainties add to the effect of the defocus envelope (see above) and obviously have a detrimental influence on the overall resolution. Interestingly, we found a systematic 6° tilt of the Gatan specimen holder of our side-entry Philips CM200 FEG microscope in our measurements (Fig. 6). On that instrument we now routinely collect data with the holder tilted to $-6°$ to compensate for this effect.

   The CTF correction is applied to each extracted particle separately, based on the defocus parameter associated with its position in the micrograph. The correction is performed by flipping the phases of the negative lobes of the CTF. At this stage of the processing the images are limited by noise and it is thus not advisable to boost the amplitudes in the areas in Fourier space where the CTF curve is small. In a typical data set, the molecular images cover a relatively broad range of defocus values. Thus, in areas in Fourier space where some images have a minimum transfer of information, others will have a maximum and overall one will attain a smooth effective CTF. One great advantage of working with data that have been CTF corrected in this early phase of the analysis is that all further processing will be with images that have synchronised phases. Only in the final part of the processing, where one wants to filter the resulting 3D reconstructions to correct for the low strength of the high-frequency components, will the issue of the CTF correction surface again.

3.4 Precision of CTF determination

The precision with which the exact defocus and astigmatism parameters can be determined is of crucial importance for the CTF correction. It is not simple to measure the CTF in low-contrast vitreous-ice specimens. Any flaw in the measurement of these parameters adds a corresponding component to the defocus envelope discussed above (Fig. 1(*e*) and (*f*)) which can cause severe suppression/elimination of the high-resolution image components while integrating all measured data into the final 3D reconstruction. We have designed an algorithm – entirely based on correlation techniques – for finding the defocus parameters using all available information as efficiently as possible. This algorithm is typically applied to the spectra of relatively large ($\sim 2048 \times 2048$ pixel) extracted patches covering much of the

Fig. 6. Position dependent PhCTF determination. This graph illustrates a typical nominal '0' tilt on our CM200 cryo-EM/Gatan cryo-holder system. The system exhibits a systematic 6° tilt with respect to the nominal tilt angles that, if not corrected for, causes a defocus spread of almost 3000 Å. After our diagnostic analysis, the holder is now systematically used at a nominal −6° tilt to compensate for this effect. However, the defocus difference between front and back of the plot of ∼ 600 Å – perpendicular to the tilt axis of the goniometer – is not correctable with the current set up. Moreover, due to a recent repair of this particular holder, the nominal '0°' tilt position requires recalibration.

original micrographs. (This general algorithm can also be applied to sums of spectra of patches of the original image or to the extracted molecular images themselves.) For each patch, the amplitude spectrum is calculated and the resulting 2D spectrum is then high-pass filtered to remove the characteristic background ramps associated with spectra of EM images. Note that this filtering is performed by 2D-Fourier transforming the spectra and then multiplying the transforms by a rotationally symmetric Gaussian high-pass filter. At the same time, a broad Gaussian low-pass filter is applied to reduce the large amount of noise in the direct spectra. The theoretical CTF is then fitted to the treated spectra of the patches. The full cross-correlation coefficients between the measured spectra and theoretical CTF functions are calculated for all reasonable defocus parameters. In this correlation calculation all spectrum areas (and not just those of the zero crossings) contribute to the fitting procedure. The maximum correlation coefficient indicates the correct defocus parameter.

When the image is not free of astigmatism, the elliptically shaped spectra are first aligned with the main axes pointing vertically. This operation is performed by finding the rotational alignment angle between each spectrum and a mirrored version of itself. Rotation of the spectrum by half of the rotation angle then places the main axes of the ellipsoids along the X- and Y-axes of the image frame. Another copy of the spectrum, that is orthogonal to the first, is then anisotropically scaled along the x- and y-axes until it matches the first. Thus, not only the orientation but also the magnitude of the ellipticity is obtained. This information is used to elliptically average the 2D spectrum into a 1D curve, which is then fitted to a theoretical non-astigmatic CTF curve. Our CTF correction algorithm will be published elsewhere in full detail (Patwardhan *et al.* 2000b). The final CTF correction is performed on each extracted

particle separately, by applying the CTF parameters found for the environment from which each particle was extracted. This method gives much better results than methods where one assumes all particles from one micrograph to require the same CTF correction. Further improvements may be required for achieving resolution levels better than 4 Å. Our current implementation does not take into account the fact that individual particles may be positioned at different heights in the ice layer and it also does not (yet) take into account the defocus differences existing within the height of the particle itself (see Section 8, below).

4. Single particles and angular reconstitution

Angular reconstitution, in its combination with multivariate statistical techniques to classify and average the characteristic views of a molecule, forms a complete, self-contained methodology for the high-resolution 3D structure analysis of uncrystallised macromolecules by electron microscopy (van Heel, 1987). No crystals need to be grown for single particle approaches, and thus the specimen preparation and data collection can be fast and simple. The molecules in solution are not constrained by the crystal contacts, which may introduce artefacts. The time scale of the standard specimen preparation techniques for cryo-EM is measured in seconds under normal circumstances and may be reduced to milliseconds in order to visualise short-lived functional states of the molecules (Unwin, 1995). In contrast to these values are the time scales of sometimes weeks associated with crystallisation procedures for X-ray diffraction. The one advantage of X-ray crystallography is that of the superior resolution. Almost all larger macromolecules and complexes that have been solved to atomic resolution have been solved by X-ray crystallography. Much would be gained if the single-particle methodologies would be improved to attain resolution levels allowing atomic interpretation of the resulting maps.

Single molecules in solution are not held in fixed orientations, and they thus have six degrees of freedom: three translational ones; X, Y, Z, and three rotational ones corresponding to three 'Euler angles'; α, β, and γ. The image of the vitrified solution created in the electron microscope is a projection along the Z direction (by choice of the co-ordinate system). If we assume for the time being that the EM provides us with an ideal projection along Z (assuming an idealised infinite depth of focus), we have five parameters to determine for every particle image: the two 'in-plane' translational degrees of freedom, X and Y; the in-plane rotational degree of freedom, α; and two out-of-plane degrees of rotation β and γ. The in-plane degrees of freedom can be removed by various forms of alignment such as the alignment by classification (see below). The out-of-plane rotations of the molecules are the ones that provide the 3D information. The two degrees of freedom of out-of-plane rotations of the individual molecules must be identified before one can perform a 3D reconstruction. The processing of the single-particle images is aimed at determining these parameters for each of the individual particles in the data sets.

4.1 Preliminary filtering and centring of data

The raw CTF-corrected images of the individual macromolecules first need to be filtered and normalised (van Heel & Stöffler-Meilicke, 1985). We routinely band-pass filter the molecular images mainly to suppress disturbing low spatial frequencies. The presence of low frequency

components in the images (often associated with density ramps, etc.) may seriously disturb alignment procedures using cross correlation functions. The very high spatial frequencies may also be suppressed during certain phases of the analysis (for example, during alignments) but these spatial frequencies, although very noisy, also contain the finest details one hopes to retrieve from the data. For the first iterations of 3D structure determination, it may be necessary to suppress the high frequencies in order to find a first low-resolution preliminary 3D model for the later refinement procedures. During these refinements, the original high-frequency information in the data may be reintroduced.

The band-pass filter we normally apply is the product of a wide Gaussian low-pass filter to cut away the high spatial frequencies, and a narrow Gaussian high-pass filter to cut away the disturbing low frequencies. One may choose to leave in a fraction of the very low frequency components in the data (say, 0·001) in order to, at a later stage, be able to restore these components in the images. (In practice, however, these very low frequency components are never of any relevance to the interpretation of the final maps.) After filtering, the unwanted background is removed by imposing a (soft-edged) circular mask to the filtered images. the data within the circular mask is then normalised to a zero average density and an arbitrary variance value of 100 (van Heel & Stöffler-Meilicke, 1985). These pre-treated images all reside in a single IMAGIC file (van Heel *et al.* 1996) and are used as the starting point for all subsequent alignments.

4.2 Alignments using correlation functions

Raw electron microscopical images of the individual particles are extremely noisy and difficult to appreciate visually; one can often barely see the molecules in the raw micrographs. It is virtually impossible to judge the quality of the molecular images without first applying averaging procedures to improve the signal-to-noise ratio (SNR). To average the raw images – or to compare them by any computational means – it is necessary to first align the molecular images with respect to each other so as to eliminate the 'in-plane' translational and rotational degrees of freedom among the images. Aligning a set of images with respect to a reference image is best performed using correlation function-based alignment procedures (Saxton & Frank 1977; Steinkilberg & Schramm, 1980; Frank *et al.* 1981a; van Heel *et al.* 1992a). Conventional cross correlation functions ('CCF') are 'squared' correlation functions which are biased towards the predominant (typically the low) frequency components in the data (van Heel *et al.* 1992a). Instead of the conventional CCF we often rather use the mutual correlation function (MCF) for alignments, in particular in later phases of the processing. In the later phases of the processing, one wants to focus in on the weak high-resolution data components, and boost these to statistically significant levels.

4.3 Choice of first reference images

Aligning a set of images with respect to a reference image using correlation procedures tends to bias the data set towards the properties of the reference image used (Boekema *et al.* 1986). Thus, 'reference-free' alignment procedures (Schatz & van Heel, 1990) are preferred for obtaining a first inventory of the various typical molecular views present in the data. With

our preferred reference-free alignment by classification (Dube *et al*. 1993), the molecular images are centred by translational alignment relative to the rotationally averaged total sum of all images in the set (iterated two or three times). Because only a translational alignment relative to the overall averaged 'blob' is performed, the rotational orientation of all molecules remains arbitrary. Multivariate statistical classification procedures (see below) are then used to find similar images, in similar rotational orientations. The resulting class averages are then compared visually or computationally, to find similar images in different in-plane rotational orientations. These systematically present views are good unbiased reference images with which to start the multi-reference alignment procedures (van Heel & Stöffler-Meilicke, 1985).

An additional advantage of this procedure is that the eigenimages (see below) of the centred data set allow one to objectively find the symmetry properties of the constituent molecules (Dube *et al*. 1993; van Heel *et al*. 1996). Earlier symmetry analysis approaches were based on finding the predominant rotational symmetry components of one single image at a time (Crowther & Amos, 1971). When first aligning a whole set of images with respect to one single image with a strong, say 12-fold symmetry component, the 12-fold symmetry property will be imposed upon the data set by the reference-bias effect (Boekema *et al*. 1986). When the resulting average is then analysed by the rotational analysis, the 12-fold symmetry becomes overwhelming. There are ample examples of such self-fulfilling prophecy processing in the literature. After the first 'reference-free' alignment procedures (Schatz & van Heel, 1990) were proposed, other proposals in this direction followed. Unfortunately, the alignment procedure (Penczek *et al*. 1992), named 'reference free', which has been quoted many times in the literature, actually uses a (changing) global average as reference, and thus does not belong in this category of methods.

4.4 Multi-reference alignment of data

For the angular reconstitution approach we need a data set of randomly oriented single particles in which 'all' different projections of a structure are mixed. For such data sets we also need many different reference images in order to perform a good alignment of the molecules. Correlation-function based alignment procedures require the image to be aligned to resemble the reference image: they need to 'correlate'. Thus a (large) number of reference images are required to align the mixed population of noisy molecular projection images. Multi-reference alignment (MRA) procedures (van Heel & Stöffler-Meilicke, 1985) have long since been in routine use for achieving the best overall alignment of a data set, and the associated algorithms have passed through a few thorough re-organisations. Multi-reference alignment procedures are used in various phases of the iterative 3D analysis, not only in the preparative phase we have just described. (MRA procedures have even been used for quasi-automatic particle selection (Lata *et al*. 1995).) Most of the computing resources needed for the single-particle cryo-EM are invested in MRA procedures, and we will therefore go into some detail in describing the algorithm.

The (tens-of-thousands of) individual molecular images are first all placed one after the other in a single huge file (Fig. 7). After their correlation alignment with respect to reference no. 1, each aligned image will have an associated correlation coefficient (or other similarity measure (van Heel *et al*. 1992a)). The data set is subsequently aligned with respect to reference no. 2. Again, an aligned data set results with associated correlation coefficients. Using both
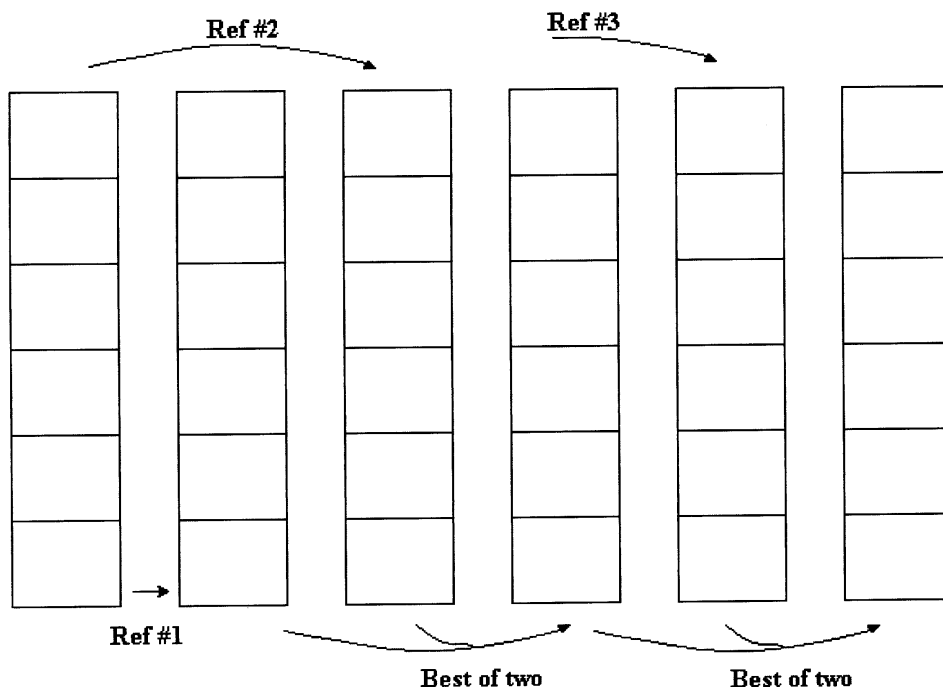
Fig. 7. Multi-reference alignment (MRA) procedure. MRA procedures allow large sets of molecular images to be optimally aligned with respect to a large set of reference images. Correlation alignments are performed with respect to many reference images whereby the aligned image with the best alignment quality – typically the correlation coefficient – survives in the final output data set. For details see main text.

the data sets aligned with respect to reference nos 1 and 2 respectively; one then creates a merged data set containing the aligned images associated with the highest correlation coefficient ('best of two'). The procedure is then repeated with reference image no. 3, etc. Originally, the MRA procedures in IMAGIC were indeed organised exactly as explained above. However, in its practical implementation this sequence of events has the disadvantage that all molecular images are read from and written to the hard disk once for every reference image to which it is aligned. This procedure thus produces a large number of input/output operations, causing the computer system to often wait for the data to become available, thus seriously hampering the efficient use of the available computational resources.

Since the first MRA implementations, computers have been through dramatic improvements. It is now possible to hold hundreds of reference images in the central memory ('RAM') of a standard computer allowing for a far more efficient organisation of the MRA algorithm. In its current implementation, the memory of the computer (or computers, see below) is filled using as many reference images as possible. Once one image is read from the input file, that image is aligned with respect to all reference images that are held 'in core'. The aligned version of the input image is thus only written to disk again after a massive number of alignments have been performed. As a consequence of removing this 'I/O bottleneck', the current IMAGIC 'MRALIGN' program may use more than 98 % of the CPU resources of a machine, even when the large image files are not stored locally on the machine, and the fast local I/O channels can thus not be used. We return to the implementation issue
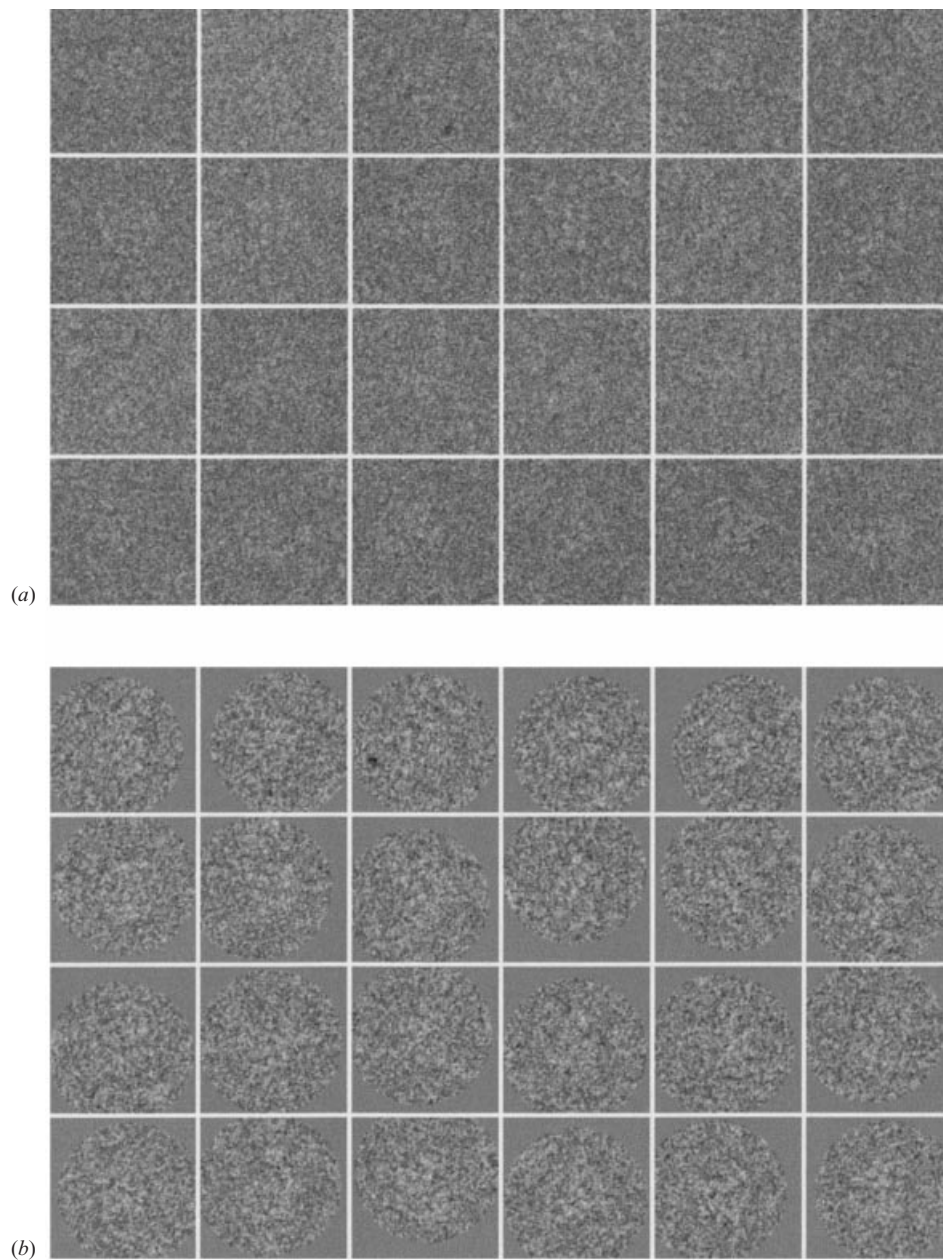
Fig. 8. Molecular images before and after alignment. (*a*) A small sub-set of a molecular images data of the *E. coli* large ribosomal subunit ('50S'). These 256×256 pixel images were scanned at 1·6 Å/pixel. (*b*) After alignment, the shifts applied to the very noisy images are most obvious from the shift of the circular masks used to suppress the background.

when discussing software parallelisation below. As an example, we show a number of ribosomal images after a final alignment with respect to a large number of reference images (Fig. 8). The aligned images are inspected visually to check whether the procedures have been completed successfully.

4.5 MSA eigenvector/eigenvalue data compression

Multivariate statistical (MSA) data compression, introduced in EM in 1980 (van Heel & Frank, 1981) is now an integral part of many EM procedures. With the MSA techniques one considers images as a linear combination of the main eigenvectors ('eigenimages') of the set, thus reducing the total amount of data and facilitating its interpretation. Although the eigenvector analysis was originally performed using the χ^2 as metric of 'correspondence analysis', we now prefer to use the modulation metric (Borland & van Heel, 1990). This metric can deal with phase-contrast EM images, which may have a zero average density, a situation that cannot be dealt with in strict correspondence analysis. Strict correspondence analysis uses a χ^2-metric, designed for positive data only (histogram data).

The full mathematics of the MSA procedures has been described in Borland & van Heel (1990). The IMAGIC MSA program (written by MvH in the mid-1980s) is optimised for efficiently finding the predominant eigenvectors/eigenvalues of extremely large sets of images. The basic principle of the MSA algorithm is the old and relatively simple 'power' procedure (Golub & van Loan, 1996). Here, we give a simplified version of the underlying mathematics, following the nomenclature of Borland & van Heel (1990). Excluded from the overall mathematics is the 'metric', the sole distinction between the 'correspondence analysis', 'principal component analysis' and 'modulation analysis' (Borland & van Heel, 1990).

A randomly chosen vector $\mathbf{r}_1$, when multiplied through a symmetric variance co-variance matrix $\mathbf{A}$, will yield a new vector $\mathbf{r}_1'$:

$$\mathbf{A} \cdot \mathbf{r}_1 = \mathbf{r}_1'. \tag{2}$$

If this resulting vector is then (after normalisation) successively multiplied through the matrix $\mathbf{A}$ again, the resulting vectors will gradually converge towards the first (largest) eigenvector $\mathbf{e}_1$ of the system, for which, per definition, the following equation holds:

$$\lambda_1 \mathbf{e}_1 = \mathbf{A} \cdot \mathbf{e}_1. \tag{3}$$

The variance co-variance matrix $\mathbf{A}$ is calculated as the matrix multiplication of the data matrix $\mathbf{U}$ and its transposed $\mathbf{U}'$:

$$\mathbf{A} = \mathbf{U}' \cdot \mathbf{U}. \tag{4}$$

The data matrix $\mathbf{U}$ contains, as its first row, all of the pixels of image #1; its general ith row contains all the pixels of image #i. The MSA algorithm operates by multiplying a set of randomly generated vectors (because of the nature of the data also called 'eigenimages') $\mathbf{r}_1$, $\mathbf{r}_2$, etc., through the data matrix $\mathbf{U}$ and its transposed $\mathbf{U}'$ respectively. The variance-covariance matrix $\mathbf{A}$ is thus never calculated explicitly since that operation is already too expensive in terms of its computational burden. The working of the MSA algorithm (Fig. 9) is as follows: a set of randomly generated images is orthonormalised and used as the first approximations to the desired 'eigenimages'. The typical number of eigenimages used is 24–69 and they are symbolised by a set of two in the illustration (top of Fig. 9). Then, the inner product between these images and the N images of the input data set is calculated. This calculation leads to coefficient vectors of length N as indicated in the left-hand side of Fig. 9. The next step is then to calculate weighted sums over the input image stack, using the different coefficient vectors as the weights for the summing. A new set of eigenimage approximations is thus generated as shown in the lower part of Fig. 9. New approximations
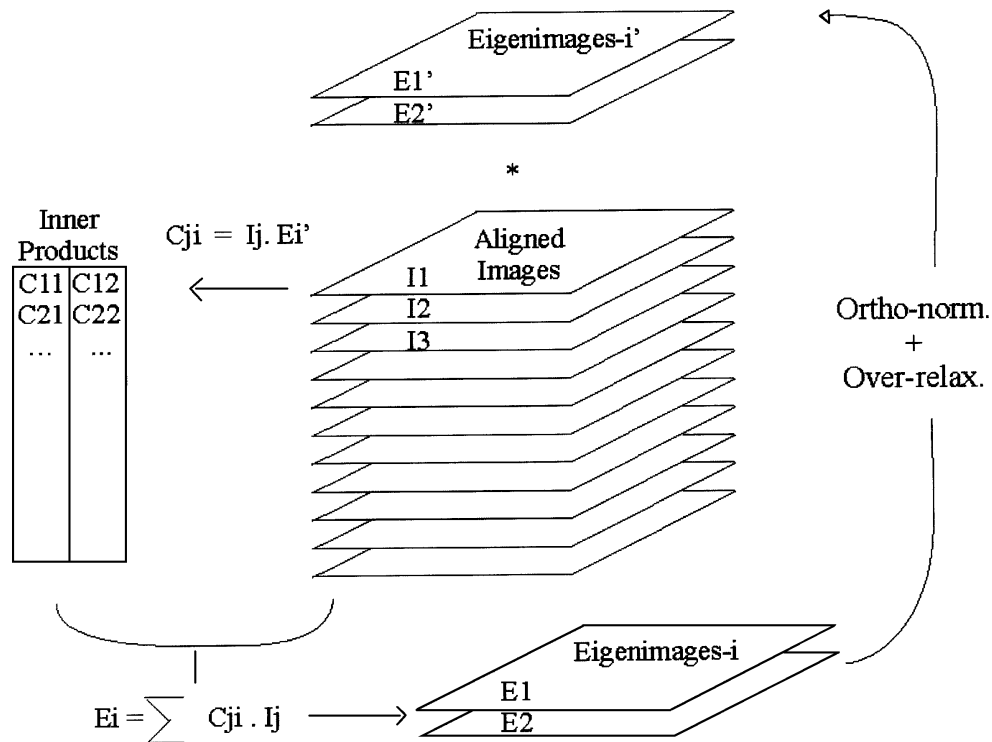
Fig. 9. MSA eigenvector eigenvalue algorithm. The IMAGIC MSA program is optimised for finding the most significant 'eigenimages'/eigenvaues of extremely large image data sets. The basic principle of the MSA algorithm is the old 'power' procedure: a randomly chosen vector, when iteratively multiplied through the data, as indicated, will yield gradually approximate the first 'eigenimage' of the system. A whole set of eigenimages can be found by making sure, after each iteration of the procedure, that each eigenimage remains orthogonal to all preceding eigenimages. This process is indicated by 'orthonormalisation' in this algorithmic diagram. Over-relaxation is a classical trick in numerical mathematics to make an algorithm converge faster: by exaggerating the changes occurring between successive iteration of the procedure, one (successfully) guesses in which direction the final solution will lie.

are generated from this set by orthonormalisation and over-relaxation with respect to the previous set. The algorithm rapidly converges (within 10–20 interations) to the most important eigenimages of the data set.

An important property of this algorithm is its efficiency for large numbers of images 'N': its computational requirements scale proportionally to '$N.P$', assuming the number of active pixels in each image to be 'P'. Many eigenvector–eigenvalue algorithms require the variance-covariance matrix as input. The calculation of the variance–covariance matrix, however, is itself a computationally expensive algorithm requiring computational resources almost proportional to 'N^3'. (This number is actually: $\mathrm{Min}(N^2P, NP^2)$.) The MSA program produces both the eigenimages and the associated eigenpixel-vectors in the conjugate data space as described in Borland & van Heel (1990). One of the intuitive charms of this fast disk-based eigenvector–eigenvalue algorithm is that it literally sifts through the image information while finding the main eigenimages of the data set. The programs have been used on data sets consisting of up to 100000 individual images. An example of the predominant eigenimages of a set of aligned ribosomal images is shown in Fig. 10(*a*).

(a)

(b)
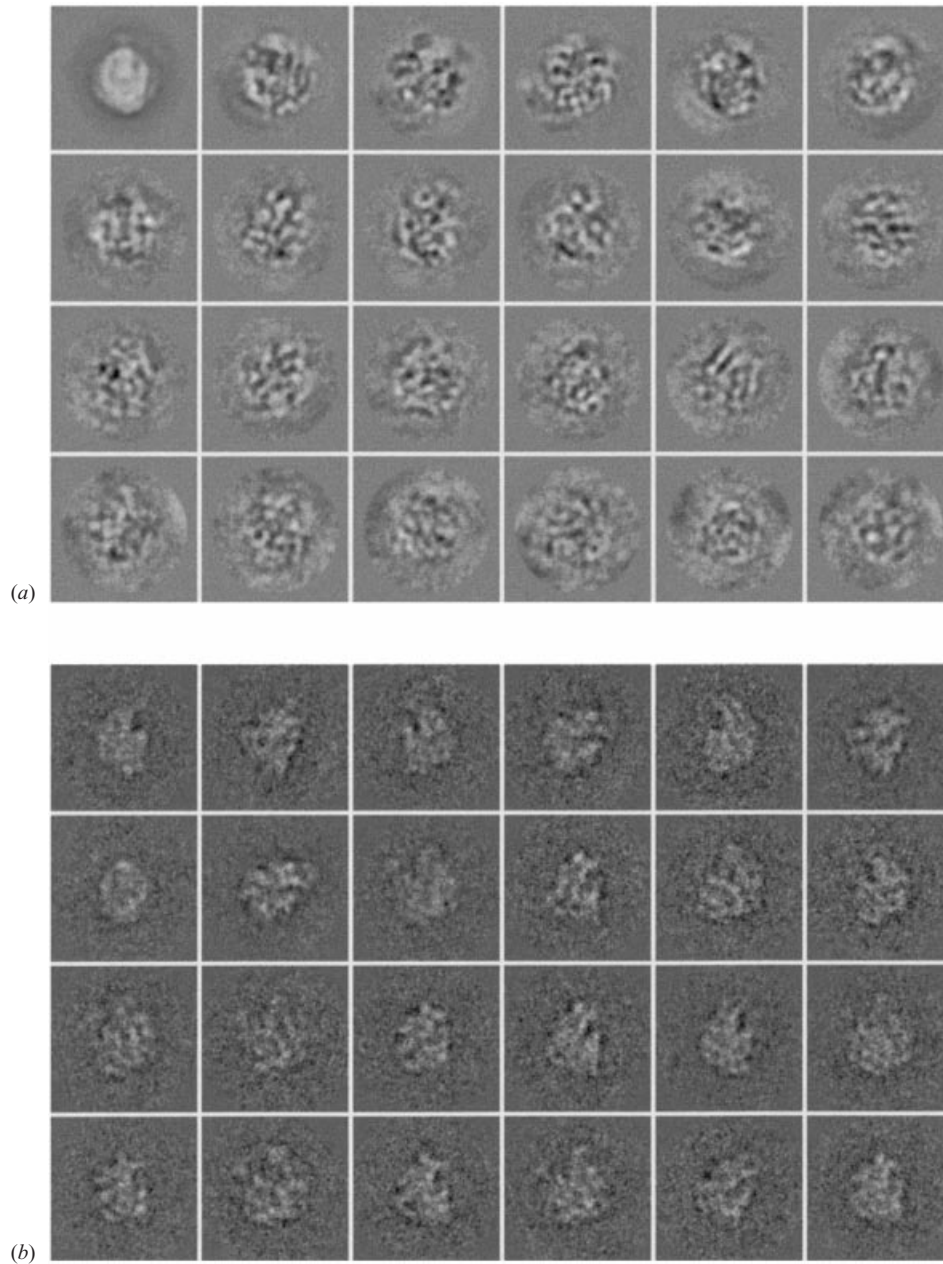
Fig. 10. Eigenimages and class average. (*a*) The 24 predominant 'eigenimages' of the aligned data set of *E. coli* large ribosomal subunits (Fig. 8) as determined using the IMAGIC MSA algorithm. (*b*) Some typical class averages. Each of these averages contains some 10–20 original images (Fig. 8(*a*)).

4.6 MSA classification

After the MSA data compression, an automatic hierarchical ascendant classification in combination with a moving elements post-processor is performed. This process has been described in a number of publications (van Heel, 1984a, 1989; Borland & van Heel, 1990). The algorithm is aimed at minimising the internal variance of the classes ('intra-class
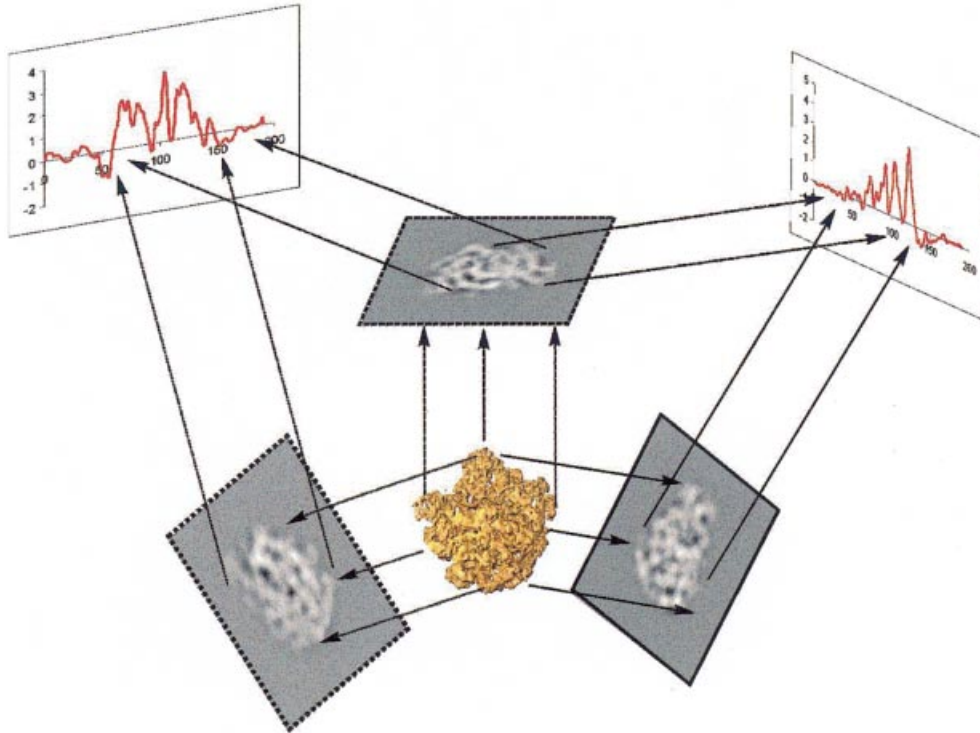
Fig. 11. The angular reconstitution technique is based on the common line projection theorem stating that two different two-dimensional (2D) projections of the same 3D object always have a one-dimensional (1D) line projection in common. From the angles between such common line projections, the relative Euler-angle orientations of set projections can be determined *a posteriori* (van Heel, 1987). For an entirely asymmetric particle like this 50S ribosomal subunit, at least three different projections are required to solve the orientation problem. For details see main text.

variance') while at the same time maximising the inter-class variance between the centres of mass of the class averages. Originally each image is a 'class' by itself and the classes are then merged two at a time, until at the end of the 'hierarchical ascendant classification' (HAC), all images become a member of one huge class containing all images in the data set. At any level of the classification process, the program will try to merge two classes 'i' and 'i'' together, which – at that level of the procedure – will be associated with the minimal added variance. The added intra-class variance criterion is known as the 'Ward' criterion (Ward, 1982; van Heel, 1984a, 1989).

$$\text{Add. Var.} = \frac{(w_i \cdot w_{i'})^2}{(w_i + w_{i'})} \cdot D_{ii'} \tag{5}$$

In this formula $d_{ii'}$ is the Euclidean distance (within the compact factor space resulting from the MSA data compression discussed above) between the classes i and i' having masses (weights) w_i and $w_{i'}$ respectively. To obtain a predefined number of classes from the process, one then cuts the 'tree' of the history of class mergers at the appropriate level. The partition obtained with HAC is post-processed to reach a local minimum of intra-class variance by

allowing members of all classes to migrate to any class where they are happier in terms of the variance criterion (van Heel, 1984a, 1989). This post processor significantly improves the quality of the partitions obtained. A number of quality criteria are included in the classification procedures that can be used to eliminate poor images and classes from the processing. Other classification schemes have been proposed but the HAC scheme has emerged as the most robust (for a review see : Frank (1990)). All images that have been assigned to the same class in the classification phase are averaged together; this averaging of images leads to a large improvement in SNR (Fig. 10(*b*)). The SNR improves proportionally to the number of images averaged, under the idealised assumption that all images averaged are identical, apart from the noise, which is different from image to image. The new class averages may be used as references for a new MRA/MSA classification iteration round. After a few interations, good class averages with improved signal-to-noise ratios can be obtained. The high SNR values obtained are of great importance for an accurate assignment of Euler angles to these projection images.
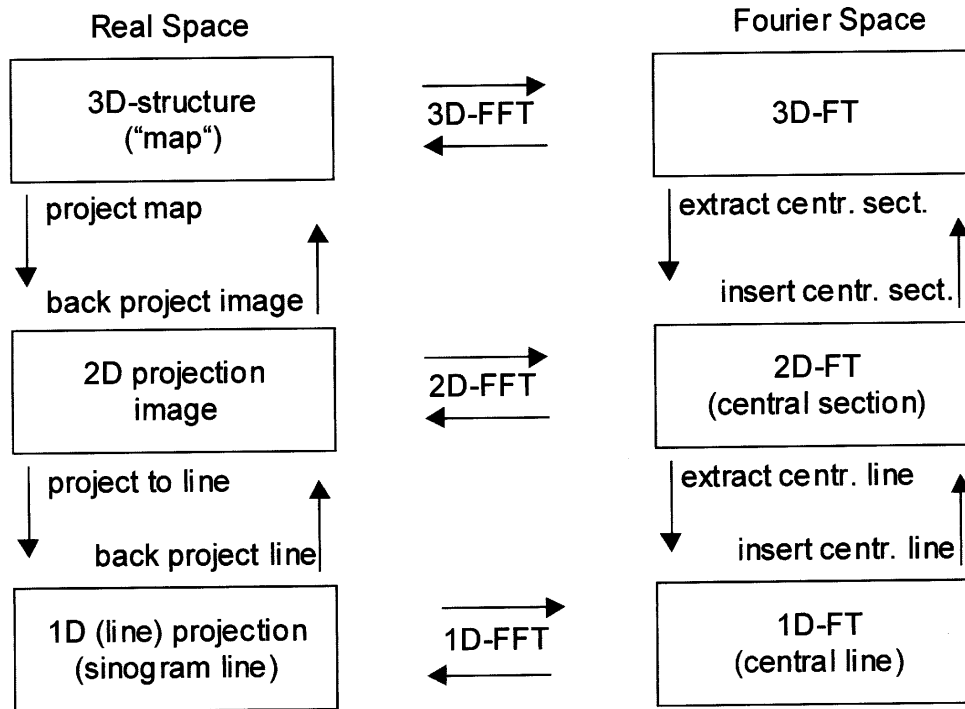
4.7 Euler angle determination ('angular reconstitution')

The class averages resulting from the MSA classification procedures are 2D projections of a 3D structure in different projection directions. After having found a sufficiently large set of good 2D projection images, we can reconstruct the 3D structure to high resolution if we know the orientational relationships between all projection images. The angular reconstitution technique allows one to *a posteriori* find the relative orientations of the 2D projections of a 3D structure (van Heel, 1987; Goncharov & Gelfand, 1988; Farrow & Ottensmeyer, 1992; Radermacher, 1994). The technique is based on the common line projection theorem stating that two different 2D projections of the same 3D object always have a one-dimensional (1D) line projection in common (Fig. 11) (van Heel, 1987). From the angles between such common line projections, the relative Euler-angle orientations of set projections can be determined. For an entirely asymmetric particle, at least three different projections are required to solve the orientation problem (van Heel, 1987; Goncharov & Gelfand, 1988). The common line projection theorem is the real-space equivalent (Fig. 12) of the Fourier-space common lines theorem (DeRosier & Klug, 1968). The Fourier-space approach has been formulated into computer programs aimed at solving the structures of icosahedral particles (Crowther, 1971).

  Similarities between line projections generated from 2D projection images are the central issue in the angular reconstitution approach. The first step when a 2D projection image is to be given an Euler-angle orientation (van Heel, 1987) is to find all line projections of that 2D projection image. The collection of all line projections of a 2D image is called a 'sinogram' (Fig. 13) following a traditional nomenclature from the field of medical computer tomography ('CT'). Conceptually, the sinogram lines can be calculated by summing all the image lines in the 2D projection image, after first rotation that image over 0°, 1°, 2°, etc. The analytical version of the sinogram concept is called the 'Radon transform' (Radon, 1917). The sinogram is not invariant to a shift of the projection image within its frame.

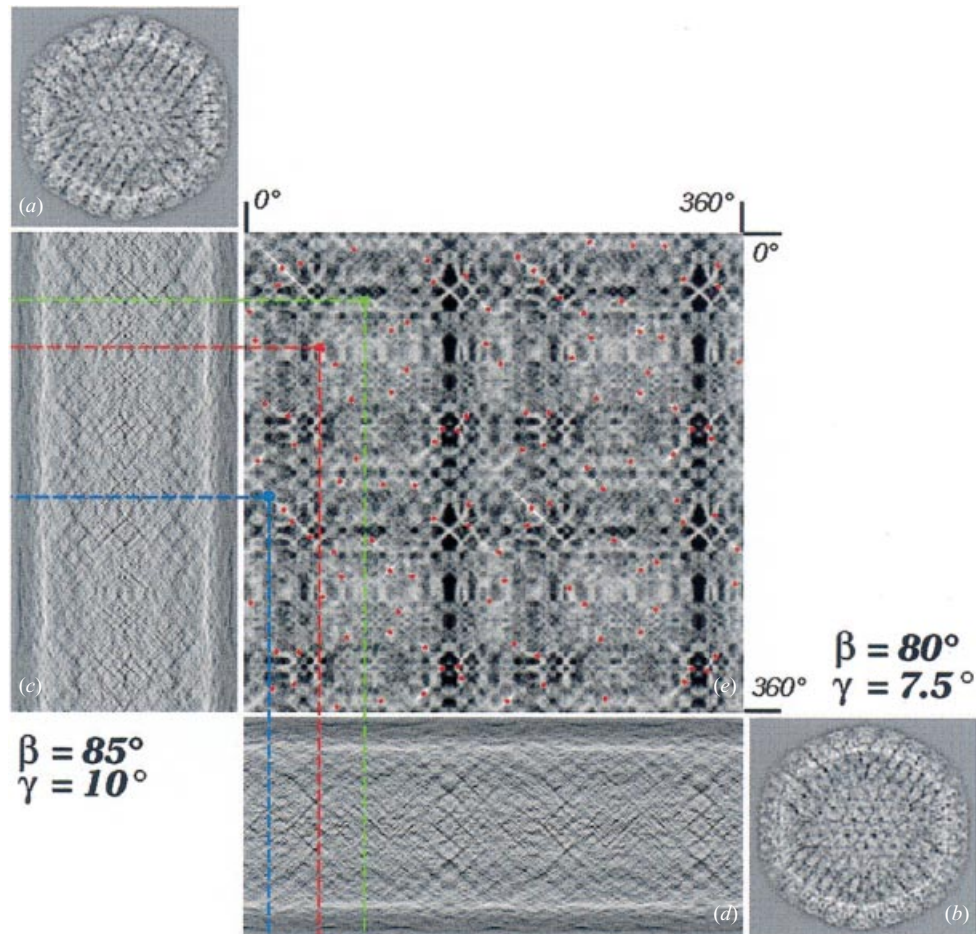## 4.8 Sinograms and sinogram correlation functions

To find the common line projection(s) between two 2D projection images, their sinograms are compared line-by-line 'sinogram correlation functions' (SinCF; van Heel (1987)). At a

Real Space   **Fourier Space**

| 3D-structure ("map") | ⟷ 3D-FFT ⟶ / ⟵ | 3D-FT |

↓ project map   ↑ back project image   ↓ extract centr. sect.   ↑ insert centr. sect.

| 2D projection image | ⟷ 2D-FFT ⟶ / ⟵ | 2D-FT (central section) |

↓ project to line   ↑ back project line   ↓ extract centr. line   ↑ insert centr. line

| 1D (line) projection (sinogram line) | ⟷ 1D-FFT ⟶ / ⟵ | 1D-FT (central line) |

Fig. 12. The equivalence between the real-space common line projection and the Fourier-space common central line concepts is illustrated here. A three-3D structure ('map') can reversibly be Fourier transformed to its 3D Fourier transform (3D-FT) and both representations contain exactly the same information. A 2D projection of the 3D map is an image like the projection images generated by the transmission electron microscope. A set of 2D projection images can be used to re-create ('reconstruct') the 3D structure by various algorithms. These 3D reconstruction schemes can be 'real-space' filtered back-projection algorithms or 'Fourier-space' reconstruction algorithms based on filling in 'central sections' in the 3D FT of the object to be reconstructed. After the data is interpolated into the 3D-FT, the 3D map can be calculated by an inverse 3D Fourier transform. The whole idea can also be applied one dimension lower: projecting a 2D image onto a single line in a direction along the plane of the image produces one line projection. The 1D Fourier transform of this line projection corresponds to a central line in the 2D-FT of the 2D image. Note that the collection of all possible line projections forms a 'sinogram' (van Heel, 1987). Whereas the real-space and Fourier-space algorithms are equivalent in principle, implementation details in both conjugate spaces can differ considerably. In general the real-space algorithms are more transparent in their usage. The 'real-space' nomenclature may be somewhat misleading in this context since many real-space algorithms like the exact-filter back-projection algorithm (Harauz & van Heel, 1986a, b) may include more Fourier transforms than actual Fourier-space algorithms.

position corresponding to a pair of shared line projections, the sinogram correlation function has a maximum. If the molecules exhibit a specific point-group symmetry, the sinogram correlation function (SinCF) shows a number of corresponding, symmetry-related peaks. Highly symmetric molecules, such as icosahedral viruses (Patwardhan *et al*. 2000c), are somewhat easier to process than asymmetric particles (Stark *et al*. 1995; Matadeen *et al*. 1999) as will be discussed below.

Some details of the line-by-line correlation of the sinograms deserve further attention. Correlation functions and correlation coefficients are 'squared' similarity measures that may be over-weighted by the predominant low frequency components in the data (van Heel *et al*.

Fig. 13. Sinograms and sinogram correlation functions. This illustration provides a graphical overview of the relations between a 2D class average (noise-reduced projection images), their 'sinograms', and the sinogram correlation function between two sinograms. The images shown here (*a*, *b*) are class averages deduced a large data set of Herpes Simplex Virus Type 1 (HSV1) cryo-EM images. Each line of the sinogram images (*c*, *d*) is generated from the 2D projection image by summing all 1D lines of the 2D images, from top to bottom, after rotation of the image over angles ranging from 0° to 360°. Equivalently, the lines of the sinograms are 1D projections of the 2D images in all possible directions ranging from 0° to 360°. Each point of the sinogram correlation function contains the correlation coefficient of two lines of the two sinograms one is comparing (*e*).

1992a). The SinCFs are thus preferably calculated only after an 'amplitude square root filtering' (ASF) of the sinogram lines (van Heel *et al*. 1992b). Moreover, the sinogram lines, after the ASF operation, are normalised to a standard variance value. This normalization is performed so that the correlation coefficient becomes a more reliable measure of similarity between line projections. The reasoning behind the ASF procedure in the SinCF calculations is virtually identical to the reasoning behind the use of the MCF rather than the CCF, discussed under 'alignments' above. All correlation coefficients between the first line of the first sinogram and all lines of the second sinogram are then used to fill the first line of the SinCF between the two sinograms. The second line of this two-dimensional function is filled

with the correlation coefficient values of the second line of the first sinogram with respect to all lines of the second sinogram, etc. An example of a SinCF is shown in Fig. 13.

To introduce a new projection into a set of projections that have already been assigned an Euler-angle orientation, a 'brute-force' search is performed over all Euler-angle orientations corresponding to the full 'asymmetric triangle' for the given point-group symmetry (Schatz *et al.* 1995; Serysheva *et al.* 1995). The search over all possible orientations leads to a complex system of predicted peak positions in all SinCFs of all earlier projections with respect to the newly introduced projection image. The normalised standard deviation of the peak heights among all corresponding peaks serves as an internal consistency check. Not only should the sum of all predicted SinCF peak positions be at a maximum for the best Euler-angle orientation, also all peaks should consistently be high, with only a small standard deviation from the average value. The normalised standard deviation is a sensitive quality criterion, which may be used to exclude poor projection images, i.e. to exclude projection images that do not match well to an otherwise consistent, high quality set of projection images. It was only after we implemented the brute-force algorithm that searches over all possible orientations of symmetric or asymmetric particles (van Heel *et al.* 1992b) that the algorithmic approaches stabilised and led to rapid convergence of the procedures.

For an asymmetric particle like the ribosome, the asymmetric triangle covers the entire unit sphere. For an icosahedral particle, however, the 'triangle' actually has the shape of a triangle and covers 1/60th of the surface of the unit sphere. For a 'D5' symmetric structure like that of the molluscan hemocyanins (Orlova *et al.* 1997), the asymmetric triangle reaches from the 'north pole' of the unit sphere down to the equator, where it spans $360/5 = 72°$ along the equator. Once a 3D reconstruction exists (see below) one moves away from original projection images to find the orientation of a new projection. 'Re-projections' are used to generate an 'anchor set' of projection images to be used for this purpose (Orlova & van Heel 1994; Schatz *et al.* 1995) (and see below). Originally, we believed our angular approach to be somewhat limited by preferred orientations of the molecules with respect to the support film (van Heel, 1987). Even in 1992 we still envisaged use of small tilts of the specimen holder in order to further randomise the orientation of the molecules in the specimen (van Heel *et al.* 1992b). However, once the brute-force search algorithms reached maturity (Schatz *et al.* 1995; Serysheva *et al.* 1995) preferred orientations turned out in practice to hardly represent a problem at all. The typical data set did contain a wide range of orientations that had simply not been recognised visually (van Heel *et al.* 1997; Schatz *et al.* 1997). The exact-filter 3D reconstruction algorithm discussed below absorbs most of the remaining effects of unequal distributions of projection orientations over the asymmetric triangle in the resulting 3D reconstruction.

4.9 Exploiting symmetry

The angular reconstitution approach requires at least three different projections of an asymmetric particle in order to find their relative orientations (van Heel, 1987; Goncharov & Gelfand, 1988). Assume we have three projections P_1, P_2, and P_3, and assume one has found all the applicable common line projection (CLPs) from their SinCFs. The latter implies we have found all the relative rotational orientations of the projections, that is: $< P_1 P_2 P_3$, $< P_2 P_3 P_1$, and $< P_3 P_1 P_2$; these angles being the angles of the triangle $\triangle P_2 P_2 P_3$ on the unit sphere. For simplicity, let us then start to constrain the position of this triangle by choosing

the direction of P_1 on the equator and at the zero meridian (after all, the centre of our co-ordinate system is London). In our standard (IMAGIC) co-ordinate system this corresponds to $\beta_1 = 90°$ (defining the equator), and $\gamma_1 = 0°$. What is missing in the Euler angle assignment of projection P_1 is then the third Euler angle (α_1) defining in the 'in-plane' rotation of projection P_1 with respect to our co-ordinate system. Whereas we can, in principle, choose α_1 of the fist projection arbitrarily, we can also choose it to be such that the P_2 projection also falls on the equator ($\beta_2 = 90°$). This choice places the CLP between projections P_1 and P_2, parallel to the 'z-axis' of our chosen co-ordinate system (Fig. 11). This choice thus fixes α_1, $\beta_1, \gamma_1, \alpha_2, \beta_2$, but not yet γ_2. What is also fixed is the shape of the spherical triangle $\triangle P_1 P_2 P_3$ by the fact that its three enclosed angles $< P_1 P_2 P_3$, $< P_2 P_3 P_1$, and $< P_3 P_1 P_2$, are known. Placing the spherical triangle $\triangle P_1 P_2 P_3$ onto the unit sphere with $P_1 P_2$ along the equator, and P_1 at the zero meridian, thus solves the problem of the relative orientations of P_1, P_2, and P_3. Once a stable set of three projections have been defined, new projections can be included as described above.

This ideal 'analytical' solution to the angular reconstitution problem (van Heel, 1987) for three projections of an asymmetrical particle is complicated to achieve in practice (van Heel *et al*. 1992b, 1997). Certainly when one is looking at a new structure for the first time, this may not be the simplest way of starting the analysis. For example, when one or more of the largest peaks in the SinCFs of the set of three projections is a spurious peak, the solution the program comes up with may not be the correct one. However difficult it may be to get things started up correctly at this level, the intrinsic stability of the 3D reconstruction problem makes that the procedures converge rapidly. Indeed, it was shown that, even when the starting angles are created by a random number generator, iterative refinements can lead to good final Euler-angle assignments (van Heel, 1984b). Once one starts the angular reconstitution iterative refinements discussed below, the convergence of the procedures becomes obvious between successive iterations of the 3D reconstructions. Thus, although special algorithms are available for this phase of the processing (Farrow & Ottensmeyer, 1992), the robustness of the angular reconstitution (AR) refinement procedures will normally cope with the problem.

The convergence of the early phases of the AR procedures is particularly rapid when dealing with highly symmetrical particles such as the D6-symmetrical annelid hemoglobins (Fig. 20) or icosahedral viruses (Fig. 18). With highly symmetrical particles, the minimum of three independent projections required for solving the Euler-angle orientational problem are already present in one single class average. Starting an icosahedral reconstruction based on a number of good class averages is thus very simple. A typical procedure is to find the Euler-angle orientation of all available class averages based exclusively on the internal information of each class average. The quality criteria available can then be used to find the best centred class average in which the icosahedral symmetry is preserved best. All other class averages can then be assigned Euler angles with respect to that first class average. After a number of other classes are given an Euler-angle assignment, a 3D reconstruction can be calculated (see below).

With highly symmetric structures, the redundancy in the data not only makes the programs converge rapidly; the quality criteria, measured over a large number of symmetry-related peaks, became very sensitive to departure from an ideal behaviour. Thus, poor class averages can readily be identified and subsequently excluded from further processing. From a historical perspective, it was only *after* the programs had been used extensively for studying symmetric

structures – we used the *Lumbricus terrestris* hemoglobin structure (Schatz *et al.* 1995) for methodology development – that we successfully revisited the case of the asymmetric molecule, the 70S ribosome of *E. coli* (Stark *et al.* 1995). The procedures are so stable when using icosahedral particles that the Euler-angle assignments often work on the raw electron microscopical images directly, that is, without using noise-reduced class averages. The AR approach then becomes equivalent to the established common lines approach.

4.10 Three-dimensional reconstruction

The problem of reconstructing a 3D object from a set of 2D projections was in principle solved analytically by Radon at the beginning of the twentieth century (Radon, 1917). The principles of 3D reconstruction were re-invented in the late 1960s and early 1970s, with the introduction of computerised tomography in medicine and 3D reconstruction techniques in electron microscopy. Techniques for extracting 3D information from electron micrographs were introduced mainly by the Cambridge group around Aaron Klug (DeRosier & Klug, 1968). Important in this work was the 'projection theorem'; a 2D projection of a 3D object corresponds to a central section of the 3D Fourier transform of the object (Fig. 12). These 3D reconstruction ideas were first applied to helical fibres (DeRosier & Moore, 1970), images of which correspond to a full tomographic tilt series of projections of the helix, and to icosahedral viruses (Crowther, 1971), exploiting their 60-fold symmetry.

The concept of the central section and its width play a fundamental role in all 3D reconstruction algorithms independent of whether they are implemented in real space or in Fourier space or independent of whether the protagonists of the various techniques are aware of this. The central section theorem was – as was the Radon transform – introduced as an analytical mathematical concept. Thus, the projection operation through the 3D object was assumed to extend from minus infinity to plus infinity, corresponding to an infinitely thin central section (a mathematical plane) through the 3D Fourier transform of the object. It is, however, important to realise that the 3D object one wants to study is of limited spatial extent. Assuming that the largest linear dimension of the 3D object is 'D' (for the *E. coli* ribosome $D \approx 250$ Å), the 2D projection operations need only to extend over a maximum distance of 'D'. As a consequence – through the well known 'reciprocity theorem' of the Fourier transform – the corresponding central section is no longer an infinitely thin plane but rather a 'slab' with a thickness of $\sim 1/D$ (van Heel & Harauz, 1986a, b; Harauz & van Heel, 1986b). Because central sections have a width, they overlap over more than just a single line in Fourier space (Fig. 14), and this overlap determines how one must go about correctly reconstructing a 3D object from its projections and what resolution levels can theoretically be achieved in the process.

As can be directly understood from Fig. 14, the closer to the origin in Fourier space one gets, the more different central sections overlap. Therefore, when one performs a 3D reconstruction by simply adding data to the corresponding central section (slab) in Fourier space for each available projection image, the low frequency components of the 3D volume will receive too much weight. Equivalently, in real space, simply back projecting all available 2D projections through the 3D reconstruction volume will lead to a 'blurred' reconstruction due to the overly strong emphasis of the low frequency components in the overlapping central sections. To avoid this blurring problem, high-pass filtering of the projections ('filtered back-projection') was introduced in medical CT a long time ago (cf. Harauz & van
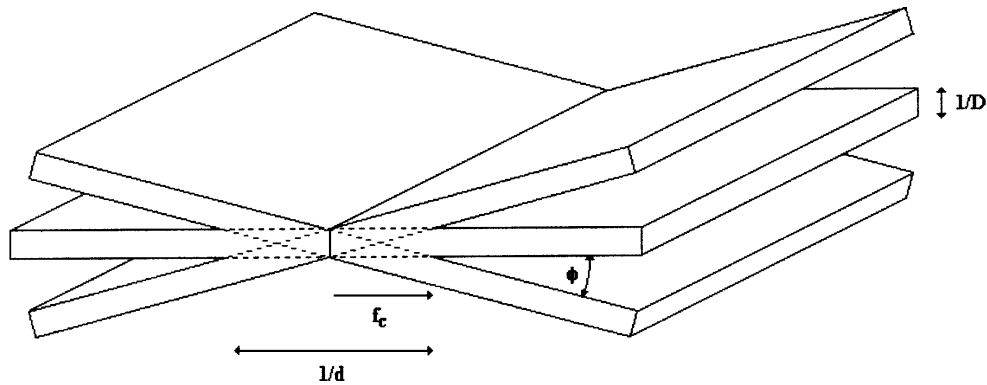
Fig. 14. Overlapping central sections. Fourier-space central sections, associated with 2D projections a 3D object of linear size 'D', have a width '$1/D$'. Central sections in Fourier space always overlap at very low frequencies, that is, close to the origin. Neighbouring central section, separated by an angle ϕ, cover largely the same information up to spatial frequency 'f_c'. The overlap of central sections is fundamental in both 3D reconstruction algorithms and in determining the highest isotropic resolution achievable for a given 3D reconstruction geometry. The areas of the 3D Fourier space volume not covered by the central section 'slabs' are not measured and are referred to as 'missing cone' or 'missing wedge' depending on the 3D reconstruction geometry.

Heel, 1986b). However, these analytical filters are only a rough approximation of the real filter function required, and they certainly do not cover the random orientations one finds in vitreous-ice embedded particles. Exact-filter algorithms were then introduced (Harauz & van Heel, 1986b; Radermacher, 1988) in which a specific filter is computed for each projection in the reconstruction problem at hand.

For a detailed comparison of the methods, and for objective comparisons of the quality of 3D reconstruction algorithms we refer to Harauz & van Heel (1986b). In spite of the serious warnings in that paper against the use of simple 'phantoms' and simple similarity criteria for judging the quality of 3D reconstruction algorithms, and of algorithms filling in missing 3D information by 'Maximum Entropy' or 'Projection Onto Convex Sets' (POCS), their misuse continues. We have yet to see the first publication in which a 3D reconstruction algorithm is actually shown – and not just claimed – to perform better than an exact-filter algorithm by the objective measures discussed. The same is true for 'data massage' programs claiming to produce better results (than obtained by just placing zeroes) in the areas of the 3D FT of the reconstruction volume that have not been measured (Fig. 14). Computational experiments for thoroughly testing such algorithms and the associated claims are easy to design but are – unfortunately – typically not part of the presentations and the claims are thus never substantiated.

The concept of the overlapping central sections allows us to answer a different question: what is the maximum resolution (spatial frequency) '$f_d = 2/d$' one can achieve for a given set of 'n' projection images of the 3D object of linear side 'D'? Formulated differently: what is the minimum number 'n' of different projections of the 3D object of linear side 'D' one needs to achieve a desired resolution of f_d? The answer, for the simple case where we have 'n' projections evenly spaced at $\phi = \pi/n$ angular intervals around a single rotation axis, is relatively simple. We choose to define the maximum resolution as that spatial frequency, where neighbouring central sections no longer overlap.

$$1/f_d \leqslant \pi D/n \cdot N_{\text{SYM}}. \tag{6}$$

This formula, first derived in 1968 (in a slightly different form (DeRosier & Klug, 1968)) for single axis tilt experiments, has an equivalent for randomly oriented projections, which differs only in the multiplicative constant (van Heel & Harauz, 1986). The N_{SYM} component is added to account for symmetry in the particles. This factor equals unity for an asymmetric particle and equals 60 for an icosahedral virus. Thus, to achieve an isotropic resolution of $f_c = 1/3$ Å on a ribosome of 250 Å diameter, we need a minimum of only ~ 200 randomly oriented projections! For an icosahedral virus of 300 Å diameter this minimum number is only 1 or 2! It must be noted, however, that this is a lower limit assuming that there is no noise in the data. In practice, even the class averages resulting from the MSA procedures described above are still quite noisy, in particular in the high-frequency end of the spectrum. To achieve statistical significance at high resolution, a much higher number of projections are typically required.

This resolution criterion is based exclusively on the completeness of the sampling in Fourier space. This criterion is associated with the 'instrumental transfer function' of the 3D reconstruction process (van Heel & Harauz, 1986) and is comparable to the theoretical resolution limit of an electron microscope associated with its CTF (Fig. 1(*b*)). The fact that the CTF extends to, say, 1·5 Å does not necessarily mean that one will actually achieve that resolution level in a practical experiment due to the noise levels in the images. Such resolution limits rather reflect what would be possible under ideal situations, in the absence of noise in the data. The resolution one can achieve in an actual 3D reconstruction experiment is normally measured by the Fourier shell correlation (FSC) and is limited by factors other than the transfer function of the 3D reconstruction operation (see below).

4.11 Euler angles using anchor sets

Once a first 3D reconstruction of a molecule is available, new possibilities present themselves. This first 3D map can be used to create an 'anchor set' by reprojection over different Euler-angle directions within the asymmetric triangle (Orlova & van Heel, 1994; Schatz *et al.* 1995; Serysheva *et al.* 1995). Re-projections from a 3D density are necessarily consistent with each other and will contain less noise than the original input images (the class averages) due to the implicit averaging taking place during the 3D reconstruction calculation. Instead of comparing each new input projection (class average) with every other projection available, the orientational search in the later phases of the analysis is performed only with respect to the anchor set projections. The anchor-set Euler-angle assignment is more sensitive and more precise than the Euler-angle assignment with respect to other classes because no inconsistencies exist within the anchor set. The quality criteria associated with the Euler angles obtained thus only reflect the inconsistencies between the input projection and the anchor-set projections. Again, poor projections can be rejected based on these quality criteria, and a new 3D reconstruction may be performed. This operation has an exact Fourier-space equivalent in the form of the cross-common-lines orientation determination (Crowther *et al.* 1994; Baker *et al.* 1999).

4.12 Iterative refinements

The reader may have noticed already that throughout the above text iterative refinements were mentioned. Iterative refinements are so interwoven at various levels of the angular
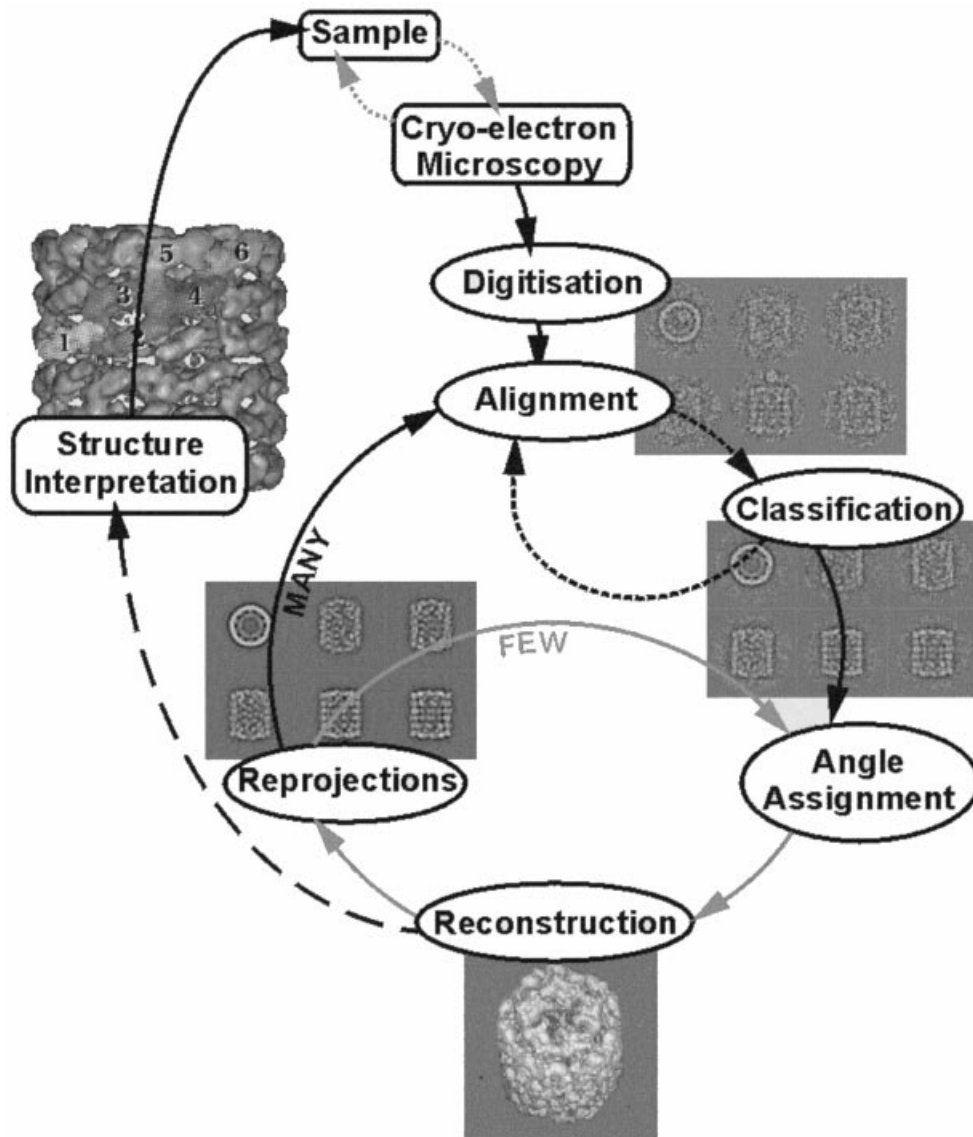
Fig. 15. Single particle cryo-EM by angular reconstitution. Using data from ongoing keyhole limpet hemocyanin (KLH1) research, we here present an overview of the various iterative refinements used in structure determination of single particles by the combination of cryo-EM and angular reconstitution. Starting at the top and proceeding in a clockwise fashion. (*a*) The quality of the pure sample is first assessed, and if necessary optimised, by EM observations. (*b*) Once good quality samples are obtained, micrographs are collected for 3D analysis using a high-resolution cryo-EM (200–300 kV, FEG); the quality of the data judged by optical diffraction. (*c*) The set of good micrographs is subsequently digitised using a high-resolution densitometer with sampling apertures of typically between 3 μm and 10 μm. (*d*) After a first CTF correction, the particles selected from the micrographs are aligned with respect to a large set of reference images ('MRA', see main text). (*e*) The class averages obtained by MSA classification serve as reference images for new rounds of MRA. (*f*) Class averages are assigned Euler angles by angular reconstitution *per se*, and these Euler angles allow a 3D reconstruction to be calculated. (*g*) Once a 3D map is obtained it can be re-projected into (a few) Euler-angle directions covering the asymmetric triangle to create an 'anchor set'. This set is used for refining the Euler-angle assignments of the class averages and a new, refined 3D reconstruction can thus be calculated. If

reconstitution procedures that it is difficult to consecutively explain each logical step of the algorithms without having already mentioned the various levels of iterative refinements. The anchor set Euler-angle assignment just discussed is performed as soon as a preliminary 3D reconstruction is available. That means that directly after one has performed a 3D reconstruction based on class averages, one creates re-projections to serve as an anchor set, which is subsequently used to refine the earlier Euler-angle assignments. With these refined Euler-angle assignments to the same set of class averages, one can directly calculate a new, refined 3D map. From this refined 3D one may then proceed to calculate a new anchor set, etc. This iterative refinement of the 3D reconstruction is only small detail (Fig. 15) in a large scheme, which itself is subject to iterative refinements.

Once the iterative refinement of the Euler-angle assignments with (a few) anchor sets and 3D reconstructions has stabilised, one creates (many) re-projections uniformly covering the asymmetric triangle. This large set of re-projections may contain as many as 500–1000 re-projections, which are then used as reference images for the alignment (MRA) of the full data set of molecular images. At this stage we can calculate what all (strongly noise-reduced) projection images of the object look like. Thus here we can create the full set of 2D projection images needed for a comprehensive alignment of all possible molecular views in the data. As was already mentioned above, this procedure is the most CPU intensive part of the computational procedures. We now turn our attention to the computer hardware aspects.

5. Computational hardware/software aspects

Computer hardware has undergone revolutionary developments during the past 20 years. In 1980 computers rated at ~ 20 kflops (kilo-floating-point-operation-per-second), equipped with 64 kbytes (kilo-bytes) of RAM (for historical reasons still called 'core' memory in those days), and 64 Mbyte hard disk would cost $\sim$ US\$100 000. For that amount of money, today, one can buy a 'farm' of loosely coupled computers, totalling some 20 CPUs delivering in excess of 20 gflop (1 giga-flop $= 10^6$ kflop). Such a system would be equipped with a total of ~ 10 gigabyte of RAM, and a total of 1 tbyte (1 tera-byte $= 10^6$ mega-byte) of disk space. Overall, we have seen an improvement of price/performance ratio of at least four–five orders of magnitude over the past 20 years. The typical life cycle of computer hardware over the past 20 years has been only about three–four years, and that short lifetime is also reflected in the lifetime of the computer manufacturers. Twenty years ago, one dealt with companies like Data General, Norsk Data, and Digital Equipment, names that have now largely disappeared from the market.

In contrast to the rather rapid turnover in the data processing hardware, the software for EM image analysis has a much longer lifetime. Software packages like MRC (Crowther *et al.* 1996); IMAGIC (van Heel & Keegstra, 1981; van Heel *et al.* 1996); SPIDER (Frank *et al.* 1981, 1996); and SEMPER (Saxton, 1996), made their first appearance more than 20 years ago. Most

---

necessary, this process can be further iterated until the results are stable. (*h*) The refined 3D is then used to create many re-projections covering the full asymmetric triangle, i.e. covering all possible unique Euler-angle orientation. This large set of re-projections is then used for a new round of MRA of the full original set of raw molecular images. (*i*) Once the resulting 3D reconstruction has stabilised, one starts the interpretation process. The interpretation is followed by....

of these software packages are still in active duty today. The software packages, of course, have undergone continuous developments in those 20 years but their basic design philosophies have remained the same. Software issues like the long-term maintenance and support for these packages are clearly more important issues than the pure hardware issues. Single-particle structure analysis by cryo-EM is a relatively complicated business, and some of the methodologies applied are still in full development. The learning curves for understanding the basic principles and specifics of any image processing system in EM are considerable; they represent a substantial manpower investment. Careful choices must be made when starting a project using this technology. Fortunately, the investments required in computer hardware and data analysis software are very modest when compared to the cost involved in setting up a high-resolution cryo-EM laboratory with advanced cryo-microscopes. At the same time, the researcher using single-particle approaches will typically spend at least an order of magnitude more time on the computer than on the electron microscope. Thus, although the computer hardware issues discussed here are obviously very important for achieving results quickly, the more fundamental issues for successfully pursuing single-particle projects are those of software transparency and ease of use, and software support.

5.1 The (IMAGIC) image processing workstation

Irrespective of the computing resources available in the background, the image processing user will normally be based at the standard desktop workstation. The workstation will be the first entry point to the overall system and all (graphical) user interactions will take place here. Typically, this computer will also house the massive data associated with the user's project(s). The single workstation, until just a year ago, would have typically been a UNIX workstation from one of the major workstation manufacturers (DEC, SGI, SUN, HP, etc.). During the past few years, however, the price/performance ration of 'PC' hardware has improved dramatically when compared that of the traditional workstation manufactures. Today, even the *absolute* performance of a PC may exceed that of the standard UNIX workstation. Thus, the entry point workstation today (in the year 2000) will be a PC with three or four 30–50 gbyte disks, running LINUX and/or Windows 98/NT/2000. Twenty years ago, the US$100 000 system mentioned above would also have been the smallest system on which one could possibly have operated the data analysis software. Today, the entry-level machine will be the same as any standard home computer or notebook computer, priced at less than US$1000. The physically smallest computer so far running IMAGIC that we have seen is a SONY VAIO measuring only $\sim 25 \times 20 \times 1\cdot 5$ cm. Although its keyboard is hardly large enough for typing, the machine fully out-performs the room-filling computers of 20 years ago. The miniature VAIO machine, in fact, had two IMAGIC versions running, one in its Windows 98 partition and the other in its LINUX partition, both residing on its 8 gbyte hard disk.

5.2 Operating systems and GUIs

The lifetime of the operating system platforms under which the image processing software operates is considerably longer than that of the hardware, but not necessarily as long as that

of the EM systems. The EM software systems need to be updated/adapted to the prevailing operating systems. Twenty years ago, the first interactive operating systems like VAX-VMS, SINTRAN-III, and UNIX, running on room-filling 'minicomputers' took over from the then prevailing batch-oriented operating systems running on main-frame computers. The user would operate one terminal (VDU: visual display unit) and directly talk to the computer. This technology represented a quantum-leap improvement in user interaction compared to the punched-card operations prevailing in university central computing facilities in the 1970s. The next such quantum leap was the introduction of windows-based graphical user interfaces, driven mainly by the X-11 windows standard developed by software teams at MIT in Boston in the 1980s. On one single graphical workstation, the user was now able to run many different virtual VDU-type terminals and view various image display systems at the same time, e.g. showing results of different phases of the analysis process. The IMAGIC system, which operates as many small programs that are started from a supervisor program, is particularly well suited for this type of graphical multi-process, multi-tasking environment. Many different IMAGIC supervisors may be started at the same time and the user can thus perform many image operations in parallel.

The next changes in operating system ('OS') were dictated by market forces more than by a change of concepts. The very stable VAX-VMS OS from Digital Equipment Corporation (DEC) – once the predominant OS for scientific computations – was gradually overwhelmed by the strong proliferation of UNIX-based operating systems from a broad range of manufacturers. What happened to the VMS OS now also appears to be happening to UNIX, with the unprecedented proliferation of Microsoft's Windows systems. The personal computer hardware had become so cheap and so easily available, that implementations on PC-based operating systems became necessary. First, a public domain version of the UNIX-like OS, LINUX was implemented to run on PC hardware (based on INTEL XX86 CPU chip series). With UNIX and X-11 available on the PC hardware platform, it became simple to convert the workstation versions of the EM software to the PC hardware environment. Still, many scientists find the learning curve for the very stable and universal UNIX OS too steep and prefer to work in their familiar MS Windows environment. Windows 98/NT implementations of EM software are now becoming more abundant. The next Apple Mac operating system will be Linux-based – according to the rumours – and we thus anticipate that the EM software will gradually be available on that platform too. The software support required for farm-based or shared-memory parallel processing (see below), is currently only available for UNIX-based systems (including Linux on PC hardware). Microsoft Windows 2000 (successor of the earlier Windows NT 4.0) also supports multiple CPUs per PC computer. This platform thus will also become available for shared-memory parallel computing in EM (see below).

The gradual change in the scientific computing community from 'punched-cards' batch operations, towards terminal-based, and then towards Windows-based, 'pull-down-menu' type of operations has been reflected in EM image processing. The EM-oriented data processing systems are all – at least originally – command line oriented programs. The user type commands in 'plain' English and the system then performs some operation on one (or many) images. Most systems allow the use of batch files, which was the only way to interact with the programs in the generation of 'batch orientated' EM software systems. Whereas batch operations are still the appropriate means of organising processes requiring massive computations with no user interaction, highly interactive operations (such as particle

selection) are best placed in a more interactive 'GUI' (Graphic User Interface) environment in a 'Windows' or X-11 environment. A number of examples of especialised interactive programs can be found in the special issue of the *Journal of Structural Biology* in EM software (Vol. **116**, 1996).

5.3 Computational logistics

Single-particle cryo-EM at high resolution is limited by the statistics of averaging, and thus by the mere size of data sets collected. Assuming that the high-resolution information is noise limited, the amount of information collected is directly proportional to the number of (good) images in the data set. The larger the data set and the larger the number of pixels in the images, the better. The logistics of the computing system and of the data storage devices are of central importance for achieving high-resolution results. With the new generation of densitometers discussed above, massive data sets may indeed be collected and subsequently stored onto high-density optical media such as CDs (650 mb) and DVDs (~ 5 gbyte). Processing of the data requires equally powerful – preferably parallel – computer systems. Fortunately, the hardware developments in the computer industry have been very rapid. Hard drives with a storage capacity of 36 or 72 gbyte disks have become over-the-counter items (priced at $\sim$ US$500, in the beginning of year 2000). The raw computational power of standard computers has also improved by many orders of magnitude. The standard computer used for running EM image processing software, like the MRC package (Crowther *et al*. 996), or our own IMAGIC system (van Heel *et al*. 1996), is a Compaq Alpha (running UNIX) workstation. High-end PCs, with an Intel Pentium III or an AMD Athlon chip (running Linux or Win98) may still be slower by a factor of 2 than the high-end Compaq (DEC) workstations, but these computers now match or exceed most standard UNIX workstations in speed.

Still, the computing resources required for aligning tens of thousands of noisy images with respect to hundreds of (re-projected) reference images are phenomenal. For close-to-atomic resolution levels, one needs to sample the images at 1 Å per pixel. If the object is, say, the 70S ribosome with a maximum diameter of 250 Å, the raw images may be as large as 512×512 pixels. Measured in 32-bit REALS, such an image is 1 mbyte in size. Aligning a data set of 100 000 such images (100 gbyte) with respect to 500 reference images remains a huge task that will keep even the most powerful single-CPU computer busy for weeks. Fortunately, most of the computational tasks in EM data processing are 'embarrassingly parallel'. Instead of aligning all 100 000 images on one single computer we can distribute the task over one hundred computers each aligning 1000 images. Such a parallelisation of the process leads to an almost 100-fold increase in turnover. Different types of parallel computer architectures exist and the different architectures are best suited for performing specific tasks.

5.4 Shared memory machines

The principle of the shared memory computer is that multiple CPUs (Central Processing Unit) operate on a large shared high-speed memory (Fig. 16). In this illustration, we have subdivided the shared memory into smaller blocks closer to each of the eight CPUs to indicate
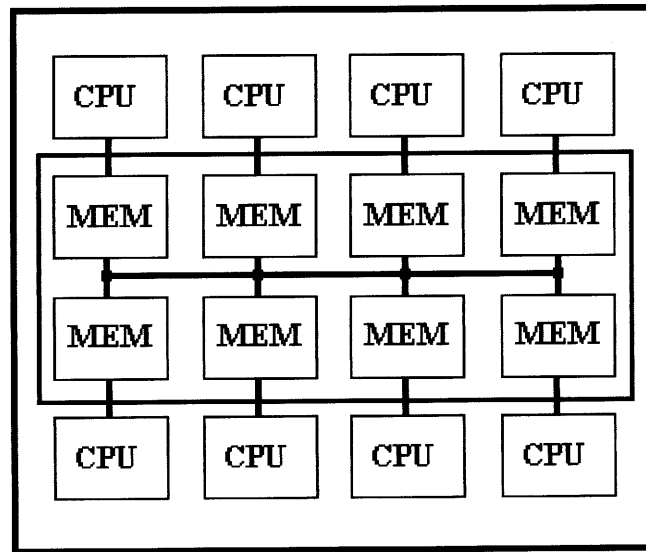
Fig. 16. The shared memory parallel computer. In a shared memory machine, many CPUs (eight here) have read/write access to a large memory bank which contains the shared data. However, the massive memory bank is typically subdivided into smaller, interconnected memory domains, and the speed of memory access depends on which CPU is trying to access which memory area. To (partially) circumvent such memory latency problems, expensive special-purpose hardware is required. The shared memory computer is thus relatively expensive per 'Giga-Flop', compared to a farm of workstations (Fig. 17).

that not all memory is necessarily close to all CPUs. At the current state of development, even single CPU machines are often limited by the speed with which they can read and write into the memory. To suppress this memory latency, virtually all computers are equipped with (expensive) high-speed memory ('cache'), which buffers the transfer to the normal memory and may dramatically reduce the average memory access time. When multiple CPUs access the same memory, the 'memory latency' problems are even more severe and need to be addressed with great care (and special hardware) by the computer manufacturers. Consequently, shared memory computers are expensive compared to single CPU machines.

The tasks that shared memory machines are good at performing are tasks that require large amounts of data to reside in the shared central memory and that simultaneously require a substantial amount of communication between the processes running on the various CPUs. One EM data-processing task that directly comes to mind here is 3D reconstruction. 3D reconstruction algorithms merge 2D projection images into a common 3D reconstruction volume (Harauz & van Heel, 1986b). Although the various 2D projections can be inserted into the reconstruction more or less independently by largely independent processes, these processes need to communicate with each other in order to avoid clashes while accessing the shared data. On the other hand, this algorithm may also be tailored for operation on a farm (see below) by having different members of the farm calculating different sections of the reconstruction volume. Refinements of the 3D reconstruction volume, including orientational refinements, may also be implemented on such shared memory parallel computes (Zhou *et al*. 1998). If the data need not be updated by all CPUs simultaneously but only by one process running on one of the CPUs from prolonged periods of time, the effectiveness of the shared
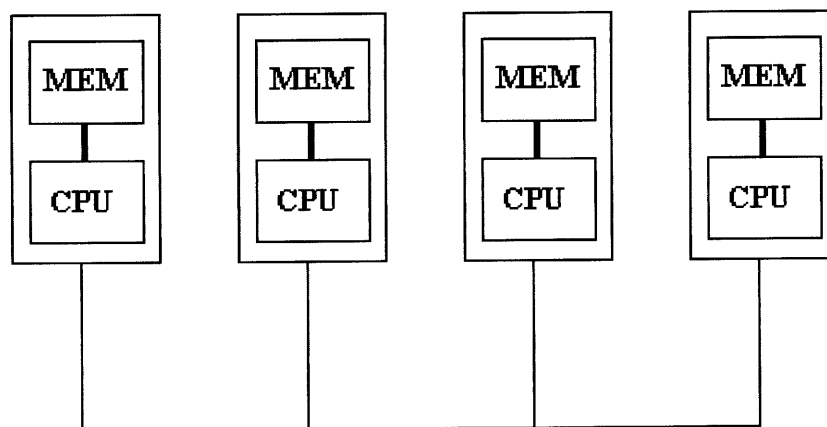
Fig. 17. A 'farm' of loosely coupled single workstation. Simple workstations coupled by Ethernet can form a very cost-effective parallel farm of computers. The most CPU-greedy procedures in single particle cryo-EM are optimally suited for this parallel hardware environment.

memory computer may actually increase because of the reduced memory latency. An example of such a task is the MRA algorithm described above. However, for such tasks, a 'farm' of individual computers is likely to be by far the most cost-effective solution.

5.5 Farming on loosely coupled computers

A collection of individual standard computers, each equipped with its own memory and disk, coupled through a standard net connection (typically Ethernet) may operate as a 'farm' on a single large-scale computing problem (Fig. 17). If the algorithm one seeks to parallelise is suitable for this type of large-grain parallelisation, the price/performance ratio of this solution in unsurpassable. The MRA algorithm described above is such an algorithm. It was explained above that a large number of reference images are stored in the central memory of one computer and the algorithm only reads and writes a raw molecular image once for the full alignment with respect to all reference images that fit into the memory. In the farming implementation of the MRALIGN program, the set of reference images are distributed to all computers working together on the problem. If the farm consists of 10 identical single-CPU computers, and the data set consists of 10 000 molecular images, the data set is distributed over the 10 computers, each aligning a different set of 1000 molecular images with respect to the same set of reference images. When this program was parallelised for farm operation in 1999 (it had been implemented on shared-memory machines first) the efficiency of the procedure was amazing. We had expected to have to distribute the input and output files over all machines in order to exploit the large parallel I/O bandwidth of all the individual computers. In reality, when the test job was distributed over five single-CPU 500 MHz alpha computers, the CPU usage of each of the computers went to over 99 %. This efficiency was due to earlier program reorganizations aimed at minimizing the necessary number of I/O operations.

When the MRA procedure is used to align a data set with respect to a large set of re-projections of a 3D reconstruction, the Euler-angle orientations of the re-projection which

best match the molecular image are written into its header. Thus, after the MRA, the aligned data set can directly serve as input for the 3D reconstruction algorithm. The projection-matching approach (a predecessor – discussed above – of the angular reconstitution approach) is thus automatically built into the alignment procedures. The better approach remains to continue the next round of MSA classification and the next anchor-set Euler-angle assignment/refinement. This angular reconstitution *per se* algorithm is also optimal for parallelisation on a farm (see below).

5.6 Implementation using MPI protocol

When programming for parallel machines, it is important that the same software runs on all platforms without any modification. The efforts that are invested in software development are very costly in terms of manpower. Maintaining different versions of a program for parallel and single-CPU machines is a software developer's nightmare. Also, from the user's perspective, the software system must remain fully transparent: a user working on a single CPU machine should never even notice he/she is running a parallel program in any other aspect than the increased speed of the calculations when running it on parallel computer hardware. In this context, it is important to note that both the shared memory machines and the farming computers may be programmed using the same 'MPI' protocol (Message Passing Interface). This protocol is a public-domain protocol (details can be found on various WWW pages including: http://www-unix.mcs.anl.gov/mpi/mpich/), which is available for all predominant (parallel) computer systems. Note that parallel computing may already start at the level of a 2CPU PC running Linux or Windows NT/2000 (shared-memory machine) or a network of two or more independent PCs running these operating systems (a simple 'farm').

5.7 Software is what it's all about

We have already entered the discussion of which programs are best suited for which parallel environments. As mentioned above, the MRA algorithm is by far the most CPU-intensive program required for single-particle cryo-EM and it is optimally suited for a farm environment. It will also run very effectively on a shared memory machine, but that parallel environment will not be very cost-effective. Another algorithm that is optimally suited for parallel processing on a farm is the Euler-angle assignment using anchor sets. The number of projection images in the anchor set will typically be 10–100. This small number of anchor-set images can be placed in each of the computers of the farm. Typically 100–1000 class averages need to be assigned an Euler-angle orientation in the process. These projection images can then be distributed equally over all available computers, which will each start assigning Euler angles to a subset of class averages. As is the case for the MRA algorithm discussed above, the Euler-angle assignment algorithm is optimally suited for a farm of loosely coupled computers; this problem too will scale almost proportionally to the number of available computers.

   Further algorithms that will eventually require parallelisation are the MSA eigenvector eigenvalue determination and the CLASSIFY automatic classification program operating on the

compressed data space generated by the MSA program. We have designed a parallel code for the MSA algorithm which requires little communication between the various processes and which should therefore be well suited for the farm parallel-computing environment. The CLASSIFY program is well suited for the shared memory computing environment. However, this program does not currently represent a major computing bottleneck so we have not yet seriously considered its parallelisation.

As was discussed above, we have already implemented the IMAGIC MRA program MRALIGN to operate on a farm: the MRA operations are associated with more than 80 % of the total amount of CPU-usage in our single-particle analysis procedures. The Euler-angle assignment program is currently being converted for operation on farms. These parallel programs also exploit shared memory environments or mixed environments in which series of, say, dual-CPU computers are organised into a farm. The farm, as was mentioned above, yields unparalleled value for money for these algorithms. The other algorithms mentioned will gradually be parallelised in the foreseeable future. In all cases, we must re-emphasise the importance of software, its maintenance and continuity. Almost more important than the fact that a given program runs very fast on any specific computing hardware is that exactly the same code will work both on single-CPU machines and on parallel environments. The code must be well organised and understandable by more than just the one individual who wrote the original program. The software 'machines' will be operational for decades and they must be amenable to further streamlining and optimisation within a team of software developers, not necessarily including the original author.

6. Interpretation of results

Interpretation of 3D reconstructions is the fun part of structural biology research by cryo-EM. It is also at this stage of the analysis that the quality of results must be assessed and the results of different research groups working on similar problems may be compared. One important quality criterion is the resolution achieved in a given 3D reconstruction and this resolution is represented by a single number. Although a single numerical value can hardly describe the richness of a detailed 3D map – important biological results are achieved at all resolution levels – resolution criteria have been the subjects of many controversies. Understanding the wealth of structural detail in a large 3D map may also represent a major effort. For example, the ribosome is so complex that even once one has an atomic-resolution map available it can still take many months to describe the detailed structural interactions revealed by the map. Once one understands the structural interactions, presenting the results in didactical form is yet another major time investment. Excellent graphical programs are available on all major computer platforms, which allow one to create good visual presentations of the results (see: *Journal of Structural Biology* in EM software (Vol. **116**, 1996) and (Vol. **125**, 1999)).

6.1 Assessing resolution: the Fourier Shell Correlation

Various techniques for assessing the resolution obtained in a reconstruction have been proposed in the past, some of which are based on an older X-ray crystallographic tradition. Because of the longer tradition and the (much) higher resolution one could achieve by X-ray

crystallography, electron microscopists tend to believe that the X-ray crystallographic resolution measures (cf. 'R-factor') are better than the resolution criteria used in EM. Yet, since in electron microscopy one has both the amplitude and the phases of the data available at all times, much more direct resolution criteria can actually be defined in cryo-EM (van Heel, 2000a, b). In contrast, in X-ray crystallography the primary experimental observables are the diffraction amplitudes (no phases are recorded). The R-factors then measure how well the calculated amplitudes and phases agree with the observed amplitudes. This advantage of the EM data processing has not always been sufficiently appreciated, and some of the first resolution criteria in EM were thus mere variants of earlier X-ray crystallography phase residuals.

An interesting harmonisation has taken place in recent years in the use of resolution criteria for EM. After a long controversy, the differential phase residual 'DPR' has been entirely superseded by the Fourier Shell Correlation criterion 'FSC' (Harauz & van Heel, 1986b; Beckmann *et al.* 1997; Böttcher *et al.* 1997; Conway *et al.* 1997; Orlova *et al.* 1997). The FSC is the 3D generalisation of the 2D FRC, independently introduced by Saxton & Baumeister (1982) and van Heel *et al.* (1982). These two papers directly followed an earlier paper on the DPR (Frank *et al.* 1981a) in which the concept of splitting the (2D) data set in two halves was introduced. Each of the two half data sets was then averaged and the two (2D) averages compared by the DPR approach. The FRC was introduced to correct for some obvious shortcomings in the definition of the DPR. It unfortunately took some 15 years for the 'DPR' school to (implicitly) admit its flaws by changing to the better FSC criterion. One of the criticisms (van Heel, 1987; Orlova *et al.* 1997) against the DPR has been its fixed 45°-threshold value. This fixed criterion can be too liberal (close to the origin in Fourier space, where too few pixels are involved to achieve statistical significance) or too stringent (far from the origin where statistical significance is achieved at much lower phase residual values).

Unfortunately, some authors have now started to use a 0·5 FSC value as the resolution-defining threshold (Beckmann *et al.* 1997; Böttcher *et al.* 1997). Again, a *fixed* threshold cannot account for the varying number of voxels in a Fourier 'shell' and this threshold must be considered a step back from the 2σ-threshold criterion, originally introduced by Saxton & Baumeister (1982) in connection with their definition of the FRC (actually called 'spatial frequency correlation function' by these authors). We have realized only a few years ago that the 3σ threshold (three standard deviations over the random noise value) that we use routinely, must be multiplied by the square root of the number of asymmetric units within the given pointgroup symmetry (Orlova *et al.* 1997). The reasons for this correction factor is that the number of independent voxels in the two Fourier-space 3D shells one compares is lower for symmetrical particles than it is for asymmetrical ones. Thus, the significance threshold for an icosahedral structure is $3\sigma \times 7{\cdot}75$ ($\sqrt{60} = 7{\cdot}75$), and for a D5 structure: $3\sigma \times 3{\cdot}16$ ($\sqrt{10} = 3{\cdot}16$). In other words, whereas the 0·5 threshold for icosahedral particles (Böttcher *et al.* 1997; Conway *et al.* 1997) may not be all too far above the symmetry-corrected 3σ threshold for asymmetrical particles like the ribosome becomes excessively stringent. Even so, an overly stringent resolution criterion need not necessarily impede the biological interpretation of the 3D maps. If one believes too much in the outcome of this resolution test and masochistically cuts away all high frequency information beyond this 'conservative' resolution estimate (Gabashvili *et al.* 2000), the fine image details will, of course, necessarily disappear from the data.

A number of problems can surface during the application of resolution criteria like the FSC

(or corrected phase residuals (van Heel, 1987)) and these problems have often interfered with the discussion about the quality of resolution criteria. These problems include the following.

(a) Masking the two 3D reconstructions with the same sharp-edged mask. This leads to similarities in the very high frequency components of both reconstruction volumes due to the sharp mask edges being the same in both volumes. Under these circumstances, the FSC curve may again cross the 3σ-threshold curve at unreasonably high resolution levels, and remain above the threshold level up to the Nyquist frequency (double the sampling step: for data sampled at 5 Å, the Nyquist frequency is 1/10 Å).

(b) Interpolation may also introduce similarities in the very high frequency domain. Because the same program (and interpolation routine) is used for calculating both 3D reconstructions, the high-resolution information in both reconstruction volumes may be correlated due to systematic rounding-off artefacts. The FSC curve, in general, may rise again close to the Nyquist frequency due to this effect alone.

(c) Under-sampling the data causes a host of problems. In under-sampled data, the systematic image information extends too close to the Nyquist frequency. Because of, for example, the joint effects of (A) and (B), and the presence of genuine high-frequency information, the FSC curve in under-sampled data may actually never even cross the threshold curve. It is thus important to have the sampling rate of the data so high that the image information stays sufficiently away from the Nyquist danger area.

In a recent paper we emphasised that any resolution claim should remain under approximately two-thirds of the Nyquist frequency (Orlova *et al.* 1997). Resolution claims significantly beyond this limit – a common mistake – should be avoided. We cannot resist the temptation to quote one typical example. In Malhotra *et al.* (1998) the data sampled at 4·8 Å and the '0·5' FSC threshold yielded a resolution of 14·9 Å. This claim is tenable while it is just above the 14·4 Å level (two-thirds of the Nyquist frequency). These authors then go on to criticise the 3σ FSC criterion, since that criterion would have yielded a 11·75 Å resolution value, which – as these authors reason – would have been too close to the Nyquist frequency. Although these authors are correct in not accepting resolution values too close to the Nyquist frequency, they are wrong for criticising the FSC criterion for that reason. The 3σ criterion applied correctly to their (under-sampled) data would have given them a resolution of $\sim 14\cdot4$ Å at best, that is, approximately two-thirds of the Nyquist frequency (Orlova *et al.* 1997). In very recent work (Patwardan *et al.* (2000c); and see below) the β-sheet structures in an icosahedral virus was completely visualised using almost exclusively the information beyond the '0·5 resolution cut-off' (Malhotra *et al.* 1998). When, however, the data are low-pass filtered such that the information beyond the '0·5 resolution cut-off' is removed, all β-sheet information disappears from the graphical renderings. This simple experiment, by *ad absurdum* reasoning, refutes the argumentation in Malhotra *et al.* (1998) and in the appendix to that paper (Penczek, 1998).

In short, when the '0·5' FSC threshold criterion is applied to asymmetric particles, rather than applying judicious conservatism one actually 'throws out the baby with the bathwater' (i.e. the signal with the noise). The worst effect of the '0·5 criterion' (for asymmetric particles) may well be its self-fulfilling prophecy character: the criterion indicates that the data have been sampled at a sufficiently high sampling rate in cases where under-sampling is the main resolution-limiting factor. It is interesting to note that when cryo-EM results are compared to similar results obtained by X-ray crystallography (Nogales *et al.* 1998; van Heel, 2000a;

Patwardhan *et al.* 2000c), the quality of the phases in EM compare favourably to their X-ray counterparts at comparable levels of resolution. Where the 3σ criterion has been used in these comparisons, the resolution values obtained for the cryo-EM data are in the same range as those obtained by standard X-ray crystallographic measures.

6.2 Influence of filtering

Filtering of the final reconstruction results is a controversial issue. Filtering may be used to emphasise the high-frequency details which typically suffer a substantial reduction of power during the EM imaging. Any filtering operation (CTF correction is also a filtering operation) may affect the subsequent graphical representation of the results. In a recent review (van Heel, 2000a), it was emphasised that the information content of a 3D data set is not influenced by normal filtering operations. (The 'normal' filtering operations exclude filters that explicitly set the amplitudes of specific frequency ranges to zero.) The information content of a 3D reconstruction at a given spatial frequency 'f' is directly related to the signal-to-noise ratio ($S_f^2/N_f^2 = SNR_f$) of the data (transformed to 3D Fourier space) at that resolution level through the formula (van Heel, 2000b; Sherman *et al.* 1998):

$$I_f = {}^2\log\left(1 + SNR_f\right). \tag{7}$$

Filtering changes the balance of the low-versus the high-frequency components in the data by multiplication with a rationally symmetric filter function in 3D Fourier space. A multiplicative filter will affect the signal S_f and the noise N_f to the same extent at any given spatial frequency and will thus not affect the SNR_f or the information content I_f of the data at the spatial frequency 'f'. Thus, filtering with well-behaved filter functions will not affect the information content of a 3D map. Why then, are CTF and temperature-factor correction such controversial issues? Whereas the information content remains unaffected, the representation of the data can change drastically with the amount and the specific details of the CTF or temperature factor corrections and with the threshold value applied. The representation, in turn, directly precedes the interpretation of the 3D reconstruction, and interpretation of the results apparently takes place very close to the emotional centre of the human brain.

6.3 Rendering

The human visual system is not geared towards interpreting 3D density distributions directly. In normal life one deals with opaque 3D objects which we have learned to appreciate by their surface texture in the course of evolution. Thus, for human consumption, one first needs to define the 'surface' of the 3D-density map and then to 'render' that artificial surface in order to generate understandable (stereo) images. We visually judge the quality of a structure by looking at the high-resolution texture of the molecular surface. The high-resolution texture may, however, be missing from the representation even though the information is present in the data (see: 'resolution'), just because of the balance of the high-resolution data components with respect to the low-frequency ones. When the temperature factor has not been sufficiently corrected for, it can thus be difficult to see the high-resolution details in the data. The same may be true when the very low frequencies (1/100 Å–1/30 Å), which are naturally suppressed by the low magnitude of the CTF in these frequency bands, are boosted to the maximum theoretical value of unity.

Although the fine details may not be visible in such surface renderings of the maps, it makes sense to threshold/interpret such data at 100 % (or even up to 120 %) of the expected volume of the molecular assembly (calculated from the molecular mass). Under normal circumstances, however, it will not make sense to threshold a structure at much more than the expected volume. Representing a ribosome, for example, at $\sim 200\%$ of the expected volume will make this structure, with an expected solvent content of 55 % (Stark *et al*. 1995) and thus full of 'hollows, voids, gaps and tunnels' (Yonath & Berkovitch-Yellin, 1993), look like an Idaho potato rather than like a Swiss cheese. On the other hand, when one wants to concentrate on the high-frequency details it is essential to emphasise those high-frequency details by an appropriate filtering operation. Threshold values corresponding to sometimes as little as 25 % of the expected volume may give the best representation of the structure's fine details. In X-ray crystallography one normally collects the information of a band of frequency components from ~ 2 Å to ~ 20 Å. The low-frequency components are missing from the measurement but those are also not normally relevant for fitting an atomic-resolution model into the 2 Å map. All one publishes are the atomic co-ordinates and the amount of expected density enclosed by the contour representation is not an issue of relevance.

The new series of low-resolution (by X-ray standards) structures of the ribosome and its components (Ban *et al*. 1998, 1999; Cate *et al*. 1999; Clemons *et al*. 1999; Culver *et al*. 1999; Tocilj *et al*. 1999) cover new territory in crystallography. The surface renderings presented in these papers most likely represent less than 50 % of the expected volume. Similarly, in the highest-resolution *E. coli* ribosomal structure (Matadeen *et al*. 1999), the surface representation filtering and threshold values are optimised to highlight the rRNA helices in the representation of the ribosome (Fig. 22). There is no real 'correct' way to filter the data and there is no 'correct' way to threshold the data, all one can do wrong is to interpret the map incorrectly (van Heel, 2000a). Filtering does not influence the information content of the map, but it does directly influence the visual and/or computational correlation techniques one uses for interpreting the maps.

6.4 Searching for known sub-structures

Single-particle cryo-EM is an excellent technique for studying large macromolecular assemblies such as ribosomes or icosahedral viruses. Much structural knowledge has already been collected on small parts of such systems, including atomic-resolution structures of some ribosomal proteins and ribosomal RNA segments. Known sub-structures are a blessing in more than one sense. Finding known sub-structures in the ribosome helps one interpret the vast biochemical knowledge base available on the ribosome and its components. At the same time, obtaining a picture of a macromolecular assembly, which includes a known atomic-resolution structure, helps one understand the resolution level achieved in a reconstruction. Given the never-ending discussions on resolution criteria, the direct comparison of cryo-EM results, with the same structure determined by entirely independent means, provides an independent internal calibration of the data.

Filtering does not affect the information content of a map but it does affect the visual fitting of known structures into the maps using interactive programs such as 'o' (Jones & Kjeldgaard, 1997) and 'ERNA-3D' (Mueller & Brimacombe, 1997a). Filtering may equally affect automatic search algorithms like 'ESSENS' (Kleywegt & Jones, 1997) that perform real-

space correlations between known atomic-resolution structures and intermediate resolution maps (van Heel, 2000a). Correlation coefficients are not the objective measures of similarity they are believed to be. Real-space correlation coefficients are not invariant to filtering of a map and must be used with care (van Heel *et al*. 1992a). There still is room for better similarity measures for use in the (automatic) interpretation of the 3D cryo-EM maps (van Heel, 2000b).

6.5 Interpretation

The most important part of the structure analysis is the interpretation of the resulting maps in terms of their biological significance. An incorrect interpretation is the only mistake that one can make with any good 3D map. Even when one computationally searches for a known sub-structure in a map and finds its best position, that result need not necessarily be correct. Different filtering parameters may place the optimum correlation peak in a different area of the map. An increase of the resolution of the map may also move the position of the substructure. Moreover, the sub-structure will always undergo some form of conformational change because of its incorporation into the larger structure. No existing program that we are aware of is able to search the map whilst introducing conformational changes to the substructure to find the best possible fit. Thus, the interpretation will remain an interactive procedure for a long time. For example, conformation of the large ribosomal subunit protein L9, in its 50S-bound state (Matadeen *et al*. 1999), is somewhat different from its 'isolated' crystal/NMR structure (Hoffman *et al*. 1996). Thus, an ESSENS search or C-terminal domain of L9 in the 50S *E. coli* map can be successful; yet, the search for the full L9 may go astray.

Of course, the interactively or computationally obtained results of searches for substructures must make sense in terms of the available biological information. No program will take over the task of highlighting the biological relevance of the results. In the examples given below, some typical issues are highlighted that have surfaced recently at the resolution levels of 5–15 Å. With the continuously increasing resolution levels of the results, the type of biological issues that can be elucidated change. Once the technology goes 'atomic', the types of biological questions one can answer will not differ from those currently answered by X-ray crystallography and NMR spectroscopy.

7. Examples

Among the first single-particle 3D structures to be determined by electron microscopy were icosahedral viruses (Crowther, 1971). As was already discussed in the introduction, 3D analysis is greatly facilitated by the fact that each viral image represents 60 different views of the asymmetric unit. Icosahedral particles can thus be seen as the 'micro-crystal' extreme of single particle methodologies. Three eye-opening papers (Böttcher *et al*. 1997; Conway *et al*. 1997; Trus *et al*. 1997) represented a significant jump in resolution relative to the typically 20–40 Å resolution levels that were achieved with earlier icosahedral reconstructions.

Why this jump? Apart from the improved cryo-EM experiments using highly coherent FEG sources, the statistics of the data collection were improved drastically from the 20–60 virus particles used traditionally to 6400 (Böttcher *et al*. 1997) or 600 (Conway *et al*. 1997). In the Böttcher *et al*. case, $60 \times 6400 = 384\,000$ individual asymmetric units of the icosahedral structure were averaged into the 3D results. The Böttcher *et al*. (1997) structure has since been
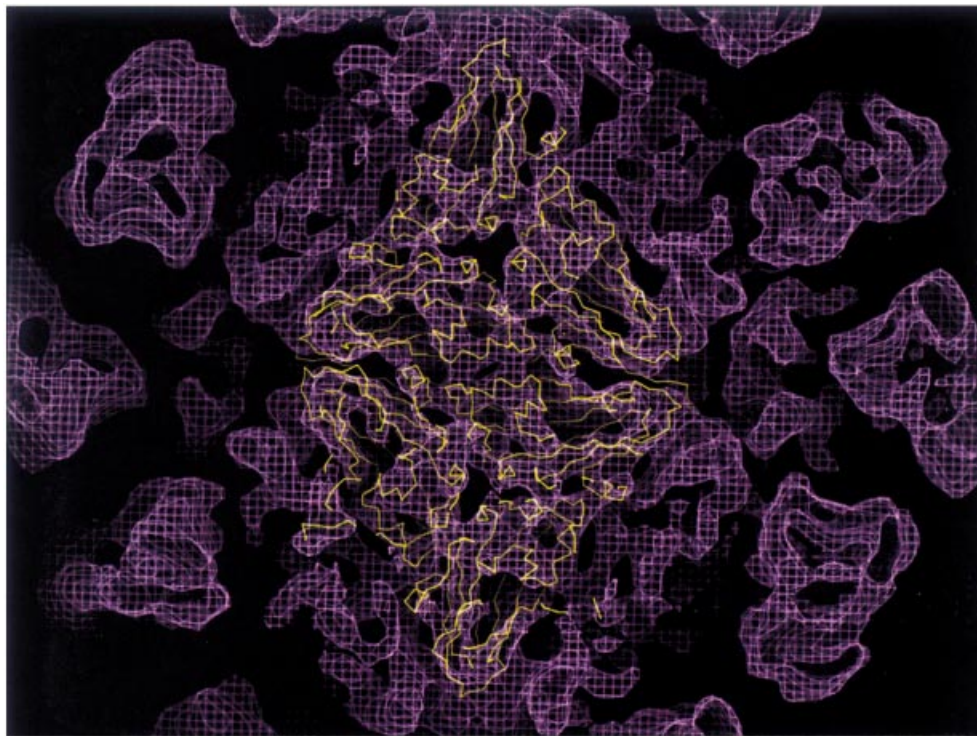
Fig. 18. Structure of an icosahedral virus (TBSV) at 5·9 Å resolution. The densities displayed here in wire-frame representation have been calculated using ∼ 6000 images of TBSV embedded in vitreous ice. The images were taken on a Philips CM200 FEG and processed by angular reconstitution as described in the main text. Inserted into this section through the cryo-EM map (purple) are backbone traces of six copies of the capsid protein, three of which form the (unique) asymmetric triangle of the icosahedral pointgroup symmetry.

used to phase X-ray crystallographic diffraction patterns of 3D crystals of the same assembly to atomic resolution (Wynne *et al.* 1999). Three-dimensional reconstructions derived from projection images alone, without experimental tilts, lack direct handedness information. Indeed, Böttcher *et al.* (1997) and Conway *et al.* (1997) show the structure with opposing hands. This issue has been solved with the availability of the X-ray structure.

7.1 Icosahedral symmetry: TBSV at 5·9 Å resolution

The single-particle processing programs in IMAGIC are formulated for all pointgroup symmetries, including the '532' icosahedral pointgroup symmetry. However, IMAGIC has only occasionally been used to tackle spherical viruses (Stewart *et al.* 1997). In order to test the resolutions achievable with single-particle cryo-EM, we chose to exploit the 532 pointgroup symmetry with its very favourable statistics. We thus embarked on a study of the Tomato Bushy Stunt Virus (TBSV), the first results of which were obtained recently (Patwardhan *et al.* 2000c). The structure of the virus capsid was solved more than two decades ago by X-ray crystallography (Harrison *et al.* 1978). It is obviously of great advantage to study a known structure while exploring the limits of a new technology: the X-ray structure serves as an external standard with which to independently gauge the quality of the cryo-EM/angular reconstitution results. The cryo-EM images were collected on a Philips CM200
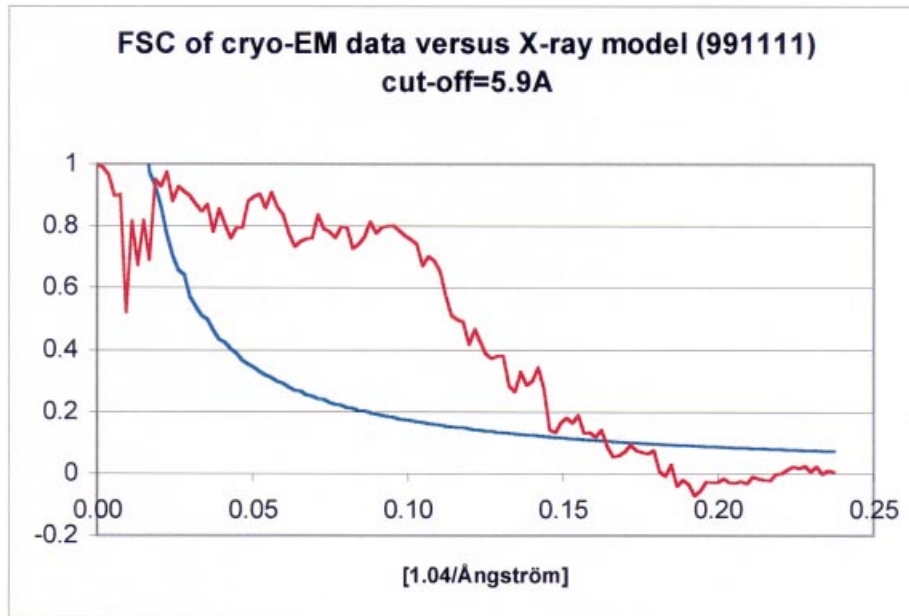
Fig. 19. TBSV at 5·9 Å: Fourier Shell Correlation. This FSC is calculated between the full cryo-EM 3D reconstruction of the TBSV particles and the X-ray structure of the icosahedral capsid. The cryo-EM reconstruction was masked with a soft-edged spherical-shell mask to exclude all material inside the capsid as well as all material beyond the spikes on the TBSV capsid surface. The exact crossing point between the FSC curve and the 3σ curve (multiplied by a factor of 60, to correct for the icosahedral symmetry (Orlova *et al.* 1997)) lies at 5·83 Å. The horizontal scaling (1·04) is 'funny' because a magnification discrepancy of 4% was found between the X-ray data and the cryo-EM magnification. We here assumed this to mean that our EM magnification calibration needed correction.

FEG cryomicroscope and digitised using the Image Science patchwork densitometer, described above (Fig. 5), using a step size of sampling step of 1·01 Å/pixel, subsequently coarsened to 2·02 Å/pixel by averaging 2×2 neighbouring pixels. The CTF correction and the 3D data analysis by angular reconstitution were performed as described above. The final reconstruction (Fig. 18), was determined to 5·9 Å resolution (Fig. 19).

Of the 3D reconstruction, a single section is shown, cut approximately perpendicular to the icosahedral two-fold axis. In the cryo-EM map, shown as a purple wire frame, the C_α co-ordinates of the X-ray structure of this entirely β-sheet structure are depicted. TBSV has a $T = 3$ capsid implying that there are three copies of the capsid protein in the asymmetric triangle of the icosahedron. Two such triangles – on two sides of the two-fold axis – are seen in this illustration. Most interestingly, the spatial-frequency information used for creating this map ranges from 4 to 11·5 Å exclusively. If the spatial-frequency information beyond 8·5 Å is removed by low-pass filtering of the map, the β-sheet information disappears from the graphical representations. The FSC curve between the cryo-EM results and a map generated from the 2·8 Å X-ray data is shown in Fig. 19. The curve indicates that the consistent information goes up to 5·9 Å if one applies the 3σ as the resolution criterion (Orlova *et al.* 1997). The FSC curve actually only departs from the threshold curve to drop to very low correlation values as a resolution level around 5·3 Å.

Normally in cryo-EM, one applies the FSC criterion for comparing two half data sets with each other. Here, the full cryo-EM data set is compared to the entirely independent X-ray

data. Thus, the noise statistics of the cryo-EM data in this comparison are better than would be the case for the comparison of one half of the data set with respect to the other. Interestingly, at the low-resolution end of the comparison, the correlation values are poorer than the typical behaviour for two halves of a cryo-EM data set with respect to each other. This may be associated with the fact that the low-resolution X-ray data are not measured in this experiment but rather created from the atomic PDB co-ordinates. We have shown here that our angular reconstitution programs are capable of achieving resolution levels exceeding those hitherto achieved by the earlier common lines programs (Böttcher *et al*. 1997). Single-particle cryo-EM may soon become an alternative technique to X-ray crystallography for solving the atomic structure of icosahedral viruses.

In spite of the aesthetically pleasing high level of symmetry of icosahedral capsids, life, in its most elementary DNA/RNA form, is one-dimensional and we thus cannot describe the life cycle of even the simplest spherical virus when we are restricted to studying icosahedral structures. For example, one is obviously interested in how a linear DNA/RNA molecule is packed within an icosahedral capsid. At the same time, some parts of the genome are likely to be anchored on specific structural components of the capsid and such parts may locally still adhere to the icosahedral symmetry of the environment and one can thus sometimes get away with an icosahedral analysis. The angular reconstitution formalism in IMAGIC (van Heel *et al*. 1996) is, as was mentioned above, formulated for general pointgroup symmetries and it allows a relaxation away from icosahedral symmetry to lower pointgroup symmetries (D5, D2, or even to C1), thus opening new opportunities for studying viral life cycles.

7.2 The D6 symmetrical worm hemoglobin at 13 Å resolution

The angular reconstitution methodology was first applied to the giant hemoglobin of *Lumbricus terrestris* (Schatz *et al*., 1995). This hemoglobin, also known as 'erythrocruorin', has a molecular mass of about 3·5 MDa (Lamy *et al*. 2000), and is known to have a D6 (622) pointgroup symmetry from earlier X-ray diffraction (Royer & Hendrickson, 1988) and conventional negative-stain EM studies (Boekema & van Heel, 1989). A controversy existed about the number of heme-containing myoglobin-type domains that are contained in this giant hemoglobin; two models existed, one with 144 heme domains (Vinogradov *et al*. 1991) and one with 192 (Ownby *et al*. 1993). A D6 pointgroup-symmetry structure consists of 12 identical asymmetric units, '1/12th' subunits. In the first cryo-EM 3D reconstruction a local three-fold axis was found within the 1/12th subunit (Schatz *et al*. 1995). This axis is compatible with 12 heme domains in the asymmetric unit but not with 16. The three-fold axis was later confirmed in a 3D reconstruction of *L. terrestris* hemoglobin at the 15 Å resolution level (Dube *et al*. 1995). The local three-fold axis in the worm hemoglobin was also confirmed in a large number of subsequent cryo-EM studies on the hemoglobins of *Lumbricus* and of other related annelids (de Haas *et al*. 1996; Taveau *et al*. 1999) at a 22–28 Å resolution level.

A very nice example of the complementarity of cryo-EM and X-ray crystallography is the recent determination of the X-ray structure of the *L. terrestris* hemoglobin structure at 5·5 Å (Royer *et al*. 2000). This structure was determined using a low-resolution cryo-EM structure as a phasing starting point. The X-ray structure revealed the detailed shape and position of the 144 myoglobin domains and an intricate network of linker chains. In particular, the 1/12th subunits are connected to the central torus of the structure by a linker chain structure

that approximately follows the local three-fold axis. The connecting structure consists of a novel triple coiled-coil structure. This coiled coil is offset from the local three-fold axis. The most intricate and appealing structure revealed is not yet at a level of resolution allowing an atomic interpretation but it certainly provides new insights into the assembly of large oligomeric structures. Moreover, at this level of resolution, the myoglobin folds can clearly be seen, allowing a direct determination of the absolute handedness of this giant hemoglobin. The angular reconstitution approach requires external information to allow for the determination of the absolute hand such as tilt series information. This information did exist (Cejka *et al.* 1991; Schatz *et al.* 1995; de Haas *et al.* 1996) but the results were in conflict with each other.

We have recently revived our cryo-EM structural elucidation efforts on this molecule and a preliminary result – based on only ~ 1000 molecular images (CM200 FEG) – is presented here at ~ 13 Å resolution (Fig. 20). At this level of resolution, we are able to see enough structural details on the surface of each of the myoglobin domains to directly determine the correct hand of the structure. As was already mentioned above, when reconstructions are performed from projection images without *a priori* known angular relationships, a left and right-handed structure fit equally well to the data. In X-ray crystallography, this problem is solved by, for example, looking at the hand of the α-helices in the electron density maps. The hand in the reconstruction (Fig. 20) is consistent with that of the X-ray structure. We expect to eventually be able to match the resolution levels currently achieved by X-ray crystallography purely by cryo-EM. The reason for reviving this research is, however, not necessarily to achieve atomic resolution but rather to assess conformational changes occurring in this intriguing complex. Cryo-EM may prove a unique technique to study the conformational changes occurring upon co-operative oxygen binding by this huge extra-cellular oxygen transporter.

## 7.3 Functional states of the 70S *E. coli* ribosome

In contrast to the TBSV and the worm hemoglobin, ribosomal structures are entirely asymmetrical. It is, in principle, more complicated to achieve high resolution levels for ribosomal structures by cryo-EM than for the highly symmetrical ones. The first 'zero tilt' cryo-EM reconstructions of the *E. coli* ribosome were published in 1995 (Frank *et al.* 1995; Stark *et al.* 1995), at 25 Å and 23 Å resolution respectively, and these were the first reliable ribosomal structures available. The experimental simplicity of the zero-tilt procedures allows one to concentrate on the biology of the ribosome and rapid progress was made in elucidating specific functional states of the ribosome. More has been learned about the structure of the various functional states of the *E. coli* ribosome in the past few years than in the preceding decades.

The first 3D reconstructions of the ribosome in defined functional states (Fig. 21), e.g. pre- and the post-translocational states, in which individual tRNAs can be seen, were published recently (Stark *et al.* 1997a). This was followed by the elucidation of the delivery of an amino acid to the translating ribosome by the ternary complex of EF-Tu and an aminoacyl-tRNA (Stark *et al.* 1997b). The known X-ray structure of the ternary complex (Nissen *et al.* 1995) could readily be recognised in the kirromycin-stalled 70S *E. coli* ribosome. Similarly, elongation factor G (Agrawal *et al.* 1998; Stark *et al.* 2000), has been visualised on the ribosome using the antibiotics fusidic acid and thiostrepton respectively. An early attempt to
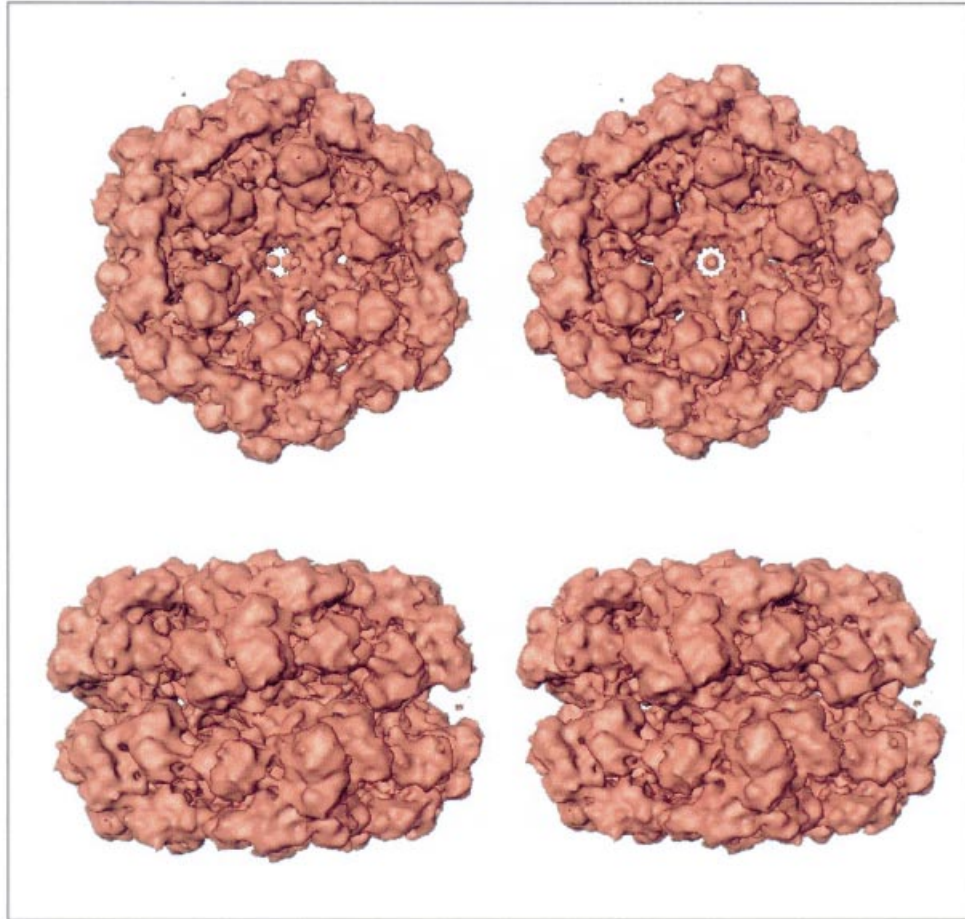
Fig. 20. Stereo pairs of *Lumbricus terrestris* hemoglobin at ∼ 13 Å resolution. This preliminary cryo-EM 3D reconstruction is the first in which the resolution is sufficiently high to see the shape of myoglobin fold directly in the reconstruction allowing a direct observation of the handedness of the structure. The handedness confirms the handedness of the X-ray structure (Royer *et al.* 2000). The 144 myoglobin domains are organised as 72 dimers. The 1/12th subunit contains six such dimers, organised around the local three-fold axis of the 1/12th subunit (Schatz *et al.* 1995).

find the a-, p- and e-site tRNA positions in the ribosome from an assay in which three tRNAs (a non-physiological situation) are known to bind simultaneously to the ribosome (Agrawal *et al.* 1996) revealed an additional mass in the inter-subunit gap, visible only when using low threshold levels for the graphical rendering. The tRNA positions interpreted from this mass, however, were flawed (Stark *et al.* 1997a; Cate *et al.* 1999). The main tunnel through the large ribosomal subunit had been suggested to correspond to the exit channel of the ribosome (Stark *et al.* 1995; Frank *et al.* 1995). This idea, which had been discussed in the literature for decades (Yonath *et al.* 1987), has seen substantial further confirmation, once based on the binding of Sec61 to yeast ribosomes (Beckmann *et al.* 1997), and once based on the structural homology between the mammalian and the bacterial ribosome (Dube *et al.* 1998a). This channel/tunnel ends on the lower-back side of the large subunit on the 'flat ribosomal surface' (Dube *et al.* 1998a) – compatible with docking onto a membrane – at the approximate position where early antibody labelling experiments (Bernabeau & Lake, 1982) had
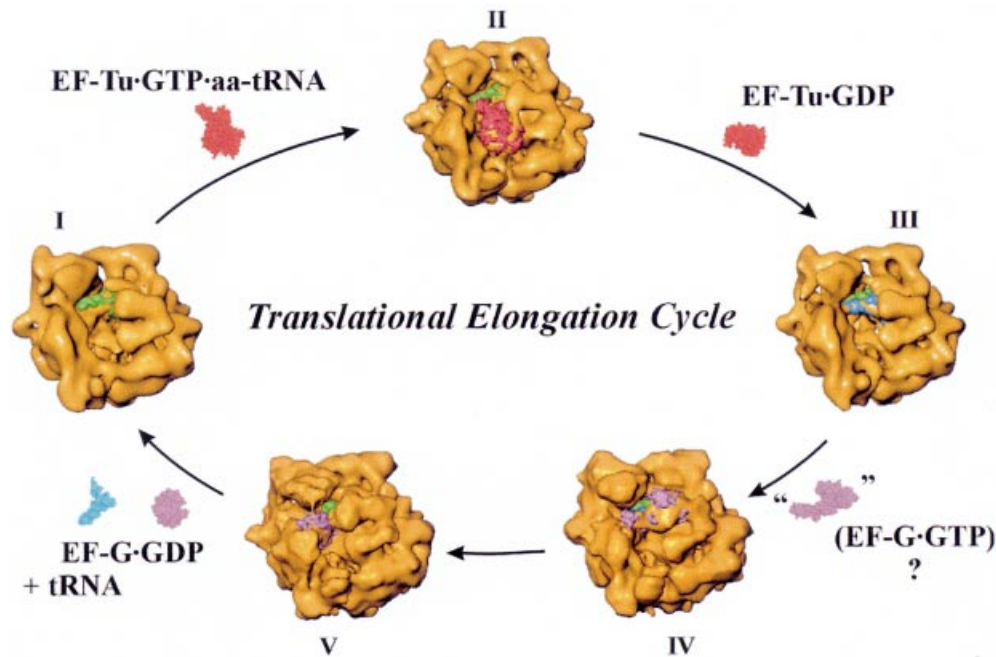
Fig. 21. Structures associated with the ribosomal elongation cycle (adding one amino acid to a nascent polypeptide chain). Shown schematically (clockwise), staring at the left: **I.** Post-translocational ribosome with P-site peptidyl-tRNA in the P site (Stark *et al.* 1997a). **II.** Delivery of an AA to the ribosome in ternary complex (Stark *et al.* 1997b); **III.** After peptidyl transfer, the A-site tRNA holds the nascent chain (Stark *et al.* 1997a); **IV.** Ribosome-bound EF-G before translocation (Stark *et al.* 2000); **V.** Ribosome-bound EF-G after translocation (Stark *et al.* 2000); the elongation cycle is completed after EF-G and the E-site tRNA leave the ribosome.

positioned the nascent polypeptide chain. The power of single-particle cryo-EM to elucidate specific functional states of the ribosome is undisputed. This is associated with the very simple and rapid specimen preparation techniques used for cryo-EM.

7.4 The 50S *E. coli* ribosomal subunit at 7·5 Å resolution

The resolution level achievable by cryo-EM reconstructions is rapidly improving. Fitting of 3D ribosomal RNA models and ribosomal protein structures into the maps are now possible with an ever-improving precision (Stark *et al.* 1995; Mueller & Brimacombe 1997; Matadeen *et al.* 1999; Gabashvili *et al.* 2000; Mueller *et al.* 2000). The model building technique has allowed one to start integrating the large body of biological information into the cryo-EM maps. The approach works well thanks to the scores of external biochemical constraints that help to define the structure. In a series of three back-to-back papers Brimacombe and co-workers have presented a model for the 16S rRNA folded in the 30S subunit (Mueller & Brimacombe, 1997; Mueller *et al.* 1997a, b). These 3D rRNA models can obviously not be considered atomic-resolution models, but it is clear that they will gradually refine, as higher and higher resolution cryo-EM or X-ray crystallographic data becomes available. With the availability of 5·5 and 4·5 Å resolution maps of the small ribosomal subunit more refined models are becoming available (Ban *et al.* 1999; Clemons *et al.* 1999). A 23S rRNA model has been constructed (Mueller *et al.* 2000) based on a 7·5 Å high-resolution cryo-EM map

(Matadeen *et al.* 1999). Cross-links to the N termini of nascent peptides of various lengths mapped to positions on the 23S rRNA that in turn mapped to positions along the channel (Mueller *et al.* 2000). This noteworthy result confirms the overall quality of the 23S rRNA model since this cross-link data had not been used a constraint for building the model in the first place. An earlier version of this 23S rRNA model was used to help interpret the positions of the expansion segments within a 21 Å cryo-EM map of the mammalian ribosome (Dube *et al.* 1998b).

After the initial successes of cryo-EM, and partially using the earlier cryo-EM maps to solve initial phasing problems, X-ray crystallography has recently been very successful in elucidating ribosomal structures (Ban *et al.* 1998, 1999; Cate *et al.* 1999; Clemons *et al.* 1999; Culver *et al.* 1999; Harms *et al.* 1999; Tocilj *et al.* 1999). These spectacular results at a resolution level of 4·5–9 Å are not yet interpretable at the atomic resolution level, but the resolution is sufficient to localise a large number of ribosomal protein with known X-ray or NMR structures, such as, for example, the proteins L2, L14, and L11 (Ban *et al.* 1999). The new X-ray structures, which include the large ribosomal subunit of *Haloarcula marismortui* (Ban *et al.* 1998, 1999) and of *T. thermophilus* (Cate *et al.* 1999), also show the exit channel at identically the same position as was seen previously in *E. coli* and in mammals (Dube *et al.* 1998a). The resolution levels in the X-ray crystallographic studies or ribosomal subunits are rapidly improving. The first atomic-resolution structures are probably going to be published this year. The X-ray studies concentrate on halophilic and thermophilic ribosomes, which were found to yield good crystals (Yonath & Berkovitch-Yellin, 1993).

A very recent cryo-EM reconstruction of the biochemically best characterised *E. coli* 50S subunit, at 7·5 Å resolution (Matadeen *et al.* 1999) (Fig. 22) shows that cryo-EM is still in 'the race to the ribosome structure' (Pennisi, 1999). A number of ribosomal proteins with known structures have been fitted in this cryo-EM map for the first time, including L9, L25 (Matadeen *et al.* 1999), and L25/5S rRNA, L5, L4 (Pape *et al.* 2000). It is noteworthy that some of these proteins (L9, L25) do not exist in archaebacteria. The positions found for other ribosomal proteins, for which analogues do exist between *H. marismortui* and *E. coli*, the X-ray positions could largely be confirmed. However, some contradictions in the interpretation of the data still exist which will be solved automatically when atomic resolution maps become available. For a recent review of these aspects of ribosomology, see van Heel (2000a).

The details of the 3D analysis of the *E. coli* 50S ribosomal subunit (Matadeen *et al.* 1999) have been used as examples of the processing in this review. The vitrified 50S samples were imaged in a Philips CM200 FEG microscope with a Gatan side-entry cryo-holder and cryo-transfer system at liquid-nitrogen temperatures. The micrographs were collected at underfocus values of $\sim 10\,000$–$20\,000$ Å at a magnification of 38 000. The best micrographs (checked by OD) were digitised using the Image Science patchwork densitometer with an aperture of 5 μm × 5 μm, corresponding to 1·32 Å × 1·32 Å on the specimen scale. For the final analysis, this sampling distance was re-interpolated to an effective sampling of 1·76 Å × 1·76 Å by masking the Fourier transforms of the raw images from 256 × 256 to 192 × 192 pixels. A total of 16 000 particles from seven micrographs entered the processing and led to the 3D map shown in Fig. 22. As an example of its richness in detail, the position of the L9 protein is depicted. A 70S initiation-like complex was reconstructed to 15 Å (Malhotra *et al.* 1998) and recently refined to 11·5 Å (Gabashvili *et al.* 2000). Similarly, the 70S kirromycin-stalled ribosome with the EF-Tu-tRNA ternary complex bound (Stark *et al.* 1997b), has been refined to resolution levels of ~ 13 Å (unpublished results). These studies

indicate that the 70S complexes are also amenable to high-resolution structure analysis and may thus equally reach ∼ 7·5 Å resolution and beyond. The comparison of the structure of bound 50S subunit in these complexes versus free 50S may help us understand the substantial conformational changes occurring in the ribosomal subunits upon binding to the 70S ribosome (Matadeen *et al.* 1999). To facilitate such comparative studies, a liberal exchange of maps must be encouraged (Liljas, 1999).

8. Perspectives

We are in a cryo-EM boom time. Many spectacular novel results have emerged in recent years. At the Gordon conference on 3D electron microscopy in 1997, where such beautiful structures as the atomic resolution structure of tubulin based on the analysis of 2D crystals (Nogales *et al.* 1998) were presented, some electron microscopists were already speaking of the 'global domination' of cryo-EM in structural biology. The sentence 'let's chase the train' – a word pun indicating it is time to start 'tracing the polypeptide chain' – was heard regularly in the bar area. This enthusiasm was triggered in part by the ever-higher resolution levels achieved by single-particle techniques, but only electron crystallography had already achieved this goal. Electron crystallography has been successful in elucidating a number of relatively small molecules that from good 2D crystals.

At the intermediate resolution levels of 10–20 Å achieved regularly by single particle methodologies life is exciting too, given the ease with which macromolecules can now be imaged in different conformational states. Equally important, intermediate-resolution structures can be of tremendous help to phase X-ray diffraction patterns to atomic resolution as is illustrated by almost all the examples given above. Nevertheless, there is a major challenge in pushing the single-particle methodologies to resolution levels better than 4 Å. The biological significance of being able to picture large complexes in three dimensions, in different conformational states, at resolution levels at which atomic models can be fitted in the resulting maps, can hardly be overestimated. For example, a detailed, atomic-resolution understanding of the conformational changes occurring in the bacterial and the eukayotic ribosomes during the elongation cycle may inspire the development of conceptually whole new families of antibiotics. What must still be done to reach this dream goal? We have already discussed many important issues that are important: the choice of defocus, the densitometry, CTF correction, etc. To further improve on the achievable resolution, some issues need further attention and a number of these issues are again associated with the CTF correction.

One issue concerns the standard CTF correction, but now looked at from a real space perspective rather than from a Fourier space one. Imposing a CTF on an image in Fourier space, has the effect in real space of convoluting the image with a point spread function (PSF). The PSF is simply the Fourier transform of the CTF. When an object is imaged at Scherzer focus (∼ 800 Å underfocus for the CM300 FEG; Fig. 1(*a*)), the corresponding PSF will be very narrow and will extend only over a few (1 Å) pixels in all directions (Fig. 2(*a*)). In Fig. 2, the PSFs are visualised by calculating the influence on a test image that consists of a single 1 Å pixel in the middle of a 256×256 pixel image. When the micrographs are taken at defocus values where one has some chance to see the particles, say, at 0·6 μm or 1·5 μ defocus (Figs 2(*b*) and 2(*c*), respectively), the PSF spreads the information of a single pixel over a large area. An image of the ribosome, on this scale, would almost fill the 256×256 frame. Thus, the information of an object like the ribosome, convoluted by the 1·5 μ PSF, spreads out over
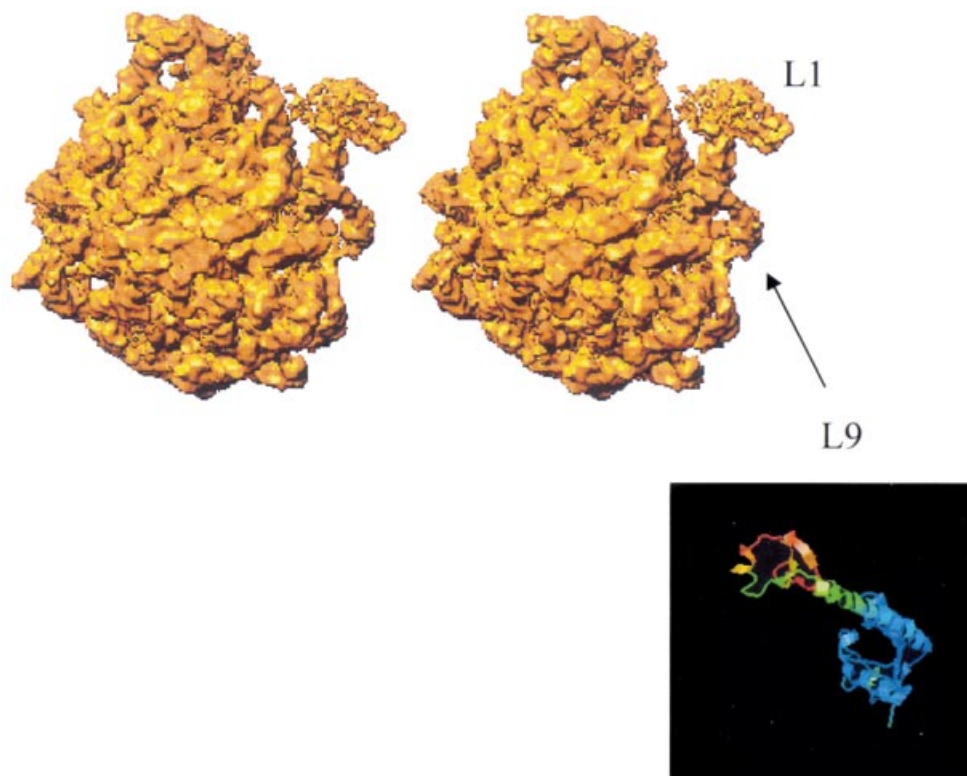
Fig. 22. The *E. coli* 50S ribosomal subunit at 7·5 Å resolution. This stereo pair of the bacterial large ribosomal subunit depicts the subunit from the back, towards its L1 side. As an example of the richness of detail the L9 protein is highlighted which contains a long α-helix connecting the C-terminal and the N-terminal part of the protein.

an area of the micrograph that covers the equivalent of at least 512×512 pixels in the original micrograph. This, in turn, implies that one must extract a sufficiently large area around each particle in order to be able to apply the CTF correction without losing information. In addition, one should not apply narrow masks around the particles prior to the CTF correction. After the CTF correction (Fig. 2(*d*)), the information is again concentrated around the pixel from which the information originally stems and thus much narrower masks become permissible.

From these considerations, it is obviously important to correct for the effects of the CTF directly as the first step of the processing as was discussed above. If the CTF is not corrected for right away, the information spread over large areas outside of the directly visible particle boundaries will persist throughout the processing. Large image files are then required throughout the entire processing. This large-file requirement would also apply to the 3D reconstruction volumes used. Although this traditional way of processing appears feasible even at high resolution, there is a direct disadvantage to organising the processing in this manner. The areas that are outside of the particles, and that need to be included in the processing, also contain noisy background and/or neighbouring particle information. That noise will hamper the overall processing of the micrographs.

A further, new CTF issue is that of the correction of astigmatism in the micrographs. Astigmatism is normally seen as a – not always avoidable – flaw in the data. We now rather

tend to see a certain amount of astigmatism in the single-particle cryo-EM data as a blessing rather than as a curse. In an image with astigmatism, two particles corresponding to the same view (the same out-of-plane degrees of rotation β and γ) will generally have a different in-plane rotational orientation α. Thus, these two molecular images within the same micrograph will represent information *de facto* taken with different defoci. Thus, all particles corresponding to the same view but different αs, after the astigmatic CTF correction, will contribute to a data collection effectively without CTF zeroes. This novel appreciation of astigmatism and its exploitation for achieving atomic-resolution cryo-EM structures preferably requires a calibration of the stigmators of the microscopes in absolute 'Å units' like the calibration of the defocus value itself. Since all modern electron microscopes are computer controlled, such an absolute calibration of the stigmators control should be simple to implement by the EM manufacturers.

It was mentioned in the introduction that, in contrast to X-ray crystallography, there is no phase problem in cryo-EM because one measures images rather than diffraction patterns. Nevertheless, single-particle cryo-EM at atomic resolution is all about retrieving the high-resolution phase information. The phase problem that we face in EM is that of *phase coherency* at high resolution. The extensive alignment algorithms and the CTF determination schemes discussed above are all aimed at making the high-frequency data components coherent. In averages of poorly aligned or poorly CTF-corrected molecular images, the high-frequency data components are essentially summed incoherently and whatever information had originally been hidden in the noise, it will average out rather than emerge during the summing procedures. All (iterative) refinement procedures described above are aimed at making the high-resolution phases emerge from the noise.

Iterative refinements already play a central role in the processing, but further refinement methods may still be required to achieve our goals. One obvious candidate for further improvement is a reduction of the defocus envelope discussed above. This envelope consists of two components. The first component is given by the 'Z' position of the particle in the vitreous-ice layer. In the methods outlined above, the defocus parameter of a single particle is assigned based on the larger environment of the particle. A 'Z-alignment' may be required for each particle individually, which implies determining (refining) the defocus parameters for each particle individually. The other component of the defocus envelope is due to the differences in defocus occurring within the individual particles. This effect is predictable and in principle correctable, but new software will have to be developed for the purpose (DeRosier, 2000). Finally, some of the refinement approaches that are in routine use in X-ray crystallography such as solvent flattening may become useful in the field of cryo-EM once one starts building atomic models.

Achieving atomic resolution for non-crystallised particles may still be a dream but there are no fundamental reasons why it cannot be achieved (Henderson, 1995) and we are getting closer and closer to its realisation. This dream is likely to first come true for small icosahedral viruses that are so regular that they may be considered 'micro-crystals'. Achieving atomic resolution is, of course, not the only issue at stake. Three-dimensional reconstructions of the ribosome in different functional states at ~ 7 Å resolution may be at least as exciting in terms of advances in our understanding of biology as an atomic resolution structure of an icosahedral capsid. With single-particle cryo-EM one can potentially have it all. One can elucidate different conformations of, say, a ribosome at 10 Å resolution first, and then slowly refine the resolution of the reconstructions to finally yield atomic co-ordinates of the different

functional states. Cryo-EM has already delivered large amounts of information on crucially important macromolecular complexes and it still bears great promises for the near future. Cryo-EM is thus a hot spot of modern structural biology. Many scientific institutions worldwide are establishing new cryo-EM laboratories and, consequently, we are facing a shortage of electron microscopists: a truly rosy perspective for our current generation of Ph.D. students.

9. Acknowledgements

10. References

ADRIAN, M., DUBOCHET, J., FULLER, S. D. & HARRIS, J. R. (1998). Cryo-negative staining. *Micron* **29**, 145–160.

ADRIAN, M., DUBOCHET, J., LEPAULT, J. & McDOWALL, A. W. (1984). Cryo-electron microscopy of viruses. *Nature* **308**, 32–36.

AGRAWAL, R. K., PENCZEK, P., GRASSUCCI, R. A. & FRANK, J. (1998). Visualization of the elongation factor G on *E. coli* 70S ribosome: the mechanism of translation. *Proc. natn. Acad. Sci. USA* **95**, 6134–6138.

AGRAWAL, R. K., PENCZEK, P., GRASSUCCI, R. A., LI, Y., LIETH, A., NIERHAUS, K. H. & FRANK, J. (1996). Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science* **271**, 1000–1002.

BAKER, T. S. & CHENG, R. H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. struct. Biol.* **116**, 120–130.

BAKER, T. S., DRAK, J. & BINA, M. (1988). Reconstruction of the three-dimensional structure of simian virus 40 and visualization of chromatin core. *Proc. natn. Acad. Sci. USA* **85**, 422–426.

BAKER, T. S., OLSON, N. H. & FULLER, S. D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol. molec. Biol. Rev.* **63**, 862–922.

BAN, N., FREEBORN, B., NISSEN, P., PENCZEK, P., GRASSUCCI, R. A., SWEET, R., FRANK, J., MOORE, P. B. & STEITZ, T. A. (1998). A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* **93**, 1105–1115.

BAN, N., NISSEN, P., HANSEN, J., CAPEL, M., MOORE, P. & STEITZ, T. A. (1999). Placement of protein and RNA structures into a 5 Å-resolution maps of the 50S ribosomal subunit. *Nature* **400**, 841–847.

BECKMANN, R., BUBECK, D., GRASSUCCI, R., PENCZEK, P., VERSCHOOR, A., BLOBEL, G. & FRANK, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* **278**, 2123–2126.

BELLARE, J. R., DAVIS, H. T., SCRIVEN, L. E. & TALMON, Y. (1988). Controlled environment vitrification system: an improved sample preparation technique. *J. elect. Microsc. Tech.* **10**, 87–111.

BERNABEAU, C. & LAKE, J. A. (1982). Nascent polypeptide chains emerge from the exit domain of the large ribosomal subunit: immune mapping of the nascent chain. *Proc. natn. Acad. Sci. USA* **79**, 3111–3115.

BOEKEMA, E. J. & VAN HEEL, M. (1989). Molecular shape of *Lumbricus terrestris* erythrocruorin studied by electron microscopy and image analysis. *Biochim. biophys. Acta* **957**, 370–379.

BOEKEMA, E. J., BERDEN, J. A. & VAN HEEL, M. (1986). Structure of mitochondrial F1-ATPase studied by electron microscopy and image processing. *Bioch. biophys. Acta* **851**, 353–360.

BOETTCHER, C., LUDWIG, K., HERRMANN, A., VAN HEEL, M. & STARK, H. (1999). Structure of influenza haemagglutinin at neutral and at fusogenic pH by electron cryo-microscopy. *FEBS Lett.* **46**, 255–259.

BOETTCHER, C., STARK, H. & VAN HEEL, M. (1996). Stacked bilayer helices: a new structural organization of amphiphilic molecules. *Ultramicroscopy* **62**, 133–139.

BORLAND, L. & VAN HEEL, M. (1990). Classification of image data in conjugate representation spaces. *J. opt. Soc. Am.* **A7**, 601–610.

BÖTTCHER, B., WYNNE, S. A. & CROWTHER, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88–91.

CATE, J. H., YUSUPOV, M. M., YUSUPOVA, G. Z., EARNEST, T. N. & NOLLER, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104.

CEJKA, Z., SANTINI, C., TOGNON, G. & GHIRETTI MAGALDI, A. (1991). The molecular architecture of the extracellular hemoglobin of *Ophelia bicornis*: analysis of two-dimensional crystalline arrays. *J. struct. Biol.* **107**, 259–267.

CHANZY, H., FRANC, J. M. & HERBAGE, D. (1976). High-angle electron diffraction of frozen-hydrated collagen. *Biochem. J.* **153**, 139–140.

CHENG, Y. & TAYLOR, K. A. (1998). Characterization of the low magnification performance of a Philips CM300-FEG. *Ultramicroscopy* **74**, 209–220.

CLEMONS, W. M., MAY, J. L., WIMBERLY, B. T., McCUTCHEON, J. P., CAPEL, M. S. & RAMAKRISHNAN, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5·5 Å resolution. *Nature* **400**, 833–840.

CONWAY, J. & STEVEN, A. C. (1999). Methods for reconstructing density maps of 'single particles' from cryoelectron micrographs to subnanometer resolution. *J. struct. Biol.* **128**, 106–118.

CONWAY, J., CHENG, N., WINGFIELD, P. T., STAHL, S. J. & STEVEN, A. C. (1997). Visualisation of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **385**, 91–94.

CROWTHER, R. A. (1971). Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Phil. Trans. R. Soc. Lond.* B **261**, 221–230.

CROWTHER, R. A. & AMOS, L. A. (1971). Harmonic analysis of electron images with rotational symmetry. *J. molec. Biol.* **60**, 123–130.

CROWTHER, R. A., HENDERSON, R. & SMITH, J. M. (1996). MRC image processing programs. *J. struct. Biol.* **116**, 9–16.

CROWTHER, R. A., KISELEV, N. A., BOTTCHER, B., BERRIMAN, J. A., BORISOVA, G. P., OSE, V. & PUMPENS, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**, 943–950.

CULVER, G. M., CATE, J. H., YUSUPOVA, G. Z., YUSUPOV, M. M., & NOLLER, H. F. (1999). Identification of an RNA–protein bridge spanning the ribosomal subunit interface. *Science* **285**, 2133–2136.

DEROSIER, D. J. (2000). Correction of high-resolution data for curvature of the Ewald sphere. *Ultramicroscopy* **81**, 83–98.

DEROSIER, D. J. & KLUG, A. (1968). Reconstruction of three-dimensional structures from electron micrographs. *Nature* **217**, 130–134.

DEROSIER, D. J. & MOORE, P. B. (1970). Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J. molec. Biol.* **52**, 355–369.

DIERKSEN, K., TYPKE, D., HEGERL, R., KOSTER, A. J. & BAUMEISTER, W. (1992). Towards automatic electron tomography. *Ultramicroscopy* **40**, 71–87.

DUBE, P., BACHER, G., STARK, H., MUELLER, F., ZEMLIN, F., VAN HEEL, M. & BRIMACOMBE, R. (1998b). Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 Å resolution. *J. molec. Biol.* **279**, 403–421.

DUBE, P., STARK, H., ORLOVA, E. V., SCHATZ, M., BECKMANN, E., ZEMLIN, F. & VAN HEEL, M. (1995). 3D structure of single macromolecules at 15 Å resolution by cryo-microscopy and angular reconstitution. In: *Microscopy and Microanalysis 1995*, eds. Bailey, G. W., Ellisman, M. H., Henigar, R. A. & Zaluzec, N. J., pp. 838–839. New York: Jones & Begell Publishing.

DUBE, P., TAVARES, P., LURZ, R. & VAN HEEL, M. (1993). Bacteriophage SPP1 portal protein: a DNA pump with 13-fold symmetry. *EMBO J.* **15**, 1303–1309.

DUBE, P., WIESKE, M., STARK, H., SCHATZ, M., STAHL, J., ZEMLIN, F., LUTSCH, G. & VAN HEEL, M. (1998a). The 80S rat liver ribosome at 25 Å resolution by electron cryomicroscopy and angular reconstitution. *Structure* **6**, 389–399.

DUBOCHET, J., & ADRIAN, M., CHANG, J.-J., HOMO, J.-C., LEPAULT, J., McDOWALL, A. & SCHULTZ, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228.

ERMANTRAUT, E., WOHLFAHRT, K. & TICHELAAR, W. (1998). Perforated support foils with pre-defined hole size, shape and arrangement. *Ultramicroscopy* **74**, 75–81.

FARROW, N. A. & OTTENSMEYER, F. P. (1992). *A posteriori* determination of relative projection directions of arbitrarily oriented macromolecules. *J. opt. Soc. Am.* **A9**, 1749–1760.

FRANK, J. (1990). Classification of macromolecular assemblies studied as 'single particles'. *Q. Rev. Biophys.* **23**, 281–329.

FRANK, J. & WAGENKNECHT, T. (1984). Automatic selection of molecular images from electron micrographs. *Ultramicroscopy* **12**, 169–176.

FRANK, J., RADERMACHER, M., PENCZEK, P., ZHU, J., LI, Y., LADJADJ, M. & LEITH, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. struct. Biol.* **116**, 190–199.

FRANK, J., SHIMKIN, B. & DOWSE, H. (1981b). SPIDER – A modular software system for electron image processing. *Ultramicroscopy* **6**, 343–358.

FRANK, J., VERSCHOOR, A. & BOUBLIK, M. (1981). Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* **214**, 1353–1355.

FRANK, J., ZHU, J., PENCZEK, P., LI, Y., SRIVASTAVA, S., VERSCHOOR, A., RADERMACHER, M., GRASSUCCI, R., LATA, R. K. & AGRAWAL, R. K. (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* **376**, 441–444.

FUJIYOSHI, Y., MIZUSAKI, T., MORIKAWA, K., YAMAGISHI, H., AOKI, Y., KIHARA, H. & HARADA, Y. (1991). Development of a superfluid helium stage for high-resolution electron microscopy. *Ultramicroscopy* **38**, 241–251.

GABASHVILI, I. S., AGRAWAL, R. K., SPAHN, C. M. T., GRASSUCCI, R. A., SVERGUN, D. I., FRANK, J. & PENCZEK, P. (2000). Solution structure of the *E. coli* 70S ribosome at 11·5 Å resolution. *Cell* **100**, 537–549.

GOLUB, G. & VAN LOAN, C. (1996). *Matrix computations*, third edition. London: The Johns Hopkins University Press Ltd.

GONCHAROV, A. B. & GELFAND, M. S. (1988). Determination of mutual orientation of identical particles from their projections by the moments method. *Ultramicroscopy* **25**, 317–328.

DE HAAS, F., ZAL, F., YOU, V., LALLIER, F., TOULMOND, A. & LAMY, J. N. (1996). Three-dimensional reconstruction by cryoelectron microscopy of the giant hemoglobin of the polychaete worm *Alvinella pompejana*. *J. molec. Biol.* **264**, 111–120.

HARAUZ, G. & OTTENSMEYER, F. P. (1984). Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs. *Ultramicroscopy* **12**, 309–319.

HARAUZ, G. & VAN HEEL, M. (1986a). Direct 3D reconstruction from projections with initially unknown angles. In *Pattern Recognition in Practice II*, eds. Gelsema, E. S. & Kanal, L., pp. 279–288. Amsterdam: North-Holland Publishing.

HARAUZ, G. & VAN HEEL, M. (1986b). Exact filters for general geometry three dimensional reconstruction. *Optik* **73**, 146–156.

HARMS, J., TOCILJ, A., LEVIN, I., AGMON, I., KÖLLN, I., STARK, H., VAN HEEL, M., CUFF, M., SCHLÜNZEN, F., BASHAN, A., FRANCESCHI, F. & YONATH, A. (1999). Elucidating the structure of ribosomal particles: an interplay between electron-cryo-microscopy and X-ray crystallography. *Structure* **7**, 931–941.

HARRISON, S. C., OLSON, A., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). Tomato bushy stunt virus at 2·9 Å resolution. *Nature* **276**, 368–373.

HEBERT, H., SCHMIDT-KREY, I., MORGENSTERN, R., MURATA, K., HIRAI, T., MITSUOKA, K. & FUJIYOSHI, Y. (1997). The 3·0 Å projection structure of glutathione transferase as determined by electron crystallography of p2₁2₁2 two-dimensional crystals. *J. molec. Biol.* **272**, 751–758.

HENDERSON, R., BALDWIN, J. M., CESKA, T. A., ZEMLIN, F., BECKMANN, E. & DOWNING, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. *J. molec. Biol.* **213**, 899–929.

HENDERSON, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193.

HOFFMAN, D. W., CAMERON, C. S., DAVIES, C., WHITE, S. & RAMAKRISHNAN, V. (1996). Ribosomal protein L9: a structure determination by the combined use of X-ray crystallography and NMR spectroscopy. *J. molec. Biol.* **264**, 1058–1071.

HOPPE, W., GASSMANN, J., HUNSMANN, N., SCHRAMM, H. J. & STURM, M. (1974). Three-dimensional reconstruction of individual negatively stained yeast fatty acid synthetase molecules from tilt-series in the electron microscope. *Hoppe-Seyler's Z. Physiol. Chem.* **355**, 1483–1487.

JONES, T. A. & KJELDGAARD, M. (1997). Electron-density map interpretation. *Method Enzymol.* **277** B, 173–207.

KIMURA, Y., VASSYLYEV, D. G., MIYAZAWA, A., KIDER, A., MATSUSHIMA, M., MITSIKOA, K., MURATA, K., HIRAI, T. & FUJIYOSHI, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* **389**, 206–211.

KISSEBERTH, N., WHITTAKER, M., WEBER, D., POTTER, C. S. & CARRAGHER, B. (1997). emScope: a tool kit for control and automation of a remote electron microscope. *J. struct. Biol.* **120**, 309–319.

KLEYWEGT, G. J. & JONES, T. A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta crystallogr.* D **53**, 179–185.

LAMY, J., KUCHUMOV, A. M., TAVEAU, J.-C., VINOGRADOV, S. N. & LAMY, J. N. (2000). Reassembly of *Lumbricus terrestris* hemoglobin: a study by matrix-assisted laser desorption/ionization mass spectrometry and 3D reconstruction from frozen-hydrated specimens. *J. molec. Biol.* **298**, 633–647.

LATA, K. R., PENCZEK, P. & FRANK, J. (1995). Automatic particle picking from electron micrographs. *Ultramicroscopy* **58**, 381–391.

Liljas, A. (1999). Function is structure. *Science* **285**, 2077–2078.

Malhotra, A., Penczek, P., Agrawal, R. K., Gabashivili, I. S., Grassucci, R. A., Junemann, R., Burkhardt, N., Nierhaus, K. H. & Frank, J. (1998). *E. coli* 70S ribosome at 15 Å resolution by cryo-electron microscopy: localization of fMet-tRNA$_{fMet}$ and fitting of L1 protein. *J. molec. Biol.* **280**, 103–115.

Matadeen, R., Patwardhan, A., Gowen, B., Orlova, E. V., Pape, T., Mueller, F., Brimacombe, R. & van Heel, M. (1999). The *E. coli* large ribosomal subunit at 7·5 Å resolution. *Structure* **7**, 1575–1583.

Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. (1999). Nicotinic acethylcholine receptor at 4·6 Å resolution: transverse tunnels in the channel wall. *J. molec. Biol.* **288**, 765–786.

Mueller, F. & Brimacombe, R. (1997). A new model for the three-dimensional folding of *Escherichia coli* 16s ribosomal RNA, II. The RNA-protein interaction data. *J. molec. Biol.* **271**, 545–565.

Mueller, F., Sommer, I., Baranov, P., Matadeen, R., Stoldt, M., Wöhnert, J., Görlach, M., van Heel, M. & Brimacombe, R. (2000). The 3D arrangement of the 23S and 5S rRNA in the Escherichia coli 50S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7·5 Å resolution. *J. molec. Biol.* **298**, 35–59.

Mueller, F., Stark, H., van Heel, M., Rinke-Appel, J. & Brimacombe, R. (1997a). A new model for the three-dimensional folding of *Escherichia coli* 16s ribosomal RNA, I. Fitting the RNA to a 3D electron density map at 20 Å. *J. molec. Biol.* **271**, 524–544.

Mueller, F., Stark, H., van Heel, M., Rinke-Appel, J. & Brimacombe, R. (1997b). A new model for the three-dimensional folding of *Escherichia coli* 16s ribosomal RNA, III. The topography of the functional centre. *J. molec. Biol.* **271**, 566–587.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. & Nyborg, J. (1995). Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472.

Nogales, E., Wolf, S. G. & Downing, K. H. (1998). Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, 199–203.

Oettl, H., Hegerl, R. & Hoppe, W. (1983). Three-dimensional reconstruction and averaging of 50S ribosomal subunits of *Escherichia coli* from electron micrographs. *J. molec. Biol.* **163**, 431–450.

Orlova, E. & van Heel, M. (1994). Angular reconstitution of macromolecules with arbitrary point group symmetry. *Proceedings ICEM 13–Paris 1994*, Vol. **1**, 507–508.

Orlova, E. V., Dube, P., Harris, J. R., Beckmann, E., Zemlin, F., Markl, J. & van Heel, M. (1997). Structure of Keyhole Limpet Hemocyanin Type 1 (KLH1) at 15 Å resolution by electron cryo-microscopy and angular reconstition. *J. molec. Biol.* **271**, 417–437.

Orlova, E. V., Serysheva, I. I., van Heel, M., Hamilton, S. L. & Chiu, W. (1996). Two structural configurations of the skeletal muscle calcium release channel. *Nature struct. Biol.* **3**, 547–552.

Ownby, D. W., Zhu, H., Schneider, K., Beavus, R. C., Chait, B. T. & Riggs, A. F. (1993). The extracellular hemoglobin of the earthworm *Lumbricus terrestris*: determination of subunit stoichiometry. *J. biol. Chem.* **268**, 13539–13547.

Patwardhan, A., Schatz, M. & van Heel, M. (2000a). A high-resolution patchwork densitometer. *Ultramicroscopy*, in press.

Patwardhan, A., Stark, H. & van Heel, M. (2000b). CTF Correction of low-dose micrographs of vitreous-ice embedded specimens. *Ultramicroscopy*, in press.

Patwardhan, A., Stark, H. & van Heel, M. (2000c). Cryo-EM of icosahedral viruses at 5·9 Å resolution: phasing in β-sheets, phasing out resolution criteria. *Structure*, submitted.

Penczek, P. (1998). Measures of resolution using Fourier shell correlation. *J. molec. Biol.* **280**, 115–116.

Penczek, P., Radermacher, M. & Frank, J. (1992). Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **40**, 33–53.

Pennisi, E. (1999). The race to the ribosome structure. *Science* **285**, 2048–2051.

Radermacher, M. (1988). Three-dimensional reconstruction of single particles from random and non-random tilt series. *J. elect. Microsc. Tech.* **9**, 359–394.

Radermacher, M. (1994). Three-dimensional reconstruction from random projections: orientational alignment via Radon transforms. *Ultramicroscopy* **53**, 121–136.

Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. (1987). Three-dimensional structure of the large ribosomal subunit from *Escherichia coli*. *EMBO J.* **6**, 1107–1114.

Radon, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Manningfaltigkeiten. Berichte über die Verhandlungen der Königlich Sachsischen Gesellschaft der Wissenschaften zu Leipzig. *Math. Phys. Klasse* **69**, 262–277.

Royer, W. E. & Hendrickson, W. A. (1988). Molecular symmetry of *Lumbricus erythrocruorin*. *J. biol. Chem.* **263**, 13762–13765.

Royer, W. E., Strand, K., van Heel, M. & Hendrickson, W. A. (2000). Structural hierarchy in the giant assemblage of annelid erythrocruorin. *Proc. natn. Acad. Sci. USA*, in press.

Saxton, W. O. (1996). Semper: distortion compensation, selective averaging, 3-D reconstruction, and transfer function correction in a highly programmable system. *J. struct. Biol.* **116**, 230–236.

SAXTON, W. O. & BAUMEISTER, W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. *J. Microsc*. **127**, 127–138.

SAXTON, W. O. & FRANK, J. (1977). Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy* **2**, 219–227.

SCHATZ, M. & VAN HEEL, M. (1990). Invariant classification of molecular views in electron micrographs. *Ultramicroscopy* **32**, 255–264.

SCHATZ, M., ORLOVA, E. V., DUBE, P., JÄGER, J. & VAN HEEL, M. (1995). Structure of lumbricus terrestris hemoglobin at 30 Å resolution determined using angular reconstitution. *J. struct. Biol*. **114**, 28–40.

SCHATZ, M., ORLOVA, E. V., DUBE, P., STARK, H., ZEMLIN, F. & VAN HEEL, M. (1997). Angular reconstitution in 3D electron microscopy: practical and technical aspects. *Scanning Microscopy* **11**, 179–194.

SERYSHEVA, I., ORLOVA, E. V., SHERMAN, M, CHIU, W., HAMILTON, S. & VAN HEEL, M. (1995). Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle calcium release channel. *Nature struct. Biol*. **2**, 18–24.

SHANNON, C. E. (1948). A mathematical theory of communications. *Bell System Tech. Jl* **27**, 379–423, 623–656.

SHERMAN, M. B., SOEJIMA, T., CHIU, W. & VAN HEEL, M. (1998). Multivariate analysis of single unit cells in electron crystallography. *Ultramicroscopy* **74**, 179–199.

STARK, H., MUELLER, F., ORLOVA, E. V., SCHATZ, M., DUBE, P., ERDEMIR, T., ZEMLIN, F., BRIMACOMBE, R. & VAN HEEL, M. (1995). The 70S ribosome at 23 Å resolution: fitting the ribosomal RNA. *Structure* **3**, 815–821.

STARK, H., ORLOVA, E. V., RINKE-APPEL, J., JÜNKE, M., MUELLER, F., RODNINA, M., WINTERMEYER, W., BRIMACOMBE, R. & VAN HEEL, M. (1997a). Arrangement of tRNAs in pre- and post-translocational ribosomes revealed by electron cryomicroscopy. *Cell* **88**, 19–28.

STARK, H., RODNINA, M. V., RINKE-APPEL, J., BRIMACOMBE, R., WINTERMEYER, W. & VAN HEEL, M. (1997b). Visualisation of elongation factor Tu on the *Escherichia coli* ribosome. *Nature* **389**, 403–406.

STARK, H., RODNINA, M. V., WIEDEN, H. J., VAN HEEL, M. & WINTERMEYER, W. (2000). Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* **100**, 301–309.

STARK, H., ZEMLIN, F. & BOETTCHER, C. (1996). Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. *Ultramicroscopy* **63**, 75–79.

STEINKILBERG, M. & SCHRAMM, H. J. (1980). Eine verbesserte Drehkorrelationsmethode für die Strukturbestimmung biologischer Makromoleküle durch Mittelung elektronen-mikroskopischer Bilder. *Hoppe-Seylers Z. Physiol. Chem*. **361**, 1363–1369.

STEWART, P. L., CHIU, C. Y., HUANG, S., MUIR, T., ZHAO, Y., CHAIT, B., MATHIAS, P. & NEMEROW, G. R. (1997). Cryo-EM visualization of an exposed RGD epitope an adenovirus that escapes antibody neutralisation. *EMBO J*. **16**, 1189–1198.

TAVEAU, J. C., BOISSET, N., VINOGRADOV, S. N. & LAMY, J. N. (1999). Three-dimensional reconstruction of *Lumbricus terrestris* hemoglobin at 22 Å resolution: intramolecular localization of the globin and linker chains. *J. molec. Biol*. **289**, 1343–1359.

TAYLOR, K. A. & GLASER, R. M. (1976). Electron microscopy of frozen hydrated biological specimens. *J. ultrastruct. Res*. **55**, 448–546.

TOCILJ, A., SCHLUNZEN, F., JANELL, D., GLUHMANN, M., HANSEN, H. A., HARMS, J., BASHAN, A., BARTELS, H., AGMON, I., FRANCESCHI, F. & YONATH, A. (1999). The small ribosomal subunit from *Thermus thermophilus* at 4·5 Å resolution: pattern fittings and the identification of a functional site. *Proc. natn. Acad. Sci. USA* **96**, 14252–14257.

TRUS, B. L., RODEN, R. B. S., GREENSTONE, H. L., VRHEL, M., SCHILLER, J. T. & BOOY, F. P. (1997). Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution. *Nature struct. Biol*. **4**, 413–420.

UNWIN, N. (1995). Acetylcholine receptor channel imaged in the open state. *Nature* **373**, 37–43.

UNWIN, P. N. T. & HENDERSON, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. molec. Biol*. **94**, 425–440.

VALENTINE, R. C. (1965). Characteristics of emulsions for electron microscopy. *Lab. Invest*. **14**, 596–602.

VAN HEEL, M. (1978). On the imaging of relatively strong objects in partially coherent illumination in optics and electron optics. *Optik* **47**, 389–408.

VAN HEEL, M. (1982). Detection of objects in quantum noise limited images. *Ultramicroscopy* **8**, 331–342.

VAN HEEL, M. (1984a). Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* **13**, 165–183.

VAN HEEL, M. (1984b). Three-dimensional reconstructions from projections with unknown angular relationship. *Proc. 8th Eur. Cong. on EM*, *Budapest*, Vol. **2**, 1347–1348.

VAN HEEL, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111–124.

VAN HEEL, M. (1989). Classification of very large electron microscopical image data sets. *Optik* **82**, 114–126.

VAN HEEL, M. (2000a). Unveiling ribosomal structures: the final phases. *Curr. Opin. struct. Biol*. **10**, 259–264.

VAN HEEL, M. (2000b). Images and information. *Ultramicroscopy*, in preparation.

VAN HEEL, M. & FRANK, J. (1981). Use of multivariate

statistics in analyzing the images of biological macromolecules. *Ultramicroscopy* **6**, 187–194.

van Heel, M. & Harauz, G. (1986). Resolution criteria for three dimensional reconstructions. *Optik* **73**, 119–122.

van Heel, M. & Keegstra, W. (1981). IMAGIC: a fast, flexible and friendly image analysis software system. *Ultramicroscopy* **7**, 113–130.

van Heel, M. & Stöffler-Meilicke, M. (1985). The characteristic views of *E. coli* and *B. stearothermophilus* 30S ribosomal subunits in the electron microscope, *EMJO J.* **4**, 2389–2395.

van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. struct. Biol.* **116**, 17–24.

van Heel, M., Keegstra, W., Schutter, W. & van Bruggen, E. F. J. (1982). Arthropod hemocyanin structures studied by image analysis. In: *Life Chemistry Reports*, *Suppl. 1*, '*The Structure and Function of Invertebrate Respiratory Proteins*', EMBO workshop, Leeds 1982, ed. E. J. Wood, pp. 69–73.

van Heel, M., Orlova, E. V., Harauz, G., Stark, H., Dube, P., Zemlin, F. & Schatz, M. (1997). Angular reconstitution: high resolution structures from un-crystallized macromolecules. *Scanning Microscopy* **11**, 195–210.

van Heel, M., Schatz, M. & Orlova, E. V. (1992a). Correlation functions revisited. *Ultramicroscopy* **46**, 304–316.

van Heel, M., Winkler, H. P., Orlova, E. V. & Schatz, M. (1992b). Structure analysis of ice-embedded single particles. *Scanning Microsc.* Suppl. **6**, 23–42.

Vinogradov, S. N., Sharma, P. K. & Walz, D. A. (1991). Iron and heme contents of the extracellular hemoglobin and chlorocruorins of annelids. *Comp. Biochem. Physiol.* **98**b, 187–194.

Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P. & Engel, A. (1997). The three-dimensional structure of aquaporin-1. *Nature* **387**, 624–627

Ward, J. H. (1982). Hierarchical grouping to optimize an objective function. *J. Am. statist. Assoc.* **58**, 236–244.

Wynne, S. A., Crowther, R. A. & Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol. Cell* **3**, 771–780.

Yonath, A. & Berkovitch-Yellin, Z. (1993). Hollows, voids, gaps and tunnels in the ribosome. *Curr. Opin. struct. Biol.* **3**, 175–181.

Yonath, A., Leonard, K. R. & Wittmann, H. G. (1987). A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science* **23**, 813–816.

Zemlin, F., Beckmann, E., Charlé, K.-P., Schatz, M. & Schlögl, R. (1999). Solid nitrogen: electron microscopy and beam damage quantification at 4 K. *Ultramicroscopy* **80**, 153–161.

Zemlin, F., Beckmann, E. & van der Mast, K. D. (1996). A 200 kV electron microscope with Schottky field emitter and a helium-cooled superconducting objective lens. *Ultramicroscopy* **3**, 227–238.

Zhou, Z. H., Chiu, W., Haskell, K., Spears, H. Jr, Jakana, J., Rixon, F. J. & Scott, L. R. (1998). Refinement of herpesvirus B-capsid structure on parallel supercomputers. *Biophys. J.* **74**, 576–588.