Scanning Microscopy

Volume 1992
Number 6 *Signal and Image Processing in Microscopy and Microanalysis*

Article 3

1992

Structure Analysis of Ice-Embedded Single Particles

M. van Heel
Fritz Haber Institute of the Max Planck Society, Germany

H. Winkler
Duke University

E. Orlova
Academy of Sciences, Russia

M. Schatz
Image Science Software GmbH, Germany

Scanning Microscopy Supplement 6, 1992 (pages 23-42) 0892-953X/92$5.00+.00
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

STRUCTURE ANALYSIS OF ICE-EMBEDDED SINGLE PARTICLES

M. van Heel*, H. Winkler[1], E. Orlova[2] and M. Schatz[3]

Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany
[1]Current address: Dept. Cell Biology, Duke Univ. Med. Center, Durham NC, USA
[2]Permanent address: Inst. Crystallography of the Acad. of Sciences, Leninsky Pr. 59, Moscow, Russia
[3]Current address: Image Science Software GmbH, Mecklenburgische Str. 27, D-14197 Berlin, Germany

Abstract

The conventional negative-stain preparation method for electron microscopy, in which biological macromolecules are contrasted using heavy metal salts (such as uranyl-acetate), is a simple and fast technique which has helped visualize hundreds of different molecular structures. Computer analysis of such negatively stained images of individual (i.e., non-crystalline) macromolecules using statistical pattern-recognition techniques has revealed considerable new structural information. Negative staining, however, has some disadvantages: the specimens are often severely flattened (as much as 25%-75%), they often exhibit strong preferential attachment of the molecules to the supporting carbon foil, and the molecular images may be difficult to interpret due to the relatively complex nature of the interaction between molecules and stain. Embedding biological macromolecules in a layer of vitreous ice (actually: "vitreous water") represents an attractive alternative preparation method which mimics the natural environment of these molecules. The processing of ice-images often requires special computational approaches such as: multivariate statistical classification of aligned images or of "invariant functions" derived from the unaligned images; alignment of images belonging to a specific class of images, determination of the spatial orientations of the projection images relative to each other ("angular reconstitution"). In this paper, we discuss our own overall single-particle structure analysis approach and highlight some new methodological developments in this context.

Key Words: Multivariate statistical analysis, correlation functions, three-dimensional reconstruction, protein structure, vitreous-ice embedding.

*Address for correspondence:
Marin van Heel
Fritz-Haber-Institut, Abt. EM / ERB
Faradayweg 4-6, D-14195 Berlin Germany
Telephone number: (49-30) 8305 380
FAX number: (49-30) 8305 509

Introduction

Electron microscopy represents a very direct method for obtaining structural information on biological macromolecules. "Direct" here means that we can IMAGE the macromolecules in the microscope and are not restricted to recording only diffraction patterns of the object of interest, as is the case with X-ray crystallography (cf. [Blundell and Johnson, 1976]). In other words, there is no "phase problem" in electron microscopy: by Fourier transforming the electron image, we not only have the diffraction amplitudes available but we also have the corresponding phases which are lost in X-ray crystallography. In spite of the substantial number of other problems that exist in electron microscopy such as the strong radiation sensitivity of biological material, electron crystallography, based on images of two-dimensional protein crystals, has now advanced so far that its results slowly start to match the quasi-atomic resolution results of X-ray crystallography (Henderson et al., 1986, 1990; Kühlbrandt and Neng Wang, 1991; Jap et al., 1990; Saß et al., 1989).

In electron microscopy (EM), as in X-ray crystallography, one attempts to extract structural information from a very large number of copies of the biological macromolecule being studied, such that each individual molecule suffers minimally from the interaction with the hard radiation with which one probes the sample (principle of "shared suffering"). In EM we produce many (extremely) noisy images of different individual molecules and we then apply some form of averaging (c.f. [Misell, 1978]) to obtain results with good signal-to-noise ratio (SNR). This averaging principle is the basic idea behind both the electron crystallographic approach and the single-particle approach in electron microscopy. Molecules that are not crystallized in two dimensions (2D) have, in principle, five degrees of freedom (translations in x and y directions, plus three Euler-angle rotations) more than molecules that are conforming to a given crystallographic environment. Consequently, sets of individual molecular images

("single particles") are more difficult to process than electron images of 2D-crystals of macromolecules.

One of the major breakthroughs in the analysis of 2D crystals which made it possible to reach quasi-atomic resolution was undoubtedly the application of correlation alignment (cf. [Henderson et al., 1986]) techniques to overcome the imperfections that even the best images of 2D-crystals possess. These correlation alignment ("correlation averaging") techniques had originally been developed (cf. [Saxton and Frank, 1977]) for the analysis of individual molecular images where they were required to bring the different molecular images into (x-y) register prior to averaging ("building a crystal" in the computer). "Single-molecule electron crystallography", however, is certainly not only justified by the spin-off it has generated to the more classical analysis of 2D protein crystals. Many macromolecules cannot be crystallized at all or only after investing a large and an unpredictable amount of time needed to convince the molecules to behave in an orderly way. Moreover, many interesting biological questions concern, for example, the interactions between molecules (e.g. monoclonal antibodies), flexibilities within a molecule, dissociation and reassociation of oligomeric assemblies. For solving such problems, crystallization of the molecules may not even be desirable. The resolutions so far reached with the single particle approach certainly do not yet match those reached by the crystallographic approach, but there actually is no specific reason why the single-particle approach should not also reach a resolution sufficiently high to follow the amino-acid chain in a protein. We may still be far from reaching such quasi-atomic resolution with single particles, but, on our way to this long-term goal, many important questions in structural biology on the 1-4 nm level are likely to be clarified.

Single Particles in Vitreous Ice: Introduction

The first two obvious problems related to the analysis of non-crystallized or individual macromolecules concern the orientation of the particles in the images relative to each other. First, let us assume that two individual molecules have exactly the orientation relative to the support film, like two identical chairs standing upright on the floor of the same kitchen. The microscopical images of the molecules then correspond to projection images of the two molecules onto the plane of the supporting surface, which, translated to our kitchen-floor model, corresponds to "X-ray" projections of the chairs onto the kitchen floor. It is obvious that the two projections are identical (apart from noise), but that they are rotated (one angular degree of freedom) and shifted (two translational degrees of freedom: x-y shifts) relative to each other. To average such "identical" projection images in a data set (for noise reduction) we must thus first align these images relative to each other within the projection plane. Bringing two such molecular images into register [Frank et al., 1978; Steinkilberg and Schramm, 1980; van Heel and Stöffler-Meilicke, 1985] is an operation we refer to as "planar alignment" or just "alignment".

The other type of orientational problem associated with the analysis of single molecules concerns the two remaining degrees of freedom of (Euler) rotation of the particle relative to the plane of the support film. In our kitchen-floor model, this corresponds to the positions that the (identical) chairs can assume after a substantial family dispute. Chairs can lie on their sides, fronts, backs, i.e., in many orientations other than their "preferred" upright orientation. The different projections of the chairs in these different orientations onto the plane of the kitchen floor are fundamentally different and cannot be brought into register by simple planar alignment. Such different types of "molecular" projections ("views"), resulting from different rotational orientations of the molecules relative to the plane of the support film, make that such a mixed set of molecular images needs to be treated with care. Averaging of noisy molecular images for the purpose of improving the SNR only makes sense if we know that the subpopulation of a (mixed) total population of images which we choose to average belong to the same group (cf. [van Heel, 1984, 1989]) since we otherwise risk averaging "cows" and "horses". It was for sorting out such mixed populations of images that Multivariate Statistical Analysis ("MSA") techniques were first introduced to electron microscopy [van Heel and Frank, 1981]. After an initial collaboration on MSA matters between Frank and van Heel, the lines of research of the Albany and the Berlin group have diverged.

The research of the Berlin group ("we") is directed towards first carrying out exhaustive searches to find all the characteristic views in a data set, and then using these views to reconstruct the molecular structure in three dimensions. The basic idea is that one can exploit the natural "random tilt-series" formed by the different views of the molecule on the support film to reconstruct the molecule in three dimensions without actually tilting the specimen holder. The experimental and theoretical advantages of this approach are considerable: each individual molecule is illuminated only once, there are no problems with defocus differences within a single micrograph, flattening of the specimen onto the support film does not (to a first approximation) affect the results, and there is no "missing cone" if the molecules really exhibit almost

random orientations. The first priority was to develop MSA eigenvector and classification algorithms [van Heel 1984, 1989] that allowed the analysis of a very large number (of the order of ten of thousands of images) of molecular images simultaneously so as to be able to find even very rare views with sufficient statistical significance. Almost equally important for sorting out our complex molecular image sets are different improved alignment schemes such as our multi-reference alignment approach, our reference-free invariant classification scheme [Schatz and van Heel, 1990, 1992], and our new reference-free "alignment by classification" approach [Dube et al., 1993], which will all be discussed in more detail below. The mathematical basis of the correlation alignment approach, the cross-correlation function itself, was recently reinvestigated and improved. Having found the different characteristic views of a molecule, a method was needed to find a-posteriori the relative orientations between the various molecular projections. This technique of "angular reconstitution" is, in part, a generalization of the earlier common-lines approach [Crowther, 1971]. Our recent improvements of the angular reconstitution approach are discussed below. The projection directions found with this technique can subsequently be used to reconstruct the molecule in three dimensions from its projections using the exact filter algorithm [Harauz and van Heel, 1986; Radermacher et al., 1986].

The Albany group approach followed a different strategy, which is aimed at explicitly exploiting preferred orientations of the molecules on the support film. Their elegant technique of "random conical tilt" [Radermacher et al., 1987a; Radermacher, 1988], in particular, exploits the random (planar) rotation orientation of molecules lying in one single preferred orientation relative to the support film. Upon tilting, the specimen holder, each of these different rotational orientations of the molecule convert into really different projections through the structure which can subsequently be used to reconstruct the molecule in three dimensions. In terms of our kitchen-floor model, tilting the kitchen floor relative to the projecting X-ray beam will convert the chair projections on the floor (originally all identical) into all kinds of different projections of the chairs, each of which carries different information about the projected object. This information can be used for reconstructing the original object in three dimensions. More recently, the Albany group has started to first perform an automatic search (by classification, see above) to find different predominant views in the data and then to use these subgroups of preferred orientations to perform independent 3D reconstructions which can later be combined into a single merged 3D reconstruction (cf. [Penczek et al.,1992]) (provided there is not too much flattening, see below).

Both approaches have their strong and their weak points. Although we originally had hoped that molecules prepared in negative stain would show the molecules in a large number of different views, the more systems we (and other groups) analyzed, the clearer it became that the negative stain technique is associated with relatively stable preferred orientations (cf. [van Heel and Stöffler-Meilicke, 1985]). A low number of different projections restricts the resolution attainable in the subsequent three dimensional reconstruction [DeRosier and Klug, 1968; Harauz and van Heel, 1986] , and this was one of the reasons for us to concentrate our research on ice-embedded specimens [Adrian et al., 1984; Stewart and Vigers, 1986; Vogel et al., 1986; Schatz et al., 1990a]. In particular, we normally use specimens without a supporting carbon film to avoid preferential attachment of the molecules onto the carbon surface. Preferred orientations of the macromolecules with respect to the ice surface, however, remain a concern, and we envisage imaging some of our ice embedded specimens using small tilts of only 20°-30° to surmount this problem (without ever taking a second "0°" micrograph). Our progress with the analysis of ice-embedded specimens was also somewhat stalled due to some remaining methodological problems, which we discuss in this paper.

The Albany approach is based on the analysis of one strictly preferential orientation of the specimen on the support film. Two electron images are made (which can already be too destructive for ice-embedded specimens), one with, say 50° tilt, and one without any tilt, i.e., 0° tilt. The rotation orientations of the different molecules (all identical views) in the 0° tilt are used to calculate the Euler projection angles from which each molecule is seen in the 50° tilted image; these Euler angles are subsequently used for the 3D reconstruction. The approach was already applied successfully to a number of structures including the 50S E.coli ribosomal subunit [Radermacher et al., 1987a, 1987b] and the 70S E.coli ribosome [Wagenknecht et al., 1989]. The first problem is that not all molecules show a single preferred orientation or that some perceived characteristic view such as the E.coli 50S ribosomal subunit can actually consist of two closely related views [Harauz et al.,1988; van Heel, 1989] which cannot be distinguished visually [Radermacher et al., 1986, 1987a, 1987b]. A practical problem with the approach is that only a narrow band of images around the tilt axis (homogeneous defocus conditions) can be used such that one is limited to using only a relatively small number of molecular images (see, however, [Zemlin, 1989; De Jong and

Typke, 1990]. One of the most serious problems with the negative stain preparation technique seems to be the extreme amount of flattening (up to even around 75%, cf. Kellenberger et al., 1982]) onto the support film that the molecules appear to suffer, making such reconstructions often unnaturally flat. Thus reconstructions based on one specific view of the molecule can often not be related to a reconstruction based on a different view of the same molecule [Carazo and Frank, 1989]. A comparative study of this effect, based on different reconstructions of the giant hemoglobin of the common earthworm, is forthcoming [Z. Cejka, personal communication].

Other approaches to the analysis of single particles have been published. The mathematically elaborate techniques of Kam [Kam, 1980], and of Vogel and Provencher [Vogel et al., 1986; Provencher and Vogel, 1988] are not discussed here since there have been no further developments with these approaches. One of the first approaches to 3D-reconstructions of single particle, conventional tomography applied to a single molecule (cf. [Öttl et al., 1983]), is not applicable to ice-embedded specimens because of the high electron dose needed to register the many different projection images of the same individual molecule. The two main approaches discussed in this section are applicable to ice and require only one image (Berlin) or only two images (Albany) of the same specimen area. The single-particle approach that actually led to the first publication of a 3D reconstruction of ice-embedded single molecules was that of Vogel and Provencher [Vogel et al., 1986] exploiting the 60-fold symmetry of the icosahedral virus. All other ice-embedded viral reconstructions were done using the common-lines approach [Crowther, 1971], the first of which was published by Fuller [Fuller, 1987]. With the common-lines technique one explicitly exploits the 60-fold symmetry of the icosahedral viruses to find the 3D orientation of each individual virus image prior to any further analysis. For lower molecular symmetries there is usually too little information in a single molecular image of the ice-embedded molecule to find the orientation of the noisy particle. In the Berlin group approach the averaging is performed over all projections belonging to one view prior to any 3D reconstruction; the 3D reconstruction is thus performed using largely noise-free projections. In the Albany-group approach, averaging takes place when all individual noisy projections are inserted into the 3D reconstruction volume.

Microscopy, Densitometry and Computing Environment

With the currently successful technique of 2D electron crystallography, the structural information is extracted from tens of thousands of unit cells for each of the different tilted views processed (cf. [Henderson et al., 1990]). Since we cannot expect ever to get anything for nothing, we will ultimately need about the same total number of molecular images for non-crystallized specimens to achieve comparable results. Current microscopical and image processing technology, however, restricts the resolution attainable with single particles so that we aim at collecting data sets containing 5,000-10,000 molecular images, distributed over all possile random tilt angles. Assuming that each micrograph contains about 100 usable molecular images (images not overlapping with those of neighboring molecules), this implies that typically about 100 good micrographs are needed for one data set.

We do not want to go into the details of the art of sample preparation and cryo-microscopy with vitreous-ice embedded specimens (cf. [Adrian et al., 1984; Dubochet et al., 1988]). What we want to emphasize is that computer-controlled microscopy will represent an important improvement over the techniques that are currently in routine use for collecting the microscopical data. Low-dose techniques used for minimizing specimen irradiation can best be implemented on a computer controlled microscope. Reproducibility optimizing the instrumental parameters (defocus, astigmatism, instrument alignment) can also best be done on such systems (see the papers in this issue that specifically deal with computer-controlled microscopes). Also, computer-driven spot-scan illumination [Downing, 1991] will be imperative in achieving high-resolution results with ice-embedded single molecules in the way these techniques are essential for electron crystallography. Automatic photography of the full surface of a good area of the specimen over many plates will facilitate the data collection. If low-tilt images (20°-30°) are necessary, dynamic focussing [Zemlin, 1989; De Jong and Typke, 1990] will substantially enlarge the areas in the micrographs that can be used for further processing.

Densitometry of hundreds of micrographs is obviously a laborious task. Conventional flat-bed densitometers are high-precision instruments which move mechanically to each point of the negative so as to digitize the micrographs point-by-point, and they are thus too slow to routinely digitize large numbers of images to a resolution of around 2400x3600 pixels per micrograph resulting in some 10 Mbyte of data each. CCD (charged coupled device) densitometers which (electronically) scan

one image line at a time and then (mechanically) move to the next line are much faster and thus much more appropriate for this large task. Our CCD line scanner (produced in 1984 by Datacopy Corporation, Mountain View, California, controlled by an IBM personal computer coupled to our workstation computers through ethernet) is slowly aging and, with its resolution of 1728 pixels per line, no longer capable of matching our higher-resolution requirements (the device is also no longer in production). Although this type of CCD line densitometer is still commercially available with a resolution of 4096 pixels per line, we have decided to build a new densitometer based on a standard video camera with a 512x512 pixels CCD chip. The camera digitizes a small area of the negative at a time and the micrograph is subsequently mechanically moved (an x-y stepper-motor controlled microscope table) to the next part of the negative. The many 512x512 subfields are subsequently computationally inserted into a large (say, 8000x8000) image on the workstation computer which controls the whole set-up. With this approach, we expect to obtain a moderately priced, fast, and high-resolution microdensitometer.

For the analysis of such large microscopical data sets, the computer system to be used for the data-processing has to fulfill a number of requirements. For example, a raw data set of 100 negatives requires about 1 Gigabyte of data storage, hence mass-storage devices are of primary importance. One or two erasable optical disks as well as various Gigabytes of conventional magnetic disk are a prerequisite for handling such data sets. Our current computer system consists of a cluster of some 5-7 MicroVAX 3100 (Digital Equipment Corporation; running under the VMS operating system) and a FORCE SPARC-1 station (under SUN-OS, a UNIX V-4 operating system) which will drive our new densitometer. A total of about 8 Gigabytes of magnetic diskspace is available in the net.

All image processing, including the densitometry, is performed in the context of the IMAGIC-5 image processing system [van Heel and Keegstra, 1981]. This general-purpose software package, which was developed in an electron microscopical research environment, was, after being upgraded to function in an X Windows environment, commercialized in 1990 by the Image Science Software company in Berlin [Image Science, 1991]. The properties of the IMAGIC-5 image processing system are important for the type of processing of very large numbers of individual images we are describing. For example, without explicitly formulating a "DO" loop by the user, all programs will automatically process all images present in a data-set (a multi-image IMAGIC file) unless the program is told otherwise. Some other image-processing systems maintain one file per image in the data-set which can lead to substantial organizational and speed problems if 10,000 files or more are required to reside in a given directory. It is not our intention to compare the properties of the different available image processing systems in detail here, although the issue is of primary importance.

All distinct particles that are not overlapping or in close contact with other particles are selected interactively from the digitized micrographs using the mouse to point at the particles in the digitized images as displayed in an "X" window of a workstation-type computer. This system of selecting the thousands of individual molecular images normally takes a few days. Only the coordinates of the molecules within the raw images are stored which are subsequently used to extract the desired particles into small nxn images; n ranges typically from 48 to 100 pixels, depending on the size of the object and the desired resolution. Automatic particle selection algorithms have been proposed [van Heel, 1982; Frank and Wagenknecht, 1984], but these have not yet been put into routine use.

The many single molecular images selected are stored in a single IMAGIC file and then pretreated by band-pass filtering to suppress the very low and very high spatial frequencies (high-frequency noise). In particular, the very low spatial frequencies (often associated with unwanted information such as a density-gradient running over the image) have considerable power and are thus often very disturbing for subsequent alignments and multivariate statistical analysis. The frequency limits of band-pass filtering vary somewhat from experiment to experiment but they are normally coupled to the size of the molecule (low-frequency) and expected resolution (high-frequency). Moreover, while downweighting the low-frequency components, these are not completely removed but, rather, set to a fraction of their original values (typically around 0.1%) so that they can be restored at a later stage by inverse filtering. After filtering, the particles are surrounded by a circular mask to cut away unnecessary background; within this mask the image statistics are standardized to zero average density and to a standard deviation (arbitrary) of 10 [van Heel and Stöffler-Meilicke, 1985; Harauz et al., 1988].

Planar Alignment

Planar alignment (see above) of individual molecular images is a prerequisite whenever the single particles need to be averaged to improve the SNR. The issue has been important since the very beginning of single particle analysis [Crowther and Amos, 1971; Frank et al., 1978]. Even if in a

specimen the individual molecules exhibit just one single preferred orientation relative to the support film, we still need to compensate for the rotational and the two translational degrees of freedom that the molecules have. To achieve planar alignment, we use alignment algorithms based on the use of correlation functions [Frank et al., 1978; Saxton and Frank, 1977; Steinkilberg and Schramm, 1980; van Heel et al., 1992], to bring all images in a data set into optimal register with a carefully chosen (see below) reference image.

Correlation functions are relatively simple functions which measure how well two images match when we overlap them; the correlation alignment procedure is (largely) equivalent to manually overlapping two negatives of the same object over a light box. The computationally cheapest known way to calculate exhaustive (=testing all possible relative shifts) correlation functions is through Fourier space thanks to the efficiency of the Fast Fourier Transform (FFT, cf. [Cooley and Tuckey, 1965; Singleton, 1969]). The cross-correlation function (CCF) between two images is calculated as follows: both the reference image and the image to be aligned are Fourier-transformed, then these transforms are (complex conjugate) multiplied point by point, and the inverse Fourier transform of the resulting function is the desired CCF. The rotational correlation function (RCF) between image and reference image is calculated similarly after first converting both images to a polar coordinate system.

The most commonly used alignment algorithm [Steinkilberg and Schramm, 1980] is based on iteratively repeating translational and rotational correlation alignments. The technique is very sensitive and can cope with very noisy images; we have successfully performed millions of alignments using this procedure. In this alignment procedure, one implicitly assumes that the molecules to be aligned belong to the same view, i.e. are in the same preferred orientation relative to the support film. Although this assumption may be justified in some cases, such as with the rather flat and symmetric (622 pointgroup) glutamine synthetase [Frank et al., 1978; Kunath et al., 1984], it is normally not justified even with flat structures like the giant hemoglobin of the common earthworm (another 622 pointgroup structure) [Boekema and van Heel, 1989]. The most important factor affecting the results of this and of other correlation-alignment approaches is the choice of the reference image.

For a data set from a macromolecule of initially unknown structure, it is not clear which projection view is most suitable as an initial reference image, and we are faced with a bootstrapping problem: we need the end result of the analysis, i.e. the noise-free characteristic views, before we even start the procedure. This bootstrapping problem can be solved iteratively with a multi-reference alignment procedure [van Heel and Stöffler-Meilicke, 1985]. In this approach, we first choose a set of "n" reference images from the raw data set. Then, each molecular image is aligned "n" times with respect to the "n" references (and normally also another "n" times relative to the mirror versions of the reference images). From these "2n" alignment parameters, those associated with highest normalized CCC (cross-correlation coefficient, i.e., the best fit) are the ones finally used to rotate and shift each particular image. The reference images are themselves treated as part of the data set and are therefore also aligned relative to each other. In the multi-reference approach, one iterates between rounds of alignments and MSA classifications (see below) which provide new classes to be used as reference images for the iteration of the procedures. This overall approach can be seen as an algorithm aimed at reaching an optimal global alignment for an entire heterogeneous data set [van Heel et al., 1982].

However, the very use of one or more reference images to align a data set is associated with the danger of biasing the entire data set towards the properties of the reference(s) [Boekema et al., 1986]. In a simple computer experiment we once have aligned a set of 256 computer-generated random noise images to an image of Einstein [Schatz, 1992]. The total sum of the aligned random-noise images unmistakably showed the face of this famous scientist. Alignment thus can pull an entire data set towards the reference, a bias problem which can become serious when the molecules to be investigated are small or when the images are very noisy (ice-embedded macromolecules). We have thus initiated a search for new "reference-free" analysis procedures which will be discussed below.

MSA Techniques

Multivariate statistical analysis (MSA) techniques can be seen as perfect averaging tools in electron microscopy (and in most other field of research where one deals with complex and noisy data sets). In spite of the fact that the interesting information is often entirely buried in noise, we can, using such techniques, reproducibly discriminate between subsets of images differing in astonishingly subtle details. MSA techniques were first introduced to electron microscopy in 1980/1981 [van Heel and Frank, 1980; van Heel and Frank 1981] and have since been used in hundreds of investigations. The MSA approach can be subdivided into two major sections: first, the eigenvector eigen-

value data-compression, and second an automatic unsupervised classification operating on the compressed data set (cf. [Lebart et al., 1977, 1984]).

Eigenvector Data Compression: Conceptually, in the MSA approach we first picture each of our measurements as a point in a multidimensional space with as many dimensions as there are data points in the measurements. Each image of, say, 64x64 pixels can thus be represented as a point in an 64x64=4096 dimensional ("hyper") space, and the entire set of images thus converts to a "cloud" of points in this space. Similarities between images translate to proximities between the corresponding points in the data cloud. The basic idea behind the data-reduction procedure is to then rotate the coordinate system of the multidimensional space such that the new coordinate system follows the shape of the data cloud as closely as possible. Systematic inter-image differences translate in the hyper-space representation into an elongation of the data cloud in a given hyper-space direction. In the rotated coordinate system, one tries to optimally (in a least-squares sense) place the unit vectors in the different elongation directions of the data cloud. The unit vectors of the rotated coordinate system (determined through an "eigenvector eigenvalue" algorithm) are themselves points in the hyperspace and hence have the character of images like all the points in the data cloud. The unit vectors (eigenvectors) of the rotated coordinate system are thus often called eigenimages (cf. [van Heel, 1989]).

The first eigenimage represents the direction of greatest variance in the data set. This direction (depending on the metric used, see below) often corresponds to the direction from the origin of the coordinate system to the center of the cloud. The second eigenimage ("orthogonal" to the first) is actually the first direction describing important differences between images. The third eigenimage describes the direction of the largest remaining differences between the images (orthogonal to eigenvector #1 and #2), etc. The cloud of images can now be described with respect to this new coordinate system. One then normally works with the coordinates of each image with respect to only the most significant ("principal") data cloud components (typically 12 to 24). By disregarding the higher-order components, the images can be considered as points in a much lower-dimensional (for example 12 to 24, rather than 4096) space. One thus achieves a large reduction in the amount of data to be analyzed as well as a very significant reduction of the effects of random noise which are typically more associated with the higher eigenvectors of the system. The data reduction is very important for the subsequent "exhaustive" comparison of all images with each other for the purpose of classification, which can be computationally very intensive. Moreover, the availability of the eigenimages allows for an often very illustrative visual inspection by the investigator of the main trends or "principal components" of the data set.

The choice the metric or the distance measure: Principal components analysis (PCA: see any book on multivariate statistics, for example, [Lebart et al., 1977, 1984]) is undoubtedly the best known and most proliferate MSA technique. There are, however, some problems associated with PCA and the conventional Euclidian distances on which it is based. If signal (measurement, image) "B" is equal to two times signal "A", then signal B will lie on the hyper-space line connecting the origin "O" to point A, at twice the distance OA. Although the two signals differ only in a multiplicative factor (and are thus essentially identical) they are located far from each in the cloud (and will thus be classified differently), and, due to the least-squares principle behind the MSA techniques, signal B will contribute four times more to the determination of the eigenvectors than will signal A.

Correspondence analysis ("CA") stemming from or at least made popular by the French school of statistician around Prof. Benzécri [Benzécri, 1977, 1980a,b], was designed to circumvent this problem of heterogeneous statistics of the measurements in the data set. By using chi-square distances, in which the measurements are divided by the total sums of each measurement (like the total density in an image), signal A is made to coincide with signal B in hyper-space. Moreover, since in CA the signal B will have twice the "mass" of signal A associated with it, signal B will also ONLY contribute twice as much (in terms of variance) to the determination of the eigenvectors as will signal A. During the first years of the application of MSA in electron microscopy, CA was the dominating technique.

The correspondence analysis Chi-square metric, however, when applied to signal-processing is also associated with a fundamental problem. The technique of dividing by the total sum of a measurement is designed for the analysis of (positive) histogram data. In contrast to the rule in the social sciences, in signal and image-processing the measurements need not be positive. Signals often have, or are normalized to, zero average density. We thus have introduced a new MSA variant based on "modulation distances", a distance measure in which the signals are divided by the standard deviation of each signal [Borland and van Heel, 1990]. The MSA variant with modulation distances we call "modulation analysis", and this MSA technique shares the generally favorable properties of CA, yet, as is the case in PCA, allows for the processing

of zero-average-density signals. This MSA technique has already been in routine use for a number of years and has, in our research, entirely superseded the CA approach. The practice of thresholding image data to positive values only, as is necessary for applying CA to signals that are not positive everywhere, is not always justifiable and - at the very least - leads to wasting of half of the signal for zero-average-density measurements. Another practice, in which the negative values are rendered positive by adding a constant to the data also has far-reaching consequences. The strong negative densities will end up as small positive densities and will have very little contribution to the total variance of the data set, whereas the large positive values will become very large positive values with disproportionallly strong contribution to the total variance. Although there are thus significant theoretical and practical differences between MSA analyses using different metrics, the differences between these approaches are miniscule compared to the difference between using MSA techniques or not using them at all.

Classification: The data compression achieved by projecting the data into the space spanned by the predominant eigenimages facilitates the grouping (classification) of images. In the early days of MSA in EM, the classification of the images was performed visually on "MSA-maps" or scatter diagrams showing the data-cloud projected onto two eigenvectors at a time. We now use automatic (unsupervised) classification schemes which are indispensable, particularly, when the relevant image information is spread over more than just two predominant eigenimages. Our best experience has been with hierarchical ascendant classification based (HAC) schemes (cf. [Lebart et al., 1984; van Heel, 1984; van Heel 1989; Borland and van Heel, 1990]. In this approach each class is first filled with one individual image, and then similar classes are merged together to form larger and larger classes until finally one class containing all the images is obtained; the history of the classification is stored in a "tree". As merging criterion, we use the Ward criterion [Ward, 1982], which is variance-oriented. Two classes are merged if the associated "added intra-class variance" of the particular merging operation is the lowest possible one at the given level of the HAC procedure.

The actual partitioning of the data set is performed after the decision to cut the classification tree at a given level (associated with a predetermined number of classes or "branches" of the tree). The number of desired classes depends entirely on the problem to be solved, and can range from just a few, for a data set of molecular images exhibiting just a few preferred orientations, up to a few hundred for a data set of ice-embedded molecules expected to show a full scale of random orientations. The goal of the classification process is to reach the optimal partition defined as that partition (for a given number of classes) for which the sum of all internal variances of the classes (intra-class variances) is minimal (cf. [Lebart et al., 1984; van Heel, 1984, 1989]).

The partitions obtained through the HAC procedure are, however, necessarily suboptimal. The fact that we take two classes (or images) together at a certain level of the HAC procedure binds the members of the two classes together forever. After the initial partitioning with the HAC scheme we therefore allow the individual members of the classes to migrate to neighboring classes using an iterative migration post-processor. If a particular image is found to fit better (in terms of its variance contribution to the class) in a different class after all, it is allowed to migrate to that class and the intra-class variances of the original and the new class are recalculated. The procedure is iterated until no further inter-class migrations take place and the procedure has thus stabilized. For further technical details about our preferred classification approaches, see [van Heel, 1989]. The aligned images in each of the classes obtained are summed together to give class-averages with an enhanced SNR, and which represent the desired "views" present in the data-set (if all has gone well). While summing the images belonging to one class, one has a last chance to reject outliers which contribute too much to the internal variance of the class [Unser et al., 1989; van Heel, 1989]. By the way, classifications can also be performed in the "conjugate" space [Borland and van Heel, 1990] in which every "pixel-vector" is depicted as a point in a conjugate hyper-space. The classification results here tell us which areas of the images behave similarly throughout the data set, or equivalently, which specific areas of the images are responsible for the inter-image differences.

Invariant Classification

As was discussed above, the choice of the reference image(s) to be used for alignment of the images in the data set can substantially influence, or even bias, the final results. This finding has triggered a search for analysis procedures which do not depend on the choice of individual references ("reference-free" procedures). The ultimate goal of the whole analysis is to obtain aligned sets of images from raw data. In our first reference-free experiments, however, our aim was first to find classes of similar images without having ever aligned the images of the raw data set. If we know

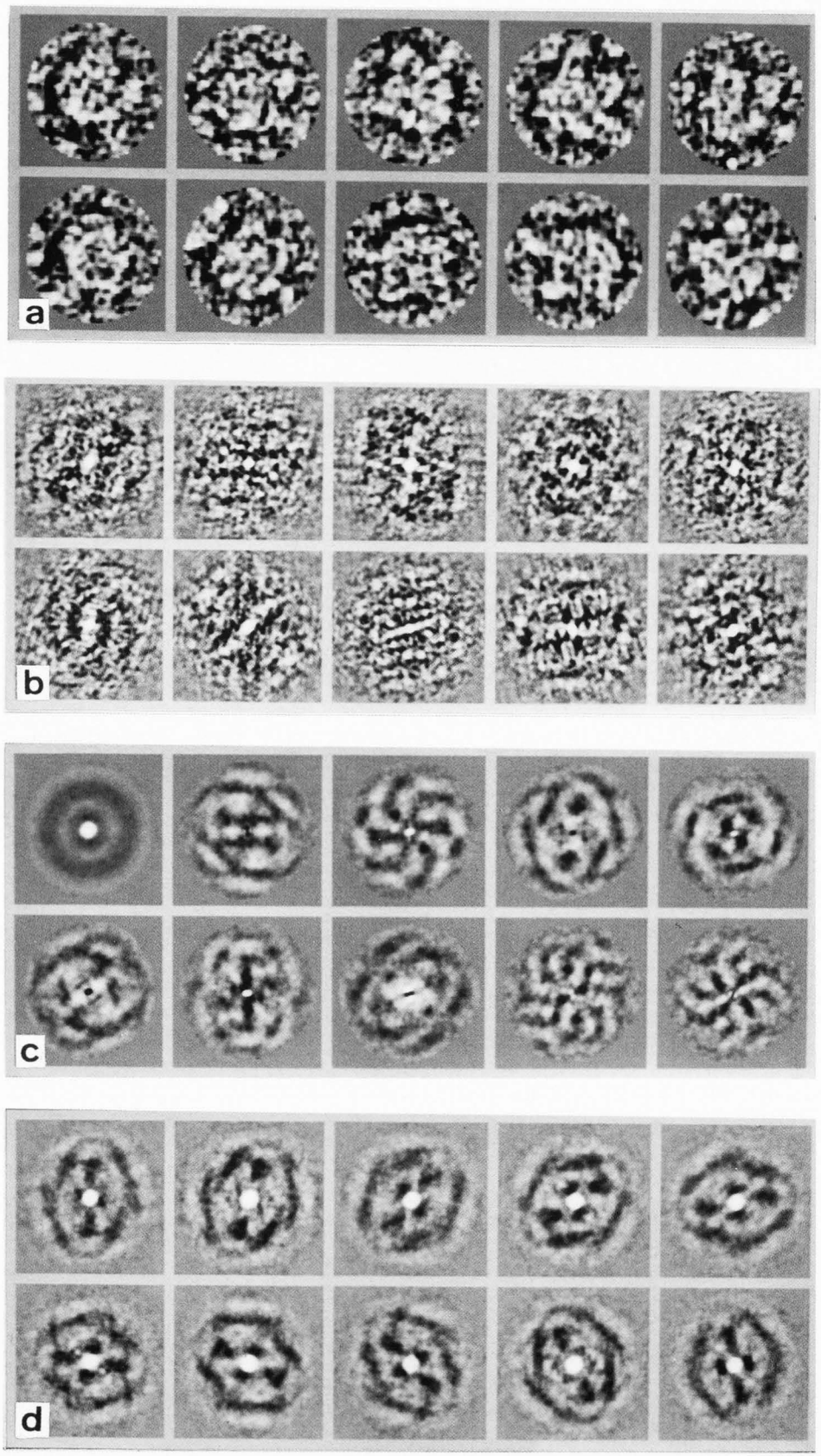


Figure 1. Alignment by classification. A set of 333 band-pass filtered molecular images (all crown views of 50S *E. coli* ribosomal subunits, only 10 shown in Fig. **1a**) are used as input to this feasibility study. After conversion to the translational-invariant SCF form (Fig. **1b**), the data set is compressed by eigenvector analysis (the first 10 "eigen-images" are shown in Fig. **1c**) and then partitioned into 10 classes by means of automatic classification. The resulting 10 SCF class averages, shown in Figure **1d**, are sorted by their different rotational orientations. The experiment demonstrates that an alignment-by-classification using real microscopical data is possible.

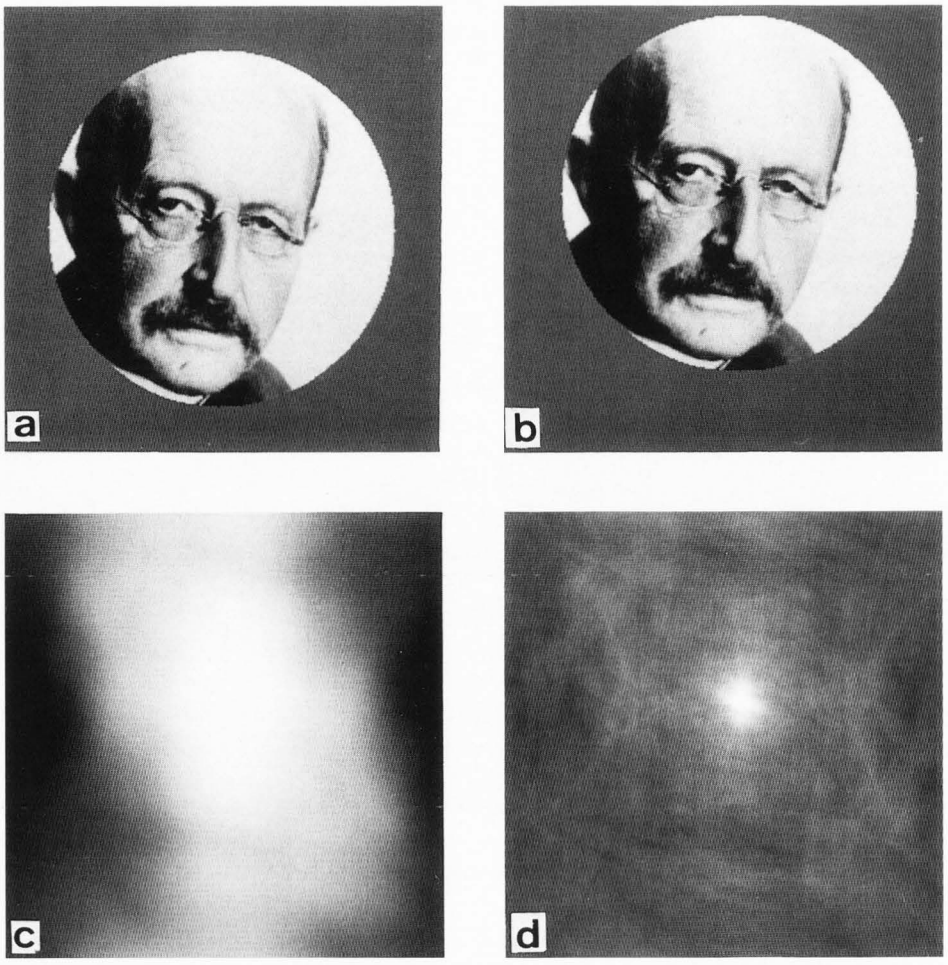
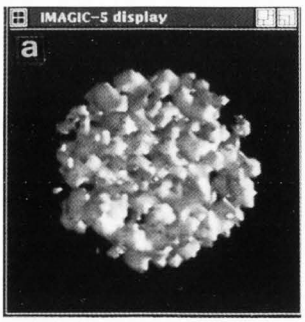


Figure 2. CCF versus MCF. The cross correlation function (Fig. **2c**) between two shifted images (Figures **2a**, **2b**) is dominated by the strong frequency components in the images, typically the low-frequency ones. This problem is circumvented in the "Mutual correlation function" (MCF). For details see text.

that given subset of our full population of image belongs to one single characteristic view, this knowledge can be exploited in subsequent intra-class alignment procedures [Schatz and van Heel, 1990, 1992]. For this purpose, we derive "invariant functions" from the raw images, and apply MSA to these secondary functions rather than to the images themselves. Depending on the given problem, the invariant functions can be chosen such that they are invariant to translational shifts and/or to rotations of the molecular images within their frames.

Invariant functions have played a significant role in EM for some time. The power spectrum, and its real-space equivalent the autocorrelation function (ACF), are both functions which are invariant to a lateral shift of the input image yet convey information about the image and its focus and astigmatism-correction conditions. One of the very first single-particle alignment algorithms ("ACF alignment", cf. [Frank et al., 1978, 1981; Frank, 1980]) exploits the translational invariance of the ACF by first performing a rotational alignment between the ACFs of two molecular images, and only then performing translational alignment between the rotated originals.

In our reference-free double correlation classification scheme [Schatz and van Heel, 1990], a (mixed) set of input images is first converted to a corresponding set of autocorrelation functions. The ACFs are then subjected to a second autocorrelation operation along rings (i.e. in the tangential direction in polar coordinates), resulting in a double autocorrelation function (DACF). The only time we ever applied the DACF approach was to model data, since we immediately realized there was a serious problem with ACFs in general, which became much more serious when cascading ACFs for the purpose of calculating the DACF. The problem is that for calculating the ACF, we square the Fourier components of the image prior to calculating the inverse transform (the ACF is thus a "squared correlation function"). As a consequence, the strong Fourier components (typically low-frequency ones) tend to entirely overwhelm the weaker components, often the fine high-frequency details we are really interested in. Instead of working with the ACFs and DACFs of the input images we thus work with the self-correlation function (SCF) which we define as being the inverse Fourier



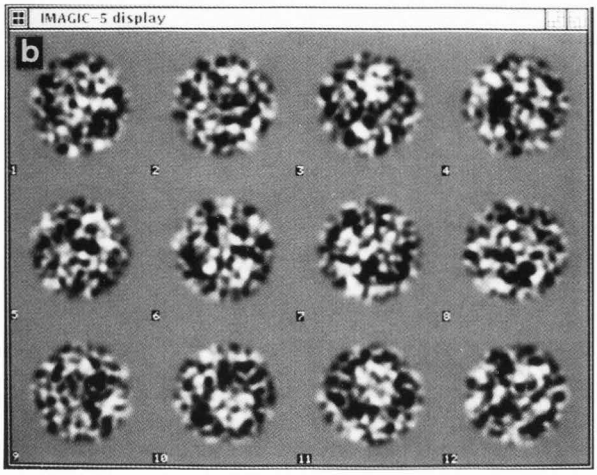
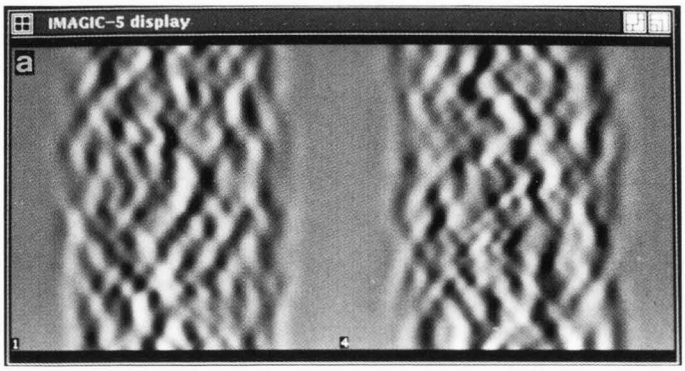


Figure 3. Angular reconstitution. A 3D model object (Fig. **3a**) is projected into a large number of different Euler directions leading to a set of 2D projections, of which 12 (numbered, from left to right, 1-4, 5-8, and 9-12, respectively, in top, middle, and bottom rows) are depicted in Fig. **3b**.



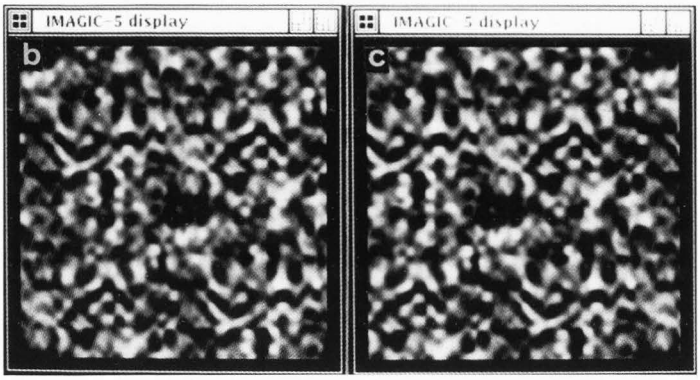


Figure 4. The sinecorr function. Each 2D projection image (Fig. 3b) is projected onto a single line (1D projection). The 1D projections in possible directions are mounted into a "sinogram". Two sinograms, calculated from projections #1 and #4, are shown in Figure **4a**. The search for the common line projection (common tilt-axis) between two 2D projections can best be performed by correlating all lines of the two corresponding sinogram images with each other in the sine-correlation function ("sinecorr", Fig. **4b**). A planar rotation (the α-angle) of one of the input images corresponds to a cyclical shift of the corresponding sinogram in the vertical direction. As a consequence, all sinecorr functions calculated from the rotated input image will be cyclically shifted (Fig. **4c**) in the horizontal or in the vertical direction.

transform of the amplitude spectrum of the image; the DSCF (double SCF) is defined analogously to the DACF yet replacing the ACF calculations by a corresponding SCF one [Schatz and van Heel, 1992].

Once a set of molecular images is converted into a set of DSCFs, we can apply data-compression and classifications to these secondary invariant functions rather than to the images themselves. Compact classes of DSCFs found in the procedure contain mainly DSCFs derived from original images which belong to one given type of projection through the molecule (one characteristic view). Theoretically, an infinite number of different input images may lead to identical output DSCFs due to the throwing away of the phase information in the SCF calculations (in the DSCF calculation one throws away phases even twice). In practice one must really resort to mathematical tricks to achieve identical DSCFs starting from different (artificial) images, and the non-uniqueness of the DSCF is thus not a real problem. A relatively conventional intra-class alignment procedure is then applied to the images associated with each DSCF class (the image closest to the center of the DSCF class is used as reference). This DSCF approach was successfully applied to images of ice-embedded hemoglobins of the common earthworm [Schatz et al., 1990b; Schatz and van Heel, 1992] to presort the different view present in the data set.

Alignment By Classification

A new idea, expressed in a recent paper [Dube et al., 1993], is to avoid the use of references by

performing "alignments by classification" rather than by correlation techniques. Within a classification one compares everything with everything and thus there cannot be a bias relative to any one particular image in the data set. In that work, a set of about 800 top-views of the oligomeric portal protein of bacteriophage SPP1 was first centered relative to a rotationally symmetric mask. The images were then subjected to MSA/classification without having undergone any rotational alignment relative to each other. The class averages resulting from the procedures clearly showed the different rotational orientations the molecules happen to have on the support film. The purpose of the SPP1 portal-protein study was primarily to find the symmetry of the structure by integrating information from a large number of molecular images simultaneously. Earlier symmetry analysis procedures [Crowther and Amos, 1971] operated on only one single molecular image at a time. The SPP1 averages showed an unexpected 13-fold rotational symmetry; portal proteins were thought to generically possess 12-fold symmetry. A reanalysis of the well-studied portal-protein assembly of Φ-29 bacteriophage [Carrascosa et al., 1982] using these new reference-free techniques again revealed a 13-fold symmetric structure [Dube et al., 1993] without any explicit rotational alignment of the images relative to each other.

We here want to emphasize that the alignment-by-classification approach can generally be used in microscopy, provided that a sufficient number of molecular images of a given view (in all possible planar rotational orientations) are present in the data set. The images used for our demonstration stem from an earlier study [Harauz et al., 1988], and are 333 original images belonging to one type of crown view of the E.coli 50S ribosomal subunit [Harauz et al., 1988] (two different types of crown views were found). Some of these original (band-pass filtered) images are shown in Fig. 1a. The images are then converted to SCF format (Fig. 1b) and subsequently subjected to eigenvector data-compression and automatic classification. The eigenimages of the analysis (Fig. 1c) are all centrosymmetric since the SCFs input-images are centrosymmetric. We have here requested 10 classes, and shown in Fig. 1d are all these SCF class-averages sorted by their rotational orientations. This simple experiment demonstrates that the technique is capable of automatically finding the different rotational orientations in this homogeneous data-set.

In the DSCF classification (see above) we lose the important information contained in the phases of the images twice. With the rotational alignment scheme described above, we avoid the information loss in the second rotational SCF calculation of the DSCF procedure. Having found the rotational orientation of the particles (however, with 180° ambiguity (cf. [Frank, 1980; Schatz and van Heel, 1990] due to the properties of the ACF/SCF), we can proceed with the rotationally aligned input images and then again apply classification to find the translational alignment differences between the images and to resolve the 180° ambiguity. This translational alignment by classification is expected to work best if the molecules have already been centered roughly using other techniques. Both phase losses occurring in the calculation of the translational and rotational invariant DSCFs can thus be regained provided that a sufficient number of molecular images is available for each of the characteristic views present in the data-set.

Mutual Correlation Functions

We have seen above that the ACF is a squared correlation function which can suffer severely from overweighting of strong frequency components in the data. The ACF of an image is the cross-correlation function (CCF) of the image with respect to itself. This fact already indicates that the CCF, which lies at the basis of all correlation alignment/averaging procedures, may itself suffer from the same type of shortcoming. Again, the problem is expected to be more serious when squared correlation functions are cascaded, such as is the case with the ACF alignment procedure (cf. [Frank et al., 1981]), in which one calculates the CCF along corresponding rings of ACFs derived from the images to be aligned.

There is more than one possible solution to the problem of finding a correlation function which relates to the SCF the way the CCF relates to the ACF. We have found it advantageous to use the Mutual Correlation Function (MCF) [van Heel et al., 1992], which is calculated very much the same way the CCF is calculated by complex-conjugate multiplication of the Fourier transform (FT) of the two images to be correlated. In the MCF calculation, however, the complex Fourier components of the images involved are downweighted by division by the square-root of the amplitude of the component, prior to the final reverse Fourier transform calculation. The difference between CCF and MCF, as important work-horse tools in alignment procedures, is demonstrated by a model experiment illustrated in Fig. 2. An image of Max Planck (Fig. 2a) is shifted translationally to produce Fig. 2b. In Fig. 2c, the CCF between original and shifted "Max Planck" is depicted. This image is obviously dominated by the low-frequency components in the input (unfiltered) images, although one can obtain the

correct shift parameters from the rather blurred highest peak in the CCF of these two noise-free input images. The MCF (Fig. 2d) between the input images, in contrast, shows a strong and well-defined peak at the position of perfect overlap between the images. It is obvious that if one uses the peak height of the CCF as a similarity measure between images, this similarity measure (cross correlation coefficient CCC) will also be largely dominated by the strong image frequency components (typically the low-frequency ones). The (normalized) peak-height of the MCF is generally a better inter-image similarity measure than the CCC [van Heel et al., 1992].

We have converted many of our alignment algorithms to function with the MCF instead of the CCF. The new algorithms, as expected, behave generally better than their predecessor [van Heel et al., 1992]. We thus expect the MCF to supersede the CCF for virtually all its applications in electron microscopy including correlation-averaging of 2D crystals. The MCF based algorithms do, however, exhibit a different frequency behavior than their predecessors, a fact which needs to be considered carefully. Whereas for CCF-based alignment algorithms the suppressing of low-frequency images components was often essential (cf. [van Heel and Stöffler-Meilicke, 1985]), this side of the spectrum has become much less bothersome with MCF alignments. At the same time, the high-spatial-frequency side of the image data (often consisting of mainly noise) which was of no real practical concern in CCF alignments, now sometimes really needs to be suppressed by low-pass filtering to obtain the best alignment results.

Angular Reconstitution

The class sums or characteristic views obtained by the alignment/MSA techniques discussed above provide us with distinct projections of the macromolecular complex under study, and the logical next step is to combine these into a 3D reconstruction. In most problems of 3D reconstruction from projections, e.g., computerized medical tomography, one knows from the experimental set-up the angular relationships between the projections from which the reconstruction is to be calculated. Here, however, we obtain the characteristic views of molecules without a-priori knowing how the different projection directions of the different views relate to each other. The method of "angular reconstitution" [van Heel, 1987; Vainshtein and Goncharov, 1986 (in russian)] allows for the a-posteriori determination of the relative angular orientations of the projections.

When a 3D object, like the test object depicted in Fig. 3a (a 3D "general phantom" with predetermined power-spectrum properties [Harauz and van Heel, 1986]), is projected into different directions, a number of 2D projections are generated (Fig. 3b) which correspond to the different views of a molecule obtained in the EM. Although two different 2D projections of such a 3D object may look very different, they always have at least (for asymmetric objects) one line projection in common. This common line projection corresponds to the direction of the tilt-axis along which we need to tilt the original 3D object to convert one of the 2D projection directions into the other. A line projection of a 2D view is calculated by projecting all density values in the 2D image onto a line in a given direction. One normally calculates all possible line-projections in all possible directions and mounts all these line-projections on top of each other in a "sinogram" [van Heel, 1987]. The sinograms for projections #1 and #4 (Fig. 3b) have been calculated as examples (Fig. 4a). The statement that two 2D projections have one of their line-projections in common implies that one of the lines in sinogram #1 is identical to one of the lines in sinogram #4. The search for the correct line-pair is performed by calculating the CCC between all lines of sinogram #1 and sinogram #4 and storing all the CCC-values in a two-dimensional images called a sinogram correlation function ("sinecorr", see Fig. 4b). (Our earlier negative comments about CCCs apply here too, and consequently we are modifying our SINECORR program to operate instead with new correlation coefficients.) The highest peak in the sinecorr function tells us in which direction in image #1 lies the common tilt-axis shared with image #4 and vice-versa. The orientation of the common tilt-axis (common line-projection) between images #1 and #4 fix the relative orientation of these two views in 3D space. With three different 2D projections (not related by a tilt around a single rotation axis) the relative Eulerian orientation of all three projection is fixed, and can be determined analytically [van Heel, 1987]. Note that the concept of the common line-projection or the common tilt-axis is the real-space equivalent of the concept of the common line [Crowther, 1971] shared by two central sections in 3D Fourier space. For highly symmetric objects (like icosahedral viruses), the angular reconstitution approach becomes the real-space equivalent of the earlier common-lines approach by Crowther [Crowther, 1971].

A problem occurs in the angular reconstitution approach when the fourth (and higher) projections are to be included in the solution. Ambiguities are introduced since the problem becomes overdetermined and the new projection can give different angular reconstitution solutions relative to each group

of two other projections. Overdetermination is no real problem in noise-free computer simulations, but becomes serious with real experimental data. Least-squares solutions are not the correct approach since errors do not manifest themselves as small deviations from the correct solution but rather lead to entirely erroneous angular directions. What is needed is rather an approach leading to the best solution relative to all earlier projections for each new projection introduced into the set. Since, once such a technique is available, one can always extract any projection from the set and reintroduce it as a "new" projection, such an approach can be iterated to give a best solution for all available projections as a whole. We introduce a brute-force, yet efficient, approach to the problem of introducing a new projection into a set of already "angled" earlier projections. The idea is to simply try all Euler-angle combinations for the new projection, and to see how well the different assumed projection directions perform relative to the orientations of earlier projections. To define the Euler angles: α is a rotation in the plane of the projection, β the angle the projection direction makes relative to the north-pole direction of the 3D reconstruction space, and γ the angle of the projection direction measured along the equator relative to an arbitrary zero meridian. To search for the best (α,β,γ) we first need to search for the best (β,γ) combination; the best α-value can be determined last. The influence of rotating a projection image in the plane of the projection (changing α) is to cause a cyclical shift of the sinecorr function (Fig. 4c) of the current projection image relative to an earlier projection.

For each (β,γ) combination attempted, we can predict where in the sinecorr function of the new projection relative to each of the old projections we expect the maximum peak, assuming that the correct α angle was zero. These predicted peaks are thus all systematically misplaced (in the horizontal direction, for example) by a distance corresponding to the correct α value. To find the best α value for a given (β,γ) combination, we extract all the predicted rows from the different sinecorr function, and cyclically shift each of them such that they all have the predicted $(\alpha=0)$ peak at the origin. After summing all these 1D curves, we can search for the maximum along the sum curve. The position of the peak gives us the best α value for the given (β,γ) choice; the peak-height tells us how good this (β,γ) combination is relative to all other possible ones. Since we simultaneously look at all sinecorr functions associated with the new projection, we will only find the best overall orientation for the new projection, and will not be distracted by spurious local maxima in one of the sinecorr functions studied. This approach was applied to a set of 50 projections (of which the first 12 are shown in Fig. 3b) calculated from the general (128x128x128 pixels) phantom shown in Fig. 3a. The algorithm was capable of reproducing the Euler angles used to generate the projections within a mean error of about 1.5°.

Some further complications and corresponding methodological refinements exist with the angular reconstitution approach which we intend to discuss in more detail in a future publication. The first complication is that of centering of the projections and the corresponding 3D reconstruction. With the 3D reconstruction methodology described here one assumes that all projections are translationally aligned relative to a common origin. This is not always simple to achieve with (for example) the center-of-mass technique, so we have experimented with reconstruction schemes in which we first reconstruct the 3D SCF of the object based on the 2D SCFs of the various projections. This procedure exploits the translational invariance of the 2D/3D SCFs, and thus allows us to first determine the Euler orientations (α,β,γ) of the projections (albeit with a "180°" ambiguity due to the centrosymmetry of the 2D/3D SCF) and only then to determine the x-y shifts needed to bring each of the 2D projections to a common origin.

Refinements of the approach are also needed for reconstructing symmetric objects: each new projection has common projection directions not only with respect to the earlier projections but also relative to itself due to the symmetry of the molecular structure. In the sinecorr function of the new projection relative to itself, the α angle orientation of the projection causes a cyclical shift in both the directions of the rows and of the columns, i.e., causes a net cyclical shift along the diagonal direction of the sinecorr function. These effects, which need to be taken into account when analyzing symmetric oligomeric molecules, correspond to the common-lines and cross-common lines in the analysis of icosahedral structures [Crowther, 1971].

Discussion and Concluding Remarks

A set of micrographs taken from a preparation of randomly oriented biological macromolecules can contain thousands of molecular images from which much information about the 3D structure of the molecule can be revealed by advanced image processing techniques. The key techniques are MSA techniques capable of finding similarities and dissimilarities existing within the large and heterogeneous set of molecular images. Invariant functions and correlation functions are important tools for retrieving the noise-free characteristic views of molecules being studied. Our recent developments

with these techniques will facilitate the routine analysis of individual (randomly oriented) biological macromolecules in electron microscopy, particularly for specimens embedded in vitreous-ice.

It has long been the consensus that the single-particle approach was limited fundamentally by the need to start the procedure by aligning two noisy molecular images relative to each other by means of correlation alignment. Such alignments break down at relatively low noise levels [Saxton and Frank, 1977; van Heel, 1982]. We have shown here that alignment can be seen ("alignment by classification") as an overall statistical property of a full set of images; increasing the size of the set will automatically improve the quality of the alignments. Alignments between two noisy individual images play no role in this approach, and certainly represent no fundamental lower limit. Counting entirely on MSA techniques to find both the different views and the planar rotational/translational alignment differences for each view in a data set, however, may require one or two orders of magnitude more molecular images than our current average of 5000 images per data-set, and may thus exceed what we currently consider to be manageable numbers. Invariant-function classifications can be used to presort sets of images into sub-groups containing mainly one characteristic molecular view. The problem of aligning the molecular images belonging to one more or less homogeneous sub-population of images can be solved with relative ease, as was demonstrated above in our example of the 50S ribosomal subunit.

The combination of techniques discussed in this paper form a complete and self-contained methodology for molecular structure analysis based on randomly oriented single particles embedded in vitreous-ice. However, our first analyses of such specimens (unsupported vitreous-ice over holes) have shown a significant amount of preferred orientation relative to the ice/air surface. To further "randomize" the different views available, in spite of preferred orientations, we envisage imaging these specimens using small tilts of only about 20°-30° using dynamic focussing [Zemlin, 1989] to obtain constant focus over the full area of the micrographs. By introducing a tilt into our experiments (one still just takes one image per object area), the methodological differences from the approach of the Albany group are somewhat reduced. However, a tilt in our approach is introduced to enhance the randomness of the molecular orientations in the preparation, whereas in the Albany school tilt is introduced to tilt a given set of molecules known (from the 0° image) or assumed to belong to a single pure preferred orientation.

We expect that the importance of finding the domain or quaternary structure of large biological molecules will grow in the future. With better schemes to predict the secondary/tertiary structure of a protein from the primary sequence alone (cf. [Kolata, 1986; Blundell et al., 1987; van Heel, 1991], the boundary conditions imposed by a measured low-resolution (say, 2nm) 3D structure obtained with EM techniques may be sufficient to start building an atomic-resolution model of the protein under investigation. In this context, we wish to draw attention to a multivariate statistical sequence analysis and structure prediction approach which was an immediate consequence of our methodological studies in electron microscopy [van Heel, 1991].

Acknowledgements

The authors are grateful to Elmar Zeitler for his continued support of our work. One of us (E. Orlova) was supported by a Max Planck grant for stimulation of international scientific collaborations.

References

M Adrian, J Dubochet, J Lepault and AW McDowall (1984) Cryo-electron microscopy of viruses. Nature 308: 32-36.

JP Benzécri, M Danech Pejouh, T Moussa and J-P Romeder (1977) Analyse descriminante et Analyse Factorielle (Discrimination analysis and factorial analysis). Les Cahiers de l'Analyse des Données 4: 369-406.

JP Benzécri (1980a) L'Analyse des Données. Volume 2: L'Analyse des Correspondences (Data analysis. Vol. 2: Correspondence analysis). Dunod, Paris.

F Benzécri (1980b) Introduction à la Classification Automatique d'Après un Example de Données Medicales (Introduction to automatic classification using medical data as an example). Les Cahiers de l'Analyse des Données 5: 311.

TL Blundell and LN Johnson (1976) Protein Crystallography. Academic Press, New York.

TL Blundell, BL Sibanda, MJE Sternberg and JM Thornton (1987) Knowledge-based prediction of protein structures and the design of novel molecules. Nature 326: 347-352.

EJ Boekema, JA Berden and M van Heel (1986) Structure of mitochondrial F_1-ATPase studied by electron microscopy and image processing. Biochim. Biophys. Acta 851: 353-360.

EJ Boekema and M van Heel (1989) Molecular shape of Lumbricus terrestris erythrocruorin studied by electron microscopy and image analysis. Biochim. Biophys. Acta 957: 370-379.

L Borland and M van Heel (1990) Classification of image data in conjugate representation spaces.

J. Opt. Soc. Am. A 7: 601-610.

JM Carazo and J Frank (1989) Three-dimensional matching of macromolecular structures obtained from electron microscopy: an application to the 70S and 50S E.coli ribosomal particles. Ultramicroscopy 25: 13-22.

JL Carrascosa, E Viñuela, N García and A Santisteban (1982) Structure of the head-tail connector of bacteriophage Φ29. J. Mol. Biol. 154: 311-324.

LW Cooley and TW Tuckey (1965) An algorithm for the machine calculation of complex Fourier series. Math. Comput. 19: 297-301.

RA Crowther (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. Philosophical Transactions of the Royal Society of London B 261: 221-230.

RA Crowther and LA Amos (1971) Harmonic analysis of electron images with rotational symmetry. J. Mol. Biol. 60: 123-130.

MJC De Jong and D Typke (1990) Dynamic focussing of tilted specimens in the CM TEM via remote control. In: LD Peachey and DB Williams, eds., Proceedings of the XIIth International Congress for Electron Microscopy in Seattle. San Francisco Press, San Francisco, 150-151.

DJ DeRosier and A Klug (1968) Reconstruction of three dimensional structures from electron micrographs. Nature 217: 130-134.

KH Downing (1991) Spot-scan imaging in transmission electron microscopy. Science 251: 53--59.

P Dube, P Tavares, R Lurz and M van Heel (1993) Bacteriophage SPP1 portal protein: a DNA pump with 13-fold symmetry. EMBO J. 12: 1303-1309.

J Dubochet, M Adrian, J-J Chang, J-C Homo, J Lepault, AW McDowall and P Schultz (1988) Cryo-electron microscopy of vitrified specimen. Q. Rev. Biophys. 21: 129-228.

J Frank (1980) The role of correlation techniques in computer image processing. In: PW Hawkes, ed., Computer processing of electron microscope images. Springer, Berlin, 187-221.

J Frank, W Goldfarb, D Eisenberg and TS Baker (1978) Reconstruction of glutamine synthetase using computer averaging. Ultramicroscopy 3: 283-290.

J Frank, A Verschoor and M Boublik (1981) Computer averaging of electron micrographs of 40S ribosomal subunits. Science 214: 1353-1355.

J Frank and T Wagenknecht (1984) Automatic selection of molecular images from electron micrographs. Ultramicroscopy 12: 169-176.

S Fuller (1987) The T4 envelope of Sindbis virus is organised by interactions with a complementary T3 capsid. Cell 48: 923-934.

G Harauz and M van Heel (1986) Exact filters for general geometry three dimensional reconstruction Optik 73: 146-156.

G Harauz, E Boekema and M van Heel (1988) Statistical image analysis of electron micrographs of ribosomal subunits. Methods in Enzymol. 164: 35-49.

R Henderson, JM Baldwin, KH Downing, J Lepault and F Zemlin (1986) Structure of purple membrane from halobacterium Halobrium: recording, measurement and evaluation of electron micrographs at 3.5Å resolution. Ultramicroscopy 19: 147-178.

R Henderson, JM Baldwin, TA Ceska, F Zemlin, E Beckmann and KH Downing (1990) A model for the structure of Bacteriorhodopsin based on high resolution cryo-electron microscopy. J. Mol. Biol. 213: 899-929.

Image Science (1991) The IMAGIC-5 software package is distributed by: Image Science Software GmbH, Mecklenburgische Str. 27, D-14197 Berlin, Germany, Tel: (49-30) 824 50 15; Fax: (49-30) 824 87 79.

BK Jap, KH Downing and J Walian (1990) Structure of PhoE Porin in projection at 3.5Å resolution. J. Struct. Biol. 103: 57-63.

Z Kam (1980) The reconstruction of structure from electron micrographs of randomly oriented particles. J. Theor. Biol. 82: 15-39.

E Kellenberger, M Häner and M Wurtz (1982) The wrapping phenomenon in air-dried and negatively stained preparation. Ultramicroscopy 9: 139-150.

G Kolata (1986) Trying to crack the second half of the genetic code. Science 233: 1037-1039.

W Kühlbrandt and Da Neng Wang (1991) Three dimensional structure of plant light harvesting complex determined by electron crystallography. Nature 350: 130-134.

W Kunath, K Weiss, H Sack-Kongehl, M Kessel and E Zeitler (1984) Time-resolved low-dose microscopy of glutamine synthetase molecules. Ultramicroscopy 13: 241-252.

L Lebart, A Morineau and N Tabard (1977) Techniques de la Description Statistique (Techniques of Statistical Description). Dunod, Paris.

L Lebart, A Morineau and KM Warwick (1984) Multivariate descriptive statistical analysis. Wiley, New York.

DL Misell (1978) Image Analysis, Enhancement and Interpretation. North-Holland, Amsterdam.

H Öttl, R Hegerl and W Hoppe (1983) Three-dimensional reconstruction and averaging of 50S ribosomal subunits of Escherichia coli from electron micrographs. J. Mol. Biol. 163: 431-450.

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

SV Provencher and RH Vogel (1988) Three dimensional reconstruction from electron micrographs of disordered specimens. I. Method. Ultramicroscopy 25: 209-223.

M Radermacher, T Wagenknecht, A Verschoor and J Frank (1986) A new reconstruction scheme applied to the 50S ribosomal subunit of E.coli. J. Microsc. 141: RP1-RP2.

M Radermacher, T Wagenknecht, A Verschoor and J Frank (1987a) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50s ribosomal subunit of Escherichia coli. J. Microsc. 146: 113-136.

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

M Radermacher (1988) Three dimensional reconstruction of single particles from random and non random tilt series. J. Electr. Microsc. Techn. 9: 359-394.

J Saß, E Beckmann, G Büldt, DL Dorset, M van Heel, J Rosenbusch, F Zemlin, E Zeitler and A Massalski (1989) Densely packed ß-structure at the protein-lipid interface of Porin is revealed by high resolution cryo-electron microscopy. J. Mol. Biol. 209: 171-175.

WO Saxton and J Frank (1977) Motif detection in quantum noise-limited electron micrographs by cross-correlation. Ultramicroscopy 2: 219-227.

M Schatz and M van Heel (1990) Invariant classification of molecular views in electron micrographs. Ultramicroscopy 32: 255-264.

M Schatz, J Jäger and M van Heel (1990a) Lumbricus terrestris erythrocruorin embedded in vitreous ice, in: G Préaux and R Lontie, eds., Invertebrate dioxygen carriers, Leuven University Press, Leuven, 245-250.

M Schatz, J Jäger, and M van Heel (1990b) Molecular views of ice-embedded Lumbricus terrestris erythrocruorin obtained by invariant classification. In: LD Peachey and DB Williams, eds., Proceedings of the XIIth International Congress for Electron Microscopy in Seattle. San Francisco Press, San Francisco, 450-451.

M Schatz (1992) Invariante Klassifizierung elektronenmikroskopischer Aufnahmen von eiseingebetteten biologischen Makromolekülen (Invariant classification of electron microscopical micrographs of ice-embedded biological macromolecules). Thesis. DHS 454: 53-55.

M Schatz, and M van Heel (1992) Invariant recognition of molecular projections in vitreous-ice preparations. Ultramicroscopy 45: 15-22.

RC Singleton (1969) An algorithm for computing the mixed radix fast Fourier transform. IEEE Trans. Audio Electroacoust. AU-17: 93-103.

M Steinkilberg and HJ Schramm (1980) Eine verbesserte Drehkorrelationsmethode für die Strukturbestimmung (An improved rotational correlation method for structural determination). Hoppe-Seyler's Z. Physiol. Chem. 361: 1363-1369.

M Stewart and G Vigers (1986) Electron microscopy of frozen-hydrated biological material. Nature 319: 631-636.

M Unser, AC Steven, and BL Trus (1989) Odd men out: A quantitative objective procedure for identifying anomalous members of a set of noisy images of ostensibly identical specimens. Ultramicroscopy 19: 337-348.

BK Vainshtein, and AB Goncharov (1986) Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections. Doklady Akad. Nauk. USSR 287: 1131-1134 (in Russian).

M van Heel and J Frank (1980) Classification of particles in noisy electron micrographs using correspondence analysis. In: ES Gelsema and L Kanal, eds., Pattern Recognition in Practice I. North-Holland Publishing, Amsterdam, 235-243.

M van Heel and J Frank (1981) Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicroscopy 6: 187-194.

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

M van Heel (1982) Detection of objects in quantum noise limited images. Ultramicroscopy 8: 331-342.

M van Heel, JP Bretaudière and J Frank (1982) Classification and multi-reference alignment of images of macromolecules. In: JB Le Poole, E Zeitler, G Thomas, G Schimmel, C Weichan and K von Bassewitz, eds., Proceedings of the 10th International Congress on electron microscopy, Hamburg, Volume 1, 563-564.

M van Heel (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). Ultramicroscopy 13: 165-184.

M van Heel and M Stöffler-Meilicke (1985) Characteristic views of E. coli and B. stearothermophilus 30S ribosomal subunits in the electron microscope. EMBO J. 4: 2389-2395.

M van Heel (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy 21: 111-124.

M van Heel (1989) Classification of very large electron microscopical image data sets. Optik 82: 114-126.

M van Heel (1991) A new family of powerful multivariate statistical sequence analysis techniques. J. Mol. Biol. 220: 877-887.

M van Heel, M Schatz, and E Orlova (1992) Correlation functions revisited. Ultramicroscopy 46: 307-316.

RH Vogel, SW Provencher, C-H von Bonsdorff, M Adrian, and J Dubochet (1986) Envelope structure of Semliki Forest virus reconstructed from cryo-electron micrographs. Nature 320: 533-535.

T Wagenknecht, JM Carazo, M Radermacher, and J Frank (1989) Three-dimensional reconstruction of the ribosome from E. coli. Biophys. J. 55: 455-64.

JH Ward Jr. (1982) Hierarchical grouping to optimize an objective function. J. Am. Statist. Assoc. 58: 236-244.

F Zemlin (1989) Dynamic focussing for recording images from tilted samples in small-spot scanning with a transmission electron microscope. J. Microsc. Techn. 11: 251-257.

Discussion with Reviewers

J. Frank: The advantages of the "Berlin" approach cannot be discussed without spelling out with what it is being compared to; upon closer inspection, and naming the competing technique, none of these advantages hold up:

(i) "each of the molecules is illuminated only once" is equally true for the random-conical approach where only the first, tilted image is used for reconstruction;

(ii) "no problems with defocus differences" -- no problems either in the random-conical approach, since the elimination of the micrograph margins can hardly be considered a problem;

(iii) flattening affects each view differently, and is actually one of the sorest points of the authors' favorite technique, hardly an "advantage" over a technique that combines **only** particles exhibiting the same view; and

(iv) "there is no missing cone" if indeed, if the molecules "really" exhibit "almost" random orientations. The reality of experiments has never produced such conditions. Specifically the greater propensity of ice-embedded molecules to exhibit more randomness in orientation than air-dried stained samples has been largely a conjecture (of many people pursuing single particle studies, including us), which is not borne out by experiments. With negative staining, on the other hand, it has proved impossible to merge data from different views without degrading the 3D results, according to Carazo and Frank (1989).

Authors: (i) In contrast to the statement above, the random conical tilt approach (now renamed to the SECReT approach) requires two exposures of the specimen.

(ii) In contrast to the statement above, the defocus differences within one image tilted over, say 45°, for use with the SECReT approach leads to very considerable defocus differences within the data set.

(iii) Flattening occurs to a large extent ("first approximation") in the direction perpendicular to the support film in negative stain preparations which does not affect the '0°' 2D projection of the structure which is the only projection used in our method. In contrast, the SECReT approach reconstructs the actual flattened 3D structure and they, thus, very often look like pancakes.

(iv) Indeed, even in ice one often encounters some very preferred orientations. However, with relatively thick ice layers and a large number of molecular images in the data set, we manage to bring even the rare views to statistical significance.

J. Frank: The usage of small tilt of 20-30 degrees is suggested as a remedy for the possible scarcity of views; This brings the whole idea of using only one picture into question. Since in this new scenario, a tilted picture is used anyway, why not use the 0 degree view to collect useful information about orientation? This would, of course, introduce a random-conical data collection through the back door, to the detriment of the "we" versus Albany world view.

Regarding the destructiveness of the second image in the random conical approach, this concern surfaces immediately after the paragraph in which the use of a second (less destructive?) image is proposed as part of the "Berlin" approach. If there is such a concern, then it should be expressed in each context where more than one image are needed. Fortunately, however, the concern is misplaced for at least two reasons: (i) It is obviously possible to divide the maximum tolerable dose between the two exposures; and (ii) only tilted projections enter the reconstruction, and the practical requirement of rotation search and MSA in the 0-degree view can be met with the second image even if this shows the particles slightly damaged. What is called "the first problem" is not a problem at all: it is not necessary for the molecule to show a single preferred orientation; it could show several, to the benefit of better data coverage. Among numerous particles investigated, the 30S and 60S subunits were the only ones, presumably due to their shapes, that exhibited a continuum of barely distinguishable views. Another part of the "first problem" cited is not a problem either, according to the work the authors cite themselves: if a "view" can be shown to be actually two closely related views, as in Harauz *et al.* (1988), then it is possible, by the same method, to separate these and proceed with independent reconstructions. This is actually our working method. The fact that Radermacher *et al.* (1986, 1987a, b) did not find this data division by using MSA is not a "problem" of the random-conical

method, but presumably due to the fact that Harauz *et al.* investigated 50S particles prepared with the **single**-layer method, while Radermacher *et al.* reconstructed particles prepared with the **double**-layer method. There is the implied message in this paragraph of the review ("some perceived characteristic view") that the random-conical method relied on the use of visual criteria only. This is neither an intrinsic characteristic of the method (this should be obvious), nor is it the working praxis: as was elaborated as part of an unfortunate exchange with Van Heel [Frank *et al.*, J. Microscopy **159**, 117-122 (1990)] we use two stages of screening for economic reasons, the main idea being that we want to get 3D reconstructions **done**: only the first stage is the visual selection of a view or view range, the second is the use of MSA and classification to narrow down this range by using objective methods.

The next problem on this discouraging list is the fact that only a "small band of images" can be used due to the change of defocus. If one aims at 3 nm resolution, then the entire micrograph (at 36,000x mag) can in fact be used. Also, the authors themselves point to the complete solution of the problem: dynamic focusing. Next, "One of the most serious problems with the negative stain preparation technique" is not a problem specific to the random-conical method, but should give the authors pause to wonder how on earth their approach to 3D reconstruction could have worked at all for negatively stained molecules -- an aim they have pursued for years. In fact, of all techniques in which projections of different particles are combined, the random-conical technique works best in these circumstances, because it combines only particles sitting in the same orientation.

Authors: The referee has misunderstood our approach with a 20-30° tilt. We use that image as the **only** image without an additional 0° image (the SECReT approach). There is no "second (less destructive) image" and no random-conical data collection "introduced through the back door, to the detriment of the 'we' versus the Albany world view". Sorry.

Yes, we agree with the referee that defocus variations over a tilted specimen can be corrected using dynamic focussing. Actually, we state this fact in the next sentence in the paper.

J. Frank: The normalization of all the images in the same way (adjusting, averages and variances to chosen values) should be justified: is this a good idea? Could different views in the data set not have different statistics?

Authors: The normalization of the statistics of the individual molecular images (after band-pass filtering) is important, but not as important as the band-pass filtering itself.

J. Frank: The "practice of thresholding image data to positive values only" is rightly dismissed, but without any reference to an actual usage of this strange idea. Did anybody really use this nonsensical idea?

Authors: The idea is not "nonsensical" and is necessary if negative values occur in the data. In EM, with phase-contrast as the main contrast mechanism, one essentially has zero average density data and thus has to think seriously about this problem. It is an equally "strange" (disastrous?) idea to add an arbitrary constant to the data to avoid negative values (the routine procedure in SPIDER, we were told) since the strong negative densities will end up as small positive densities and will have very little variance contribution whereas the positive values will become very large positive values with a very strong contribution to the total variance of the data set. We have been aware of these fundamental problems with Correspondence Analysis for many years and have introduced a better technique already many years ago and published the new approach [Borland and Van Heel, 1990].

J. Frank: The approach to the classification described is strange. It consists of two steps: during the first one the hierarchical classification (HAC) is applied, during the second one the post processing is applied to minimize the intra-class variance. From the description included, we find that the second step is simply the k-means procedure with the initial partition incorporated from the result of HAC. Since HAC does not attempt to minimize the intra-class variances, why use it as the preprocessor in the first place? The k-means procedure converges only to the local minimum of the variance criterion, thus to use the partition resulting from a procedure that is **not** aimed at this goal can result in a bias and unnecessary long calculations.

Authors: The issue of the HAC and our moving elements post-processor was discussed *in extenso* in earlier publications and we do not think this is the place to repeat that discussion. It is very different from a k-means post-processor as was discussed previously [Van Heel, 1989].

J. Frank: ACF alignment was first used in this context by Frank *et al.* (1978). The whole reasoning behind the introduction of the "self-correlation function" (an unfortunate term, by the way, because "auto" and "self" are the same) is dubious. It is not true that "for calculating the ACF', we square the Fourier components ...". The auto-correlation function is defined in **real** space for infinite functions and the Fourier operation mentioned is only one possible way to calculate it. Since for the finite signals the correspondence between real

and Fourier space operations is not obvious, some care should be taken to calculate a meaningful ACF using FFT. The problems of padding, average adjusting and normalization are not mentioned, yet incorrect application of any of this operations can result in the problems described in the text, or can aggravate such problems.

Even if we agree with the argumentation that the "proper" ACF is **not** what we want to use, the authors should discuss the meaning of the suggested modification in terms of SNR, optimality, and so on ...

Authors: We have meanwhile published a number of papers of the SCF and MCF methods alone in which the issues mentioned by the referee were discussed. It does not make much sense to repeat those discussions here. Although there are other equivalent ways of defining and calculating the ACF, the real-space path suffers from exactly the same problem and this criticism (?) thus seems futile and "dubious".