Scanning Microscopy Vol. 11, 1997 (Pages 131-145) 0891-7035/97$5.00+.25
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

EFFECTS OF ANGULAR OVERSAMPLING IN THREE-DIMENSIONAL RECONSTRUCTION

Nicolas Boisset[1], Pawel A. Penczek[2], Jean-Christophe Taveau[1], Valérie You[1], Felix de Haas[1] and Jean Lamy[1]*

[1]Groupe d'Analyse des Structures Antigéniques, Université François Rabelais, Tours, France
[2]Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY, USA

Abstract

For three-dimensional reconstruction of single particles from cryoelectron microscope images, one has to carefully check directions of projections sampled by the image data set. A gap in angular coverage might produce a stretching of the structure (i.e., missing-cone artifact, when using the random conical tilt series reconstruction technique). This problem may be solved by merging volumes obtained from various types of electron microscope views or by taking into account the possible point-group symmetry of the particle. In case of very low signal-to-noise ratio an oversampling in one direction of projection may produce a similar effect. A fast solution, termed topological selection, is to remove overabundant images to regain the even angular distribution, but other solutions allowing one to keep the entire data set for 3D reconstruction are discussed. These situations are simulated on a test volume and are described in experimental examples.

Key Words: Cryoelectron microscopy, three-dimensional reconstruction, single particle molecular architecture.

*Address for correspondence:
Jean Lamy
Groupe d'Analyse des Structures Antigéniques, Université François Rabelais
2bis Boulevard Tonnellé,
F-37032 Tours Cedex, France
Telephone number: (33) 02-47-37-66-84
FAX number: (33) 02-47-36-61-29
E-mail: lamy@em3d.med.univ-tours.fr

Introduction

In the field of single-particle three-dimensional (3D) reconstruction from frozen-hydrated specimen images, a current trend is to first calculate a low-resolution reference structure and then to submit it to a refinement method yielding a more accurate structure. The aim of the refinement is to assign three Eulerian angles (φ, θ, Ψ) to each image, in order to faithfully reflect its direction of projection (DOP) and to improve its centering. For the angular refinement, we use the 3D projection alignment method (Penczek *et al*., 1994), but other approaches can be applied as well (Radermacher, 1994; van Heel, 1984; Harauz and Ottensmeyer, 1984a, b). Our reliance on the 3D projection alignment method is based on its simplicity and robustness. Projections are compared in real space, thus it is easy to verify the quality of the match. Also, the procedure requires very little interaction from the user. Provided that the reference volume is available, the experimental data set is subjected to the alignment and after a prescribed number of iteration steps the final volume can be examined. The 3D projection alignment has become a standard tool in the single particle reconstruction field and its application has resulted in a number of new structures solved, nonsymmetric (Penczek *et al*., 1994; Frank *et al*., 1995; Agrawal *et al*., 1996; Verschoor *et al*., 1996; Lata *et al*., 1996) as well as structures with low symmetries (Lambert *et al*., 1994a,b; Lambert *et al*., 1995a,b,c; Boisset *et al*., 1995; de Haas *et al*., 1996a,b,c).

That the method is almost fully automatic requires special attention to be paid to possible artifacts with convergence. The most common situation is an uneven distribution of particle views collected from the micrographs. The standard approach is to take single exposure untilted pictures, attempt to calculate a 3D reconstruction, and, if the angular range of projections is not covered sufficiently, take additional pictures at low tilt (30°) and repeat the 3D reconstruction (Penczek *et al*., 1994). Due to many physical effects, e.g. interaction with the supporting grid and with the water surface, that are difficult to control and predict, the angular distribution of projections usually remains uneven, resulting in angular oversampling in certain regions. In this work, we present several examples demonstrating

List of Abbreviations

3D three-dimensional
DPR differential phase residual
DOP direction of projection
EM electron microscope
SIRT simultaneous iterative reconstruction technique SNR signal-to-noise ratio

---

that oversampling can either produce artifacts during 3D reconstruction or can disrupt convergence of the 3D projection matching procedure. We reproduce these artifacts by simulation and we propose solutions to reduce their consequences.

Experimental Procedures

Cryoelectron microscopy

Samples were applied on a 300 mesh copper grid covered with a thin holey carbon film. After removing excess solution, the grids were rapidly frozen by the blotting method (Dubochet *et al*., 1982; Adrian *et al*., 1984), using liquid ethane as the freezing agent. Cryoelectron microscopy was done with a Philips CM12 electron microscope (EM) (Philips Electron Optics, Eindhoven, The Netherlands) equipped with a low dose kit, an eucentric goniometer, Gatan (Pleasanton, CA) cryo-stage, and a Gatan cryo-transfer device. Each specimen was imaged twice, under low dose conditions (less than 10e-/A^2) at a defocus close to 2.5 μm with the grid tilted at 45° and 0°. For these experiments, an acceleration voltage of 100 kV was used, and images collected on Kodak (Rochester, NY) SO 163 electron microscope films. The exact magnification of the microscope was calculated using a catalase crystal test grid (Wrigley, 1968). For computation of refined structures some pictures were also taken at a defocus of 1.5 μm with the specimen grid untilted.

Digitization and particle selection

Micrographs were digitized on an Optronics (Chelmsford, MA) P1000 microdensitometer, using a 25 μm square scanning aperture. Windowing of particle images was performed simultaneously on tilted (45°) and untilted (0°) digitized micrographs, using an interactive selection program (Radermacher, 1988). At this step, relative orientations of each pair of micrographs in terms of tilt angle, direction of tilt axis, and translation were calculated from the particle coordinates. Selected images were submitted to a contrast inversion and normalization process as described earlier (Boisset *et al*., 1993).

Image processing

Digitized images were processed according to the method of *random conical tilt series* (Radermacher *et al*., 1986; Radermacher, 1988), which was originally designed to make use of preferred orientations of single particles on a support carbon film. Untilted-specimen images are only used once, at a preliminary stage, to determine in-plane rotation angles of the particles and to select homogeneous image sets by multivariate statistical analysis and automatic clustering. Corresponding tilted-specimen images are used exclusively for the 3D recon-structions.

For image processing, we followed the strategy described in Penczek *et al*. (1994). The whole set of untilted images was subjected first to a reference-free alignment, then homogeneous image classes were determined by correspondence analysis and hierarchical ascendant classification using Ward's criterion (Ward, 1963). All 3D reconstructions were calculated using the Simultaneous Iterative Reconstruction Technique (SIRT) (Herman, 1980), and orientations between individual 3D reconstructions were determined by direct search in real space (Penczek *et al*., 1992). After orientations were established the merged 3D reconstruction was calculated.

A refined 3D reconstruction was obtained at the final step of the image processing using the whole set of tilted-specimen images in the 3D projection alignment method (Penczek *et al*., 1994). When possible, the refined 3D reconstruction was used as a reference and the same approach was used on new sets of tilted or untilted-specimen images yielding a final volume.

The topology sphere

The critical step of a 3D reconstruction is the determination of Eulerian angles (φ, θ, Ψ), corresponding to the DOP of each cryo-EM image. To visualize the angular coverage of an image set, we designed a 3D representation of their assigned Eulerian angles, termed the topology sphere. In this representation, each DOP intersects a unitary sphere at a single point defined by the first two Eulerian angles (φ and θ). At the location of this point a triangle is placed and the rotation of this triangle corresponds to the third Eulerian angle (Ψ). A visual examination of the sphere allows the detection of missing or overabundant DOPs, which correspond respectively to empty zones and to overlapping triangles. The topology sphere is a modified version of the diagram of distribution of projections proposed in Penczek *et al*. (1994), where only two Eulerian angles (φ and θ) were visualized. Therefore, with the topology sphere one can visualize angular sampling as a 3D map around the surface rendering of the corresponding 3D recon-struction volume. To complete the description of angular coverage, we also calculate the histogram of the average number of images per DOP as a function of the angle θ.

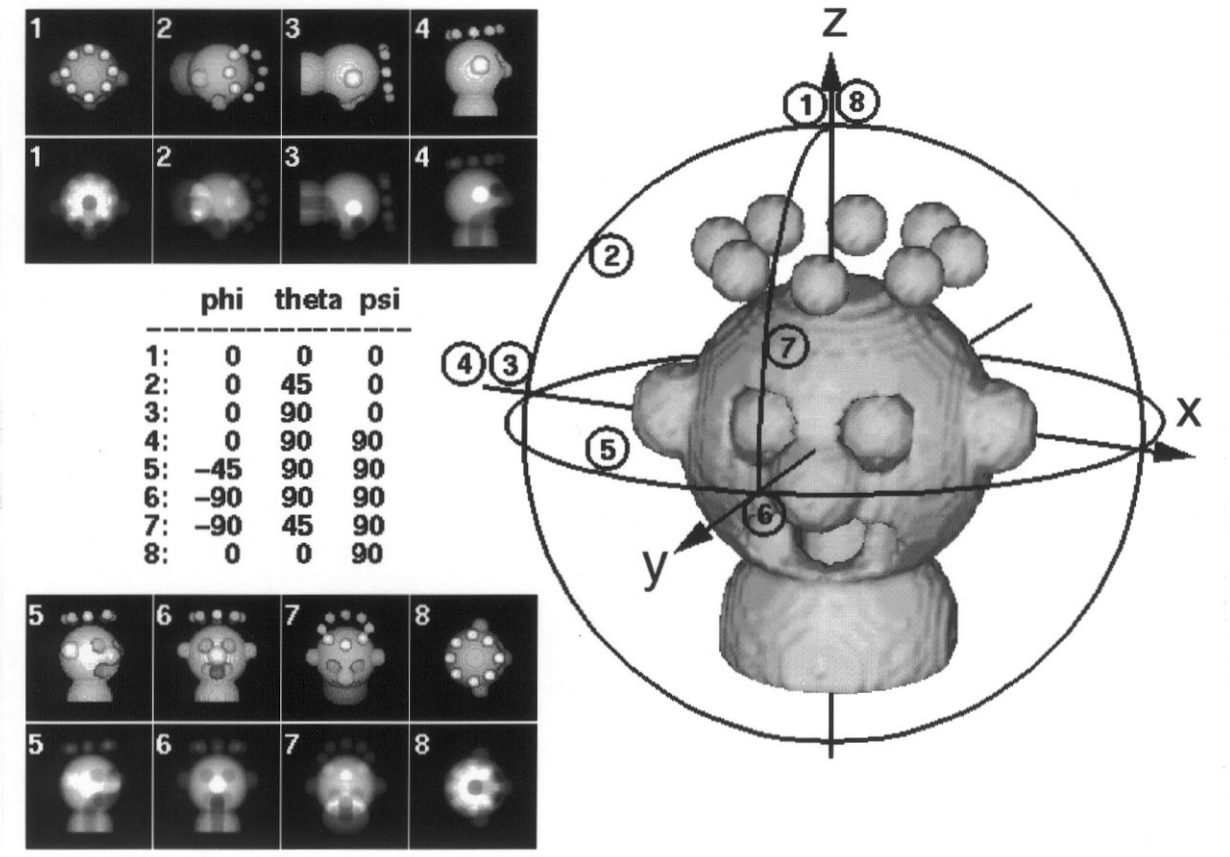


Figure 1: Convention of the Eulerian angles. The observer is moving around the object in a way defined by the Eulerian angles φ, θ, and Ψ. Eight characteristic positions are listed and the corresponding surface renderings and 2D projections of the head phantom are shown.

The Eulerian angles system and the simulation method

To illustrate the artifacts, we performed a number of simulations using a human head phantom as a test volume. The phantom is built from spheres of various radii and densities equal to 0 or 1. Roughly, it is composed of two large spheres: one for the skull and another for the neck. The nose and ears are defined by medium-size spheres of density 1, while the eyes, mouth, and trachea are composed of a set of spheres of density 0, allowing the formation of holes within the head. On the top of the skull, we added a crown defined by eight small isolated spheres. The resulting structure contains numerous bumps and notches, thus it can easily reveal deformations resulting from any 3D reconstruction artifacts.

As shown in Figure 1, specific views of the head phantom (images No. 1 to 8) are defined by sets of Eulerian angles (φ, θ, Ψ). The convention of rotation used follows that of Carazo and Frank (1988), but rather than subjecting the head phantom to a series of rotations around the z and y axes, it is easier to place the object at the center of a sphere in a fixed position and let the observer move around it. For example, a variation of θ from 0° to 90° corresponds to a move of the observer from the pole to the equator of the sphere (Fig. 1, images No. 1 to 3), while a variation of (φ from 0° to -90° corresponds to a progression of the observer from the right side to the front of the head (Fig. 1, images No. 4 to 6). Finally, a variation of Ψ does not change the perspective of the observer but only induces an in-plane rotation of the 2D projection (Fig. 1, images No. 3, 4 or 1, 8). The experimental examples presented here are all cylindrical particles. For clarity of description, they are always oriented with the cylinder axis parallel to axis Z. Therefore, in polar coordinates, circular top views of the cylinder correspond to θ values of 0° or 180°, while rectangular side views have a θ angle of 90°. Consequently, in the topology spheres shown below (Figs. 3-6), top views are represented by small triangles located near the intersections of axis Z with the sphere (poles of the sphere), while side views correspond

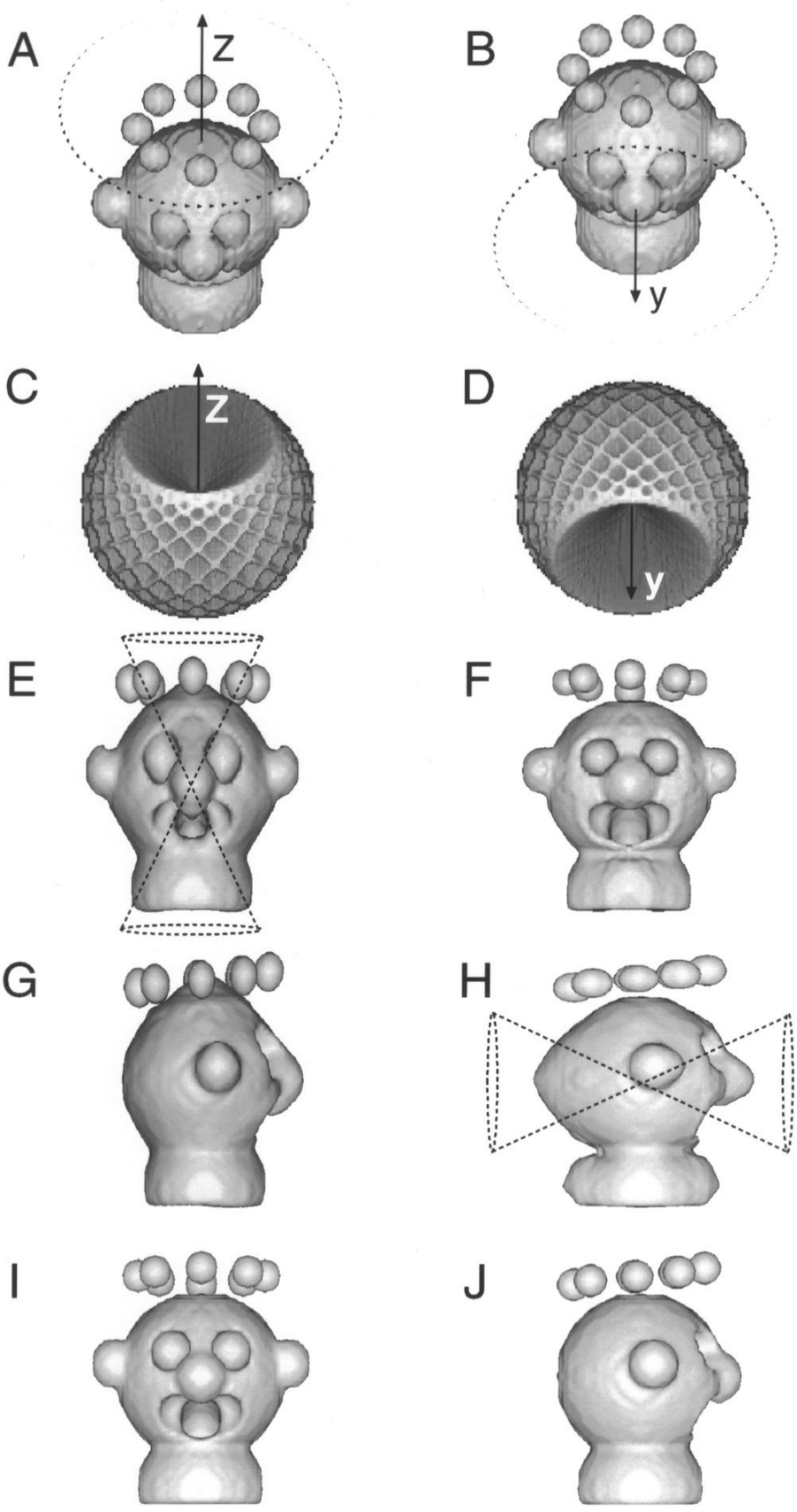


Figure 2: Simulation of the missing-cone artifacts. (**A, B**) Topology spheres of two perpendicular conical tilt series. (**C, D**) Schematic representation of coverage of 3D Fourier space corresponding to the conical tilt series **A** and **B**, respectively. (**E, G**) Surface renderings of the 3D reconstruction corresponding to the conical tilt series **A**. (**F, H**) Surface renderings of the 3D reconstruction corresponding to the conical tilt series **B**. (**I, J**) surface renderings of the 3D reconstruction corresponding to the merged conical tilt series **A** and **B**.

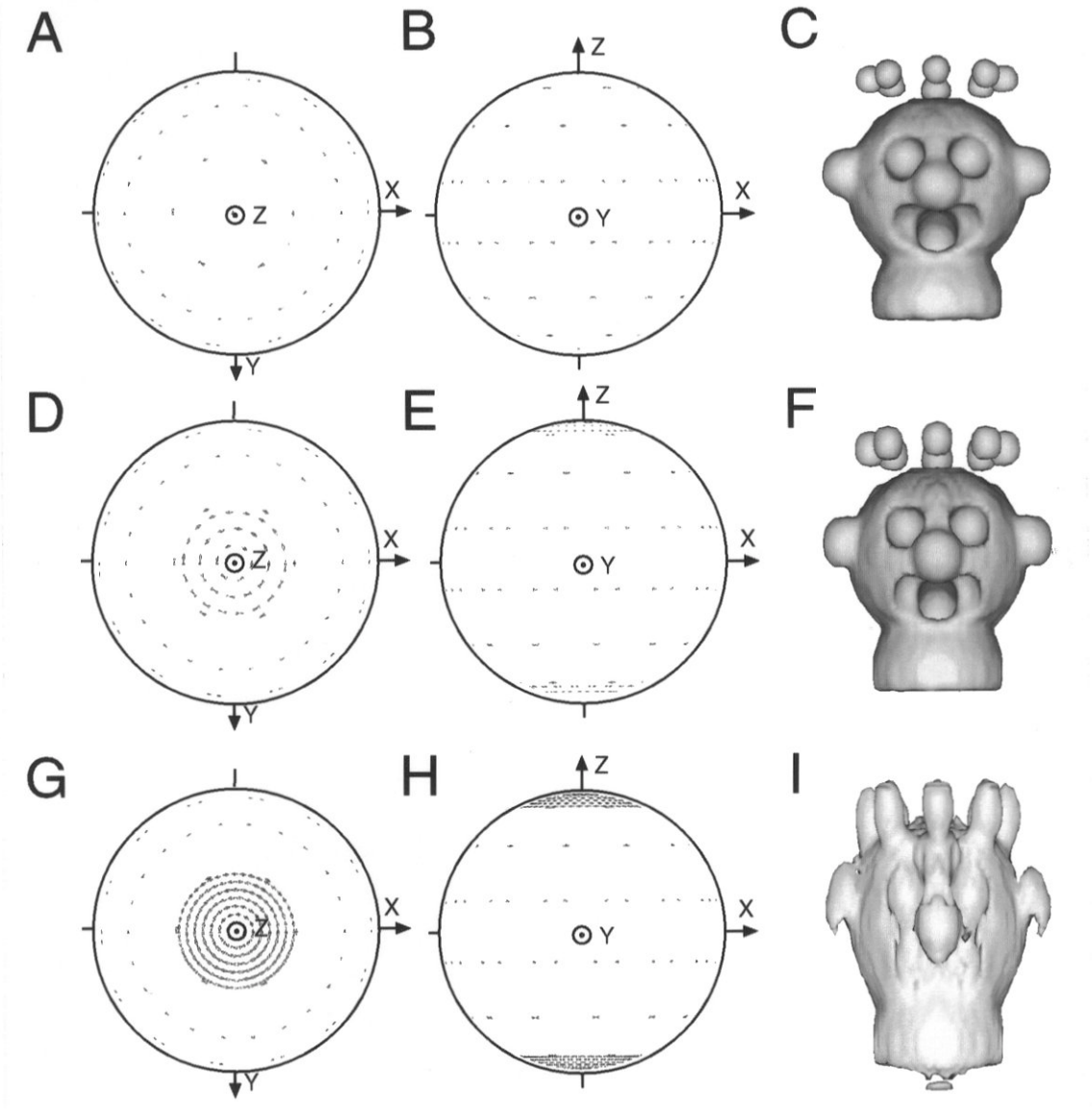


Figure 3: Simulation of overabundance of a single type of EM views. (**A, D, G**) Topology spheres seen from the pole. (**B, E, H**) Topology spheres seen from the equator. (**C, F, I**) Surface renderings of 3D reconstructions calculated from the three series of projections. (**A, B, C**) 64 projections with an even angular coverage. (**D, E, F**) 80 top view projections are added to the first image set. (**G, H, I**) 266 top view projections are added to the first image set.

to small triangles located away from intersections of the sphere with the axis Z (equator of the sphere).

Results

Simulations of reconstruction artifacts

Missing-cone artifact. The topology spheres of Figures 2A and 2B illustrate a situation where two conical tilt series of projections of the head phantom were obtained. The triangles representing DOPs of the image sets form two circles located near the pole (Fig. 2A) or the equator (Fig. 2B) of the sphere. In this case, the coverage of the Fourier space is more important, as shown in Figures 2C and 2D for the respective data sets. In these schematic representations each 2D projection corresponds to a disk located at the center of the sphere and perpendicular to its DOP. The disks partially fill the spheres but two opposite cones remain empty. These gaps in the data collection, termed missing-cones, are responsible for the deformations visible in the 3D reconstructions calculated from each projection set. In

the first case (Fig. 2A) the deformations are vertical (Figs. 2E and 2G), while in the second case, the deformations are horizontal (Figs. 2F and 2H). Thus, as expected, the direction of elongation coincides with the main axis of the missing-cone. These artifacts can be easily overcome by merging the two image sets to calculate a multicone 3D reconstruction (Figs. 2I and 2J). In this test case, the two cones were perpendicular, thus it was possible to fill the gaps in the data collection (Figs. 2C and 2D) entirely.

Overabundant images around one direction of projection. With a reference structure free of any missing cone, one can proceed with the 3D projection alignment procedure using all available projection data. The dangerous situation occurs when all DOPs are represented, but their distribution is highly uneven. In this case, the statistical relevance of the DOPs varies strongly. This has to be taken into account, in particular when one uses an iterative 3D reconstruction technique, as well as other reconstruction methods. To illustrate this problem, three reconstructions of the human head phantom were calculated from image sets with an increasing imbalance of projections, but with all DOPs present. A first volume (Fig. 3C) was calculated from a set of 64 images corresponding to evenly distributed DOPs (Figs. 3A and 3B). In Figures 3D-3F, a second volume was obtained after adding 80 images to the 64 original ones. The number of iterations of the SIRT procedure was kept the same (60 iterations). As visible in Figures 3D and 3E, the DOPs of these additional images were restricted near the poles of the topology sphere, so that we have created an overabundance of top views. Despite this imbalance of projections, the 3D reconstruction still resembles the original phantom. Although some features, such as shapes of the crown and the ears, and contour of the eyes, start to show a vertical elongation. Finally, we increased the number of top views so that each pole of the topology sphere contains 133 images (Figs. 3G and 3H). The 3D reconstruction was calculated using the same number of 60 iteration steps as in the two previous tests. Now the result (Fig. 3I) shows a strong vertical stretching. Its structure is blurred along the vertical axis and resembles the one with the vertical missing-cone artifact (Fig. 2E). The only noticeable differences between the two volumes are that with overabundant top views, the crown is linked to the skull and the ears are falling down (Fig. 3) with the vertical missing-cone artifact, the crown is also stretched but does not reach the skull and the ears are pointing up (Fig. 2E).

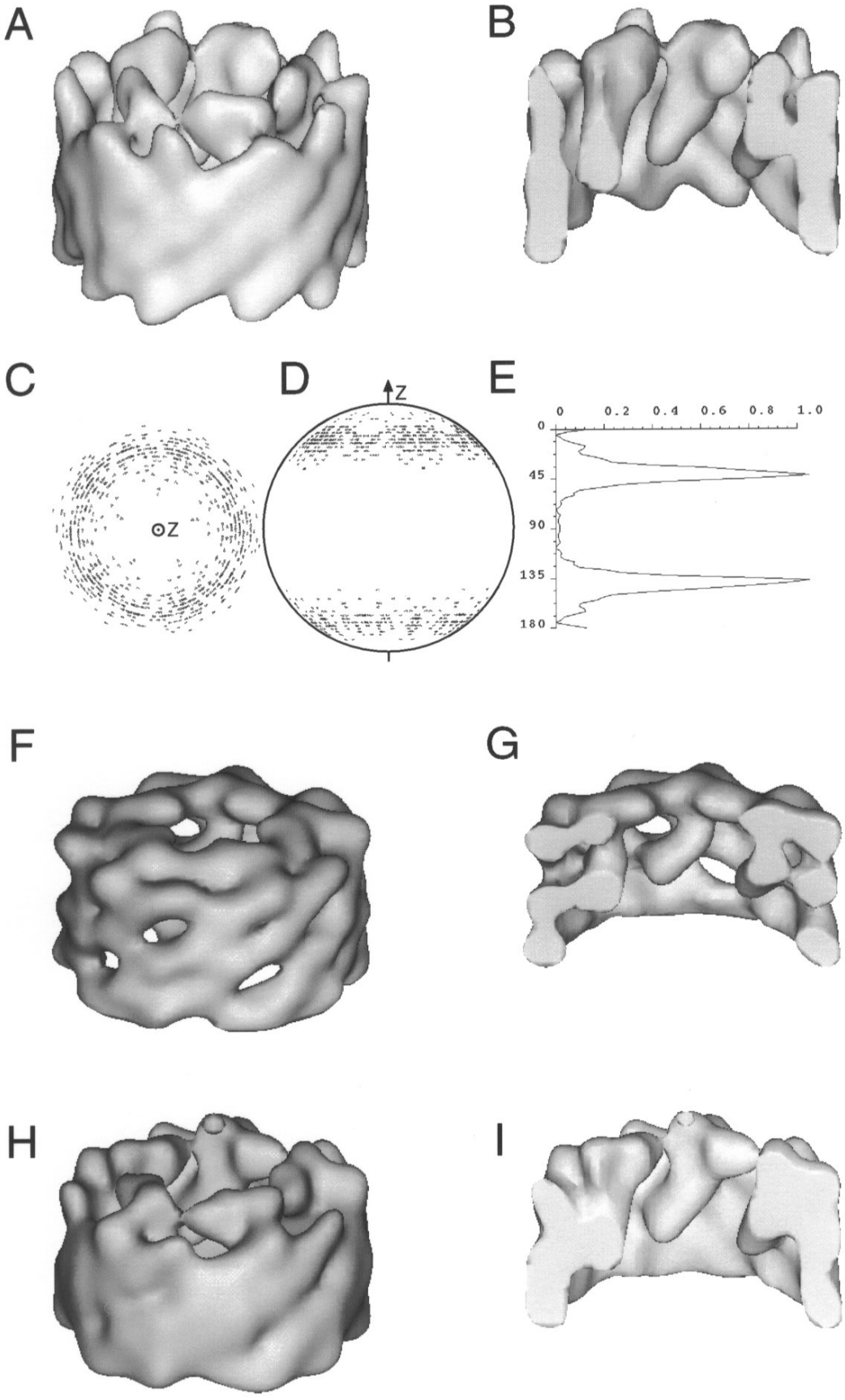
Authentic examples of deformations due to uneven image sampling

Missing-cone artifact

· **Case of chiton hemocyanin**: The respiratory pigment of the chiton *Lepidochiton sp*. is a decameric complex showing many similarities with other molluscan hemocyanins (Hc). A 3D reconstruction was carried out from tilted pairs of electron micrographs (Lambert *et al*., 1994a). The particle has a C_5 pointgroup symmetry and is composed of a hollow cylinder with five luminal oblique arches at one of its extremities (Figs. 4A and 4B). In vitreous ice, the molecule exhibited almost exclusively circular top views and seldom rectangular side views. The rare occurrence of side views is responsible for incomplete angular coverage, which produces a strong missing-cone artifact in the 3D reconstruction. The topology sphere of the data set (Figs. 4C and 4D) clearly shows that the tilted-specimen images correspond to DOPs located within two annular portions of the sphere. In addition, the histogram of the average number of images per angle θ (Fig. 4E) indicates that the only sampled DOPs are those with angle values close to 45° or 135°. The 45° angle simply reflects the tilt applied to the grid, while the presence of images with a θ angle of 135° (or 180°-45°) results from the fact that some particles were oriented upside down within the ice layer. The sorting of these two antiparallel EM views was done using multivariate statistical analysis on tilted-specimen images (Lambert *et al*., 1994a).

One year later, the architecture of another gastropod Hc, from the snail *Helix pomatia*, was resolved by the same technique (Lambert *et al*., 1995a). As this particle produced numerous top, side, and intermediate views the merged 3D reconstruction volume obtained from the whole image set was free from any missing-cone artifact. The whole molecule (a didecamer) is twice as big as the chiton hemocyanin, but there were reasons to assume that the half-molecule could resemble the decamers of chiton. For example, Figures 4F and 4G show that the half-cylinder of *Helix* Hc also possesses a C_5 point-group symmetry and is composed of a hollow cylinder with oblique arches. However, one can see many new features such as a continuous collar complex that partially closes one opening of the cylinder, five

Figure 4 (*on facing page*): Exploring the missing-cone artifact on chiton hemocyanin. (**A, B**) Surface representation of the hemocyanin of *Lepidochiton sp*. (**C, D**) Topology spheres of tilted specimen images used for 3D reconstruction of *Lepidochiton* hemocyanin, seen from the pole and the equator, respectively. (**E**) histogram of average number of images as a function of Eulerian angle θ. (**F, G**) Surface rendering of the hemocyanin half-molecule from *Helix pomatia*. This volume was used to calculate a series of 2D projections corresponding to the Eulerian angles represented in **C** and **D**. The volume calculated from these projections is represented in **H** and **I**. (**A, F, H**) Whole molecules. (**B, G, I**) Molecules with the front half removed.


A
B
C
D
E
Z
⊙Z
0
0.2
0.4
0.6
0.8
1.0
45
90
135
180
F
G
H
I

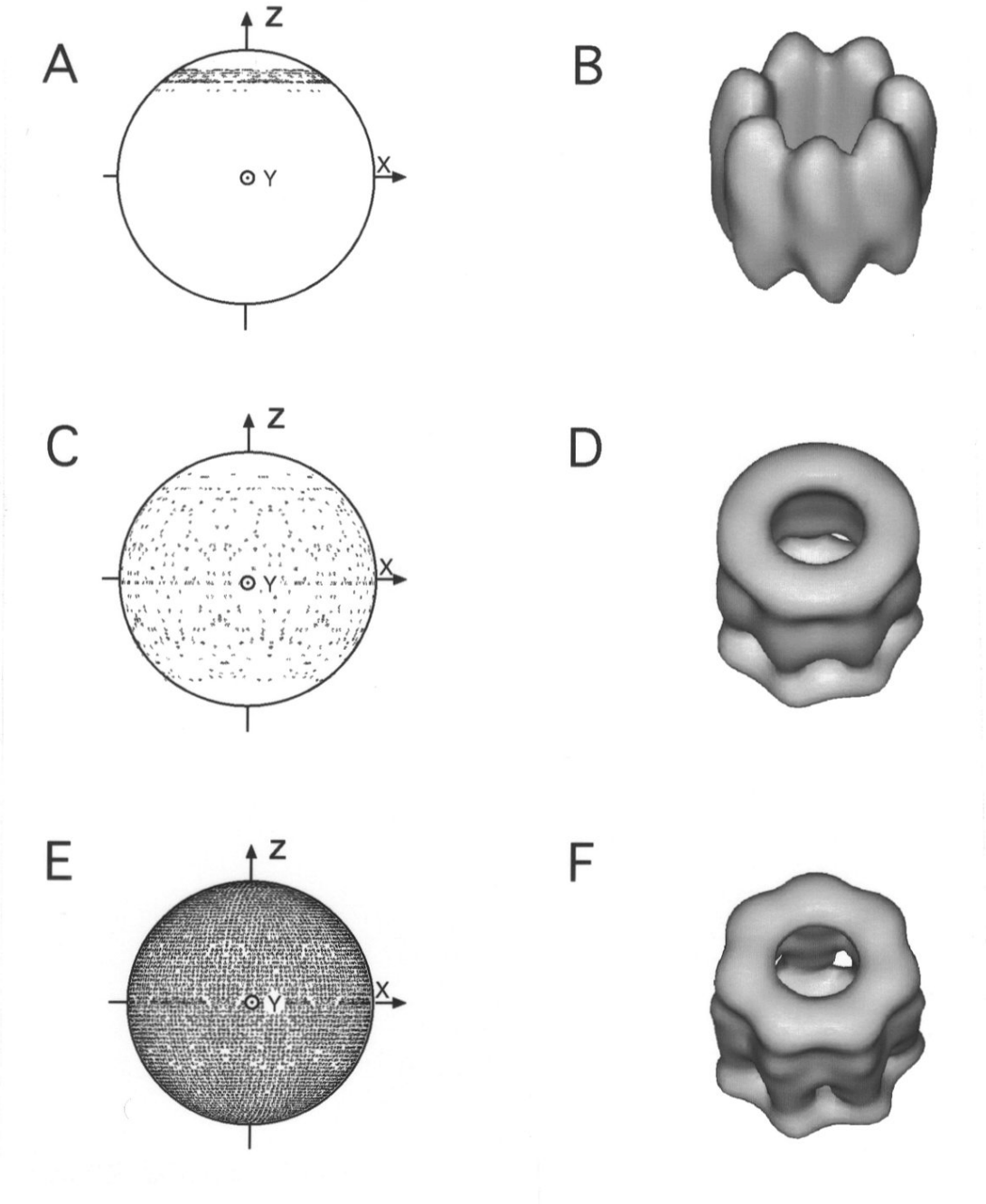


Figure 5: Impact of the symmetry of GroEL chaperonin on the missing-cone artifact. (**A, C, E**) Topology spheres seen from the equator. (**B, D, F**) surface renderings of GroEL reconstructed from tilted-specimen images of top views (**A, B**), side views (**C, D**) and after complete coverage of the topology sphere (**E, F**). (**G**) Surface rendering of GroEL reconstructed from untilted-specimen images recorded at a lower defocus (close to 1 micron) yielding a structure with a better resolution (26 A according to the differential phase residual criterion). The volume (**F**) was used as a starting reference for the refinements of this new set of images.

discontinuous oblique clefts within the cylindrical wall, and an elbow-bend shape of the oblique arches allowing them to link the collar complex to the lumen of the cylinder.

Could the missing-cone in the 3D reconstruction of chiton Hc be responsible for the structural differences with the half-complex of *Helix*? To test this hypothesis, we introduced a missing-cone artifact in the 3D reconstruction of the decameric fragment of *Helix* by projecting it along DOPs of Figures 4C and 4D. The 2D projections were used to emulate experimental cryo-EM images and to reproduce the exact gap in the data collection of chiton Hc. A 3D reconstruction was calculated from this incomplete data set. The resulting structure (Figs. 4H and 4I) is in complete agreement with the 3D reconstruction of chiton (Figs. 4A and 4B): the clefts of the hollow cylinder disappeared, the continuous collar complex is almost separated into five pieces, and the elbow bend of the five arches is not visible anymore.

· **Case of GroEL chaperonin**. The chaperonin GroEL from *E. coli* illustrates the variations of angular coverage encountered in single-cone 3D reconstructions of particles possessing a C_n point-group symmetry. The GroEL particle is composed of 14 subunits assembled into a hollow cylinder that shows a C_7 point-group symmetry. In the untilted-specimen grid, particles in the vitreous ice layer show equal amounts of circular top views, rectangular side views, and intermediate orientations. Corresponding images extracted from tilted-specimen images generate a conical set of projections corresponding to an annulus on the topology sphere. Hence, the tilted-specimen images corresponding to the top views form an annular disposition centered on the pole of the topology sphere, while the ones corresponding to the side views produce an annular disposition centered on a point located at the equator of the topology sphere.

As the molecule has a sevenfold symmetry axis, each projection can be attributed to seven equivalent DOPs. When one enforces the C_7 point-group symmetry of the particle during the 3D reconstruction process, each image is used seven times with seven different sets of Eulerian angles. In our convention system the cylinder axis (also symmetry axis) of the particle corresponds to the polar (z) axis of the topology sphere. Therefore, when taking into account the C_7 point-group symmetry of the particle, the corresponding topology sphere is copied seven times and regularly rotated around its z axis (by $n*(360^\circ/7)$, with n = {1, 2, 3, 4, 5, 6, 7 }). Such spheres are visible in Figures 5A and 5C for the tilted specimen images corresponding to the top and side views, respectively. One clearly sees that in the first case (Fig. 5A), the rotated annuli are superimposed so that the angular gap inducing the missing-cone artifact remains unchanged, while in the second case triangles are spread on every part of the topology sphere, except at the poles. Therefore, the single-cone 3D reconstruction calculated with an enforced C_7 point-group symmetry from tilted specimen images of top views (Fig. 5B) is still affected by a strong missing-cone artifact and seems stretched along the cylinder axis. Conversely, the equivalent volume calculated from tilted specimen images of side views (Fig. 5D) shows no obvious deformation, as it is not affected by the missing-cone anymore. Once all DOPs are taken into account, after merging and angular refinement of all the EM views (Fig. 5E), the final volume (Fig. 5F) shows some improvements but still shares common features with the single-cone 3D reconstruction obtained from the tilted set of side views (Fig. 5D). Therefore, when a particle possesses a C_n point-group symmetry, the missing-cone is counterbalanced in the single-cone 3D reconstruction volumes calculated from a complete conical tilt series of side views.

Oversampling artifact: case of *Riftia pachyptila* hemoglobin. While studying an extracellular hemoglobin extracted from the deep sea worm *Riftia pachyptila*, we could observe the artifact induced by an overabundance of one type of EM view (de Haas *et al*., in press). This molecule resembles the hemoglobin of *Macrobdella decora* (de Haas *et al*., 1996a) that shows a hexagonal bilayer structure composed of 12 hollow globular substructures and a linker complex. In the untilted-specimen grid, particles produced mainly hexagonal top views and rectangular side views, while corresponding tilted-specimen images contained intermediate orientations. A first 3D reconstruction was obtained from tilted-specimen images only. The topology sphere corresponding to these images was almost completely covered, except at the poles and the equator, which corresponds to the top and side views, respectively. As these orientations were commonly observed in untilted-specimen images, we decided to collect such images and to subject them to the 3D projection alignment procedure. The topology sphere of the whole image set (tilted and untilted) is almost completely covered (Figs. 6A and 6B). However, the histogram of images per DOP as a function of angle θ (Fig. 6C) indicates an overabundance of top views corresponding to DOPs with a θ angle close to 0° and 180°. We calculated two 3D reconstructions, using the same value of relaxation parameter and the same number of iterations of the SIRT. The first 3D reconstruction was calculated using all available images. It shows a strong stretching along the six-fold axis of the particle and a disappearance of the central toroid (a component of the linker complex) (Figs. 6D-6F). The second reconstruction was calculated using only one image per DOP, always choosing the image that showed the best correlation coefficient. The topology sphere of the selected image set (Figs. 6G and 6H) shows that the angular coverage is complete and the corresponding histogram of DOPs illustrates even distribution and projections (Fig. 6I). The reconstruction is no longer stretched (Figs. 6K and 6L)

and the hexagonal toroid is regained (Fig. 6J). This approach is a simple solution to the artifact related to overabundant projections. The main advantage is that the final reconstruction has an even distribution of signal-to-noise ratio (SNR) in 3D Fourier space, thus the resolution is isotropic. However, the rejection of a certain number of projections decreases to a certain extent the quality of the 3D reconstruction, and corresponds to a fast, but "brute force" approach to the problem. In the following section, we will explore one, among many possible ways, to minimize the artifact encountered when using the SIRT algorithm on projection sets with uneven angular distributions.

Discussion

Theoretical analysis

The reason for distortions caused by the overabundant images around one DOP is that we created heavy oversampling in Fourier space while keeping the number of iterations of SIRT procedure constant. In this situation, oversampled portions of Fourier spaced retained a higher weight compared to undersampled regions. The effect is immediately understandable considering the convergence properties of SIRT. One step of this procedure is as follows:

$$g^{(n+1)} = g^{(n)} + \lambda P^T (d - Pg^{(n)}) \quad (1)$$

where d is a vector containing all the available 2D projections, $g^{(n)}$ is a vector containing the solution of 3D structure during the *n-th* iteration step, P is a projection matrix containing information about spatial relations between projection data and 3D structure. Thus, it contains the same information as the whole set of Eulerian angles. In addition, λ is a relaxation parameter controlling convergence speed. It can be demonstrated that

$$0 < \lambda < (2/\delta) \quad (2)$$

where δ is so called *proper value* of the projection matrix P (Herman, 1980). The value of δ is proportional to the order of matrix P and usually $\delta << 1$. Thus, an increased number of projection requires lowering the value of λ and increasing the number of iterations. Iterations defined by Equation 1 converge in a least square sense to the true solution $g(\infty)$. Termination of the iterative process after a finite number of steps corresponds to the low pass filtration of the final structure (Norton, 1985).

To investigate angular oversampling, let us introduce the partial vector d_k containing data from the *k-th* projection and corresponding to the partial projection matrix P_k (which is derived from DOP given by Eulerian angles (φ_k, θ_k, Ψ_k). In this notation Equation 1 becomes

Figure 6 (*on facing page*): Artifacts caused by an overabundant type of projection (top views) in the 3D reconstruction of *Riftia* hemoglobin. (**A-F**) 3D reconstruction calculated using the whole data set with an excess of top views. (**G-L**) 3D reconstruction calculated using one image per direction of projection. The same number of iteration steps in the SIRT reconstruction procedure were used in both cases. Topology spheres seen from the pole (**A, G**) and from the equator (**B, H**). Histograms of average number of images as a function of Eulerian angle θ, before (**C**) and after (**I**) topological selection. Surface renderings of both 3D reconstruction volumes oriented in their top (**D, J**), intermediate (**E, K**), and side (**F, L**) views, respectively.

$$g^{(n+1)} = g^{(n)} + \lambda \sum_{k=1}^{p} (P_k^T (d_k - P_k g^{(n)})) \quad (3)$$

where p is the total number of projections. We assume for simplicity that only one projection (for $k = 1$) is overabundant and that we have its m copies. In this case,

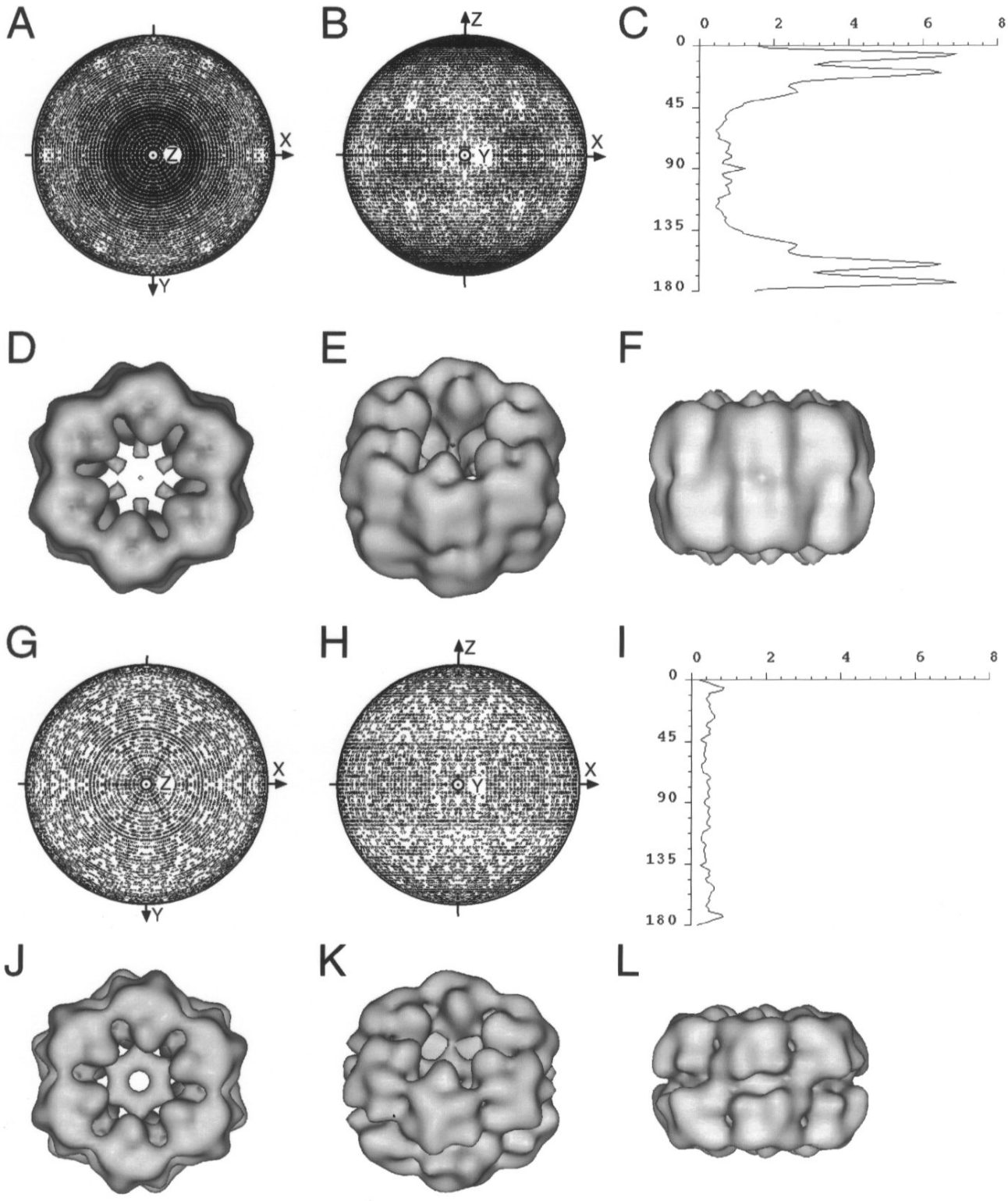
$$g^{(n+1)} = g^{(n)} + \lambda \{ \sum_{q=1}^{m} (P_1^T (d_1 - P_1 g^{(n)})) + \sum_{k=2}^{p} (P_k^T (d_k - P_k g^{(n)})) \} \quad (4)$$

which can be simplified to

$$g^{(n+1)} = g^{(n)} + \lambda m \{ P_1^T (d_1 - P_1 g^{(n)} \} + \lambda \sum_{k=2}^{p} \{ P_k^T (d_k - P_k g^{(n)}) \} \quad (5)$$

Equation 5 demonstrates that in this case the effective relaxation parameter for single projections is equal to λ, while for overabundant DOP it is $m\lambda$. Thus, the convergence rate for this particular DOP is much higher (the analysis of the rate of convergence of iterative reconstruction algorithms can be found in Norton (1985). To assure even convergence for all projections, one has to decrease the value of λ and increase the number of iteration steps approximately log$m$ times (assuming a geometrical convergence rate). Failure to account for the overabundance of projections will result in artifacts, as illustrated in Figure 3I.

In the case when the distribution of directions of projections is highly uneven or when the majority of input projections have more than one copy, the SIRT algorithm in the version given by Equation 1 very quickly becomes inefficient. This is demonstrated by Equation 4, where the


A
Z
X
Y
B
Z
Y
X
C
0
2
4
6
8
0
45
90
135
180
D
E
F
G
Z
X
Y
H
Z
Y
X
I
0
2
4
6
8
0
45
90
135
180
J
K
L

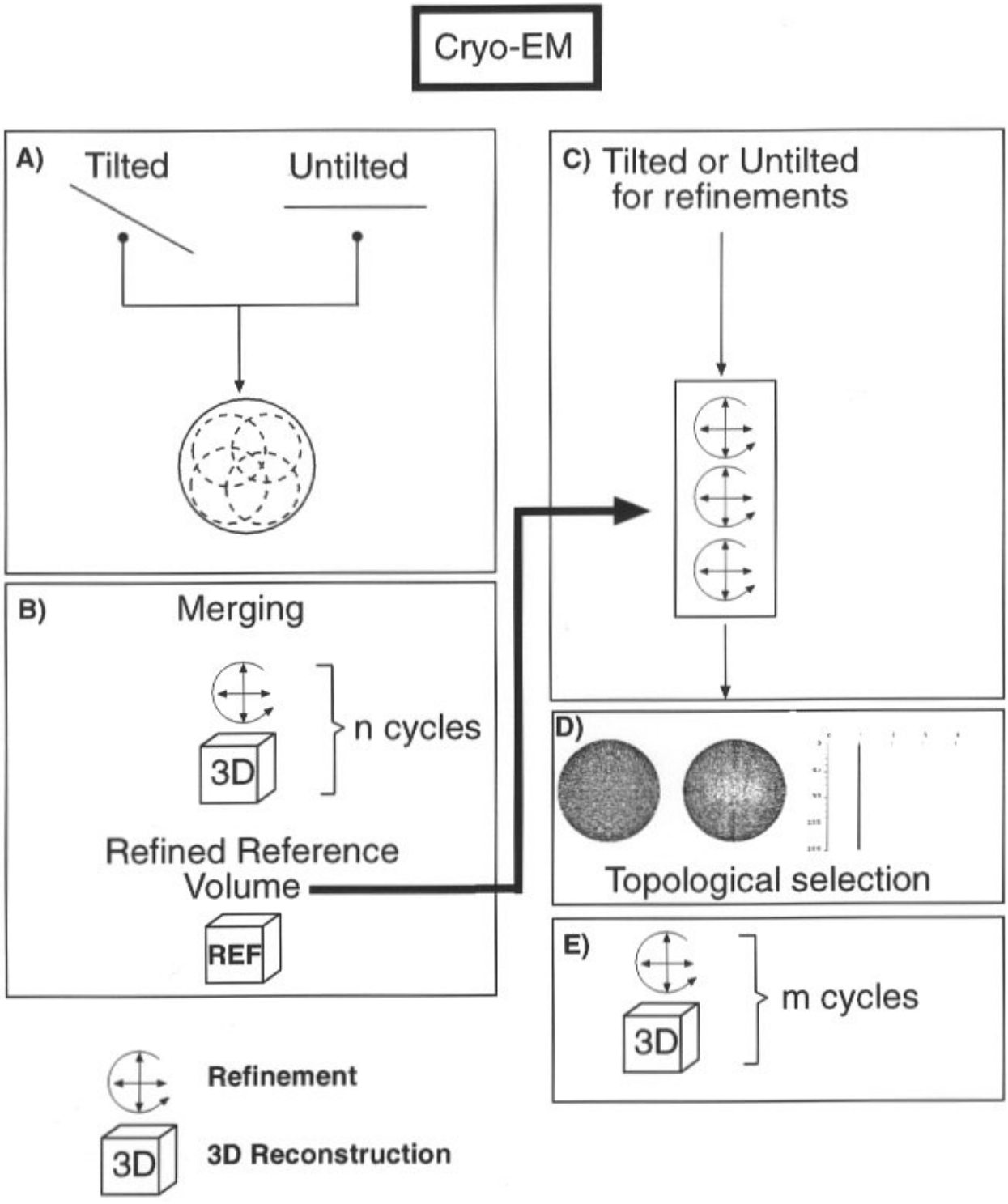


Figure 7: Strategy of refinement for single particle 3D reconstruction. (**A**) Recording of tilted- and untilted-specimen images at the cryoelectron microscope and calculation of separate single-cone 3D reconstructions. (**B**) Merging of these 3D reconstructions to fill the missing-cone followed by *n* cycles of angular (curved arrow) and shift (crossed arrows) refinements and 3D reconstructions. (**C**) Collection of additional cryo-EM images for refinement, using the final volume from step B as a reference. (**D**) Topological selection. (**E**) *m* cycles of refinement yielding the final volume.

first term denotes *m* projections and back-projections calculated in the sarne DOP given by P_q. This problem can be easily corrected in the actual implementation of the algorithm by performing only one projection and back-projection per each unique DOP and multiplying corrections to the structure by the corresponding number of projections at this DOP.

The problem of overabundant projections is by no means restricted to iterative 3D reconstruction techniques. In the filtered back-projection reconstruction methods the uneven distribution of DOPs calls for a construction of so called *general weighting filters* (Radermacher *et al*., 1986; Harauz and van Heel, 1986), which will account for any uneven sampling in Fourier space. Here we will demonstrate that the same general weighting method has to be applied in case of overabundant projections. Making the same assumption as in the derivation of Equation 5 (only one projection has *m* copies) and following the notation of Harauz and van Heel (1986), we can write schematically expression for the weighting function of *j-th* projection as

$$w_j(\omega) = \frac{1}{1 + \sum_{\substack{k=1 \\ k \neq j}}^{p} I_{jk}(\omega)} \qquad (6)$$

where $I_{jk}(\omega)$ is an *interaction function* describing an overlap between *j-th* and *k-th* projection (Harauz and van Heel, 1986). Taking into account that the first projection has *m* copies we obtain:

$$w_j(\omega) = \frac{1}{1 + \sum_{q=1}^{m} I_{j1}(\omega) + \sum_{\substack{k=2 \\ k \neq j}}^{p} I_{jk}(\omega)} = \frac{1}{1 + m\, I_{j1}(\omega) + \sum_{\substack{k=2 \\ k \neq j}}^{p} I_{jk}(\omega)} \qquad (7)$$

Thus, contribution of the overabundant projection ($k = 1$ in this case) to the weight of every other projection is *m* times higher than the contribution of single projections. This is similar to the result of Eq. 5, where it was demonstrated that the effective convergence rate of the overabundant projection to the convergence is increased *m* times. As for iterative algorithms, construction of the weighting function becomes inefficient for overabundant projections, the process can be speeded up by considering only one projection per DOP and including the corresponding number of projections in the algorithm.

So far we restricted our considerations to noise-free projection data. The EM data contains a high level of noise (usually at the level of SNR = 1), thus this fact has to be taken into account while adjusting parameters of the 3D reconstruction program (for discussion of projection variance see (Liu and Frank, 1995).

With low levels of SNR, the heavy over-representation of DOPs will result in uneven distribution of SNR in 3D Fourier space and in anisotropic resolution. One can make an attempt to increase SNR by replacing original projection data by group averages and use these in the alignment and 3D reconstruction procedures (as suggested in Schatz *et al*., 1995) and applied in other cases (Stark *et al*., 1995). Unfortunately, even though it results in substantial increase of speed, this method is not strictly correct. Indeed, by replacing *m* overlaps with overabundant projections with one intersection with the average of these projections we obtain:

$$w_j(\omega) = \frac{1}{1 + I_{j1}(\omega) + \sum_{\substack{k=2 \\ k \neq j}}^{p} I_{jk}(\omega)} \qquad (8)$$

Thus, the contribution to the weight of the *j-th* projection from the overlap with the average of overabundant projections has weight one instead of *m*. This will cause the distribution of SNR in 3D Fourier space to be incorrect. The same holds for the iterative algorithm. By replacing *m* copies of the overabundant projection d_1 by the average $\langle d_1 \rangle$ in Equation 1, its relaxation parameter, and thus its relative weight in iterations, will become *m* times lower.

Topological selection

Another possibility to achieve an even distribution of SNR and isotropic resolution is to use topological selection and to exclude surplus projections from the data set. In this way, the efficiency of the 3D reconstruction process can be improved and the risk of producing the artifact minimized.

The example of *Riftia* hemoglobin (see **Results, Oversampling artifact: case of *Riftia pachyptila* hemoglogin**) suggested a systematic approach to the 3D reconstruction and refinement of cryo-EM images, that was successfully applied to several biological macromolecules (de Haas *et al*., 1996a,b,c). The first goal is to produce a reference volume free of missing-cone artifacts. For this purpose, we use the random conical tilt method (Fig. 7A). At this stage, advantage is taken of preferred orientations of particles within the ice layer. Each such orientation produces a single cone initial 3D reconstruction. Depending on the possible symmetries, some of these structures will be corrupted by missing- cone artifacts, while some others will not (e.g., see example of GroEL with the conical tilt series corresponding to top and side views). Then, after finding

the orientations of the primary volumes, a merged 3D reconstruction is obtained from the corresponding combined image sets (Fig. 7B). Merging should give a better angular coverage and eliminate missing-cone artifacts, so that the merged 3D reconstruction volume can be used as a first reference structure for *n* cycles of 3D projection alignment of the whole image set.

If gaps are detected in the angular coverage, additional cryo-EM images are collected (Fig. 7C). Since a reference volume is available, there is no need to collect tilt-pairs of micrographs. Moreover, if the particle exhibits large numbers of intermediate EM views, tilting may not be required. Therefore, one can concentrate on the quality of the micrographs collected and use lower defocus to achieve higher resolution (Fig. 5G). These new images are subjected to several cycles of 3D projection alignment. If a strong imbalance in angular coverage is observed one can apply a topological selection in order to remove overabundant images and get an even distribution of the SNR in the 3D reconstruction (Fig. 7D). The criterion used for topological selection is the following: When several images correspond to the same DOP, the image showing the highest correlation coefficient with the corresponding 2D projection of the reference volume is kept for further processing. The set of selected images is then submitted to a last series of 3D projection alignment and 3D reconstruction, using as a reference structure the reconstruction volume of the previous cycle (Fig. 7E).

Conclusion

The combination of well established techniques (random conical tilt method, 3D projection alignment) allows fast 3D reconstruction of single particles from cryo-EM images. The strategy described was used for asymmetric as well as symmetric particles. In cases where a low-resolution reference 3D reconstruction is available, the 3D projection alignment technique alone can produce new results quickly and with minimal interaction from the user. In this context, we presented two types of artifacts that may arise while working with image sets showing an uneven angular distribution. The missing-cone artifact, resulting from a gap in angular coverage, can usually be solved by computation of a multicone or merged 3D reconstruction volume. When using iterative algebraic 3D reconstruction algorithms, a slower convergence may occur when overabundant images produce an unevenness of angular distribution. In this situation, we have successfully applied an approach of topological selection to obtain an even distribution of projections. For images corresponding to the same DOP, the correlation coefficient with the corresponding 2D projection of the reference volume (Penczek *et al*., 1994) was used as the criterion of selection.

A more subtle approach consists, when using the SIRT algorithm, in keeping the whole set of original images, but increasing the number of iterations and lowering the relaxation parameter. A theoretical analysis of the distortions resulting from the oversampling indicates that similar effects could be encountered when using other 3D reconstruction algorithms than the SIRT.

Acknowledgements

We are grateful to Dr. Olivier Fayet (Toulouse) for providing a sample of GroEL. This work was supported by: the Comite d'Indre et Loire of the French Ligue Nationale Contre le Cancer (to N.B. and J.C.T.), grant IFREMER 519529 ref SFC/CD/96239 (Dorsales program) (to J.L., Primary Investigator), grant NIH lR01 GM 29169 (to Joachim Frank, Primary Investigator). Dr. Felix de Haas and Mrs. Valerie You thank the RegiQn Centre (France) for the grants that enabled them to work in Tours as a postdoctoral fellow and doctoral student, respectively.

References

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

Agrawal RK, Penczek P, Grassucci RA, Li Y, Leith A, Nierhaus KH, Frank J (1996) Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. Science **271**: 1000-1002.

Boisset N, Penczek P, Pochon F, Frank J, Lamy JN (1993) Three-dimensional architecture of human alpha 2-macroglobulin transformed with methylamine. J Mol Biol **232**: 522-529.

Boisset N, Penczek P, Taveau JC, Lamy J, Frank J, Lamy JN (1995) Three-dimensional reconstruction of *Androctonus australis* hemocyanin labeled with a monoclonal Fab Fragment. J Struct Biol **115**: 16-29.

Carazo JM, Frank J (1988) 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.

de Haas F, Boisset N, Taveau JC, Lambert O, Vinogradov SN, Lamy JN (1996a) Three-dimensional reconstruction of *Macrobdella decora* (leech) hemoglobin by cryoelectron microscopy. Biophys J **70**: 1973-1984.

de Haas F, Taveau JC, Boisset N, Lambert O, Vinogradov SN, Lamy JN (1996b) Three-dimensional reconstruction of the chlorocruorin of the polychaete annelid *Eudistylia vancouverii*. J Mol Biol **255**: 140-163.

de Haas F, Zal F, Lallier F, Toulemon A, Lamy JN (1996c) Three-dimensional reconstruction of the hexagonal bilayer hemoglobin of the hydrothermal vent tube worm

Riftia pachyptila by cryoelectron microscopy. Proteins **26**: 241-256.

Dubochet J, Lepault J, Freeman R, Bernman JA, Homo J-C (1982) Electron microscopy of frozen water and acqueous solutions. J Microsc **128**: 219-237.

Frank J, Zhu J, Penczek P, Li Y, Srivastava S, Verschoor A, Grassucci R, Lata RK, Agrawal RK (1995) A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. Nature (London) **376**: 441-444.

Harauz G, Ottensmeyer FP (1984a) Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs. Ultramicroscopy **12**: 309-320.

Harauz G, Ottensmeyer FP (1984b) Nucleosome reconstruction via phosphorus mapping. Science **226**: 936-940.

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

Herman GT (1980) Image Reconstruction From Projections. Academic Press, New York. p 310.

Lambert O, Boisset N, Taveau JC, Lamy JN (1994a) Three-dimensional reconstruction from a frozen-hydrated specimen of the chiton *Lepidochiton sp*. hemocyanin. J Mol Biol **244**: 640-647.

Lambert O, Boisset N, Penczek P, Lamy J, Taveau JC, Frank J, Lamy JN (1994b) Quaternary structure of *Octopus vulgaris* hemocyanin. Three-dimensional reconstruction from frozen-hydrated specimens and intramolecular location of functional units $O_v e$ and $O_v b$. J Mol Biol **238**: 75-87.

Lambert O, Boisset N, Taveau JC, Preaux G, Lamy JN (1995a). Three-dimensional reconstruction of the $á_{D-}$ and $â_{C-}$hemocyanins of *Helix pomatia* from frozen-hydrated specimens. J Mol Biol **248**: 431-448.

Lambert O, Taveau JC, Boisset N, Lamy JN (1995b) Three-dimensional reconstruction of the hemocyanin of the protobranch bivalve mollusc *Nucula hanleyi* from frozen-hydrated specimens. Arch Biochem Biophys **319**: 231-243.

Lambert O, Boisset N, Taveau JC, Lamy J (1995c) Quaternary structure of *Sepia officinalis* hemocyanin three-dimensional reconstruction from frozen-hydrated specimens. Arch Biochem Biophys **316**: 950-959.

Lata KR, Agrawal RK, Penczek P, Grassucci R, Zhu J, Frank J (1996) Three-dimensional reconstruction of the 30S *Escherichia coli* ribosomal subunit in ice. J Mol Biol **262**: 43-52.

Liu W, Frank J (1995) Estimation of variance distribution in three-dimensional reconstruction. I. Theory. J Opt Soc Am **A12**: 2615-2627.

Norton SJ (1985) Iterative reconstruction algorithms: Convergence as a function of spatial frequency. J Opt Soc Am **A2**: 6-13.

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

Penczek PA, Grassucci RA, Frank J (1994) The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy **53**: 251-270.

Radermacher M, Wagenknecht T, Verschoor A, Frank J (1986) A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. J Microsc **141**: RP1-RP2.

Radermacher M (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J Electron Microsc Tech **9**: 359-394.

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

Schatz M, Orlova EV, Dube P, Jager J, van Heel M (1995) Structure of *Lumbricus terrestris* hemoglobin at 30 A resolution determined using angular reconstitution. J Struct Biol **114**: 28-40.

Stark H, Mueller F, Orlova EV, Schatz M, Dube P, Erdemir T, Zemlin F, Brimacombe R, van Heel M (1995) The 70S *Escherichia coli* ribosome at 23 A resolution: Fitting the ribosomal RNA. Structure **3**: 815-821.

van Heel M (1984) Three-dimensional reconstructions from projections with unknown angular relationship. Proc. 8th Eur. Cong. Electron Microsc. Csánady Á, Röhlich P, Szabó D (eds). Programme Comm Eighth Eur Congress Electron Microsc. Vol. 2, pp. 1347-1348.

Verschoor A, Srivastava S, Grassucci R, Frank J (1996) Native 3D structure of eukaryotic 80S ribosome: morphological homology with the *E. coli* 70S ribosome. J Cell Biol **133**: 495-505.

Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Amer Statist Assoc **58**: 236-244.

Wrigley NG (1968) The lattice spacing of crystalline catalase as an internal standard of length in electron microscopy. J Ultrastruct Res **24**: 454-464.