Journal of Structural Biology
doi:10.1006/jsbi.2001.4338, available online at http://www.idealibrary.com on IDEAL®

The Effect of Overabundant Projection Directions on 3D Reconstruction Algorithms

C. O. S. Sorzano,* R. Marabini,† N. Boisset,‡ E. Rietzel,§ R. Schröder,§ G. T. Herman,† and J. M. Carazo*

**Centro Nacional de Biotecnología (CSIC), Universidad Autónoma Madrid, 28049 Madrid, Spain; †Center for Computer Science and Applied Mathematics, Temple University, Philadelphia, Pennsylvania 19122; §Max Planck Institut für mediznische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany; and ‡Laboratoire de minèralogie cristallographie, Equipe Systemes moleculaires et biologie structurale, Université Pierre et Marie Curie Paris 6-4, Tour 16, 2eme etage, case 115, 4, Place Jussieu, 75252 Paris Cedex 05, France*

Received December 6, 2000, and in revised form February 9, 2001

The experimental process of collecting images from macromolecules in an electron microscope is such that it does not allow for prior specification of the angular distribution of the projection images. As a consequence, an uneven distribution of projection directions may occur. Concerns have been raised recently about the behavior of 3D reconstruction algorithms for the case of unevenly distributed projections. It has been illustrated on experimental data that in the case of a heavily uneven distribution of projection directions some algorithms tend to elongate the reconstructed volumes along the overloaded direction so much as to make a quantitative biological analysis impossible. In answer to these concerns we have developed a strategy for quantitative comparison and optimization of 3D reconstruction algorithms. We apply this strategy to quantitatively analyze algebraic reconstruction techniques (ART) with blobs, simultaneous iterative reconstruction techniques (SIRT) with voxels, and weighted backprojection (WBP). We show that the elongation artifacts that had been previously reported can be strongly reduced. With our specific choices for the free parameters of the three algorithms, WBP reconstructions tend to be inferior to those obtained with either SIRT or ART and the results obtained with ART are comparable to those with SIRT, but at a very small fraction of the computational cost of SIRT.

Key Words: **3D reconstruction algorithms; algorithm optimization; algorithm comparison; collection geometry.**

1. INTRODUCTION

In 3D reconstruction from projections the distribution of projection directions (expressed in the coordinate system fixed to the experimental object) is called the collection geometry. Knowledge of the collection geometry is crucial for the reconstruction process. The Projection Theorem (Herman, 1980) allows us to identify regions of the Fourier space in which there is no information regarding the Fourier transform of the object under study (missing cones, missing wedges, . . .).

A number of data collection strategies are possible in the field of 3D electron microscopy (3DEM) of biological macromolecules. If a low-resolution reference volume is available, a fruitful approach is to collect untilted field micrographs from which a wide range of different views can be obtained. By direct comparison of the experimental images with the reference volume, their directions of projection can be determined and then input to a 3D reconstruction algorithm (Frank, 1996). A shortcoming of this approach is that depending on the characteristics of the specimen and of its preparation for microscopy, the number of images corresponding to the different projection directions may vary quite significantly among the different directions. An experimental situation of this type was first investigated by Boisset *et al.* (1996) and de Haas *et al.* (1996) in their studies of the giant hemoglobin of *Lumbricus terrestris* and of the human α-2 macroglobulin. Indeed, the strong differences that they observed in the performance of the different reconstruction methods, as well as the marked discrepancies between the results obtained in an unevenly distributed situation and an evenly distributed situation, have been the primary motivation of the present study. Similar problems were encountered in a number of cases concerning cylindrical particles (e.g., CCT chaperonin, molluscan hemocyanins) oriented mostly in their circular top views within the ice layer and were overcome either by a careful interactive selection of the particles in the digitized micrographs or by taking additional tilted-specimen images.

1047-8477/01 $35.00

In principle, nothing worse than a reconstruction with anisotropic resolution was initially expected for the case of the experimental situation described in the previous paragraph. However, this is not the observation reported by Boisset *et al.* (1998), who showed qualitatively that some reconstruction algorithms used in the field tend to elongate the produced volumes along the overabundant projection directions while others do not.

Furthermore, the oversampling of some projection directions is becoming an important issue for single-particle 3D reconstruction, as most high-resolution projects require the collection of large untilted-specimen image sets at specific defocuses. This trend in the data collection is triggered by the necessity to correct for the contrast transfer function of the microscope and requires the collection of at least 10 times as many images. However, an even distribution of orientations is hard to obtain for single particles observed under such experimental conditions and all 3D reconstruction algorithms need to be thoroughly tested. This issue is particularly important for low-symmetry objects such as cylindrical structures with Cn or Dn point-group symmetries, which are among the most common shapes found in soluble biological macromolecular assemblies (e.g., chaperonins, proteasomes, helicases, extracellular respiratory pigments) and in membrane-associated proteins solubilized with detergents (e.g., calcium release channel, plant photosystem complexes, α-latrotoxin, and pore-forming toxins).

In this work we develop a quantitative approach to study this phenomenon extending the analysis of Boisset *et al.* (1998) to algebraic reconstruction techniques (ART) with blobs (Marabini *et al.*, 1998), simultaneous iterative reconstruction techniques (SIRT) (Penczek *et al.*, 1992), and weighted back-projection (WBP) (Radermacher, 1992). For this purpose, we define a methodology for quantitative comparisons and apply it to the specific case of unevenly distributed projections. This methodology (similar to that proposed by Matej *et al.* (1994, 1996) but with some modifications) performs a fair comparison between algorithms following a task-oriented approach. Each reconstruction algorithm is first adjusted so that its performance is optimized and then the results obtained by the different algorithms are compared. A random set of realizations of a parametric family of computer-generated objects (so-called phantoms), projected according to the desired collection geometry and reconstructed with optimized parameters, is used for the comparison. A number of quantitative measures of goodness are taken for each reconstruction and, finally, a statistical comparison of the merits of the different reconstruction algorithms is performed.

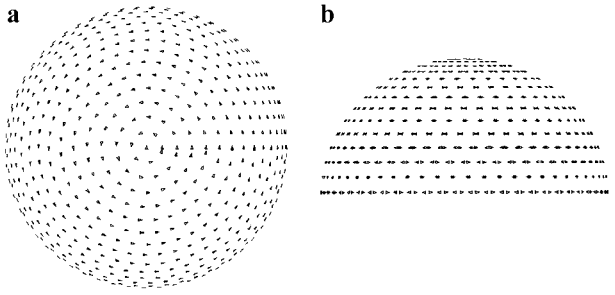

FIG. 1. Top (a) and side (b) views of the topology sphere corresponding to an even distribution of projection directions.

2. MATERIALS AND METHODS

In this section we develop the objective comparison strategy to be used in this work. This approach rests on the assumptions that the "true" object to be reconstructed is known and that the measures of quality of algorithm performance should be defined in terms of the similarity of the reconstruction to the true object. These measures of quality are tailored to well-defined and problem-specific tasks and are quantified through numerical observers commonly referred to as Figures of Merit (FOMs). In such an approach, the work must be based on data sets whose 3D structure is known. In the present study we use geometrical phantoms and quantum mechanical simulations. Our goal is a fair comparison of algorithms, so our measures must be as independent as possible from any particular phantom and/or noise realization used during the simulations.

Typically, there are a few (maybe just one) adjustable parameters that must be specified for a 3D reconstruction algorithm. In our work we optimize this set of parameters for each particular type of angular distribution. This is done by letting one of the FOMs be a training FOM and selecting the parameters of an algorithm so that this training FOM is maximized for each particular family of phantoms, noise, and directions of projections. Considering that the training FOM represents just one of the set of the measures of quality in which we are interested and that the simulations are only an approximation of the experimental reality, we devised a method that provides, instead of a unique set of optimum parameter values, a set of regions that contains them. Our evaluation process proceeds by generating a large set of projection simulations, performing the reconstruction with randomly picked parameters from their optimal regions, calculating a battery of FOMs on the reconstructed volumes, and, finally, providing a statistical comparison over the set of reconstructions obtained from all the algorithms under study.

2.1. Collection Geometry Representation

In the coordinate system attached to the particle to be reconstructed, a projection direction can be represented by a unit vector that is anchored at the origin and points into the upper half-space and that is parallel to the direction of electrons that generated that projection. All such unit vectors lie on a hemisphere. A collection geometry can be illustrated by a display of small triangles attached to the ends of the unit vectors representing all the projection directions (called a "topology sphere" by Boisset *et al.* (1998)). An even distribution of projections is characterized by a constant distance between nearest triangles. Notice that, except for a few exceptions, it is not possible to distribute evenly a given number of unit vectors. However, a fairly good approximation can be achieved as is shown in Figs. 1a and 1b.

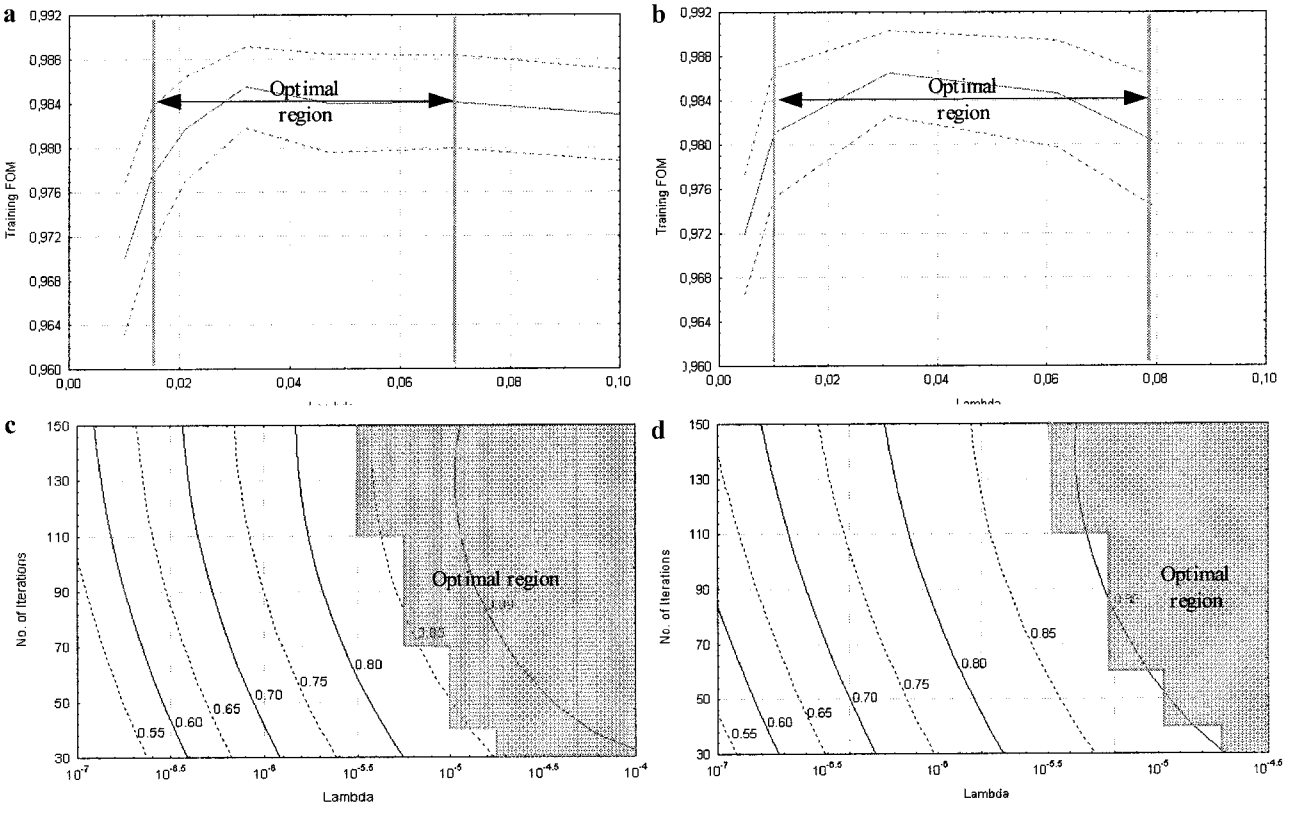

FIG. 2. tFOM versus relaxation parameter at the end of the first cycle through the data for ART in the even case (a) and the uneven case (b); the mean tFOM and its standard deviation are plotted. tFOM isocurves for SIRT are represented for the even case (c) and the uneven case (d).

2.2. Parameter Selection

Most reconstruction algorithms require the specification of a number of parameters. Some of these may be considered fixed, in the sense that a given value has been found appropriate for them, based either on some prior knowledge or on general considerations. However, there are other parameters, referred to as free parameters, whose selection is considered to be dependent on the particular specifications of the problem at hand. In this work, we are considering two clearly distinguishable types of collection geometry: the case of an even distribution of projection directions and the case of a highly uneven distribution. Therefore, our first goal is to specify the free parameters of the different reconstruction algorithms that will be tested for each of these two types of collection geometry.

A FOM is a measure of the similarity (from some particular point of view) between the object to be reconstructed (a "phantom") and its reconstruction from projection data. For any particular ensemble of phantoms and data collection methodology, we wish to select the free parameters of an algorithm so as to maximize the expected value of some Training Figure of Merit (tFOM). We do this training experimentally. We first explain our approach in the case of only one free parameter.

We randomly pick a set of phantoms and simulate the projection data generation. Let $m(\lambda)$ be the mean and $\sigma(\lambda)$ be the standard deviation of the values of the tFOM for the reconstructions of these phantoms when the free parameter has the value λ. Our method approximates the interval of those values of λ for which the hypothesis that the expected value of the tFOM is maximal at λ cannot be rejected with 99.5% confidence based on the experimentally observed means and standard deviations. In Figs. 2a and 2b we show the plot of $m(\lambda)$, as well as the plots of $m(\lambda) + \sigma(\lambda)$ and $m(\lambda) - \sigma(\lambda)$, together with the "optimal region" estimated by our algorithm. The algorithm is based on a Fibonacci search that starts with a large interval that is certain to include the optimal region and then narrows this interval to a desired level of accuracy. A similar approach has been suggested by Matej *et al.* (1996), but without taking into consideration the standard deviations in determining the optimal region.

The extension to the case of two free parameters can be done in two different ways: the first consists of selecting an optimal region that is an area in the plane, instead of an interval, defined by the two parameters (see Figs. 2c and 2d); the second, inspired in Obi *et al.* (2000), selects different values of λ for each iteration; thus, the λ for the first iteration is chosen randomly from the optimal region determined by Figs. 2a and 2b, while the free parameter for the second iteration is optimized following the procedure described in this section. This extension procedure can be applied for any number of iterations.

2.3. Phantoms

2.3.1. Analytically generated phantoms. In this paper we deal with analytically generated phantoms of the following kind. A relevant part (in the shape of a box) of space is identified and is subdivided into cube-shaped abutting voxels. Some continuous geometrical objects (in this paper these will be cylinders) are specified. A density is assigned to each voxel as follows. Consider the voxel to be subdivided into eight equal subvoxels and let n be the number of centers of these subvoxels that are inside one of the

geometrical objects. Then the density of the voxel is defined to be $n/8$. We use V to denote the set of all voxels, F (for foreground) to denote those voxels whose density is 1, and B (for background) to denote those voxels whose density is 0.

The phantoms used in this article were designed following the principles proposed by Marabini *et al.* (1998). They consist of two identical cylinders aligned along the Z axis. The position, height, radius, and separation are random to avoid a possible dependence of the Figures of Merit used on any of these variables.

2.3.2. Phantoms obtained by quantum mechanical simulation. As in previous work (Marabini *et al.*, 1997) a phantom that resembles a real protein complex was constructed for quantum mechanical image simulations. G-actin atomic coordinates (pdb entry 1ATN by Holmes *et al.*, 1990) were used to obtain a small cube of a protein-like atomic distribution (carbon, oxygen, and nitrogen atoms). To obtain spheres as shown in Fig. 3, eight such small cubes were put together to form a larger cube from which a sphere is extracted. Twelve identical spheres were then put together to form the phantom protein complex (Fig. 3). Such atom distribution phantoms were rotated according to the given collection geometry and embedded into a distribution of oxygen atoms with a density equal to that of water. The final atom distributions of protein phantom and oxygen atoms were chosen as input for image simulations using the program YaMS (Dinges and Rose, 1995). The images were simulated by calculating the propagation of an electron wave through the scattering potentials formed by the atoms in the distribution. In our present simulations only the elastic potentials were considered.

2.4. Figures of Merit

Various FOMs have been previously proposed both in medical imaging (Matej *et al.*, 1994) and in 3DEM (Marabini *et al.*, 1997). In this work we put forth a more complete set of FOMs, including those that evaluate elongations. Two broad families of FOMs have been developed; we refer to them as structural consistency measures and structural separability measures.

2.4.1. Structural consistency measures. Structural consistency refers to the correspondence between the voxel values in the reconstruction and in the phantom. It can be measured using magnitudes such as absolute errors, squared errors, differences in the means, differences in standard deviations, and differences in minimum and maximum values. The region over which these differences are taken can be the whole volume (for global measures), the background only, the foreground only, or the regions around the foreground.

We need to introduce some notation. We use R to denote the region over which the FOMs are measured. Inside the region there are N_R voxels, which are numbered from 1 to N_R. The value of the voxel number i of this region is denoted as $p_{R,i}$, while the corresponding value in the reconstruction is denoted as $r_{R,i}$. The mean voxel value in the region R in the phantom is denoted by μ_{p_R} and in the reconstruction μ_{r_R}; σ_{p_R} and σ_{r_R} denote the corresponding standard deviations. The maximum voxel value in the region R in the phantom is denoted by M_{p_R} and in the reconstruction by M_{r_R}; m_{p_R} and m_{r_R} denote the corresponding minimal values. Based on these definitions the following structural consistency FOMs have been developed:

- Mean squared error FOM (a measure of consistency between voxel values, it gives significantly more importance to large errors):

$$\mathrm{scL2}_{\mathrm{FOM}}(R) = 1 - \frac{1}{N_R}\sum_{i=1}^{N_R}\left(\frac{p_{R,i} - r_{R,i}}{2}\right)^2.$$

- Mean absolute error FOM (another measure of consistency between phantom and reconstruction voxel values with the importance of errors proportional to their size):

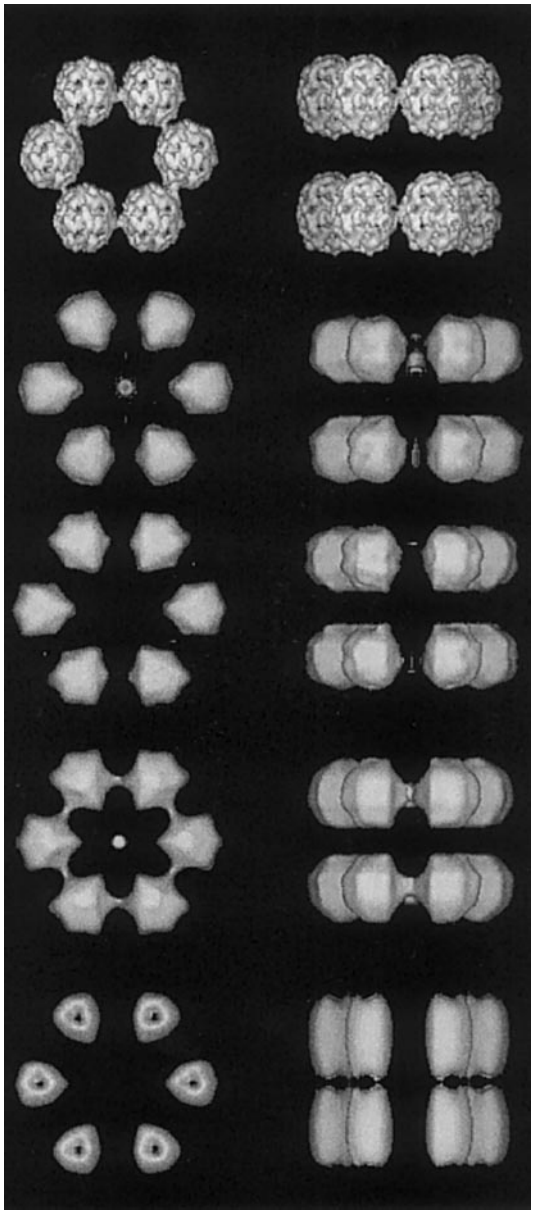

FIG. 3. 3D reconstructions of a realistic phantom simulated at the atomic level using an uneven distribution of projection directions with 3099 projections. From top to bottom: original phantom, reconstruction with ART, reconstruction with SIRT, reconstruction with WBP, reconstruction with SIRT using non-optimal parameters.

$$\mathrm{scL1}_{\mathrm{FOM}}(R) = 1 - \frac{1}{N_R}\sum_{i=1}^{N_R}\left|\frac{p_{R,i} - r_{R,i}}{2}\right|.$$

- Mean density value FOM (measures the correspondence between the density averages over the whole region):

$$\mathrm{sc}\mu_{\mathrm{FOM}}(R) = 1 - \frac{1}{2}\left|\mu_{p_R} - \mu_{r_R}\right|.$$

- Density standard deviation FOM (checks if the standard

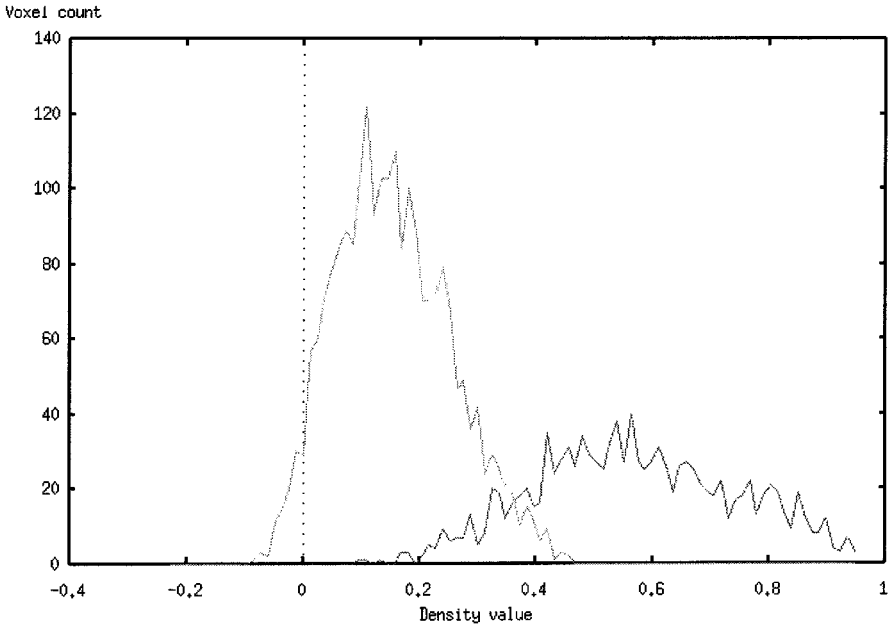

FIG. 4. Typical histograms of the reconstructed values in the foreground and in the surrounding background.

deviation in the phantom is well reproduced in the reconstruction):

$$\mathrm{sc}\sigma_{\mathrm{FOM}}(R) = 1 - |\sigma_{p_R} - \sigma_{r_R}|.$$

- Range FOM (the agreement between the voxel value ranges in the phantom and in the reconstruction):

$$\mathrm{sc}\Delta_{\mathrm{FOM}}(R) = 1 - \frac{1}{2}\left(|M_{p_R} - M_{r_R}| + |m_{p_R} - m_{r_R}|\right).$$

Below we report on these FOMs for three different regions: the whole volume V, the foreground F, and the background B. We also report on other structural consistency measures that have been designed so that errors are weighted differently depending on where they appear.

- Blurring FOM (emphasizes accuracy near the foreground using a division by d_i, which is the Euclidean distance of the center of the ith voxel in the background from the voxel in the foreground that is nearest to it):

$$\mathrm{scbl}_{\mathrm{FOM}} = 1 - \frac{1}{N_B}\sum_{i=1}^{N_B}\frac{1}{d_i}\left(\frac{p_{B,i} - r_{B,i}}{2}\right)^2.$$

- New masses appearing FOM (emphasizes accuracy far from the foreground and should thus be able to detect long elongations):

$$\mathrm{scap}_{\mathrm{FOM}} = 1 - \frac{1}{N_B}\sum_{i=1}^{N_B} d_i\left(\frac{p_{B,i} - r_{B,i}}{2}\right)^2.$$

- Radon FOM (measures accuracy of averages in horizontal layers $\pi(k)$ of voxels, $1 \le k \le K$, and is thus appropriate to detect vertical elongations):

$$\mathrm{scrt}_{\mathrm{FOM}} = 1 - \frac{1}{K}\sum_{k=1}^{K}|\mu_{p_{\pi(k)}} - \mu_{r_{\pi(k)}}|.$$

Notice that this measure is equivalent to the comparison of the Radon transforms of the phantom and the reconstruction along the vertical direction.

2.4.2. Structural separability measures. Structural separability is the issue of how well the foreground can be distinguished from the surrounding background. Foreground/background separation cannot be achieved by a simple comparison of voxel density values since, as Fig. 4 shows, it is common that in a reconstruction an overlap exists between the distributions of voxel densities in the foreground and in the background. We have designed a number of FOMs to measure relevant aspects of foreground/background separability in the reconstructions of our analytically generated phantoms.

- Foreground mean separability

We wish to measure the distinguishability of the foreground from the immediately surrounding background. For this purpose we associate with each of the two cylinders of the phantom four parts: an inner shell, a medial shell, an outer shell, and a surrounding background (Fig. 5 shows these four parts for one of the two cylinders of the phantom). We intend to investigate the separability of the reconstructed voxel values in the inner shell (respectively in the outer shell) from those in the surrounding background.

To do this, let I consist of all the voxels whose center is in the inner shell of a cylinder, O consist of all the voxels in the foreground whose center is in the outer shell of a cylinder, and S consist of all the voxels in the background whose center is in the surrounding background of a cylinder (as defined in Fig. 5). Assuming that voxel densities follow a gaussian distribution, a standard measure of separability of two means can be calculated using a Student's t test (Vardeman, 1994, pp. 307–311). Thus, we obtain the FOMs

$$\mathrm{hsin}_{\mathrm{FOM}} = \frac{|\mu_{\mathrm{I}} - \mu_{\mathrm{S}}|}{\sqrt{(\sigma_{\mathrm{I}}^2/N_I) + (\sigma_{\mathrm{S}}^2/N_S)}} \quad \text{and}$$

$$\mathrm{hsbr}_{\mathrm{FOM}} = \frac{|\mu_{\mathrm{O}} - \mu_{\mathrm{S}}|}{\sqrt{(\sigma_O^2/N_O) + (\sigma_S^2/N_S)}}.$$

- Detectability error FOM

This FOM quantifies the error that would have been committed if the foreground were segmented from the background based

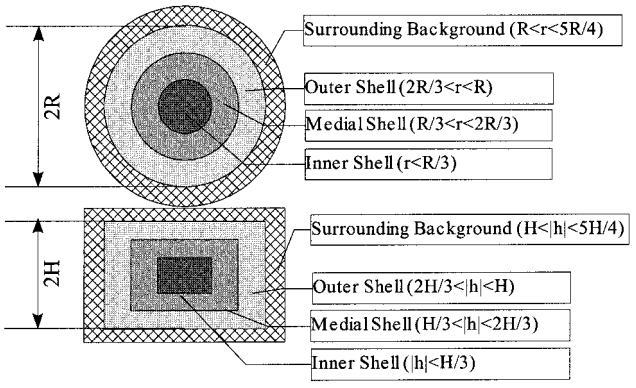

FIG. 5. Horizontal and vertical central sections of a cylinder.

only on the voxel density values. This measure is evaluated by calculating dterr, which is the total area under both the histogram of reconstructed voxel values in the background and the histogram of the reconstructed voxel values in the foreground after both histograms have been normalized so that the area under the histogram is unity:

$$\text{hsdt}_{\text{FOM}} = 1 - \text{dterr}.$$

- Vertical resolution FOM

The vertical resolution FOM is based on the one proposed by Marabini *et al.* (1997). Given the two cylinders in the phantom, three planes can be defined, two of them intersecting the cylinders through their centers and the third one between them (see Fig. 6). Let P_1 and P_2 be the set of voxels in the foreground that are intersected by planes 1 and 2, respectively. Let P_3 be the set of voxels (necessarily in the background) that are intersected by plane 3. The vertical resolution measure is the defined as

$$\text{hsvr}_{\text{FOM}} = \frac{\mu_{P_1} + \mu_{P_2} - 2\mu_{P_3}}{\sqrt{\sigma^2_{P_1} + \sigma^2_{P_2} + 4\sigma^2_{P_3}}}.$$

3. RESULTS

In this section we report on our experimental study of the behavior of the three algorithms (in the cases of even and uneven distributions of projections), first, using analytical phantoms, then quantum mechanical simulations, and, finally, real data.

The goal of the work with the analytical phantoms is twofold. On the one hand, it is vital to identify optimal ranges of the free parameters of the reconstruction algorithms that are being compared in this study (ART, SIRT, and WBP). On the other hand, we seek a set of evaluations using a broad range of FOMs (those defined in Section 2.4) that indicate the relative merits of the different methods that are being compared. After identifying the optimal range of parameters and carrying out a phantom level comparison, we apply the reconstruction algorithms to the other two cases: to data obtained by quantum mechanical simulation and to the set of data that was first used by Boisset *et al.* (1998) in the work on the effects of an uneven distribution of projection

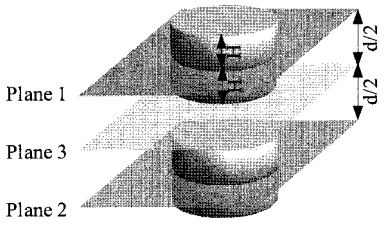

FIG. 6. Plane definition for the vertical resolution FOM. Planes 1 and 2 are orthogonal to the cylinder direction and pass through the centers of each cylinder. Plane 3 is defined parallel to 1 and 2 and passing through the middle point between the two cylinders.

directions on 3D reconstructions obtained by the different algorithms.

3.1. Statistical Analysis Performed Using Analytical Phantoms

Phantoms as described in Section 2.3.1 were generated with a random center between (−5, −5, −5) and (5, 5, 5), a random cylinder radius and height between 5 and 8, and a random separation of cylinders between 3 and 6. Projections were calculated for 541 evenly distributed directions (even case) and 1000 additional directions that were randomly distributed on the small sector with a tilt angle of 15° shown in Fig. 7 (uneven case). Thus, nearly 70% of the projections are within the overloaded region. The calculated projected density values as well as the geometrical information available to the algorithm regarding the location of the projection plane were corrupted by noise as it was done by Marabini *et al.* (1998). Concisely, zero-mean Gaussian noise was added with a SNR of around 0.66 for the density values, with a standard deviation of 2 pixels for the center positions and with a standard deviation of 5° for the angular locations. The even case is used as a control study in order to understand better the behavior of the reconstruction algorithms under conditions of uneven distributions: if FOMs are to de-

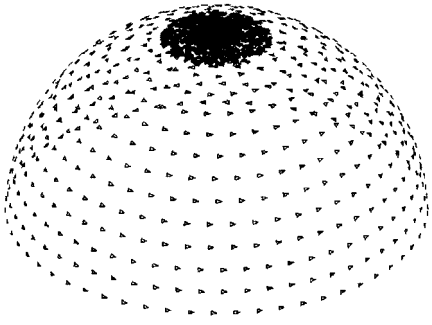

FIG. 7. Perspective view of the topology sphere corresponding to an uneven distribution of projections.

tect an elongation, we need to know what these FOMs are in case there is no elongation present.

For the purpose of finding the optimal regions of free parameters, in this study we selected $scL2_{FOM}(F)$ to be the tFOM. For each choice of the free parameters a total of 30 reconstructions were carried out, changing each time the phantom and the noise realization. The amount of computation involved in this approach has been very large, reaching the order of 420 h of a SGI Power Challenge with MIPS 10 000 at 200 MHz in the case of ART and 1000 h for SIRT.

The results of this search for the optimal region of free parameters for ART and SIRT are shown in Fig. 2. For ART, the relaxation parameter λ was optimized for each cycle through the data, observing small changes between the training FOM for one and two iterations; therefore, we did not proceed with further iterations. The optimal intervals are also similar for both cycles through the data (even case: $\lambda_1 \in [0.015, 0.07]$, $\lambda_2 \in [0.1, 0.6]$; uneven case: $\lambda_1 \in [0.01, 0.78]$, $\lambda_2 \in [0.01, 0.7]$) although for the second cycle through the data, the optimal range is smaller in both cases. Figures 2a and 2b show the optimal regions in the first cycle through the data for the case of an even distribution of projection directions and for the case of an uneven distribution, respectively.

For SIRT, a combination of two factors, relaxation parameter and number of iterations, need to be considered, too. Since previous experience indicates that SIRT needs a much larger number of cycles through the data, consideration of computational cost suggests the use of a constant relaxation parameter for all iterations. Figures 2c and 2d show the optimal regions for the even distribution and the uneven distribution. In both cases, the left boundary of the optimal region is similar. However, the right side of the optimal region in the even case is wider ($\lambda = 10^{-4}$) than in the uneven case ($\lambda = 10^{-4.5}$). This is a consequence of the SIRT implementation in SPIDER (Frank *et al.*, 1996), which is what we used in our experiments. A point that must always be considered in comparisons is that SIRT will typically require at least an order of magnitude more computer time than either ART or WBP.

WBP has a free parameter related to some form of low-pass filtration. In this work we have used an implementation based on the development by Radermacher (1992) for which we had already demonstrated (Marabini *et al.*, 1998) that the low-pass filter parameter does not play a substantial role under the conditions in which the simulations were done. Low-pass filter parameters within the optimal region shown in Marabini *et al.* (1998) have been selected in the evaluation experiments reported below.

For evaluation, a series of 120 test reconstructions with each method (ART, SIRT, and WBP) was performed using the analytical phantoms. The free parameters used for the 120 tests were randomly picked from their optimal region using a uniform distribution (this approach was proposed by Matej *et al.*, 1996). The performance of ART, SIRT, and WBP is presented in Tables I and II. Specifically, Table I presents the mean and standard deviation of the values of each FOM for each of the three algorithms under an even and an uneven distribution of projection directions. Also, the cases in which the algorithm performs significantly better when the number of projections is larger despite the fact that the distribution of projection directions was uneven are indicated by an up-arrow; the cases in which the results deteriorate when the angular distribution of projections is not even are indicated by a down-arrow.

A ranking of the performance of ART, SIRT, and WBP with respect to the set of FOMs defined in Section 2.4 is presented in Table II, both for the even and for the uneven distributions of projection directions.

3.2. Qualitative Algorithm Comparison

Accurate simulations of the data collection in 3DEM were obtained using the quantum mechanical approach described in Dinges and Rose (1995). Images for normal EM conditions were simulated for an artificial protein complex embedded in ice as described in Section 2.3.2. A large set of projections (3099) was computed for the same distribution of projection directions (shown in Fig. 7) as that reported by Boisset *et al.* (1998) in their study of the giant hemoglobin of *L. terrestris*. The following set of conditions was used in the simulations: elastic image contrast (i.e., no inelastic interactions are taken into account), 2 μm of defocus, acceleration voltage 120 kV, spherical aberration 2.0 mm, source spread 0.75 eV, aperture 12.0 mrad, focus spread 0.0 μm, ice thickness 100 nm $\pm$ 30 Å. The choices of free parameters for this experiment were $\lambda_1 = 0.047$, $\lambda_2 = 0.03$ for ART, and for SIRT two different cases were considered: the first with a worse choice ($\lambda_{SIRT} = 5 \times 10^{-7}$, 100 iterations) and the second with a better choice ($\lambda_{SIRT} = 10^{-5}$, 100 iterations). The results of the 3D reconstruction using ART, SIRT, and WBP are presented in Fig. 3, and they show effectively the elongation effect obtained with SIRT with inferior parameters. At the same time, the nonelongated reconstruction produced by ART and WBP can be seen to have a small artifact in the middle of both rings.

TABLE I

FOM Values for the Three Reconstruction Algorithms in the Even (Upper Row) and Uneven (Lower Row) Cases

FOM	ART		SIRT		WBP	
$scL2_{FOM}(V)$	0.99784 ± 0.00026		0.99775 ± 0.00087	↑	0.99577 ± 0.00075	
	0.99771 ± 0.00034		0.99858 ± 0.00028		0.99554 ± 0.00076	
$scL1_{FOM}(V)$	0.9651 ± 0.0020		0.9682 ± 0.0098	↑	0.9547 ± 0.0026	↓
	0.9645 ± 0.0023		0.9790 ± 0.0038		0.9503 ± 0.0030	
$sc\mu_{FOM}(V)$	0.9726 ± 0.0013	↑	0.9976 ± 0.0069		0.9649 ± 0.0019	
	0.9736 ± 0.0023		0.9996 ± 0.0003		0.9645 ± 0.0012	
$sc\Delta_{FOM}(V)$	0.836 ± 0.030		0.778 ± 0.102	↑	0.715 ± 0.062	↑
	0.819 ± 0.045		0.882 ± 0.048		0.848 ± 0.015	
$scL2_{FOM}(B)$	0.99824 ± 0.00023		0.99841 ± 0.00092	↑	0.99632 ± 0.00070	↑
	0.99815 ± 0.00028		0.99930 ± 0.00023		0.99726 ± 0.00036	
$scL1_{FOM}(B)$	0.9671 ± 0.0018		0.9713 ± 0.0102	↑	0.9573 ± 0.0023	
	0.9667 ± 0.0021		0.9827 ± 0.0039		0.9560 ± 0.0016	
$sc\mu_{FOM}(B)$	0.9692 ± 0.0015		0.9945 ± 0.0070		0.9622 ± 0.0023	↓
	0.9701 ± 0.0023		0.9959 ± 0.0009		0.9560 ± 0.0016	
$sc\sigma_{FOM}(B)$	0.9435 ± 0.0064		0.9261 ± 0.0243	↑	0.9904 ± 0.0114	↓
	0.9392 ± 0.0099		0.9483 ± 0.0086		0.9438 ± 0.0077	
$sc\Delta_{FOM}(B)$	0.658 ± 0.031		0.605 ± 0.110	↑	0.398 ± 0.085	↑
	0.634 ± 0.052		0.699 ± 0.044		0.667 ± 0.027	
$scL2_{FOM}(F)$	0.9847 ± 0.0032		0.9740 ± 0.0132		0.9773 ± 0.0055	↓
	0.9842 ± 0.0027		0.9732 ± 0.0074		0.9325 ± 0.0085	
$scL1_{FOM}(F)$	0.901 ± 0.013		0.865 ± 0.038		0.871 ± 0.012	↓
	0.900 ± 0.010		0.859 ± 0.026		0.749 ± 0.017	
$sc\mu_{FOM}(F)$	0.916 ± 0.017		0.874 ± 0.041		0.922 ± 0.033	↓
	0.917 ± 0.016		0.862 ± 0.029		0.751 ± 0.019	
$sc\sigma_{FOM}(F)$	0.829 ± 0.023		0.834 ± 0.032		0.740 ± 0.053	↑
	0.821 ± 0.021		0.841 ± 0.016		0.875 ± 0.008	
$sc\Delta_{FOM}(F)$	0.595 ± 0.058		0.573 ± 0.118		0.461 ± 0.099	
	0.576 ± 0.051		0.617 ± 0.059		0.467 ± 0.096	
$scbl_{FOM}$	0.99936 ± 0.00012		0.99951 ± 0.00021		0.99828 ± 0.00057	↑
	0.99931 ± 0.00015		0.99964 ± 0.00009		0.99914 ± 0.00021	
$scap_{FOM}$	0.9867 ± 0.0019		0.9870 ± 0.0091	↑	0.9798 ± 0.0028	↓
	0.9867 ± 0.0021		0.9960 ± 0.0020		0.9786 ± 0.0008	
$scrt_{FOM}$	0.9719 ± 0.0012		0.9952 ± 0.0059		0.9653 ± 0.0017	↓
	0.9709 ± 0.0018		0.9971 ± 0.0006		0.9565 ± 0.0016	
$hsin_{FOM}$	85.52 ± 29.39		75.65 ± 31.92		24.69 ± 9.17	↑
	90.55 ± 36.99		88.44 ± 27.26		32.35 ± 13.63	
$hsbr_{FOM}$	110.38 ± 33.96		94.01 ± 46.75		54.67 ± 26.71	
	109.39 ± 30.29		110.21 ± 32.79		65.38 ± 38.82	
$hsdt_{FOM}$	0.999999 ± 0.000005		0.999998 ± 0.000008		0.999989 ± 0.000023	
	0.999998 ± 0.000008		0.999999 ± 0.000004		0.999991 ± 0.000002	
$hsvr_{FOM}$	6.70 ± 0.51		5.70 ± 0.66	↑	2.89 ± 0.46	
	6.34 ± 0.64		6.24 ± 0.45		2.91 ± 0.43	

Note. Arrows indicate whether the uneven value is significantly better (↑) or significantly worse (↓) than the even case.

The second experimental data set used in the present work for qualitative comparisons corresponds to a series of 3099 images of *L. terrestris* extracellular respiratory pigment. These images were recorded on a Philips CM200 field emission gun cryoelectron microscope with an acceleration voltage of 200 kV, a magnification of 63 240×, and a defocus ranging from 2 to 3.5 μm. For unknown experimental reasons two-thirds of the particles were oriented in the vitreous ice layer with their sixfold symmetry axis perpendicular to the air–water interface. Therefore, this experimental data set corresponds to a typical case of uneven angular sampling, with overabundant hexagonal top views. In fact, this was the original data set for which the first concerns about uneven projection distributions were raised by Boisset *et al.* (1998). The angular distribution and algorithm parameters are exactly the same as in the previous atomic phantom. An additional experiment was performed with SIRT in this particular case: the inferior λ ($\lambda_{SIRT} = 5 \times 10^{-7}$), but a larger number of iterations (1000). Figure 8 shows the results obtained. The three reconstructions with the correctly chosen parameters seem to be equivalent, while the two SIRT reconstructions with a nonoptimal relaxation parameter appear to be elon-

TABLE II
Algorithm Ranking under the Different FOMs and Cases

	Even case			Uneven case		
FOM	1st	2nd	3rd	1st	2nd	3rd
$scL2_{FOM}(V)$	ART SIRT		WBP	SIRT	ART	WBP
$scL1_{FOM}(V)$	SIRT ART		WBP	SIRT	ART	WBP
$sc\mu_{FOM}(V)$	SIRT	ART	WBP	SIRT	ART	WBP
$sc\Delta_{FOM}(V)$	ART	SIRT	WBP	SIRT	WBP	ART
$scL2_{FOM}(B)$	SIRT ART		WBP	SIRT	ART	WBP
$scL1_{FOM}(B)$	SIRT	ART	WBP	SIRT	ART	WBP
$sc\mu_{FOM}(B)$	SIRT	ART	WBP	SIRT	ART	WBP
$sc\sigma_{FOM}(B)$	ART	SIRT	WBP	SIRT	WBP	ART
$sc\Delta_{FOM}(B)$	ART	SIRT	WBP	SIRT	WBP ART	
$scL2_{FOM}(F)$	ART	SIRT WBP		ART	SIRT	WBP
$scL1_{FOM}(F)$	ART	WBP SIRT		ART	SIRT	WBP
$sc\mu_{FOM}(F)$	WBP ART		SIRT	ART	SIRT	WBP
$sc\sigma_{FOM}(F)$	SIRT ART		WBP	WBP	SIRT	ART
$sc\Delta_{FOM}(F)$	ART SIRT		WBP	SIRT	ART	WBP
$scbl_{FOM}$	SIRT	ART	WBP	SIRT	ART	WBP
$scap_{FOM}$	ART SIRT		WBP	SIRT	ART	WBP
$scrt_{FOM}$	SIRT	ART	WBP	SIRT	ART	WBP
$hsin_{FOM}$	ART SIRT		WBP	ART SIRT		WBP
$hsbr_{FOM}$	ART SIRT		WBP	SIRT ART		WBP
$hsdt_{FOM}$	ART SIRT		WBP	SIRT ART		WBP
$hsvr_{FOM}$	ART	SIRT	WBP	ART SIRT		WBP

Note. When two methods share a place, it means that neither is significantly better than the other; however, they are ordered by FOM value even if they are not significantly different.

gated in the overloaded direction. This elongation artifact is stronger for the reconstruction done with a smaller number of iterations.

4. DISCUSSION

An objective comparison based on FOMs indicates that, for the particular variants of the algorithms that we investigated, ART and SIRT outperform WBP in both cases studied in this work (Table II). ART and SIRT perform quite similarly, although Table II seems to indicate that for the uneven case SIRT tends to behave slightly better than ART (note that the computational cost of SIRT is more than an order of magnitude greater than that of ART; this raises the possibility of improving the performance of ART by running it longer, but still at a cost much less than that of SIRT). Also, SIRT shows an improvement that is consistent with the fact that the uneven distribution uses 1000 more projections than the even distribution, which means that there is more information and, therefore, the reconstruction should be better. Nevertheless, WBP shows an unstable behavior when changing from the even to the uneven case: some FOMs improve and others deteriorate. As for ART, its performance remains essentially unaffected by the use of additional data, except for one FOM that indicates significant improvement.

The present study demonstrates that the parameter selection used in previous structural studies (Boisset *et al.*, 1998) using SIRT was not appropriate. The elongation artifact that was reported in the article cited was explained as the result of giving more importance to the views in the preferred direction than to the others. While this intuitive interpretation is probably correct for the early stages of the algorithm, it is no longer true at the later stages. Indeed, when SIRT is run with the appropriate relaxation parameter, the artifacts do not appear. Further, even when using too small values for the relaxation parameter—as was done by Boisset *et al.* (1998)—by letting the algorithm run for a very large number of iterations, the artifact fades away. We conclude that (a) the proper choice of the free parameters of the algorithms is critical and (b) if the parameters are well chosen, an overabundant knowledge of the volume from one direction will not result in inferior reconstructions.

5. CONCLUSIONS

In this work we have extended our previous task-oriented evaluation of algorithmic behavior toward an explicit incorporation of the variance of the Figures of Merit used to evaluate the various tasks and, at the same time, a more complete set of Figures of Merit has been devised. This methodology has been

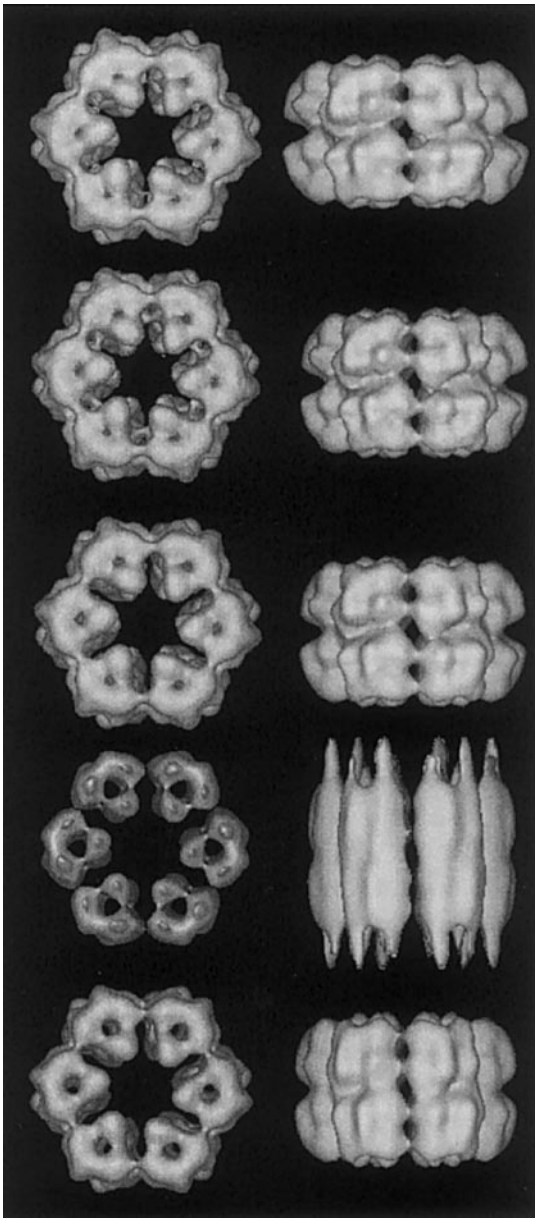

FIG. 8. 3D reconstruction of the giant hemoglobin of *Lumbricus terrestris* using an uneven distribution of projection directions with 3099 projections. From top to bottom: reconstruction with ART, reconstruction with SIRT, reconstruction with WBP, reconstruction with SIRT using nonoptimal parameters, reconstruction with SIRT with the same nonoptimal parameter but with 1000 iterations.

applied to the assessment of the relative performance of three reconstruction algorithms, ART, SIRT, and WBP, under conditions in which the angular distribution of projection directions has been either even or uneven. We have shown that previous reports of SIRT behaving very poorly for the case of uneven distribution of projection directions were based on a nonoptimal choice of the free parameters of the algorithm (relaxation parameter and number of iterations). It is thus established that free parameters have a strong influence on the final reconstruction and so it is undoubtedly worthwhile to optimize them for a particular type of application. Furthermore, we have shown that, for the case of an uneven angular distribution of projections, both ART and SIRT outperform WBP when the free parameters have been properly selected, although ART does so at a fraction of the computational cost (between one and two orders of magnitude) required by SIRT.

We are grateful to Dr. Uwe Lüken (FEI) and Dr. Dieter Typke (Max-Planck Institut für Biochemie) for letting us use the Philips CM200 FEG cryoelectron microscope in Martinsried (Germany). Dr. Marabini is partially supported by a grant from the Spanish "Ministerio de Educación y Ciencia." Partial support was provided by the "Comisión Interministerial de Ciencia y Tecnología" of Spain through Project BIO98-0761, the Intergovernmental French/Spanish Collaborative Action HF1998-0065, and National Institutes of Health Grant HL28438.

REFERENCES

Boisset, N., Penczek, P., Taveau, J. C., You, V., de Haas, F., and Lamy, J. (1998) Overabundant single-particle electron microscope views induce a three-dimensional reconstruction artifact, *Ultramicroscopy* **74**(4), 201–207.

Boisset, N., Taveau, J. C., Pochon, F., and Lamy, J. (1996) Similar architectures of native and transformed human alpha 2-macroglobulin suggest the transformation mechanism, *J. Biol. Chem.* **271**(42), 25762–25769.

de Haas, F., Zal, F., Lallier, F. H., Toulmond, A., and Lamy, J. (1996) Three-dimensional reconstruction of the hexagonal bilayer hemoglobin of the hydrothermal vent tube worm *Riftia pachyptila* by cryoelectron microscopy, *Proteins* **26**(3), 241–256.

Dinges, C., and Rose, H. (1995) Simulation of filtered and unfiltered TEM images and diffraction patterns, *Phys. Stat. Sol.* **150**(1), 23–29.

Frank, J. (1996) Three-Dimensional Electron Microscopy of Macromolecular Assemblies, Academic Press, San Diego.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields, *J. Struct. Biol.* **116**(1), 190–199.

Herman, G. T. (1980) Image Reconstructions from Projections: The Fundamentals of Computerized Tomography, Academic Press, New York.

Holmes, K. C., Popp, D., Gebhard, W., and Kabsch, W. (1990) Atomic model of the actin filament, *Nature* **347**(6288), 44–49.

Marabini, R., Herman, G. T., and Carazo, J. M. (1998) 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs), *Ultramicroscopy* **72**(1–2), 53–65.

Marabini, R., Rietzel, E., Schroeder, R., Herman, G. T., and Carazo, J. M. (1997) Three-dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt schema: Application of a new three-dimensional reconstruction algorithm and objective comparison with weighted backprojection, *J. Struct. Biol.* **120**(3), 363–371.

Matej, S., Furuie, S. S., and Herman, G. T. (1996) Relevance of statistically significant differences between reconstruction algorithms, *IEEE Trans. Image Proc.* **5**(3), 554–556.

Matej, S., Herman, G. T., Narayan, T. K., Furuie, S. S., Lewitt, R. M., and Kinahan, P. E. (1994) Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms, *Phys. Med. Biol.* **39,** 355–367.

Obi, T., Matej, S., Lewitt, R., and Herman, G. T. (2000) 2.5-D simultaneous multislice reconstruction by series expansion methods from Fourier-rebinned PET data, *IEEE Trans. Med. Imaging* **19**(5), 474–484.

Penczek, P., Radermacher, M., and Frank, J. (1992) Three-dimensional reconstruction of single particles embedded in ice, *Ultramicroscopy* **40**(1), 33–53.

Radermacher, M. (1992) Weighted back-projection methods, *in* Frank, J. (Ed.), Electron Tomography, Plenum, New York.

Vardeman, S. B. (1994) Statistics for Engineering Problem Solving, IEEE Press, New York.