Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Structural Biology 146 (2004) 381–392
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

A multiresolution approach to orientation assignment in 3D electron microscopy of single particles

C.O.S. Sorzano,[a,b,c,*] S. Jonić,[b] C. El-Bez,[d] J.M. Carazo,[c] S. De Carlo,[e] P. Thévenaz,[b] and M. Unser[b]

[a] *Escuela Politécnica Superior, Universidad San Pablo-CEU, Campus Urb., Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, Spain*
[b] *Biomedical Imaging Group, Swiss Federal Institute of Technology Lausanne., CH-1015 Lausanne VD, Switzerland*
[c] *Biocomputing Unit, Centro Nacional de Biotecnología (CSIC), Campus Universidad Autónoma s/n, 28049 Cantoblanco, Madrid, Spain*
[d] *Laboratoire d'analyse ultrastructurale, Université de Lausanne., CH-1015 Lausanne VD, Switzerland*
[e] *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 1, rue Laurent Fries, BP 10142, 67404 Illkirch Cedex, France*

Received 19 November 2003, and in revised form 13 January 2004

Abstract

Three-dimensional (3D) electron microscopy (3DEM) aims at the determination of the spatial distribution of the Coulomb potential of macromolecular complexes. The 3D reconstruction of a macromolecule using single-particle techniques involves thousands of 2D projections. One of the key parameters required to perform such a 3D reconstruction is the orientation of each projection image as well as its in-plane orientation. This information is unknown experimentally and must be determined using image-processing techniques. We propose the use of wavelets to match the experimental projections with those obtained from a reference 3D model. The wavelet decomposition of the projection images provides a framework for a multiscale matching algorithm in which speed and robustness to noise are gained. Furthermore, this multiresolution approach is combined with a novel orientation selection strategy. Results obtained from computer simulations as well as experimental data encourage the use of this approach.

1. Introduction

The spatial distribution of the Coulomb potential of macromolecular complexes is crucial in structural biology and provides key information about the way macromolecules interact. It can be recovered by three-dimensional (3D) electron microscopy (3DEM) which addresses biological structures with sizes ranging from 100 Å to 1 μm (Frank, 1996, 2002; Kuhlbrandt and Williams, 1999; Unger, 2001).

The highest-resolution 3D models (3–10 Å) are obtained by electron crystallography (Kuhlbrandt and Williams, 1999; Unger, 2001). However, this technique is not suitable in case of large macromolecular complexes that often resist crystallization or that can only be partially crystallized, after removing their flexible parts (Frank, 2002). In this case, single-particle techniques are commonly used. These techniques have resulted in reconstructions with medium-to-high resolutions (5–15 Å) for a number of different specimens (van Heel et al., 2000).

The 3D reconstruction of a macromolecule following the single-particle approach is based on the information provided by thousands of two-dimensional (2D) projections of objects that are nearly identical but that have completely unknown and independent orientations. However, the relative orientation of each particle is needed by the reconstruction algorithm.

The orientation or pose (as also known in the Computer Vision community) of a particle in the microscope is specified by a projection direction and by an in-plane pose, which, in turn, is given by an in-plane rotation and translation. The projection direction as well as the in-plane rotation are usually coded using three Euler angles (two out-of-plane rotations defining the projection direction, and the in-plane rotation).

The techniques used for determining the pose of single-particle projections can be categorized in two

* Corresponding author. Fax: +34-91-585-4506.
E-mail address: coss@cnb.uam.es (C.O.S. Sorzano).

1047-8477/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2004.01.006

main families: reference-free and reference-based. The method of moments (Basu and Bresler, 2000a,b; Gelfand and Goncharov, 1990; Goncharov, 1990; Salzman, 1990) is reference-free and uses a known relationship between the area moments of the 3D object and the moments of its 2D projections. The common-line method is another reference-free algorithm based on the property that two projections of the same object from different directions must share a common line in Fourier space (Central-Slice Theorem). The search of the common line can be done by comparing the projection sinograms in the image space (Goncharov, 1990; Penczek et al., 1994; van Heel, 1987) or in the Fourier space (Goncharov, 1990; Lauren and Nandhakumar, 1997). The approach of Lauren and Nandhakumar (1997) is particularly interesting since shifts and noise are explicitly considered during the common-line search. However, these techniques remain very sensitive to noise. When used in single-particle electron microscopy, they are usually combined with classification algorithms because of the noise reduction effect observed in the class averages or representatives (van Heel et al., 2000).

The second group of techniques relies on a reference 3D model. The underlying idea of this family of methods is to match the projections of the reference model with those obtained in the microscope. The two reference-based methods most commonly used in single-particle electron microscopy are the ones developed by Penczek et al. (1994) and by Radermacher (1994). The method proposed by Penczek et al. (1994) is based on a library of projections of the reference volume (from now on referred to as reference projections) that are evenly distributed over the entire range of the two out-of-plane angles. It computes one-dimensional cross-correlation functions in polar coordinates between the experimental images and all the reference projections. The two out-of-plane rotation angles are determined by the largest cross-correlation coefficient and the in-plane rotation angle is determined by the position of the maximum in the corresponding cross-correlation function. Notice that the reference images and the experimental ones are considered to have been previously translationally aligned. In this paper, we will refer to this method as the *real-space matching*. The method introduced by Radermacher (1994) computes a five-dimensional (5D) cross-correlation function (function of five parameters: three angles and two translations) between the 2D Radon transform (RT) of each experimental image and the 3D RT of the volume. The parameters for which this function achieves its maximum define the pose of the particle image. This method will be referred to as the *Radon-based assignment*.

All techniques exposed so far assume that the angular-assignment step is performed independently of the reconstruction step. One advantage of doing so is that any reconstruction algorithm can be used. However, the two steps are sometimes performed simultaneously, as in Provencher and Vogel (1988) and Vogel and Provencher (1988).

Our algorithm is also independent of the reconstruction step. It is based on a library of projections of the input volume, like in the real-space matching method. The correlation coefficient between each experimental image and all the reference projections is computed in a coarse-to-fine fashion using a discrete wavelet transform. This reduces the computation complexity and increases robustness with respect to noise. Furthermore, the selection of the pose is not based on the sole correlation maximum, but also on a set of highly correlated reference images. How to build this set of reference images that correlate well with the experimental image will be explained in more detail later. We will refer to the proposed method as the *wavelet-space matching*.

Saad and Chiu (2000) also proposed the use of wavelets to compute the similarity between two images. However, the wavelet decomposition was used in that work to assign different weights to frequency components. The reference projections were first classified into classes. Any experimental image was first compared with a representative of each class and, then, compared to all the reference images within the three classes with highest correlation.

In this paper, we introduce a wavelet-space matching algorithm that is completely different from the approach of Saad and Chiu (2000), and so is the comparison strategy. We compare its performance with the real-space matching and the Radon-based assignment as implemented in the SPIDER package (Frank et al., 1996). This comparison is performed using synthetic data. The performance of our algorithm is also shown on experimental data. Our method is available in the Xmipp package (Marabini et al., 1996) (http://www.cnb.uam.es/~bioinfo/).

2. Angular-assignment algorithm

The proposed algorithm is based on the alignment of the experimental images to those in a library of simulated projections computed using a reference model. The alignment is done by comparing the correlation coefficient of the experimental image whose angles are to be assigned with all the reference projections as in Penczek et al. (1994). Here is a brief description of the algorithm:

1. For each reasonable in-plane pose of the experimental image do:
   1.1. Search the image in the library that best matches (based on correlation) the experimental image.

2. Select the translational and angular parameters that correspond to the most likely image pose.

The main departures from the approach of Penczek et al. (1994) are:

- The reference projection that better matches a given in-plane pose of the experimental image at hand (step 1.1 in the algorithm) is searched in a coarse-to-fine fashion using a wavelet transform. The coarse-to-fine strategy is used to compute partial approximations to the correlation coefficient. The experimental image at a given in-plane pose is compared with all the reference projections at the coarsest resolution. Only those reference projections with a sufficiently high partial correlation pass to the next stage. At the next stage, the resolution is increased and another sub-band of the images are compared at this resolution level. Again, only those with a sufficiently high partial correlation progress to the next stage. This process is iterated until all sub-bands within the same resolution level have been included. At this point, the resolution is increased again, and new sub-bands are considered. When all sub-bands have been considered, we select the reference projection with the highest correlation. The advantages of this approach are twofold: first, most of the computation burden is avoided since, for most reference projections, the correlation is computed only at low resolution; second, this strategy is more robust to noise.
- Step 1 in our algorithm provides the best reference projection for each one of the in-plane poses of the experimental image explored. However, a decision must be taken about which one of the in-plane poses is the most likely. This is done in step 2. The best-matching reference projection, and therefore the in-plane pose of the experimental image, is selected based on the angular information of several of the best-matching reference projections, instead of simply assigning the pose parameters of the best-matching reference image. The underlying idea is to consider the angular distribution of a few good-matching images instead of a single one.

The computation of the correlation coefficient in the wavelet space is introduced in Section 2.1. The multiresolution properties of the wavelet transform are exploited in Section 2.2. Finally, the novel pose selection strategy is presented in Section 2.3.

2.1. Correlation in the wavelet space

The goal of this section is to derive a formula that allows the computation of the correlation between two images in the wavelet space. It will be shown that under some assumptions the correlation can be easily computed in terms of the coefficients of the wavelet expansion of the input images. For simplicity purposes, we will consider one-dimensional signals here. The extension to higher dimensions is straightforward.

The correlation function between two signals f and g is usually defined as $R_{f,g}(\tau) = \int_{-\infty}^{\infty} f(t)g(t+\tau)\,\mathrm{d}t$ and is interpreted as the correlation of the function f with every shifted version of g. Without loss of generality we will concentrate on computing the correlation function at the origin

$$R_{f,g}(0) = \int_{-\infty}^{\infty} f(t)g(t)\,\mathrm{d}t = \langle f, g\rangle_{L_2}. \tag{1}$$

Given a wavelet function ψ, any signal $f \in L_2$ (L_2 is the space of functions with bounded square integral) can be expressed as

$$f(t) = \sum_{i,s\in\mathbb{Z}} d_{f,s}[i]\psi_{i,s}(t), \tag{2}$$

where $\psi_{i,s}(t) = \frac{1}{\sqrt{2^s}}\psi((t-2^s i)/2^s)$ and $d_{f,s}[i] = \langle f, \psi_{i,s}\rangle$ are the corresponding expansion coefficients of f Aldroubi and Unser (1996). The parameter s is referred to as the scale; it controls the resolution of the wavelet function while i determines its position. The defining property of an orthonormal wavelet basis is

$$\langle \psi_{i,s}, \psi_{j,s'}\rangle = \delta_{i-j}\delta_{s-s'},$$

where $\delta_i = 1 - |\mathrm{sign}(i)|$ is Kronecker's delta.

Plugging Eq. (2) into Eq. (1) and using the orthogonality property, it can be easily shown that

$$R_{f,g}(0) = \sum_{s\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} d_{f,s}[n]d_{g,s}[n] = \sum_{s\in\mathbb{Z}}\langle d_{f,s}, d_{g,s}\rangle_{l_2}. \tag{3}$$

In other words, the correlation between the two continuous functions f and g can be computed as the sum over all scales of the discrete correlation of their corresponding wavelet expansion coefficients. In the case of 2D signals, with a proper indexing scheme (Aldroubi and Unser, 1996) the proof is straightforward, the correlation between two continuous images can also be computed as the sum of the sub-band discrete correlations (coefficients corresponding to a particular selection of scale and orientation, see Fig. 1).

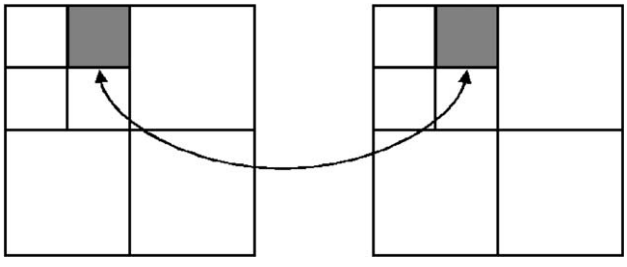

Fig. 1. The correlation between two continuous images is computed as the sum of their sub-band discrete correlations. Sub-bands are typically arranged by the wavelet transform as depicted in this figure: each of the squares represent a sub-band (low frequency sub-bands occupy smaller areas). The figure shows two corresponding sub-bands in two different images.

Finally, the correlation between the images is commonly normalized by their respective energy. This normalized correlation is referred to as the correlation coefficient and is defined as

$$\rho = \frac{R_{f,g}(0)}{\sqrt{R_{f,f}(0)R_{g,g}(0)}}. \quad (4)$$

We use the correlation coefficient to evaluate the similarity between two images.

2.2. Multiresolution projection matching

At this point we have an alternative way to compute the similarity between two images using their wavelet representations. However, the possibilities of the wavelet decomposition for multiresolution data processing have not been fully exploited yet. Multiresolution can help us to discard the reference projections that do not match the experimental projection at coarse data resolutions. Coarse resolutions provide a rough representation of the particle shape. Thus, if two projections do not match at these resolutions, then the particle shapes in the two images are clearly different. On the contrary, if they correlate well at these resolutions, finer details are added to the computation of the correlation (see Eq. (3)) in order to check whether this match still holds at finer resolutions.

Therefore, in our algorithm, Eq. (3) is not computed at once for each reference image, but the different inner products participating are summed sub-band by sub-band. At each sub-band, the images with the worse partial correlations are removed from the list of possible candidates for the best-matching reference. This process goes on until all sub-bands have been processed (see Fig. 2). When the last sub-band has been considered, there are typically only few reference projections surviving. At this stage the one with the highest correlation coefficient is selected as the one that best matches a specific in-plane pose of the experimental image. The proportion of discarded images at each sub-band is called the "discarding factor" and its effect on accuracy and computing time will be explored in Section 3.1.3. This procedure reduces the computational complexity since the sub-band correlations at coarser scales come much cheaper than those at finer scales. Similar approaches have been shown to be more robust to noise in other image-processing applications (Dengler, 1989; Desco et al., 2001; Thévenaz et al., 1998).

2.3. Pose selection strategy

The procedure described in the previous sections identifies the best-matching reference projection in the library for each reasonable in-plane pose of the experimental image. The next step is to select the most likely of all these poses. A common approach is to just select the pose with maximum correlation. However, this may be misleading because the correlation objective maximum can be very close to other correlation values. In this case, it is not so clear which pose to select since the correlation values are perturbed by the noise present in the experimental images.

The solution adopted in this work is to produce an angular assignment that is based on a few high-correlation values rather than on the sole correlation maximum. Let us consider the list of in-plane poses with highest correlations and their corresponding reference projections. It is possible that, in this list, we have clusters of reference projections with similar projection directions. If this is the case, it is likely that the true pose of the particle at hand is one that comes from the most populated cluster, even if the correlation maximum is not within this group.

To measure the similarity of the projection directions of two different reference images we make use of their corresponding Euler matrix. Given three Euler angles,

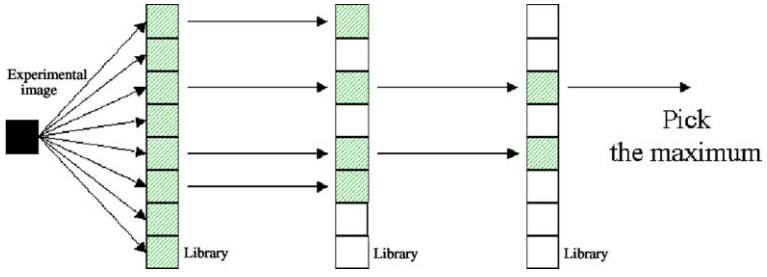

Fig. 2. In a first step, the coarsest sub-band of the experimental image is compared with the coarsest sub-bands of all the reference projections in the library. Then, the reference images with lowest correlations are discarded (in this figure the discarding factor is 50%). In the next stage, the comparison is updated using another sub-band. Again, those images with the lowest correlations are discarded. This process goes on until all the sub-bands are considered. In the last stage, the reference image with maximum correlation is selected as the best matching for a given translation and in-plane rotation of the experimental image.

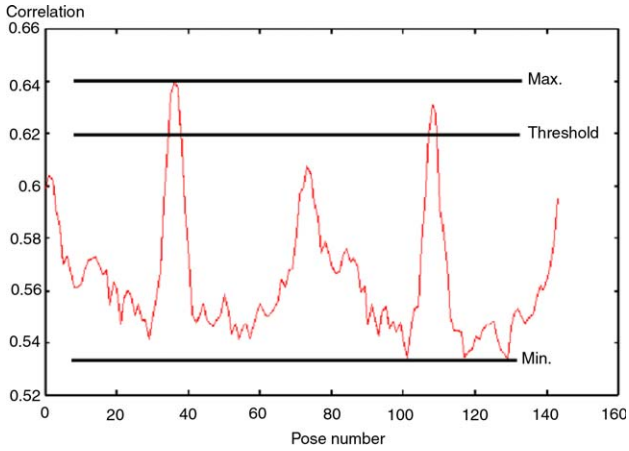

Fig. 3. The correlation coefficient as a function of the in-plane pose of the experimental image. The maximum and minimum correlations, as well as the threshold above which correlations are considered as high, are marked in the plot.

ϕ, θ, and ψ, representing rotations around z, y, and z, respectively, the associated Euler matrix is defined in Eq. (5)

$$E_{\phi,\theta,\psi} = \begin{pmatrix} \cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi & \cos\psi\cos\theta\sin\phi + \sin\psi\cos\phi & -\cos\psi\sin\theta \\ -\sin\psi\cos\theta\cos\phi - \cos\psi\sin\phi & -\sin\psi\cos\theta\sin\phi + \cos\psi\cos\phi & \sin\psi\sin\theta \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix} \quad (5)$$

The third row of this matrix provides the projection direction. The best-matching reference projections are clustered according to the Euclidean distance between their projection directions. In this way, image groups are formed such that the distance between any two projection directions in the group does not exceed a certain threshold. Once the most populated cluster has been identified, we pick from it the projection direction and in-plane rotation with the highest correlation coefficient.

We use the following procedure to build the list of highest-correlated reference images. The correlation range for an experimental image is defined as the difference between the maximum and minimum correlations found in Section 2.2. A correlation coefficient is said to be high if it is higher than 80% of the correlation range (see Fig. 3).

3. Results

We evaluated the performance of our algorithm in a fully controlled simulated environment using objective measures of the quality of the assignment. We also tested its performance using electron-microscopy experimental data. In particular, we used the cryo-negative staining data of GroEL obtained by De Carlo et al. (2002). This particle was selected because its atomic model is available; therefore, the assignment quality can also be established objectively.

3.1. Computer-simulated experiments

For the computer-simulated experiments, several volumes known at atomic coordinates were taken from the Macromolecular Structure Database (Protein Quaternary Structure query, PQS, Boutselakis et al., 2003). Thousand projection images were simulated with random projection directions (ϕ and ψ were uniformly distributed in $[0, 360]$, and θ was uniformly distributed in $[0, 180]$). The applied Euler angles were stored for posterior comparison. The effect of the microscope contrast transfer function (CTF) on the images was also considered within the simulation (accelerating voltage = 200 kV, defocus = −27,700 Å, spherical aberration = 2 mm, convergence cone = 0.21 mrad) (Frank, 1996; Velázquez-Muriel et al., 2003). A signal-to-noise ratio of 1/3 was simulated. These simulated projections were input to the angular-assignment algorithm under study.

3.1.1. Measure of the assignment quality

The assigned poses were compared with the true poses. The shift parameters were compared measuring the Euclidean distance of the assigned shift $(\hat{x}, \hat{y})$ to the true one (x, y), $\sqrt{(\hat{x}-x)^2 + (\hat{y}-y)^2}$. Notice that this accuracy is measured in pixels.

The quality of the angular assignment was evaluated using the following measure: given any three Euler angles (ϕ, θ, ψ), the rows $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ of their corresponding Euler matrix define a coordinate system attached to the rotated particle before projection (see Eq. (5)). We computed the average angular error between the axes determined by the true $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ and assigned $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ angles as $\frac{1}{3}\sum_{i=1}^{3} \arccos\langle \mathbf{e}_i, \hat{\mathbf{e}}_i \rangle$. Notice that this accuracy is measured in degrees. A perfect assignment should produce a zero-average angular error among axes.

A third indicator of the quality of an assignment is given by its robustness, that is, the proportion of particles that are reasonably assigned. A particle is reasonably assigned if its average angular error does not exceed a certain threshold. The 95% confidence interval of the proportion of reasonably assigned particles is reported for each experiment (Vardeman, 1994). The reported accuracy measures correspond to the average angular and translational errors for only the reasonably assigned particles.

3.1.2. Comparison experiments

We compared the performance of the proposed method with the performance of two angular-assignment algorithms commonly used in single-particle electron microscopy: projection-matching in real-space (Penczek et al., 1994) and angular assignment via Radon space (Radermacher, 1994). We performed four different experiments to show the properties of our assignment method.

For each experiment, a library of reference projections was built from a reference volume with an angular step of 5°. This library was used at the same time for the real-space projection-matching algorithm and for the proposed algorithm. The Radon transform was also sampled every 5°. The angular search was extended to the whole projection sphere (complete search) and the search for the particle center was restricted to a radius of 2 pixels around its ideal position (the shift step is 1 pixel). Since the angular sampling step was $\Delta = 5°$, we considered that the particles with an average angular error within 3Δ were reasonably assigned. For the selection of the most likely pose of the experimental image (Section 2.3), reference projections were considered to form a cluster if the distance between any two projection directions in the cluster was smaller than 3Δ. For the wavelet transform, we used Daubechies-12 wavelets as implemented in Press et al. (1992).

The *Halobacterium halobium* bacteriorhodopsin (PQS entry: 1BRD, Henderson et al., 1986) was used for the first experiment. The same volume was used for the simulation of the experimental images as well as a reference for the three algorithms being compared. The robustness of the algorithms and their accuracy are shown in Table 1.

To explore the influence of the quality of the reference model on the assignment accuracy, the following experiment was designed: the experimental images were simulated from the extracellular segment of *Homo sapiens* integrin Avb3 (PQS entry: 1L5G, Xiong et al., 2002). The projection library was built upon a filtered version of the phantom (low-pass filtered to $0.025\,\text{Å}^{-1}$). Table 2 shows the corresponding results.

To illustrate the performance of the algorithm on less-cooperative data, we simulated the experimental images using the human adenovirus type-5 hexon (PQS entry: 1RUX, Rux and Burnett, 2000, see Fig. 4). The results of the assignments are shown in Table 3. Fig. 5 shows the angular error vs. the correlation for each of the 1000 images and for the three compared algorithms.

Finally, we show an experiment in which the reference model does not correspond to the particle being imaged. Three extra masses were added to the bacteriorhodopsin (see Fig. 6) to simulate the experimental projections. These masses emulate the presence of additional features like antibodies. The bacteriorhodopsin without extra masses was used as reference volume for the assignment process. Table 4 summarizes the resulting assignment for each of the three methods. Fig. 7 shows the angular error vs. the correlation for each of the 1000 images and for the three compared algorithms.

3.1.3. Effect of the discarding factor

Let us explore the effect of the discarding factor introduced in Section 2.2. The first experiment of the preceding section (bacteriorhodopsin with a perfect reference volume) was repeated varying the discarding factor. The angular accuracy of the assignment as well as the computing time is shown in Table 5. For speed reasons, the translational parameters were assumed to be known. The angular accuracy has approximately an L-shaped curve while the computing time reduces slowly as the number of discarded images is increased.

3.2. Separate effect of the two proposed improvements

Finally, we were interested in recognizing which of the two novel features (multiresolution projection matching (see Section 2.2) or the pose selection strategy

Table 1
Accuracy of the reasonably assigned particles and its proportion for the bacteriorhodopsin

Method	Accuracy (°)	Accuracy (pixel)	Robustness
Radon-based assignment	5.05 ± 3.09	0.50 ± 0.53	[56.8%,63.0%]
Real-space matching	3.99 ± 2.77	0.36 ± 0.22	[92.2%,95.3%]
Wavelet-space matching	5.47 ± 2.95	0.58 ± 0.56	[97.8%,99.3%]

Table 2
Accuracy of the reasonably assigned particles and its proportion for the extracellular segment of integrin Avb3 showing the influence of the quality of the reference model

Method	Accuracy (°)	Accuracy (pixel)	Robustness
Radon-based assignment	5.09 ± 2.66	0.83 ± 0.75	[99.0%,99.8%]
Real-space matching	4.36 ± 2.90	0.95 ± 0.80	[90.7%,94.1%]
Wavelet-space matching	4.79 ± 2.87	0.71 ± 0.61	[97.2%,98.9%]

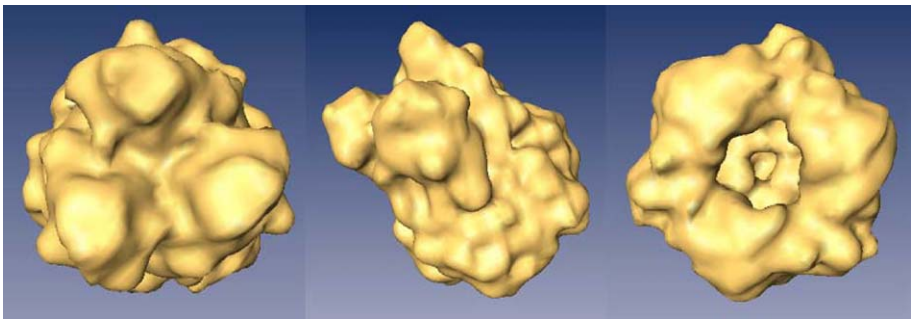

Fig. 4. Isosurface representation of the adenovirus type-5 hexon.

Table 3
Accuracy of the reasonably assigned particles and its proportion for less-cooperative data (adenovirus type-5 hexon)

Method	Accuracy (°)	Accuracy (pixel)	Robustness
Radon-based assignment	11.06 ± 2.83	1.44 ± 0.65	[1.3%,3.2%]
Real-space matching	3.37 ± 2.50	0.62 ± 0.46	[73.6%,79.0%]
Wavelet-space matching	4.88 ± 2.63	0.63 ± 0.55	[90.7%,94.1%]

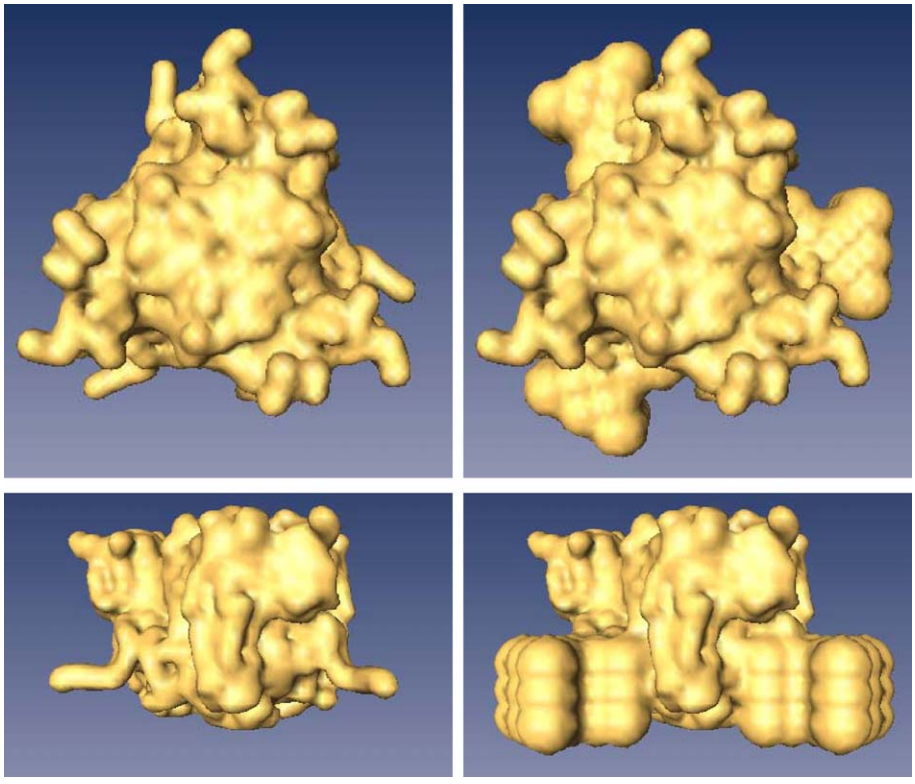

Fig. 6. Top (upper part) and side views (lower part) of the isosurfaces of the bacteriorhodopsin and of the bacteriorhodopsin with extra masses.

(see Section 2.3)) was more important for the increased robustness observed. For this task the adenovirus type-5 hexon was reassigned using different combinations of the two proposed modifications. The case *real-space matching with selection of the maximum* corresponds to what in this paper has been referred to as space-matching method and was run through Spider. This algorithm performs a 2D + 3D search of the pose parameters. In order to avoid a possible suboptimal result with respect to a full 5D search, the translational parameters were provided in all the assignments. Therefore, all the assignments are performing a simple 3D

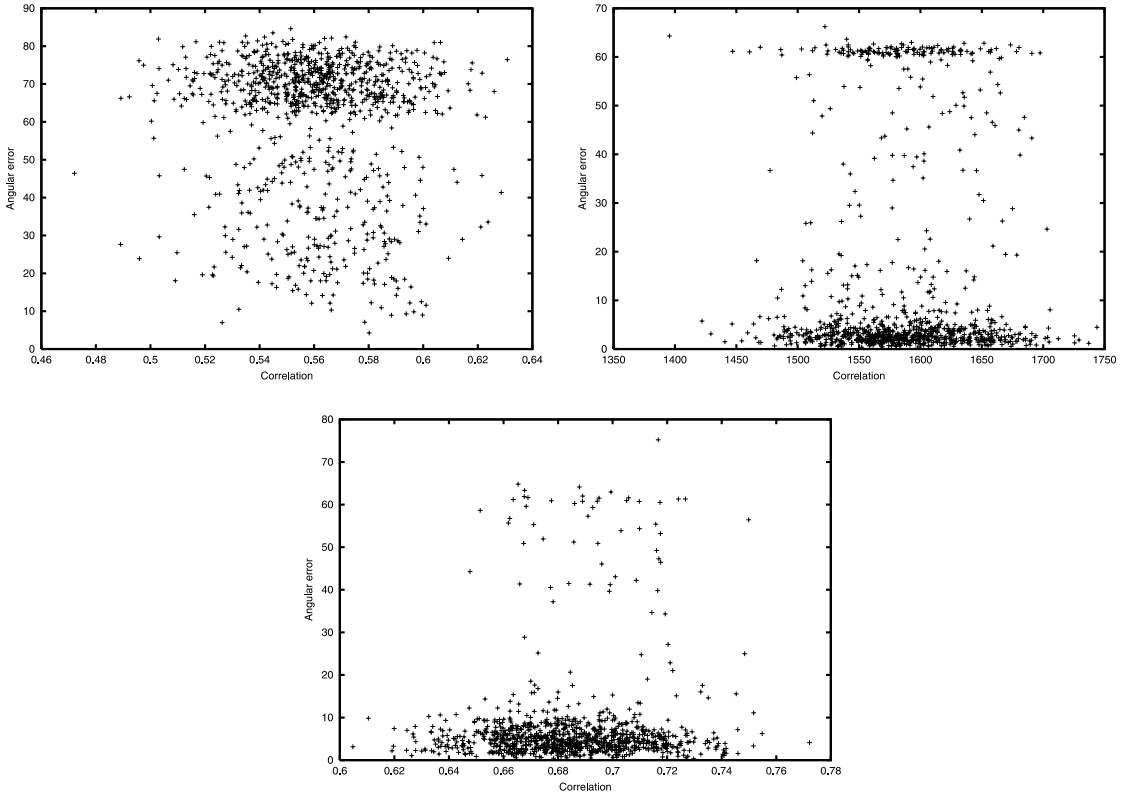

Fig. 5. Angular error vs. correlation for the pose assignment of the adenovirus type-5 hexon. From top to bottom: Radon-based assignment, real-space assignment, and wavelet-space assignment.

Table 4
Accuracy of the reasonably assigned particles and its proportion for the bacteriorhodopsin with extra masses showing the influence of a reference model that does not correspond to the particle being imaged

Method	Accuracy (°)	Accuracy (pixel)	Robustness
Radon-based assignment	5.29 ± 2.99	0.98 ± 0.77	[30.2%,36.1%]
Real-space matching	5.31 ± 3.44	1.22 ± 0.84	[51.4%,57.7%]
Wavelet-space matching	5.80 ± 3.02	0.98 ± 0.62	[62.6%,68.5%]

search. The discarding factor used in the multiresolution approach of the correlation was 40%. Table 6 shows the corresponding results.

3.3. Results on experimental data

In order to test the applicability of the proposed algorithm to experimental electron-microscopy data, we have used the cryo-negative micrographs of GroEL taken by De Carlo et al. (2002). Cryo-negative staining is a novel technique that preserves well the native state of the sample while providing high contrast. The reconstruction published by De Carlo et al. (2002) with these data had a resolution of 14 Å. The angular assignment in that case was performed using the projection-matching algorithm building the library of reference projections every 3°.

In this experiment, we performed a pose assignment based on the reference model used in the last iteration of De Carlo et al. (2002). Our reference library was built every 10°. The average angular difference between Euler axes of the original pose assignment and the one performed with the proposed algorithm is smaller than 3° for 35% of the particles and smaller than 10° for the 64%. For validating our assignment we performed a 3D reconstruction. Since the X-ray model of GroEL is available in PDB (Berman et al., 2000, PDB entry: 1GRL, Braig et al., 1994, 1995), we computed the Fourier Shell Correlation (Frank, 1996) between the X-ray model and the reconstruction using ART + blobs (Marabini et al., 1998) with our pose assignment. Fig. 8 shows the Fourier Shell Correlation obtained by the assignment for the last iteration produced in De Carlo et al., 2002 (this reconstruction was also done with ART + blobs) and the one

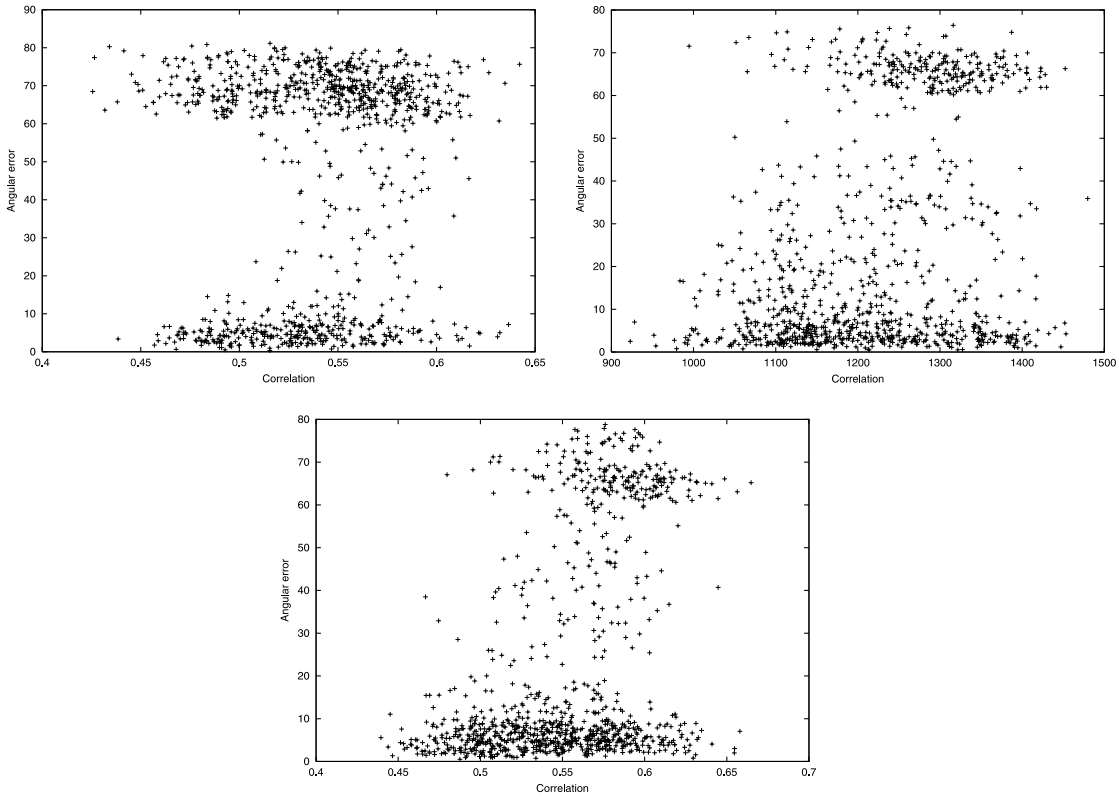

Fig. 7. Angular error vs. correlation for the pose assignment of the bacteriorhodopsin with extra masses. From top to bottom: Radon-based assignment, real-space assignment, and wavelet-space assignment.

Table 5
Computing time of the proposed algorithm and angular accuracy as a function of the discarding factor

Discarding factor (%)	Time (min)	Accuracy (°)
40	24.57	4.68
45	22.07	4.68
50	20.05	4.68
55	18.58	4.68
60	17.50	4.70
65	16.80	4.77
70	16.23	4.85

by the proposed algorithm. The frequencies at which each of the curves fall below 0.5 are 0.038 Å^{-1} (26 Å) and 0.040 Å^{-1} (25 Å), respectively.

As is usually done in single-particle electron microscopy, the set of projections was randomly split in two halves for each of the assignments. Four different reconstructions were produced (two for the previous assignment and two for the proposed assignment). The Fourier Shell Correlation corresponding to each pair of volumes is shown in Fig. 9.

4. Discussion

In the experiments carried out, we have compared the performance of the proposed algorithm with two widespread angular-assignment algorithms as they are implemented in Spider (Frank et al., 1996). We find that, for some proteins, all algorithms exhibit similar performances in terms of accuracy and robustness (see Table 2). It is interesting to notice that, for non-pathological cases, the angular accuracy is close to the angular sampling step for the three algorithms (see Tables 1, 2, and 4).

Table 6
Assignment accuracy and proportion of reasonably assigned particles for the adenovirus type-5 hexon using different combinations of the two proposed features

Selection strategy	Space	Accuracy (°)	Proportion
Maximum	Real space	3.48 ± 2.59	[72.8%,78.2%]
Maximum	DWT	4.17 ± 2.09	[78.4%,83.4%]
Most populated	Real space	4.33 ± 2.22	[86.1%,90.2%]
Most populated	DWT	4.32 ± 2.18	[90.3%,93.7%]

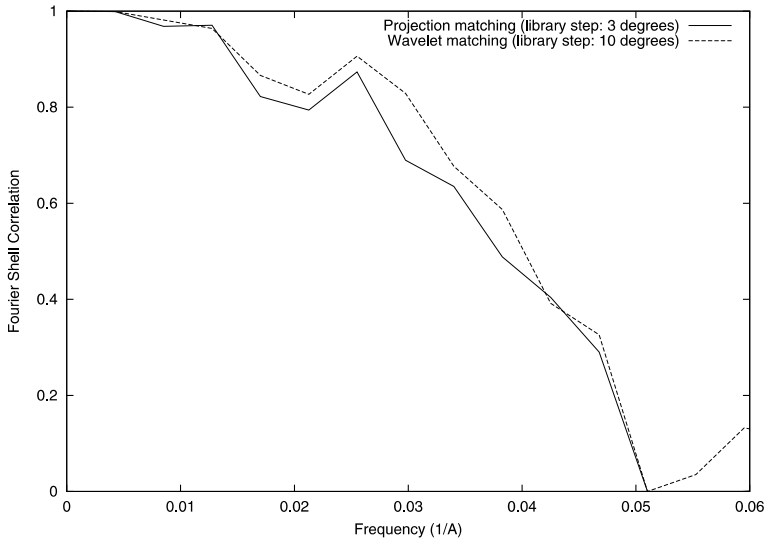

Fig. 8. Fourier Shell Correlation of two reconstructions of GroEL using experimental cryo-microscopy data and two different assignment algorithms with the X-ray model of GroEL.

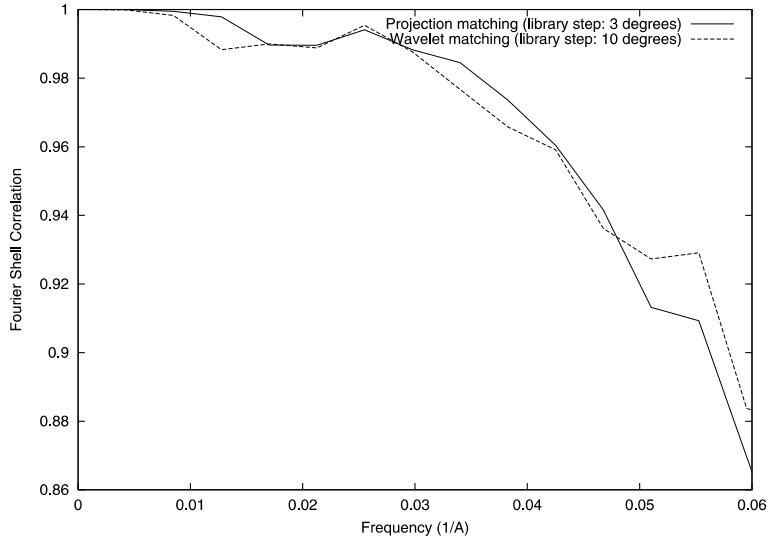

Fig. 9. Fourier Shell Correlation of two reconstructions of GroEL using experimental cryo-microscopy data and two different assignment algorithms when the data set is split in two halves.

However, there are some "more difficult" proteins for which the angular-assignment process is more likely to be mislead (see, for instance, the proportion of the reasonably assigned particles for the bacteriorhodopsin and the adenovirus type-5 hexon, Tables 1 and 3, respectively). This extra difficulty mainly translates into a reduced robustness while the accuracy of reasonably assigned particles is maintained. This effect may take dramatic dimensions as is the case for the Radon-based assignment of the adenovirus type-5 hexon. Our algorithm also shows a smaller number of reasonably assigned particles in this case, although it is more robust than the other two approaches.

The experiment with extra masses (Table 4) is particularly interesting since it simulates a common situation in which the reference model does not match the macromolecule being studied. As should be expected, the robustness drastically decreases for all angular-assignment methods.

The correlation coefficient used by the pose-assignment algorithm for comparing a single experimental image with a set of reference projections is commonly used as a measure of the quality of the assignment. In this way, the correlation coefficients of different experimental images are compared and those experimental images with low correlation coefficient are excluded

from the reconstruction process. However, as shown in Figs. 5 and 7, for none of the three algorithms studied in this paper is there a clear relationship between the correlation among different experimental images and their assignment accuracy. These graphs point out a topic in which more research must be carried out: how to identify the wrongly assigned images when the ground truth is unknown.

The discarding factor is a free parameter of our algorithm. When the discarding factor is varied, the average angular error shows a flat region until 55% (see Table 5). This means that, for the experiment carried out, we could have safely dropped up to the 55% of the images with lower sub-band correlation. Beyond this discarding factor, we are sometimes dropping the best-matching reference projection. The computing time reduces with increasing discarding factors, but this reduction is smaller for high discarding factors than for low ones. This can be explained by the fact that the experimental image is compared with all reference images at least at coarse resolution, even if the discarding factor is high.

The experiment carried out in Section 3.2 is quite informative since it reveals the relative importance of the individual contributions of each of the modifications proposed to the increase of robustness. For the experiment carried out, the most important contribution is that of the pose selection strategy. However, the multiresolution computation of the correlation accounts for an important improvement of the robustness even with a simple pose selection strategy. Finally, the combination of both techniques reveals a positive interaction yielding a robustness that cannot be achieved by any of the modifications independently.

The computing time for the complete-search experiments performed in this work varies from one algorithm to another. Projection matching in the space domain is clearly the fastest algorithm (about 20 min). The main reason is that the translational search (2D) and the angular search (3D) are performed independently; thus, the problem complexity is highly reduced (2D + 3D). Projection matching in the wavelet domain took about 8 h. However, it must be taken into account that the proposed algorithm performs a full 5D search and has to make a wavelet transform for every considered in-plane pose of the experimental image since the wavelet transform lacks a shift property similar to the one of the Fourier transform (Simoncelli et al., 1992). Finally, the Radon-based assignment took about 50 h for the complete search. The Radon-based assignment also performs a 5D search, but the Radon transforms of the images and the volume can be precomputed. The previous time measures were performed on a single processor of a Cluster Alpha Server (5 nodes ES 45) with 1 GHz Alpha EV68 microprocessors and 8 GB of RAM memory per node. The operating system is Tru64 5.1.

The proposed algorithm has been tested on experimental electron-microscopy data. In particular, the GroEL has been used since its atomic model is available and, therefore, allows the comparison of the reconstructions with a volume very close to the ground truth. The results of our algorithm are similar to those obtained in previous works (De Carlo et al., 2002) with a much-finer reference library. The resolution achieved in each case was around 25 Å, although the Fourier Shell Correlation at most of the lower frequencies was slightly higher for the reconstruction obtained using the newly proposed algorithm (see Fig. 8). We also computed the standard self-consistency measure in single-particle electron microscopy by computing the Fourier Shell Correlation between two volumes that were reconstructed from two random halves of the particle set. For the experiment carried out, the self-consistency of the previous assignment is slightly better than the one of the proposed algorithm (see Fig. 9), as was expected since the assignment using wavelet-space matching was performed upon a much-coarser reference library.

5. Conclusions

A new algorithm for assigning the pose of single-particle electron-microscopy images has been introduced. It requires a reference 3D model. It is robust with respect to noise since it uses a multiresolution strategy to compute the correlation between the experimental images and the library of projections of the reference volume. The final decision about the pose of the particle is taken considering a set of the highly correlated reference projections instead of simply the most correlated one. The resulting algorithm appears to be more robust than the current standards in the field in a number of simulated experiments. The algorithm has been shown to work with electron-microscopy experimental data as well.

The algorithm presented in this paper discretizes the angular and translational space (parameter space). This allows us to perform a complete exhaustive search on the parameter space in a finite time. However, there exist approaches (Jonić et al., 2003; Provencher and Vogel, 1988; Vogel and Provencher, 1988) that perform a continuous search in the parameter space, thus providing a higher assignment accuracy. The main drawback of these continuous alternatives is that they need a reasonable initial guess for the particle pose. We are currently developing a continuous parameter algorithm for angular assignment that uses the assignment produced by the algorithm proposed in this paper as initialization.

Acknowledgments

Partial support is acknowledged to the "Comisión Interministerial de Ciencia y Tecnología" of Spain through Project BIO98-0761, and BIO2001-1237 and to the NIH through Grant HL70472. Partial support is also acknowledged to the "Universidad San Pablo-CEU" through Project 17/02.

References

Aldroubi, A., Unser, M., 1996. Wavelets in Medicine and Biology. CRC Press, Boca Raton.

Basu, S., Bresler, Y., 2000a. Feasibility of tomography with unknown view angles. IEEE Trans. Image Process. 9, 1107–1122.

Basu, S., Bresler, Y., 2000b. Uniqueness of tomography with unknown view angles. IEEE Trans. Image Process. 9, 1094–1106.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Boutselakis, H., Dimitropoulos, D., Fillon, J., Golovin, A., Henrick, K., Hussain, A., Ionides, J., John, M., Keller, P.A., Krissinel, E., McNeil, P., Naim, A., Newman, R., Oldfield, T., Pineda, J., Rachedi, A., Copeland, J., Sitnov, A., Sobhany, S., Suarez-Uruena, A., Swaminathan, J., Tagari, M., Tate, J., Tromm, S., Velankar, S., Vranken, W., 2003. E-MSD: The european bioinformatics institute macromolecular structure database. Nucleic Acids Res. 31, 458–462.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., Sigler, P.B., 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. Nature 371, 578–586.

Braig, K., Adams, P.D., Brunger, A.T., 1995. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. Nat. Struct. Biol. 2, 1083–1094.

Dengler, J., 1989. A multi-resolution approach to the 3d reconstruction from an electron microscope tilt series solving the alignment problem without gold particles. Ultramicroscopy 30, 337–348.

Desco, M., Herández, J.A., Santos, A., Brammer, M., 2001. Multiresolution analysis in fMRI: sensitivity and specificity in the detection of brain activation. Hum. Brain Mapp. 14, 16–27.

De Carlo, S., El-Bez, C., Alvarez-Rúa, C., Borge, J., Dubochet, J., 2002. Cryo-negative staining reduces electron-beam sensitivity of vitrified biological particles. J. Struct. Biol. 138, 216–226.

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

Frank, J., 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. Annu. Rev. Biophys. Biomol. Struct., 303–319.

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

Gelfand, M.S., Goncharov, A.B., 1990. Spatial rotational alignment of identical particles given their projections: theory and practice. Transl. Math. Monogr. 81, 97–122.

Goncharov, A.B., 1990. Three-dimensional reconstruction of arbitrarily arranged identical particles given their projections. Transl. Math. Monogr. 81, 67–95.

Henderson, R., Baldwin, J.M., Downing, K., Lepault, J., Zemlin, F., 1986. Structure of purple membrane from *Halobacterium halobium*: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. Ultramicroscopy 19, 147–178.

Jonić, S., Thévenaz, P., Unser, M., 2003. Multiresolution-based registration of a volume to a set of its projections. In: Proceedings of the SPIE Symposium on Medical Imaging, vol. 5032. San Diego, pp. 1049–1052.

Kuhlbrandt, W., Williams, K., 1999. Analysis of macromolecular structure and dynamics by electron cryo-microscopy. Curr. Opin. Struct. Biol. 3, 537–543.

Lauren, P.D., Nandhakumar, N., 1997. Estimating the viewing parameters of random, noisy projections of asymmetric objects for tomographic reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. 19, 417–430.

Marabini, R., Masegosa, I.M., Martín, M.C.S., Marco, S., Fernández, J.J., de la Fraga, L.G., Vaquerizo, C., Carazo, J.M., 1996. Xmipp: an image processing package for electron microscopy. J. Struct. Biol. 116, 237–240.

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

Penczek, P.A., Grasucci, R.A., 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.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. Numerical Recipes in C, second ed. Cambridge University Press, Cambridge.

Provencher, S.W., Vogel, R.H., 1988. Three-dimensional reconstruction from electron micrographs of disordered specimens. I. Method.. Ultramicroscopy 25, 209–222.

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

Rux, J.J., Burnett, R.M., 2000. Type-specific epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. Mol. Ther. 1, 18–30.

Saad, A., Chiu, W., 2000. Hierarchical wavelets projection matching for orientation determination of low contrast electron cryomicroscopic images of icosahedral virus particles. In: Proceedings of the IEEE ICASSP'00, vol. 4. pp. 2270–2273.

Salzman, D.B., 1990. A method of general moments for orienting 2D projections of unknown 3D objects. Comput. Vis. Graph. Image Process. 50, 129–156.

Simoncelli, E.P., Freeman, W.T., Adelson, E.H., Heeger, D.J., 1992. Shiftable multiscale transforms. IEEE Trans. Inform. Theory 38, 587–607.

Thévenaz, P., Ruttiman, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. 7, 27–41.

Unger, V.M., 2001. Electron cryomicroscopy methods. Curr. Opin. Struct. Biol. 11, 548–554.

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

van Heel, M., Gowen, B., Matadeen, R., 2000. Single-particle electron cryo-microscopy: towards atomic resolution. Q. Rev. Biophys. 33, 307–369.

Vardeman, S.B., 1994. Statistics for Engineering Problem Solving. Academic Press, Boston.

Velázquez-Muriel, J.A., Sorzano, C.O.S., Fernández, J.J., Carazo, J.M., 2003. A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjusting. Ultramicroscopy 96, 17–35.

Vogel, R.H., Provencher, S.W., 1988. Three-dimensional reconstruction from electron micrographs of disordered specimens. II. Implementation and results. Ultramicroscopy 25, 223–240.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., Arnaout, M.A., 2002. Crystal structure of the extracellular segment of integrin $\alpha V^{\beta}3$ in complex with an Arg-Gly-Asp ligand. Science 296, 151–155.