



ELSEVIER

Ultramicroscopy 93 (2002) 111–121

ultramicroscopy

www.elsevier.com/locate/ultramic

# A trajectory-based algorithm to determine and refine Euler angles of projections in three-dimensional microscopy. Improvements and tests

Pier Luigi Bellon<sup>a</sup>, Francesca Cantele<sup>a</sup>, Sacha De Carlo<sup>b</sup>,  
Salvatore Lanzavecchia<sup>a,\*</sup>

<sup>a</sup> *Dipartimento di Chimica Strutturale e Stereochimica Inorganica, Università' degli Studi, Via G. Venezian 21, I-20133, 20129 Milano, Italy*

<sup>b</sup> *Centre de Microscopie Electronique de l'Université de Lausanne, 27, rue du Bugnon, 1005 Lausanne, Switzerland*

Received 21 November 2001; received in revised form 5 March 2002

## Abstract

An improvement of the trajectory matching algorithm is presented, which is based on the use of the derivative of trajectories and of the projection of experimental sinogram lines in the factor space determined by sinogram lines of projections of a model. The algorithm performance is illustrated by use of different phantom structures, to show the effect of symmetry on trajectory matching. A GroEL complex has also been reconstructed from both cryo-negatively stained and unstained frozen-hydrated samples. The refinement of this structure has been carried out by the trajectory matching algorithm as well as by conventional cross-correlation methods. Slight differences among the two results are discussed. The improved trajectory matching algorithm, based on  $\chi^2$  distances, runs much faster than correlation analysis and looks satisfactory as for the quality of the reconstructed structures. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Correspondence analysis; Trajectories in factor space; Sinograms; Projection alignment

## 1. Introduction

When a 'complete set' of projections of a three-dimensional object is submitted to correspondence analysis (CA) [1], the representative points in factor space spread along a trajectory, i.e., along a closed and complicated line. By complete set we

mean a collection obtained with some rule in which the projecting directions are obtained by varying an angle in the interval  $0-2\pi$ . A set of this kind is obtained in a random conical tilt experiment [2] in which the varying angle is the azimuth. Frank and Radermacher have published, for projections of the ribosome obtained with this technique, the first closed path of points in factor space, though the observation reported was limited to two main eigenvectors [3]. A similar path has been reported for a single axis tilt series [4] though in this case the trajectory remains open owing to the missing wedge.

\*Corresponding author. Tel.: +39-02-5031-4441; fax: +39-02-5031 4454.

E-mail address: salvatore.lanzavecchia@unimi.it (S. Lanzavecchia).

We have recently published our finding about trajectories obtained from another class of functions, the sinograms [5]. Sinograms are a collection of line integrals of 2D images computed, line by line, by rotating stepwise from 0 to  $2\pi$  a vector orthogonal to the lines of integration. Given a set of images, say TEM projections, their sinograms are computed and CA is performed on the collection of all line integrals, each regarded as separated item. Each line becomes a point in factor space and all points form a trajectory. The points of a trajectory are found in the same sequence as the lines in the sinogram. A whole set of sinograms yield a set of trajectories.

Sinogram trajectories are alternative to CA of whole images. In an experiment of three-dimensional microscopy, sinograms must be obtained at some point for different purposes: to align equivalent images [6], to find common lines shared by different projections [7,8] and to perform a three-dimensional reconstruction via the Radon transform [9]. On the other hand, these functions can be conveniently computed by direct Fourier methods [10]. A great advantage of CA carried out on sinogram lines is the strong decrease of computational complexity compared to an analysis carried out in image- or pixel space. A detailed analysis of this aspect is beyond the purpose of this paper. We shall limit ourselves to compare the dimensions of the matrices submitted to diagonalisation in the two cases. As an example, let us consider a set of 4000 images ( $\approx 2^{12}$ ),  $64^2$  pixels in size ( $2^{12}$  pixels). Its CA would be based on a  $2^{12} \times 2^{12}$  matrix, both in image- and pixel space. Conversely, the matrix required in a sinogram line analysis is only  $64 \times 64$  elements in size ( $2^{12}$ ), that is  $2^{12}$  times smaller. At first sight, such a dramatic reduction of the problem dimension might appear as a sacrifice of information. This is not true: instead of obtaining points representing individual images in factor space, we obtain trajectories which exactly correspond to sinograms and hence to images. Therefore, this kind of CA can greatly alleviate the computational burden of multivariate image statistics.

The advantages of using sinogram trajectories rather than images is by no means limited to the convenience in performing CA. Trajectories are

actually versatile entities whose potentialities deserve to be explored in depth. Some applications of the novel instrument have been already demonstrated in a previous article [5]. We have illustrated the power and low complexity of common lines detecting algorithm and of a determination of Euler angles of experimental projections, based upon sinogram trajectories. The former application is based on the fact that the sinograms of two different projections share a line and, correspondingly, their trajectories must cross each other. This means that two points of one either coincide or lie at narrow distance from two points of the other. The second consists of determining which experimental trajectory matches at best with a computed counterpart, that is, which of the former has all its points closest to those of the latter. Both tasks require the evaluation of distances among points in a  $\chi^2$  metric whose dimensions are reduced to few main eigenvectors. The advantage is that cross-correlation algorithms are replaced by simple computation of distances.

In the present article we wish to present recent improvements of the process of matching trajectories of experimental- and computed projections. The novel alignment strategy can be used in exhaustive (or almost so) refinements of three-dimensional structures of macromolecular assemblies. We shall present and discuss the results obtained with phantom structures and a trajectory-based refinement of the chaperonin GroEL [11].

## 2. Sample preparation

The GroEL chaperonin (stock solution, 18 mg/ml) was diluted in a buffer solution composed of 20 mM MOPS, 100 mM KCl, 4 mM  $\text{MgCl}_2$ , pH 7.4, to a final concentration of  $\sim 1$  mg/ml. GroEL particles were prepared for cryo-electron microscopy and cryo-negative staining as follows. Home-made holey carbon films were used. The films were mounted on 200-mesh copper grids, and a thin layer of PtC was sputtered on one side of the grid. For cryo-electron microscopy, the sample was immediately vitrified after blotting

[12]. For cryo-negative staining, the sample was applied on the staining solution droplet for 30 s. The specimen was then removed and vitrified in liquid ethane as described in [13].

The micrographs used in this work were recorded at a temperature of  $-180^{\circ}\text{C}$  and at a calibrated magnification of  $\times 52780$  in a Philips CM100 (LaB<sub>6</sub> cathode source operated at 100 keV) using a Gatan 626 cryo-holder (Gatan Inc., Warrendale, PA, USA).

### 3. The method

#### 3.1. From sinogram to trajectories

Given a 3D object, like that in Fig. 1a, and a set of  $n$  projections  $64 \times 64$  pixels wide, some of which are shown in Fig. 1b, their sinograms (two-dimensional Radon transforms) are represented as arrays with 128 rows and 64 columns as in Fig. 1c. The arrays are further organised, one beneath the other, in an observation matrix  $\mathbf{X}$ , with  $128 \times n$  rows and 64 columns (Fig. 1d). This matrix is submitted to CA which consists of three steps: (i) a covariance matrix  $\mathbf{C} = \mathbf{X}^T \mathbf{N} \mathbf{X} \mathbf{M}$  ( $\mathbf{N}$  and  $\mathbf{M}$  being diagonal normalising matrices) is computed, with  $64 \times 64$  coefficients. Note that if  $\mathbf{M}$  and  $\mathbf{N}$  are omitted (no normalisation), the covariance matrix is the same as used in the principal component- or Karhunen–Loeve transformation [14,15]; (ii) the secular equation  $\mathbf{C}\mathbf{U} = \mathbf{U}\mathbf{A}$  is solved,  $\mathbf{U}$  being an orthonormal

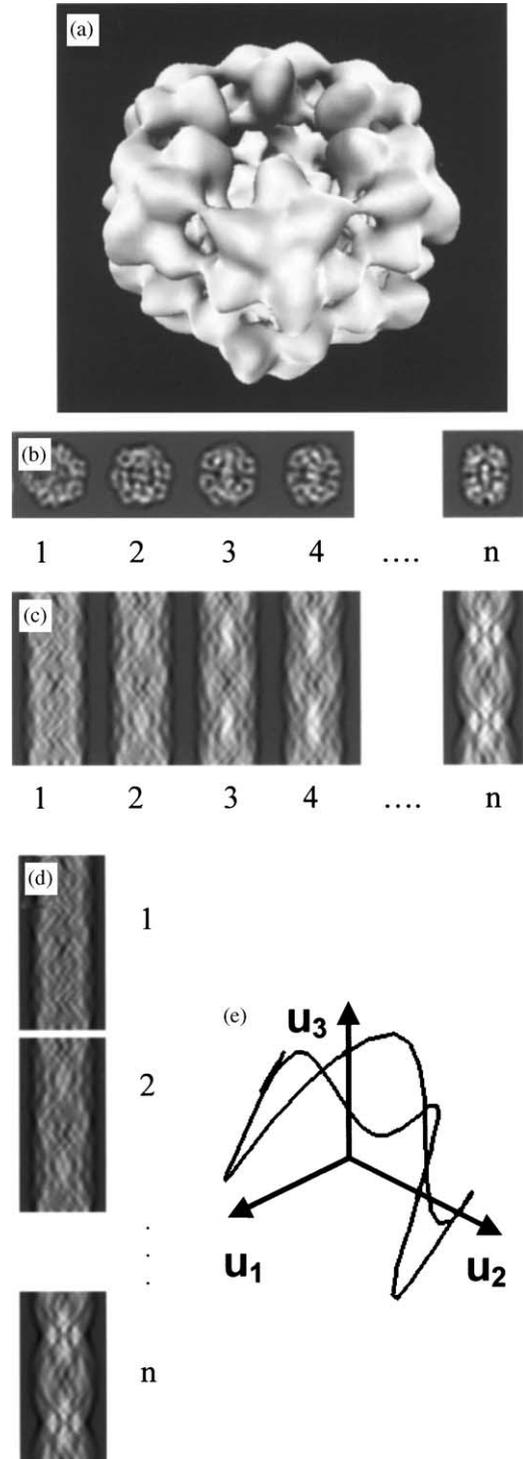


Fig. 1. From a structure model to sinogram trajectories. The three-dimensional model (a) yields the set of projections  $1 \dots n$  shown (b), sizing  $64 \times 64$  pixels and their sinograms, shown (c), which are numerical arrays sizing 64 columns  $\times$  128 rows (rows span the angular range  $0-2\pi$ ). In (d) an array is formed by placing sinograms  $1, \dots, n$  one beneath the other. This large array with 64 columns and  $128n$  rows is the observation matrix  $\mathbf{X}$  which, once normalised and pre-multiplied by the transpose  $\mathbf{X}^T$ , yields an inertial tensor  $64 \times 64$  coefficients wide used to solve the secular equation. Different trajectories, one per sinogram and occurring in the same order as do the sinograms in  $\mathbf{X}$ , are observed in factor space. Their points occur in the same sequence as the lines in a sinogram. A trajectory is shown (e), in a representation limited to three main eigenvectors.

matrix of eigenvectors which specify the directions of inertial axes of the line distribution and  $A$  being a diagonal matrix of eigenvalues representing the inertia associated with the eigenvectors; (iii)  $U$  is used to operate a ‘rotation’ of all vectors of  $NXM$  along the directions prescribed by eigenvectors. There is a rapid decrease of inertia along the diagonal of  $A$  so that significant features of images or of sinogram lines are adequately represented in a space of the first few eigenvectors. The number of eigenvectors considered is such that the corresponding eigenvalues encompass not less than 90% of total inertia. In our experience this reduces to

6–10 the dimensions of factor space. Hence, a trajectory can be viewed in this low-dimensional space and, if one so wishes, converted back to its sinogram and to the image by omitting all eigenvectors containing featureless information. This form of statistical filtering has been already investigated in the case of whole images [16]. An example of a trajectory represented in the space of three main eigenvectors is shown in Fig. 1e.

In Fig. 2 three different analytical phantoms are shown, possessing symmetry  $C_1$ ,  $D_4$  and  $O$ . The group orders are 1, 8 and 24, respectively, that is the phantoms contain 1, 8, and 24 equal

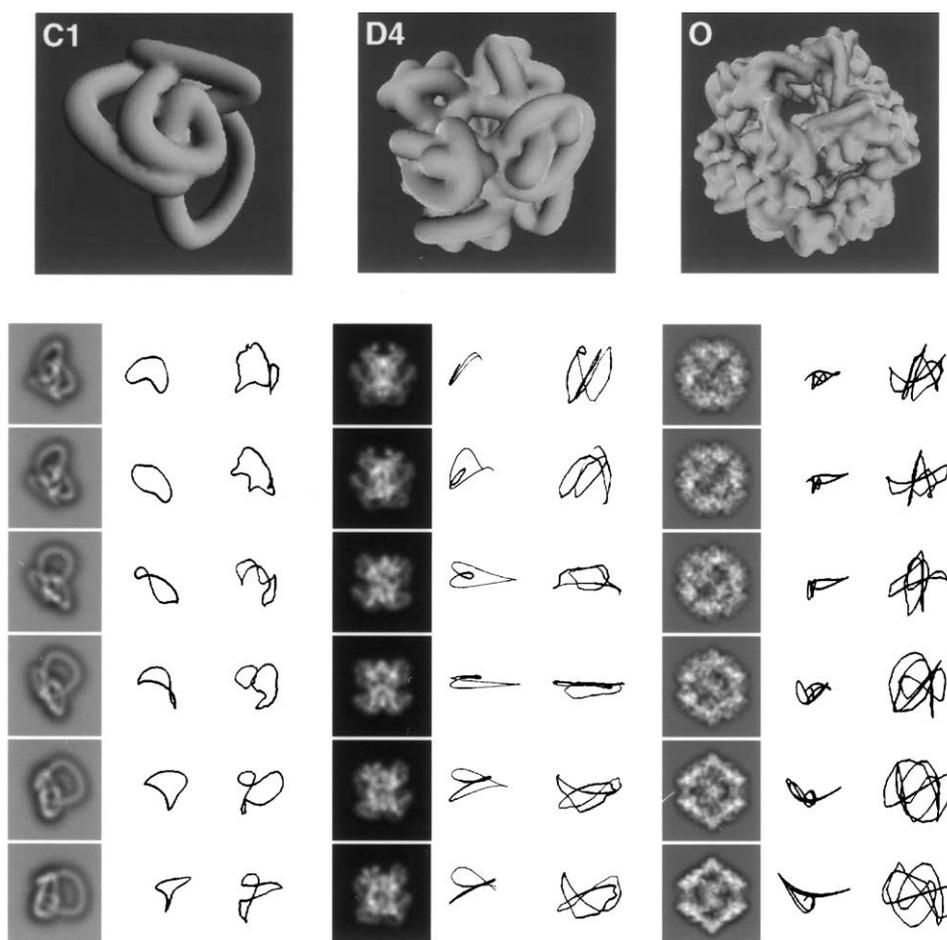


Fig. 2. The three main columns of images show, at the top, three different analytical phantoms with symmetries  $C_1$ ,  $D_4$  and  $O$ . Beneath each of the three models are, at left, some noise-free projections accompanied by their trajectories represented in the space of two main eigenvectors. Two versions of trajectories are shown: original and, in rightmost columns, after derivation.

asymmetric units replicated according to the group symmetry. Fig. 2 shows the three phantoms and, beneath each of them, some of their projections and of the trajectories obtained. As can be seen, as the group order increases trajectories of higher complexity are obtained. Owing to the presence of symmetry, the same projection can be obtained along different directions so that a given line is present in multiple copies in a sinogram. This introduces crossing points along the trajectory, and some problems arise in determining the line shared by two sinograms [5]. It may, however, become an advantage in a projection matching process.

### 3.2. The trajectory matching

Projection matching [17] is the process by which Euler angles are assigned to experimental images. It is an iterative process, commonly performed by cross-correlating experimental data with projections of an updating structural model and involves the use of a two-dimensional Fourier algorithm. In the novel approach, the trajectories of experimental images are compared with those of the model, in order to assess which matches which. The matching is evaluated by computing a distance between two trajectories, defined as the sum of point-to-point distances in factor space. Note that rotating a projection corresponds to an angular shift of the sinogram and of the trajectory, i.e., to a cyclic permutation of line indices. Given two trajectories, if point 1 in the first is near to point  $m$  in the second, point 2 will be near to point  $m+1$  and so on. Thus, we need to determine the cyclic permutation which minimises the sum of distances. A  $\chi^2$  map of distances of points of a trajectory from all points of the other is first obtained. This map is equivalent to the sinogram correlation [8] whose values represent all cross-correlation maxima between lines. The sum of a diagonal line of this map corresponds to a sum of point-by-point distances, related to a given shift  $m$ . The index corresponding to the minimal sum determines the mutual rotation of the two maps. Eq. (1) defines the distance between two trajectories  $S(i,j)$  and  $T(i,j)$ , where  $i$  spans the number  $N$  of lines and  $j$  the dimensions of factor

space ( $J$ ):

$$d_{S-T} = \min \left\{ \sum_i \left[ \sum_j (S_{ij} - T_{i+m_j})^2 \right]^{1/2} \right\}. \quad (1)$$

The trajectory matching requires that each experimental trajectory be compared with all computed ones to determine the minimum distance. Once this is done, an experimental projection is eventually assigned with the two Euler angles used to project the model and with a third angle, an in-plane rotation, obtained from the index  $m$ .

If a whole set of sinogram lines of computed and experimental projections are submitted to CA, two main problems arise, which were illustrated in our previous article. The first is that the two clouds of experimental and computed trajectories are separated along the first main eigenvector. The other effect is that experimental trajectories are smaller than their computed counterpart. A preliminary shift and scaling operation was therefore devised to make the comparison possible. This solution was a bit naive. In the method exposed below both effects are avoided in a more rigorous way.

### 3.3. The new trajectory matching algorithm

A remedy for the separation of the two clouds of trajectory consists of carrying a CA on the lines of sinograms of computed projections alone, to obtain a matrix  $U_c$  of eigenvectors and a diagonal  $A_c$  matrix of eigenvalues. The lines of experimental sinograms are projected in this new space. The equation to project the  $i$ th line  $s_i$  of a new sinogram in the factor space and to obtain the new co-ordinate  $S_{ij}$  [18] is

$$S_{ij} = \frac{1}{M_s \sqrt{\lambda_j}} \sum_k s_{ik} P_{kj}, \quad (2)$$

where  $S_{ij}$  is the  $j$ th co-ordinate of line  $S_i$ ;  $s_{ik}$  is the  $k$ th pixel of line  $s_i$  whose total sum is  $M_s$ ;  $\lambda_j$  is the  $j$ th eigenvalue in the matrix  $A_c$  and  $P_{kj}$  are the pixel coordinates in the line space.

Still, experimental trajectories obtained in this way are smaller than their computed counterparts. The problem is solved by comparing the derivatives rather than the originals of both types of

trajectories. Derivatives are independent from the position of the origin and from the relative dimensions of trajectories though they preserve the shape-related information. Consider the length of each segment of a trajectory  $S$ :

$$\ell_i = \|S_i - S_{i-1}\|_J = \left( \sum_j (S_{ij} - S_{i-1j})^2 \right)^{1/2}$$

the discrete derivatives, to be computed for each significant eigenvector, will be of the type:

$$S'_{ij} = \frac{S_{ij} - S_{i-1j}}{\ell_i}.$$

Derivatives  $S'$  are defined in the same space as trajectories  $S$  and are used to compute the distances the same way as is done for trajectories. Some examples of derivatives of experimental trajectories are shown in Fig. 2. This treatment allows one to substantially overcome the problems mentioned above.

## 4. Results

### 4.1. Numerical simulations

A set of 1000 projections with random orientations were computed for each phantom of Fig. 2. The images were corrupted, down to  $S/N = 1$ , with Gaussian noise (Fig. 3, top row) and with noise digitised from featureless areas of micrographs of ice embedded samples (Fig. 3, bottom

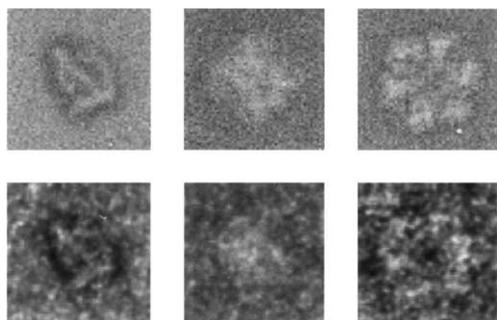


Fig. 3. Typical projections of the three phantoms shown in Fig. 2, corrupted with noise ( $S/N = 1$ ). Top row: images corrupted with Gaussian noise. Bottom row: images corrupted with noise digitised from structure-free portions of micrographs of ice embedded phantoms.

row). The sets of reference projections had viewing directions homogeneously distributed in the independent part of the unit sphere. Four hundred and twenty-two projections spaced by  $10^\circ$  along two Eulerian angles were computed for the phantom with symmetry  $C_1$ , with the third angle set at random. Analogous projections were obtained for phantoms with symmetries  $D_4$  and  $O$ ; 223 and, respectively, 74 references spaced by  $5^\circ$  were obtained in the independent angular range dictated by symmetry. The model reconstruction was performed in all cases with a fast and accurate algorithm based on the Radon transform [9].

In order to compare the relative performances, the projection matching was carried out using both trajectory-based and cross-correlation algorithms [6]. We have compared the Euler angles of computed projections with those determined by the two methods. Table 1 gives the results obtained. The accuracy attainable is related to the angular spacing of viewing directions of the reference projections; the data in the table indicate the percentage of noisy projections correctly aligned within one or two sampling steps of the independent part of the unit sphere. The discrepancy [19] of reconstructed phantoms with respect to the models, reported in the Table 2 as *rmsd*, is essentially identical for both algorithms. As for the angle determined, however, the conventional cross-correlation method looks a bit more accurate but it takes much a longer time ( $\approx 6 \times$ ).

### 4.2. Trajectory matching with actual experimental data

Trajectory matching has been tested on a real structure, obtained from ice-embedded samples of the GroEL protein complex [11]; the results have been compared with those obtained with a standard correlation algorithm. Two set of projections, stained [13] and unstained, were digitised with about 1600 molecules per set. Some of them are shown in Fig. 4. Each type of images was divided into two subsets of about 800 images. The performances of the methods were evaluated by Fourier shell correlation (FSC) [20] to estimate the cross-resolution of two independently refined models.

Table 1

Percentage of angles correctly assigned to 1000 analytical random projections of phantoms by correlation- and trajectory-based algorithms

1	Phantom	C1				D4				O			
2	Noise type	Gauss		Exp.		Gauss		Exp.		Gauss		Exp.	
3	Accuracy	10°	6°	10°	6°	5°	3°	5°	3°	5°	3°	5°	3°
4	Correl. (%)	100	83	96	68	100	82	96	68	100	84	99	85
5	Traject. (%)	100	87	93	62	100	81	91	58	100	84	99	75
6	Correl.time	6 h 30'				3 h 30'				1 h 10'			
7	Traject.time	1 h				33' 30"				13'			

Row 1: symmetry of the phantom; row 2: type of noise; row 3: criterion of accuracy, i.e., maximum errors accepted for correct assignments; row 4 and 5: % of orientations correctly determined within the angular spacing of row 3 with use of correlation- and, respectively of trajectory based algorithms. Row 6 and 7: elapsed times for correlation and, respectively, trajectories.

Table 2

Discrepancy between original phantoms and models reconstructed from 1000 noisy projections according to the Euler angles detected by trajectories and correlation algorithms

1	Phantom	C1		D4		O	
2	Noise type	Gauss	Exp.	Gauss	Exp.	Gauss	Exp.
3	Correl.	0.125	0.170	0.064	0.085	0.154	0.157
4	Traject.	0.122	0.172	0.068	0.086	0.158	0.163

Row 1: symmetry of the phantom; row 2: type of noise; row 3,4: discrepancy for correlation and, respectively, trajectories.

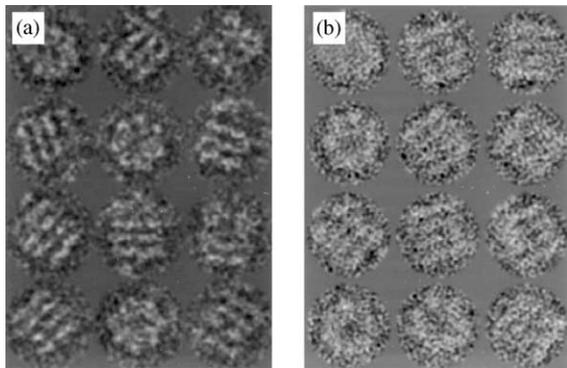


Fig. 4. Some images digitised from micrographs of GroEL preparations. (a) Molecules in cryo-negatively stained preparations; (b) molecules from unstained preparations.

These protein complexes lay essentially in two preferred orientations. Classification and class averages evidenced an almost exclusive presence of top and side views. The images were well centred during the classification process. From

these two views a preliminary model for projection matching was obtained. This model and the first refinements have been projected at regular steps of 5° in viewing direction angles to produce the data set for classification. In the last steps of refinement a different criterion was adopted. Based upon the Euler angle already assigned to each projection, a set of close viewing directions have been computed at steps of 1° and the relative projections obtained from the refining model. Their sinogram lines were projected on the lines of the 'classification set' according to Eq. (2) and a set of reference trajectories spaced by 1° were generated. Therefore, each trajectory coming from experimental image was compared with computed trajectories up to 1° of angular resolution. A similar approach was adopted also for the refinement by correlation; in this case, the algorithm was able to take into account also shift misalignments.

Stained molecules were quickly oriented, in about 5–6 cycles, by both algorithms. Sections *a*

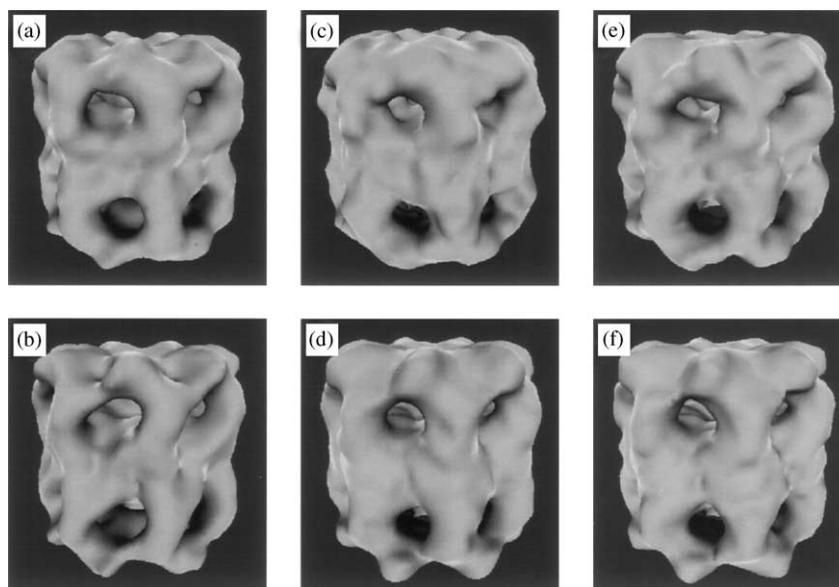


Fig. 5. Three-dimensional models of GroEL obtained by refining the projection angles of sets of 1600 projections. Top row: results of refinement by correlation matching. Bottom row: results of refinement by trajectories. Models (a) and (b) are obtained from cryo-negatively stained samples. Models *c* and *d* are obtained from unstained samples in a refinement starting from a preliminary reconstruction from top and side views selected within the unstained set. Models (e) and (f) result from unstained samples with the refinement based upon models (a) and (b), respectively.

and *b* of Fig. 5 show the models refined by correlation and by trajectory matching, respectively (trajectories limited to 6–7 eigenvectors). Much slower convergence, with both methods, was observed for unstained molecules. Using a preliminary model obtained from unstained samples, 10 cycles were necessary to reach good FSC agreement between the results obtained from two subsets (see Fig. 5c and d). Starting from the refined model obtained from stained samples, however, convergence arrived quickly. The results are shown in Fig. 5e and f. This behaviour suggests that contrast enhancing by staining may offer a substantial help in reconstructing difficult molecules. The structures shown in Fig. 5 yield the FSC diagrams of Fig. 6, which indicate a spatial resolution of about 2.2 nm, in agreement with the position of the first zero of the contrast transfer function (CTF) in the micrographs. Slight differences among the reconstructed molecules are discussed below.

## 5. Discussion

In the present study we are introducing a matching among derivatives of experimental- and computed sinogram trajectories. A key feature of the method is that experimental trajectories are obtained by projecting experimental sinogram lines in a factor space defined for a noise-free model. Experimental trajectories, whatever obtained, are smooth and well-behaving curves due to the fact that the image noise is reduced by line integration in the process of obtaining sinograms. Additional smoothing is also possible by applying a moving average to the trajectory points before computing the derivatives.

The two applications of trajectories presented in our previous study were dealing with common line detection and projection matching. Both applications were tested with use of phantom structures. In this paper we have tested the behaviour of trajectories as symmetry increases and observed

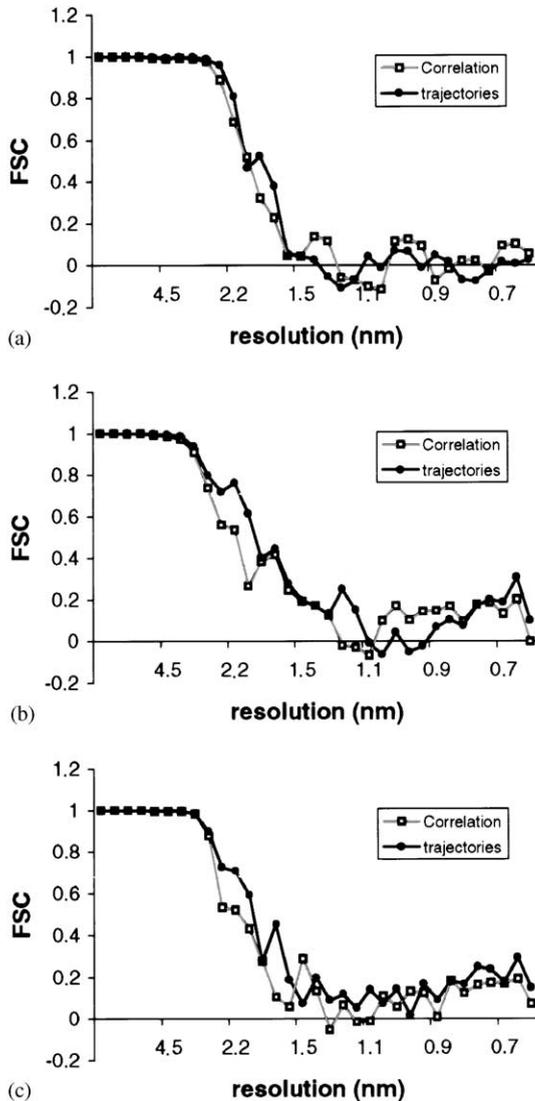


Fig. 6. FSC diagrams for reconstructed models of GroEL presented in Fig. 5. (a) analysis of cryo-negatively stained samples; (b) analysis of unstained samples in a refinement starting from a preliminary reconstruction from top and side views selected within the unstained set; (c) analysis of unstained samples with the refinement based upon models created from stained samples. The diagrams show that the resolution attained in all cases is about 2.2 nm, corresponding the first zero of the CTF.

that their complexity grows up as the group order increases. Our experiments indicates that symmetry has no adverse effects in a trajectory matching

algorithm, especially if derivatives are adopted. The two tables indicate that the percentage of viewing directions correctly determined does not depend upon the symmetry. Moreover, the results are nearly equivalent to those of correlation algorithms. Thus, symmetry does not represent a problem in matching processes: the larger the complexity, the probability of wrong matching diminishes.

With trajectory matching we nicely reconstruct all phantoms, so as to obtain results visually indistinguishable from the original. As for discrepancy indices, which represent unbiased error estimates, it is still true that cross-correlation is a bit better than trajectory matching. This is not surprising since the accuracy of the novel instrument needs to be refined further. What looks appealing is computation time which is reduced to one-sixth of that required by cross-correlation.

The overall performance of trajectory matching has been further evaluated in a real experiment aiming at reconstructing a protein complex from ice embedded samples. In this case the experiment consisted of an exhaustive assignment of viewing directions down to  $1^\circ$  of accuracy. The quality of the results have been judged on the basis of the FSC between the models obtained from two independent subsets of projections. The FSC diagrams indicate that the agreement obtained by trajectory matching is equivalent or even better than that obtained by correlation. This is a valuable indication, FSC being the commonly adopted criterion to assess the correctness of reconstructed models [21]. The GroEL molecule has been reconstructed from two different preparations yielding high- and low signal-to-noise ratio (S/N) images. As expected, both methods converge quickly in the case of high S/N images and slowly in other cases. The models reconstructed from high- and low S/N images are somehow different, with both methods, the difference being more pronounced for trajectory matching. The model reconstructed from cryo-negatively stained samples by trajectory matching and correlation look slightly different in spite of similar FSC diagrams (Fig. 5a and b). On the contrary, unstained samples yield quite similar refined

models (Fig. 5c and d). The difference between the two models from stained molecules is worth being investigated further. One hypothesis deals with a possible heterogeneity of the preparation. We have not attempted to discriminate between different classes within the same type of view (see e.g. [22]) so that all molecules selected have been used for the reconstruction. It might be possible that the two algorithms have found a different solution depending upon the type of molecules which drove the process.

Fig. 5e and f shows the models quickly obtained from unstained molecules, using the reconstruction obtained from stained samples. Interestingly, the reconstruction obtained with trajectories (5f) is almost indistinguishable from that shown in 5d, refined from a starting model coming from unstained images. On the contrary, the results of correlation (5e) seems to preserve some features of the model (5a). After iterated cycles of projection matching the reconstruction of Fig. 5e converges to 5c. This might indicate that trajectory matching is less sensitive to the starting model, being mostly driven by the symmetry properties of original projections. This could be an attractive peculiarity to be tested in the reconstruction of molecular complexes slightly modified with respect to already known structures.

An aspect not fully investigated yet deals with the reliability of trajectories in the presence of significant displacements of projections from a common origin. At present, trajectory matching works only if projections are reasonably well centred. This condition should be fulfilled if the images require preliminary classification, as was the case for the GroEL data. Otherwise the projections must be shift-aligned in advance.

It has been shown that also standard alignment algorithms work more efficiently if the shift and rotation processes are divided. This means that the images should be first centred and then rotationally aligned by use of angular correlation [23], in agreement with the strategy adopted here. Our experience suggests that in images  $64 \times 64$  pixels wide, displacements by 1–2 pixels from the common origin is not prejudicial in the first stages of a refinement. A low complexity alignment based on sinograms [6] could be used from cycle to cycle

of refinement before entering the CA step. In this way, after a coarse image alignment, sinograms and sinogram lines become the only functions used throughout the entire analysis, even in the reconstruction step if the Radon transform method [9] is used. In any event, a strategy to manage the effect of little shifts on the trajectory shapes is presently under study.

## 6. Conclusions

Exhaustive refinement of molecular models with use of trajectories described by sinogram lines in factor space, proves to be an attractive alternative to cross-correlation methods as for speed and accuracy, even though some aspects of the novel technique need to be further investigated. The main advantage of this strategy is represented by a time gain so that, at present, a trajectory-based refinement might be carried out exhaustively, with only a final step left to the time-consuming correlation algorithm. The method, demonstrated here with simulations and with a real molecular model, looks interesting mainly because it represents a frontier advancement in the applications of correspondence analysis. This multivariate statistical method, till now used in EM for mere image classification, proves more and more to be a far reaching tool in the process of creation of preliminary models and in refining their reconstruction.

## Acknowledgements

P.L.B. and S.L. wish to acknowledge the financial support of the Italian Ministry of University and Research (ex 60% and 40%). The GroEL sample was kindly donated by F.U. Hartl, Max Planck Institute for Biochemistry, Martinsried, Germany.

## References

- [1] J.P. Benzecri, *Correspondence Analysis Handbook*, Marcel Dekker Inc., New York, 1992.

- [2] M. Radermacher, T. Wagenknecht, A. Verschoor, J. Frank, *J. Microsc.* 146 (1987) 113–136.
- [3] J. Frank, M. Radermacher, T. Wagenknecht, A. Verschoor, *Ann New York Acad. Sci.* 483 (1986) 77.
- [4] N. Bonnet, E. Simova, S. Lebonvallet, H. Kaplan, *Ultramicroscopy* 40 (1992) 1.
- [5] P.L. Bellon, F. Cantele, S. Lanzavecchia, *Ultramicroscopy* 87 (2001) 187–197.
- [6] S. Lanzavecchia, L. Tosoni, P.L. Bellon, *Comp. Appl. Biosci.* 12 (1996) 531.
- [7] A.B. Goncharov, B.K. Vainshtein, A.I. Ryskin, A.A. Vagin, *Sov. Phys. Crystallogr.* 32 (1987) 504.
- [8] M. van Heel, *Ultramicroscopy* 21 (1987) 111.
- [9] S. Lanzavecchia, P.L. Bellon, M. Radermacher, *J. Struct. Biol.* 128 (1999) 152.
- [10] S. Lanzavecchia, P.L. Bellon, *Bioinformatics* 14 (1998) 212–216.
- [11] K. Braig, Z. Otwinosky, R. Hegde, A.T. Boisvert, A. Joachimiak, A.L. Horwich, P.B. Sigler, *Nature* 371 (1994) 578–586.
- [12] M. Adrian, J. Dubochet, J. Lepault, A.W. McDowell, *Nature* 308 (1984) 32–36.
- [13] M. Adrian, J. Dubochet, S.D. Fuller, J.R. Harris, *Micron* 29 (1998) 145–160.
- [14] K. Karhunen, *Ann. Acad. Sci. Fennicae, Ser A* 137 (1947).
- [15] M. Loève, *Fonctions aléatoires de second ordre*, in: P. Lévy (Ed.), *Processus Stochastique et Mouvement Brownien*, Herman, Paris, 1948.
- [16] J.P. Breaudiere, J. Frank, *J. Microsc.* 144 (1986) 1–14.
- [17] G. Harauz, F.P. Ottensmeyer, *Science* 226 (1984) 936–940.
- [18] L. Lebart, A. Morineau, N. Tabard, *Technique de la description statistique*, Dunod, Paris, 1977.
- [19] G.T. Herman, A. Lent, S.W. Rowland, *J. Theor. Biol.* 42 (1973) 1–32.
- [20] M. van Heel, *Ultramicroscopy* 21 (1987) 95.
- [21] S.J. Ludtke, P.R. Baldwin, W. Chiu, *J. Struct. Biol.* 128 (1999) 82.
- [22] S. Falke, M.T. Fisher, E.P. Gogol, *J. Struct. Biol.* 133 (2001) 203.
- [23] L. Joyeux, P.A. Penczek, *Ultramicroscopy*, in press.