Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 156 (2006) 546–555
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

On the nature of 2D crystal unbending

Debora Gil [a], Jose Maria Carazo [a], Roberto Marabini [b,*]

[a] Centro Nacional de Biotecnología. Ciudad Universitaria de Cantoblanco, Calle Darwin, 3, 28049 Madrid, Spain
[b] Escuela Politécnica Superior, Ciudad Universitaria de Cantoblanco, Calle Francisco Tomás y Valiente, 11, 28049 Madrid, Spain

Received 25 April 2006; received in revised form 24 July 2006; accepted 25 July 2006
Available online 4 August 2006

Abstract

Crystal unbending, the process that aims to recover a perfect crystal from experimental data, is one of the more important steps in electron crystallography image processing. The unbending process involves three steps: estimation of the unit cell displacements from their ideal positions, extension of the deformation field to the whole image and transformation of the image in order to recover an ideal crystal. In this work, we present a systematic analysis of the second step oriented to address two issues. First, whether the unit cells remain undistorted and only the distance between them should be changed (rigid case) or should be modified with the same deformation suffered by the whole crystal (elastic case). Second, the performance of different extension algorithms (interpolation versus approximation) is explored. Our experiments show that there is no difference between elastic and rigid cases or among the extension algorithms. This implies that the deformation fields are constant over large areas. Furthermore, our results indicate that the main source of error is the transformation of the crystal image.

Keywords: Electron microscopy; Unbending; 2D crystal; Interpolation; Approximation

1. Introduction

Transmission electron microscopy (TEM) analysis of biological material is inconvenienced by sample sensitivity to electron radiation. In order to minimize the damage caused by radiation, the electron dose is kept low and consequently the signal-to-noise ratio of the images is poor. Image processing methods have been developed to overcome this problem. Although, in theory, these methods can be applied to a TEM image of any object, they are usually most powerful for objects in which subunits are arranged in a regular manner, such as two-dimensional crystals. In fact, electron crystallography is currently the only way to reach atomic resolution using electron microscopy (Henderson, 2004). However, this is seldom attained and, as in many other techniques, the quality of the experimental data is a major limitation that prevents crystallographic studies reaching atomic resolution. This limitation is in part due to the disorder in the crystal, induced mainly by stretching and bending of the crystal or the presence of cracks, absences or dislocations of the lattice.

In this work, we focus on the correction of lattice defects. In particular, we analyze the importance of the different schemes that can be used to recover an ideal crystal from the experimental one. From the theoretical point of view, crystal unbending fits into either an approximation or an interpolation scheme. Interpolation methods are better suited for crystals that present homogeneous deformations on large areas, while approximation techniques are optimal in the case of highly varying deformations. We have explored representative algorithms for each scheme, including the most popular in the electron microscopy field as implemented in the MRC package (Crowther et al., 1996).

We have also explored the importance of (i) modifying the unit cells so that they suffer the same deformation as the whole crystal (elastic case) versus (ii) modifying only the distance between the unit cells centers (rigid case).

* Corresponding author. Fax: +34 91 3720112.
E-mail address: roberto@cnb.uam.es (R. Marabini).

1047-8477/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2006.07.012

The classical implementation of crystal unbending follows (i) but, if (ii) is close to what really happens during crystal formation, it might introduce a distortion in the unit cells, which would limit the maximum resolution achievable.

The work is divided as follows. Section 2 summarizes the traditional method for lattice correction and introduces the mathematical background needed for understanding the different approaches. Sections 3 and 4 describe the experiments performed, unbending assessment with phantoms in Section 3 and experimental data in Section 4. Finally, Section 5 discusses the results.

2. Theoretical background

2.1. Crystal unbending

In electron crystallography, Fourier transformation is usually employed because the Fourier transform (FT) of a perfect 2D crystal is confined to isolated points. In contrast, the contribution from other sources (i.e., noise) is essentially distributed at random. The standard methodology for 3D reconstruction using 2D crystals consists of the following stages: (i) signal-to-noise ratio (SNR) enhancement by filtering, (ii) correction of the lattice defects (iii) Fourier synthesis: FT normalization, determination of the phase origin and 3D reconstruction. Image processing in crystallography starts by acquiring projections of the crystal at various angles. The 2D projections obtained in the TEM need to be enhanced to increase their SNR. This is done by calculating their FTs and setting to zero all those frequencies not related with the periodic signal (*optical filtering*). In practice, all the Fourier components not in a vicinity of the reciprocal lattice points are forced to have zero value.

Biological crystals are seldom perfect, they present distortions and stretching, are limited in extent and disordered. These imperfections degrade the FT and reduce high resolution information. The degradation of the FT produces a broadening of the peaks at the reciprocal lattice (i.e., they are no longer a delta function but a Gaussian) plus an attenuation of the high frequency terms. As filtering (in Fourier space) is equivalent to averaging all the unit cells in the crystal (in real space), this degradation makes the average unit cell blurred. In other words, crystal imperfections translate into small misalignments of the different unit cells and the final average then turns out blurred.

Lattice unbending was developed by Henderson et al. (1986), to correct these distortions and recover high resolution information. This method identifies the position of each unit cell (usually by cross-correlation) and interpolates a new image with smaller distortion. The new image is further refined, usually through a few rounds of "unbending".

The main concern of crystal unbending is to correct the crystal distortions, that is, its deviation from regular lattice repetition. The standard approach is a three-steps algorithm:

(1) Extract experimental displacements for each unit cell. A region in the crystal image that appears to have the least distortion is selected as a reference area and cross correlated with the original image. Then, the position of the cross correlated peaks is used to generate a list of vectors that map the experimental image into an ideal crystal made by repetitions of the reference area. Each peak is mapped to the nearest theoretical position. Therefore, displacements larger than half a unit cell are not properly computed. The output of this first step is a list of scattered shifts (F_k) at some of the grid points P_k.
(2) Extend experimental shifts. Since the crystal should be corrected or "unbent" by applying the shifts computed in the previous step, the experimental shifts must be extended to produce a deformation field at each image pixel. The way pixel shifts are computed determines the unbending.
(3) Deform crystal image. The original crystal is unbent according to the inverse of the deformation field. The computation of the unbent image involves interpolation of the original gray values to handle values at non integer positions.

Crystal unbending can be stated as a particular case of function extension (Evans, 1993). In 2D crystallography, the goal is finding a vector field, $f(x, y) = (f_x, f_y)$, defined on the whole image from the experimental displacements, $F_k = (F_x, F_y)_k$, given at the grid points P_k (Fig. 1(a)).

A usual requirement in extension processes is that the target unknown function must fulfill some regularity. Such regularity is usually expressed in terms of a minimum internal energy (such as in the case of snakes (Kass et al., 1987) or radial–kernel extensions (Wendland, 2005)) and is related to the way information distributes on the domain. Depending on the similarity between $f(P_k)$ and the given values F_k, there are two main schemes for crystal unbending: interpolation and approximation. The choice between one of the methods depends on the nature of the experimental values.

2.2. Interpolation vs approximation

Interpolation methods (Amidror, 2002) require that the target function passes through the experimental positions (i.e., $f(P_k) = F_k$ as in Fig. 1(b)). From the practical point of view, interpolation schemes have two main advantages. On one hand, they ensure that the experimental information is preserved. On the other, they have an efficient implementation in terms of a weighted sum of the values F_k. Although interpolating functions are continuous, the condition $f(P_k) = F_k$, bounds their global variability by that of the discrete values F_k. Such limitation on the overall regularity of the function f can lead to an artificial blurring of the unit cell if the experimental shifts are highly noisy.

Approximation algorithms (Kass et al., 1987) reach a compromise between similarity to the original F_k values

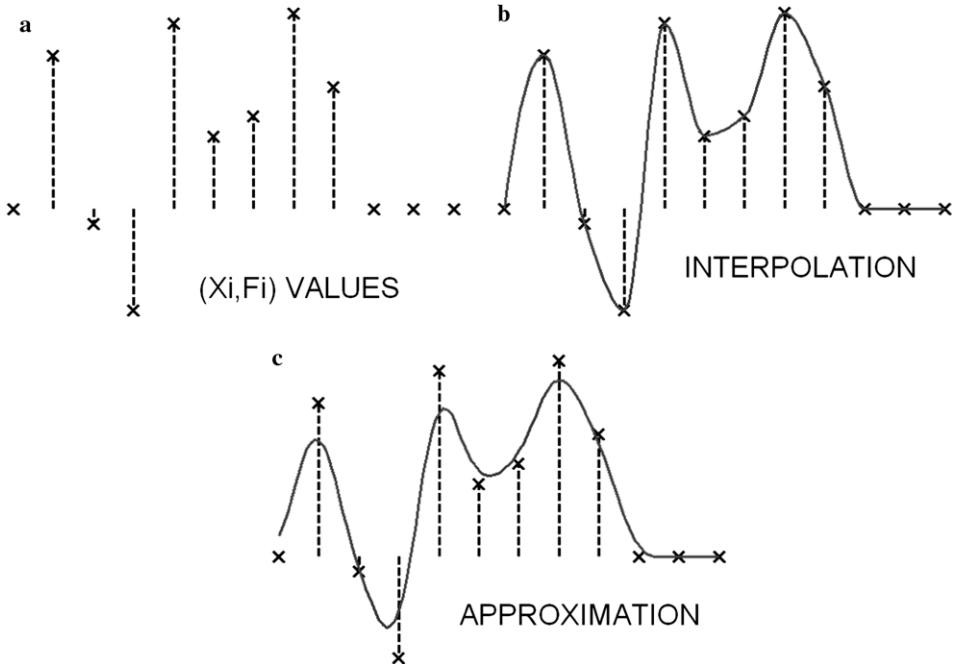

Fig. 1. Interpolation vs approximation: (a) scattered values, (b) interpolation and, (c) approximation.

(accuracy to experimental data) and smoothness (robustness against noise) of the target function (Fig. 1(c)). Usually the above condition is achieved by minimizing an energy functional (see Appendix A.1 for details). Since there is no bound on f regularity, the impact of noisy shifts is suppressed. Unfortunately, depending on the way we solve the energy equation, the influence of the values F_k might be minimum and the approximating function might not resemble the original values. The difference between the approximating function and the values F_k is called *fitting error*.

Fig. 2 illustrates the artifacts that the interpolation and approximation approaches might present. Fig. 2a shows the original image, a continuous deformation field

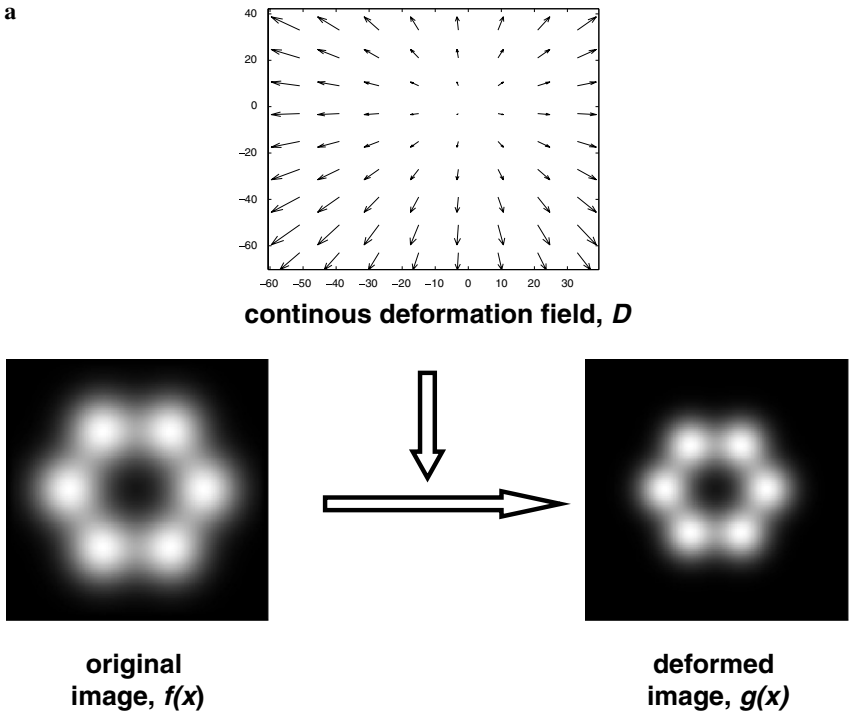

Fig. 2. Interpolation and approximation artifacts.

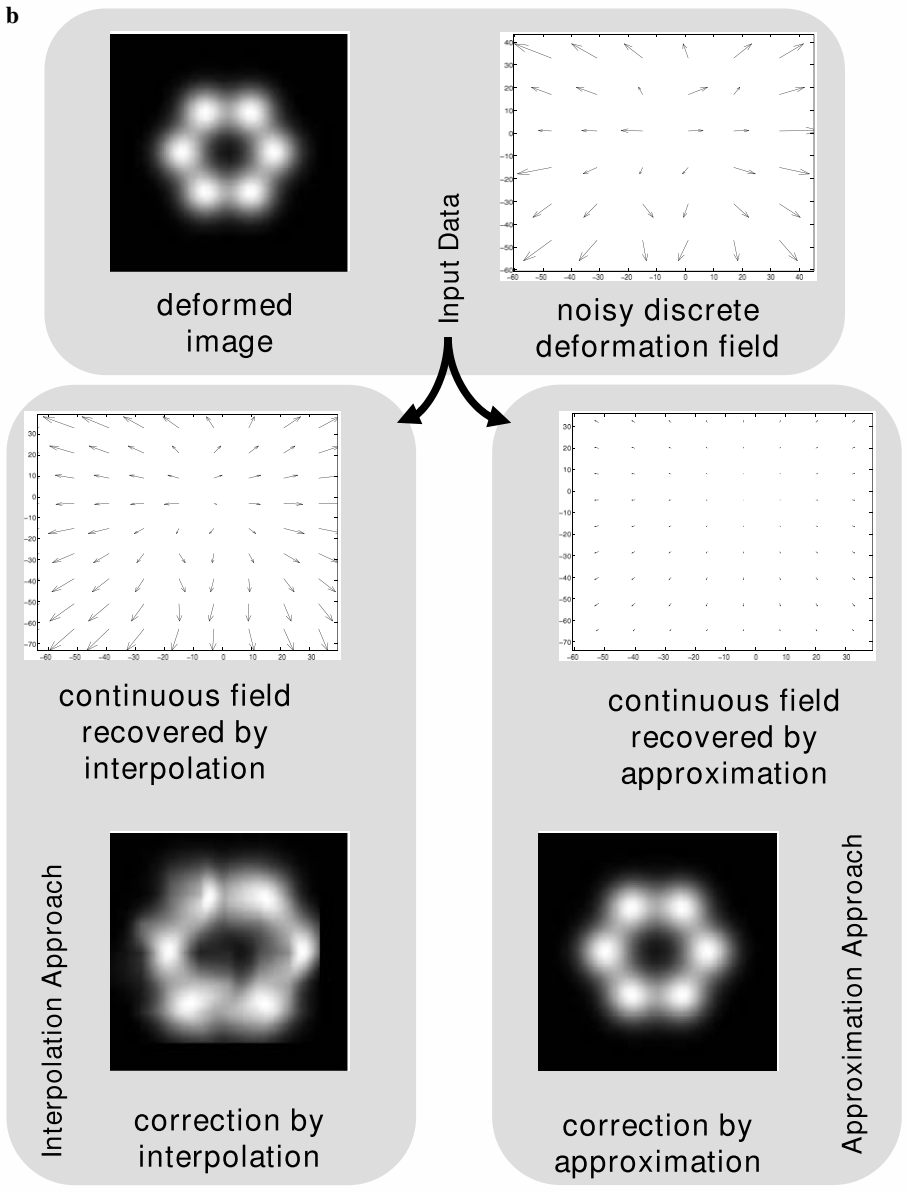

Fig. 2 (*continued*)

and the result of applying the continuous field to the original image, we note that given an original image $f(x)$ and a deformation field D, a deformed image $g(x)$ is defined as $g(x) = f(Dx)$. Therefore, an expansion deformation as D shrinks the signal $f(x)$. In Fig. 2b we start from the deformed image and a discrete sampling of the continuous field with added noise (this field, in the context of crystal unbending, represents the experimental shifts at the unit cell center). The goal is to restore the continuous field from the noisy discrete sampling either by interpolation or approximation (second row Fig. 2b) and, thus, recover the original image (third row Fig. 2b). Since the interpolated field still keeps the noisy nature of the sampling, the transformed image (labelled as *correction by interpolation*) is distorted. Meanwhile the approximated field, although it avoids image distortions, leaves the transformed image (labelled as *correction by approximation*) unchanged.

We have selected two representative methods (see Appendix A.1 for details) of each family (interpolation and approximation). The selected algorithms range from a high precision but low robustness against noise to a lower precision but higher robustness to noise:

- *Linear interpolation* is the most popular way of computing interpolations (Amidror, 2002). It is the fastest accurate algorithm but the most sensitive to noise.
- *Bessel interpolation* provides a more regular interpolation. It is robust against global irregularities but not to local noise.

- *Bessel approximation* guaranties robustness against noise while keeping a low fitting error (precision). Its main drawback is the computational expenses (a few hours for a large crystal).
- *MRC approximation* is a B-spline based method and the standard *de facto* in 2D electron crystallography (Henderson et al., 1986). B-spline based methods require a user defined parameter (the number of control points, see Appendix A.2 for details) that allows controlling the fitting error. As the number of control points increases B-splines get closer to an interpolation scheme.

We note that MRC implementation assumes small displacements. Although in theory should be able to process shift as large as half unit cell, this may require editing the code. This is not the case for the other schemes described.

3. Unbending of synthetic data

3.1. Phantom data generation

Two phantoms have been used in this work. The first phantom is a simple structure: a hollow cylinder surrounded by small spheres (Fig. 3a). This trivial phantom allows a clear identification of algorithm-induced artifacts and, thus, a more reliable algorithm comparison. The second phantom (Fig. 3b) was produced from the crystal structure of a protein (the histone fold heterodimer of the chromatin accessibility complex, PDB entry: 2BYM). In our model the atoms are replaced by blobs (Lewitt, 1990).

Both phantoms are mathematically defined by the crystal vectors, the crystal size (the number of unit cells), and the parameters that define each feature (either a cylinder, a sphere or a blob) contained in the unit cell. From these phantoms, projections with several noise levels were generated. Special care was taken to generate realistic noise by considering the addition of noise not only to the pixel values of the projections, but also to the distances between neighboring unit cells (lattice distortion). We also incorporated feature skipping, so that for each unit cell of the phantom and for each feature (cylinder, sphere or blob), the feature was not taken into account when computing the projection if a random number was below a given threshold. The actual value of the lattice distortions at each unit cell center was computed from experimental data (human AQP2 (Schenk et al., 2005)) and assumed constant within the unit cell (rigid case).

The MRC suite of programs (Crowther et al., 1996) (in our experiments also helped by SPECTRA (Schmid et al., 1993)) was used to process the data and compute the correlation maxima needed to perform the unbending. In our experience image processing of crystals is very robust against noise added to pixels values and feature skipping. Therefore, we only report on the case of lattice distortion noise. In this case, different signal-to-noise ratios only affect the number of experimental shifts (i.e., number of crosscorrelation peaks above certain threshold) considered in the extension step. The impact of the crystal quality has been assessed by creating three noise cases (high, average and low) defined by a threshold value. For each case only those correlation peaks with a correlation above the threshold are taken into account in the computation of experimental shifts. The threshold values are set so that the high quality crystal uses 99% of the correlation peaks, the average one 75% and the low one the extreme value of just 10%.

3.2. Experimental set-up

Our experiments have been designed to address the following topics:

(1) *Elastic vs rigid.* The goal is to check whether the unit cell is deformed during crystal unbending (elastic case) or not (rigid case). The exact phantom shifts at the unit cell centers are used to compute the unbending with two different interpolations schemes. Nearest neighbor interpolation of the deformation field is used to simulate a rigid block-wise unbending, while a linear interpolation is used for a generic elastic unbending.

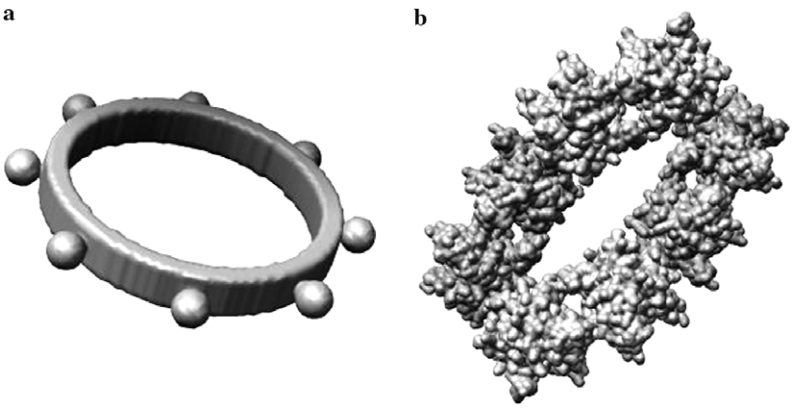

Fig. 3. Computer generated rendering of (a) a phantom consisting of a hollow cylinder (inner radius = 27 Å, outer radius = 32 Å) and 8 small spheres (radius = 4 Å). The spheres are located around the outer surface of the cylinder at each 45 degrees. The value of their z coordinate is 0 Å. (b) Phantom based on the PDB entry: 2 BYM. Sampling rate is 1 Å per pixel.

(2) *Interpolation vs approximation.* In order to determine which is the best scheme, four methods have been tested: linear and Bessel interpolation, and Bessel and MRC-spline approximation. Since MRC performance depends upon the number of control points, we have run MRC with control points in the range [7,60] sampled every 8 units.

We note that, although different approximation/interpolation schemes are used to recover the continuous deformation field, in all cases, the unbent crystal image is computed using bilinear interpolation. The comparison protocol focuses on lattice distortions and follows these steps. First, we produce a perfect crystal without noise and process it with the MRC package skipping the unbending step. This yields a reference unit cell that is compared to the one obtained from a noisy crystal after unbending. Accuracy in resolutions is quantified with the following standard measures:

- *L^2-relative error (RE).* This is an overall normalized measure of dissimilarity between two images, I_1 and I_2, given by:

$$\mathrm{RE} = \frac{\sum_{i,j}(I_1(i,j) - I_2(i,j))^2}{\sum_{i,j} I_1(i,j)^2}. \quad (1)$$

- *Fourier Ring correlation (FRC).* This is a local measure of similarity (correlation) between two images in a given Fourier frequency range (van Heel, 1986). If FT_1 and FT_2 are the Fourier transforms of the images to be compared, then the formula of FRC for a ring defined by the range I_R is: W_1, W_2

$$\mathrm{FRC} = \frac{\sum_{\|W_k\| \in I_R} \mathrm{FT}_1(W_k)\overline{\mathrm{FT}_2}(W_k)}{\sqrt{\sum_{\|W_k\| \in I_R} \|\mathrm{FT}_1(W_k)\|^2}\sqrt{\sum_{\|W_k\| \in I_R} \|\mathrm{FT}_2(W_k)\|^2}} \quad (2)$$

where $\overline{\mathrm{FT}_2}(W_k)$ is the conjugate of a complex number and W_k is the frequency of the different spots.

In the case of methods depending on user defined parameters (i.e., the number of control points in MRC approximation), we have optimized them following two different criteria. The first one is minimizing the global relative error, the second one is maximizing the number of good spots in the Fourier transform after unbending. Spot quality is usually given in terms of the signal-to-noise ratio and it is reported in a logarithmic scale as the IQ number (Henderson et al., 1986).

3.3. Results

Plots in Fig. 4 show FRC curves for a high (Fig. 4a) and an average quality crystal (Fig. 4b) in the case of the cylinder phantom. The plot labelled "Op. Fil" has been produced computing the unit cell without unbending and therefore sets a lower bound. On the other hand, plots labelled "Elast. TrueU" and "Rigid TrueU" are based on interpolation of the exact shifts used to create the phantom. They set the upper bounding for the resolution achievable. The FRC minimum at 0.35–0.40 Å^{-1} is related with the absence of energy in the phantom and does not indicate a real loss of resolution.

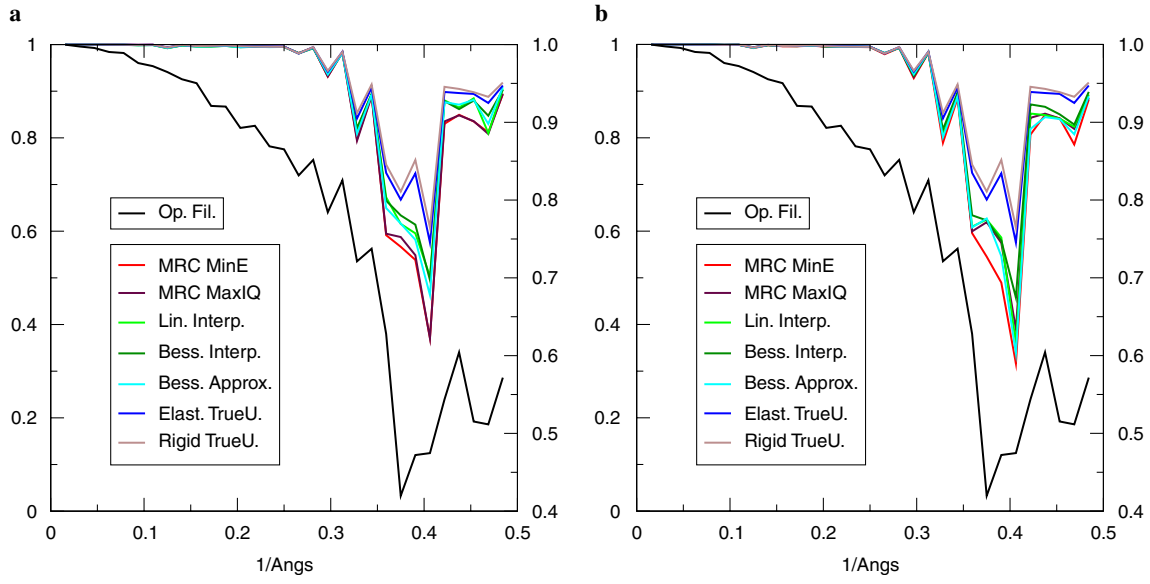

Fig. 4. FRC for the unit cell obtained without unbending (Op. Fil.), the cells produced with MRC approximation (MRC MinE and MRC MaxIQ), linear interpolation (Lin. Interp.), Bessel interpolation (Bess. Interp.), Bessel approximation (Bess. Approx.) and using the true phantom shifts for the elastic (Elast. TrueU) and the rigid schemes (Rigid TrueU). Plots have been computed from (a) a high quality crystal and (b) an average one. In order to emphasize the differences a second y-axis on the right side of the figures has been introduced. The left y-axis corresponds to the "Op. Fil." plot and the right one applies to the rest.

Visual inspection of the FRC plots in Fig. 4 indicates that differences between elastic and rigid unbending are negligible. The phantom used in this tests is a particularly favorable case for the rigid approach, since: (i) the phantom itself has been created following a rigid approach and (ii) the particle (the structure made with the hollow cylinder and spheres) is surrounded by zeros and, therefore, the rigid unbending of the image does not create overlapping at the unit cell border. Even in this ideal situation for the rigid scheme, it does not outperform the elastic approach.

The FRC plots for the different unbending methods ploted are the linear (''Lin. Interp'') and Bessel (''Bess. Interp.'') interpolations and the Bessel (''Bess. Approx.'') and MRC (''MRC MAxIQ'' and ''MRC MinE'') approximations. The label ''MRC MinE'' follows the minimum error criterion while ''MRC MaxIQ'' follows the IQ criterion for parameter optimization. The plots show that there is no significant difference among the algorithms, regardless of the crystal quality and the deformation field.

Table 1 contains the global relative errors given by equation (1) setting I_1 to the phantom unit cell and I_2 to the cell obtained after crystal unbending except for the first line (labelled ''Op. Fil.'') where only optical filtration has been performed. The unbending schemes tested are labelled as in Fig. 4. Those methods that can be applied to experimental data are shown in boldface. The table shows an increase in errors for the cylinder phantom with respect to the PDB-based one that is due to the presence of higher frequencies in the former (i.e., sharp transitions from black to white at the cylinder border).

Another interesting piece of information that can be extracted from the table is the sensitivity of the error in the rigid scheme (Rigid TrueU) to the different phantoms. It compares to the elastic scheme (Elast. TrueU) for the cylindrical phantom but it increases in the case of the PDB-based one. A rigid unbending is a piece-wise constant deformation arranged into blocks centered at the unit cell center (a rigid motive, in general). Therefore it is prone to produce discontinuous deformation fields with discontinuities located at each block border. This can distort unit cells, as in the case of the PDB-based phantom, if the block does not coincide with the rigid motive. The impact of the exact location of the rigid motive is much less important for the cylindrical phantom as the structure is surrounded by zeros. An elastic unbending ensures a continuous deformation but it deforms the unit cell if the unbending is a function of the distance between cell centers. Since this does not depend on the unit cell, the similar error between elastic and rigid unbending for the cylinder phantom, prompts that this is not the case.

Finally, our tests contain some interesting results connected to the MRC unbending method. The MRC method depends on a user supplied parameter (the number of control points) that affects its performance. Its optimal value can be found by exhaustive search of FT spots (MRC MaxIQ), that is, using only experimentally available data. As already introduced, MRC behaves as an approximation algorithm if the number of control points is low, otherwise it works like an interpolation. Fig. 5 shows the relative error versus the number of control points. The plots are not monotonically increasing or decreasing with respect to the number of control points and vary between crystals. This ill-posed behavior implies that the optimal number of control points must be determined for each case by exhaustive search of the error function (MRC MinE). Furthermore, MRC algorithm is very sensitive to the crystal quality (percentage of correlation peaks used as input for the approximation). When all peaks are available the error decreases with the number of control points (Fig. 5 first column), but as the crystal quality lowers, local oscillations appears (Fig. 5 second and third column). We also note that MRC error is very similar for high (interpolation) and low (approximation) number of control points.

4. Unbending of experimental data

In addition to the tests made with synthetic data, a quantitative comparison of the performance of the different algorithms with experimental data has been made. Images from human aquaporin AQP2 (Schenk et al., 2005) were used as test specimen.

As in the case of phantom data, the comparison is based on the resolution obtained with the different approaches. Since in this case the ideal reconstruction is not available the resolution has been estimated (as is traditional in the

Table 1
Relative error for the different unbending strategies

	99% peaks		75% peaks		10% peaks		Total	
	CYL.	2BYM	CYL.	2BYM	CYL.	2BYM	CYL.	2BYM
Op. Fil.	0.59	0.37	0.59	0.37	0.59	0.37	0.59	0.37
MRC MaxIQ	**0.26**	**0.15**	**0.30**	**0.16**	**0.39**	**0.21**	**0.32**	**0.17**
MRC MinE	0.26	0.15	0.30	0.15	0.38	0.18	0.31	0.16
Bess. Approx.	**0.28**	**0.15**	**0.30**	**0.17**	**0.36**	**0.21**	**0.31**	**0.18**
Lin. Interp.	**0.28**	**0.15**	**0.30**	**0.16**	**0.35**	**0.21**	**0.31**	**0.17**
Bess. Interp.	**0.28**	**0.15**	**0.30**	**0.16**	**0.33**	**0.20**	**0.30**	**0.17**
Elast. TrueU	0.20	0.13	0.20	0.13	0.20	0.13	0.20	0.13
Rigid TrueU	0.18	0.26	0.18	0.26	0.18	0.26	0.18	0.26

For high (interpolation) and low (approximation) number of control points.

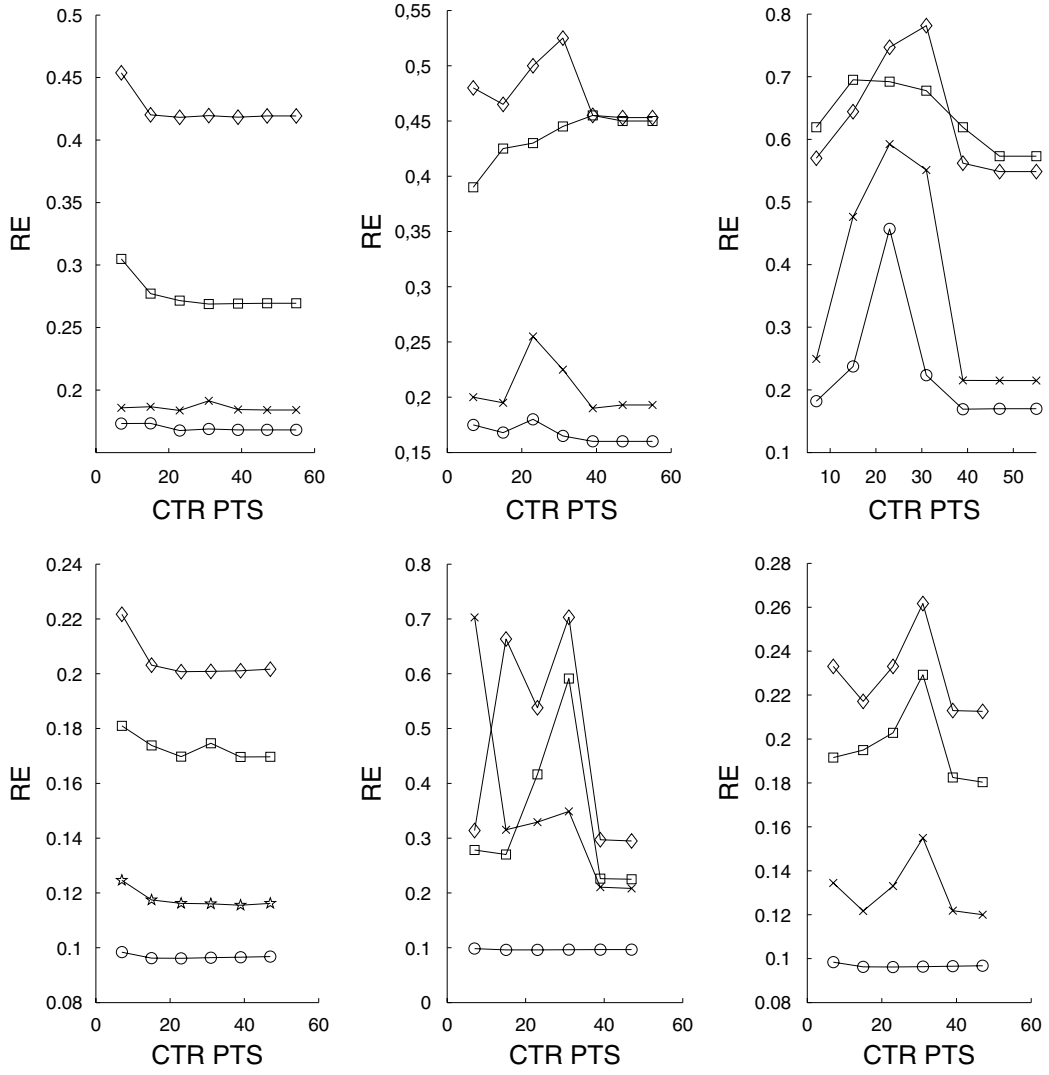

Fig. 5. Relative error dependency on the number of MRC control points for different crystal qualities: high (99% PEAKS), average (75% PEAKS) and low (10% PEAKS). Plots in the first row correspond to the cylindrical unit cell and in the second one to the PDB-based. Each graphic shows the error for four different electron micrographs identified with a distinct marker: circles, squares, crosses and diamonds.

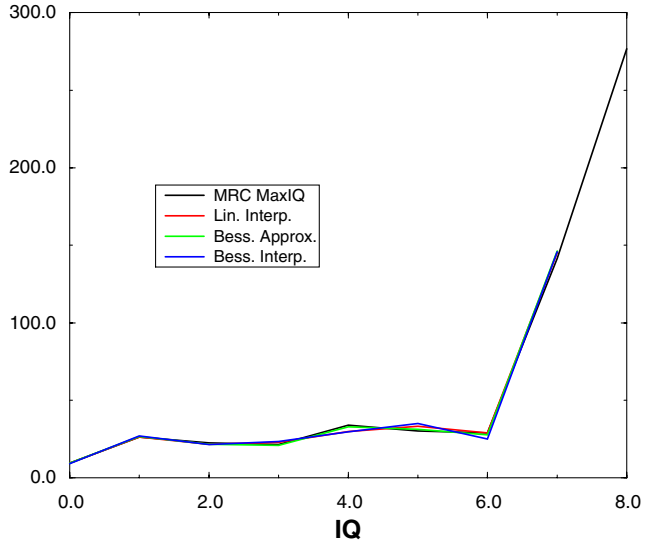

Fig. 6. Average histograms for spots IQ. Labels as in Table 1.

2D-crystal field) by using IQ values Henderson et al., 1986. The overall quality of the crystal is assessed by computing, for each micrograph, the histogram for IQ. The average histograms for the methods considered (MRC MaxIQ, Lin. Interp., Bess. Interp. and Bess. Approx.) are plotted in Fig. 6. As expected from the synthetic data experiments, all methods that can be applied to experimental data behave equally.

5. Discussion and conclusions

This paper addresses two main issues. First, whether the unit cells should be kept undistorted or suffer the same deformation than the whole crystal. Second, determine the unbending algorithm that best adapts to the nature of the crystal deformations.

To answer the first issue, synthetic data was generated using a rigid deformation so that the unit cells remain

always undistorted but not the distance between them. And two different kinds of unit cells were produced: a PDB-based one and a cylinder surrounded by zeros. The latter represents a particularly well suited case for rigid unbending, since by setting to zero the value at the borders, unit cell overlapping is no longer a problem.

The results show that the deformation fields are such that elastic and rigid cases are equivalent, since even for the cylinder phantom, the rigid scheme does not outperform the elastic one. An optimal performance of the rigid unbending requires identification and isolation of each of the rigid motives, which is far from trivial. Therefore we conclude that the elastic approach is the most sensible option.

Regarding the second issue, the performance of the different unbending approaches is very similar as far as parameters are optimum. Although a criterion based on the IQ values allows a reliable parameter optimization, the best approach in terms of computational efficiency is a parameter-free interpolation.

The conditions in which interpolation and approximation schemes give different results are stated in the literature (Amidror, 2002). Interpolation schemes performance decreases for noisy vector fields while approximation algorithms are robust against noise but fail to correct the true deformation if the vector field is highly varying. The equal performance of the different unbending strategies (using experimental displacements) implies that deformation fields are neither noisy nor highly varying. In fact, we can extract two interesting conclusions on the nature of 2D crystal deformation. First, the deformation fields are very regular because interpolation and approximation behave similarly. Second, since elastic and rigid approaches perform equally only for constant deformation fields, we conclude that the deformation fields are constant over large areas.

The sources of error in the unbending process come from the extraction of the experimental shifts, the computation of the continuous deformation by an unbending algorithm and the deformation of the crystal image. The coincidence in the resolutions achieved by the different unbending algorithms tested indicates that the unbending algorithm itself is not the main source of error. Therefore, the sources of error in the unbending process are in order of magnitude: the transformation of the crystal image and the computation of the experimental shifts.

Acknowledgments

The authors thank Andreas Engel and Ansgar Philippsen, for providing AQP2 data, and Drs. San Martín, Scheres, Núñez and Sánchez Sorzano for kindly revising the manuscript. Work supported by Grants GEN 2003-20235-C05-05, BFU2004-00217/BMC, NRCC 050402040003, EU 3D-EM Network of Excellence FP6-502828, EU -512092, NIH-1R01HL67465-01, FIS-04/0683 and MEC-TIN2005-00447.

Appendix A. Function extension algorithms

Function extension algorithms (Evans, 1993) consist in defining a (probably multi-valued) function, $f(p)$, on the whole N-dimensional domain from values, F_k, given at discrete positions P_k. In the context of crystal unbending, since P_k do not need to be evenly spaced, crystal unbending is within the framework of scattered data problems (Wendland, 2005).

A.1. Interpolation schemes

Interpolation methods (Amidror, 2002) split into two main categories: global and local approaches. Global schemes compute a value at a point by averaging all F_k values, while the local ones (linear methods, for instance) only use neighboring values. The first methods require solving a system of equations involving all experimental points in order to obtain F_k weights. This limits their efficiency in the case of large data sets. The second ones need computing a (triangular) mesh of the scattered data points in order to define neighborhoods, which decreases their applicability as the domain dimension increases. In a 2D-domain with a large amount of scattered data, the best choice is a local scheme (Amidror, 2002).

Some local interpolation schemes compute the interpolating function, $f(p)$, at a 2D point, p, by using its barycentric coordinates (Amidror, 2002) to average the nearest F_k values. That is, if a point p is on a triangle given by $\widehat{p_0p_1p_2}$ (see Fig. 7), then:

$$f(p) = a_0F_{p_0} + a_1F_{p_1} + a_2F_{p_2} \tag{3}$$

for a_i the barycentric coordinates of the point p with respect to $\widehat{p_0p_1p_2}$. In the function interpolation context, the triangle is given by the Delanauy triangulation (de Berg et al., 1997) of the scattered data P_k.

Geometric arguments (Amidror, 2002) indicate that barycentric coordinates can be computed in terms of the area ratios $a_i = A_i/A$, for A the area of the triangle $\widehat{p_0p_1p_2}$ and A_i the area of the subtriangle opposite (see Fig. 7) to

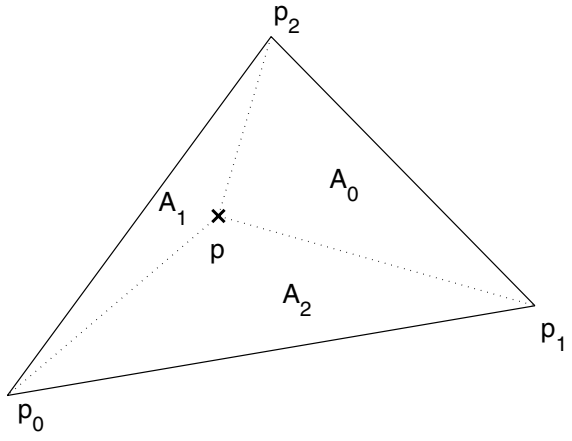

Fig. 7. Interpolation scheme on a triangle.

the vertex p_i. Triangle areas are given by the square norm of the vector product between two of their sides as $A = \|\vec{v}_1\|^2\|\vec{v}_2\|^2 \sin(\alpha)^2$, for α the angle between the vectors $\vec{v}_1 = p_1 - p_0$ and $\vec{v}_2 = p_2 - p_0$. From this geometric point of view, it follows that the interpolation is determined as soon as a norm on the vector space $\mathbb{R}^2$ is given (Lang, 1971). We have explored the following representative algorithms:

(1) *Linear interpolation.* This is the usual way of computing local interpolations (Amidror, 2002) based on area weights. It is given by using the standard Euclidean norm

$$\|\vec{v}\|^2 = v_x^2 + v_y^2$$

to compute the area. The resulting function is piecewise linear and at most continuous unless all values (P_k,F_k) lie on a plane.

(2) *Bessel interpolation.* It corresponds to using the Bessel kernel (Evans, 1993) to compute triangle areas:

$$\|\vec{v}\|^2 = B(\|\vec{v}\|) = \frac{1}{2}\int \frac{e^{-t}e^{-\|v\|^2/(4t)}}{t}dt. \quad (4)$$

Since $B(\|\vec{v}\|) \to 0$ as $\|\vec{v}\| \to \infty$ and $B_0 := B(0) > 0$, we normalize the Bessel kernel between 0 and 1. The norm we use to compute areas is given by $B_N(\|\vec{v}\|) := 1 - B(\|\vec{v}\|)/B_0$. Because in the discrete domain, B_N is of compact support, the degree of differentiability at the vertexes P_k is $\mathscr{C}^\infty$. This translates into an increase in the smoothness of the interpolating function without increasing the computational complexity when compared with linear interpolation.

A.2. Approximation schemes

Approximation schemes are based on an energy functional involving a similarity term and an internal energy. The internal constrains are usually given in terms of the square L^2 norm of the first derivatives of f (Evans, 1993). Similarity is given by straight correlation between the target fitting function and the original values to be approximated. In the case of a discrete data set, this yields:

$$E(f) = \sum (f(P_k) - F_k)^2 + \int_{\mathbb{R}^2} (\partial_x f)^2 + (\partial_y f)^2 dx dy \quad (5)$$

where $\partial_x f$, $\partial_y f$ are the partial derivatives of f. The minimum of the functional E computed by its Euler–Lagrange equations (Evans, 1993) leads to solving a partial differential equation (PDE). The values F_k might be ignored unless f is expressed as a linear combination of basis functions (Wendland, 2005). In this case, the computational efficiency of the approach is not optimal because a system of (linear) equations involving the inversion of a large sparse matrix must be solved.

Among the different strategies that take especial care of scattered data, we have tested the following approaches:

1. *MRC B-spline approximation.* B-spline snakes are a standard way (Brigger et al., 2000) of obtaining a compact representation of functions in terms of N control points. The number of control points is a parameter that controls the tradeoff between accuracy to experimental values and robustness to noise. As the number of control points increases, B-splines get closer to an interpolation scheme. B-splines as implemented by MRC (Henderson et al., 1986) allow a maximum number of 60 control points.
2. *Bessel approximation.* This is a non parametric strategy designed to minimize the impact of the fitting error. The main problem in equation (5) is that the similarity term is of discrete nature and the smoothness constraints are continuous and, thus, they are not comparable. This can be solved, in practice, by changing the discrete function F_k by a continuous function interpolating the values. This converts (5) into a diffusion-reaction equation, which solution is given (Evans, 1993) by convolving the interpolating function with the two dimensional Bessel Kernel (4). By using a continuous model of the discrete sampling the impact of the fitting error is minimized without the need of adjusting any parameter.

References

Amidror, I., 2002. Scattered data interpolation methods for electronic imaging systems: a survey. J. Electron. Imaging 11 (2), 157–176.

Brigger, P., Hoeg, J., Unser, M., 2000. B-spline snakes: a flexible tool for parametric contour detection. IEEE. Trans. Image Process. 9 (9), 1484–1496.

Crowther, R.A., Henderson, R., Smith, J.M., 1996. MRC image processing programs. J. Struct. Biol. 116, 9–16.

de Berg, B., van Kreveld, M., Overmars, M., Schwarzkopf, O., 1997. Computational Geometry—Algorithms and Applications. Springer.

Evans, L., 1993. Partial differential equations. Berkeley Math. Lect. Notes.

Henderson, R., Baldwin, J.M., Downing, K.H., 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.

Henderson, R., 2004. Realizing the potential of electron cryo-microscopy. Q. Rev. Biophys. 37, 3–13.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes, active contour models. Int. J. Comput. Vis. 1 (4), 321–331.

Lang, S., 1971. Linear Algebra, second ed. Addison Wesley.

Lewitt, R.M., 1990. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. J. Opt. Soc. Am. A 7, 1834–1846.

Schenk, A., Werten, P., Scheuring, S., de Groot, B., Muller, S., Stahlberg, H., Philippsen, A., Engel, A., 2005. The 4.5 Å structure of human AQP2. J. Mol. Biol. 350, 278–289.

Schmid, M.F., Dargahi, R., Tam, M.W., 1993. SPECTRA: A system for processing electron images of crystals. Ultramicroscopy 48, 251–264.

van Heel, M., 1986. Noise-limited three-dimensional reconstructions. Optik 73, 83–86.

Wendland, H., 2005. Computational aspects of radial basis function approximation. In: Jetter, K. et al. (Eds.), Topics in Multivariate Approximation and Interpolation. Elsevier.