

Averaging Data Derived from Images of Helical Structures with Different Symmetries

David DeRosier^{1*}, David L. Stokes² and Seth A. Darst³

¹W. M. Keck Institute for Cellular Visualization
Rosenstiel Basic Medical Sciences Research Center
Department of Biology
Brandeis University, Waltham
MA 02454, USA

²Skirball Institute of Biomolecular Medicine
Department of Cell Biology
New York University School of Medicine, New York
NY 10016, USA

³The Rockefeller University
Box 224, 1230 York Avenue
New York, NY 10021, USA

There are many examples of macromolecules that form helical tubes or crystals, which are useful for structure determination by electron microscopy and image processing. Helical crystals can be thought of as two-dimensional crystals that have been rolled into a cylinder such that two lattice points are superimposed. In many real cases, helical crystals of a particular macromolecule derive from an identical two-dimensional lattice but have different lattice points superimposed, thus producing different helical symmetries which cannot be simply averaged in Fourier-space. When confronted with this situation, one can select images corresponding to one of the observed symmetries at the expense of reducing the number of images that can be used for data collection and averaging, or one can calculate separate density maps from each symmetry, then align and average them together in real-space. Here, we present a third alternative, which is based on averaging of the Fourier-Bessel coefficients, $g_{n,l}(r)$, and which allows the inclusion of data from all symmetry groups derived from a common two-dimensional lattice. The method is straightforward and simple in practice and is shown, through a specific example with real data, to give results comparable to real-space averaging.

© 1999 Academic Press

Keywords: electron microscopy; image analysis; helical crystals; Fourier-Bessel coefficients; circumferential vector

*Corresponding author

Introduction

Electron microscopy (EM), combined with image processing, is becoming an increasingly powerful tool for structural biology. Structures from non-crystalline specimens of large macromolecular assemblies can be determined to resolutions approaching 7 Å for highly symmetric structures like isocahedral viruses (Bottcher *et al.*, 1997; Conway, *et al.*, 1997) or around 15 Å for asymmetric structures (Malhotra *et al.*, 1998). Currently, for higher resolution information, crystalline specimens are required. There are generally two types of such crystalline specimens, either single-layered, two-dimensional (2D) crystals or helical assemblies. While the largest impact of EM and image processing to date has come from the analysis of 2D crystals (Henderson *et al.*, 1990; Jap *et al.*, 1991; Kimura *et al.*, 1997; Kühlbrandt *et al.*, 1994; Nogales *et al.*, 1998), helical processing methods are developing rapidly and have resulted in struc-

tures in the 10-8 Å resolution range, as in the work on the acetylcholine receptor (Unwin, 1993, 1995), the Ca²⁺-ATPase of sarcoplasmic reticulum (Zhang *et al.*, 1998), and the structure of the bacteria flagellar filament of *Salmonella* (Mimori *et al.*, 1995; Morgan *et al.*, 1995). Higher resolution structures are anticipated in the near future.

In many cases, macromolecules form helical crystals that fall into groups or families described by different helical symmetries. Examples include helical crystals of acetylcholine receptor (Toyoshima & Unwin, 1990; Unwin, 1993), Ca²⁺-ATPase of sarcoplasmic reticulum (Toyoshima *et al.*, 1993; Zhang *et al.*, 1998), and *Escherichia coli* RNA polymerase (Polyakov *et al.*, 1995). To see how these families arise, recall that a helical lattice can be derived by rolling a 2D lattice into a cylinder and superimposing a given lattice point with some other lattice point in the lattice. The circumferential vector is defined as the vector that joins the two lattice points. Figure 1 shows a 2D lattice of points and two possible circumferential vectors, corresponding to the 5,2 vector and the 5,3 vector. To construct the helix, one cuts along two lines that pass through the ends of, and are perpendicular to, the circumferential vector (shown as broken lines

Abbreviations used: EM, electron microscopy; 2D, two-dimensional, 3D, three-dimensional.

E-mail address of the corresponding author: derosier@binah.cc.brandeis.edu

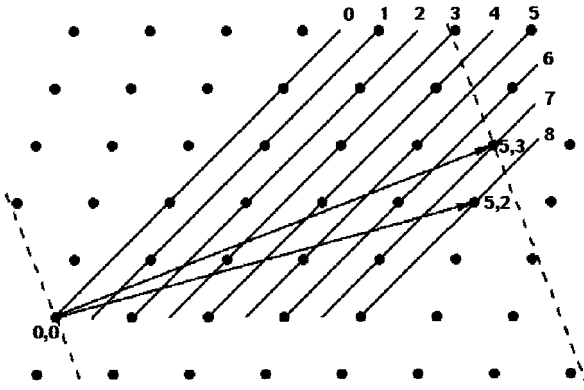


Figure 1. A two-dimensional lattice from which different helical lattices can be obtained. The 5,2 and 5,3 circumferential vectors, corresponding to two different possible helical lattices, are shown as solid vectors. A set of eight lattice lines are shown and numbered. The broken lines indicate where cuts should be made in order to roll up the 2D lattice into a helical net corresponding to the 5,3 lattice vector.

for the 5,3 circumferential vector). Then one glues the two cut edges together to make the cylindrical helix. The lattice lines of the 2D lattice become helical lattice lines. The set of lattice lines corresponding to Miller indices $h = 1$ and $k = 2$ are shown. In the Fourier transform of the 2D lattice, this set of lines gives rise to the (1,2) Bragg reflection. In the Fourier transform of the helical lattice, these helical lattice lines will give rise to one of the layer lines (Toyoshima *et al.*, 1993; Toyoshima & Unwin, 1990). The order, n , of this layer line is equal to the number of times these lattice lines cross the circumferential vector. In the case of the 5,3 circumferential vector, the order of the layer-line derived from the 1,2 lattice lines is $n = 7$, whereas in the case of the 5,2 circumferential vector, the order would be $n = 8$. There are, of course, an infinite number of circumferential vectors, and therefore an infinite number of helical indexing schemes (or symmetries) that can be derived from this one lattice.

To obtain a three-dimensional (3D) map from electron micrographs of a helical structure, one begins by taking the Fourier transform of each image and determining the indexing scheme, which specifies the circumferential vector and the position and order of each layer line in the transform. Layer lines consist of Fourier-Bessel coefficients, $G_{n_i}(R, Z_i)$. To improve the signal-to-noise ratio, the individual $G_{n_i}(R, Z_i)$ values from different individual crystals can be averaged, provided they all have the same order (n), which is equivalent to having the same circumferential vector. If different crystals have different indexing schemes (i.e. different circumferential vectors), then the $G_{n_i}(R, Z_i)$

values cannot be averaged. In practice, there have been two solutions to this problem. One solution is to discard all the images that do not belong to a particular family (Toyoshima & Unwin, 1990), which significantly reduces the amount of data that can be used from a set of micrographs. The other solution is to calculate separate 3D density maps from each set of averaged $G_{n_i}(R, Z_i)$ values (one map for each symmetry), then to scale, align, and average these maps in real-space (Zhang *et al.*, 1998). While this procedure has been shown to work well, in practice it is laborious and requires significant Fourier averaging for each individual map to ensure accurate compensation for the contrast transfer function and alignment in real-space. Here, we present an alternative solution to this averaging problem that allows the inclusion of data from all individual helical assemblies, regardless of the differences in the circumferential vector. This is made possible by converting all the $G_{n_i}(R, Z_i)$ values into $g_{n_i}(r, Z_i)$ values. The method is straightforward and only requires that the underlying 2D lattice be the same for all crystals. To demonstrate the method, we apply it to data from Ca^{2+} -ATPase helical crystals and obtain results comparable to real-space averaging (Zhang *et al.*, 1998).

Background

The 3D distribution of scattering density, ρ , is generated from its Fourier-Bessel transform, $G_{n_i}(R, Z_i)$, in two steps:

$$g_{n_i}(r, Z_i) = \int G_{n_i}(R, Z_i) J_n(2\pi i R) R dR \quad (1)$$

and then

$$\rho(r, \phi, z) = \frac{1}{c} \sum_l \sum_{n_l} g_{n_l}(r, Z_l) e^{i n_l \phi} e^{-2\pi i Z_l z} \quad (2)$$

where r , ϕ , and z are the cylindrical coordinates in real space, and n_l and Z_l denote the order and height of the layer lines of the Fourier-Bessel transform.

The symmetry of a particular helical structure is described by a helical net, a set of equivalent points around the cylindrical structure. If this net is cut along a line parallel with the helix axis, unrolled, and the resulting strip repeated side-to-side, an infinite 2D lattice is generated. This 2D lattice is characterized by two unit lattice vectors, \mathbf{a} and \mathbf{b}^\dagger . The circumferential vector, \mathbf{c} , corresponds to a lattice vector:

$$\mathbf{c} = h_c \mathbf{a} + k_c \mathbf{b} \quad (3)$$

where h_c and k_c are integers. Each family of lines in the helical net corresponds to a particular layer line, n_l and Z_l and also corresponds to a particular set of lattice lines (h, k) in the 2D lattice. What we show below is that the value of $g_{n_i}(r, Z_i)$ is the same as the Fourier coefficient $F(r, h, k)$ derived

[†] Vectors are denoted by boldface; unit vectors are denoted by $\hat{}$.

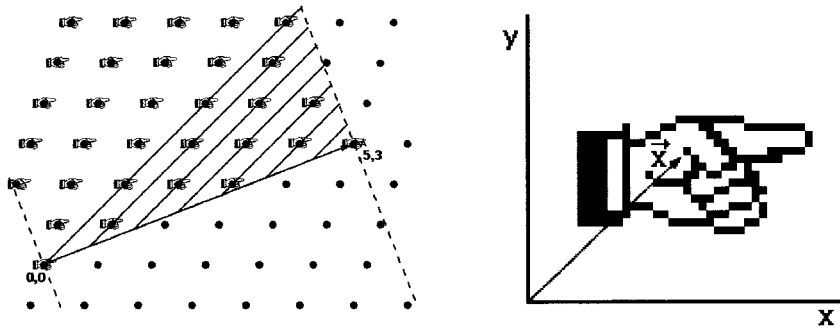


Figure 2. (a) A portion of the lattice that will generate the helical net convoluted with a subunit, in this case a hand. The helical lattice in this case corresponds to the 5,3 circumferential vector. (b) The local coordinate system used to relate the Fourier transform of the subunit to the Fourier transform of the helical structure.

from the 2D lattice. Thus, as long as the underlying 2D lattice remains unchanged, the $g_{n,l}(r)$ values for a given (h,k) can be averaged together regardless of the circumferential vector defining the helical symmetry. This analysis draws on the work of Klug *et al.* (1958), who showed the intimate relationship between a 2D crystal and a helical crystal.

Results

Theoretical

First, let us choose the density distribution at an arbitrary radius, r . When this cylindrical section is unrolled, a planar section through the resulting 2D crystal is generated (Figure 2(a)). The section through a single subunit of the 2D crystal corresponds to a 2D density distribution, $\rho(r, \mathbf{x})$, where \mathbf{x} is the vector pointing to the element whose scattering density is ρ (Figure 2(b)). The contribution of $\rho(r, \mathbf{x})$ to the Fourier coefficient, $F(r, h, k)$, is $\rho(r, \mathbf{x})e^{2\pi i \mathbf{h}^* \cdot \mathbf{x}}$, where $\mathbf{h}^* = h\mathbf{a}^* + k\mathbf{b}^*$, and \mathbf{a}^* and \mathbf{b}^* are the lattice vectors of the reciprocal lattice. The corresponding term from the helical structure, $g_{n_l}(r, Z_l)$, is $\rho(r, \phi, z)e^{i(-n_l\phi + 2\pi Z_l z)}$.

To show that these two are the same, we convert the lattice lines (h, k) to n and Z using the geometry shown in Figure 3. The spacing between and perpendicular to the (h, k) lines is d_h . The separation of lines along the circumferential vector, \mathbf{c} , is $d = d_h / \cos(\gamma)$. The order of the (h, k) lines is the number of

times they cross the circumferential vector:

$$n = \frac{|\mathbf{c}|}{d} = \frac{|\mathbf{c}| \cos(\gamma)}{d_h} \quad (4)$$

Since the vector \mathbf{h}^* points along the direction of d_h and has a length of $1/d_h$, we write:

$$n = \mathbf{h}^* \cdot \mathbf{c} \quad (5)$$

The value of Z is

$$Z = \frac{1}{p} = \frac{\sin(\gamma)}{d_h} \quad (6)$$

Since \mathbf{c}^* points along the longitudinal or helix axis (perpendicular to \mathbf{c}), we can rewrite equation (6) as:

$$Z = \frac{\mathbf{h}^* \cdot \mathbf{c}^*}{|\mathbf{c}^*|} \quad (7)$$

Now we need to convert the vector \mathbf{x} into its components, ϕ and z . The magnitude of the circumferential vector \mathbf{c} corresponds to an angle of 2π . Thus, the angle, ϕ , corresponding to \mathbf{x} is:

$$\phi = -\frac{\mathbf{x} \cdot \mathbf{c}}{|\mathbf{c}|} \times \frac{2\pi}{|\mathbf{c}|} = -2\pi \frac{\mathbf{x} \cdot \mathbf{c}}{|\mathbf{c}|} \quad (8)$$

The reason for the negative sign is a convention in which ϕ is taken along the negative direction of \mathbf{c} . This arises because the helical net is opened with the inside facing up. The value of z is just the projection of \mathbf{x} onto \mathbf{c}^* :

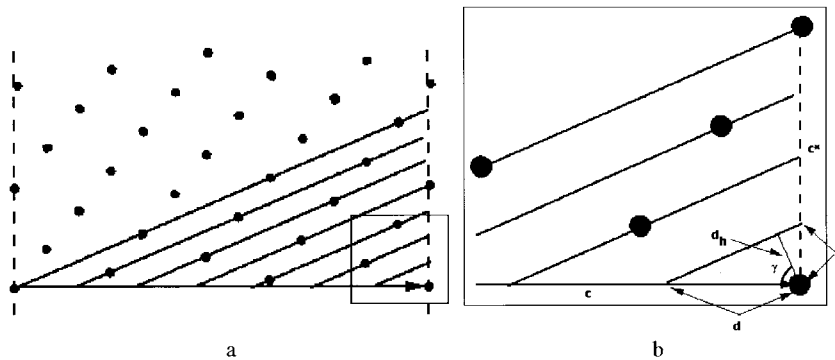


Figure 3. The helical lattice corresponding to the 5,3 circumferential lattice. (a) The unrolled helical lattice. (b) An enlargement of the portion shown in the inset. The various parameters used in the derivation are depicted.

$$z = \frac{\mathbf{x} \cdot \mathbf{c}^*}{|\mathbf{c}^*|} \quad (9)$$

Now substituting equations (5), (7), (8), and (9), we can calculate the contribution of $\rho(r, \mathbf{x})$ to g_n :

$$\rho(r, \phi, z)e^{i(-n\phi+2\pi Zz)} = \rho(r, \mathbf{x})e^{2\pi i \left(\frac{(\mathbf{h}^* \cdot \mathbf{c})(\mathbf{x} \cdot \mathbf{c})}{|\mathbf{c}|^2} + \frac{(\mathbf{h}^* \cdot \mathbf{c})(\mathbf{x} \cdot \mathbf{c}^*)}{|\mathbf{c}^*|^2} \right)} \quad (10)$$

Since:

$$\begin{aligned} \hat{\mathbf{c}} = \frac{\mathbf{c}}{|\mathbf{c}|} \quad \text{and} \quad \hat{\mathbf{c}}^* = \frac{\mathbf{c}^*}{|\mathbf{c}^*|} \frac{(\mathbf{h}^* \cdot \mathbf{c})(\mathbf{x} \cdot \mathbf{c})}{|\mathbf{c}|^2} + \frac{(\mathbf{h}^* \cdot \mathbf{c}^*)(\mathbf{x} \cdot \mathbf{c}^*)}{|\mathbf{c}^*|^2} \\ = (\mathbf{h}^* \cdot \hat{\mathbf{c}})(\mathbf{x} \cdot \hat{\mathbf{c}}) + (\mathbf{h} \cdot \hat{\mathbf{c}}^*)(\mathbf{x} \cdot \hat{\mathbf{c}}^*) \end{aligned} \quad (11)$$

where $\hat{\mathbf{c}}$ and $\hat{\mathbf{c}}^*$ are unit vectors along the equator and longitudinal axis (they form an orthogonal pair of axes). The products in the equation then correspond to the dot product of the vectors \mathbf{h}^* and \mathbf{x} :

$$\rho(r, \phi, z)e^{i(-n\phi+2\pi Zz)} = \rho(r, \mathbf{x})e^{2\pi i \mathbf{h}^* \cdot \mathbf{x}} \quad (12)$$

Thus, if we merely sum up all the elements of π , we prove our result that $g_{n_l}(r, Z_l)$ is equal to $F(r, h, k)$ where $n_l = \mathbf{h}^* \cdot \mathbf{c}$ and $Z_l = \mathbf{h}^* \cdot \mathbf{c}^*/|\mathbf{c}^*|$. This result is true for any arbitrary value of r , thus the result can be generalized to say that $g_{n_l}(r, Z_l)$ is equal to $F(r, h, k)$, which is what we set out to show.

Our next step is to show that a helical structure derived from a different circumferential vector, \mathbf{c} , generates the same functions, g . To show this is relatively straightforward since the values of the g all correspond to values of $F(h, k)$. Thus, choosing a different circumferential vector will yield the same values $F(h, k)$ but with different corresponding values of n_l and Z_l : $n'_l = \mathbf{h}^* \cdot \mathbf{c}^*$ and $Z'_l = \mathbf{h}^* \cdot \mathbf{c}^*/|\mathbf{c}^*|$. Thus, given the two circumferential vectors, it is straightforward to determine corresponding pairs n_l, Z_l and n'_l, Z'_l . For examples, see the indexing of the Fourier-Bessel transforms of the acetylcholine receptor (Toyoshima & Unwin, 1990) or the Ca^{2+} -ATPase (Toyoshima *et al.*, 1993) where the values of h and k are given.

Two details remain. First, before one can average the corresponding $g_{n_l}(r, Z_l)$ from two different structures, one must determine the corresponding values of r and r' . Let us set the 2D lattice in the middle of the subunits in the 2D crystal of subunits. When the strip determined by the circumferential vector is cut out and rolled up, the radius of these features is $r_c = |\mathbf{c}|/2\pi$. Similarly for the helix derived from the vector \mathbf{c}' , $r'_c = |\mathbf{c}'|/2\pi$. Thus, we expect that:

$$g_n(r_c + \delta r, Z) = g_n(r'_c + \delta r, Z') \quad (13)$$

δr is assumed to be the same for the two structures, since the thickness of the subunit is not expected to change on rolling the crystal into a tube. Second, one must fix a common phase origin,

which is straightforward for tubes with different symmetries. One class of tubes is chosen as the reference to which the data sets of all other classes will be fitted. One begins by resetting the values of n'_l for corresponding $g_{n_l}(r, Z_l)$ to those of the reference. After resetting the values of n , one can minimize phase residuals between the pairs of corresponding functions $g_{n_l}(r, Z_l)$ by applying phase shifts of $e^{in\Delta\phi}e^{2\pi i\Delta z Z_l}$. These functions behave just like $G_{n_l}(R, Z_l)$, and previously developed algorithms can be readily adapted (Hanein & DeRosier, 1999).

A practical example

To generate an 8 Å resolution structure of the Ca^{2+} -ATPase from sarcoplasmic reticulum, Zhang *et al.* (1998) used averaged $G_{n_l}(R, Z_l)$ values from three different helical symmetries described by $(-21,6)$, $(-22,6)$, or $(-23,6)$ circumferential vectors. Three independent 3D density maps were calculated from each of these averaged data sets, then these three maps were scaled, aligned, and averaged in real-space to generate the final structure. We started with the same three sets of averaged $G_{n_l}(R, Z_l)$ values from the three symmetries and averaged the $g_{n_l}(r, Z_l)$ values using our new method outlined above.

Briefly, $g_{n_l}(r, Z_l)$ values were calculated in 1 Å steps of r from each of the three sets of averaged $G_{n_l}(R, Z_l)$ values that had been truncated at 10 Å resolution. As was done for the earlier real-space averaging, the $(-22,6)$ data set was chosen as the "reference" symmetry, and the radial shift, $\Delta r = r_c - r'_c$, for each of the other symmetries with respect to the reference was determined by minimizing the following function with respect to Δr :

$$\text{residual}(\Delta r) = \frac{\sum_{n_l, r} \omega |g_{n_l}^{\text{ref}}(r, Z_l) - s g_{n'_l}^{\text{test}}(r + \Delta r, Z_l)|}{\sum_{n_l, r} \omega |g_{n_l}^{\text{ref}}(r, Z_l) + s g_{n'_l}^{\text{test}}(r + \Delta r, Z_l)|} \quad (14)$$

where n and l are related to n' and l' by the criteria described above; r denotes the radius range over which the calculations were performed. The individual maps were carefully inspected to choose a radius range that just included the protein density. This is because the $g_{n_l}(r, Z_l)$ values at low and high radii (i.e. inside and outside the cylindrical section that includes the protein density) contain only noise. In essence, we forced the $g_{n_l}(r, Z_l)$ values outside of the relevant radius range to zero, which is akin to the "solvent-flattening" procedure used in protein crystallography (Crowther, 1967; Wang, 1985). The scale factor, s , was determined as follows:

$$s = \frac{\sum_{n_l, r} |g_{n_l}^{\text{test}}(r, Z_l)|}{\sum_{n_l, r} |g_{n_l}^{\text{ref}}(r, Z_l)|} \quad (15)$$

For each test symmetry, a clear minimum was found, and these minima were the same whether

the weighting factor, w , was chosen as 1.0 or as $g_{n,l}^{\text{ref}}(r)$. For the $(-21,6)$ family, $\Delta r = 9 \text{ \AA}$, and for the $(-23,6)$ family, $\Delta r = -13 \text{ \AA}$. Plots of the real component of g (after scaling) as a function of $r + \Delta r$ (where $\Delta r = 0$ for the reference family) for selected layer lines are shown in Figure 4, illustrating the excellent correspondence, even at large Z .

Next, $g_{n_l}(r, Z_l)$ values were calculated from each of the three complete sets of averaged $G_{n_l}(R, Z_l)$ values, which extended to 6.7 \AA resolution. An average set of $g_{n_l}(r, Z_l)$ values was then calculated as follows:

$$g_{n,l}^{\text{avg}}(r) = g_{n,l}^{(-22/6)}(r) + s' g_{n',l'}^{(-21/6)}(r + \Delta r') + s'' g_{n'',l''}^{(-23/6)}(r + \Delta r'') \quad (16)$$

The scale factors, s' and s'' , were determined as described in equation (14), but were further multiplied by the square-root of the total number of molecules that were averaged together to generate the averaged data sets for that particular symmetry family. This weighting scheme between the symmetry groups is identical with the weighting

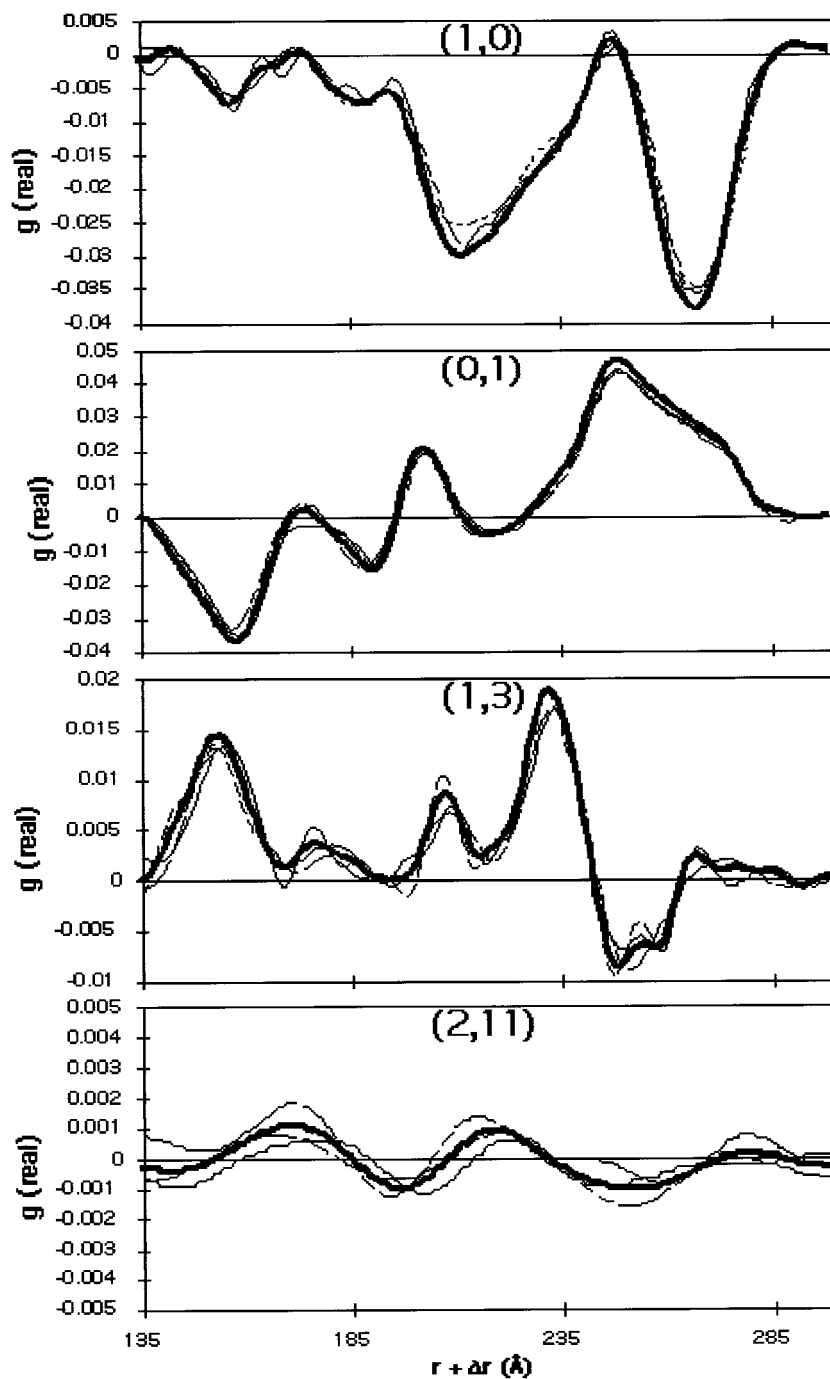


Figure 4. Plots of the real component of $g_{n_l}(r, Z_l)$ (after scaling) as a function of $r + \Delta r$ for selected layer lines. The thick black line corresponds to the scaled average ($\Delta r = 0$), the thin solid line denotes the reference $(-22,6)$ ($\Delta r = 0$), the large dashes denote $(-21,6)$ ($\Delta r = 9 \text{ \AA}$), and the smaller dashes denote $(-23,6)$ ($\Delta r = -13 \text{ \AA}$).

scheme used by Zhang *et al.* (1998) to calculate the real-space averaged structure.

Because these Ca^{2+} -ATPase crystals have $p2$ symmetry, the two-fold phase residual of the final averaged data can be used to assess the efficacy of the averaging procedure. After converting the data back into $G_{n_i}(R, Z_i)$ values, the amplitude-weighted, two-fold phase residuals for: (1) the real-space averaged structure, (2) the little- g averaged structure, and (3) the reference structure calculated from the $(-22,6)$ crystals without averaging with the other families, were compared (Figure 5). As described by Zhang *et al.* (1998), equatorial data as well as data lower than 0.1% of the maximal off-equatorial amplitude were excluded from the phase residual calculation. In contrast to Zhang *et al.* (1998), however, the data were not edited by manually specifying the radial extent of each layer-line to exclude data with poor two-fold phase residuals. This results in higher two-fold phase residuals than those reported by Zhang *et al.* (1998) but makes the different data sets more comparable. Both the real-space and little- g averaged data show substantially improved two-fold phase residuals over the best single, unaveraged data set out to about 9 Å resolution. In each resolution bin, the real-space averaged data and the little- g residuals differ by no more than 2° .

Discussion

We have shown that data from electron images of helical assemblies derived from the same 2D lattice but having different circumferential vectors can be averaged using a very straightforward method that involves converting the $G_{n_i}(R, Z_i)$ values into $g_{n_i}(r, Z_i)$ values, and directly averaging these after applying a phase shift to bring them to a common phase origin and after an appropriate radial shift to account for the different diameters of the respective helical structure. The method provides insight into the functions that are used to

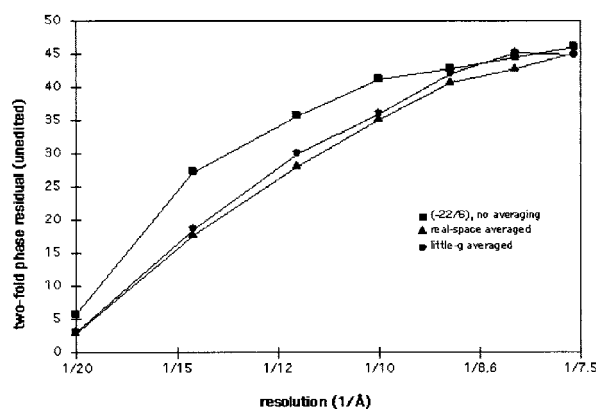


Figure 5. Two-fold phase residuals calculated in resolution bins (∞ –20.0 Å; 20.0–14.4 Å; 14.4–11.55 Å; 11.55–10.0 Å; 10.0–8.94 Å; 8.94–8.16 Å; 8.16–7.56 Å) for the $(-22/6)$ symmetry family, the real-space averaged data (Zhang *et al.*, 1998), and the little- g averaged data.

represent 3D structures with helical symmetry and provides a simple way of combining data. We have shown using a specific example, $G_{n_i}(R, Z_i)$ values collected and averaged from three symmetry families of $p2$ Ca^{2+} -ATPase crystals, that the method gives averaged data with improvements comparable to a real-space averaging method used previously (Zhang *et al.*, 1998).

Our method has several advantages (in addition to its simplicity) and disadvantages when compared with real-space averaging methods. For real-space averaging methods: (1) the helical diffraction patterns must be independently indexed correctly. (2) There must be enough tubes in each symmetry group so that a reliable, CTF-corrected structure can be determined for each group. This is so that the real-space alignment is accurate. What constitutes enough tubes depends, of course, on the signal-to-noise and the resolution of the analysis, among other things. (3) The tubes must have similar unit cells within each symmetry group (but not between symmetry groups). (4) One must be able to mask out an equivalent molecule or subunit in each independent structure and if improvements in two-fold residuals are desired, one-half of the unit cell must be treated independently from the other half during alignment and averaging.

For the little- g averaging: (1) The method can be used objectively to check the indexing of the helical diffraction patterns. Since the functions $g_{n_i}(r, Z_i)$ are expected to be the same between symmetry groups, it is not necessary at the outset to determine circumferential vectors for all classes of tubes (i.e. index all the helical diffraction patterns). Assume the indexing is known for only one class, the reference class. One can compare $g_{n_i}(r, Z_i)$ values with the reference class and thereby determine which layer lines correspond to one another. After establishing the correspondence between layer lines, one can align all the $g_{n_i}(R, Z_i)$ values radially by minimizing the residuals using the amplitudes alone. One can then assign to all classes of tubes, the orders, n_i , for the reference tube. One can finish the alignment by selecting a phase origin (ϕ_0, z_0) which minimizes the phase residual (Hanein & DeRosier, 1999). This procedure could be used either to check the indexing scheme, or to index from the beginning. (2) The method can be used even if symmetry groups contain only one tube. (3) All the tubes must have similar unit cells, even between symmetry groups. (4) No assumptions need to be made about what features comprise equivalent molecules or subunits.

A major advantage of the little- g averaging method is the ability to incorporate data from symmetry groups that contain only one tube. However, a major disadvantage is that the unit cells between the symmetry groups must be similar. The unit cell of a helical assembly can vary from crystal to crystal, even among crystals with the same circumferential vector (Yonekura *et al.*, 1997), because of small variations in angular or axial subunit spacing (i.e. non-isomorphism). As with any other aver-

aging strategy that does not involve aligning real-space densities within a unit cell, these variations are not accounted for and will destroy the coherence of the data as the resolution increases. Of course, the severity of the problem depends on a number of factors, including the resolution expected for the analysis, and the magnitude of the unit cell variability with respect to that resolution. For the specific example given here, the amount of unit cell variability was small compared to the 8 Å resolution (Zhang *et al.*, 1998) but this could perhaps account for the very slightly worse (2°) two-fold phase residuals of the little-*g* averaged data compared with the real-space averaging. Thus, in cases where sufficient data within each symmetry group are available, the real-space averaging method should be used to take into account unit cell differences. However, the little-*g* averaging method may prove beneficial in cases where such complete data sets are not available.

Acknowledgments

We acknowledge grant support from GM26357, GM35433, and the W.M. Keck Foundation (D.J.D.), GM56960 and AR40997 (D.L.S.), and GM53759, the March of Dimes, and a Pew Scholarship in the Biomedical Sciences (S.A.D.).

References

- Bottcher, B., Wynne, S. A. & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron microscopy. *Nature*, **385**, 88-91.
- Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B Virus capsid by cryo-electron microscopy. *Nature*, **385**, 91-94.
- Crowther, R. (1967). A linear analysis of the non-crystallographic symmetry. *Acta Crystallog.* **22**, 758-764.
- Hanein, D. & DeRosier, D. J. (1999). A new algorithm to align three-dimensional maps of helical structures. *Ultramicroscopy*, in the press.
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Backmann, E. & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899-929.
- Jap, B. K., Walian, P. J. & Gehring, K. (1991). Structural architecture of an outer membrane channel as determined by electron crystallography. *Nature*, **350**, 167-170.
- Kimura, Y., Vassilyev, D. G., Miyazawa, A., Kidera, A., Matsusha, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature*, **389**, 206-211.
- Klug, A., Crick, F. H. & Wyckoff, H. W. (1958). Diffraction by helical structures. *Acta Crystallog.* **11**, 199-213.
- Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature*, **367**, 614-621.
- Malhotra, A., Penczek, P., Agrawal, R. K., Gabashvili, I. S., Grassucci, R. A., Jünemann, R., Burkhardt, N., Nierhaus, K. H. & Frank, J. (1998). *Escherichia coli* 70S ribosome at 15 Å resolution by cryo-electron microscopy: localization of fMet-tRNA-fMet and fitting of L1 protein. *J. Mol. Biol.* **280**, 103-115.
- Mimori, Y., Yamashita, I., Murata, K., Fujiyoshi, Y., Yonekura, K., Toyoshima, C. & Namba, K. (1995). The structure of the R-type straight flagellar filament of *Salmonella* at 9 Å resolution by electron cryomicroscopy. *J. Mol. Biol.* **249**, 69-87.
- Morgan, D. G., Owen, C., Melanson, L. A. & DeRosier, D. J. (1995). Structure of bacterial flagellar filaments at 11 Å resolution: packing of the alpha-helices. *J. Mol. Biol.* **249**, 88-110.
- Nogales, E., Wolf, S. G. & Downing, K. H. (1998). Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature*, **391**, 199-203.
- Polyakov, A., Severinova, E. & Darst, S. A. (1995). Three-dimensional structure of *Escherichia coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell*, **83**, 365-373.
- Toyoshima, C. & Unwin, N. (1990). Three-dimensional structure of the acetylcholine receptor by cryoelectron microscopy and helical image reconstruction. *J. Cell. Biol.* **111**, 2623-2635.
- Toyoshima, C., Sasabe, H. & Stokes, D. L. (1993). Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature* **362**, 467-471.
- Unwin, N. (1993). Icotinic acetylcholine receptor at 9 Å resolution. *J. Mol. Biol.* **229**, 1101-1124.
- Unwin, N. (1995). Acetylcholine receptor channel imaged in the open state. *Nature*, **373**, 37-43.
- Wang, B. (1985). Resolution of the phase ambiguity in macromolecular crystallography. In *Methods in Enzymology* (Wyckoff, H., Hirs, C. & Timasheff, S., eds), vol. 115B, pp. 90-112, Academic Press, London.
- Yonekura, K., Stokes, D. L., Sasabe, H. & Toyoshima, C. (1997). The ATP-binding site of Ca^{2+} -ATPase revealed by electron image analysis. *Biophys J.* **72**, 997-1005.
- Zhang, P., Toyoshima, C., Yonekura, K., Green, N. M. & Stokes, D. L. (1998). Structure of the calcium pump from sarcoplasmic reticulum at 8 Å resolution. *Nature*, **392**, 835-840.

Edited by W. Baumeister

(Received 9 September 1998; received in revised form 26 February 1999; accepted 1 March 1999)