Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 13 November 2003
Accepted 29 April 2004

A graphical representation of image quality for three-dimensional structure analysis of two-dimensional crystals

Anchi Cheng[a]* and Mark Yeager[a,b]

[a]Department of Cell Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA, and [b]Division of Cardiovascular Diseases, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA. Correspondence e-mail: acheng@scripps.edu

Electron crystallography of two-dimensional (2D) crystals is a powerful approach for the analysis of membrane protein structure. Three-dimensional (3D) structures are derived by merging data from 2D crystals at varying tilt axes and tilt angles. A graphical representation was developed that incorporates the tilt geometry to display the quality of each image. Information that can be extracted from the plot includes a symbol for each image that reflects the completeness of the data for that crystal. The polar plot also includes a novel parameter that reflects the minimal variation in tilt axis to ensure overlap of data points in reciprocal space between images at the same tilt angle. The tilt geometry and image quality plot is especially useful for tracking the progress of data collection and for assessing the completeness of two data sets prior to determination of difference maps.

1. Introduction

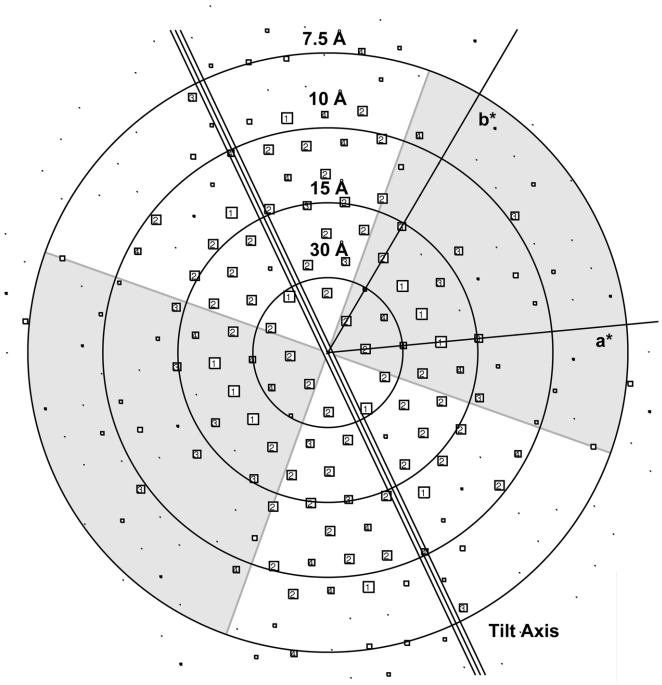
In the determination of three-dimensional (3D) structure by analysis of two-dimensional (2D) crystals, the Fourier transform of each transmission-electron-microscopic (EM) image provides reciprocal-space information for a single plane, called the central zone, that passes through the origin (Henderson & Unwin, 1975; Amos *et al.*, 1982; Henderson *et al.*, 1986). The amplitude and phase data from images with varying tilt geometries are merged to produce a 3D density map. For a given tilt angle of the EM stage, the tilt axis and the sidedness of the crystals are random. To sample as much of reciprocal space as possible, it is important to follow the course of data accumulation to guide further data collection. Therefore, we developed a graphical representation as a tool for tracking the progress of data collection.

2. Description

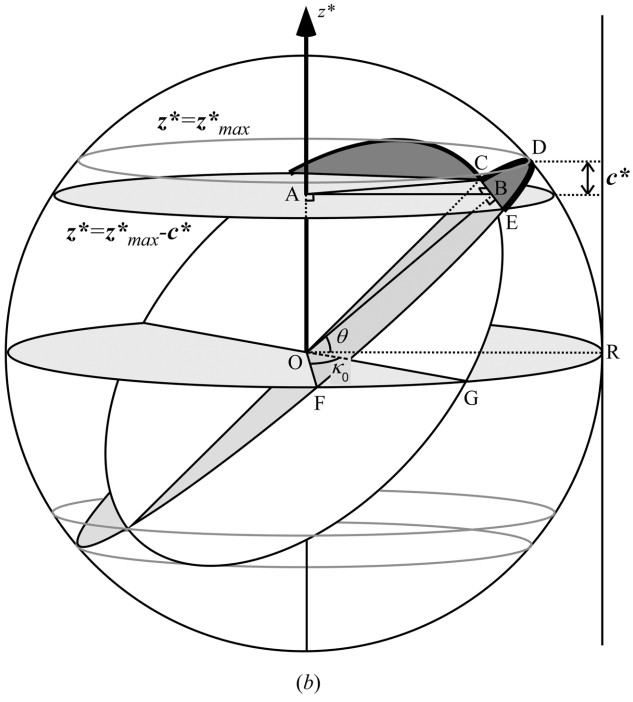
The tilt geometry follows the same symmetry operations as dictated by the two-sided plane group of the 2D crystal. We express the unique tilt geometry of a particular image by two parameters, a tilt angle and an angle from the tilt axis to the closest equivalent axis of a^*. These are not necessarily identical to the tilt angle and tilt axis relative to a^* that are assigned to the crystal during analysis of the particular image because there are, in general, several symmetrically equivalent assignments of the latter. In the process of merging images using the program *ORIGTILT* (Crowther *et al.*, 1996), the conversion of the reflection indices and phases to the asymmetric reciprocal lattice of a tilted crystal is dictated by the crystal symmetry, the tilt geometry of the image relative to the electron beam, the orientation of the crystal in the image and the defocus of the beam. To generate the tilt geometry plot, we use the same conversion algorithm since the central zone, expressed by its normal vector, has the same symmetry as reflections in reciprocal space.

The quality of an image is traditionally expressed by the highest resolution shell in which significant data are available, often determined by calculating the completeness as commonly performed in X-ray crystallography. However, the quality of a tilted image often deteriorates in the direction perpendicular to the tilt axis because of a lack of crystal flatness, specimen drift and charging that are amplified when the stage is tilted with respect to the electron beam (Fig. 1*a*). To emphasize this shortcoming, we include in our completeness test only reflections that lie within an arbitrary $\pm 45^\circ$ wedge in the zones perpendicular to the tilt axis as shown in Fig. 1(*a*). To perform the completeness test for images used in the 3D reconstruction, four uniformly divided resolution shells bounded by ∞, $4r$, $2r$, $(4/3)r$ and r are assigned, where r is the resolution limit. The quality of the image (IMQ) is then expressed by a rank of 1 to 5, where 1 is arbitrarily chosen to correspond to $\geq$50% completeness of the data in the highest resolution shell, $(4/3)r$ to r. An IMQ score of 5 means that the completeness criterion is not met in any of the four resolution shells. In the calculation, we include reflections with intensity quality (IQ) values ≤ 7, as defined in the program *MMBOXA* (Crowther *et al.*, 1996). [IQ 1 and 2 spots have peak-to-r.m.s. (root-mean-square) ratios of 7:1 and 3.5:1, respectively. IQ 7 spots have a ratio of 1. IQ 8 spots have a signal that is weaker than the background level (Henderson *et al.*, 1986).] Overall, the inclusion of reflections in the $\pm 45^\circ$ wedge for the completeness test and the ranking of completeness provide an unbiased assessment of the usefulness of a particular image to the entire data set. In the graphical representation shown in Fig. 2, the angular distribution of the images is visually summarized by displaying the IMQ for each particular image according to its tilt geometry in a polar coordinate system that spans the asymmetric part of space that covers all unique tilt geometries.

For a given reflection, varying the tilt angle causes the structure-factor data to trace the continuous molecular transform, called a lattice line, perpendicular to the reciprocal lattice of the 2D repeat. In



(*a*)



(*b*)

Figure 1
Geometric relationship that determines parameters plotted in the tilt geometry and image quality plot. (*a*) Definition of the quadrant used for the completeness test. In this example from an image of a connexin 43 crystal tilted at 28°, the quality of the reflections is displayed by their IQ values in reciprocal space (IQ values range from 1 to 7 and correspond to S/N of $\geq$7 to 1, respectively). The largest boxes with the lowest IQ values have the highest signal-to-noise ratio. Four resolution bins are tested in the shaded sectors that range from 45 to 135° with respect to the tilt axis and have a greater impact on the *z* resolution. The deterioration of completeness in the tilted quadrant is clearly visible in this image and gives an IMQ value of 3. (*b*) Untilted and tilted central zones that define the minimal overlap of tilt axis, κ_0, required by the data set. The darkly shaded arc in the two tilted central zones contains information in the z^*_{max} bin. The trigonometric relationship in right triangles *ABC* and *OBC*, where *B* is the midpoint of line *CE*, defines the relationship between κ_0, z^*_{max}, c^* and θ. See Appendices *A* and *B* for details.

determining the 3D structure, the data points in each lattice line are fitted with a continuous curve, which interpolates the transform. The fitted lattice line is then sampled by a pseudo-lattice along $\mathbf{z}^*$ to facilitate the inverse Fourier transform to generate the 3D map. To estimate the minimum tilt-axis overlap that will allow a reliable fit of the lattice line data, we consider a data set for which the resolution limit is *r*. In reciprocal space, this resolution shell forms a sphere with radius $r^* = 1/r$ as shown in Fig. 1(*b*). Plane *OFG* corresponds to the untilted central zone given by $z^* = 0$ and contains the a^* and b^* axes. A central zone with tilt angle θ can be displayed as plane *OCD* that intersects the sphere in a circle. Point *D* identifies z^*_{max}, the highest resolution in the z^* direction, which is found in this central zone within the limit of r^*. By drawing two parallel planes $z^* = z^*_{max}$ and $z^* = z^*_{max} - c^*$, where c^* is the sampling interval of the lattice line fitting, we define the area in the central zone where the structure-factor information will contribute to the highest z^* bin at the given tilt angle. This area is shown in Fig. 1(*b*) as the darker shaded area bounded by arc *CDE* and line *CE*. A second image with a small tilt-axis rotation contains data overlapped with this bin. As the central zone of the second image rotates around the z^* axis, there will be a point where the z^*_{max} bin of the second image no longer overlaps the z^*_{max} bin of the first. The 'resolved' tilt-axis angle difference, κ_0, can therefore be calculated from the geometric relationship of the two central zones by

$$\cos(\kappa_0/2) = (r^* \sin\theta - c^*)/\{\tan\theta[r^{*2} - (r^* \sin\theta - c^*)^2]^{1/2}\}. \quad (1)$$

κ_0 is the minimal threshold for overlap in tilt axis between images with the same tilt angle (see Appendix *A*). A well determined lattice line fit requires that there is sufficient overlap of the data in reciprocal space. Therefore, a more useful parameter is κ_{50}, which corresponds to the tilt-axis separation angle at which the two z^*_{max} bins are 50% overlapped (see Appendix *B*). In our plot, the κ_{50} for tilt angles of 10, 20, 30 and 40° are displayed as shaded arcs in polar coordinates, bounded by the calculated θ values and the lower tilt angle that limits

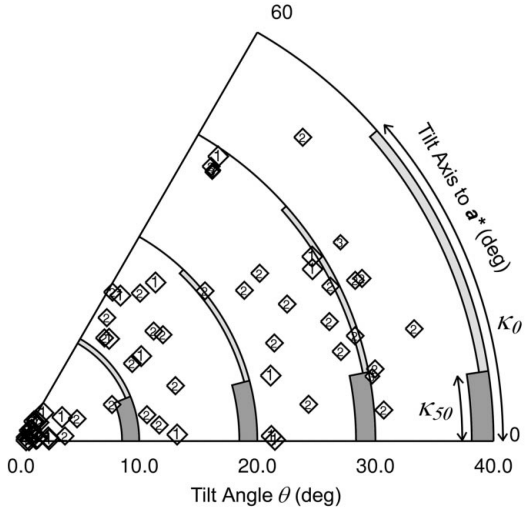


Figure 2
Example of tilt geometry and image quality plot. The 2D crystals of connexin 43 gap junction channels belong to the two-sided plane group *p6*. The resolution limit is 7.5 Å and the lattice lines are sampled at 1/300 Å^{-1} to include the known thickness of the gap junction channel. Individual image quality is displayed as IMQ values that range from 1 to 5. The size of the symbols is proportional to the significance of the data in the 3D reconstruction. The thick dark grey arcs indicate the κ_{50} values calculated at the higher tilt-angle boundary of the box and are bounded at the lower tilt angle that limits z^* to be $z^*_{max} - c^*$. The light grey narrower arcs show the extent of κ_0.

z^* to be $z^*_{max} - c^*$ (Fig. 2). For comparison, the extent of κ_0 is displayed by the narrower shaded arc.

The value of κ_{50} has the following properties. It is a property of the crystal rather than of the images used in the data merging. It decreases as the tilt angle increases and decreases as r^* increases. In addition, larger values of c^* result in higher values of κ_{50}. At very high tilt angles, such as those larger than 70°, the overlap of information in the z^*_{max} bins is so unlikely that the κ_{50} values are less than 3° in our connexin 43 test data set. Nevertheless, since only a few lattice lines pass the z^*_{max} bins at such high tilt angles, the coarsely separated tilt geometries only have a small effect on the overall structure.

The program that generates the tilt geometry and image quality plot was derived by modification of *CTFAPPLYK* and *ORIGTILTK* in the MRC package (Crowther *et al.*, 1996). The format of the input was based on the same template used for *ORIGTILTK*, the standard MRC program for image merging, and could therefore be easily applied to any merged data set. It is called *PLTILTK* and is available as part of the MRC 2D crystal processing package or by direct request to us. The κ_{50} was determined numerically by testing the inclusion of uniformly sampled points in the two z^*_{max} bin projections onto the plane $z^* = z^*_{max} - c^*$ for a series of tilt-axis separations.

3. Results and discussion

Our tilt geometry and image quality plot is similar to the data plot of the Euler angles in the 3D reconstruction of icosahedral viruses (Crowther, 1971) and for single-particle reconstruction (Frank *et al.*, 1996; van Heel *et al.*, 1996; Ludtke *et al.*, 1999) but is tailored to the unique data-collection strategy used in 2D crystal analysis. An added feature of our plot is that the symbols for each data point signify the quality (*i.e.* completeness) for that particular image. As an example, Fig. 2 shows the data distribution plot for a set of images collected from 66 crystals of connexin 43 gap junction channels. The 2D crystals belong to the two-sided plane group *p*6 and therefore range from 0 to 60° in tilt axis. Most images have significant information in the tilt quadrant as defined in Fig. 1 and are of high quality since all but four are categorized as IMQ ≤ 2. The distribution of tilt geometries is also adequate because the greatest separation between zones is not much larger than κ_{50}. On the other hand, the plot mercilessly displays several problems in the data that would otherwise not be obvious. First, it shows that the maximal tilt angle is only 35°. Furthermore, the tilt axes of the two images at this tilt angle only have a small overlap in the highest z^* resolution bin. Therefore, there is a limited gain in information at the maximal tilt angle. Second, it shows that there is a zone with almost no data between tilt angles of 20 to 30° and tilt axes between 35 to 55° from the a^* axis. Such a zone of emptiness in the data distribution reduces the redundancy for fitting certain lattice lines. As a result, improvement of the 3D construction would benefit by inclusion of images with tilt geometries that fill this hole.

The parameter κ_{50} should be considered with care. All high-tilt images contain data that fill gaps at lower z^*. Therefore, comparison of the experimental image data to κ_{50} should only be made on those groups of images with tilt angles around the maximal tilts. Also, users should be aware that tilt-axis equivalence means that a central zone with a tilt axis near 60° in a *p*6 data set is in fact quite close to another central zone whose tilt axis lies near 0°. As a result, the separation between any two tilt axes is never larger than 30° in the case of crystals with *p*6 symmetry.

For a 3D reconstruction, several parameters are used to assess the quality of the data set: (i) the number of images, (ii) the maximal tilt angle at which images are included, (iii) the azimuthal projection of the 3D data set (nicely displayed by Wolf *et al.*, 1996), (iv) phase residuals of lattice line fits for overall or high-resolution reflections, (v) the completeness of the structure-factor data used in the reconstruction, and (vi) the resolution defined by the full width at half maximum of the point-spread function in the direction within the plane and perpendicular to the 2D crystal (Unger & Schertler, 1995). We are also aware of a report in which (vii) tilt geometry was plotted using a polar coordinate system (Hebert *et al.*, 2001). The quality of the individual images was, however, not displayed in this case. Our plot summarizes the information contained in parameters (i), (ii), (iii) and (vii) and is more informative than the single completeness value obtained in (v). It does not replace parameters such as the point-spread function, which displays the global effect of the data quality on the density map. We also note that, in those cases where amplitudes are recorded using electron diffraction, the distribution of tilt geometry for the recorded patterns can also be analyzed by a similar plot but the ranking of diffraction quality will need to be defined.

4. Conclusions

We have developed a graphical representation to track the progress of data accumulation for 3D structure determination. It is a general tool for electron crystallography of 2D crystals in which images with varying tilt geometries are merged to produce a 3D density map. We have found this plot to be useful for quality control and for comparison of data sets prior to the determination of difference maps.

APPENDIX *A*
Derivation of κ_0 value

In Fig. 1(*b*), points *C*, *D*, *E*, *F*, *G* and *R* are on the sphere that limits the resolution to r^* while point *O* represents the origin of the sphere and the reciprocal-unit-cell axes. Therefore, maximal resolution is

$$r^* = \overline{OC} = \overline{OD} = \overline{OE} = \overline{OF} = \overline{OG} = \overline{OR}. \quad (2)$$

The untilted central zone contains points *O*, *F*, *G*, *R*. The tilted central zone, at a tilt angle of θ, contains points *B*, *C*, *D*, *E* and *F*. Point *D* identifies the location where z^*_{max} is found in this central zone within the limit of r^*. The central zone intersects plane $z^* = z^*_{max} - c^*$, which identifies the lower bound of the z^*_{max} bin, at the line $\overline{CE}$. Point *B* bisects line $\overline{CE}$ and is therefore collinear with *O* and *D*. *R* is chosen on the untilted central zone and on the resolution sphere so that plane *DOR* is perpendicular to the tilt axis of the tilted central zone, $\overline{OF}$. In addition, plane $z^* = z^*_{max} - c^*$ intersects the z^* axis at point *A*. The plane given by $z^* = z^*_{max} - c^*$ and the untilted central zone are parallel to each other. In addition, point *B* bisects $\overline{CE}$, and the ends of this line segment terminate on circles on the planes given by $z^* = z^*_{max} - c^*$ and the central zone. Therefore, it follows that

$$\angle OAB = \angle OBC = \angle ABC = 90^\circ \quad (3)$$

and

$$\theta = \angle DOR = \angle BOR = \angle ABO. \quad (4)$$

As shown in Fig. 1(*b*), $\kappa_0 = \angle FOG$ corresponds to the tilt-axis separation of two central zones when the highest z resolution zones, shown as the darkly shaded area, intersect at one point, *C*. Since, again, $z^* = z^*_{max} - c^*$ and the untilted central zone are parallel to each other,

$$\kappa_0 = \angle FOG = \angle CAE = 2\angle BAC. \quad (5)$$

$z^*_{\max}$ can be expressed using r^* and θ as

$$z^*_{\max} = \overline{OD}\sin\theta = r^*\sin\theta \quad (6)$$

and, for the right-angle triangle ΔAOB,

$$\overline{OA} = z^*_{\max} - c^* = \overline{AB}\tan\theta. \quad (7)$$

Therefore,

$$\overline{AB} = (r^*\sin\theta - c^*)/\tan\theta \quad (8)$$

and, for the right-angle triangle ΔOAC,

$$\overline{AC} = (\overline{OC}^2 - \overline{OA}^2)^{1/2} = [r^{*2} - (r^*\sin\theta - c^*)^2]^{1/2} \quad (9)$$

and, finally, using equation (5) and the right-angle property of triangle ΔABC, we have

$$\cos(\kappa_0/2) = \overline{AB}/\overline{AC} = (r^*\sin\theta - c^*)/\{\tan\theta[r^{*2} - (r^*\sin\theta - c^*)^2]^{1/2}\}. \quad (1)$$

APPENDIX *B*
Derivation of κ_{50} value

The estimate of κ_{50} is done numerically by testing the inclusion of uniformly sampled points in the two $z^*_{\max}$ bin projections onto the plane $z^* = z^*_{\max} - c^*$ for a series of tilt-axis separations. This is valid because, regardless of the actual z^* value, all data with the same x^*y^* coordinates contribute to the same $z^*_{\max}$ bin in these regions. For the tilted central zone discussed in Appendix *A*, the projection of circle *FEDC* on plane *ACE* with $\overline{AB}$ as its x_1 axis is an ellipse

$$(x_1/\cos\theta)^2 + y_1^2 = r^{*2} \quad (10)$$

and the line $\overline{CE}$ is given by

$$x_1 = (r^*\sin\theta - c^*)/\tan\theta. \quad (11)$$

Therefore, points inside the projection area of the central zone with the highest z resolution must satisfy

$$\begin{cases} (x_1/\cos\theta)^2 + y_1^2 < r^{*2} \\ x_1 > (r^*\sin\theta - c^*)/\tan\theta. \end{cases} \quad (12)$$

A similar equation can be written for a second central zone of tilt angle θ in its own coordinate system (x_2, y_2) as

$$\begin{cases} (x_2/\cos\theta)^2 + y_2^2 < r^{*2} \\ x_2 > (r^*\sin\theta - c^*)/\tan\theta. \end{cases} \quad (13)$$

Since the second central zone, and hence its coordinate system, can be produced by a rotation of the first by the tilt-axis separation, ω, the coordinates of a given point are related in the two coordinate systems by

$$\begin{aligned} x_2 &= x_1\cos\omega + y_1\sin\omega \\ y_2 &= -x_1\sin\omega + y_1\cos\omega. \end{aligned} \quad (14)$$

For each sampled ω, the program therefore transforms each sampled (x_1, y_1) that satisfies equation (12) to $t(x_1, y_1)$ using equations (13) and tests whether the point also satisfies equation (14). Interpolation of the percentage area overlap *versus* ω yields an estimate of κ_{50}.

We thank Richard Henderson for providing information on the MRC programs and permission for including the program in their package. We also thank Barbie Ganser-Pornillos for helpful discussions and Mark J. Daniels and Vinzenz Unger for comments on the manuscript. This work was supported by NIH NHLBI RO1 HL48908 (MY). MY is the recipient of a Clinical Scientist Award in Translational Research from the Burroughs Wellcome Fund. AC was supported by a postdoctoral fellowship from the American Heart Association during part of this work.

References

Amos, L. A., Henderson, R. & Unwin, P. N. T. (1982). *Prog. Biophys. Mol. Biol.* **39**, 183–231.
Crowther, R. A. (1971). *Philos. Trans. R. Soc. London Ser. B*, **261**, 221–230.
Crowther, R. A., Henderson, R. & Smith, J. M. (1996). *J. Struct. Biol.* **116**, 9–16.
Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). *J. Struct. Biol.* **116**, 190–199.
Hebert, H., Purhonen, P., Vorum, H., Thomsen, K. & Maunsbach, A. B. (2001). *J. Mol. Biol.* **314**, 479–494.
Heel, M. van, Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996). *J. Struct. Biol.* **116**, 17–24.
Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). *Ultramicroscopy*, **19**, 147–178.
Henderson, R. & Unwin, P. N. T. (1975). *Nature (London)*, **257**, 28–32.
Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). *J. Struct. Biol.* **128**, 82–97.
Unger, V. M. & Schertler, G. F. X. (1995). *Biophys. J.* **68**, 1776–1786.
Wolf, S. G., Nogales, E., Kikkawa, M., Gratzinger, D., Hirokawa, N. & Downing, K. H. (1996). *J. Mol. Biol.* **262**, 485–501.