GUI programs for processing individual images in early stages of helical image reconstruction—for high-resolution structure analysis

Koji Yonekura, a,b,* Chikashi Toyoshima, c Saori Maki-Yonekura, b and Keiichi Namba a,b

a Graduate School of Frontier Biosciences, Osaka University, 3-4 Hikaridai, Seika, Kyoto 619-0237, Japan
b Dynamic NanoMachine Project, ICORP, JST, 3-4 Hikaridai, Seika, Kyoto 619-0237, Japan
c Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo, 113-0032, Japan

Received 30 June 2003, and in revised form 30 July 2003

Abstract

A set of programs equipped with graphical user interface has been developed for processing individual images in early stages of the three-dimensional helical image reconstruction procedure. These programs can be used for initial screening of suitable image area, straightening the object image, determination of box parameters including the repeat distance, determination of the out-of-plane tilt and initial editing of the layer-line data. These tasks are difficult to automate and therefore very time-consuming. The programs, developed by adopting the concept of the layer-line indexing [Ultramicroscopy 84 (2000) 1–14], are effective for processing many images of filamentous molecular assemblies and especially tubular crystals having various helical classes. Using these programs, higher-resolution signals can be extracted more reliably and quickly, and the time required for processing each image can be reduced to 1/2–1/10. Here also presented is an overview on helical image reconstruction for high-resolution structure analysis. © 2003 Elsevier Inc. All rights reserved.

1. Introduction

Biological macromolecules that are assembled with helical symmetries present great advantages for three-dimensional structure analysis by electron microscopy (EM), because every single image contains many different views of constituent molecules, and hence this method does not require images recorded at different tilt angles. The first three-dimensional structure obtained by EM was by helical image reconstruction (DeRosier and Klug, 1968). Macromolecular complex structures with helical symmetries have high potentials to be analyzed at high resolution, but, until now, there were only a small number of examples where the structure was solved at around or better than 10 Å resolution. Examples include tubular crystals of three membrane proteins: nicotinic acetylcholine receptor at 4.6 Å (Miyazawa et al., 1999), Ca$^{2+}$-ATPase at 6–8 Å (Xu et al., 2002; Zhang et al., 1998), and Na$^+$,K$^+$-ATPase at 11 Å (Rice et al., 2001); and two filamentous molecular assemblies: the bacterial flagellar filament (Mimori et al., 1995; Morgan et al., 1995) and tobacco mosaic virus (Jeng et al., 1989), both at ~10 Å resolution. Very recently, the structure of nicotinic acetylcholine receptor was analyzed at 4.0 Å resolution, but it was accomplished by extraordinary intensive effort on data collection and analysis over the several years to include ~110 000 molecular images from ~360 images of the tubular crystals (Miyazawa et al., 2003).

In the structure analysis by electron cryomicroscopy (cryoEM) of frozen hydrated samples embedded in vitreous ice, the attainable resolution more or less depends on the number of molecular images averaged. The number of molecules contained in individual images of a tube or a filament is much larger than that of other highly symmetric molecular complexes, such as icosahedral spherical viruses. For example, there are more than 1000 molecules in a typical tubular crystals of Ca$^{2+}$-ATPase with a diameter of ~700 Å and a length of a few thousand Å; there are a few ~500 molecules in the bacterial flagellar filament, having a diameter of 240 Å and a length of a few thousand Å. However, in most
cases, the attained resolution was comparable to or even lower than those obtained for spherical viruses. This has been thought that the numbers of object images included in the analyses are much smaller than those used for spherical virus structures. The major factor that limits the number is that collection of many images of tubular crystals showing good diffraction patterns is much more difficult than collecting good images of viruses, mainly because the larger the object particles, the more chance they suffer from deformation by the quick-freeze procedure, or because of inherent poorness in the helical symmetry. Another factor is that, compared to single particle image analysis, the procedure of processing each image is much more complicated. Therefore, automation is much more difficult, and hence the image processing has been very time consuming. To reduce the time required for such tedious tasks, we have developed a set of programs equipped with graphical user interfaces (GUI programs). These programs can be used for initial screening of suitable image area, straightening the image, determination of box parameters, determination of the out-of-plane tilt (ω), and initial editing of layer-line data. They allow extraction of the highest possible signals from each image. By using these programs, we recently obtained a density map of the bacterial flagellar filament at ~4 Å resolution from a relatively small number of molecular images (~41 000), based on which we were able to build its complete atomic model (Yonekura et al., 2003). Here, tubular crystals of Ca\(^{2+}\)-ATPase and the R-type straight flagellar filament are used as examples to demonstrate how efficiently and easily the image processing can be done with these programs. We also describe some important points for high-resolution structure analysis by helical image reconstruction.

2. Methods

CryoEM analysis of frozen-hydrated specimens of Ca\(^{2+}\)-ATPase tubes and the straight flagellar filaments were carried out as described by Yonekura et al. (1997) and Mimori et al. (1995), respectively. Fig. 1 shows the protocol for processing individual images at the initial stage of helical image reconstruction. The programs developed in the present work (names written in bold-face capital letters) can be used for determination of box parameters, straightening the image, determination of the out-of-plane tilt (ω), and initial editing of layer-line data. They are closely linked to other programs for helical image reconstruction (e.g., HLXBOX, IMGCCF, SRCHAID, SRCH, HLXFL, and so on), full description of which was given by Toyoshima (2000). Some of the programs developed previously were modified for the new GUI programs. This program set works under the X-Window system on SGI, HP AlphaStation, and LINUX workstations and are available on request. Brief directions are given below.

2.1. XDISPHBOX

This program displays a digitised image and can be used to select an area used for HLXBOX (Toyoshima, 2000) (Fig. 2). Then, the program makes a control file for HLXBOX to obtain the initial Fourier transform. In the second mode, it can be used to select a reference and a test area and makes two control files for IMGCCF to do the cross-correlation (Toyoshima, 2000). The first one is for estimation of the bend of the helix axis, and the second one is for determination of the repeat distance and the precise in-plane inclination of the helix axis with respect to the sampling raster. The program also makes a control file for MRCPRJ (Yonekura et al., 2001) to do the running average along an approximate helix axis (see Results).

2.2. XDISPACF

This program displays a cross-correlation map generated by IMGCCF with a list of high cross-correlation peaks (Fig. 3). By supplying a peak number selected visually from the displayed list, the program makes a control file for HLXBOX with the repeat distance and the in-plane inclination of the helix axis determined from the specified correlation peak.

2.3. XDISPACFHB

This program works similarly to XDISPACF, but displays the original image with a list of high cross-correlation peaks. The program also displays areas used for box and cross-correlation as XDISPHBOX does. The displayed box is refreshed according to a specified peak in this trial of HLXBOX.

2.4. XDISPHCV

This program displays an image with its high cross-correlation peaks determined by IMGCCF (Fig. 4). By visually selecting and editing the peaks, the program calculates a spline curve along the filament image and then straightens it by quadratic interpolation.

2.5. XDISPHFFT

This program displays the Fourier transform of an selected area with all possible layer lines calculated from the helical parameters and repeat distance superimposed in the right half area and sum of amplitudes over entire horizontal pixel array on the left (Fig. 5). It also indicates the Bessel order (n) and the layer-line number (l) at the top left corner of the image window, when the cursor is
on a predicted theoretical layer line. The program can be used to collect the layer-line amplitude and phase data, and makes a control file for HLXS (Toyoshima, 2000) or HLXSX described below. This program is an extension of XDISPH, which simply displays the Fourier transform of a tube or a filament (Toyoshima, unpublished).

2.6. HLXSX

This program displays amplitude and phase profiles along layer lines specified by XDISPHFFT and can be used to pick amplitude peaks (Fig. 6A). Then, it makes a control file for SRCHAID and runs SRCHAID and
Fig. 2. GUI of **XDISPHBOX** displaying a digitized image of a tubular crystal of Ca\(^{2+}\)-ATPase. White box encloses area used for the following procedures. Red boxes enclose the reference area (the small one on the left) and the test area (the large one on the right) for cross-correlation.

Fig. 3. GUI of **XDISPACF** displaying high cross-correlation peaks (above 80% of maximum). Upper panel shows a cross-correlation map, where peaks are marked and numbered in order of the cross-correlation coefficient. The peak with the 4th highest value was used for the following analysis because its Fourier transform showed sharp and strong layer lines as shown in Fig. 5. If either of the 1st, 2nd, and 3rd peaks was used, layer lines even within low-resolution became overlapped. The peaks numbered in red indicate those already tried for determining the box length to calculate the Fourier transform. The table in the lower window shows the list of normalized correlation coefficients, together with the corresponding repeat distance and the in-plane inclination of the helix axis.

Fig. 4. GUI of **XDISPHCV** displaying an image of the R-type straight flagellar filament. (A) Red circles indicate cross-correlation peaks determined from a raw image. (B) Red and blue circles indicate cross-correlation peaks determined from the image obtained by running average along the approximate helix axis. The red line represents a calculated spline curve using the red points. The blue points were manually excluded and the two points at the both ends were manually added for the calculation. Both in (A) and (B), top 35 points with high cross-correlation coefficients are displayed where cross-correlation peaks are above 73.7 and 86% of the maximum in (A) and (B), respectively, except for the peak correlated with itself. Note that there are many false peaks in (A) because of a relatively high noise level and a poor contrast. Running average improved the accuracy of the determination of the helix axis line in (B). (C) Straightened image by the spline curve shown in (B). Axially compressed images displayed in the right panels clearly show the traces of the filament axis.
Fig. 5. Display of the Fourier transform and layer-line amplitude profile by **XDISPHFFT**. (A) Best possible one determined here. (B) One obtained by 50 g.u. shift of the box position along the helix axis and 1 g.u. extension of the repeat distance from that in (A). Red lines on the left edge of the window show the sum of amplitudes over the horizontal lines above a cut off value. Red lines on the right half show all possible theoretical layer lines within a given resolution. The horizontal line in cyan indicates the mouse position, which in this example is on the layer line of \( n = 16, l = 287 \) (as displayed at top left) at an axial spacing of \( \approx 18.6 \text{A} \). The peak amplitude of this layer line is displayed at the top left corner. Note that the peak is \( \approx 10\% \) higher in (A) than in (B) (148.37 vs. 134.87). Layer lines in (B) are not as sharp as those in (A). The lower windows show amplitude profiles of a layer line indicated by a white arrow in number display within a box of \( 15 \times 7 \) pixels. The peak amplitude of this layer line on an axial position of 222 g.u. is also \( \approx 31\% \) higher (210 vs. 160) and the amplitude profile perpendicular to the layer line is significantly sharper in (A) compared with (B). In (B), the layer-line profile spreads over a pixel above (223 g.u,) indicating that the repeat distance should be shortened.

Fig. 6. Amplitude profile and phase distribution of a layer line displayed by **HLXSX**. This particular example is a layer line \((h = 1, k = 3; n = -5)\) at an axial spacing of \( \approx 34.5 \text{A} \). Blue and multi-color solid lines in the lower panels of each window are the layer-line amplitude profiles in the left and right sides of the Fourier space, respectively. Colors represent the phase residual between two corresponding data points across the meridian (cyan \( < 30^\circ \) to green \( < 45^\circ \) to magenta \( < 70^\circ \) to red). Blue and red marks in the upper panels represent the phases in the left and right sides of the Fourier space, respectively. (A) First mode to pick up good peaks to determine the out-of-plane tilt \((\omega)\) of and the shift \((\Delta x)\) normal to the helix axis. Vertical lines indicate the position of peaks that were picked up. After selecting good peaks, ‘Do SRCH’ button is clicked to run SRCHAID and SRCH. Colors of vertical lines indicate the phase residual between the data across the meridian on this trial for minimizing \( Q \)-value. (B) Second mode to edit layer lines. The region between the two vertical lines in orange is to be extracted and used for initial alignment and averaging. Note that by applying an out-of-plane tilt \((\omega)\) of \( 2.56^\circ \) and a shift \((\Delta x)\) of \( -0.01 \text{g.u.} \), determined in the first mode, phase residuals became significantly smaller (e.g., the color of the first peak changed from red to cyan) and the data points near the meridian disappeared due to the correction for the \( \omega \) tilt. Current mouse position is indicated by the vertical line in yellow, and the multiple data values on the mouse position are displayed at the bottom of the window: the radial position in g.u.; the amplitude of the both sides average; the phase of the right side; the amplitude of the right side; and the phase difference between the two sides.
SRCH to determine the out-of-plane tilt (ω) and the shift normal to the helix axis (Δx) (Toyoshima, 2000). After applying these two parameters to correct for the tilt and the shift, the program displays the layer-line profiles again for editing the individual layer lines (Fig. 6B). Then, the program makes a control file for HLXFL to extract the layer lines (Toyoshima, 2000).

2.7. IMAGEH2PS

This program makes a postscript file of a digitised image with rectangular boxes, indicating areas used for box (HLXBOX) and cross-correlation (IMGCCF).

2.8. ACFMAP

This program makes a postscript file of a cross-correlation map generated by IMGCCF.

All the programs were written with C language. All except for IMAGEH2PS and ACFMAP need X library, and HLXSX needs Motif library as well.

3. Results

3.1. Determination of box parameters

The first task of the helical image reconstruction procedure is to determine an image area to be processed. XDISPHBOX is a program to display a digitized image in the X-Window system and can be used for initial image area selection. First, the user clicks on three points to define a box to enclose an area of straight tube or filament, as long as possible along the helix axis (Fig. 2). In this step, the rubber band extended from the selected point to the current mouse position is useful to check whether the area being enclosed is straight or not. Once, the user is satisfied, a rectangular box is defined. The program makes a control file for HLXBOX, storing the box corner positions and the in-plane inclination of the helix axis. HLXBOX is a program to carry out the following tasks: cutting out the specified area; floating the cut-out image, padding zeros in the outside area in the direction normal to the helix axis, but not along the axis, and stretching the image along the helix axis to the edge of the whole image box (Toyoshima, 2000). At the same time, HLXBOX makes the projection profile along the helix axis and store it in a file. According to this projection profile, the center position of the helix axis is adjusted (Toyoshima, 2000) and refined further by using amplitude peaks in the strong layer lines by SRCH, SRCHAID, and HLXSX (see below). If the tube or filament is highly curved and the area that can be used for analysis is too short, XDISPHCV can be used to straighten the image (see below).

The Fourier transform of the selected area is then calculated by FFTRANS and checked by XDISPH or XDISPHFFT. Checking the Fourier transform is easily done by typing ‘f’ in the image window of XDISPHBOX. If it shows sharp and strong layer lines, the next step is to determine the repeat distance and more precise in-plane inclination of the helix axis by cross-correlation within the image. It is important for the boxed area to include an exact multiple number of structural repeat to avoid layer line spreading over multiple pixels. Hence, the determination of the exact repeat distance is important and is usually accomplished by cross-correlation search within the image. XDISPHBOX can be used again to select a reference and a test area for cross-correlation. To do this, the user provides two points, the center and a corner, for both the reference and test areas (red boxes in Fig. 2). Care must be taken in selecting the reference area because the results strongly depend on it. The distance between the center positions of the reference and test areas were normally 5000–6000Å for the Ca2+-ATPase tubes and 1500–5000Å for the flagellar filaments. Then XISPHBOX makes a control file for IMGCCF, which generates a cross-correlation map. XDISPACF displays the map and finds peaks with high cross-correlation values (Fig. 3). Strong peaks at longer repeat distances indicate that the image is appropriate for high-resolution image analysis. By supplying one of the peak numbers in the list, XDISPACF makes a control file for HLXBOX with the box parameters adjusted for the specified peak. Use of a too short repeat distance or a box containing a few repeats should be avoided because these would result in overlaps of layer lines even within low resolution. To avoid those, check the Fourier transform by XDISPHFFT as described below. If an in-plane inclination of the helix axis given by a peak is far from others, such peak must be wrong and should be discarded. After peaks shown in Fig. 3 were tried, the 4th highest peak was selected in this case for the following process. The specified box parameters in the control file can be verified by feeding it into XDISPHBOX (white box in Fig. 2). In the case of the flagellar filament, XDISPACF sometimes showed a lot of cross-correlation peaks, which were randomly distributed. In such situations, it is better to see the original image superimposed with cross-correlation peaks. XDISPACFH is a program for this purpose. False peaks, which are off the helix axis, can be easily excluded by using this program. By providing ‘–f’ option, XDISPHBOX, XDISPACF, XDISPACFH, XDISPACVC (see below), and XDISPHFFT (see below) start immediately using the previous parameters. For the filing purpose, IMAGEH2PS makes a postscript file depicting the image areas used for box and cross-correlation. ACFMAP makes a postscript
file from a log file of IMGCCF, showing correlation peaks.

3.2. Straightening

Long filamentous structures tend to be curved during sample preparation on EM grids or due to the nature of samples. For such samples, straightening with cubic splines is now a well-established technique to improve data quality and increase the size of area available for image analysis (Carragher et al., 1996; Egelman, 1986; Owen et al., 1996). XDISPHEVC carries out the straightening of the images with a visual interface. To do so, the user selects a reference area by XDISPHEBOX at first. Then, XDISPHEBOX makes a control file for IMGCCF to do the cross-correlation to find the helix axis position. IMGCCF calculates a cross-correlation map for the whole image. In this procedure, a rather extreme low-pass filter on a raw image gives a good result. If the cross-correlation map is still too noisy, its running average along the approximate helix axis is effective in reducing the image-to-noise ratio (cf. Figs. 4A and B). XDISPHEBOX also makes a control file for MRCPRJ (Fig. 1; Yonekura et al., 2001) for the running average. Alternatively, a filtered image created from only strong peaks (Fig. 1; Yonekura et al., 2001) for the running average. XDISPHBOX makes a control file to display the Fourier transform of a pair of Bessel orders, \( n_{10} \) and \( n_{01} \), and layer-line numbers, \( l_{10} \) and \( l_{01} \) for the principal (1, 0) and (0, 1) layer lines, which correspond to the Miller indices for two-dimensional crystals. Detailed explanation of this concept is described by Toyoshima and Unwin (1990), and Toyoshima (2000). At this point, only \( l_{10} \) and \( l_{01} \) are required to define the heights of all the layer lines. The values of \( n_{10} \) and \( n_{01} \) can be determined afterward by using SRCH, SRCHAID, and HLXSX (see below). XDISPHEFFT shows all possible layer lines within a given radial region in the Fourier space and indicates whether or not signal peaks are well sampled on the theoretical layer line (Fig. 5). The program also shows the sum of amplitudes over individual layer lines (windows in Fig. 5) and layer-line amplitude profiles within a given box display (lower windows of Fig. 5). These functions are used to check if the axial repeat distance of the structure should be increased or reduced by a pixel or even a fraction of a pixel, so that the position of the theoretical layer lines can be adjusted accurately to high resolution. These functions were very effective in extracting data from a layer line near the meridian at an axial spacing of \( \sim 15.5 \) Å for the tubular crystals of \( Ca^{2+} \)-ATPase, and from a layer line having a Bessel order of 2 at an axial spacing of \( \sim 12.5 \) Å resolution for the R-type straight flagellar filaments. One can decide whether or not the image should be further processed after various repeat distance are tried with the reasonable inclination of the helix axis found by XDISPACF or XDISPACFHB. Then, a translation search to optimize the box position should be conducted by shifting the box position by 25–50 g.u along the helix axis and checking its Fourier transform again. Once the box parameters are determined, XDISPHEFFT can be used to pick layer lines and make a control file to be used by HLXS or HLXSX, a program described below, to display the layer-line data profiles. This control file can be fed into XDISPHEFFT as a template for other data sets, so that the position of layer lines can be easily adjusted by just changing the layer-line numbers for \( l_{10} \) and \( l_{01} \).

3.3. Refinement of box parameters using the fourier transform

To check whether or not the selected box determined as described above is suitable for further analysis, XDISPHEFFT can be used to display the Fourier transform. XDISPHEFFT requires a pair of Bessel orders, \( n_{10} \) and \( n_{01} \), and layer-line numbers, \( l_{10} \) and \( l_{01} \) for the principal (1, 0) and (0, 1) layer lines, which correspond to the Miller indices for two-dimensional crystals. Detailed explanation of this concept is described by Toyoshima and Unwin (1990), and Toyoshima (2000). At this point, only \( l_{10} \) and \( l_{01} \) are required to define the heights of all the layer lines. The values of \( n_{10} \) and \( n_{01} \) can be determined afterward by using SRCH, SRCHAID, and HLXSX (see below). XDISPHEFFT shows all possible layer lines within a given radial region in the Fourier space and indicates whether or not signal peaks are well sampled on the theoretical layer line (Fig. 5). The program also shows the sum of amplitudes over individual layer lines (windows in Fig. 5) and layer-line amplitude profiles within a given box display (lower windows of Fig. 5). These functions are used to check if the axial repeat distance of the structure should be increased or reduced by a pixel or even a fraction of a pixel, so that the position of the theoretical layer lines can be adjusted accurately to high resolution. These functions were very effective in extracting data from a layer line near the meridian at an axial spacing of \( \sim 15.5 \) Å for the tubular crystals of \( Ca^{2+} \)-ATPase, and from a layer line having a Bessel order of 2 at an axial spacing of \( \sim 12.5 \) Å resolution for the R-type straight flagellar filaments. One can decide whether or not the image should be further processed after various repeat distance are tried with the reasonable inclination of the helix axis found by XDISPACF or XDISPACFHB. Then, a translation search to optimize the box position should be conducted by shifting the box position by 25–50 g.u along the helix axis and checking its Fourier transform again. Once the box parameters are determined, XDISPHEFFT can be used to pick layer lines and make a control file to be used by HLXS or HLXSX, a program described below, to display the layer-line data profiles. This control file can be fed into XDISPHEFFT as a template for other data sets, so that the position of layer lines can be easily adjusted by just changing the layer-line numbers for \( l_{10} \) and \( l_{01} \).

3.4. Determination of the out-of-plane tilt and refinement of the shift normal to the helix axis

HLXSX is a program to display amplitude profiles and phase distributions of individual layer lines picked by XDISPHEFFT. To determine the out-of-plane tilt (\( \phi \)) of and the shift (\( \Delta x \)) normal to the helix axis, many amplitude peaks have to be picked from the strong layer
lines to calculate the amplitude weighted phase residuals between pairs of corresponding peaks across the meridian ($Q$-value) by using SRCH (DeRosier and Moore, 1970) and SRCHAID (Toyoshima, 2000). Making control files for SRCHAID for many images was a tedious task; one had to check layer-line profiles on print outs made by HLXS, mark strong peaks, edit the control file using a text editor, calculate $Q$-value by SRCH, then, remove those peaks showing relatively large phase residuals, calculate $Q$-value again, and so on. By using HLXSX, however, just a mouse click allows one to select a peak automatically. Ideally, the phases of the corresponding two peaks across the meridian are either equal (even parity) or 180° off (odd parity), depending on the $n$ associated with that particular layer line. This rule, however, never holds in practice because of a finite value of $\omega$ and $\Delta \alpha$. Hence, the criterion for selecting peaks is a similarity of the amplitude profiles between the pair of peaks across the meridian (Fig. 6A). If one clicks a false position that is not a peak, HLXSX automatically searches through the layer-line data for the nearest peak by calculating a vector average over two corresponding regions across the meridian. Then, it makes a control file for SRCHAID and calls SRCHAID and SRCH to calculate $Q$-value. Then, vertical lines appear on selected peaks, indicating phase residuals between the two corresponding points across the meridian on this trial (Fig. 6A). By clicking a mouse, one can remove peaks with relatively large phase residuals indicated by vertical lines in red and magenta, where red represents the phase residual above 70° and magenta below 70° but above 45°. The correct pair of $(n_{10}, n_{01})$ should give the minimal $Q$-value. Many helical classes can be easily tried by pushing the button ‘Change N’ in the HLXSX window (Fig. 6A). If there is a large $\Delta \alpha$ value is obtained, one should apply the value to the horizontal position in the control file of HLXBOX. Then, the Fourier transform and $Q$-value are calculated again until $\Delta \alpha$ becomes small enough (we usually allow the value $<0.02$ g.u., which corresponds to 0.02–0.05 Å). The control file for SRCHAID can be read into HLXSX as a template and easily modified for image processing of other tubes or filaments.

3.5. Editing the start and end positions of layer lines

After the out-of-plane tilt ($\omega$) was determined and the shift normal to the helix axis ($\Delta \alpha$) became sufficiently small, then one can go to the second mode of HLXSX. In this mode, amplitude profiles and phase distributions after the correction for $\omega$ and $\Delta \alpha$ are displayed. The next task is to edit the start and end positions of layer lines for initial alignment and averaging of the processed images to use only regions having small phase residuals between data pairs across the meridian (Fig. 6B). All possible layer lines within a target resolution should not be used at this stage because they are too noisy to achieve reliable alignment. It also needs a large disk space and a long time of calculation. The phase residuals are indicated in multiple colors. One can specify the start and end positions by mouse clicking, where regions to be selected should be from a trough to a trough except for the start position of the layer line with $n = 0$ or those with small $n$ for images with relatively large $\omega$. If one clicks an inappropriate non-trough position, HLXSX automatically searches for the nearest trough. The program makes a control file for HLXFL, which extracts selected regions from layer lines. The extracted layer-line data set will be used for the initial alignment and averaging by HLXFIT and HLXAVG, respectively (Toyoshima and Unwin, 1990).

4. Discussion

Compared to single particle image analysis, there are far more steps of image processing including manual intervention to process individual images for helical image reconstruction, as described by Toyoshima (2000). Many of the steps are for the determination of box parameters, and others are for the determination of the out-of-plane tilt and for initial editing of layer-line regions. These tasks are essential for high-resolution structure analysis, but tedious and time consuming. To reduce such burdens, we developed new GUI programs that work intelligently to reduce manual intervention as much as possible. The new protocol is shown in Fig. 1. The new GUI programs can be used for almost all tasks described in Fig. 5 of the paper by Toyoshima (2000). In addition, a GUI program to straighten distorted images with cubic splines has also been implemented (dotted box in Fig. 1). There are several program packages, which provide GUI tools for helical image reconstruction, e.g., the Brandeis helical package (Owen et al., 1996) and the PHOELIX package (Carragher et al., 1996). But, for analyzing tubular crystals, these packages are not suitable because tubular crystals have relatively large diameters, and almost every tube has a different selection rule of the helical symmetry. Toyoshima (2000) showed that indexing of the two basic layer lines is a promising approach. We have been developing programs adopting this concept (Toyoshima, 2000; Yonekura and Toyoshima, 2000a,b), but the implementation of GUI has been minimal.

4.1. Determination and refinement of box parameters

Our new GUI programs developed here are mainly divided into three parts. The first part is to determine box parameters, which include the repeat distance and the in-plane inclination of the helix axis. XDISHPPBOX, XDISPACF, XDISPACFHB, and XDISHPFFT are programs for this purpose.
High-resolution structure analysis essentially requires many tubes or filaments images to be aligned and averaged. XDISPHBOX provides an easy and efficient way for initial screening of many images. Unless the Fourier transform of an area initially selected by XDISPHBOX shows poor signals, one can go to the next step to determine the repeat distance. Here, the second mode of XDISPHBOX can be used to select a reference and a test area for the cross-correlation analysis to determine the repeat distance and the in-plane inclination of the helix axis. Determination of the repeat distance is the most time-consuming but important task. The PHOELIX package carries out semi-automatic determination of the repeat distance based on the intercepts of layer lines for given selection rules, without cross-correlation (Carragher et al., 1996). However, as mentioned before, it is not suitable for tubular crystals belonging to various helical classes. Even in the image processing of the R-type straight flagellar filaments, which belong to just one helical class, the repeat distances were proved to be varied from 1000 to 4000 Å by use of the cross-correlation search. Many different repeat distances can be easily tried out by XDISPACF or XDISPACFHB. More importantly, to collect highest possible signals from high-resolution layer lines, we have to check carefully whether or not higher-resolution signals are accurately sampled on the theoretically calculated layer-line positions. In such situations, automation of these tasks is very difficult and also dangerous. Hence, easy and useful interactive visualization tools for the Fourier transform are essential. XDISPHFFT displays not only the Fourier transform of tube or filament images but also useful information (Fig. 5). They are especially effective in extracting high-resolution signals by allowing the determination of the best possible box parameters (Fig. 5). After a set of averaged layer-line data is obtained, the box parameters can be further refined by using the averaged data set as a reference (HBOXREFN, Yonekura and Toyoshima, unpublished).

4.2. Straightening

The second part is the straightening of filament images by using cubic splines. For structure analyses, we usually select out only areas that look straight, but object images always have some curvatures. Correction for three-dimensional distortions has worked well for tubular crystals of membrane proteins (Beroukhim and Unwin, 1997; Yonekura and Toyoshima, unpublished). In these methods, individual tube images are divided into several short segments, and then each segment is corrected for image distortions such as the in-plane and out-of-plane tilt by fitting its layer-line data to the averaged one independently. However, these methods turned out not to work so well for the flagellar filament as for the tubular crystal of Ca$^{2+}$-ATPase (unpublished results). Because the diameter of the filament is relatively small, about 240 Å, approximately a third of the tube, dividing each repeating unit into short segments made the signal-to-noise ratio severely worse. This would result in poor fitting for each segment against the averaged dataset. In this case, straightening the filament axis by cubic splines worked straightforward. This well-established method can correct for only in-plane bending, but successive correction along the filament is possible (Carragher et al., 1996; Egelman, 1986; Owen et al., 1996). In addition, the increase in the area containing straight object available for the structure analysis can reduce the number of filament images required for a target resolution to be achieved. This is particularly important for samples that have relatively long repeat distance and tend to be bent. For example, the L-type straight flagellar filament typically has a much longer repeat distance (~4000 Å) than the R-type (~1700 Å), and images of its frozen hydrated samples show more curved filaments. Hence, straightening is necessary for high-resolution structure analysis (Maki-Yonekura et al., unpublished), while it was not used for the structure analysis of the R-type (Yonekura et al., 2003). XDISPHCV is a program to straighten the image using cubic splines. This program uses a cross-correlation map to determine the positions of the filament axis as the PHOELIX package does (Carragher et al., 1996). In this procedure, running averaging along the approximate filament axis is effective for cryoEM images of frozen hydrated filaments such as the flagellar filaments, which show a very poor contrast due to its relatively small diameter (Fig. 4). XDISPHBOX and XDISPHCV take an MRC format image file without any definition files or environmental variables. XDISPHCV carries out the straightening by quadratic interpolation and therefore produces images with very little deterioration in image intensities and hence is much more effective in extracting higher-resolution signals than those obtained by linear interpolation. We compared peak amplitudes of some layer lines from a typical image of the flagellar filament straightened by linear and quadratic interpolation. A peak amplitude of a layer line at the axial spacing of ~50 Å did not show any difference, but that of ~25 Å became ~0.6% higher and that of ~12.5 Å became ~2.8% higher in the image straightened by quadratic interpolation. This indicates that much higher gains in amplitude recovery can be expected at higher resolution and the consequential advantage in the quality and resolution of final image reconstruction would be obvious.

4.3. Determination of the out-of-plane tilt and the shift normal to the helix axis, and initial editing

The third part is the determination of the out-of-plane tilt ($\omega$), refinement of the shift normal to the helix axis ($\Delta \psi$) and initial editing of the layer-line regions. For
these purposes we developed HLXSX, which displays layer-line profiles in the X-Window system. Using this program, selection of amplitude peaks to be used for the determination of \( \omega \) and \( \Delta \phi \) can be easily done by mouse click. This task could be carried out automatically: strong peaks could be selected automatically from each layer line; then, initial \( Q \)-value be calculated by SRCH; peaks with bad phase residuals be removed and \( Q \)-value be calculated again. But, this simple protocol is not suitable for tubular crystals whose helical classes are unknown; it would remove good amplitude peaks due to bad phase residuals resulting from wrong assignment of a helical class. Using HLXSX, many different helical classes can be easily and quickly tried out (see Section 3.4). Once a control file for SRCHAID is made, it can be read into HLXSX and used for other tubes belonging to the same helical class, because peak positions on each layer line are virtually identical as long as the images were taken at the same magnification. We do not need to pick up all the peaks for each tube. We just need to feed a template and edit the positions of peaks. The same thing is also true for editing of the start and end positions of layer lines. HLXSX can reduce the time required for processing images to \( \sim 1/10 \) compared to the previous procedure by using HLX.

4.4. High-resolution structure analysis of the flagellar filament

We have recently analyzed the three-dimensional structure of the flagellar filament at \( \sim 4 \AA \) resolution and built its complete atomic model (Yonekura et al., 2003). This is one of the first atomic models of macromolecules obtained solely by image analysis. For single particle image analysis, it is generally believed that the number of images from a few tens of thousands to hundreds of thousands is required to obtain a map at \( 10–7 \AA \) resolution. We have demonstrated that a quite small number of molecular images (\( \sim 41,000 \)) is sufficient to produce a high-resolution map. The reason why such small number of images allowed us to obtain a high-resolution map can be considered as follow. First, a liquid helium-cooled and highly stable specimen stage of the electron microscope (Fujisho, 1998) produced high quality images with sufficient reduction of radiation damage in high-resolution structural information. Second, the structural order and helical symmetry of the flagellar filament is high enough to give rise to high-resolution structural data, as demonstrated by X-ray fiber diffraction patterns showing sharp layer-line reflections beyond 3 \( \AA \) resolution (Yamashita et al., 1998). Also, the helical symmetry makes initial screening of good images easy, because their Fourier transforms show sharp layer lines arranged symmetrically across the meridian. In single particle image analysis, it is generally difficult to select out only good images. As a result, incorporation of bad ones would make the quality and resolution of the final map worse than a limited number of good images are supposed to give. For the image analysis of the R-type straight flagellar filament, we selected \( \sim 100 \) good filament images recorded in \( \sim 60 \) micrographs, which are among \( \sim 1000 \) micrographs we collected (Maki et al., unpublished). Third, many new devices in the helical image reconstruction method, implemented in many newly developed programs and algorithms described above and in those to be described elsewhere, allowed accurate alignment within individual images and between images. For this purpose, we have developed many programs for box refinement, distortion corrections (unpublished), solvent flattening (Yonekura and Toyoshima, 2000b) and so on. The first part of the analysis for the individual images is described here, and the rest will be described elsewhere together with an overview of the whole procedure comparing the new and conventional methods in more detail. The GUI programs developed here have been used to extract the highest possible signals from individual images of the flagellar filament, which is now proved to be essential for accurate image alignment and high-resolution structure analysis from a relatively small number of images.

5. Conclusion

By using these GUI programs developed here, the time required for processing an image becomes shorter by twice to ten times, and also data can be extracted to significantly higher resolution.

Acknowledgments

We thank Toshiyuki Yamashita for his help in developing XDIPHFFT and HLXSX, Toshihiko Akiba for originally developing ACFMAP and IMAGE2PS, which is a base program of IMAGEH2PS.

References


