

INTRODUCTION TO ELECTRON DIFFRACTION PROCESSING USING MRC 2D-CRYSTAL PROCESSING SOFTWARE

Anchi Cheng (acheng@scripps.edu) 2005

Note: This text was prepared for a book chapter on practical guide to the MRC package Vinzenz Unger and I wrote but never published in 1997. It may be outdated in term of the program names but the concept should apply nonetheless.

Apart from recording images, electron microscope can also be used to record diffraction patterns with proper adjustment of focus length. Since the diffraction pattern is not sensitive to specimen movement, and is not modified by CTF, the amplitudes of the reflections so derived are more reliable than those derived from images. It is therefore advantageous to use the amplitudes derived from the diffraction and phases from the images together to construct the structure factor list when the former is available. However, to be able to collect diffraction data, the crystals needs to be relatively large ($> 1 \mu\text{m}$) and well ordered. The size required increases further if the unit cells are larger and/or if the packing density is lower (Wang and Kühlbrandt, 1991). So far, only handful of 3-D structures of 2-D protein crystals have been analyzed utilizing electron diffraction data. These proteins include bacteriorhodopsin (Baldwin and Henderson, 1984), PhoE porin (Walian and Jap, 1990), light harvest complex II (Wang and Kühlbrandt, 1992), tubulin (Wolf et al., 1996) and aquaporin 1 (Cheng et al., 1997; Walz et al., 1997).

The data collection procedures for low-dose electron diffraction patterns have been described in detail for several systems (Ceska and Henderson, 1990; Walian and Jap, 1990; Wang and Kühlbrandt, 1992). General procedures for collecting diffraction data on CCD camera are also available (For example, Brink and Chiu, 1994). The main focus for this section of the text is to outline the general procedures in using the analysis programs.

As in the case of processing image, the first step is to digitize the diffraction pattern. The scanning step size should be small enough that even the sharpest diffraction peak is well sampled in any direction. Further, since the distribution of the intensity is much broader in the case of diffraction than that of images, it is critical to carefully calibrate the scanner gain, offset etc. An optical density-electron dose look-up table is also required as described by Baldwin and Henderson (1984) and can be obtained by scanning linear exposure series micrograph that are developed and scanned under the same condition as the actual diffraction patterns. If a CCD camera is used for data collection, a linear look-up table is adequate here since the electron dose conversion is normally performed before the pattern is saved.

Once the diffraction pattern is scanned, the analysis follows steps which have been described previously (Baldwin and Henderson (1984), Ceska and Henderson (1990))

- The center of the pattern is first determined. A radial background curve ascribed to the inelastic and diffuse elastic scattering is generated, and the long time-dependence baseline drift and fluctuation of the densitometer is estimated (all by BACKAUTO). Both the radial background and the densitometer baseline drift will be corrected in the intensity integration step.

- The reflections are indexed and rough cell parameters are determined (AUTOINDEX).
- The integration of the reflection intensity is performed by (PICKAUTO) or (PICKPROF). Both programs refine the cell parameters and parameters describing the pincushion effect with the position distortion due to the curvature of the Ewald sphere considered. As a result of these refinements and correction, the final locations of the reflections are accurate to the highest resolution and tilts. Using the optical density-electron dose look-up table and the parameters and correction curves previously determined by BACKAUTO, the programs then integrate the spot intensities, average the intensity of the Friedel pairs, and provide useful statistics such as the number of reflections with positive and negative intensity, the agreement between Friedel symmetry related spots, which is written as R Friedel.
- With the completion of intensity integration, the parameters used in all previous steps should be evaluated and optimized for best result. In general, a strong pattern gives good statistics such as small R Friedel with any program used. Optimization of parameters becomes more crucial for tilted or weaker patterns. A better intensity integration produces smaller R Friedel, and the number and spatial distribution of the significant reflections will correspond well to our eyeballed results. After all, our eyes are superb integrators.
- Before proceeding with merging data sets, one has to reject Friedel pairs that differ widely due to dust or scratches. The very low resolution reflections which measurements are unreliable due to extremely high background from inelastic scattering should be removed from the data set, too. Both can be done with preset limits in the programs. To widely different Friedel Pairs, a cutoff in the difference between the intensity of the (h,k) and (-h,-k) reflections is defined. Such uniform cutoff is quite convenient. However, the user should carefully check what reflections are removed since that there are cases at high resolution in which genuine difference in intensity may be present as the result of z_1^* not equal to z_2^* at high resolution in the pair (h,k, z_1^*) and (-h, -k, $-z_2^*$) and that the particular lattice line has a steep gradient of intensity. These reflections can, in general, be included in the analysis (Baldwin and Henderson (1984)). Also commonly observed at this stage is that significant percentage of the diffraction intensity may have a small negative value. One source for this systematic error has been identified as the high local background produced by disorder in the crystal. The need for correcting this effect increases at higher resolution (Grigorieff and Henderson, 1995) and, since the correction requires a reliable model to describe the diffuse background, it is not commonly performed.
- The next step in the processing involves initial scaling and merging (MERGEDIFF) of data, as well as lattice line fitting (SYNCFIT or LATLINE). In the case of 3D data, assigning the correct tilt geometry is as important here as in the case of image processing, and the risks taken by bad merging are similar. Unlike for the images where changes in defocus can be used to estimate tilt geometry, all initial tilt geometry determinations in diffraction patterns are based on the lattice distortion upon tilting (Shaw and Hills, 1981). Different strategies have been used successfully in merging diffraction data: For example, Baldwin and Henderson (1984) merged only data with high tilt angle to obtain the initial reference data set so that this data set requires little or no further tilt geometry refinement. On the other hand, Wang and Kühlbrandt (1992) started merging diffraction data of LHC II from low tilts. The initial

scaling required at the point of merging is based on the total scattering intensity of the data sets, which may be further improved in the refinement process after a rough lattice line fitting is performed.

- MERGEDIFF will be used repeatedly along with a previous fitting result for refinement of tilt geometry and scaling with application of global temperature factors to each pattern so that the scaling between patterns reflects both the total and the resolution-dependent scattering power. Application of global temperature factor is common for diffraction data merging. However, it is important to realize that this should be done only to merge data of comparable quality but not to boost contribution of portions of poor data. A properly refined data set will result in improvement toward smaller merging R-factors and smaller fitting R-factors for the following lattice line fitting.

In the past, the diffraction and images are often performed independently. As the result, the structure factors of the 3-D lattice points used in map calculation contain the phases of the image lattice line fitting and the amplitudes of the diffraction intensity fitting. Alternatively, LATLINE allows simultaneous fit combining the experimental phases from the images with the intensities from the diffraction patterns and thus give a model consistent fit to both phases and amplitudes (Agard, 1983).

At last, the combined structure factor list should be checked and the users should fill in the very low resolution structure factor amplitudes with properly scaled values derived from image processing. Failure in doing so may distort the final map. For example, low-resolution amplitudes to 30 Å was taken from image transform when the 3-D map of LHC-II was calculated (Kühlbrandt and Wang, 1991).

The map calculation and its evaluation from this point on is the same as that of an image-derived structure. In general, although the features of the maps may be similar, the combined map will be better defined.

The MRC package for the electron diffraction processing has not been used as widely as that for image processing because of the limited number of existing examples. The programs for the former are therefore not as general and fast. Users should keep in mind that they may need to modify the procedure or even the programs for their special need. Anchi Cheng 1997

BACKAUTO: This program generates a few curves and values for corrections required prior to spot intensity integration. It searches for the center of the pattern in the input pattern file, $\{\text{pattern}\}.pat$, and with a few passes of filtering and smoothing, calculates a radial density curve ($\{\text{pattern}\}.bck$) which is stored for background correction. NPNTS is the radius in numbers of pixels for smoothing, and should be much larger than the width of the diffraction peaks. A background corrected output pattern file is also produced ($\{\text{pattern}\}.crt$). In addition, long time-dependent fluctuation in the densitometer performance is estimated using a curve representing the variation in Y-axis, the slow scan axis. IRMAXC and IRMINC defines the range of radius in which the variation in intensity is considered for the Y-variation curve at a given point. IRMAXC and IRMINC are therefore used to exclude the unusually noisy radii and/or the large inelastic background area from the Y-variation curve calculation.

AUTOINDEX: This is a program to perform automatic indexing of a background corrected pattern file, $\{\text{pattern}\}.crt$. Suggested raster sizes for spot integration are also generated, and the choices are visualized in the output plot $\{\text{pattern}\}.i.ps$. The lattice base vectors and the raster sizes are input into the header of the input file, $\{\text{pattern}\}.crt$. An average picture near an averaged spot is presented as $\{\text{pattern}\}.avr$ for users to examine. The only parameter need to be defined manually is the radius of circle omitted in the peak searching because it is too close to the direct beam.

PICKAUTO, PICKPROF: These are programs performing the spot intensity integration with various corrections. **PICKAUTO** uses fixed peak and background raster sizes in the integration. **PICKPROF** contains the function of **PICKAUTO** and may perform intensity integration with the help of fitting the reflection with an average profile generated from strong reflections in the pattern. **PICKPROF** also allows correction of position distortion in CCD camera. In both programs, a number of geometric corrections such as shrinkage of pixel size in one direction relative to the other and offset of even/odd scan lines can be accounted for. Corrections are made in search of the reflection centers of gravity for pincushion distortion and for curvature of Ewald sphere. In addition, the programs also correct for radial background and convert optical density to electron dose before the integration. A few parameters, when input as zero, can automatically be extracted from the headers of the input files. The intensity integration programs, **PICKPROF** in particular, has a large number of options regarding how the refinement and integration are performed. The example shown in the protocol session presents the simplest profile fitted integration procedure. The users can start from such simple setup (or, even simpler, without profile fitting) and select the most suitable options for the data once the program runs smoothly. In general, a complex option should not be chosen unless it provides significant improvement. A clear indication of proper option choosing and optimized parameter setting are a lower Friedel R-factor with more significant reflections and the distribution of the calculated integrated intensities resemble more that on the micrographs. For example, local, spot-by-spot profile fitting was used in analyzing aquaporin 1 data.

Other than the ones discussed in previous programs, the parameters are grouped and explained in the following:

(1) parameters for data selections:

XAMINE - threshold of the peak raster used to reject weak reflections for the center of gravity and lattice parameter calculations.

FRACT, ABSOL - criteria for Friedel difference rejection in term of fractions to the Friedel average and of absolute values.

(2) parameters for distortion corrections:

ANGDIS, SHRINK - angle on film (ANGDIS) perpendicular to which a shrinkage factor SHRINK is applied in cases of pixel size distortion are found in densitometer.

TILDIR - direction of tilt on microscope, +1 or -1. The choice of one or the other determines the sign of the Ewald sphere curvature correction term. The correct sign will result in peak search of a reflection sampling z^* toward the incident beam to become further away from the origin of the diffraction than what is expected of a flat Ewald sphere. The choice depends on the actual data collection condition of the microscope and densitometry, and the assignments of lattice base vector like what was described in assigning tilt geometry for images except that here +1 stands for positive tilt angle and -1 stands for negative tilt angle.

LUNBEND - (Only in PICKPROF) If .true., a position distortion table TABLEIN is used for data derived by CCD detector.

INTERLEAVE - (Only in PICKPROF) Apply offset to odd/even scan lines to eliminate densitometer zig-zag offset if .true.

(3) integration and profile fitting parameters:

LPROFIT - (Only in PICKPROF) If .true., perform profile fitting for spot integration. If .false., perform fix raster addition spot integration. The latter produce results similar to that from PICKAUTO.

LPTYPE - (Only in PICKPROF) The type of profile fitting if used. For example, LPTYPE=0 gives the simplest one profile for the whole film and 2 for spot-by-spot basis. In the latter case, spots surrounding the target spot in a radius of LRANGE are used to produce the local averaged profile. When IFLATTEN is .true., PROFMIN is used to as threshold below which the profile intensity is set to zero so that the baseline of the profile is flattened.

NXM, NYM, NXMT, NYMT - peak raster sizes for fixed raster spot integration near and far from tilt axis, respectively. They are also used in detecting spot and finding center of gravity of reflections in the step that refines lattice parameter and pincushion correction parameter B3. Because the lattice parameter

and B3 determines where the spot integration take place for every reflection, this set of raster sizes as well as the initial lattice parameter affect strongly the integration result.

NXB, NYB - background raster sizes for PICKAUTO. It stands for the maximum boundary size for profile fitting in PICKPROF and is calculated from NXM, NYM, NXMT, and NYMT if PICKPROF is used.

MERGEDIFF: This program handles sets of electron diffraction patterns. It scales and sorts the data and produces a merged list of the data. Against reference lattice lines, tilt geometry refinement is also possible and the scaling may include the use of isotropic or anisotropic temperature factors. Many of the flags assisting merging reflections that are indexed differently in each film are similar to those used in ORIGTILT. In addition, detwining option is included following the example of bacteriorhodopsin (Baldwin and Henderson, 1984).

When SORT = .true., several outputs are possible. The most important one is the sorted file, dmerge.dat, by choosing NFILE3 = .true. LSUMMARY = .true. produce a convenient summary of the parameters and results.

Several parameters in the program are related and some are forced to change when conflicts occur. The user should therefore check if the desired operation is indeed carried out. One example is the relationship between SCFDET, DETWIN, TA_REFMT, LCURVES, and LREF. For instance, a reference data is considered used (LREF = .true.) whenever a reference curve is used (LCURVES = .true.) since the curve is a kind of reference data. Scale and temperature factor (SCFDET) and/or tilt geometry (TA_REFMT) can't be refined or determined without a reference curve in place (LCURVES = .true. and LREF = .true.). Further, detwinning of the crystals (DETWIN = .true.) can only be performed when scale factors are allowed to change (SCFDET = .true.).

The spot intensity input files are identified by their unit number NIN in the program. Here, we can refine isotropic (LBISO = .true.) or anisotropic (LBISO = .false.) temperature factors when reference data is used. The anisotropic temperature factors are input as TFPAR and TFPERP for the value in the direction parallel and perpendicular, respectively, to the tilt axis. Only TFPAR is used in isotropic temperature factor refinement. If an individual scale factor is not zero, then SCFDET is set to .false. automatically.

WIDTH parameter is in Å, and WSTAR = 1/WIDTH is the general sampling step in the reciprocal step for the refinement. WSTAR will be overwritten, however, by the individual values of WIN in Å⁻¹, and further, by the sampling step in the reference curve if the latter is more loosely sampled.

LCF (T/F) determines whether to produce a pre-LCF file for difference Fourier when such a task is required (Ceska and Henderson, 1990).

SYNCFIT: This is the lattice line fitting program written for electron diffraction. It is specific for individual two-sided plane group, though. For general symmetry, LATLINE with IPAT=1 or 2 can be used. See LATLINE description for details. Users should be aware that SYNCFIT fits both positive

and negative spot intensities while **LATLINE** only accepts positive intensity or amplitude. While it is still debatable about how negative intensity should be treated, the negative values need at least be set to the smallest positive number if they are intended to be input to **LATLINE**.