

# 3D RECONSTRUCTION FROM 2D CRYSTAL IMAGE AND DIFFRACTION DATA

Andreas D. Schenk,<sup>\*</sup> Daniel Castaño-Díez,<sup>†</sup> Bryant Gipson,<sup>†</sup>  
Marcel Arheit,<sup>†</sup> Xiangyan Zeng,<sup>‡,§</sup> and Henning Stahlberg<sup>†</sup>

## Contents

1. Introduction to Electron Crystallography Data Processing	102
2. Algorithms for Electron Crystallography	104
2.1. Initialization of a project	107
2.2. Processing of crystal images	109
2.3. Processing of electron diffraction patterns	112
2.4. Merging and 3D reconstruction	113
2.5. Molecular replacement and model building	116
3. Image Processing with 2dx	117
3.1. Project initialization in 2dx	117
3.2. Processing of an individual image in 2dx	118
3.3. Merging in 2dx	120
4. Electron Diffraction Processing with XDP	120
4.1. Project initialization in XDP	120
4.2. Processing of an individual electron diffraction pattern in XDP	121
4.3. Merging in XDP	122
5. Electron Diffraction Processing in IPLT	123
5.1. Project Initialization in IPLT	123
5.2. Processing of an individual electron diffraction pattern in IPLT	123
5.3. Merging in IPLT	124
6. Conclusions	125
Acknowledgment	127
References	127

<sup>\*</sup> Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA

<sup>†</sup> C-CINA, Biozentrum, University of Basel, Basel, Switzerland

<sup>‡</sup> Department of Mathematics and Computer Science, Fort Valley State University, Fort Valley, Georgia, USA

<sup>§</sup> EON Corporation, California, USA

## Abstract

Electron crystallography of 2D protein crystals can determine the structure of membrane embedded proteins at high resolution. Images or electron diffraction patterns are recorded with the electron microscope of the frozen hydrated samples, and the 3D structure of the proteins is then determined by computer data processing. Here we introduce the image-processing algorithms for crystallographic Fourier space based methods using the Medical Research Council (MRC) programs, and illustrate the usage of the software packages 2dx, XDP, and IPLT.

## 1. INTRODUCTION TO ELECTRON CRYSTALLOGRAPHY DATA PROCESSING

Electron crystallography records electron micrographs of thin, sheet-like, two-dimensional (2D) crystals. Such 2D crystals of proteins are sometimes encountered *in vivo*, as for example the purple membrane of *Halobacterium salinarium* (Henderson and Unwin, 1975). Alternatively they can be obtained from detergent-solubilized and purified membrane proteins by reconstitution into a lipid-bilayer at very low lipid-to-protein ratio (Jap *et al.*, 1992; Kühlbrandt, 1992). Images of noncrystalline membrane-reconstituted membrane proteins usually do not provide sufficient signal-to-noise ratio (SNR) to allow detection of the protein location and orientation, so image processing by signal averaging methods fails. However, in 2D crystals the membrane proteins form a regular array in the membrane, so that image processing by Fourier filtering and averaging becomes possible. 2D membrane protein crystals are usually single-layered but in some cases, multilayered 2D crystals can be observed. Multilayered crystals often show a better crystallinity, which can be explained by the additional crystal contacts in between the different layers. If the number of crystal layers is small and defined, for example, 2 or 3, then the structure reconstruction from images of these crystals is possible, and might even benefit from additional symmetry present in the multilayered crystal. However, if the number of layers is too high and/or varying, as would be the case for 2D crystals that have a strong tendency to stack, then the structure determination from such crystals faces strong obstacles with the currently available software tools (Kulik, 2004).

Electron crystallography records images of negatively stained or frozen hydrated 2D crystals, or subjects the 2D crystals to selected area electron diffraction. Such data can be recorded from nontilted crystals, or the crystals can be imaged at a tilt angle. Due to the random orientation of the unit cell vectors in comparison to the tilt axis position, data recording at a fixed set of tilt angles can fully sample Fourier space apart from the missing cone, similar

to random conical tilt reconstructions in single-particle cryo-EM. The combination of data from different tilt angles ranging from  $0^\circ$  to  $70^\circ$  then produces the final 3D map of the protein.

Recorded diffraction patterns of 2D crystalline samples show diffraction spots, which can be evaluated to give the amplitudes of the protein structure. This can allow the determination of precise amplitude data up to very high resolution (Gonen *et al.*, 2005). If the 2D crystals were badly ordered, however, then such diffraction pattern would only show low-resolution diffraction orders and would therefore be unsuitable for structure determination by electron diffraction. Recorded real-space images of 2D crystals can be Fourier transformed, and the resulting complex-valued Fourier transforms give access to the amplitudes and *phases* of the protein structure. In case of moderate crystal defects in the 2D crystals, the recorded images can be computer-processed prior to the calculation of the Fourier transformation, to computationally correct (“unbend”) the crystal distortions in the images. This process of crystal “unbending” was introduced by Henderson and Unwin (1975) and allows to significantly improve the number and quality of high-resolution spots in the calculated Fourier transformations of the images, thereby giving access to higher resolution data.

Computer data processing for electron crystallography using the “unbending” procedure and evaluation of calculated Fourier transformations or electron diffraction patterns were implemented in the so-called MRC program suite. The MRC image-processing package was written and maintained over the past 30 years by Richard Henderson in the Medical Research Council (MRC) in Cambridge, UK, and others (D. Agard, L. Amos, T. Baker, J. Baldwin, T. Ceska, R. A. Crowther, D. DeRosier, E. Egelman, S. Fuller, T. Horsnell, P. Moore, J. M. Short, G. Vigers, and others; see Crowther *et al.*, 1996). Over many years a large set of programs has been written for processing images of 2D crystals (Henderson *et al.*, 1990; Kühlbrandt *et al.*, 1994; Murata *et al.*, 2000; Unwin and Henderson, 1975). The MRC programs partly make use of functions and routines from the CCP4 program suite (Collaborative Computational Project, 1994), and the MRC programs later became the basis for several other software packages. Parts of the original Fortran MRC software code can be found in several of today’s biological image-processing tools, including those for processing noncrystalline specimens (see, e.g., Chapter 16).

SPECTRA from the ICE package was a program that generated Unix Shell scripts that launched an early version of the MRC programs (Hardt *et al.*, 1996; Schmid *et al.*, 1993). SPECTRA contained its own image visualization program, and facilitated the use of the MRC programs. The SPECTRA software was discontinued several years ago and is not available any more. Another similar programming effort is the GRIP system by Wilko Keegstra in the University of Groningen (unpublished). In recent years, other electron crystallography image-processing solutions became available that

follow the same unbending idea. Similarly to SPECTRA, 2dx (Gipson *et al.*, 2007b) is a front-end interface for user-friendly interaction with the MRC programs and offers user-guidance and help functions. 2dx in addition offers optional full automation of the 2D and 3D processing and includes tools to organize the project, automate the workflow, merge the extracted data (Gipson *et al.*, 2007a), and apply a single-particle maximum likelihood program to the 2D crystal images (Zeng *et al.*, 2007b). XDP is another front-end software system for the MRC programs, which is mainly focused on the evaluation of electron diffraction patterns (Mitsuoka *et al.*, 1999), and which is complemented by a graphical user interface (GUI) and a set of scripts to simplify the use of the MRC programs. IPLT (Philippesen *et al.*, 2003, 2007b) is a highly modular and adaptable software package developed in C++ and Python that provides a reimplement of existing algorithms complemented by newly developed algorithms for processing images and diffraction patterns. In addition it provides a GUI to guide the user through the processing and merging of electron diffraction data. CRISP and ELD are commercial software packages for the processing of 2D crystal data (Hovmöller, 1992). XMIPP also offers certain functionality for 2D crystal image processing (Sorzano *et al.*, 2004).

Here we describe the general image-processing algorithm for the image unbending approach, the evaluation of computed Fourier transformations and electron diffraction patterns, and the 3D merging of the measured data to obtain a 3D reconstruction of the protein. We introduce the unbending algorithm as implemented in the MRC programs, and briefly introduce the 2dx, XDP, and IPLT programs and their usage.

## 2. ALGORITHMS FOR ELECTRON CRYSTALLOGRAPHY

Structural data of 2D protein crystals can either be gained by recording real-space images or electron diffraction patterns in an electron microscope. Both can be recorded on either film or CCD cameras. In practice, however, high-resolution images are best recorded on photographic film and digitized with a scanner, to make use of the better point spread function (PSF) of film compared to that of a CCD for higher voltage electrons (Downing and Hendrickson, 1999). Electron diffraction patterns are best recorded on digital CCD cameras, to make use of the higher bit-depth of the CCD pixels compared to film. In addition recording on a CCD camera offers a wider linear range and an overall increase in data recording speed as no subsequent digitalization of the recorded data is necessary.

Digitized images or diffraction patterns are then processed to extract structural data. Although image and diffraction data share certain features, the different challenges encountered during data extraction mean that two

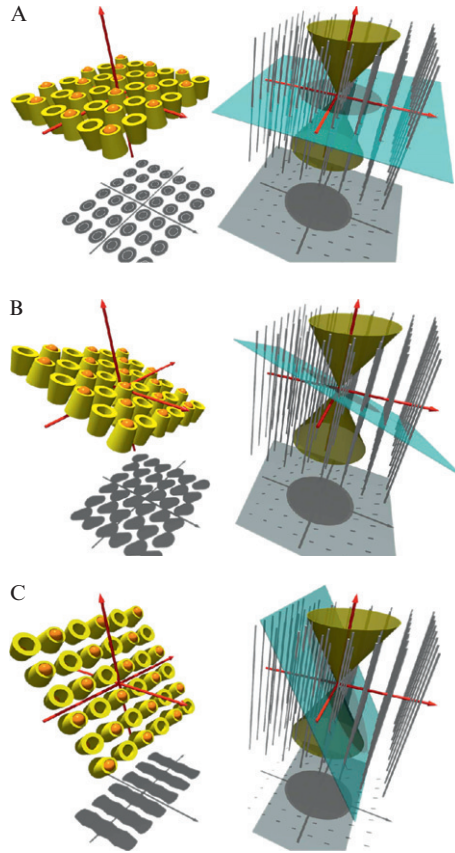
different sets of algorithms have to be applied. The processing of electron crystallography data can roughly be divided into four phases:

1. setup of the processing project and determination of the basic crystal parameters;
2. processing of the individual images and/or diffraction patterns;
3. merging of the data of individual images or patterns into one 3D dataset, refinement, lattice line fitting, and discretization of the dataset; and
4. building of an atomic model and model refinement.

For each of the phases we will outline the steps to be performed in a general fashion—also discussing some of the pitfalls and difficulties inherent to the processing—and then provide introductory protocols for the use of 2dx, XDP, and IPLT.

The general goal of the electron crystallography data processing is to determine the Fourier representation of the protein structure. This then allows calculating the real-space map by a simple Fourier back-transformation. While a 3D crystal would have as its Fourier representation a set of diffraction spots in 3D Fourier space, the 2D crystals have as their Fourier representation a set of vertically arranged “lattice lines.” These continuous lines in Fourier space lie on a regularly spaced  $h, k$  grid, which is defined by the crystal axes  $a, b$  of the 2D crystal. Because the 2D crystal has in the third dimension, the  $z$ -axis, only one layer (or a very small number of layers), its Fourier representation is continuous in the vertical  $z^*$  direction, thereby giving rise to the continuous lattice lines. Due to the limited tilt range at which 2D crystal specimens can be imaged (usually  $\leq 70^\circ$ ), a cone of Fourier space cannot be measured, as shown in Fig. 4.1.

It is interesting to note that a 2D crystal sample is limited in the vertical dimension (there is no crystalline structure above and below the 2D crystal sample), but in the horizontal membrane plane direction the crystal is generally continuous, so that no empty boundary wall can be drawn around one crystal unit cell. The task of electron crystallography is to reconstruct the protein structure that populates the volume of the crystal unit cell. This is by definition a periodically continued unit cell in the horizontal two crystal axis directions. The Fourier transformation (calculated by the FFT) also assumes periodic boundaries in the vertical direction for the 2D crystal unit cell, but nevertheless the protein structure is embedded in a larger vertical space that contains empty voxels above and below the protein. In other words, in the vertical direction the crystal structure is masked by some masking function, so that a few nanometers above and below the protein structure the unit cell volume is empty. As the density for the protein structure is free to assume any density distribution, its Fourier representation itself also has no restrictions on its individual Fourier pixels (with the exception of the Friedel symmetry). However, the fact that the real-space volume is multiplied by a mask in the vertical direction means that the Fourier space



**Figure 4.1** Cartoon depicting the real-space (left) and Fourier space (right) representations of a 2D crystal. Three sample tilt angles are shown: (A) untilted, (B)  $30^\circ$  tilt, (C)  $60^\circ$  tilt. The imaging process in the microscope produces a projection image of the 2D crystal, as symbolized by the shadow under the crystal on the left side. This image is then corrected for lattice distortions and Fourier transformed. The resulting Fourier transformation corresponds to the central section that is indicated on the right as a plane. The values for amplitude and phase on the diffraction peaks in the Fourier transformation are measured and stored along the lattice lines at the position, where the lattice lines intersect with the plane (right). The crystal tilt angle defines the tilt of that plane, thereby defining the vertical  $z^*$  height of the measurement on the individual lattice lines. Each recorded image (left) thereby contributes only one (two due to Friedel Symmetry) measurement to a certain lattice line (right), while these measurements can under certain conditions be used for several lattice lines in case of crystal symmetry. Since the 2D crystals cannot be imaged at tilt angles higher than  $\pm 70^\circ$ , the lattice line values in the indicated missing cone region in Fourier space cannot be experimentally determined, resulting in the so-called “missing cone” problem.

representation of the protein structure is convoluted with the Fourier transformation function of that vertical mask. If the vertical mask in real space is a rectangle, then the convolution function in Fourier space is a Sinc function. (And if the vertical mask is a different shape, then the Fourier space convolution function is *not* necessarily a SINC function, see below.) A convolution of the Fourier representation of the protein structure with a vertical SINC profile means that every Fourier pixel is replaced by the SINC function multiplied by the real and imaginary values of the Fourier pixel. This results in a Fourier space volume, where only in the vertical direction there now is a neighborhood correlation. This fact is the basis for a recently developed software module that will become available in the 2dx software package in the near future, which allows to predict the Fourier space values in the missing cone based on the precisely measured (i.e., oversampled) Fourier pixel values outside of the missing cone (Gipson *et al.*, 2010). For this reason, the Fourier space representation of the crystal's unit cell has in the vertical direction a neighborhood correlation of adjacent Fourier pixels, while in the horizontal direction there is no such neighborhood correlation.

The purpose of processing one image of a tilted 2D crystal is then to obtain correct values for the amplitudes and phases for the points on the lattice lines, where they intersect with the tilted plane in Fig. 4.1. Once a sufficient number of measurements are available for every lattice line, a continuous function can be fitted to these measurements, which is usually a linear combination of SINC functions. These can then be equidistantly sampled, to give an evenly sampled Fourier volume, which is then Fourier back-transformed into real space to produce the final 3D map for the protein.

## 2.1. Initialization of a project

The first step to setup the processing for a new unknown 2D crystal is to determine the unit cell size and the symmetry of the crystal. The easiest way to do this is by recording untilted images of the crystal, either negatively stained or cryo-embedded. Images of negatively stained crystals with low amount of staining allow detecting the lattice also from poorly ordered crystals, due to the higher SNR of such images. In addition, when low amounts of stain are used, the stain preferably occupies the space between the supporting carbon film and the 2D crystal, so that the 2D crystal surface structure on the side toward the carbon film is stronger contrasted than the other crystal surface. Images of such preparations then show uneven surface staining, which allows distinguishing between up- and down-oriented molecules in the unit cells. Analysis of the difference map between maps from evenly stained and unevenly stained samples can allow understanding the 3D orientation of the molecules (Chiu *et al.*, 2007). Recording

images of cryo-embedded samples, however, has the advantage of increasing precision and to eliminate the effect of uneven staining, to more reliably recognize screw axis symmetries.

The unit cell parameters of the crystals can be determined from the recorded images by one of several algorithms, as for example described in Zeng *et al.* (2007a). An automated algorithm searches for the reciprocal lattice that has the best fit to the diffraction spots of the Fourier transformation of the image of a nontilted 2D crystal. For a single-layered 2D crystal image, the algorithm generates lattice candidates from the difference vectors between the identified diffraction spots. The strongest linearly independent low-resolution difference vectors are then good candidates for defining the basis vectors of the 2D crystal lattice in reciprocal space. These reciprocal lattice basis vectors  $u^* = (u_1^*, u_2^*)$  and  $v^* = (v_1^*, v_2^*)$  can be translated into their corresponding real-space lattice basis vectors, as

$$u = \left( \frac{v_2^*}{u_1^*v_2^* - u_2^*v_1^*}, \frac{-v_1^*}{u_1^*v_2^* - u_2^*v_1^*} \right) \text{ and } v = \left( \frac{-u_2^*}{u_1^*v_2^* - u_2^*v_1^*}, \frac{u_1^*}{u_1^*v_2^* - u_2^*v_1^*} \right) \quad (4.1)$$

The same formula applies to translate real-space vectors into their correspondents in reciprocal space. These real-space unit cell parameters for vector length and included angle are calculated once for the entire 2D crystal project.

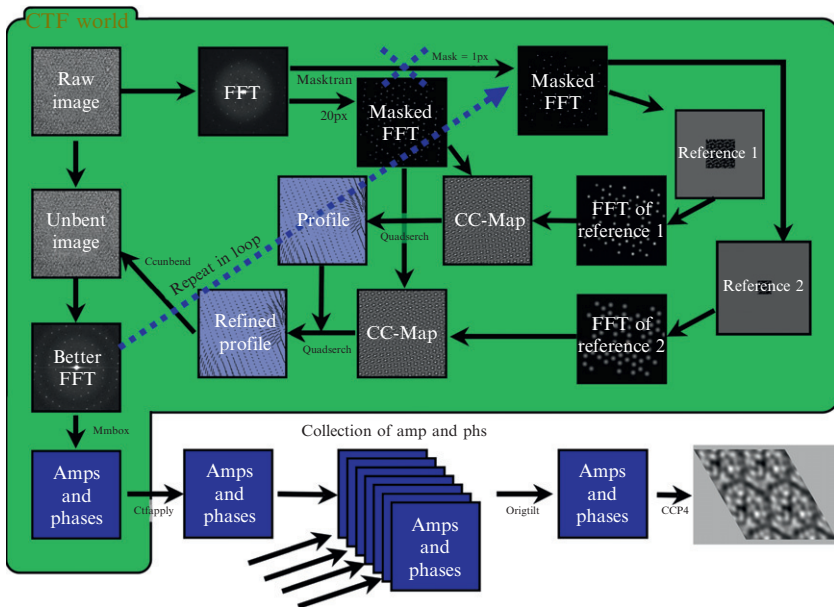
The symmetry of an unknown crystal can be determined from the evaluated diffraction spot data for amplitudes and phases, using the MRC program *allspace* (Valpuesta *et al.*, 1994). Certain crystal symmetries can be excluded in advance, depending on the angle between the two unit cell vectors and their relative length. Landsberg and Hankamer (2007) provide a comprehensive overview of the symmetries relevant to 2D crystals. Although prior single-particle studies of the protein in question can give a hint of the symmetry of the building block of a crystal, it should be noted that the crystal nevertheless could exhibit a different overall symmetry (see, e.g., Schenk *et al.*, 2005 or Vonck *et al.*, 2002). Crystals with an apparent three-, four-, or sixfold symmetry might also be affected by twinning (Yeates and Fam, 1999), which can make it impossible to process the data unless the twinning is taken into account. If the symmetry does not exhibit a twofold axis parallel to the membrane plane, the crystals can adsorb to the supporting grid in an up- or down-oriented manner. In this case, the orientation of each 2D crystal on the carbon film support has to be determined correctly, before 3D merging of the data is possible. Some proteins, as for example aquaporins (Gonen *et al.*, 2004; Hiroaki *et al.*, 2006; Kukulski *et al.*, 2005; Schenk *et al.*, 2005), tend to crystallize in double layers. This by itself is not a disadvantage as it often increases crystal quality, but it has to be

detected and the  $z$  dimension of the crystal has to be adjusted during lattice line fitting and discretization. A double-layered crystal can often be identified—even in low magnification images—by the double fringe at its edges.

## 2.2. Processing of crystal images

Real-space images of 2D crystals can be processed as a collection of densely packed single particles or as a crystalline arrangement using Fourier methods. The single-particle approach offers the possibility to classify particles from the same image into, for example, different conformations of the same protein, and in the case of severe crystal disorder can better localize and align the protein complexes to a common coordinate system than crystal unbending usually does (Zeng *et al.*, 2007b). For this review, we will focus on the unbending approach only.

The unbending of 2D crystal images attempts to position each crystal unit cell into the location that corresponds to the ideal lattice (Fig. 4.2). This



**Figure 4.2** Cartoon depicting the flow chart of the unbending algorithm for the case of a nontilted crystal sample. The raw image is modulated by the microscope's CTF (area on green background), which is typically only corrected after the unbending and evaluation of amplitudes and phases. For a detailed description see text. Figure reproduced from Gipson *et al.* (2007a).

unbending as implemented in the MRC program *ccumbend* typically moves small image blocks into the new location, and for this requires a distortion vector map that describes how each portion of the image has to be translated. This distortion vector information is obtained from a cross-correlation (CC) map between a reference map and a Fourier filtered version of the image. The reference map can be refined in several iterations, during which the reference patches become sharper and smaller, thereby allowing to trace at increasing confidence also the finer distortions in the crystal. *ccumbend* then uses this information to perform a block-wise unbending, whereby it shifts segments of the image, but does not rotate them. The alternative would be a smooth image unbending, using a spline-based pixel-wise interpolation as for example implemented in *2dx\_ccumbendh.for*, or as also available in IPLT, to produce unbent images without fractures. In case of images of materials sciences specimens that show atoms at high SNRs, such smooth unbending would appear more beautiful to the human eye, as it does not risk cutting projections of individual atoms into fragments (Morgan *et al.*, 2009). Another advantage of the pixel-wise unbending is that it compensates for rotation of parts of the image. Nevertheless, for biological 2D crystal samples, the advantage of a smooth interpolation method is not clear. With a well-aligned TEM, even in badly distorted 2D crystal images, the individual protein complexes should be imaged without any distortions within themselves. In some image cases, a spline-based pixel-wise unbending might distort the shape of the individual proteins, after which their high-resolution structure would be disturbed.

The unbending approach bears the danger to alter the high-resolution ripples of the extended PSF of the microscope, so that this could result in a loss of high-resolution information (Philippsen *et al.*, 2007a). In that case, a solution would be to first correct for the effect of the contrast transfer function (CTF), and then correct the lattice distortions (Mariani, 2009).

This is easily implemented for images of nontilted samples, but becomes difficult for images of tilted samples due to the complicated nature of the CTF-analog for tilted specimens. Due to the defocus gradient across the image, the diffraction spots in the Fourier transform of the image are split into two subpeaks separated in the direction perpendicular to the tilt axis. The distance of the two subpeaks increases with the distance from the tilt axis. For small lattice distortions the original amplitude and phase information can be extracted using the tilted transfer function (TTF) algorithm introduced by Henderson *et al.* (1986) that convolutes the local area around each peak with a corresponding peak profile to restore the original diffraction peak. The TTF algorithm is implemented in the set of *tt\*\*\** programs, as for example *ttbody*, *tmask*, and *ttrefine*. This algorithm, however, does not account for the asymmetry of the split subpeaks, which gets more prominent with increasing resolution as described by the tilted contrast transfer imaging function (TCIF) in Philippsen *et al.* (2007a). In addition the TTF

algorithm cannot correct the full image but only the diffraction peaks. Therefore, for the correction of images of noncrystalline samples (e.g., in thin-section tomography), or of badly distorted crystalline samples, where sample information outside the diffraction peak itself is desirable for optimal distortion correction, or for the correction of crystalline samples at very high resolution, where the nonsymmetric nature of the split peaks plays a role, a more general algorithm for correction of the imaging effects encountered in tilted samples would be of great value. Unfortunately, even though the work of Philippsen *et al.* provides the mathematical basis for this TCIF, a direct inversion and therefore correction of the TCIF is beyond the current computing power of today's limited computational resources.

The description of the lattice distortions is typically obtained by the MRC program *quadsrch* from a peak search within a CC map between the raw image and a reference map. The latter can be obtained by Fourier filtering the raw image and can be later refined by Fourier filtering unbent versions of the raw image. CC of this reference with the raw image then gives a CC map that contains peaks at the positions of the crystal unit cells. A peak search algorithm as implemented in the MRC program *quadsrch* tries to predict the approximate position of a peak in that CC map, based on the neighboring lattice nodes, and then locates the actual CC peak in the immediate vicinity of the predicted position only. This process can be iteratively refined, to produce a precise map of the crystal unit cells even with images of extremely low SNRs.

This list of unit cell coordinates can then be used to crop single-particle windows with the individual crystal unit cells out of the crystal image, as is offered by the 2dx software (Gipson *et al.*, 2007b). These unit cell particles can then be used for a single-particle processing, using a maximum likelihood algorithm (Zeng *et al.*, 2007b). The MRC software instead uses these coordinates to unbend the raw image, with the MRC program *ccunbend*. The Fourier transformation of the unbent image is then evaluated for the amplitude and phase values of the diffraction spots with the MRC program *mmbbox* (respectively *ttbox*), and each measurement is accompanied by a value for the average amplitude in the vicinity of the spot as background information, which then is translated into a figure of merit for the measurement. The amplitude for a spot is measured as the amplitude above the background value. The MRC software classifies the evaluated spots according to their SNR with so-called "IQ values": An IQ = 1 spot is at least seven times stronger than its background, an IQ = 7 spot is equally strong as the background, an IQ = 8 is weaker than the background, and an IQ = 9 spot has a negative amplitude, meaning that the Fourier transformation at that location has an intensity that is below the local average noise level. Spots of IQ values 1–3 are typically related to real measurements. However, since during the merging process the Fourier data are individually

weighted according to their SNR, even spots of IQ values 4–6 might contribute valid information to the merging process, while spot measurements of IQ 7–9 can usually be discarded.

### 2.3. Processing of electron diffraction patterns

The first step in processing individual diffraction patterns encompasses the determination of the beamstop position and shape and takes care of the masking of the beamstop and the center of the diffraction pattern.

The second step deals with the subtraction of the background of inelastically scattered electrons, which are dominating at lower scattering angles, and of the background from electrons that were elastically scattered from noncrystalline parts of the sample (e.g., the carbon film), which may appear also at higher scattering angles. This can be done with the MRC program *backauto*.

The next step concerns the determination of the reciprocal lattice vectors. Lattice determination for diffraction patterns is somewhat more involved than lattice determination for Fourier transforms of images, as the origin of the lattice is not in the center of the image and the lattice origin itself together with the lowest resolution diffraction spots is not visible because it is blocked by the beamstop. Depending on the software used for analyzing the diffraction data, the lattice can either be determined automatically, as offered by the MRC program *autoindex*, or as also possible with IPLT, or the user has to index the lattice manually by providing the position and index of several diffraction spots.

Collapsed vesicles—a quite common form of 2D crystals—often give rise to a set of two epitaxial, twinned lattices. Since the unit cell vectors of the two lattices do not coincide, the two lattices can be separated. Unfortunately, there is often a fair number of diffraction peaks overlapping. These spots have to be excluded from further processing. In the case of electron diffraction pattern processing, it is often easier to discard these patterns as a whole, if enough single-layered patterns are available. The indexing of the lattice also defines the tilt geometry. Due to the lack of a defocus gradient in comparison to images, the lattice distortion is the only source to define the tilt geometry for that crystal sample.

Once the lattice and the tilt geometry are determined, the diffraction intensity of all the peaks can be calculated by integrating over all the pixels contained in the individual peaks, as done by the MRC program *pickauto*. The size of the box for integration has to be optimized depending on the sharpness of the peak, which varies not only based on the sample itself but also based on the recording conditions, that is, the relation of CCD pixel size and camera length. A too large box size leads in general to an increase in noise in the data, whereas a too small box size leads to an underestimation of the peak intensity, as parts of the peak might get excluded from integration.

The peak intensities are corrected for background contribution, which most commonly is measured by calculating a background average from the area surrounding the peak.

## 2.4. Merging and 3D reconstruction

Merging of the individual measurements for image data (amplitudes and phases) or diffraction data (amplitudes only) into one merged dataset comprises the following steps: (1) symmetrization, (2) scaling, (3) merging, (4) tilt geometry refinement, (5) lattice line fitting, and (6) discretization. The first two steps 1–2, symmetrization and scaling, deal with each image or diffraction pattern individually. The third step, merging, fuses these into one combined dataset. The last three steps 4–6 benefit from the combined dataset to refine and evaluate the data.

Here we will give a general overview of the different steps. The specific protocols for merging data can be found in the descriptions of the individual software systems below.

Amplitude and phase data are present as a function of their  $h$  and  $k$  Miller indices. These, together with the information about the tilt geometry, allow to assign for each reflection also a  $z^*$  height, which describes the vertical position of these measurements along the corresponding  $h$ ,  $k$  lattice line.

### 2.4.1. Symmetrization

During symmetrization the diffraction peak measurements are moved into the unique subarea of Fourier space (called the “asymmetric unit”), according to the earlier determined symmetry. The measured peak intensities do not have to be adjusted, as they are invariant in all symmetry operations. The symmetrization reduces the number of lattice lines to be fitted and increases the number of data points per fitted lattice line. In addition, comparison of symmetry related points gives an estimate of the quality of the dataset. If the symmetry of the crystal contains a screw axis, then the Fourier space representation of the data will have systematic absences. In this case, those Fourier pixels can then be set to zero to enforce those absences.

### 2.4.2. Scaling

In contrast to X-ray crystallography, where the whole dataset is recorded from one crystal, every 2D crystal image or electron diffraction pattern is recorded from a different crystal, which is usually embedded and/or imaged slightly differently. This leads to a varying absolute scale of the amplitudes, which means that the individual image data or diffraction patterns have to be scaled against each other before merging. For this, an image dataset or diffraction pattern can be compared to the reference in the area with a similar  $z^*$  range (common line scaling) or the patterns can be scaled

according to their weighted resolution binned average intensity, similar to the well-known Wilson plot (Wilson, 1942) in X-ray crystallography. As this method does not depend on comparing individual reflections in a narrow  $z^*$  range, it avoids the necessity to collect a large fraction of diffraction data at low tilt angle to be able to scale the dataset. Potential errors in determination of the scale factors will negatively impact the lattice line fitting procedure, thereby contributing to the error in the amplitudes during lattice line fitting.

### 2.4.3. Merging

Once the individual datasets are scaled, they can be merged into one big dataset. This can be done with the MRC program *origtilt* for images and *mergediff* for diffraction patterns.

### 2.4.4. Tilt geometry refinement

An important step in merging a dataset is the refinement of the tilt geometry. For data from only slightly tilted specimens, small errors in the measured lattice vectors can falsely be interpreted as strong changes to the tilt geometry. For image data and specimen tilts  $\leq 25^\circ$ , the tilt geometry is therefore more reliably determined from the defocus gradient across the image. For electron diffraction data of specimens at low tilt, the tilt geometry determined from the lattice vectors cannot be considered reliable. At higher tilt angles  $\geq 30^\circ$ , the tilt geometry assignment for an individual dataset can be refined from the lattice distortion. After the merging step, the available 3D dataset can also be used to refine the tilt geometry for each individual image or diffraction pattern, by comparison of the data with the 3D reference dataset (MRC program *origtilt*). Potential errors in the tilt geometry lead to errors in the  $z^*$  direction later on during the lattice line fitting and discretization.

### 2.4.5. Lattice line fitting by SINC interpolation

Before the dataset can be discretized, the lattice lines have to be fitted along  $z^*$ . Most programs use a sum of SINC functions in one variation or another to fit the lattice lines (MRC program *latline*). The parameterized lattice line functions are then discretized into evenly sampled Fourier pixel values, to produce an evenly spaced Fourier volume with  $h$ ,  $k$ , and  $l$  dimensions. The sampling interval for the discretization is calculated from the thickness of the crystal. This Fourier volume is then back-transformed into real space to produce a 3D map for the protein structure.

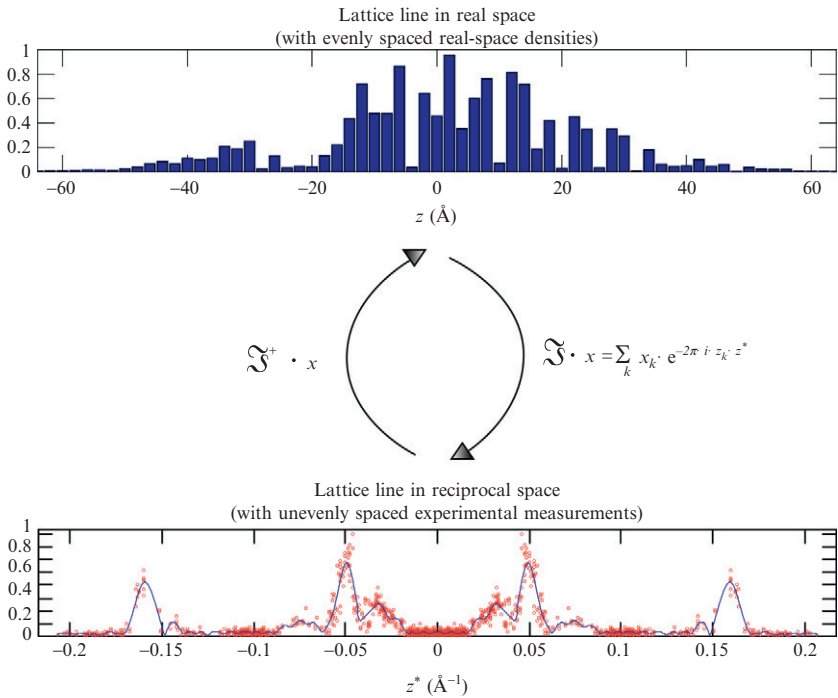
### 2.4.6. Lattice line fitting by matrix transformation and SVD

An alternative approach introduced by Gipson *et al.* (2010) does not make use of the SINC interpolation approach. Instead, Gipson *et al.* attempt for each lattice line to find a linear transformation matrix that most closely can

translate the regularly sampled pixels of the real-space form of that lattice line into the irregularly sampled measured data in Fourier space along that one lattice line, as depicted in Fig. 4.3. This matrix is determined by taking advantage of the finite thickness of the 2D crystal sample. The real-space form of that lattice line is thereby still placed in a reciprocal space concerning its  $h, k$  coordinates, so that this representation is in a mixed space, where the horizontal  $h$  and  $k$  dimensions are reciprocal, while the vertical  $z$  dimension is real. The required transformation matrix for that lattice line then transforms from this regularly sampled hybrid space into the irregularly sampled native Fourier space that contains the experimental measurements. This transformation is defined as

$$\mathfrak{F}_{z, z^*} x_z = \hat{x}_{z^*} \quad (4.2)$$

where  $z$  is referring to the real-space pixel numbers along the real-space version of that lattice line,  $z^*$  is the index referring to the (unevenly sampled) experimental lattice line measurement,  $\mathfrak{F}_{z, z^*}$  is the discrete to



**Figure 4.3** The matrix  $\mathfrak{F}$  accomplishes the linear transform between the regularly spaced real-space representation of a lattice line and the experimentally measured, unevenly spaced lattice line values in reciprocal space.

$z^*$ -sampled Fourier transform matrix,  $x_z$  is an unknown regularly spaced real-space solution, and  $\hat{x}_{z^*}$  is the vector containing the experimentally measured, unevenly sampled Fourier space lattice line values sampled at  $z^*$ .

Once in this form, the standard tools from statistics and linear analysis become available for solving this problem. The standard method currently used to find the least squares solution for this system involves truncated singular value decomposition (SVD) (Hansen, 1987). This process decomposes the Fourier matrix  $\mathfrak{F}_{z, z^*}$  into a quasi-diagonalized form

$$\mathfrak{F}_{z, z^*} = U \cdot S \cdot V^T \quad (4.3)$$

with  $S$  being a diagonal matrix of “singular values” closely related to Eigenvalues, and  $U$  and  $V^T$  being matrices to accommodate the remainder of the linear transformation.

The least squares solution to this problem can then be found by transposing the system, then inverting the nonzero values of  $S$ .

$$\mathfrak{F}_{z, z^*}^+ = V \cdot S^+ \cdot U^T \quad (4.4)$$

This “pseudoinverse” matrix  $\mathfrak{F}_{z, z^*}^+$  then gives the least squares solution to Eq. (4.2), when applied to the list of spots for a given lattice line.

$$x_z = \mathfrak{F}_{z, z^*}^+ \cdot \hat{x}_{z^*} \quad (4.5)$$

The threshold for the smallest singular value not assumed to be zero gives this method a large degree of control over noise sensitivity, resulting in a dynamically adjustable, noise-robust method for 3D structure determination.

Besides finding an optimal solution to this problem, this method provides the “real-space” terms directly, without the need for a sum of SINC functions fit and resampling of the fitted curves. These regularly spaced real-space lattice line values can then either be directly used for further processing or can be Fourier transformed, so that a regularly spaced  $h, k, l$  conventional Fourier space dataset is created.

## 2.5. Molecular replacement and model building

For well-ordered large ( $> 1 \mu\text{m}$ ) 2D crystals, electron diffraction data collection can rapidly produce a dataset of high resolution that does not contain phase information. In this case, and if a homology model is available for the structure to be determined, it is possible to apply the technique of molecular replacement (Vagin and Teplyakov, 1997) that is commonly used for structure determination of 3D crystals in X-ray crystallography. Because the measurement of the unit cell is considered to be of limited precision due

to the uncertainty in the magnification calibration and the flatness of the specimen grid, the unit cell size might have to be slightly adjusted to fit the reference structure. In addition—if the target resolution for the density map is beyond 3 Å—the molecular replacement profits from replacing the atomic scattering factors for X-rays by atomic scattering factors for electrons (Grigorieff *et al.*, 1996). If no suitable model for molecular replacement is available, the missing phase data have to be taken from the real-space image data. After processing and discretization of the image data, these can be combined with the discretized diffraction data to yield a high-resolution density map (Henderson *et al.*, 1990).

### 3. IMAGE PROCESSING WITH 2DX

In 2dx, a project is created and managed through the program `2dx_merge` (Gipson *et al.*, 2007a). The recorded real-space images are aggregated by the program and grouped in directories according to their tilt angles. The program `2dx_image` is used for the processing of an individual image.

#### 3.1. Project initialization in 2dx

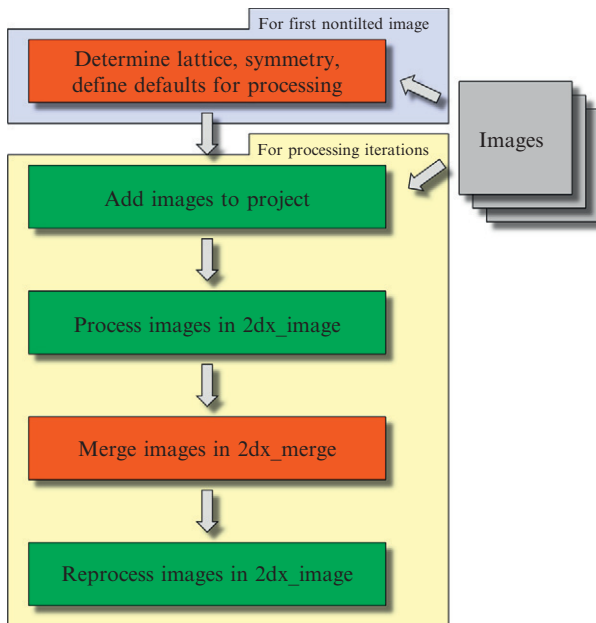
When starting `2dx_merge` the user is requested to select a project directory. Just creating a new folder starts a new project. `2dx_merge` will then create the “`2dx_master.cfg`” configuration file and the “merge” directory with its files in the project folder. After the initialization `2dx_merge` will open a GUI window that shows different panels to display data, or provide specific functions and routines.

`2dx_merge` manages the recorded real-space images of 2D crystals and displays them in the central panel of the GUI. At first, images have to be imported via the “File” pull-down menu option “Import Images...”, which lets the user add a selection of images. Image files can be in MRC or TIFF format. Information embedded in the file names of the image selection can be extracted in the import dialog by a regular expression in the “Translation” field. The regular expression is used to fill the following fields: the name of the protein, the tilt angle category, the micrograph number, and its subimage identifier. Each field is depicted by parenthesis in the regular expression. The result of the file name extraction is listed on the fly in a table of the dialog window, in which the first column can be used to set default values. The specified naming scheme is then used by `2dx_merge` to organize the image files into separate directories for each tilt category and subdirectories for each image in the project folder. Upon the import the initialization script of `2dx_image` is run on every single image to prepare it for processing. The unit cell size of the unknown crystal is

determined by processing the first individual image and can then be passed from the individual image configuration file to the “2dx\_master.cfg” configuration file.

### 3.2. Processing of an individual image in 2dx

Individual processing of a real-space image is done by 2dx\_image (Gipson *et al.*, 2007b). 2dx\_image is called by 2dx\_merge when double-clicking a listed image in the central table of the GUI. 2dx\_image can also be used as a stand-alone program. The general approach is to first process one image in 2dx\_image, to determine the lattice and other project-wide parameters, and to search for optimal processing parameters for this type of images. With these, 2dx\_image then should be able to process the majority of the following images automatically. After merging in 2dx\_merge, the obtained merged 2D or 3D dataset can be used to refine the processing with 2dx\_image, as depicted in Fig. 4.4.



**Figure 4.4** Flow chart of the image processing in 2dx. The manual steps are highlighted in light gray. The first image is processed manually, during which project-wide parameters (e.g., lattice dimensions, symmetry) and optimal processing settings are determined. 2dx can then process the remaining images largely automatically. Merging remains a manual task, after which the reconstruction can iteratively be used to refine the individual image processing.

The GUI of `2dx_image` consists of different panels displaying data and parameters as well as the image-processing routines. In the “Processing Data” panel parameters can be set, as for example the magnification of the microscope, but most values are calculated through the scripts held in the “Standard Scripts.” Each script represents an individual processing routine, which not only calls the corresponding MRC and other stand-alone programs but also determines the parameters and data that are displayed in the GUI. The scripts order as listed in the panel defines the complete workflow of 2D crystal image processing: initialization, Fourier transformation, defocus and astigmatism calculation, tilt geometry determination, automatic lattice determination, lattice refinement, spot list determination, unbending, CTF correction, and projection map generation. For each script a description can be displayed by pressing the “Help” button. An image can be processed automatically by selecting all the standard scripts and pressing the “Run” button.

Most of the scripts call corresponding MRC programs for the actual image processing, and some additional functions were added. The defocus, astigmatism, and tilt geometry are calculated using `ctffind2` (Grigorieff, 1998). The reciprocal crystal lattice is determined automatically: for images of tilted crystals, the approximate tilt geometry is first determined from the defocus gradient across the image. A hypothetical test lattice is then created from the known real-space dimensions of the crystal, which is then distorted according to the obtained approximate tilt geometry information, and translated into the reciprocal space. These test lattices are then rotated around the  $z$  direction in the image by  $360^\circ$ , while each time taking the tilt geometry distortions into account, and while slightly varying lattice parameters, tilt geometry, and magnification, to search for the best fit with the measured lattice peaks in the calculated Fourier transformation of the image of the tilted crystals, until the best fitting lattice is found (Zeng *et al.*, 2007a). For the determination of several lattices in the case of multiple overlaying 2D crystals, the above algorithm is used to first find the strongest lattice. The diffraction peaks that correspond to that lattice are then removed from the list of diffraction spots, and the algorithm is iteratively reapplied to the remaining diffraction spots, until all possible lattices are found.

The lattice is refined using the MRC program `mmlatref`, but can also be refined manually by selecting diffraction spots in the Fourier transform displayed by the integrated full-screen image browser. The user is assisted in the spot selection by a peak search in the surroundings of a mouse double-click. The unbending of the 2D crystal image is split into multiple rounds. A first reference is created by using a one-pixel diameter Fourier mask. The following unbending rounds with wider Fourier masks then retrieve the structure’s underlying signal. Iterative refinement of the reference from unbent images allows refining the unbending protocol, while each time the raw original image is unbent only once. Thanks to predictive lattice node tracking implemented in the MRC program team *quadserch* and

*ccumbend*, unit cells can be localized with good precision, and this information can be used to extract the unit cell image stacks for the single-particle maximum likelihood algorithm that is available in 2dx (Zeng *et al.*, 2007b).

### 3.3. Merging in 2dx

2dx\_merge not only assists in the logistics of managing the processing workflow but also performs the actual 3D merging of the image data. For this, 2dx\_merge offers a 2D mode, where evaluated image data (amplitudes and phases) from nontilted crystals are merged into a final 2D projection map. 2dx\_merge also offers a 3D mode, where data from tilted images can be included. This is done by first merging data from crystals at low tilt angles to the nontilted dataset, and later successively adding data from higher tilt angles. While 2dx\_image allows optionally fully automated image processing, 2dx\_merge so far only assists in the manual merging process, following the MRC program philosophy described above. Once a 3D reconstruction is produced, this can then be used to calculate synthetic reference maps for a certain tilt geometry, to be used by 2dx\_image as ideal reference (using the MRC program *maketran*) to perform an optimized unbending. This iteratively can lead to a refined 3D reconstruction, which in turn can benefit a better unbending. In its current state (as of summer 2010), 2dx does not assist in the processing of electron diffraction patterns. 2dx is freely available (Gnu public license) at <http://2dx.org>.

## 4. ELECTRON DIFFRACTION PROCESSING WITH XDP

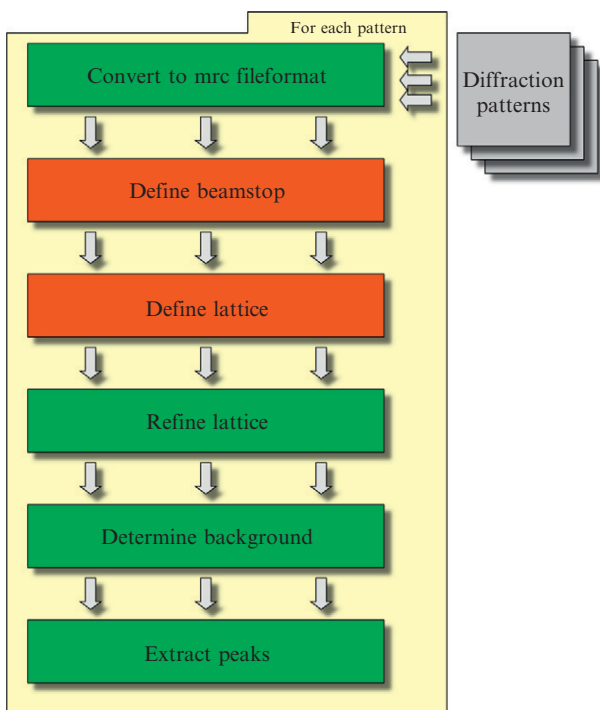
### 4.1. Project initialization in XDP

The directory setup for a XDP project is flexible and can be chosen as the user wishes. Nevertheless it is advantageous to group diffraction patterns with identical nominal tilt angles into separate directories, as it reduces cluttering of the main project directory and simplifies the organization during the merging steps. Before using diffraction patterns in XDP, they have to be converted to data only format using *em2em* or a similar tool and renamed. XDP uses as image name a two-letter prefix to identify a project followed by a six-digit number. The first two digits are defined by the nominal tilt angle and the remaining four digits can be freely chosen. Running the script *padd* generates a parameter file for the image, calculates a histogram, prepares the diffraction pattern to be used in XDP, and converts it to the *mrc* file format. At this stage all the necessary files are created for the diffraction to be processed within the XDP GUI.

## 4.2. Processing of an individual electron diffraction pattern in XDP

The general workflow to be used in XDP is illustrated in Fig. 4.5. To process the diffraction data, XDP is started with the `-c` option followed by the two-letter project prefix. Diffraction patterns can be loaded using the screen menu; the starting image being the cut file. In XDP, determination of the beamstop is done manually by selecting four points defining a quadrangle covering the stem of the beamstop (Menu Box). The part of the beamstop covering the center of the beam is ignored as the center region of the diffraction pattern is excluded from processing later on during processing anyway to avoid over saturated areas of the CCD.

The lattice is determined in XDP by manually picking and indexing one point of the lattice in each quadrant of the diffraction pattern (Menu index). The user has to take care that the overall sum of  $h$  and  $k$  indexes is not equal to 0, because otherwise the lattice determination in the subsequent step will fail.



**Figure 4.5** Flow chart of the diffraction processing in XDP. The manual steps performed in the image viewer are highlighted in light gray. Each of the processing steps has to be run for every diffraction pattern individually (indicated by multiple arrows between the processing steps).

In the next step running `diffint` refines the user-defined lattice. At this point the lattice information is still saved in a separate file. To add it to the extra header section of the `mrc` image file, `tomx` is run. Running `back` generates the mean radial density profile, which is used for a global background correction of the diffraction pattern. As soon as the global background correction is done, the background corrected file (`cmpct` file) is available in the screen menu.

`pick` performs the diffraction peak integration and the correction of the diffraction intensities for the local background contribution. The average background for a peak is determined in the area around the peak.

### 4.3. Merging in XDP

In XDP, diffraction patterns are merged starting with the best, untilted diffraction pattern as reference. The other patterns are compared to the reference, scaled, and gradually merged into the combined dataset going from low tilt angles to high tilt angles. Diffraction patterns above  $45^\circ$  should only be merged into the dataset after the initial merge is completely refined. Initially the data should only be merged at a relatively low resolution ( $\sim 5$  Å). The resolution can then be gradually increased during the refinement cycles. XDP uses the two programs `mergediff` and `syncfit` of the MRC program collection to perform the merging. `mergediff` is responsible for symmetrization, scaling, merging, and refinement and `syncfit` performs the actual lattice line fitting using a sum of `SINC` functions and discretization of the fitted lattice line. To simplify the use of these two programs, XDP provides several command line scripts acting as frontends to `mergediff` and `syncfit`.

The merging is performed in a separate directory, which is subdivided into directories for the individual tilt angles. For every dataset to be merged, the files containing the lattice vectors, the parameters, and the integrated data points have to be copied to the corresponding merge subdirectory. The script `mergint` scans the directory of an individual tilt angle and generates a list with the patterns to be merged. `mergset` can then be used to set some basic merging parameters like the maximal resolution. The following steps in processing need some template files, which can be fetched with `mgsetdif_XX` (`mgsetdif_XXT` for tilted patterns) where `XX` is the project prefix. For a new project the template files have to be adjusted as they contain project specific parameters as for example the unit cell size. `mgsetdif_XX/T` provides additional sets of templates for tilt geometry and anisotropic scaling refinement. Merging and scaling within the directory of one tilt angle can be performed using `difcomp`. This script also allows refining the tilt geometry for nominally untilted images. Once all the tilt angle directories are processed, the overall merging can be performed by `difinit`. This is followed by `syncfit-1st.com`, which fits the lattice lines based on the merged data and discretizes the fit based on the entered sample thickness. The discretized

data, which is stored in text format, can be converted to the mtz format by `f2mtz` and be used for molecular replacement and further processing in CCP4 ([Collaborative Computational Project, 1994](#)).

## 5. ELECTRON DIFFRACTION PROCESSING IN IPLT

In IPLT the project handling is done in the IPLT Diff Manager `gip1t_diff_manager`. It serves as a platform for creation of new projects, adding and processing new diffraction data, verifying processed data, and organizing the data merging. Complementary to the Diff Manager, there is a command line interface called `iplt_diff`, which can be used to process data without having to start a GUI.

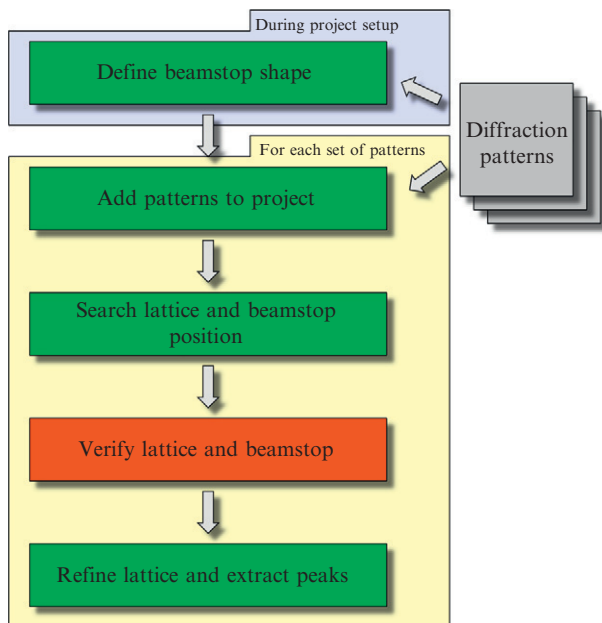
### 5.1. Project Initialization in IPLT

The general workflow for diffraction pattern processing in IPLT is depicted in [Fig. 4.6](#). Starting the Diff Manager and selecting “New Project” creates a new project. The user has to enter the main directory for the project, a project prefix, the file format of the recorded diffraction patterns, and the unit cell and spacegroup parameters. The Diff Manager will create the project directory, the main settings file (`project.xml`), and open the central processing GUI. New diffraction patterns are added using the GUI and will be placed into separate directories within the project directory. New diffraction patterns can be encoded in any of the many formats readable by IPLT, but the `dm3`, `tiff`, and `mrc` formats are probably the most commonly used.

From one of the diffraction patterns the general beamstop shape has to be determined using the “Define Beamstop Mask” tool within the Diff Manager. This information is used later on for the automatic determination of the beamstop position.

### 5.2. Processing of an individual electron diffraction pattern in IPLT

The first step in the processing of individual diffraction patterns is the determination of the beamstop position and lattice for all diffraction patterns. In case that the diffraction pattern is distorted—for example, due to lens distortions encountered with an off-axis positioned CCD—IPLT provides also an option to fit the parameters for a barrel or spiral distortion of the lattice. The automatic lattice determination also includes a global fit of the background produced by inelastically scattered electrons. This fit then allows to extrapolate the position of the highest background intensity, which coincides with the origin of the lattice. This fit, however, is not used for the background correction for the diffraction peak intensities that



**Figure 4.6** Flow chart of the diffraction processing in IPLT. The manual steps performed in the image viewer are highlighted in light gray. IPLT can process a whole set of patterns in one go (indicated by single arrows between steps). The beamstop shape has to be determined only once during the project setup.

are later evaluated, because the model for the background is too coarse to account for the exact background variations, which are influenced by many effects like sample contamination, uneven illumination conditions, lens distortions, and electron scattering of the amorphous ice.

Once the lattice and the beamstop position are determined for a pattern, they can be visually checked and if necessary corrected in the Diff Manager using the “Verify Beamstops and Lattices” option. The rest of the processing steps do not need any user interaction and can be performed by selecting the “Refine Lattices and Extract Data” option. The automatic steps consist of a refinement of the lattice vectors, the peak extraction and correction for background contribution, and the determination of the tilt geometry.

### 5.3. Merging in IPLT

Automatic merging using the default parameters and directory setup can be done by selecting the “Merge” option in the Project tab. More customized merging can be performed by running the individual steps for merging manually in the merging tab of the Diffraction Manager.

Alternatively—if no GUI is desired—the identical merging steps can also be performed using the `iplt_diff_merge` command line executable.

Before the datasets are merged, they are symmetrized according to the symmetry defined during project setup. Following symmetrization the datasets are preliminary merged to provide a first reference for scaling. They are then scaled according to their weighted resolution binned average intensity to bring them to a common relative level. The symmetrization, premerging, and scaling steps can be performed by selecting “Scale” in the merging tab of the Diffraction Manager.

Once the datasets are scaled, they can be subjected to a refinement of the tilt geometry by selecting “Refine” in the merging tab. The tilt geometry is refined by varying tilt angle and tilt axis position slightly from the determined values and minimizing the difference between the dataset of the currently refined image and the overall merged dataset, which serves as reference. The search range for the tilt geometry is defined for its normal vector on the crystal plane. If parameterized in tilt angle and tilt axis, the search range for the tilt axis is a function of the tilt angle: for a pattern from an untilted specimen the tilt axis position is searched in a range of  $180^\circ$ , whereas for a pattern from a highly tilted sample only a small deviation of the determined tilt axis position is allowed.

After refinement of the tilt geometry, lattice lines can be fitted and a discretized dataset can be generated by sampling the lattice line fit at equally spaced intervals. Selecting the “Lattice Line Fit” option in the merging tab runs both steps together. IPLT uses a sum of squared *SINC* functions to fit the intensity data. The interval spacing for discretization is determined by the thickness of the original sample. Whereas the original datapoints can be negative if the background intensity is higher than the peak intensity, the fitted line will always have a positive or zero value. To improve the reliability of the fitting errors, IPLT provides a bootstrapping mode where subsets of the data are fitted several times and the fitting errors are calculated from the difference of the individual fits. Using the lattice line fitted data as reference, the sequence of scaling, tilt geometry refinement, and lattice line fitting can be repeated in an iterative way to further refine the merged dataset.

Data files generated by IPLT can directly be used for molecular replacement and further processing in CCP4. No conversion is necessary.

---



## 6. CONCLUSIONS

The last decade has seen the consolidation of electron crystallography as a reliable alternative to X-ray crystallography and NMR for structural studies on membrane proteins, as it provides unique insights into their molecular arrangement in their lipid environment.

In practice, the realization of this potential advantage is not a trivial matter. Besides the experimental complexity, final success in structure determination is critically dependent on the image processing of the collected data. The involved image processing builds from concepts stemming from X-ray crystallography and electron microscopy, but new algorithms are required to account for the particular geometry, distortion treatment, and the integration of amplitude and phase information.

Earliest efforts to develop these specific software tools have resulted in the MRC software suite for electron crystallography. Initially conceived as a multitude of Fortran programs linked by a common file format, they were the first specific tool to be used by the community. The feedback provided by this extensive testing led to a constant improvement and eventually to the consolidation of a reliable processing pipeline.

The growing acceptance achieved by electron crystallography has driven the development of new software resources that make these already well-established algorithms available to a wider spectrum of users. Recently developed and freely distributed software packages like 2dx and IPLT envision the integration of all the different aspects of the data processing, offering the user a unified interface and reducing the requirement for computational expertise for the users. As these packages explicitly aim at attending the needs of a constantly increasing community of users, developments like the creation of graphical interfaces and production of exhaustive documentation and tutorials have been accompanied with an increase in direct contact with the users. The method of electron crystallography and image processing by the several programs including MRC, 2dx, and IPLT is now taught in the biannual series of workshops in electron crystallography (see <http://2dx.org/workshop>), and further support is ensured by the IPLT.org and 2dx.org web sites maintained at the University of Basel.

This general orientation toward an increased user-friendliness of already established computational techniques has not prevented a parallel effort in developing new algorithms that push the range of approachable problems beyond the current limitations. The maximum likelihood single-particle approach has been introduced as a powerful alternative to deal with images of 2D crystals, which is particularly strong in the case of severe crystal distortions. The recently developed method of Projective Constraint Optimization can compensate for the inherent geometrical constraints that used to limit the achievable vertical resolution of the technique, by filling data into the missing cone (Gipson *et al.*, 2010).

A further development direction is the cross-linking of the different software solutions. Experience has already identified some particular aspects of the image processing that are better resolved by some programs than others. Even if the attained file format compatibility already allows for simultaneous use of different approaches, a complete integration of the different software approaches will be advantageous. Nevertheless, the

available electron crystallography software resources already constitute a useful toolbox for the full spectrum of potential users. The workflow guidance provided to users by software systems like 2dx gives them a good level of operational confidence in a short learning period. On the same time, experienced users of 2dx have the possibility of accessing the full range of thoroughly tested and optimized algorithms in the MRC software, and can adapt them to their particular needs. Finally, developers have an extended repertoire of tools that supports the creation of new features and conceptual advances. In case of 2dx, this is possible through the simple integration of independently compiled stand-alone programs that are called through the flexible 2dx interface. In case of IPLT, users can contribute to the software at several levels, ranging from high-level python scripts to low-level C++ routines. These new features can in turn get easily integrated into the existing general software frameworks that provide an efficient channel for distribution to the whole community.

In summary, electron crystallography is a mature method (Glaeser *et al.*, 2007) that provides important contributions to our understanding of the structure and function of membrane proteins. Electron crystallography is a powerful tool to study at intermediate resolution the conformational changes that membrane proteins can undergo in the membrane embedded state and to determine at the level of atomic resolution the structure of membrane proteins and their interaction with the lipid bilayers.

## ACKNOWLEDGMENT

This work was partially supported by the Swiss National Science Foundation and by the NIH grant GM084921.

## REFERENCES

- Chiu, P. L., Pagel, M. D., Evans, J., Chou, H. T., Zeng, X., Gipson, B., Stahlberg, H., and Nimigean, C. M. (2007). The structure of the prokaryotic cyclic nucleotide-modulated potassium channel MloK1 at 16 Å resolution. *Structure* **15**, 1053–1064.
- Collaborative Computational Project (1994). The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr.* **50**, 760–763.
- Crowther, R. A., Henderson, R., and Smith, J. M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9–16.
- Downing, K. H., and Hendrickson, F. M. (1999). Performance of a 2k CCD camera designed for electron crystallography at 400 kV. *Ultramicroscopy* **75**, 215–233.
- Gipson, B., Zeng, X., and Stahlberg, H. (2007a). 2dx\_merge: Data management and merging for 2D crystal images. *J. Struct. Biol.* **160**, 375–384.
- Gipson, B., Zeng, X., Zhang, Z., and Stahlberg, H. (2007b). 2dx–User–friendly image processing for 2D crystals. *J. Struct. Biol.* **157**, 64–72.

- Gipson, B., Masiel, D. J., Browing, N. D., Spence, J., Mitsuoka, K., and Stahlberg, H. (2010). Automatic recovery of missing amplitudes and phases in tilt-limited electron crystallography of 2D crystals (manuscript in preparation).
- Glaeser, R., Downing, K., DeRosier, D., Chiu, W., and Frank, J. (2007). *Electron Crystallography of Biological Macromolecules*. Oxford University Press, USA.
- Gonen, T., Sliz, P., Kistler, J., Cheng, Y., and Walz, T. (2004). Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429**, 193–197.
- Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C., and Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438**, 633–638.
- Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. *J. Mol. Biol.* **277**, 1033–1046.
- Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M., and Henderson, R. (1996). Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421.
- Hansen, P. C. (1987). The truncated SVD as a method for regularization. *BIT Arch.* **27**, 534–553.
- Hardt, S., Wang, B., and Schmid, M. F. (1996). A brief description of I.C.E.: The integrated crystallographic environment. *J. Struct. Biol.* **116**, 68–70.
- Henderson, R., and Unwin, P. N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32.
- Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J., and Zemlin, F. (1986). Structure of purple membrane from *Halobacterium halobium*: Recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* **19**, 147–178.
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929.
- Hiroaki, Y., Tani, K., Kamegawa, A., Gyobu, N., Nishikawa, K., Suzuki, H., Walz, T., Sasaki, S., Mitsuoka, K., Kimura, K., Mizoguchi, A., and Fujiyoshi, Y. (2006). Implications of the aquaporin-4 structure on array formation and cell adhesion. *J. Mol. Biol.* **355**, 628–639.
- Hovmöller, S. (1992). CRISP: Crystallographic image processing on a personal computer. *Ultramicroscopy* **41**, 121–135.
- Jap, B. K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aebi, U., and Engel, A. (1992). 2D crystallization: From art to science. *Ultramicroscopy* **46**, 45–84.
- Kühlbrandt, W. (1992). Two-dimensional crystallization of membrane proteins. *Q. Rev. Biophys.* **25**, 1–49.
- Kühlbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614–621.
- Kukulski, W., Schenk, A. D., Johanson, U., Braun, T., de Groot, B. L., Fotiadis, D., Kjellbom, P., and Engel, A. (2005). The 5 Å structure of heterologously expressed plant aquaporin SoPIP2;1. *J. Mol. Biol.* **350**, 611–616.
- Kulik, V. (2004). Structure of bovine liver catalase solved by electron diffraction on multilayered crystals. Department of Physics, University of Osnabrueck, Osnabrueck, pp. 159.
- Landsberg, M. J., and Hankamer, B. (2007). Symmetry: A guide to its application in 2D electron crystallography. *J. Struct. Biol.* **160**, 332–343.
- Mariani, V. (2009). Transfer of tilted sample information in transmission electron microscopy. Biozentrum, University of Basel, Basel.
- Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y., and Fujiyoshi, Y. (1999). The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: Implication of the charge distribution. *J. Mol. Biol.* **286**, 861–882.

- Morgan, D. G., Ramasse, Q. M., and Browning, N. D. (2009). Application of two-dimensional crystallography and image processing to atomic resolution Z-contrast images. *J. Electron Microsc. (Tokyo)* **58**, 223–244.
- Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., and Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605.
- Philippson, A., Schenk, A. D., Stahlberg, H., and Engel, A. (2003). IPLT-image processing library and toolkit for the electron microscopy community. *J. Struct. Biol.* **144**, 4–12.
- Philippson, A., Engel, H. A., and Engel, A. (2007a). The contrast-imaging function for tilted specimens. *Ultramicroscopy* **107**, 202–212.
- Philippson, A., Schenk, A. D., Signorell, G. A., Mariani, V., Berneche, S., and Engel, A. (2007b). Collaborative EM image processing with the IPLT image processing library and toolbox. *J. Struct. Biol.* **157**, 28–37.
- Schenk, A. D., Werten, P. J., Scheuring, S., de Groot, B. L., Müller, S. A., Stahlberg, H., Philippson, A., and Engel, A. (2005). The 4.5 Å structure of human AQP2. *J. Mol. Biol.* **350**, 278–289.
- Schmid, M. F., Dargahi, R., and Tam, M. W. (1993). SPECTRA: A system for processing electron images of crystals. *Ultramicroscopy* **48**, 251–264.
- Sorzano, C. O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H., Carazo, J. M., and Pascual-Montano, A. (2004). XMIPP: A new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204.
- Unwin, P. N., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* **94**, 425–440.
- Vagin, A., and Teplyakov, A. (1997). MOLREP: An automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025.
- Valpuesta, J. M., Carrascosa, J. L., and Henderson, R. (1994). Analysis of electron microscope images and electron diffraction patterns of thin crystals of phi 29 connectors in ice. *J. Mol. Biol.* **240**, 281–287.
- Vonck, J., von Nidda, T. K., Meier, T., Matthey, U., Mills, D. J., Kühlbrandt, W., and Dimroth, P. (2002). Molecular architecture of the undecameric rotor of a bacterial Na(+)-ATP synthase. *J. Mol. Biol.* **321**, 307–316.
- Wilson, A. (1942). Determination of absolute from relative X-ray intensity. *Nature* **150**, 152.
- Yeates, T. O., and Fam, B. C. (1999). Protein crystals and their evil twins. *Structure* **7**, R25–R29.
- Zeng, X., Gipson, B., Zheng, Z. Y., Renault, L., and Stahlberg, H. (2007a). Automatic lattice determination for two-dimensional crystal images. *J. Struct. Biol.* **160**, 353–361.
- Zeng, X., Stahlberg, H., and Grigorieff, N. (2007b). A maximum-likelihood approach to two-dimensional crystals. *J. Struct. Biol.* **160**, 362–374.