High-throughput high-resolution cryo-electron crystallography

Inauguraldissertation

Zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch – Naturwissenschaftlichen Fakultät
der Universität Basel

von

Sebastian Scherer

aus Basel, Schweiz

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Henning Stahlberg, Fakultätsverantwortlicher

Prof. Dr. Volker Roth, Korreferent

Basel, den 24. März 2015

Prof. Dr. Jörg Schibler (Dekan)

Summary

High-resolution structures of membrane and soluble proteins can be obtained by cryo-electron crystallography, given highly-ordered cryo-preparations of perfectly flat 2D crystals are available. Studies of membrane proteins, which are embedded into a lipid membrane, mimicking the native cell membrane, are of particular biological interest. However there are multiple reasons why electron crystallography is far from being a mainstream protein structure determination technique. In this thesis we address three major difficulties of electron crystallography: (i) resolution loss due to not perfectly flat crystals, (ii) reliable high-throughput automatic image processing and (iii) correction of electron beam-induced motion of the sample.

The conventional electron crystallography image processing procedure assumes perfectly flat 2D crystals, which are almost impossible to obtain. Our new processing approach, described in Chapter 2, weakens this assumption tremendously. Traditional processing assigns the same tilt geometry to all proteins of one 2D crystal. Thus local tilt geometry variations, due to not perfectly flat crystals, are neglected. We developed an algorithm that optimizes the tilt geometry of each protein separately, while exploiting the correlation between neighboring proteins. The new method proves the feasibility of this approach, improves the achieved resolution and opens the doors to new studies, i.e. structural studies of membrane proteins embedded into lipid vesicles.

Recently a new generation of digital detectors tremendously changed the cryo-electron microscopy field. Beside a significantly increased signal-to-noise ratio these detectors record dose-fractionated movies of the sample under the electron beam. Previous cameras only recorded one image instead. This new exposure mode enables the computational correction for beam-induced sample movements. In Chapter 3 we describe a new "real-time" automation pipeline for electron crystallography using direct electron detectors. The novel pipeline automatically corrects for homogeneous sample drift on frame level and processes the acquired images automatically. Both, the time-to-solution and the quality of the obtained 3D reconstructions are significantly improved.

Heterogeneous beam-induced sample movements are the most severe resolution-limiting factor in modern cryo-electron crystallography. In Chapter 4 we present an algorithm, termed movie-mode unbending, which corrects for inhomogeneous beam-induced sample drift. In contrast to the previous homogeneous drift-correction, the novel algorithm can correct for locally varying beam-induced sample motion. This novel approach significantly increases the resolution of electron cryo-crystallographic studies recorded on the latest detectors.

Chapter 5 covers multiple successful applications of the methods introduced above. For instance the real-time drift-correction module of the developed automation pipeline was used for multiple near-atomic resolution single particle and helical image processing projects. The high-throughput crystal image processing was applied to different kinds of 2D crystals and enabled the qualitative assessment of different sample preparation methods. Additionally we describe the implementation of a high-throughput single particle automation pipeline, which will enable the generation of near-atomic resolution single particle cryo-electron microscopy density maps on a daily basis.

Contents

1 Introduction

Membrane proteins are nanoparticles that are embedded (integral) into or attached (peripheral) to a cell membrane. They are present in all living organisms, where they perform vital functions such as nutrient uptake, drug efflux or signal transduction [Abeyrathne *et al.* 2010]. Many disease-associated protein mutations causing miss-folding and miss-functioning, are observed in membrane proteins [Sanders & Myers 2004]. Detailed understanding of membrane protein structures at atomic level is indispensable for the functional understanding and the design of new pharmaceutical drugs.

Contents

1.1 Structure and function of membrane proteins

Amino acid residues are the building blocks of proteins. They are composed of an *amine* (-NH2) and a *carboxylic acid* (-COOH) functional group, plus an additional side-chain that is specific to each amino acid. There are 20 proteinogenic amino acids that are encoded via triplet codons in the genetic code, i.e. the *"standard" amino acids*. Additionally there are three *"non-standard" amino acids* that are not directly DNA-coded, but they are present in many non-eukaryotes and most eukaryotes.

Proteins are formed of one or multiple linear chains of amino acids, i.e. *polypeptide chain(s)*. The individual amino acids are bound by so-called *peptide bonds* between the negatively charged carboxyl group and the positively charged amino group of two neighboring amino acid residues. This linear sequence of amino acids is referred as the protein's *primary structure*. Although the primary structure is unique for each protein its 3D structure is not described by the primary structure.

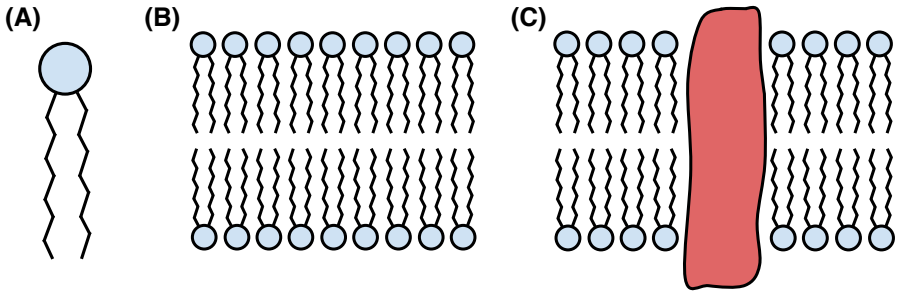

Figure 1.1: Membrane architecture: (A) Phospholipid, (B) Phospholipids arranged in a lipid bilayer, (C) integral membrane protein embedded into a lipid bilayer.

The *secondary protein structure* relates to highly regular structural elements, but does not reveal three-dimensional atom positions. The main secondary structure elements are *alpha helices* and *beta sheets*, whose formation is caused by the energetically favorable state of these specific hydrogen-bonding patterns of the peptide backbone of the protein.

The three-dimensional structure of a protein is described by the three-dimensional space coordinates of all atoms of the protein complex, the so-called *tertiary structure*. In general determining the tertiary structure of unknown biological molecules is one of the major aims of structural biology.

The three-dimensional structure of large multi-subunit protein complexes is often referred as its *quaternary structure*. For instance a complex formed of two polypeptide chains is called *dimer*, whereas complexes consisting of three subunits are called *trimers*. In general these constructs are called *multimers*. Additionally they often feature an internal symmetry. In practice the known symmetry of a protein complex turns out to be useful for three-dimensional image processing.

The aqueous compartments of a cell have to be separated from its surrounding by means of a barrier. In most forms of life a *lipid bilayer* acts as this separator. Natural lipid bilayers are composed of *phospholipids*, which feature a hydrophilic head and two hydrophobic tails (Figure 1.1A). Due to the hydrophobic effect, phospholipids self-assemble into a bilayer, where the hydrophilic heads point outwards to the water on both sides and the hydrophobic tails point into the center of the bilayer (Figure 1.1B). Lipid bilayers are impermeable to most water-soluble molecules and particularly to ions. Important functions such as salt concentration regulation or cell signaling are performed by membrane proteins, which span the entire bilayer and thus bridge between the “inside” and “outside” of the cell (Figure 1.1C).

1.2 Structural analysis of proteins

The three-dimensional structure of a protein is described by the spatial position of all its atoms. Typically these positions are stored in a so-called *PDB-file*, whose deposition in a centralized archive (*Protein Data Bank*[1]*;* PDB) is strongly encouraged. By the end of the year 2014, 105'455 released structures had been deposited in the PDB. Approximately 0.1% of all deposited structures describe membrane proteins.

In general there are different experimental methods for three-dimensional structure determination of (membrane) proteins: X-ray crystallography (XRD), nuclear magnetic resonance spectroscopy (NMR) and cryo-electron microscopy (cryo-EM). Each of these techniques follows different principles and features different requirements, benefits and limitations. The prerequisite that a suitable amount of purified protein has to be available is shared by all three methods, which are briefly discussed below:

X-ray crystallography Approximately 90% of all deposited protein structures were determined by X-ray crystallography. Holistically X-ray crystallography studies diffraction patterns of three-dimensional protein crystals. The limiting factor of X-ray crystallography is the production of three-dimensional repeating protein assemblies, the so-called *3D crystals*. In practice a lot of different crystallization conditions have to be screened to obtain 3D crystals of sufficient size and quality. Finally a good enough crystal is mounted on a *goniometer*, which is used to position and orient the crystal during the measurement. Then the sample is bombarded with a focused monochromatic beam of X-rays. The repetitive protein assembly of the 3D crystal leads to the formation of *diffraction patterns* featuring regularly spaced reflections. Two-dimensional diffraction patterns of the same crystal at different angles are combined into a three-dimensional spectrum. However diffraction patterns only contain the amplitude information of the structure factors, while their phases remain unmeasured. Thus additional techniques such as *molecular replacement* [Drenth 2007] or *heavy atom methods* [Wilson 1970] have to be used to computationally restore the missing phases to obtain the three-dimensional electron density map.

Nuclear magnetic resonance spectroscopy Roughly 9% of the deposited protein structures were solved by nuclear magnetic resonance spectroscopy. This technique uses the observation that some nuclei (1H, 13C or 15N) feature a nonzero spin to determine protein structures in solution. During the experimental procedure the protein sample is first placed in a strong magnetic field leading to a parallel (respectively anti-parallel) alignment of the nonzero spin nuclei. Subsequently the application of *radio pulses* perturbs the magnetic equilibrium.

[1] www.rcsb.org

The absorption of a radio pulse by a particular nucleus of the protein depends on its chemical environment. From the measured *magnetic resonance spectra* pairwise distance restraints can be formulated. These constraints, combined with well known general protein properties (e.g. bond lengths), are reformulated into an energy minimization problem. Finally numerical algorithms reveal the three-dimensional protein structure at atomic details.

Cryo-electron microscopy In contrast to classical light microscopy, where visible light (*photons*) is used to image a sample, cryo-electron microscopy uses electrons with a much shorter wavelength. This allows imaging of much smaller objects. *Transmission electron microscopy* (TEM) is of particular importance for structural studies of biological macromolecules. Here a high-voltage electron beam exposes a thin biological sample. Most of the electrons pass through the sample without being scattered, but some electrons interact with the probe and they are weakly scattered. After passing through the sample the spatially varying electron beam is magnified by a set of electrostatic objective lenses and finally a two-dimensional image is captured on a specialized detector. As transmission electron microscopy records two-dimensional projections of the sample, advanced computational tools are required to obtain a three-dimensional reconstruction of the protein of interest. As this thesis aims to further improve the computational part of cryo-electron microscopy, a detailed overview of the different three-dimensional image processing approaches is given in the next section.

1.3 Three-dimensional cryo-electron microscopy techniques

A sample embedded into vitrified ice at cryogenic temperature, the use of a transmission electron microscope and the computational three-dimensional reconstruction are fundamental principles shared by all different three-dimensional cryo-electron microscopy techniques. The low temperature is required to minimize beam-induced sample damage. In 1984 the group of Jacques Dubochet recorded images of a biological molecule embedded in a layer of vitrified ice [Adrian *et al.* 1984] for the first time in history. Today vitrified ice is obtained by plunge-freezing the sample into liquid ethane or liquid propane manually or by means of a specialized machine, which controls environment variables such as humidity. A temperature drop-rate of up to 10'000 K/s is required to obtain solidified featureless ice. Formation of hexagonal ice crystals would destroy the protein's fine structural details.

The matter of fact that transmission electron microscopes record noisy planar projections of three-dimensional objects requires advanced reconstruction techniques. The two-dimensional measurements of the electron microscope and the three-dimensional objects are linked by the *central slice theorem* [Mersereau & Oppenheim 1974]. This

theorem states that the Fourier transform of a 2D projection of a 3D object holds the same information as the central 2D cross-section perpendicular to the projection axis of the reciprocal 3D volume. Thus a three-dimensional reconstruction can be obtained by back-projecting all different two-dimensional projections into a 3D volume. Dependent on the target resolution and the particular nature of the biologic sample, different three-dimensional electron microscopy approaches can be used:

Single particle analysis 78% of all deposited electron microscopy maps in the *Electron Microscopy Data Bank*[2] (EMDB) were generated using *single particle analysis techniques* [Frank 1975]. Lower-resolution single particle analysis is usually performed on *negatively stained* samples whereas higher-resolution studies generally analyze samples embedded into vitreous ice. The limitation that ice-embedded protein samples are highly sensitive to radiation damage, requires recording of images by means of a moderate electron dose. Cryo-electron microscopy images of single particle samples feature a high amount of noise. The low signal-to-noise ratio makes image processing in cryo-electron microscopy a challenging process. In general it is required to record a large number of micrographs (typically more than 1000). Each image usually shows hundreds of randomly oriented copies of the target protein. In a first step the imaged particles have to be selected and extracted from the recorded micrographs. The low signal-to-noise ratio of individual proteins can be overcome by combining multiple similar views of the same protein, i.e. *2D class averaging*. The two-dimensional classification approach can also be used to clean the set of picked particles [Scheres 2014b]. The three-dimensional orientation of each particle is initially unknown. By means of comparing model projections of the three-dimensional protein structure approximation with the experimental data, the orientation of individual particles can be determined. As the three-dimensional structure is often unknown too, an iterative process has to be used. Typically the iteration is started with a low-resolution approximation of the protein structure. The resolution of the reconstruction and the accuracy of the angular assignment is increased over multiple rounds of iterative projection-matching and back-projecting.

Recent advances in processing software, such as *maximum likelihood methods* for the analysis of heterogeneous datasets [Scheres 2012b] and novel detecting devices [McMullan *et al.* 2009], significantly improved the resolution achieved by modern cryo-electron microscopy. The first near-atomic resolution single particle reconstructions were obtained from large icosahedral virus samples [Zhang *et al.* 2010]. Recently it was possible to get a three-dimensional reconstruction of a 170kDa membrane protein at 4.5Å resolution [Lu *et al.* 2014b].

[2] http://www.ebi.ac.uk/pdbe/emdb/

Results like this make single particle cryo-microscopy a promising structure determination method for smaller proteins in the future.

Electron tomography and sub-tomogram averaging Cryo-electron tomography offers the unique possibility to study nanomachines in their native environment. In contrast to single particle analysis where each area of the sample is exposed only once, electron tomography distributes a high electron dose over multiple exposures of the same area at different tilt angles. Three-dimensional registration of the different slices produces a 3D reconstruction of the sample. Typically electron tomography is used to obtain lower-resolution reconstructions of entire cellular regions or proteins embedded into their native environment. Similar to single particle analysis, it is possible to extract individual three-dimensional particles featuring a low signal-to-noise ratio. *Sub-tomogram averaging* aligns these three-dimensional particles in an iterative procedure similar to single particle analysis [Förster & Hegerl 2006, Winkler 2007]. So far the highest resolution obtained from aligning three-dimensional particles, which were boxed from their native assembly, was 8.5Å [Schur *et al.* 2013]. Due to future hardware and software developments, sub-tomogram averaging will be a powerful method to study larger membrane proteins in their cellular environment at moderate resolution.

Helical image processing In nature some proteins exist in the form of helical polymers. Helical assembled proteins are well suited for electron microscopy as in most cases, one image contains all different views of the helically arranged protein. Traditionally the *Fourier-Bessel formalism* was used to obtain three-dimensional reconstructions from mostly perfectly ordered straight filaments [Lepault 1985, Diaz *et al.* 2010]. Over the past decade, a single particle reconstruction technique was established for helical samples [Egelman 2007]. Thus segments of the helical polymers are extracted from the image and their orientations are iteratively refined similar to classical single particle analysis. Based on the latest generation of detectors, near-atomic resolution maps of helical assembled proteins are obtained routinely today.

2D crystallography Similar to helical image processing, 2D crystallography uses the particular way how proteins are assembled in the sample. For electron crystallography the protein of interest has to be crystallized in 2D first. Thin well-ordered crystalline sheets, vesicles or tubes are then imaged at various tilt angles. Due to the limited dose-resistance of small proteins, only one image per crystal can be recorded. Reciprocal space processing approaches, exploiting the regular arrangement of proteins, are used to obtain a high-resolution projection map of each crystal. Finally projection maps of different crystals at various viewing angles are combined to obtain one three-dimensional reconstruction. As this thesis addresses the improvement of image processing methods for 2D

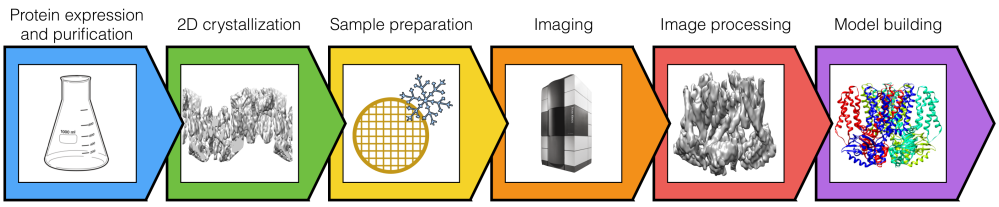

Figure 1.2: Electron crystallography workflow.

crystals, a detailed introduction into cryo-electron crystallography is given in the next section.

1.4 Electron crystallography

In 1975 Richard Henderson and Nigel Unwin reported the first three-dimensional structure of a membrane protein in history [Henderson 1977]. They analyzed two-dimensional crystalline patches of *Bacteriorodopsin* by means of electron crystallography techniques. Since then multiple membrane protein structures have been determined by this technique. So far the highest-resolution was obtained for *AQP0* by Gonen et al. in 2005 [Gonen *et al.* 2005]. Thus electron crystallography is still the cryo-electron microscopy technique holding the resolution record. There are two different flavors of electron crystallography: direct imaging and electron diffraction. As indicated by its names, direct imaging records high-resolution images of the crystals, whereas electron crystallography records diffraction patterns of 2D crystals, similar to X-ray diffraction of 3D crystals. Diffraction patterns of 2D crystals are recorded on an electron microscope by removing one objective lens. In general electron diffraction results in higher-resolution maps for large 2D crystals when compared to imaging. On the other hand, imaging of 2D crystals obtains structural insights for 2D crystals that are not suited for electron diffraction due to their limited size. This thesis focuses on processing of two-dimensional crystal images, as in practice most 2D membrane protein crystals are not large enough for electron diffraction.

Holistically the electron crystallography pipeline (Figure 1.2) includes the following major steps: (i) protein expression and purification, (ii) two-dimensional crystallization, (iii) sample preparation, (iv) data collection, (v) image processing and (vi) model building. Major aim of this thesis is to improve the applied image processing algorithm and to increase the efficiency and throughput of data acquisition.

1.4.1 Protein expression and purification

Obtaining a sufficient amount of pure and homogeneous solution of the protein of interest is the first stage of each cryo-electron crystallography project. Membrane proteins can either be extracted from native tissues or expressed in a prokaryotic or eukaryotic expression system. Since membrane proteins are embedded in a lipid bilayer, detergents are required to extract the proteins from the membranes. The choice of the appropriate detergent is crucial for the success of the later steps, such as 2D crystallization [Abeyrathne *et al.* 2010]. Then, the detergent-solubilized proteins have to be purified, prior to crystallization.

1.4.2 2D crystallization

The fundamental goal of the two-dimensional crystallization step is to reconstitute the expressed and purified proteins into large flat lipid membrane crystals. Therefore the solubilized membrane proteins in detergent are mixed with detergent *solubilized lipids* at a given *lipid-to-protein ratio*. Crystal formation is now induced by constantly reducing the detergent concentration of the ternary mixture by means of *dilution* [Rémigy *et al.* 2003], *dialysis* [Jap *et al.* 1992] or addition of *Biobeads* [Rigaud *et al.* 1997] respectively *cyclodextrin* [Signorell *et al.* 2007]. During detergent removal the hydrophobic trans-membrane part of the proteins favors to be surrounded by lipids rather than water. On the other hand the lipids favor to build clusters to "protect" their hydrophobic sites. The combination of both preferential behaviors naturally triggers the formation of two-dimensional lipid protein crystals.

In general successful two-dimensional crystallization depends on various parameters such as choice of lipids, lipid-to-protein ratio, choice of detergents, buffer conditions, possible additives, temperature and method of detergent removal. Carefully optimizing all these factors makes the successful growing of large 2D crystals a time-consuming task.

1.4.3 Sample preparation

Once large enough two-dimensional crystals are available, they have to be prepared for high-resolution image acquisition in a transmission cryo-electron microscope. To prevent the sample from being dehydrated in the electron microscope's vacuum and to reduce electron beam-induced damage, the crystals are imaged at cryogenic temperatures. Therefore the solution containing the 2D crystals is pipetted onto an electron microscopy grid first. In general there are two different sample preparation methods for 2D crystals: plunge freezing and sugar embedding. For plunge freezing,

the hydrated sample is plunged into liquid ethane at liquid nitrogen temperature in order to obtain 2D crystals embedded in vitrified ice. During sugar embedding the water molecules are replaced by trehalose prior to freezing.

1.4.4 Data acquisition

The prepared samples are then transferred into the electron microscope and imaged at cryogenic temperature to minimize beam damage. Typically the microscope is operated in low-dose mode to reduce the electron bombardment of the sample. In a first step the electron microscope is operated at a low magnification to search for good exposure locations in terms of crystal size, ice thickness and crystal arrangement. The final exposure is acquired at a higher magnification with a previously optimized electron dose and defocus value. To obtain a three-dimensional reconstruction of the protein, images of crystals at different angles are required. Tilting the stage of the electron microscope allows recording of images at different angles whereas the random orientation of the crystals on the grid ensures image acquisition in all perspectives. Due to beam-induced sample movements, charging effects and ice deformation during a multi-second exposure, the maximally possible tilt angle is limited to $50 - 60°$ in practice. For electron diffraction the acquisition of a high-magnification image is replaced by recording of a diffraction pattern. However the fundamental process of searching for good grid locations remains unchanged.

1.4.5 2D crystal image processing

Cryo-electron microscopy images of 2D crystals suffer from various resolution limiting effects such as low signal-to-noise ratio, limited crystalline arrangement and contrast transfer function effects. Advanced image processing algorithms are applied to account for the limitations mentioned above. Henderson introduced the basic 2D crystal image processing approaches in 1977 [Henderson 1977]. Releasing an advanced version of these scripts in 1996 made the method publically available [Crowther *et al.* 1996]. Working with this collection of scripts was significantly simplified in 2007 with the introduction of 2dx [Gipson *et al.* 2007b]. Gipson et al. introduced a graphical user interface and a standardized workflow for crystallographic image processing (Figure 1.3). The basic steps of image processing for 2D crystals are detailed below:

Defocus estimation Images recorded on a transmission electron microscope are not 1:1 projections of the specimen, as they are modulated by a *contrast transfer function* (CTF). The CTF describes (in reciprocal space) how different frequencies are transformed into the recorded image [Scherzer 1949]. Besides

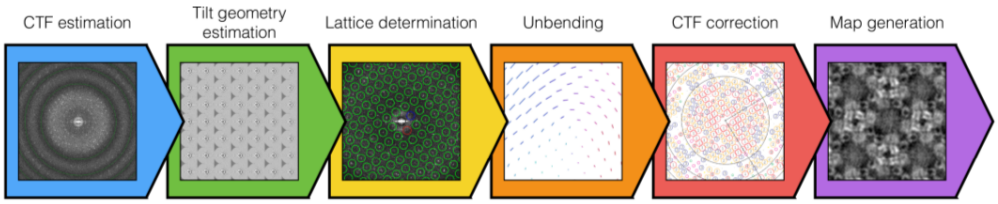

Figure 1.3: Electron crystallography image processing workflow.

imaging system dependent parameters, such as the acceleration voltage dependent wavelength and the spherical aberration coefficient of the microscope, the CTF depends on the amount of defocus used to record a particular image (reviewed in [Penczek 2010b]). Cryo-electron microscopy uses the concept of recording defocused images for the sake of contrast enhancement. Resolution zones where the CTF equals to zero are called *Thon rings*. The processing pipeline implemented in 2dx uses a CTF-model that depends on three parameters: defocus values in two orthogonal directions and the orientation of the principal axis with respect to the coordinate system of the image. 2dx uses the well-established program CTFFIND3 [Mindell & Grigorieff 2003] for defocus and astigmatism estimation. CTFFIND3 maximized the correlation between the experimental power spectrum and a parametrized simulated power spectrum by means of a grid search approach combined with a subsequent gradient minimization procedure. The actual CTF correction is done later after correcting of other resolution limiting effects.

Tilt geometry determination Three-dimensional electron crystallography requires recording of tilted images. The orientation of a 2D crystal in an electron microscope is given by the orientation of the microcopy grid and the rotational orientation of the crystal on the grid. Different areas of an image recorded on a tilted grid in an electron microscope feature different defocus values due to their varying z-height. 2dx implements two different approaches for tilt geometry determination based on the presence of a defocus gradient. The program CTFTILT [Mindell & Grigorieff 2003] uses a tile-based CTF model to determine the tilt axis and obtains the tilt angle secondly from those tiles. Alternatively the image is cut in 7x7 smaller sub-images, whose local defocus is determined by CTFFIND3. Fitting of a two-dimensional plane into the set of locally varying defoci finally reveals the tilt geometry.

Lattice determination The regular arrangement of the proteins in a 2D crystal results in the formation of diffraction spots when looking at the amplitudes of the Fourier transform of the recorded images. The later correction for crystal imperfection requires to know the crystallographic lattice of the imaged crystal. The lattice of a mono-layered 2D crystal can be parameterized in Fourier

space by two vectors $\vec{u}$ and $\vec{v}$. Specific diffraction spots are numbered according to their *Miller indices*. For instance the spot $(1, 3)$ is located at $1\vec{u} + 3\vec{v}$ with respect to the lowest frequency origin in the center of the reciprocal space image. 2dx implements two different lattice determination approaches [Zeng *et al.* 2007a]: *GetLattice* and *FindLattice*. GetLattice does not rely on previous knowledge about the lattice. Two linearly independent vectors $\vec{u}$ and $\vec{v}$ are determined based on the occurrence frequency of orientational patterns within the set of previously determined reflections. In cases of diffraction patterns with systematic absences of diffraction spots, e.g. beam-induced drift perpendicular to the tilt axis, this method is likely to fail. A more reliable algorithm is implemented in the program FindLattice. Based on the "a priory" known non-tilted lattice parameters and the previously determined tilt geometry the expected vectors $\vec{u}$ and $\vec{v}$ are predicted. After rotationally aligning the predicted lattice, a least square approach is used to fine-tune the parameters of the tilted lattice.

Correcting for crystal imperfections - *Unbending* Holistically correcting for crystal imperfections is done in three steps: (i) creation of a reference that approximates the expected crystallographic pattern, (ii) cross-correlation of the reference with the image and (iii) determining and correcting for translational imperfections. By comparing the spot brightness with the background intensity, the quality of all reflections on the lattice are quantified with a so-called *IQ-value*. Only spots above a user-defined threshold are kept for reference generation while the remaining areas of the spectrum are deleted (*Fourier filtration*). Applying an inverse Fourier transform to the Fourier-filtered spectrum reveals the first approximation of the crystal. Cross-correlation of this reference with the raw image produces a so-called *cross-correlation profile* indicating the unit-cell positions of the 2D crystal. The MRC program QUADSERCH [Crowther *et al.* 1996] is used to precisely determine the location of the cross-correlation peaks. Finally translational corrections are applied to small patches of the input image in order to maximize the regularity of the observed cross-correlation profile. 2dx divides the process of crystal unbending into multiple rounds. After the first round of unbending a more accurate reference is generated. The second reference is then used for the two consecutive final rounds of unbending. To avoid erroneous overfitting, the translational correction is applied only after the last round of unbending. The corrections from the previous rounds are used as a starting point for the later QUADSERCH runs.

CTF-correction and final map generation The MRC program MMBOX is used to convert the unbent image into a so-called *APH-file* [Crowther *et al.* 1996], which stores amplitude and phase information for all reflections. Note that these reflections are still modulated by the contrast transfer function of the

electron microscope. To compensate for CTF-effects 2dx applies phase flipping amplitude modulation [Arheit *et al.* 2013c]. The CTF-corrected and appropriately weighted reflections in the final APH-file are used to calculate the final noise-free real-space representation of one crystallographic unit-cell.

Combining multiple non-tilted images - *2D Merging* In order to obtain a high-resolution 2D projection map, the processing results from multiple different images (unit-cells) have to be combined, i.e. *merged*. The fundamental process is straightforward: (i) registration of all projection maps followed by (ii) weighted averaging (and symmetrization) of the structural information.

The translational placement of the unit-cells obtained from individual image processing of all non-tilted micrographs in the data set are random. Obviously averaging multiple unit-cells requires prior translational alignment. In practice the common origin is chosen such that the later symmetrization of the merged projection map is trivial, i.e. according to standard symmetry operations. Initially the resulting unit-cell of one crystal is selected and centered with respect to the symmetry. The following translational alignment (*phase-origin refinement*) of all remaining unit-cells is efficiently done by applying phase-shifts in Fourier space. The merged amplitude and phase information is generated by reflection-wise weighted averaging of the aligned APH-files [Henderson *et al.* 1986]. Symmetrizing the merged APH-file enforces the symmetry of the protein of interest. Similar to individual image processing, the real-space representation of the merged two-dimensional projection map is obtained by an inverse Fourier transform of the final symmetrized set of reflections. In practice the refinement is implemented as an iterative process, which uses the latest merged map as reference for the next phase-origin refinement.

3D Reconstruction The three-dimensional Fourier transform of a two-dimensional crystal shows diffraction spots in the xy-plane caused by the regular protein arrangement in this direction. The lack of such a repetitive pattern in z-direction leads to the formation of continuous lattice lines orthogonal to the xy-plane (Figure 1.4). Holistically a 3D reconstruction from 2D crystal images is obtained by the following process: (i) individual single image processing of all images, (ii) inserting the APH-information into the three-dimensional Fourier space, based on the previously determined 3D orientation of each crystal, (iii) interpolation along the lattice lines to account for unsampled regions of the Fourier transform [Shaw 1984] and (iv) inverse Fourier transform of the resulting reciprocal space representation in order to obtain a 3D real-space reconstruction (reviewed in [Arheit *et al.* 2013a]). Similar to phase-origin refinement in 2D merging, the translational alignment of all (tilted and non-tilted) crystals of one project is optimized by an iterative 3D refinement procedure [Schenk *et al.* 2010]. Optionally this iterative process can correct the tilt geom-

etry inaccuracies and resolution limiting on-axial beam-tilt effects. Typically the maximal tilt angle at which reasonable images can be acquired is $50-60°$. Thus a significant region of the reciprocal space remains unsampled, which results in the occurance of a so-called *missing cone*. Gipson et al. presented an approach that computationally restores the correct structural information in missing cone [Gipson *et al.* 2011].

Model building The final step of modern cryo-electron microscopy based structural biology is building of an atomic model into the three-dimensional reconstruction of the protein of interest. There are two different types of model building used in cryo-electron microscopy: (i) fitting of a previously determined structure into the three-dimensional reconstruction and (ii) de-novo model building in cases where no previously known structure but a high-resolution reconstruction is available.

In practice it is often the case that the structure of individual components (subunits) of the protein complex of interest have been solved in a previous study. Alternatively there are scenarios in which the protein structure has been determined by means of a different structure analysis technique, e.g. X-ray crystallography. Although the protein structure is no longer completely unknown in these cases, there are often a lot of open questions, such as what is the structure of the membrane-embedded protein. Flexible molecular dynamics fitting was introduced to fit known protein structures into three-dimensional cryo-electron microscopy reconstructions [Trabuco *et al.* 2008]. Technically, a molecular dynamics simulation is combined with an additional constraint, modeled by an additional force field, which drives the simulated structure in direction of the 3D map. In general molecular dynamics simulations require a lot of computational resources. Efficient geometry based sampling of protein structures under experimental restraints (as implemented in DireX, [Schröder *et al.* 2007]) is a valuable alternative to fit a model into an experimental density map.

Based on recent hardware developments, cryo-electron microscopy reveals reconstructions at significantly improved resolution [Kühlbrandt 2014]. In cases where a high-resolution map (typically around 3.5Å) but no structural information is available, de-novo model building can often be used successfully. De-novo model building in cryo-electron microscopy is an active field of research that is mostly inspired by X-ray crystallographic software packages and workflows [Brown *et al.* 2015, López-Blanco & Chacón 2015, Wang *et al.* 2015].

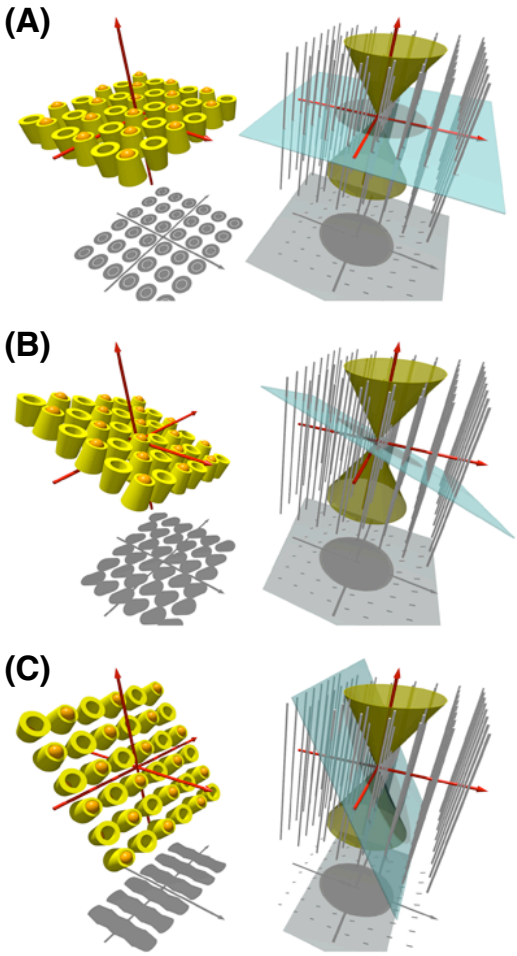

Figure 1.4: Schematic view of a 2D crystal in real space (left) and Fourier space (right) shown at three different tilt angles: (A) 0°, (B) 30° and (C) 60°. Scheme from [Schenk *et al.* 2010].
The shadows in the left panels illustrate the real-space images recorded by the transmission electron microscope. The blue slices in the right panels equal the Fourier transforms of the recorded images, which are placed into the 3D reciprocal space reconstruction based on the central slice theorem. The regular arrangement of protein in the xy-plane causes the formation of diffraction spots in the xy-plane in reciprocal space, while the absence of such a pattern in z-direction leads to the presence of lattice lines in Fourier space. Although missing values along these lines can be interpolated by lattice line fitting, the values along these lines have to be densely sampled to obtain a high-resolution 3D map. However the experimentally limited tilt range causes the formation of a missing cone, in which no information is measured at all (yellow cone in the right panels).

1.5 Direct electron detecting

In the past the highest-resolution cryo-electron microscopy data sets were acquired on photographic film (for example [Zhang *et al.* 2010]). Prior to image processing, the photographic films had to be developed and digitized by a scanner. The limited throughput of this pipeline explains the wide use of optically inferior charge-coupled devices (CCD) cameras in the field. In order to record an image, a CCD converts incoming electrons into photons via a scintillator (signal-to-light conversion). Due to its higher throughput and more convenient handling, CCD cameras were preferred over photographic film in practice. The recently introduced complementary metal-oxide-semiconductor (CMOS) chips allow direct conversion of electrons into an image, which significantly improves the signal-to-noise ratio of the images [McMullan *et al.* 2009, Ruskin *et al.* 2013]. The used CMOS technology is similar to cell-phone cameras. However the CMOS chips have to be radiation hardened to prevent electron beam damage. To prevent image blurring caused by electron scattering within the chip, the CMOS chips have to be very thin (*back-thinning*). Current sensors are approximately half as thick as a sheet of paper. Beside the improved optical performance of the new generation of detectors, they feature an enhanced detector readout frequency. Thus instead of one single image, a short movie sequence of the same sample area can be recorded. Although the samples are kept at cryogenic temperature, beam-induced sample movements cannot be eliminated completely. Recent studies showed how movie-mode image acquisition can be used to eliminate (or minimize) beam-induced sample blurring [Campbell *et al.* 2012, Brilot *et al.* 2012, Scheres 2014a]. Generally there are two different motion-correction approaches: (i) translational correction on frame level and (ii) alignment of individual regions of the frames.

Li et al. [Li *et al.* 2013a] used a global cross-correlation function to translationally align the frames of a movie sequence. This approach assumes that different regions of the imaged area feature the same drift. Locally varying drift cannot be corrected by this approach. Nevertheless the software is widely used in the field and led to multiple high-resolution reconstructions (for example [Amunts *et al.* 2014, Liao *et al.* 2013]).

Multiple recent studies reported the development of single particle analysis techniques using a per-particle drift-correction [Campbell *et al.* 2012, Brilot *et al.* 2012, Scheres 2014a]. Dependent on the particle size, the algorithmic approaches range from fitting of linear motion tracks [Scheres 2014a] up to correcting for sample rotations [Campbell *et al.* 2012] caused by beam-induced ice-layer deformations. Another recent publication reports the development of movie-mode processing capabilities for single-particle based helical image processing [Fromm *et al.* 2014].

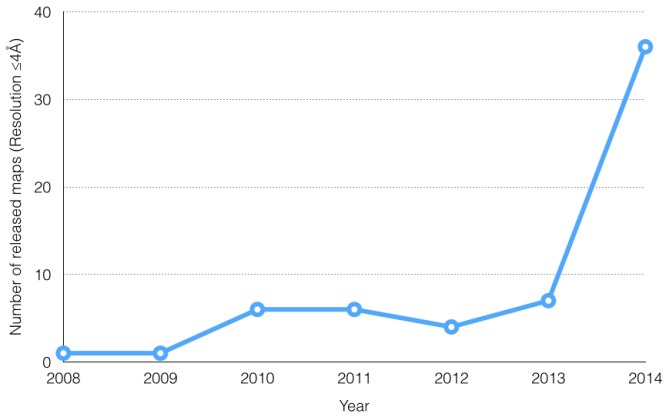

Figure 1.5: Number of high-resolution maps ($\leq$ 4Å) deposited into the EMDB from 2008 till 2014.

A hybrid beam-induced motion-correction approach, which uses both approaches consecutively (frame-wise motion correction followed by motion tracking of individual particles), obtained a near-atomic resolution reconstruction of a small membrane protein [Lu *et al.* 2014b]. With the introduction of the new direct electron detectors cryo-electron microscopy has finally become a protein structure analysis method that allows de-novo atomic model building for a large variety of samples (Figure 1.5 and [Liao *et al.* 2013, Amunts *et al.* 2014, Kudryashev *et al.* 2015] for exmaple).

1.6 Structure and aim of this thesis

The main focus of this thesis is on improving image processing approaches for 2D crystal images in order to achieve higher resolution reconstructions in shorter time. These method developments include (i) overcoming resolution-limiting assumptions made for classical image processing, (ii) implementing of an automation interface that allows a significantly improved throughput and finally (iii) developing of new software that fully exploits the potential of the latest generation of detectors. Thus this thesis consists of three method development chapters and one additional chapter summarizing multiple applications of the developed tools.

Classical image processing for 2D crystals assumes that the imaged crystals are perfectly flat and well ordered. Practically it is almost impossible to obtain these perfect crystals. In Chapter 2 we illustrate the limitations, in terms of resolution, introduced by the limited order and flatness. We present an algorithm that refines the three-dimensional orientation of each protein of a crystal. The very low signal-to-noise ratio does not allow three-dimensional refinements to be carried out on individual proteins of a 2D crystal. We solve this issue by exploiting the natural correlation between neighbouring proteins within a 2D crystal. Both, the proof that single particle approaches can be combined with 2D crystal image processing and a superior 3D reconstruction are given by the newly implemented software. To handle the large computational requirements of the novel approach, the developed software uses high-performance computing techniques.

Traditionally high-resolution studies of 2D crystals were recorded on photographic film. Due to the latest generation of digital detectors, this time consuming analog data acquisition is obsolete today. The significantly improved image quality and the fully digital nature of the entire process enables the implementation of an automatic "real-time" image processing tool presented in Chapter 3. A superior 3D reconstruction obtained by means of the novel automated processing pipeline within two working days is shown in this chapter. Previously, using the traditional film-based acquisition approach, more than 10 months of work were required to obtain a significantly lower resolution map. The developed image acquisition pipeline, featuring real-time image processing and quality insurance capabilities, was successfully applied to multiple 2D crystal projects in the meantime.

Chapter 4 covers recent algorithmic developments, which are necessary to fully exploit the possibilities enabled by the new generation of detectors. As detailed in Section 1.5, these detectors record a short movie of the sample instead of one single image. Thus it became possible to correct for beam-induced sample movements by computer programs, i.e. movie-mode data processing. A polished 3D reconstruction shows the benefits of movie-mode crystal unbending. Additionally analyzing the resolution-dependent beam-induced fading of computed diffraction spots, revealed

the optimal imaging procedure for two-dimensional crystal images recorded on the latest generation of detectors.

Finally an overview about multiple projects, all of them making use of one or multiple methods outlined above, is given in Chapter 5. The automation framework from Chapter 3 was successfully used for the structural characterization of three different membrane protein 2D crystals. Furthermore the real-time drift-correction, implemented as first stage of the automation pipeline, is helpful for all kind of cryo-electron microscopy projects. For instance the drift-correction was used in context of two remarkably high-resolution helical image processing projects. Today automatic data acquisition is indispensable for high-resolution high-throughput single particle projects. Therefore Chapter 5 additionally documents the successful implementation of such an automation system.

1.7 Publication list

Peer-reviewed and published articles

Kudryashev, M., Wang, R.Y-R., Brackmann, M., Scherer, S., Maier, T., Baker, D., DiMaio, F., Stahlberg, H., Egelman, E.H., and Basler, M., 2015. **The structure of the type six secretion system contractile sheath solved by cryo-electron microscopy.** Cell 160 (6), 952-962

Scherer, S., Kowal, J., Chami, M., Dandey, V., Arheit M., Ringler, P., Stahlberg H., 2014. 2dx_automator: **Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline.** Journal of Structural Biology 186 (2), 302-307

Scherer, S., Arheit, M., Kowal, J., Zeng, X., Stahlberg, H., 2013. **Single particle 3D reconstruction for 2D crystal images of membrane proteins.** Journal of Structural Biology 185 (3), 267-277

Conference proceedings

Kowal, J., Scherer, S., Sejwal, K., Chami, M., Baumgartner, P., Rangl, M., Scheuring S., Schröder, G., Nimigean, C., Stahlberg, H., 2014, **Cryo-electron microscopy of potassium channel membrane proteins.** Microscopy and Microanalysis 20 (S2), 1206-1207

Work in progress

Scherer, S., Kowal, J., Chami, M., Schröder G., Stahlberg H., **Movie-mode image processing for 2D crystals.** Journal of Structural Biology *(submitted)*

Lauer, M.E., Graff-Meyer, A., Rufer, A.C., Maugeais, C., von der Mark, E., Matile, H., D'Arcy, B., Magg, C., Ringler, P., Mueller, S.A., Scherer, S., Dernick, G., Thoma, R., Hennig, M., Niesor, E.J., and Stahlberg, H., **Cholesteryl ester transfer between lipoproteins does not require the formation of a ternary tunnel complex with the cholesteryl ester transfer protein.** *(submitted)*

Sborgi, L., Ravotti, F., Dandey, V., Dick, M., Mazur, A., Reckel S., Chami, M., Scherer, S., Böckmann, A., Egelman, E., Stahlberg, H., Broz P., Meier, B., Hiller, S,., **Structure and assembly of the mouse ASC filament by combined NMR spectroscopy and cryo-electron microscopy.** *(submitted)*

2 Single particle 3D reconstruction for 2D crystal images of membrane proteins

Correcting for out-of-plane tilt of individual proteins of a 2D crystal is required to achieve the highest possible resolution. In this article we investigate the impact of limited flatness and present a computational approach to deal with locally disordered crystals.

The following section has been published in:

Journal of Structural Biology
Volume 185, Issue 3, March 2014, Pages 267-277
http://dx.doi.org/10.1016/j.jsb.2013.12.011

Single particle 3D reconstruction for 2D crystal images of membrane proteins

Sebastian Scherer[a], Marcel Arheit[a], Julia Kowal[a], Xiangyan Zeng[b] and Henning Stahlberg[a,*]

a - Center for Cellular Imaging and NanoAnalytics, Biozentrum, University Basel, 4058 Basel, Switzerland
b - Fort Valley State University, 1005 State University Dr., Fort Valley, GA 31030, USA

* Corresponding Author: Henning.Stahlberg@unibas.ch

Contents

Abstract

In cases where ultra-flat cryo-preparations of well-ordered two-dimensional (2D) crystals are available, electron crystallography is a powerful method for the determination of the high-resolution structures of membrane and soluble proteins. However, crystal unbending and Fourier-filtering methods in electron crystallography three-dimensional (3D) image processing are generally limited in their performance for 2D crystals that are badly ordered or non-flat. Here we present a single particle image processing approach, which is implemented as an extension of the 2D crystallographic pipeline realized in the 2dx software package, for the determination of high-resolution 3D structures of membrane proteins. The algorithm presented, addresses the low single-to-noise ratio (SNR) of 2D crystal images by exploiting neighborhood correlation between adjacent proteins in the 2D crystal. Compared with conventional single particle processing for randomly oriented particles, the computational costs are greatly reduced due to the crystal-induced limited search space, which allows a much finer search space compared to classical single particle processing. To reduce the considerable computational costs, our software features a hybrid parallelization scheme for multi-CPU clusters and computer with high-end graphic processing units (GPUs). We successfully apply the new refinement method to the structure of the potassium channel MloK1. The calculated 3D reconstruction shows more structural details and contains less noise than the map obtained by conventional Fourier-filtering based processing of the same 2D crystal images.

2.1 Introduction

The low signal-to-noise ratio (SNR) obtained when unstained proteins embedded in a layer of amorphous ice are imaged by cryo-transmission electron microscopy (cryo-EM) hinders their structural analysis. Both the localization and orientation of smaller molecules and complexes are difficult to define precisely, and the use of averaging methods to enhance the signal is correspondingly difficult. The situation can be improved for membrane proteins as they can be incorporated in a lipid bilayer and crystallized in two dimensions (2D). This strongly restricts the orientational freedom of the macromolecule in the crystal lattice. The regular arrangement lends itself to averaging techniques and, of course, to electron crystallography.

Perfect 2D crystals are flat, distortion-free and, unfortunately, almost impossible to obtain. Rather, deviations have to be recognized and corrected for, as do possible imaging effects. A series of image processing and electron crystallography software packages tackle this problem (reviewed in [Arheit *et al.* 2013c]), the most famous being the MRC programs [Crowther *et al.* 1996]. These have allowed atomic models to be determined for different membrane proteins. The basic procedure used

to process crystal images is comprised of six steps: (i) lattice determination, (ii) determination of the tilt geometry, (iii) lattice correction (unbending), (iv) contrast transfer function (CTF) correction, (v) data averaging for each image, and (vi) merging and lattice line fitting of the data from several images in three dimensions (3D). The 2dx software [Gipson *et al.* 2007b] implements a user-friendly graphical interface to the MRC programs, and adds several additional modules for automation and additional algorithms. 2dx offers user guidance and facilitates the project management. The conventional crystallographic image processing approach addresses locally disordered crystals by crystal unbending, where small image patches (typically 25×25 pixels) are shifted in the image plane. In the present study we show that the mathematical model behind unbending fails if crystals are disordered and tilted, or are non-flat. In order to overcome these limitations, we introduce a new refinement process that resolves local crystallographic disorder in 3D.

Cryo-electron microscopy (cryo-EM) single particle analysis [Frank 1975] is a technique used to determine the structure of proteins without any crystallization. Randomly oriented single particles are imaged under the electron microscope leading to different views of the protein under investigation. The orientation of the recorded projections is initially unknown. The approximate orientation of each particle is determined by comparing the particles with a set of trial projections obtained from an initial model for instance obtained by random-conical tilt reconstruction [Radermacher 1988]. Back-projecting all particles into a 3D volume based on their updated orientations leads to an improved 3D model. This iterative process is repeated until the 3D model no longer changes. A broad variety of single particle reconstruction software packages is available today, for instance: EMAN [Ludtke *et al.* 1999], BSoft [Heymann 2001], Xmipp [Sorzano *et al.* 2004], Sparx [Hohn *et al.* 2007], Frealign [Grigorieff 2007] Relion [Scheres 2012b], or Simple [Elmlund & Elmlund 2012].

Here we present a single particle-based reconstruction method for 2D crystals that is implemented as an optional extension of the 2dx software. The approach applies the principle of single particle processing to 2D crystals, exploiting neighborhood correlation between adjacent proteins in the crystal in order to overcome the limitations originating from the low SNR. High performance computing hardware is used to handle the computational costs of the new procedure. The major differences of the new method to classical single particle analysis are the new local averaging step, the significantly finer angular search space sampled during the refinement and a new crystal-based particle selection procedure validating the alignment parameters of each particle.

2.2 Approach

In the MRC programs, the mathematical model behind local 2D correction of crystal disorders by in-plane shifting of small crystalline patches, does not cover all the possible crystal imperfections present in a 2D crystal. 3D rotational disorder is not addressed at all. The simplification of the unbending approach is partially valid for non-tilted specimens but not for tilted specimens Figure 2.1. 3D disorder is parameterized by three Euler angles and two in-plane translational parameters, and must be considered for each protein of a 2D crystal individually to fully exploit the data and achieve reconstructions with the best possible resolution.

Below we introduce a single particle-based reconstruction method for 2D crystals that refines the orientation locally for each particle, while using the correlations of neighboring particles in the crystal to overcome limitations imposed by the low SNR of cryo-EM images. The new processing approach (Figure 2.2) is a refinement of the conventional crystallographic image processing pipeline described by [Arheit *et al.* 2013a, Arheit *et al.* 2013c] (Figure 2.2A), and is implemented as an option in the open source 2dx (version 3.5.0) software. This new procedure can be applied to any previously recorded 3D cryo-electron crystallography dataset.

The initial stages of the classical approach (Figure 2.2A) include determination of image defocus values with the program CTFFIND3 [Mindell & Grigorieff 2003], lattice definition, and correction for translational disorder in the crystal, i.e., image unbending. The latter involves cross-correlation of the raw image with an iteratively improved reference image to define the position of individual unit-cells using the MRC program Quadserch, and translational adjustment of small crystal patches using the MRC program CCUnbend. Finally, the structural data from all unit cells of one crystal image are combined into one resulting unit cell image at a much higher SNR. The ensemble of data from multiple crystals at different specimen tilts is merged into one 3D reconstruction (MRC programs Origtilt, Latline, and others).

Our new refinement procedure (Figure 2.2B and C) uses the results obtained by the classical method as a starting point, i.e., unit cell positions, CTF parameters of each image and the final 3D model, to improve the resolution of the 3D reconstruction. We propose a new interactive and a new automatic procedure to refine the classically determined tilt geometry of entire 2D crystals (Figure 2.2B). Subsequently the 3D orientation of each protein of each crystal is iteratively optimized using our new local refinement tool (Figure 2.2C), taking the correlation between neighboring particles into account.

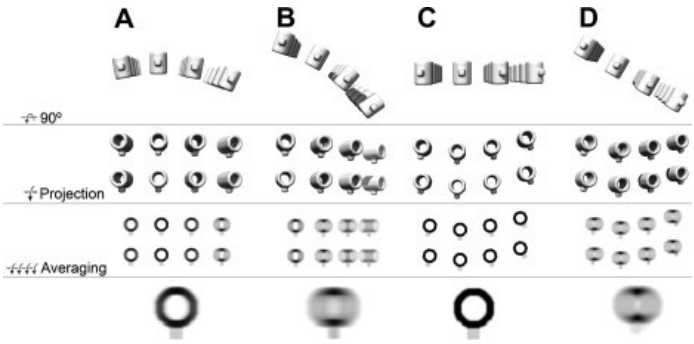

Figure 2.1: Effect of sample flatness and local disorder on resolution in crystallographic processing. Artificial crystals composed of multiple copies of a tube with a protrusion are considered. From top to bottom: perspective side-view of the crystal; top-view of the crystal (only two protein rows shown for the sake of clarity); the orthographic projection recorded by a transmission electron microscope; the outcome of an optimal crystal unbending procedure as could be achieved by the classical crystallographic processing. (A) Nominally non-tilted, yet locally bent 2D crystal with perfectly in-plane aligned particles. Although the local disorder cannot be retrieved perfectly by unbending the crystal, the merged projection map (bottom row) resembles the true projection of the structure with little resolution loss. (B) The crystal in (A) tilted by 30°. Due to the membrane curvature, the regularity of the crystal is lost in the recorded projection. As a result, high-resolution spots perpendicular to the tilt axis in the Fourier transform of the crystal disappear, which strongly limits the resolution of the projection map (bottom row). Thus the presence of ultra-flat preparations is required when imaging tilted 2D crystals and processing them with the classical lattice unbending approach. (C) Non-tilted perfectly flat crystal with rotational disorder. This is the situation in which the classical unbending procedure performs optimally. Note that even the rotational disorder can be corrected by using an optimal patch size in the unbending step. (D) The perfectly flat crystal in (C) is tilted by 30°. Here, the classical unbending-based processing cannot deal with this situation as the projection image contains multiple different views of the proteins in the crystal. The classical processing fails to align the difference projections from slightly different viewpoints. Taken together, the unbending-based processing performs well on non-tilted crystals, but encounters serious limitations for tilted crystals, even if they are perfectly flat but rotationally disordered.

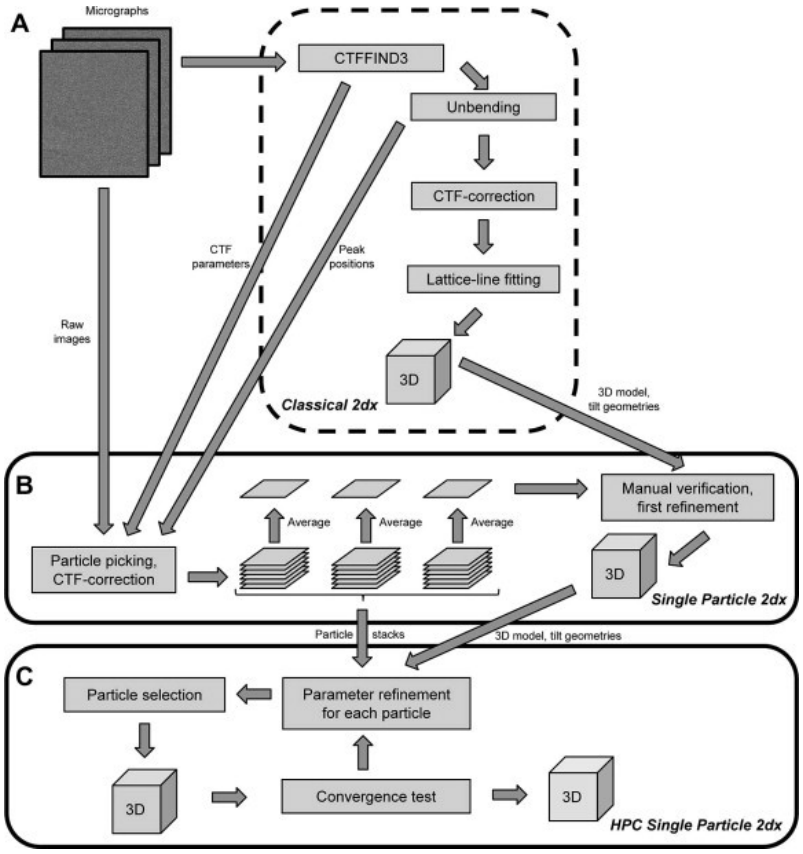

Figure 2.2: Single particle refinement for 2D crystals. The presented approach refines classically processed cryo-electron crystallographic datasets. (A) All micrographs are processed and merged classically with 2dx [Gipson *et al.* 2007b]. The unit-cell positions determined while unbending the crystals, the defocus values at the center of each micrograph and the final MRC-reconstruction are then forwarded to the new single particle module of 2dx. (B) The particles are cut out from the original micrographs based on the unit cell positions determined previously with the MRC program Quadserch [Crowther *et al.* 1996]. Dependent on their position, the tilt geometry and the central defocus parameters of the crystal from which the particles were originally picked, they are CTF-corrected and finally stored in per-crystal particle stacks. In order to verify and correct the initial tilt geometry of the particles, one averaged particle is calculated for each crystal. The orientation of this average with a much higher SNR is refined either manually by means of a new interactive tool or automatically by an iterative single particle reconstruction procedure. (C) Refining the local tilt geometry variations, is computationally demanding. Therefore, the particle stacks, the refined initial model and the configuration file are uploaded to a high-performance computing cluster.

2.2.1 Particle localization and picking from 2D crystals

The picking of single particles within 2dx as originally introduced by Zeng et al. [Zeng *et al.* 2007b] was expanded by an additional step that reduces the number of false positive particles in the background based on estimates of the local particle density of the crystalline structure.

The new particle-picking tool of the 2dx single-particle processing module (Figure 2.2B), uses the unit-cell positions determined by the MRC program Quadserch to extract patches centered on the cross-correlation peaks in the cross-correlation profile generated by the 2dx-script UnbendII, which indicates the unit-cell position on the raw micrograph. The screw axis symmetry present in a larger number of 2D membrane protein crystals requires inverting the initial tiltgeometry for all protein particles picked from the oppositely facing population of particles. One example is shown in Figure A.1. In the current implementation, in case of a screw axis present in the crystal symmetry, particles picked from the two differently oriented conformations are processed independently, and thus no correlation between these two subsets is generated during the processing. Due to this initial tilt geometry inversion of one sub-population of particles in such a case, the remaining symmetry enforcement such as 4-fold symmetrization during the reconstruction step is straight-forward. The advantage of using Quadserch via 2dx_image is that 2dx allows interactive verification and refinement of the unbending parameters (e.g., reference position and Fourier diffraction spots contributing to the reference) in order to improve the reliability of the cross-correlation profile.

As some false positives in non-crystalline regions are nevertheless unavoidable, the new particle-picking tool uses a particle density-based selection method to ensure that erroneously selected particles corresponding to false correlation peaks in the background are rejected. In a first step the center of mass (CM) of the crystal is calculated. The n ($n \approx 40$) particles closest to the CM are considered as the central region, and a density is calculated for this region based on the number of particles per surface. A region is then defined around each particle based on the n nearest neighbors of the particle, and the unit density of the region is calculated (local density). Particles with a local density significantly deviating from the density of the densely packed central region around the CM of the crystals are rejected, as they most likely correspond to an erroneous noise induced peak rather than a particle induced peak.

2.2.2 Local CTF-correction

Particles that passed the density-based selection procedure are still affected by unfavorable CTF effects, which depend on (i) the sample tilt of the entire crystal and (ii) the specific position of each particle in the image. Using basic trigonometry, the local defocus and therefore its associated CTF parameters for each particle are calculated from the CTF values at the center of the image determined by CTFFIND3 [Mindell & Grigorieff 2003] and manually verified within 2dx. The particle defocus values were not further refined during the following processing. A detailed explanation of the local CTF correction can be found in the supplementary material (A) of this article. Finally, the CTF is corrected by phase flipping and the corrected particles are stored in per-crystal particle stacks.

The particle position in the z-direction changes if the orientation of regions of the crystal is changed because the 2D crystal is not flat (Figure 2.1A&B). However defocus values are not refined in the current implementation. Following Zhang and Zhou [Zhang & Hong Zhou 2011] the resolution limit imposed by a defocus determination that is inaccurate by 124nm at 300kV acceleration voltage is at 7Å. This resolution limit is beyond the resolution of the test dataset used in the results section (Section 2.4). Further, by basic trigonometry and the values reported by Zhang and Zhou, it follows that changing the global tilt angle of an entire non-tilted 1μm sized crystal by up to 14° during the refinement does not limit the resolution. In the here processed dataset, we have not observed an angular change in this range. We therefore chose to not refine the defocus values of individual particles in our current implementation.

2.2.3 Initial tilt geometry: determination, verification and refinement

The tilt geometries of entire crystals determined by classical crystallographic processing may still be slightly off (up to 5°). Thus, we have implemented a new user interface (Figure 2.3) that allows manual verification and correction of the tilt geometries and in-plane particle positions of an entire dataset, and the use of an automatic routine that optimizes the crystal orientations iteratively based on the results obtained by the classical method.

For this verification, all particles of one crystal are averaged in real-space in order to generate one average with a higher SNR for each crystal, i.e., a 'per-crystal' average. The orientations of the crystals serve as the initial orientation of the averages. The per-crystal averages are then compared with the images obtained by projecting the final 3D model generated using 2dx in the respective nominal directions. The new user-interface (Figure 2.3) enables interactive refinement of the initial orientations

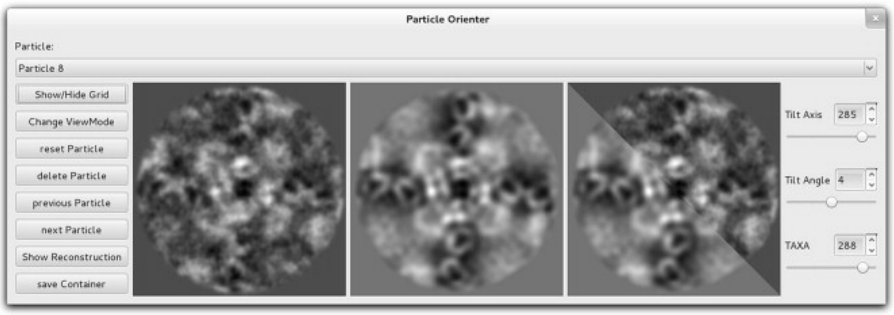

Figure 2.3: Interactive graphical user interface to verify and refine the particle orientation. The goal is to find the three Euler angles that lead to the projection of the 3D reconstruction (central window) that matches best the per-crystal average (left window). Changing the sliders on the right immediately updates the projection shown in the central window. For a more convenient comparison between projection and average, the right window is used to toggle between the two view modes, which are shown side by side. An updated starting model can be achieved by back-projecting the averages along their refined orientations. This refined model can be inspected in 3D by means of UCSF Chimera [Pettersen *et al.* 2004] launched directly from the new interface.

determined by conventional processing. Particles from highly inconsistent crystals or from crystals with insufficient structural information can then be removed from the dataset.

Instead of manually refining the dataset, one can run an iterative self-consistent single particle refinement [Frank 1975] of the per-crystal averages. A set of trial projections of the initial volume around the nominal orientation of each per-crystal average is calculated, and correlated with the per-crystal averages. The orientation of the projection with the highest correlation value is then assigned to the average and, thus, to all particles contributing to this average. Subsequently, the 3D reconstruction is updated by back-projecting the averages into a 3D volume. The described iterative procedure is repeated until the refinement converges, i.e., the orientation parameters of the particles change less than a given threshold. In practice we measured the average angular change per average and stopped the refinement once this average change was below a user-defined threshold of 0.2°. Note that the proposed refinement of the initial orientations resembles a classical real-space single particle refinement of the per-crystal averaged particles. The computational efficiency is significantly improved in case of an approximately flat 2D crystal, which then allows the angular search space to be restricted to ±5°.

The reliability of the in-plane alignment of the particles has a large effect on the quality of the final reconstruction. A better 3D model will lead to a more precise in-plane alignment of the individual particles. After the first round of initial model refinement, our approach optimizes the two in-plane shift parameters of all individual particles. In order to do this, each particle image is cross-correlated with the projection of the initial 3D model into its nominal direction. The vector from the center of the particle to the maximal cross-correlation value is then remembered as an additional shift value for the particle. Note that the described shift refinement resembles synthetic unbending in 2dx, where projections of the 3D model are used as a reference for the translational unit-cell position correction [Kunji *et al.* 2000]. More reliably in-plane aligned particles produce better per-image averages, which further improves the initial tilt geometry determination.

Dependent on the specific dataset, it can be difficult to refine the initial tilt geometries uniquely based on averages showing one isolated protein. Using image patches that show the central protein surrounded by its crystalline neighbors, generally makes the orientation determination more reliable as it increases the amount of information that can be used for the refinement. In order to stay computationally efficient, we bin the extracted patches once. After the global tilt geometries of all crystals have been optimized based on the binned patches, the refined orientations are applied to the unbinned particle stacks picked from the same dataset.

2.2.4 Locally averaged single particle orientation parameter refinement

The major aim of the present study is to resolve the locally varying tilt geometry caused by smoothly bent membrane protein crystals. This is achieved by means of an adapted single particle algorithm. The new procedure (i) overcomes the limitations caused by the low SNR and (ii) reduces the computational cost significantly by exploiting the fact that all particles originate from ordered 2D crystals.

The aim of conventional single particle processing [Frank 1975] is to find the optimal set of alignment parameters $(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s_x}, \hat{s_y})_i$, for each particle that maximizes the following expression:

$$(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s_x}, \hat{s_y})_i = \underset{(\theta,\psi,\phi,s_x,s_y)\in\mathcal{T}}{\arg\max} \operatorname{sim}\Big(P_i, \mathcal{P}(A, \theta, \psi, \phi)\Big) \tag{2.1}$$

where $\operatorname{sim}(\cdot,\cdot)$ measures the similarity (i.e., cross-correlation) between the particle P_i and a reference projection $\mathcal{P}(A, \theta, \psi, \phi)$ that depends on the 3D reference A and the three Euler angles (θ, ψ, ψ) describing the orientation of the particle in 3D. Once the best matching orientation has been determined for each particle, a more

accurate 3D structure is obtained from them using back-projection along the updated orientations. The updated structure serves as a reference for the subsequent iteration step. This self-consistent optimization is continued until convergence is reached, i.e., until (i) angular changes are smaller than a certain threshold or (ii) the structure A changes less than a given threshold from one iteration to the next one. Similar to the initial model refinement the iteration is considered to be converged, if the average angular change over all particles is below 0.2° from one iteration to the next. In practice, the optimization of Equation 2.1 is carried out over a finite set of trial projection $\mathcal{T}$ uniformly covering the entire angular search space. Note that the optimization also includes an in-plane shift (s_x, s_y) alignment of the particle P_i. As this optimization can be done efficiently by determining the offset between the cross-correlation peak and the center of particle P_i, the shift parameters are not separately mentioned in the optimization target function in Equation 2.1.

Both, the significantly lower SNR and the correlation between neighboring particles, are essential differences between classical single particle approaches and our new application. The application of a single particle refinement-based version of Equation 2.1 to individual particles picked from cryo-EM images of unstained 2D membrane protein crystals was notsuccessful in our hands, likely due to the too low SNR induced by the small electron dose used on imaging, and by disturbances from overlapping projections of neighboring particles in images of tilted 2D crystals (data not shown). Conventional single particle processing of randomly oriented particles can increase the SNR by refining the orientations of class averages instead of the individual particles. Similarly, here we exploit the correlation between neighboring particles in order to increase the SNR. Orientations of two neighboring particles will be very similar, while those of distant particles may deviate due to lattice bends and the membrane curvature that our approach aims to resolve (Figure 2.1). Accordingly, the orientations of close and distant particles should not be treated as if they are directly coupled. Instead of using individual particles for the alignment step we use locally averaged, radially Gaussian weighted particle averages to determine the angular orientation of each particle. Therefore, we replace the objective function of the underlying optimization problem Equation 2.1 by the following expression:

$$(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s_x}, \hat{s_y})_i = \underset{(\theta,\psi,\phi,s_x,s_y)\in\mathcal{T}_i}{\arg\max}\ \mathrm{sim}\Big(P_i + \sum_{j\in<\mathcal{N}_i>} \lambda_j P_j, \mathcal{P}(A,\theta,\psi,\phi)\Big) \tag{2.2}$$

where the similarity measure, $\mathrm{sim}(\cdot,\cdot)$ is the same as in Equation 2.1. The major novelty is that the similarity is now calculated between the projection and the weighted sum of the central particle P_i and particles P_j belonging to the neighborhood $< \mathcal{N}_j >$ associated with particle P_i. Note that the projection trial space $\mathcal{T}_i$ depends on the initial orientation of the crystal and thus has an additional subscript index i. Typically, we use a dozen neighbors and a sampling cone of a few degrees.

In this way, the computational complexity can be massively reduced and a finer search space can be sampled. The weights λ_j are following a Gaussian distribution, where the particles closer to the central particle contribute more to the average:

$$\lambda_j = \exp\left(-\frac{1}{2}\left(\frac{R_j}{\sigma}\right)^2 \right) \tag{2.3}$$

where R_j equals the distance between the central particle P_i and the adjacent particle P_j. The parameter σ is used to tune the weight. Our current implementation adjusts σ so that the particle furthest from the center contributes ten times less to the average than the central particle. Once the locally summed particles are aligned, the original central particles are back-projected into a 3D volume in order to generate a new reference for the following iteration step. The procedure is repeated until convergence is reached.

2.2.5 On the fly likelihood-based particle selection

Due to crystal defects or fallacious noise peaks, some particles of each crystal are incorrectly aligned and thus should not be used for the reconstruction step. Therefore, we introduce a particle selection method that uses the iteratively refined alignment parameters of each particle together with the neighborhood correlation stemming from the crystalline arrangement to judge the plausibility of each particle image.

Based on the alignment we define a 'fingerprint vector' for particle P_i:

$$\vec{x_i} = \left(\theta_i, \psi_i, \phi_i, (s_x)_i, (s_y)_i, \mathrm{cc}_i\right)^T \tag{2.4}$$

consisting of the three Euler angles $(\theta_i, \psi_i, \phi_i)$, the two shift parameters $(s_x)_i, (s_y)_i$ and the corresponding highest cross-correlation value cc_i determined during the refinement process. As suggested by Figure A.2 the components of the particle fingerprint vectors follow a 6-fold multivariate normal distribution. Based on this assumption the log-likelihood function $\log f(x_i|\mu_{\mathcal{C}(i)}, \Sigma_{\mathcal{C}(i)})$, which describes the likelihood (or 'probability') of finding a single observation $\vec{x_i}$ given the estimated mean vector $\mu_{\mathcal{C}(i)}$ and covariance matrix $\Sigma_{\mathcal{C}(i)}$) (with $\Sigma_{mn} = (x_m - \mu_m)(x_n - \mu_n)$), is given by [Anderson & Olkin 1985]:

$$\log f(x_i|\mu_{\mathcal{C}(i)}, \Sigma_{\mathcal{C}(i)}) = -\frac{1}{2}\log(|\Sigma_{\mathcal{C}(i)}|) - \frac{1}{2}(\vec{x_i} - \mu_{\mathcal{C}(i)})^T \Sigma_{\mathcal{C}(i)}^{-1} (\vec{x_i} - \mu_{\mathcal{C}(i)}) \tag{2.5}$$

Here, $\mathcal{C}(i)$ indicates that the mean vector and the covariance matrix depend on the crystal from which the particle P_i was originally picked. $|\Sigma_{\mathcal{C}(i)}|$ represents the determinant of the covariance matrix. Note that a normalized particle finger print vector is used for the likelihood estimation. The mean vector and covariance matrix are estimated separately for each crystal. Particles with a log-likelihood value below a certain threshold are not used for the reconstruction step in order to improve the quality of the structure.

While the parameter refinement process optimizes different parameters serially, the particle fingerprint vector evaluation considers all parameters associated with the alignment simultaneously. This is important as it allows the correlations between the alignment parameters to be taken into account. This is not possible when the individual alignment parameters are verified sequentially, in which case the sequential thresholding and assessment will most likely fail if, e.g., the in-plane orientation worked out fine but the angular assignment failed. Our new method conceptually outperforms approaches that serially judge the plausibility of individual alignment variables and do not account for their correlations.

2.3 General implementation details

2.3.1 Software design and 2dx integration

The presented single particle extension for 2dx has been implemented in version 3.5.0 of the open source software package, which is available for download at `www.2dx.org`. Following 2dx's philosophy, we have extended the set of available image processing scripts by an additional script section dedicated to single particle processing (Figure A.3). These new single particle scripts are used to launch the different image processing programs of the single particle backend, which is implemented in `c++`. The new manual angular refinement and verification user interface (Figure 2.3) is natively integrated into 2dx's frontend realized with Qt4 (`www.qt-project.org`). Pieces of code of the new single particle library use functions implemented in EMAN2 [Tang *et al.* 2007], i.e., projection, back-projection and symmetry operations. Therefore, the new executables are linked with the EMAN2 backend library libEM2.a. The new package makes intensive use of `libboost`, a collection of high quality `c++` utility libraries (`www.boost.org`), for low-level subroutines such as parsing files or multi-dimensional or sophisticated data structures.

2.3.2 High performance computing

Automatic image acquisition procedures, such as Leginon [Carragher *et al.* 2000], in combination with the latest generation of direct electron detectors [Bai *et al.* 2013, Campbell *et al.* 2012, Li *et al.* 2013a] that can record movies of dose-fractionated image series instead of single images, will increase the size of future datasets tremendously. In order to process these future datasets in a reasonable amount of time, supporting high-performance computing hardware is indispensable for our new software.

Due to the large number of proteins present in each imaged crystal (up to 3000 particles/crystal, 131Å $\times$ 131Å unit-cell size, 2 proteins/unit-cell, up to 3 μm crystal side-length) the memory consumption of the per-protein refinement step is in most real world datasets higher than the amount of available memory on a regular desktop computer. Thus we perform the conventional MRC-based processing, particle picking and initial model refinement on a local machine, while running the intensive per-protein refinement on a remote high performance cluster computer that should have significantly more memory and computing power. All desktop-based pre-processing steps (Section 2.2.1-2.2.3) are accelerated by means of a shared memory parallelization scheme based on OpenMP [Dagum & Menon 1998] and follow a highly adapted data-flow mechanism that minimizes the memory requirements by sequentially loading the particle stacks to fulfill the memory boundaries of traditional workstations. Hence the initial pre-processing can be performed on powerful workstations within a reasonable time.

Today's super computers consist of a large number of multi-core processors that are connected by a hierarchical high-bandwidth network. In order to use these machines as efficiently as possible, the per-protein local refinement tool features a hybrid parallelization approach [Jost *et al.* 2003]. Consequently, the set of particles to be refined is distributed equally over a large number of multi-core processors (here called compute nodes) that communicate through the message-passing interface (MPI). Each compute node additionally stores a copy of the actual 3D reconstruction, which serves as reference for the alignment described in Section 2.2.4. The parallelization approach is visualized in Figure 2.4.

As particles originating from the same crystal share a considerable number of the required trial projections, pre-calculating and storing the projections decreases the computational complexity of the later cross-correlation maximization significantly compared to an on-the-fly projection calculation scheme. Thus, each compute node uses the orientation of the particles and the current approximation of the reconstruction to pre-calculate the projection gallery required for the subsequent alignment step (Section 2.2.4). After disabling inconsistent particles based on the procedure detailed in Section 2.2.5 all particles stored on a compute node are back-projected

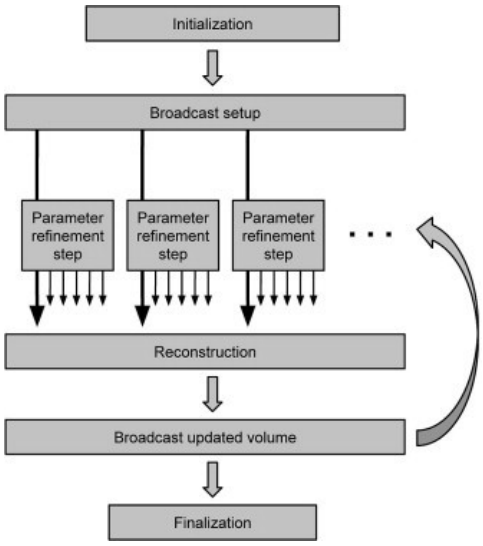

Figure 2.4: Hybrid parallelization approach. The application is launched on a large number of multi-core compute nodes connected by a high-bandwidth network. Each of the node runs only one MPI rank on one of its cores and keeps the rest of the cores initially idle. In order to setup the refinement, one of the MPI nodes (master node) decides on the optimal particle distribution over all available nodes and broadcasts the load-balancing scheme to all MPI ranks. Each node of the cluster now loads the required particles from the storage system attached to the high performance cluster. A projection gallery is calculated locally on each node based on the orientations of the particles assigned to the node. As there is only one MPI job running per multi-core node all of the per-node computational tasks can be performed in parallel by means of a shared memory parallelism scheme. After the neighborhood-dependent local orientation refinement, all particles stored on a node are back-projected into a 3D volume on each node. In order to update the global reconstruction, all local reconstructions are sent to the master MPI rank, which assembles the global reconstruction. This updated 3D model is subsequently broadcast to all nodes in the system and serves as reference for the next iteration step. Once convergence of the refinement procedure is reached, the optimal orientations are stored to the file system and some statistical parameters are stored for a later quality analysis of the refinement, e.g., the cross-correlation values or assigned angles.

into a local 3D volume. As the updated global 3D structure depends on all particles spread over a large number of nodes, all local reconstructions are gathered on the master MPI-node and finally merged by weighted Fourier averaging into one volume. Our approach uses Fourier-placing on each node and synchronizes the cumulated interpolation weights towards the master MPI-node to account for non-uniformly sampled voxels. The central slice theorem, storing the Fourier space representation of all objects, and the absence of any per-particle weights, allow us to distribute the reconstruction step over multiple compute nodes. The updated structure is used as a new reference model for the next alignment step and thus is broadcasted to all nodes of the cluster.

The projection and cross-correlation steps per node are accelerated by an OpenMP shared memory parallelization scheme. To achieve this we encapsulated the EMAN2 projection method so that multiple projections can be calculated simultaneously in a thread-safe way, i.e., multi threads can execute a function simultaneously without influencing each other. All together, our hybrid parallelization approach maximizes the performance on homogeneous clusters by using an optimized parallelization paradigm on each level of parallelism.

2.3.3 GPGPU accelerated projection method

The performance bottleneck of the initial model refinement (Section 2.2.3) performed on a desktop computer is the calculation of the projection gallery rather than the alignment of the per-crystal averaged particles. We successfully accelerated the projection method by using a general-purpose graphic processing unit (GPGPU) running the c++ language extension CUDA provided by Nvidia. The use of graphic processors significantly accelerates image processing algorithms related to cryo-EM [Castaño-Díez *et al.* 2008]. Our implementation makes efficient use of the texture memory and can be used simultaneously by multiple host threads in a thread-safe manner. By means of the hypercube technology introduced in CUDA-5.0, multiple device kernels can run simultaneously on the same graphic processing unit (GPU) by utilizing multiple execution streams (implemented in hardware), which once again boosts the projection calculation. Our software additionally supports multi-GPU systems in order to distribute the work over several hardware accelerators. GPU-support on high-performance computing clusters will be realized in a future version.

2.4 Results and discussion

2.4.1 MloK1 – an experimental test dataset

We recently determined two different confirmations of a ligand-modulated bacterial potassium channel MloK1 [Kowal *et al.* 2014]. This tetrameric ion channel features cyclic nucleotide binding domains (CNBDs) and putative voltage sensor domains (VSDs) on each of its monomers. In the context of the study presented here we used images of the ligand-bound MloK1 2D crystals to benchmark the new refinement procedure. The test dataset consisted of 73 low dose images recorded on a Philips CM200-FEG cryo-transmission electron microscope, operated at an acceleration voltage of 200kV and a nominal magnification of 50′000×, using a defocus range of 500–2500nm. The crystal samples were vitrified in the holes of holey carbon film. Images were recorded on photographic films (Kodak SO163), which were digitalized using a drum scanner with a step size of 5μm, resulting in a pixel size of 1Å at the specimen level and in a crystallographic unit-cell size of 131Å. All crystal images were processed and merged into one 3D structure (Figure 2.5A, mesh, blue) with 2dx, yielding a final resolution at 7Å resolution in xy-direction and 12Å orthogonal to the imaging plane [Kowal *et al.* 2014].

2.4.2 Experimental procedure: single particle refinement of a test dataset

In a first step we extracted 1000 patches (300 × 300 pixels) from each crystal. The patches were CTF-corrected, binned by a factor two (resulting in patches of 150×150 pixels) and averaged in order to generate per-crystal average patches. Each patch contained the central protein tetramer and four neighboring tetrameric proteins. The presence of these neighbors simplified the orientation determination. The orientation of the per-crystal averages was manually refined against the initial 3D structure by means of the new interactive particle alignment tool (Figure 2.3).

Based on the positions of the unit-cells determined while unbending the crystals with 2dx, we extracted 244′604 un-binned particle images (150 × 150 pixels) from the micrographs. Each particle image was masked so that it contained one unique central protein tetramer. Subsequently we removed 34′066 false positive particles selected in the background, by means of the new density based particle selection tool (Section 2.2.1). All subsequent refinement steps were performed with the remaining 210′538 individual membrane proteins from 73 crystals. The automatic iterative initial orientation refinement described in Section 2.2.3 was executed within a ±5° range until the model did not significantly change further. All of the automatically determined orientations were verified manually before any further refinement was performed. After the first initial model refinement, we optimized the in-plane posi-

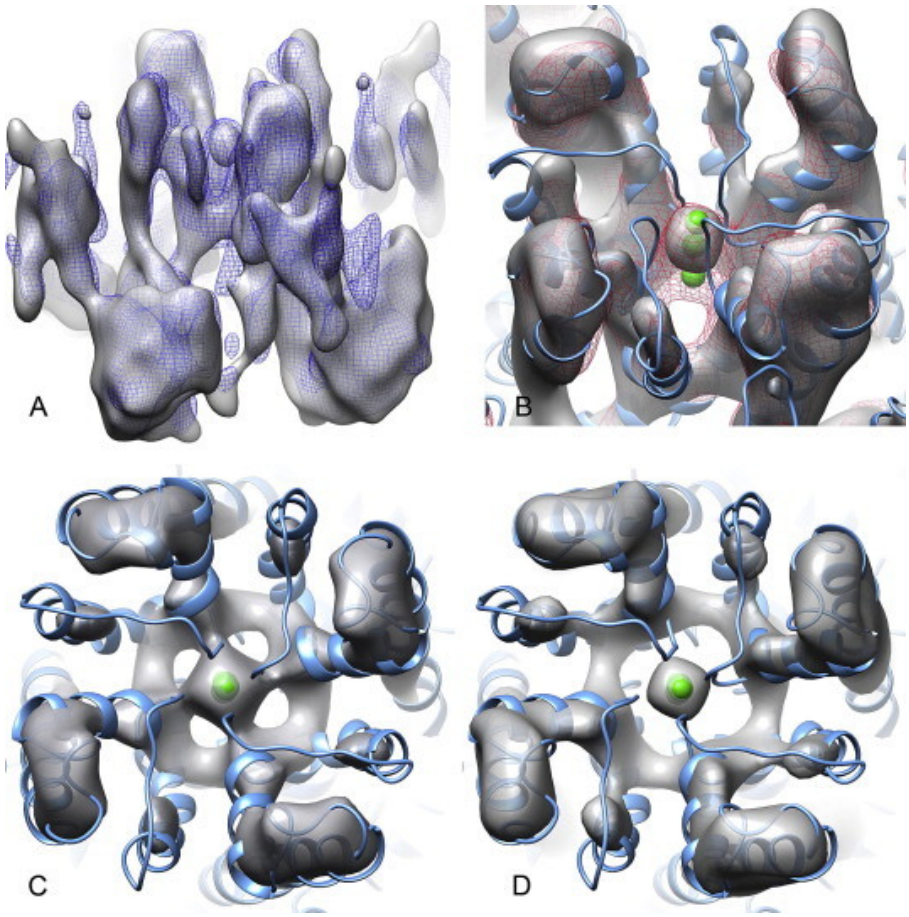

Figure 2.5: Results obtained with the 3D local tilt geometry refinement procedure on a cryo-EM data set. (A) Reconstruction obtained by classical unbending processing within 2dx (mesh, blue) compared to the locally refined structure obtained using our new algorithm (solid, grey). (B) Closer look at the pore region of the membrane protein where the quality difference is largest. The atomic model shown was obtained by flexible molecular dynamic fitting using the cryo-EM map generated by the classical unbending approach as an additional constraint. The locally refined structure (solid, grey) resolves the helices much better than the conventionally generated map (mesh, red). (C and D) Comparison of the top-views of the pore region obtained by unbending (C) and the new procedure (D). The iso-contour levels were set so that the long perpendicular helices of the pore region have the same dimensions in all maps shown.

tion of each particle within the 21×21 pixel window by allowing individual particles to shift by maximally 10 pixels in any direction and recalculated the per-crystal particle averages. This was done by applying the local averaging scheme described in Section 2.2.4 for the in-plane alignment, i.e., instead of individual noisy particles the weighted sums of nearby particles were used to determine the optimal in-plane correction. Although we propose to use the reconstruction obtained by classical processing as starting model, our software can deal with any starting model provided as MRC-format file. Using a large in-plane shift search space allows the procedure to correct for translational offsets between experimental data and the starting model obtained by another procedure. After the shift correction another round of iterative initial model refinement was applied to the updated per-crystal averages.

The local tilt geometry refinement for each particle was performed in two consecutive refinement rounds. In the first phase of the coarse refinement, the local orientations were fine-tuned within a $\pm 12^\circ$ cone in which the angular distance between the individual trial direction was maximally 3°. In the second phase, the in-plane rotation (orientation of the crystal on the grid) of each particle was optimized over a $\pm 3^\circ$ range in 1° steps. For this initial coarse refinement, information in the resolution range from $12-75$Å was used for the cross-correlation calculation. Averaging the central particle with the 16 nearest neighboring particles increased the SNR enough for a consistent orientation determination to be made, i.e., the orientations determined for neighboring regions of the crystal were in a reasonable range (a few degrees). In our hands the refinement characterized by the resulting parameter set showed a remarkably fast convergence. After the third iteration the 3D structure did not visibly change anymore, consequently we limited the number of coarse iteration steps to three. The subsequent fine-grained second refinement was based on the optimal alignment parameters found at the coarser refinement level. In order to resolve finer local variations, we limited the trial projection space to a $\pm 3^\circ$ cone with 1° separated trial directions. The in-plane rotation was optimized over $\pm 1.5^\circ$ in 0.5° steps. By reducing the number of neighboring particles contributing to the average used for the alignment to 8, we allowed larger local fluctuations. The second refinement phase used information in the resolution range $10 - 50$Å and the resolution of the iteratively refined structure was limited to 10Å.

In order to improve the convergence and stability of the algorithm, we linearly combine the $(n-1)$ reconstruction with the new back-projection based on the refined parameters determined in step (n) to obtain the new reconstruction used as the reference in step $(n+1)$. The most successful refinements were achieved when 30% of the $(n-1)$ structure was retained. The likelihood-based particle selection (Section 2.2.5) was used to select the 85% most consistent particles per crystal for the back-projection step. Convergence of the second iteration procedure was observed after 5 rounds. A comparison of the structure obtained by classical processing and the new single particle-based reconstruction is shown in Figure 2.5A. The proba-

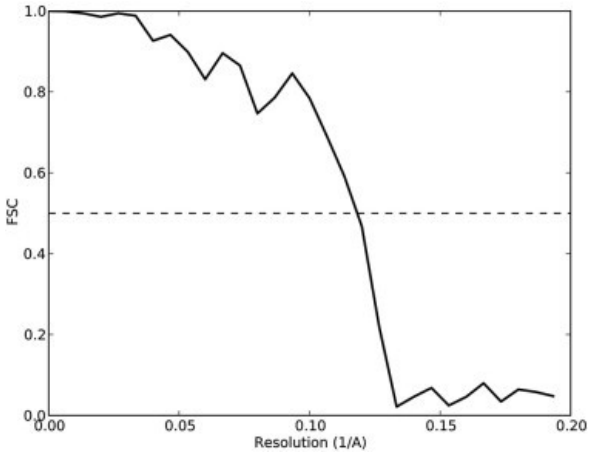

Figure 2.6: Fourier shell correlation of the final single particle based refined 3D structure indicating a resolution of 8.3Å (0.5 FSC-criterion, pixel size 1Å).

bility distributions of the refined parameters of the crystals (three angles and two shift values) show approximately Gaussian distributions around the nominal parameters of the crystals (Figure A.2A–E), suggesting that our method resolves the true variations in membrane flatness of the 2D crystals in 3D.

2.4.3 Resolution measure, prevention of overfitting

From the Fourier shell correlation (FSC) shown in Figure 2.6 we conclude (0.5 criterion) that the final reconstruction obtained by means of the presented algorithm is 8.3Å. This one-dimensional resolution limit is an average value for the anisotropic resolution of the reconstruction due to the missing cone and the beam induced resolution loss for highly tilted samples. A direct comparison between the resolution achieved by the new tool and the resolution of the conventionally generated reconstruction is not possible as there is no universal definition of resolution valid for both crystallography and single particle reconstructions. In electron crystallography the resolution is deduced from the frequency at which the phase residual drops below a commonly accepted value [Unger 2013], whereas the resolution of a single particle reconstruction is commonly calculated by splitting the dataset into two halves from which two independent reconstructions are generated. The investigation of the frequency-dependent decay of the correlation between these two reconstructions allows us to report a unique number as the resolution of the reconstruction [Penczek 2010a]. Instead of trying to conceal the two different resolution determination methods, we here estimate the improvement gained by the new refinement procedure by visual comparison of the obtained reconstructions (Figure 2.5).

A major issue of commonly used single particle reconstruction approaches is their tendency to overfit the data due to the alignment of noise instead of the true structure [Stewart & Grigorieff 2004]. The choice of appropriate low-pass filter parameters can partly avoid overfitting. However, the careful determination of suitable constraints requires a lot of time even for experienced image processing specialists, and novices generally fail. A more reliable approach is based on the calculation of a so-called gold standard FSC [Scheres & Chen 2012], where the entire dataset is split before the first round of refinement. The iterative optimization is then performed on the two halves independently. After each iteration, the FSC curve between the two independent reconstructions is used to construct an optimal filter without any user-defined parameters.

Both, the correlation between neighboring particles used during our new alignment procedure and the contribution of all crystals to the starting model, violate the independency assumption on which the gold-standard method is based. Therefore, the realization of the gold-standard procedure is not straight forward for our application. We prevented overfitting by carefully applying conventional low-pass filters with user-defined parameters. Nevertheless, the gold-standard procedure will be implemented in future versions of the presented algorithm in one or another way. For instance, exploiting only a sub-group of the crystallographic symmetry present in most of the datasets could be used to cross-validate the reconstruction and its resolution.

2.4.4 Structural improvements

The crystals of the presented test dataset were relatively well-ordered and flat, which is favorable for classical MRC processing. Nevertheless, the new refinement procedure improves the resolution of the final reconstruction and shows more structural details (Figure 2.5), although some protein backbone loops are not resolved, likely due to the anisotropic resolution and the basic CTF-correction scheme employed here. From Figure 2.5A, where the classically MRC-based merged structure (mesh, blue) and the new local refinement based single particle reconstruction are shown (solid, grey), we can conclude that the algorithm presented here resolves densities that cannot be seen in the reconstruction generated by classical MRC processing. The most significant differences are in the pore region of the membrane protein (Figure 2.5B–D). The atomic model shown for comparison was devised based on X-ray structures from parts of the molecule, which were subjected to flexible molecular dynamic fitting [Schröder *et al.* 2007] to the cryo-EM reconstruction generated by classical MRC-based methods. Although this procedure introduced a bias to the atomistic model from the MRC-based reconstruction, the resulting atomic model matches best to the structure generated using our new single-particle procedure.

For instance the new single-particle reconstruction resolves better the isolated helix S6 next to the central channel, as well as the densities for trapped ions in the central axis of the channel in the selectivity filter region (Figure 2.5C and D). These and further manual observations suggest that our new refinement procedure outperforms the conventional MRC-based processing on this dataset.

An adequate choice of the iso-contour levels used to display cryo-EM reconstructions is crucial for a reliable comparison between two different structures. The correct way is to select the thresholds so that the shown volumes equal the molecular mass of the protein divided by the average protein density. Due to the missing cone and strongly anisotropic resolution commonly affecting 2D crystallography datasets, such an approach fails in the present situation. Thus, we selected the thresholds so that the long helices of the pore region of the protein perpendicular to the membrane have the same dimensions in both volumes.

2.4.5 Performance evaluation of the likelihood-based particle selection approach

Figure 2.7 demonstrates the efficient skipping procedure developed to avoid the use of particles that are inconsistently aligned due to imaging artifacts or crystal defects. For instance the tilt angle distribution of the particles (black circles) of the crystal region shown in Figure 2.7 shows regions of strong inconsistency (Figure 2.7A; black ellipse). Based on the procedure described in Section 2.2.5 we calculate the log-likelihood to find each particle given the mean vector and the covariance matrix of the particle fingerprints of the entire crystal. The resulting log-likelihood values indicating the consistency of the particle alignment parameters are shown in Figure 2.7B. The $5 - 15\%$ of the particles with the lowest consistency values were not used for the subsequent reconstruction (Figure 2.7C), which led to a further improvement of the reconstruction.

2.4.6 Computing performance of the refinement program

The processing presented here was done on a desktop machine featuring two Intel Xeon E5-2660 3.0 GHz 8-core CPUs, 64 Gb memory and two Nvidia GeForce GTX Titan with 2688 CUDA cores and 6 Gb DDR5 RAM each. Solid-state hard drives were used for faster writing and loading of the particle stacks.

As mentioned in Section 2.3.3 the bottleneck of the initial model refinement is the calculation of the projection gallery. Table 2.1 shows the computing time required to generate the projection gallery used for one initial model refinement step applied to the presented test dataset. Because the GPU-implementation efficiently uses all

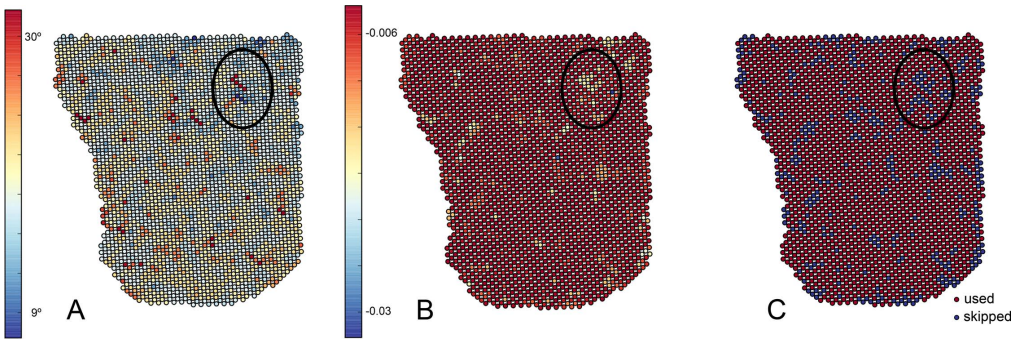

Figure 2.7: Likelihood-based particle selection within one 2D crystal. (A) Tilt angle distribution within the crystal. The area marked with a black ellipse shows inconsistently aligned particles as the tilt geometry varies more than 10° within the distance of two particles. (B) The likelihood values of the particles. Based on the particle finger print vector Eq. (5), we can calculate the likelihood to find each particle given the mean vector and covariance matrix of all particle finger print vectors of the crystal. (C) Retained and rejected particles. Particles with the lowest likelihood values (also called consistency) should not contribute to the reconstruction as their alignment most likely failed. In the present example 10% of the particles were removed (blue particles) whereas the remaining 90% of the particles were used for the back-projection step.

available host threads and all available hardware execution and data streams of the GPUs, the initial model refinement is now an interactive processing task that is 40 times faster than the fastest CPU implementation of the same task (Table 2.1). The calculation of one single projection in the 'full parallel GPU-setup' took 0.14 ms on average.

Table 2.1: Performance measure of the calculation of 54,166 projections, each 150×150 pixels. Speedup_1 states how many times faster the more sophisticated implementation is compared to the single threaded basic implementation, whereas Speedup_2 compares the GPGPU realizations with the fastest CPU implementation running on all available compute cores.

	Wall-clock time (s)	Speedup_1	Speedup_2
Single core (1 x Intel Xeon)	4943	-	-
16 cores (2 x Intel Xeon)	313	15.8	-
Single GPU, 1 CPU thread	128	38.6	2.4
Single GPU, 16 CPU cores	13.3	372	23.5
Dual GPU, 16 CPU cores	7.8	634	40.1

One iteration step of the first phase (coarse refinement level, Section 2.4.2) of the local tilt geometry refinement of the $210'538$ particles with a box size of 150×150 pixels took about 100 min wall-clock time using the machine mentioned above. During the second phase of refinement, the cone used for the trial projection calculation depends on the previously optimized local orientation of the particles. Additionally, due to the smaller angular step size much more projections have to be computed. Thus the required projection gallery no longer fits into the memory of our machine. Instead of caching the projections during the second refinement phase, they are calculated on the fly. Due to the GPGPU acceleration of the projection calculation the fine-grained local refinement can be done within a reasonable time, resulting in 300 min wall-clock time for one step of the second refinement phase.

The software was successfully installed and run on several high-performance computing clusters. Refinements using up to 512 processor cores we performed. The hybrid parallelization approach was benchmarked on up to 128 cores and showed a fairly good scaling behavior (Figure A.1).

2.5 Conclusions

The implementation of the described procedure in the 2dx package and as a standalone high-performance computing cluster program, allows variations in the local tilt geometry of a 2D membrane protein crystal to be taken into account in order to increase the resolution and the amount of structural details visible in the final 3D structure. The concept of local averaging (central and neighboring particles) makes it possible to iteratively refine further imaging parameters, such as local magnification changes, deviations of the defocus from the pre-calculated values, or beam-tilt induced phase distortions. Up to now, the low SNR of low dose cryo-EM images has prohibited the successful fitting of these imaging parameters based on projections of single membrane proteins picked from a crystal.

The high-performance computing implementation of the here-presented software prepares it for the processing of large datasets recorded automatically on next generation direct electron detectors. Instead of recording one image with the full electron dose, these detectors can be used to distribute the dose over multiple images (frames) in order to track beam-induced distortions. As recently shown [Bai *et al.* 2013, Campbell *et al.* 2012], beam-induced motion of individual particles can be resolved by means of processing individual dose-fractionated sub-frames from a direct electron detector operated in "movie mode". A future extension of the here presented approach to resolve beam-induced motion of the crystal lattice should further improve the reconstruction quality.

The here presented alignment procedure is based on cross-correlation maximization, which is prone to noise bias. Several recent studies demonstrated superior results (in both resolution and reliability) obtained by gold-standard maximum likelihood approaches compared to classical cross-correlation maximization [Lyumkis *et al.* 2013, Scheres 2012b]. A maximum likelihood approach that considers the correlation between neighboring particles as prior knowledge due to the presence of crystals is therefore expected to also further improve the single particle reconstruction introduced here.

Although our new algorithm weakens the assumptions made about sample flatness, it still requires a successful pre-processing based on the classical MRC-based 2D crystallography pipeline. Thus, the presented procedure is a refinement tool rather than an independent reconstruction pipeline. As the entire conventional processing pipeline is based on the assumption of perfectly flat and well-ordered 2D crystals, growing them is still crucial. Large well-ordered mono-layered 2D crystals are extremely difficult and time consuming to obtain. In early crystallization trials one often observes 'polycrystals' that at least locally, within smaller patches, show some minimal crystallinity, e.g., of 10×10 crystal unit cells. In combination with the processing procedure presented here, a tool that picks particles from polycrystals

would allow such badly ordered 2D membrane protein crystals to be processed. It would then be possible to determine the structure of membrane proteins whose 2D crystallization does not fulfill the stringent order and flatness requirements we know today.

Acknowledgments

We thank D. Castaño-Díez and S.A. Müller for insightful discussions and critically reading of the manuscript. We thank T. Robinson (Swiss National Supercomputing Center, CSCS), U. Borstnik (ETH Zurich, Informatikdienste, Cluster-support) and M. Jacquout (University Basel, Universitätsrechenzentrum) for their support and discussion in context of the high-performance realization of the here presented procedure. This work was supported by the Swiss National Science Foundation (Grants 315230_146929, 205320_144427, and the NCCR TransCure) and by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID d19.

All the here-described procedures are implemented in 2dx-3.5.0, which will be available on `www.2dx.org`. The entire set of tilted images of MloK1 used to benchmark the software, together with the 2dx processing parameters is available at the EBI at `http://www.ebi.ac.uk/~ardan/aspera/em-aspera-demo.html` under accession number 10006. 2dx/MRC-derived maps of MloK1 are available at the EMDB at accession codes EMD-2526 and EMD-2527. The fitted molecular models are available at the PDB under accession codes 4CHV and 4CHW.

Supplementary data

Supplementary data associated with this article can be found in A.

3 *2dx_automator*: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline

The introduction of a new detector generation in the field of cryo-electron microscopy manifested the need of an automatic image processing pipeline for 2D crystals. The here presented automation system significantly increases the throughput of image acquisition and data analysis in cryo-electron crystallography.

The following section has been published in:

Journal of Structural Biology
Volume 186, Issue 2, May 2014, Pages 302-307
http://dx.doi.org/10.1016/j.jsb.2014.03.016

_2dx_automator_: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline

Sebastian Scherer, Julia Kowal, Mohamed Chami, Venkata Dandey, Marcel Arheit, Philippe Ringler and Henning Stahlberg*

Center for Cellular Imaging and NanoAnalytics, Biozentrum, University Basel, 4058 Basel, Switzerland

* Corresponding Author: Henning.Stahlberg@unibas.ch

Contents

Abstract

The introduction of direct electron detectors (DED) to cryo-electron microscopy has tremendously increased the signal-to-noise ratio (SNR) and quality of the recorded images. We discuss the optimal use of DEDs for cryo-electron crystallography, introduce a new automatic image processing pipeline, and demonstrate the vast improvement in the resolution achieved by the use of both together, especially for highly tilted samples. The new processing pipeline (now included in the software package 2dx) exploits the high SNR and frame readout frequency of DEDs to automatically correct for beam-induced sample movement, and reliably processes individual crystal images without human interaction as data are being acquired. A new graphical user interface (GUI) condenses all information required for quality assessment in one window, allowing the imaging conditions to be verified and adjusted during the data collection session. With this new pipeline an automatically generated unit cell projection map of each recorded 2D crystal is available less than 5 min after the image was recorded. The entire processing procedure yielded a three-dimensional reconstruction of the 2D-crystallized ion-channel membrane protein MloK1 with a much-improved resolution of 5Å in-plane and 7Å in the z-direction, within 2 days of data acquisition and simultaneous processing. The results obtained are superior to those delivered by conventional photographic film-based methodology of the same sample, and demonstrate the importance of drift-correction.

3.1 Introduction

The introduction of direct electron detectors (DEDs) featuring a radiation-hardened CMOS sensor [Bammes *et al.* 2012, Milazzo *et al.* 2011] to the field of cryo-electron microscopy (cryo-EM) has brought enormous advantages. The detective quantum efficiency (DQE) of these devices is significantly higher than that of charge-couple device (CCD) cameras or even photographic film [McMullan *et al.* 2009, Veesler *et al.* 2013]. In general, images recorded on DEDs have a higher signal-to-noise ratio (SNR) and the highest contrast ever achieved. The highest DQE is thereby achieved by Gatan's K2 Summit camera due to its ability to count single electrons from its internal chip readout every 2.5 ms [Li *et al.* 2013b, Ruskin *et al.* 2013]. The fast readout speed also makes it possible to record movie sequences rather than single images, allowing to split the electron dose over multiple exposures of the same region, resulting in a sequence of images. Although the SNR of the individual frames is much lower, the sample motion during exposure can still be fol-

lowed and compensated for by frame-alignment and averaging [Brilot *et al.* 2012, Campbell *et al.* 2012], leading to impressive resolution improvements in cryo-EM [Liao *et al.* 2013]. Here, we discuss application of a DED to two dimensional (2D) membrane protein crystals.

Processing individual images of 2D crystals generally involves (i) defocus estimation, (ii) determination of the tilt-geometry, (iii) lattice estimation, (iv) lattice correction (unbending) to correct for crystal imperfections, (v) contrast transfer function (CTF) correction and (vi) averaging the data of one crystal into one unit-cell [Arheit *et al.* 2013c]. When recording data on a CCD camera or on film, both the low SNR and the presence of beam-induced sample drift, make automatic processing of crystal data challenging and, in some cases, impossible. Manual processing and assessment of the quality of individual images is still the major bottleneck. Here, we present an automation pipeline that mostly eliminates this tedious work and ensures high-throughput, while exploiting the potential of DEDs, i.e., drift-correction of dose-fractioned exposure series. The processing pipeline is executed automatically at the EM, immediately after each movie/image has been acquired. The quality of the recorded images is displayed on a single user-friendly GUI via which the operator of the electron microscope (EM) can continuously optimize the microscope settings based on the results obtained. Both, the quality and the productivity of data acquisition sessions are tremendously increased by this fast, automated procedure. The performance of the new pipeline, in terms of time to solution and achieved resolution, is superior to that of any previous 2D crystal processing procedures.

3.2 Automation pipeline setup

The automation pipeline (Figure 3.1) consists of five different stages: (i) movie/image acquisition, (ii) initial quality assessment on the microscope computer, (iii) automatic drift-correction (movie mode only), (iv) automatic image processing of individual crystals, and (v) manual 3D merging of the entire 2D crystal dataset. Storing a raw sequence of dose-fractioned movie frames (later called a stack) in a particular folder on the storage server triggers the processing procedure outlined above. The new pipeline automates the major bottleneck of the classical 2D crystal pipeline, namely the processing of a large number of 2D images (ideally more than 100). 3D merging is not a time intensive process, and requires operator decisions that can vary greatly between projects, such as detecting conformational variability among the 2D crystals, or choosing optimal lattice line fitting parameters. 3D merging therefore still is a manual process in our pipeline.

In our setup the operator of the EM records and stores the raw movie stack from the acquisition computer onto a network-attached storage server. Both, drift-correction and the automatic image processing run on a dedicated additional workstation. In

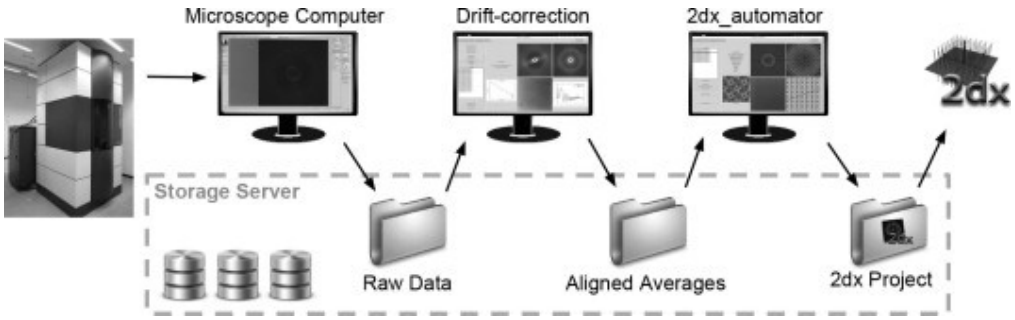

Figure 3.1: Automation pipeline. Images are recorded with the EM using a DED operated in movie mode. Saving image stacks to the storage server triggers the drift-correction implemented by [Li *et al.* 2013a]. The corrected average is automatically moved to another folder on the storage server, which triggers the automatic image processing by 2dx_automator. 2dx adds the drift-corrected image into the existing project and processes the image based on the project-wide parameters. Once all individual images have been processed and if required manually optimized, 2dx_merge is used to merge the entire dataset in 3D.

this way, a 2D projection map is generated from a raw stack (25 frames, $\sim$ 4k $\times$ $\sim$ 4k pixels) in less than 5min.

3.3 Data acquisition

The presented software was developed and optimized for the online processing of 2D crystal images recorded on a FEI Titan Krios equipped with a Gatan K2 Summit DED. In the experiments we report, the detector was operated in the electron counting-mode with an image size of $\sim$ 4k $\times$ $\sim$ 4k pixels. Following [Li *et al.* 2013b], we distributed the cumulated dose over 25 frames recorded over 5 sec. We operated the microscope in low-dose mode at a nominal magnification of $22'500\times$ at the screen level, resulting in an effective magnification of $\sim 37'000\times$ on the K2 chip (5μm physical pixel size), which converts to a pixel size of 1.34Å on the sample level.

The pixel electron dose rate, used to image the sample was kept below 6 counts/pixel/second, in order to avoid coincidence loss due to undercounting [Li *et al.* 2013b]. The cumulative dose applied to the sample for recording one image stack remained below $\sim$ 16 electrons per Å^2.

The Gatan K2 summit detector can also be operated in the so-called 'super-resolution mode', where the physical $\sim$ 4k $\times$ $\sim$ 4k optical chip is modeled as an $\sim$ 8k $\times$ $\sim$ 8k device by sub-pixel precision hit-location interpolation. Li et al. record images in

super-resolution mode and apply a "2 × 2-binning" by cropping in Fourier space to avoid aliasing [Penczek *et al.* 2014], before drift-correction and further image processing [Li *et al.* 2013a]. Both, the better detector quantum efficiency of the K2 summit detector operated in super-resolution mode [Ruskin *et al.* 2013] and the fact that Fourier cropping of an ∼ 8k × ∼ 8k image into an ∼ 4k× ∼ 4k image prevents unfavorable aliasing of super-resolution data in Fourier space, favor the use of the super-resolution mode. Our drift-correction automation tool can deal with both stacks of ∼ 4k × ∼ 4k frames as well as stacks of ∼ 8k × ∼ 8k frames, as long as the latter are automatically binned before processing [Liao *et al.* 2013]. It corrects the drift of entire frames during the exposure. Use of super-resolution imaging on a Gatan K2 summit DED would further improve the DQE at high resolution, which would also benefit single particle refinement to correct for local varying tilt geometries of the 2D crystals [Scherer *et al.* 2013].

3.4 Drift-alignment of individual frames

Sample drift during exposure to the electron beam can be corrected in the movie mode by aligning all frames and then averaging to obtain a drift-corrected average. There are different algorithmic approaches to the frame-alignment problem. The alignment method developed by Li et al. [Li *et al.* 2013a] features some algorithmic advantages compared to the straight-forward alignment of adjacent frames integrated into Gatan's acquisition software. Instead of aligning adjacent frames with respect to each other, Li et al. determine the local offsets between adjacent and also non-adjacent frames, which leads to an over determined system of linear equations. The optimal shifts with respect to a 'global cross-correlation' can be determined from the least square solution of this system of equations. In our hands the method developed by Li et al. showed more stable behavior than Gatan's acquisition software. In order to overcome the beam-induced motion of the sample at the beginning of an exposure and 'shutter opening' artifacts, like Li et al. we recommend removing the first two frames from the raw stacks and only aligning and averaging the 23 remaining frames [Li *et al.* 2013a].

To enable real-time drift-correction, we developed new automation software equipped with a graphical user interface (GUI; Figure 3.2), which employs the drift-correction toolkit by [Li *et al.* 2013a]. Our new tool periodically checks for arrival of a new raw image stack in an input directory, and then launches drift-correction on the stack to produce a frame-aligned average image in an output directory. The GUI of this tool (Figure 3.2) shows (i) the power spectrum of the raw average, (ii) the power spectrum of the drift-corrected average, (iii) the drift-corrected average extracted from the corrected stack, (iv) the rotationally averaged power spectra of the raw and corrected averages as one-dimensional plots, and (v) the applied shifts as a

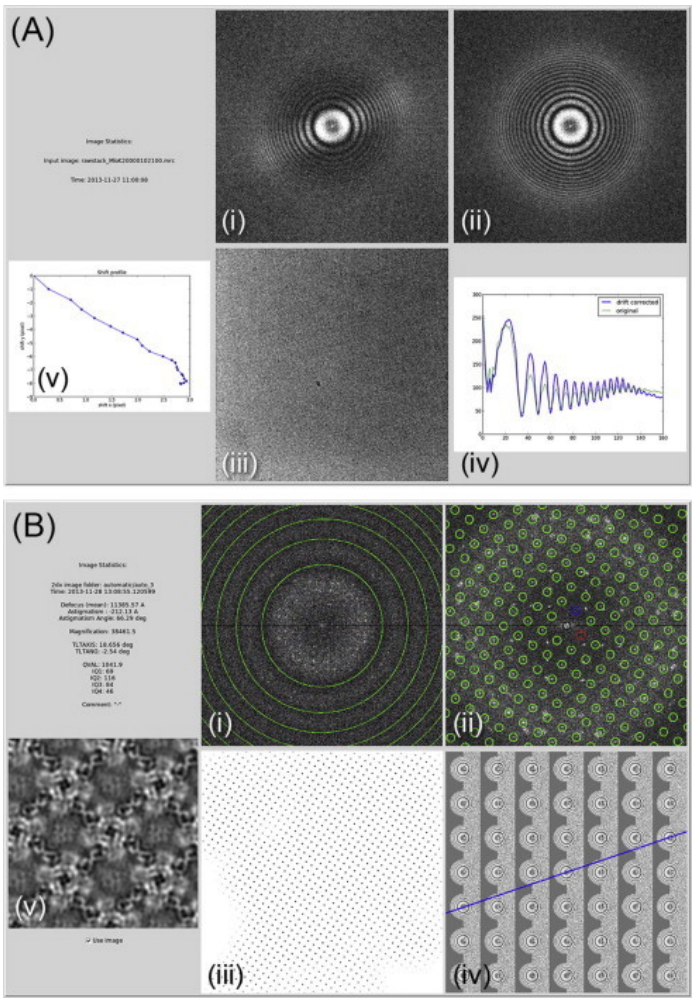

Figure 3.2: Automation GUIs. (A) Drift-correction GUI (assessment part). (i) power spectrum of the raw average, (ii) power spectrum of the drift-corrected average, (iii) drift-corrected average image, calculated from the corrected image stack, (iv) rotationally-averaged power spectrum of the raw and corrected image averages as one-dimensional plots, and (v) the applied shifts as a 2D plot. Based on the power spectra of the raw and drift-corrected averages, the user can judge the quality of the recorded image and the effect of the drift-correction. Additionally we provide a plot of the applied in-plane shift and a binned version of the corrected average. (B) Diagnostic view of the 2dx_automator GUI. (i) Fourier transform with Thon rings and (ii) fitted lattice, (iii) peak profile from the unbending step used to mask the image, (iv) locally estimated defocus values, (v) final 2D projection map.

2D plot. This new automation GUI can also be used to automatically drift-correct movie-mode images recorded for single particle and helical reconstruction projects.

3.5 Automatic image processing in 2dx

Drift-corrected images are then automatically processed by 2dx. The starting point for the presented automatic crystal image processing is hereby a configuration file providing generic parameters applicable to all images of the project. These parameters are obtained by manually processing a few images with 2dx_image and defining the optimized parameters as default values for subsequent automatic runs. The software offers a set of exemplary configuration files and allows loading of own configuration files from other projects, which facilitates the initial tuning significantly. The use of DEDs significantly simplifies finding commonly applicable processing parameters. Automation was successful for all so far tested membrane protein projects ($n > 3$), while it often failed before the introduction of DEDs due to the lower quality of the images.

The first step of image processing in 2dx is to estimate the defocus values of each micrograph. This is done using a slightly adapted version of CTFFIND3 [Mindell & Grigorieff 2003] that is fully integrated in 2dx. We suggest searching for astigmatic defocus values within $0.5 - 5.0$ μm in steps of 250Å. Based on this range, CTFFIND3 was able to find the correct defocus parameters for all tested images.

As the automation does not know 'a priori' whether a particular image shows a tilted or non-tilted crystal, our automation treats all images as if they show a tilted sample and determines the tilt-geometry of each as detailed in [Arheit *et al.* 2013b]. Due to the new detector and the significant improvements of the version from 2013 of CTFFIND3, the success rate of this procedure is almost 100% (Section 3.6).

The remaining processing tasks: (i) lattice determination [Zeng *et al.* 2007a], (ii) crystal unbending, (iii) CTF-correction and (iv) map-generation, are automatically conducted in exactly the same way as for a classical 2dx-project [Arheit *et al.* 2013c].

The automation pipeline also offers the possibility to test for potential second crystallographic lattices in the crystal images, which can then be processed within 2dx either fully automatically or manually. As most crystals do not cover the entire image, masking during the processing improves the final 2D projection map. Automatic masking is now included in the pipeline and achieved by applying binary morphologic operations to the peak-profile generated by the 2dx-script UnbendII.

Even some images recorded on a DED cannot be processed automatically and require manual intervention or fine-tuning at some points. In order to check the result of the automatic processing pipeline, we have implemented a GUI (termed 2dx_automator) that displays several useful diagnostic graphs and figures (Figure 3.2); these include (i) an overview of processing parameters such as the determined tilt geometry and IQ statistics, (ii) Fourier transform with fitted Thon rings, and (iii) fitted lattice, (iv) peak-profile from the unbending step used to mask the image, (v) locally estimated defocus values indicated by Thon ring fits from local CTFFIND3 runs on 7×7 sub-tiles of the image, and (vi) the final 2D projection map. From this information, the operator can judge the reliability and quality of the automatic processing. Further, when processing fails the provided information can be used to detect, which particular processing step needs manual intervention.

To provide a complete overview of the data recorded and automatically processed, we have implemented a 'project statistics overview' chart (Figure B.1). This informs the user about the recorded sample tilts, the recorded defocus values dependent on sample tilt, and about the determined quality values (QVal) [Gipson *et al.* 2007a, Gipson *et al.* 2007b]. From this overview the user operating the EM can see which data are still missing with respect to tilt angles and defocus settings. Thus, our software greatly facilitates the collection of data at all of the required settings, leading to high quality 3D reconstructions in terms of resolution and completeness.

3.6 Application to a MloK1 membrane protein dataset

In order to test the new automation pipeline, we imaged 2D crystals of the potassium channel membrane protein MloK1 (131×131Å unit cell size, P_{4212} symmetry, 1–3μm diameter of 2D crystals, vitrified on thin carbon films supported by holey carbon film; [Kowal *et al.* 2014]) using a Titan Krios microscope equipped with a Gatan K2 summit DED as described above. The automation pipeline was setup so that storing a raw stack to a particular folder triggered automatic drift-correction followed by automatic image processing by 2dx_automator. We recorded and automatically processed 148 MloK1 images with $0° - 45°$ nominal sample tilt within two days. We observed a 100% success rate for drift-correction, defocus determination and tilt-geometry determination based on the measured defocus gradient. In $\sim$ 4% of all cases (6 images out of 148), manual lattice determination was required, because the automatic lattice determination based on the program FindLattice [Zeng *et al.* 2007a] failed. The average time required to obtain a 2D projection map was below 5min.

The processing results obtained using images of MloK1 crystals recorded on photographic film and by DED (without and with drift-correction) are shown and compared in Figure 3.3 and Figure B.4. Crystal images recorded on film show diffraction

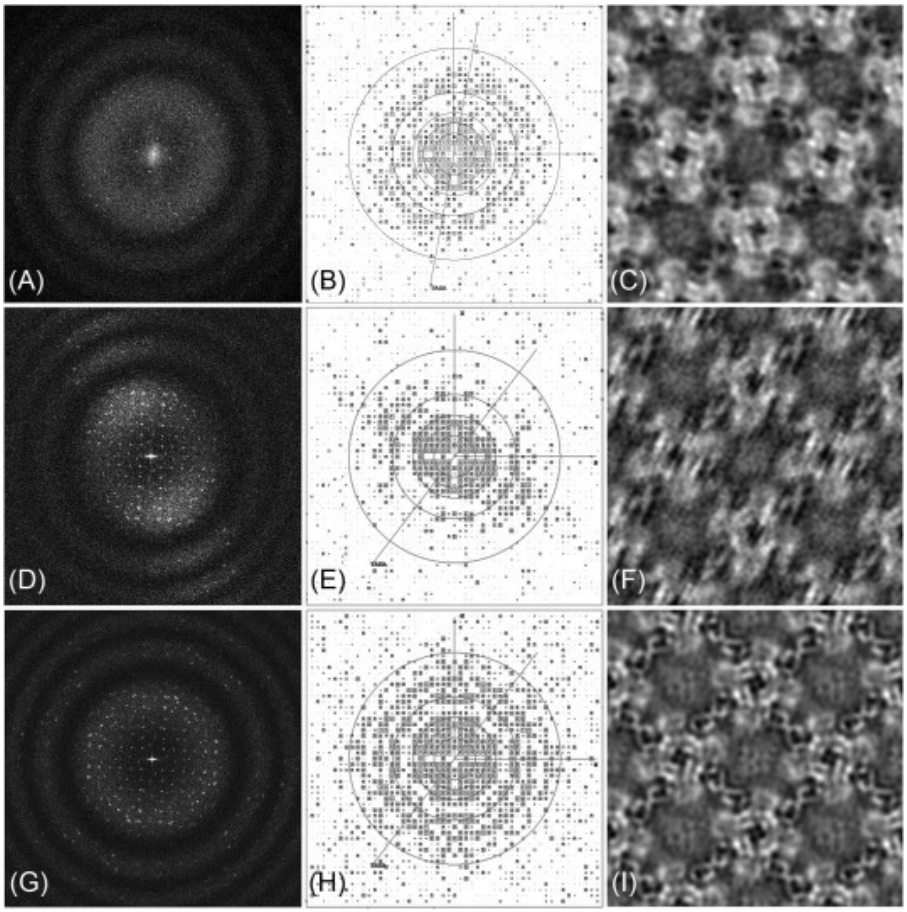

Figure 3.3: Impact of DEDs and drift-correction on 2D electron crystallography projection maps generated from a non-tilted crystal. (A–C) are for a crystal imaged on photographic film. (A) Power spectrum of an image. (B) Resolution circle plot from the canonical lattice showing the quality of the diffraction spots. Circles correspond to 36Å, 24Å, 18Å, 12Å and 7Å, the border is at 5Å. (C) Reconstructed unit-cell (2×2 unit cells are shown). (D–F) same as for (A–C) but a DED recording without drift-correction. The resolution loss due to drift is visible. (G–H) raw data used for (D–F) but with drift-correction before crystallographic image processing. The results for the crystals recorded on the DED (D–F) were obtained without manually tuning the processing.

spots with acceptable SNRs and with a fairly high completeness up to 9Å, whereas the drift-corrected image recorded on Gatan's K2 summit DED show similar completeness out to 5Å. Projection maps generated from images recorded on the DED show more features and allow easier identification of individual helices. These data also show that drift-correction on tilted 2D crystals is central to the success of electron crystallography.

3.7 Towards overcoming the beam-induced resolution loss for highly tilted samples

Recording and aligning movie images does not overcome beam-induced resolution loss completely, as local movements within the crystals are not addressed by this approach [Glaeser *et al.* 2011]. Beam-induced movement has a stronger impact on images of tilted samples than of non-tilted samples, due to a beam-induced physical up- or downwards movement of the sample and due to charge-induced deflection of the beam. Drift-correction therefore has a much greater impact still on image sequences recorded on tilted 2D crystals (Figure B.2), where we often obtained isotropic resolution up to at least 7Å in all directions also on tilted samples. Correcting the drift between dose-fractioned frames of one exposure is an easy and effective way to almost eliminate beam-induced resolution loss. Crystal unbending applied to the individual frames, i.e., 'movie-mode unbending', would further overcome the beam-induced local movements within the imaged 2D crystals [Glaeser *et al.* 2011].

3.8 3D reconstruction

We manually merged the best 65 of the 148 recorded and automatically processed micrographs to obtain a 3D reconstruction, following the approach described in [Arheit *et al.* 2013a]. Different quality assessments are shown in Table B.1 and Figure B.3. The resolution of the DED reconstruction (5/7Å) is significantly higher than the resolution obtained when photographic film was used (7/12Å) [Kowal *et al.* 2014], as also evident in Figure 3.4. The most important difference, however, is the time required to produce these two reconstructions: Several months (> 10) of work in data collection and processing over a period of 3 years were required for the data recorded on photographic film, whereas the DED study took only two days of data collection and processing, demonstrating the superiority of the new procedure.

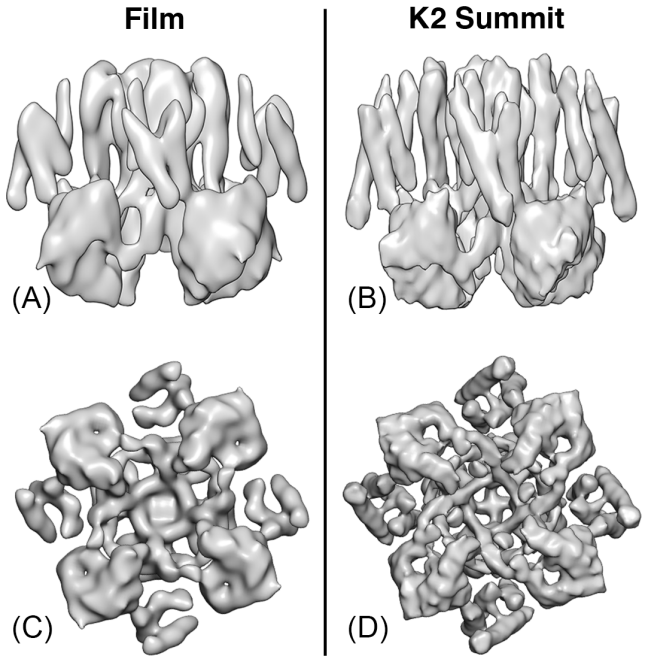

Figure 3.4: Comparison of the MloK1 3D reconstructions obtained from photographic film (A, C; > 10 months of work) and DED (B, D; 2 days of work). A, B: Membrane view. C, D: Cytoplasmic view of the reconstruction. The resolution obtained using the DED was higher.

3.9 Conclusions and future developments

The presence of high-order diffraction spots in the power spectra of the drift-corrected average, even for highly tilted samples, the almost 100% success rate of the automatic processing pipeline, and the quality of the 3D reconstruction, prove that the new automation pipeline and the new software outperform all previously available procedures in terms of usability, time to solution, and achieved resolution.

The drift-correction method used in the new automation pipeline accounts for displacements of entire frames, but not for local movements within sub-areas of the image. Correcting these local displacements, so-called movie-mode unbending, would further increase the achieved resolution. When targeting a resolution significantly below 5Å for non-tilted samples, careful microscope alignment, parallel sample illumination and computational correction for on-axial beam-tilt effects are indispensable. Images of crystals recorded at high sample tilt have to be corrected for titled-transfer function effects [Henderson *et al.* 1986], the use of which is implemented in 2dx. Limitations due to not perfectly flat crystals, especially when working with 2D crystals vitrified in buffer solution, can be overcome by applying a 3D single particle processing approach to individual proteins of a 2D crystal [Scherer *et al.* 2013]. Any limited order in the imaged 2D crystals, however, will likely remain the resolution-limiting factor in the structure determination.

Acknowledgments

We thank Shirley A. Müller for insightful discussions and critically reading of the paper, and Pawel A. Penczek, Werner Kühlbrandt, Cristina Paulino, and Deryck Mills for fruitful discussions. This work was supported by the Swiss National Science Foundation (grants 315230_146929, 205320_144427, and the NCCR TransCure). We acknowledge generous support by the FAG Basel.

All the here-described procedures are implemented in the 2dx package, which is available on `www.2dx.org`

Supplementary data

Supplementary data associated with this article can be found in B.

4 Movie-mode image processing for 2D crystals

Beam-induced sample motion is the most severe resolution-limiting factor in electron crystallography. Here we present an algorithm that computationally resolves heterogeneous sample motion based on dose-fractionated exposure series recorded on the latest generation of detectors.

The following section has been submitted for publication in:

Journal of Structural Biology

Movie-mode image procssing for 2D crystals

Sebastian Scherer[a], Julia Kowal[a], Mohamed Chami[a], Gunnar F. Schröder[b,c], and Henning Stahlberg[a*]

a - Center for Cellular Imaging and NanoAnalytics, Biozentrum, University Basel, 4058 Basel, Switzerland
b - Forschungszentrum Jülich, Institute of Complex Systems, ICS-6: Structural Biochemistry, 52425 Jülich, Germany
c - Department of Physics, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany

* Corresponding Author: Henning.Stahlberg@unibas.ch

Contents

Abstract

Direct electron detectors (DED) feature both a better detector quantum efficiency (DQE) resulting in images with a higher signal-to-noise ratio (SNR) and an enhanced readout frequency, which enables movies to be recorded instead of single images. The latter allows beam-induced sample movement to be corrected, thereby eliminating a major limiting factor in high-resolution cryo-electron crystallography. Although for two-dimensional (2D) crystal samples, the sample movements may vary from one region of a crystal to another, all motion-correction approaches applied so far only correct for global translational offsets between the movie frames. Here we present an algorithm, termed movie-mode unbending, that corrects for crystal imperfections and locally-varying beam-induced sample deformations, while preventing overfitting and computationally optimizing the electron dose used for each resolution range individually. We applied the new procedure to a three-dimensional (3D) dataset consisting of 97 images of the voltage-gated ion-channel MloK1 crystallized in 2D. The QVal, quantifying the quality of an image by one single number dependent on the spectral SNR of the computed diffraction spots, was increased by 30% on average over all test images. The 3D density map and its quality measurements improved significantly. The refined map shows more distinct densities than the 3D reconstruction obtained by classical crystal unbending and is easier to interpret.

4.1 Introduction

Direct electron detectors (DEDs) equipped with a radiation-hardened complementary metal-oxide-semiconductor (CMOS) sensor tremendously increase the signal-to-noise ratio (SNR) and thus improve the quality of cryo-electron microscopy images [Bammes *et al.* 2012, Glaeser *et al.* 2011, Milazzo *et al.* 2011]. Instead of converting incoming electrons into light via a scintillator, these new devices detect incoming electrons directly, which significantly reduces detector background noise and explains their name. Beside the superior detector quantum efficiency (DQE) of DEDs [Ruskin *et al.* 2013], this new generation of detectors feature an enhanced sensor readout frequency. This makes it possible to record movies of the sample while it is continuously exposed to the electron beam, rather than just a single image. Li et al. introduced an algorithm that efficiently measures and corrects for translational offsets between dose-fractioned movie-frames recorded on DEDs before the drift-corrected stack of image frames is averaged [Li *et al.* 2013a]. The impact of sample movements such as those caused by specimen stage-drift, can thereby be strongly reduced, which increases the efficiency of data acquisition significantly. These new hardware and software developments have led to several recent breakthroughs in high-resolution (< 4Å) cryo-EM 3D structure determinations [Allegretti *et al.* 2014, Amunts *et al.* 2014, Liao *et al.* 2013]. Cryo-EM has finally

become a powerful and fast method for near-atomic resolution protein structure determination.

We recently presented a semi-automated electron crystallography pipeline, where we combined a movie-mode capable DED with our image processing automation software 2dx_automator [Scherer *et al.* 2014]. This pipeline involved the following steps: (i) recording stacks of dose-fractioned movie-frames on a DED, (ii) on-the-fly drift-correction of the stacks, (iii) averaging the drift-corrected movie-frames to obtain one drift-corrected average image per stack, (iv) automated real-time image processing of the average image to obtain one crystallographic unit-cell per acquired stack, and finally (v) manual merging of all unit-cells featuring different orientations to obtain a 3D reconstruction. Using a DED and the 2dx_automator software significantly reduced the required time to record data and improved the resolution of the density maps obtained. Furthermore we demonstrated that drift-correction is central to the success of processing images of highly tilted 2D crystals.

Besides sample quality, beam-induced specimen movement is assumed to be the major limiting factor in cryo-electron microscopy. This especially applies to electron crystallography, if highest-resolution images of highly tilted 2D crystal samples are to be recorded [Glaeser *et al.* 2011]. The behavior of the sample under the electron beam was investigated in multiple recent studies [Bai *et al.* 2013, Brilot *et al.* 2012, Campbell *et al.* 2012, Glaeser *et al.* 2011, Veesler *et al.* 2013]. By analyzing virus or ribosome particles during the exposure, these studies report locally correlated movements that differ among the particles of an image. This was also observed for 2D crystal samples before (Anchi Cheng, The Scripps Research Institute, La Jolla, CA, USA, personal communication). It is commonly believed that these local movements are due to irreversible deformations of the ice layer and also the sample, caused by the electron beam [Brilot *et al.* 2012]. The amount of beam-induced motion depends on the microscope settings and the properties of the sample and grid. For 2D crystals, Glaser et al. proposed firmly attaching the crystals to an appropriate strong and conductive support in order to overcome beam-induced sample movements [Glaeser *et al.* 2011]. However, such treatment may not be optimal for membrane proteins that show structural alterations when adsorbed to carbon film. A more general approach is to allow the 2D crystal membrane to move under the beam, and then computationally retrieve the structural data from recorded movies. Here, algorithms accounting for universal motion of all movie-frames, e.g., caused by stage drift, cannot correct for spatially varying beam-induced motion as they neglect locally varying movements within the frames. An image processing approach that locally treats sub-regions of a 2D crystal sample in each movie frame is needed.

Another resolution-limiting factor in electron microscopy is electron beam-induced radiation damage. A high cumulative electron dose destroys fine structural details, whereas a low cumulative electron dose preserves the high-resolution information

but reduces the SNR of the recorded image. Before detectors capable of movie-mode imaging were available, the full electron dose was captured within one single image and the EM operator had to carefully choose the electron dose depending on the target resolution. In 2010, before the advent of DEDs, Baker et al. recorded dose-fractionated image series of crystals and analyzed the resolution dependent vanishing of computed diffraction spots. Their findings suggest to record dose-fractionated image series of the sample and add these frames together using resolution-dependent frequency filters, in order to optimally exploit each image frequency at its maximal SNR [Baker *et al.* 2010].

2D membrane protein crystals provide a very low SNR in cryo-EM images for each single membrane protein. However, the fact that a 2D crystal contains many regularly arranged identical copies of the same protein allows electron crystallography to overcome this limitation by using Fourier-filtering methods. Processing individual micrographs of 2D crystals generally involves the following six steps [Arheit *et al.* 2013c]: (i) Defocus and (ii) tilt-geometry estimation, (iii) lattice determination, (iv) correcting for crystal imperfection, termed unbending, (v) Fourier extraction of amplitude and phase values from computed reflections, (vi) Contrast transfer function (CTF) correction and (vii) generating a projection map of one unit cell at higher SNR. So far, this procedure was applied to the drift-corrected average gained from a movie-mode exposure of a 2D crystal and thus did not account for locally varying beam-induced sample movements.

In the present study we describe an algorithm, termed movie-mode unbending, that combines the approaches listed above for 2D crystals, by correcting for locally varying beam-induced sample deformations and computationally optimizing the electron dose for each frequency, at the same time reducing the risk of overfitting the data at low SNR. The new procedure is implemented as an additional unbending script in 2dx [Gipson *et al.* 2007b](software available for download at `www.2dx.org`). It first processes the globally drift-corrected average images, and then uses the determined parameters to refine the processing with movie-mode unbending of individual subsets of the globally drift-corrected movie-frames. We present the application of the new software to an experimental dose-fractioned membrane protein dataset recorded on a FEI Titan Krios equipped with a Gatan K2 summit detector, which yielded an improved 3D reconstruction of the ion-channel MloK1. The quality of the images, quantified by a single number per image, the so-called QVal, was increased ∼30% on average over the whole test dataset.

4.2 Approach

Application of the here presented algorithm to 2D crystal images confirms the hypothesis of different local movements within regions of a single 2D crystal during

imaging (Figure 4.1). We observed local movements of up to 40Å within micrometer-sized 2D crystal regions during the exposure required to record a movie (typically featuring a total dose of 40 electrons/Å^2 distributed over 40 frames). This observation underlines the need for per-frame movie-mode image processing for high-resolution electron crystallography.

4.2.1 Correction of local movements during the exposure

The major step in single image processing for 2D crystallography is correcting crystal imperfections, termed unbending (detailed in [Arheit *et al.* 2013c]). A Fourier-filtered reference is generated based on the crystallographic lattice determined in reciprocal space, and cross-correlated with the micrograph. Peaks in the resulting cross-correlation profile (localized by the MRC program Quadserch) correspond to crystallographic unit-cell locations. The MRC program CCUnbend is subsequently used to translationally adjust small patches (e.g., 25×25 pixels in size) of the micrograph to maximize the regularity of the crystal in order to achieve high- resolution projection maps. The image unbending is usually refined over several rounds, resulting in an improvement and sharpening of the computed diffraction spots. To avoid malicious noise accumulation or the domination of reference bias, the raw image is never unbent multiple times, but instead a so-called distortion-vector field or "ERROR field" is continuously improved. The ERROR field obtained in the previous round of unbending is used as the starting solution for the current round of refinement. Generally after three rounds of refinement, the final ERROR field is applied to unbend the raw micrograph in one step. The Fourier-transform of the unbent image is then evaluated to obtain the amplitude and phase values at the lattice reflection sites, and the data are CTF-corrected. If only global drift correction factors (x- and y-shifts) were applied to each frame before the averaged raw image was generated this unbending procedure will give a lower-resolution result, since all local sample movements within the 2D crystals are ignored.

The fundamental idea of movie-mode unbending is to process each frame of a movie individually and thereby correct for beam-induced crystal distortions that occurred while the image stack was recorded. This involves (i) correctly exploiting the smooth motion of consecutive frames, (ii) accounting for radiation damage during the increasing exposure (Section 4.2.2), (iii) dealing with the extremely low SNR of one movie-frame, and finally (iv) preventing of overfitting (Section 4.2.3). Again, the use of global factors alone is not sufficient.

Our implementation (Figure 4.2) starts by processing the average image obtained when a translational drift-correction is applied to each frame of the movie. This single averaged image has a high SNR and is used to determine the defocus, tilt-geometry and crystallographic lattice for the 2D crystal. The Fourier-filtered unbent

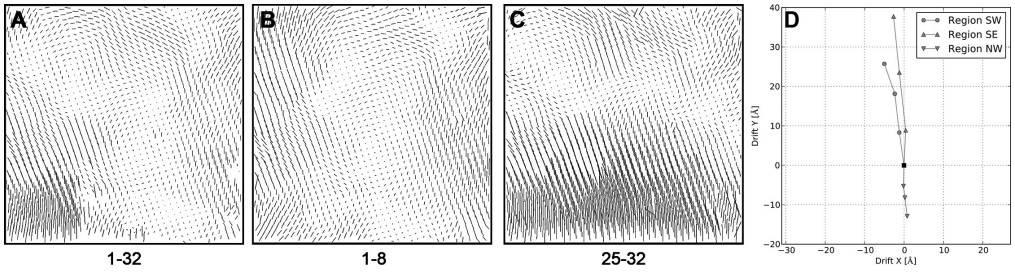

Figure 4.1: Electron dose dependent variation of distortion-vector ERROR-fields showing the necessity of beam-induced motion-correction. (A) Unbending profile (4k × 4k pixels) obtained by processing the average image generated by real-space averaging of all 32 movie-frames. For display purposes, the motion vectors have been elongated ten times. By averaging eight subsequent frames we obtained a new movie consisting of only four frames, showing the crystal at different exposure times (averaging frames 1-8, 9-16, 17-24, and 25-32, after discarding the first two and the last 6 frames of the entire dataset). The four resulting averaged frames were then individually unbent with the algorithm introduced in Section 4.2 and illustrated in Figure 4.2. The distortion-vector ERROR field of the first average (frames 1-8) is shown in (B), the one for the fourth average (frames 25-32) in (C). (D) Drift-profiles quantifying the averaged local movements over the whole sequence of frame averages relative to the first frame average. Fifty neighboring trajectories were averaged for each quadrant of the images. The drift of the northeast quadrant of the crystal was omitted because of visualization difficulties due to negligible local drift (< 2Å). The lower area (southern quadrants) of the crystal is drifting upwards during the exposure whereas the movements of the upper part are smaller and in another direction. An approach that corrects for sample movements at frame level by translationally aligning entire frames is not able to correct for such inhomogeneous and electron dose-dependent local crystal deformations.

image produced by the 2dx_image script "Unbend II" is used to generate a high-SNR reference image. The cross-correlation map between this reference image and the raw image indicates the unit-cell positions in the crystal. The MRC program Quadserch is used to localize peaks in the cross-correlation map, thereby determining the crystal unit cell locations. Their deviations from the predicted, ideal unit cell locations are determined by Quadserch and represented by the lattice-distortion ERROR field.

Once the above is complete, the frames of the movie are processed individually, following the same approach. However, several parameters such as defocus, tilt geometry, crystal lattice, as well as the reference map for cross correlation with each movie frame, are automatically forwarded to movie-mode unbending from the processing run described above. In the version distributed with 2dx (Figure 4.2), the program 2dx_quadserch can consecutively refine the determined ERROR field. When processing the first frame of a movie stack, we use the ERROR-field generated by the script "Unbend II" as the starting ERROR field for refinement. For all subsequent frames, the ERROR field generated while processing frame (n-1) is used as the starting ERROR field for the peak localization when unbending frame (n). This approach exploits the smooth correlation between the crystal distortions among subsequent movie-frames to ensure a continuous and physically comprehensible motion-correction. Finally the MRC program CCUnbend is used to correct each movie frame according to its corresponding ERROR field by shifting small image patches. The unbent images are then averaged and corrected for non-uninform Fourier space sampling, yielding one combined output image. Using the same global reference image for all frames ensures that the unbent frames are in register, which is a prerequisite for the final real-space averaging step. The MRC program MMBoxA is finally employed to evaluate the averaged unbent frames, and to produce one APH-file from the averaged individually-unbent frames. Among other things, this APH-file contains amplitude, phase and SNR information for all computed diffraction spots. After correcting for CTF-effects, a 2D projection map is generated and finally the resulting APH-file can be used for 3D merging.

4.2.2 Accounting for resolution dependent radiation damage

As demonstrated in detail for catalase crystals by Baker et al. [Baker *et al.* 2010], electron beam damage is a crucial limit to high-resolution in cryo-EM. Based on the electron dose-dependent fading of diffraction spots, they showed how the optimal electron dose depends on the targeted resolution. For instance, to record data with an optimal SNR, a target resolution of 3Å requires a cumulative electron dose of 11 electrons per $Å^2$, while a target resolution of 27Å requires a total dose of 22 electrons per $Å^2$. Generally, a higher cumulative electron dose destroys the high-resolution

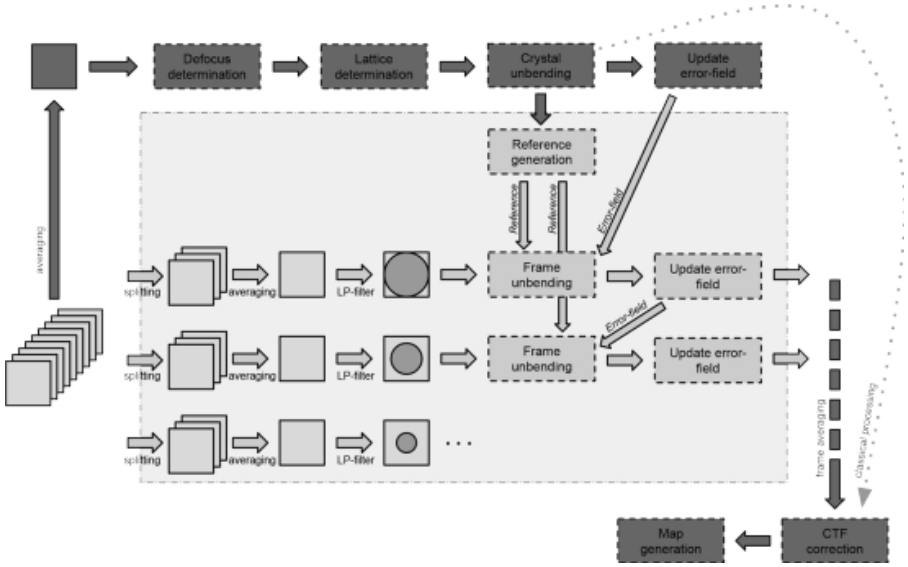

Figure 4.2: Movie-mode unbending algorithm implemented in 2dx. Initially the drift-corrected stack of movie-frames is averaged and classically processed as illustrated in the uppermost pathway [Arheit *et al.* 2013c]. Subsequently subsets of movie-frames are extracted from the full stack, averaged into so-called frame averages and optimally low-pass filtered to account for radiation damage (Section 4.2.2). Based on the previously obtained Fourier-filtered and unbent image, a reference image is generated, which is later used for the sequential unbending of the frame averages. Using the same reference for all frame averages ensure the translational alignment of the resulting unit cells. The distortion-vector ERROR field used to unbend the averaged image is used as an initial approximation to determine the ERROR field to unbend the first frame average. The later frame averages are unbent based on the ERROR fields refined for the preceding average. Finally all unbent frame averages are combined into one unbent image by weighted averaging in Fourier space, taking the resolution limitations applied to the frame averages into account. After CTF-correction, a single projection map is generated from this movie, and is then ready for later 3D merging.

details, but amplifies low-resolution features, such as overall shape and position of the particles.

DED movie-mode data acquisition allows dose fractionation, i.e., the distribution of a high electron dose over a large number of movie-frames. In this work, we recorded data with a cumulative dose of 40 electrons/Å^2 distributed over 40 movie-frames. Averaging a certain number of movie-frames during image processing, allows average images corresponding to various desired electron doses to be generated. In practice, this has made it possible to assess the impact of electron beam damage on different amino acids [Allegretti *et al.* 2014]. Nevertheless, such a-posteriori averaging of movie-frames still suffers from the trade-off between the optimal electron dose required to obtain the low and high-resolution information present in the images. [Baker *et al.* 2010, Bammes *et al.* 2010] investigated the dose-dependent fading of computed diffraction spots of thin 3D catalase crystals with 2dx. We conducted the same measurements of dose-effects for lipid membrane 2D crystals (Figure C.2) to verify that the previous findings for thin 3D crystals are applicable here. Based on the previous studies and our measurements we can apply optimal low-pass filters that vary with the accumulated electron dose. In this way, we can ensure that all structure frequencies contribute to the resulting projection map at their optimal dose-dependent SNR.

For the first couple of frames, we let all frequencies contribute fully to the average, whereas for the later frames, where some high-resolution components of the structure have suffered from radiation damage, only the lower frequencies contribute to the average. The cut-off frequency $r(N)$ of the low-pass filters, expressed in absolute frequency units within the interval $[0, 0.5]$, was chosen such that $r(N) = A\exp(BN)$, where N equals the frame number. The coefficients A and B where chosen according to the decay reported by Baker et al. [Baker *et al.* 2010]. Based on the assumption that for our dataset each movie-frame corresponded to an electron dose of approximately 1e/Å^2, we set $A = 0.8$ and $B = -0.05$. Although the parameters were determined by Baker et al. [Baker *et al.* 2010] for a different kind of crystal, our experiments (Figure C.2) confirm that they are a good approximation for the here presented application. This procedure improves the stability and reliability of frame-wise crystal unbending by ensuring that all information present in the image frames is optimally available for processing.

Sequential low-pass filtering of movie-frames prior to frame-wise unbending is an efficient way to computationally optimize the electron dose for each resolution range. Application of an appropriate weighing scheme while finally averaging the damage compensated frames is required to ensure correct sampling in Fourier space and to not disturb the subsequent 3D merging. This allows EM movies to be recorded with a higher cumulative dose than usual to ensure that no low-resolution information is "missed".

4.2.3 Prevention of overfitting

The tendency to accumulate noise into interpretable structural elements is a major problem of iterative algorithms used in cryo-EM. This so-called overfitting arises from the combination of the low SNR generally present in cryo-EM images and over optimistic resolution cutoffs. When we applied the above algorithm to all movie-frames of an image-stack consecutively (electron dose of $\sim$ 1 electron per Å^2 per frame) we observed a serious overfitting issue indicated by discontinuous distortion-vector-fields and artificial ripples in the resulting projection maps (Figure 4.3C). Due to the low signal present in each frame, the cross correlation profile lacked distinct correlation peaks, which significantly reduced the reliability of the MRC program Quadserch to track the correct crystal unit cell locations. Increasing the ambiguity of cross correlation peaks can only be achieved by increasing the SNR of the raw data or the reference image. Using a larger reference would give a higher SNR for the cross-correlation map. However, cross-correlation with a reference that is too large reduces the possibility to track fine crystal distortions i.e., any distortions that are smaller than the reference dimensions. On the other hand, increasing the SNR of the frames by using a higher electron dose per frame leads to beam-induced blurring within the frames caused by sample instabilities in the EM. Thus, a compromise has to be made. Here, we chose to average a given number of consecutive movie frames and process the resulting sub-frame averages like a movie. Such frame averages overcome the SNR-induced limitations described above. However, averaging multiple sub-frames and thus reducing the "electron dose resolution" (see above) limits the possibility to capture all beam-induce motion. Nevertheless, the situation can be improved by applying a global frame-drift correction to each low SNR frame prior to the averaging step to compensate for global stage drift and homogeneous beam-induced sample movement.

This poses the question how many frames should optimally be averaged and processed together. We tested the movie-mode unbending algorithm with all conceivable numbers of averaged sub-frames and using the calculated parameter QVal [Gipson *et al.* 2007b] to assess the quality obtained (Figure 4.3A); a higher QVal value indicates that the image has more and/or sharper diffraction spots. In addition, we inspected the distortion ERROR fields and the projection maps obtained to see if overfitting could be visually detected (Figure 4.3C-E). Each of the movies analyzed consisted of 40 frames with a cumulative electron dose of 40 electrons/Å^2. As cryo- EM samples have the tendency to move most at the beginning of the exposure, we completely omitted the first two frames. Further, the last six frames were only included in the initial high-SNR average calculated after global drift correction of the individual frames. To avoid the effects of beam-induced sample damage, these six frames were not otherwise used. Hence we fully processed 32 frames of the movies with a cumulative electron dose of $\sim$ 34 electrons/Å^2, including the two

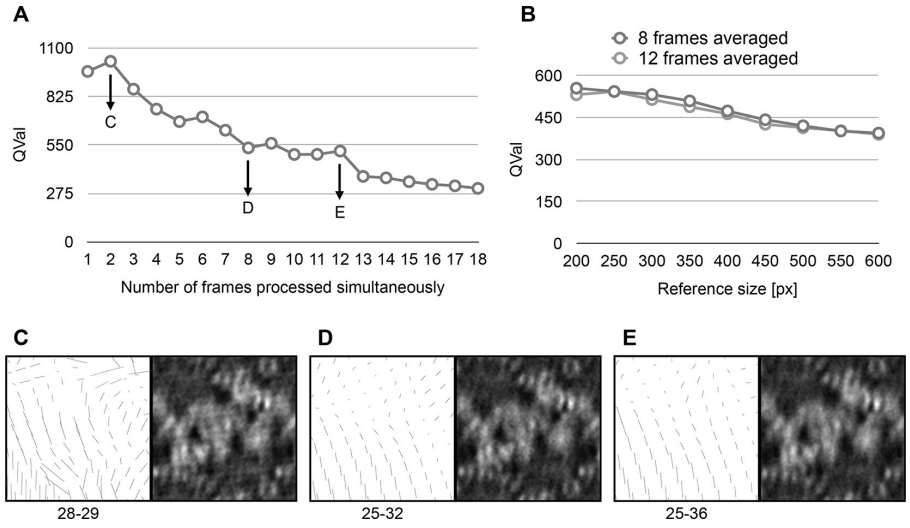

Figure 4.3: Development of movie-frame unbending using the MloK1 test dataset. For this test dataset classical processing resulted in a QVal of 232.2. (A) Relationship between the number of averaged movie-frames and the measured QVal. During drift-correction at frame level, the two first frames were removed and thus omitted from the movie- mode unbending. Three different regimes were found: overfitting regime when 1-7 frames were averaged, indicated by artificially high QVals; working range of the processing procedure when 8-12 frames were averaged; and stiffness region when 13- 18 sub-frames were averaged, where the detection of beam-induced motion-correction is limited due to large number of averaged frames. (B) Impact of the reference size on the measured QVal analyzed for two different averaging schemes. (C-E) Visual comparison of the distortion-vector ERROR field (left) and the resulting projection map (right) when processing averages consisting of 2 frames (C), 8 frames (D), or 12 frames (E) were processed. The chaotic ERROR field in C is accompanied by striation artifacts in the image. D shows an optimal approach. E shows increased blurring in the image, due to a too smooth ERROR field.

omitted frames at the beginning of each movie. This procedure complies with the settings previously used for movie-frame processing of single particle cryo-EM samples [Allegretti *et al.* 2014, Bai *et al.* 2013]. Note that ensuring equal divisibility of the data for the analysis meant that the total number of frames included varied slightly. The minimal changes in the total electron dose this caused, did not influence the results significantly. On analyzing the effect of averaging different numbers of frames before movie-mode unbending, we observed three different regimes where the performance and reliability of the latter procedure differed significantly (Figure 4.3A). There were signs of overfitting when only a small number (1-7) frames were averaged and processed, and the procedure was no longer able to resolve the beam-induced motion accurately enough if too many (13-18) frames were averaged. Averaging 8-12 frames gave optimal results. Movie-mode unbending of the resulting movies comprised of 4 or 3 averaged frames, did not produce any signs of overfitting and still allowed beam-induced crystal alterations to be reduced. Although our observations were consistent over the whole MloK1 test dataset, the optimal batch size presumably depends on the sample as well as on the imaging conditions. It must be stressed that although the final movie used for movie-mode unbending was only comprised of 4 high SNR frame averages, the 40 individual frames originally recorded were essential to allow global drift-corrections to be made over a fine dose raster. This important correction would not be possible if just 4 higher dose, and thus higher SNR, frames were initially recorded per movie.

A similar experiment to the above was performed to determine the optimal reference size. As expected, Figure 4.3B confirms that the use of a larger reference leads to a lower QVal. From analyzing the QVal measurements for different reference sizes and visually inspecting the distortion-vector ERROR fields, we conclude that for our dataset a reference size of 300×300 pixels ($\sim 3\times \sim 3$ unit-cells) prevented overfitting and allowed beam-induced sample deformations to be accurately corrected. The final, algorithm including frame averaging, optimal low-pass filtration, and sub-frame-average movie-mode crystal unbending is illustrated in Figure 4.2.

4.3 Results and discussion

Cryo-EM images of 2D crystals of the cAMP-modulated potassium channel MloK1 (130×130 Å unit cell, P4212 symmetry, vitrified on thin carbon film supported by holey carbon film) were used to test the presented algorithm. This four-fold symmetric ion channel undergoes conformational changes upon cyclic adenosine monophosphate (cAMP) binding [Kowal *et al.* 2014]. Here, we used the ligand-free (closed channel) conformation (PDB Model: 4CHW).

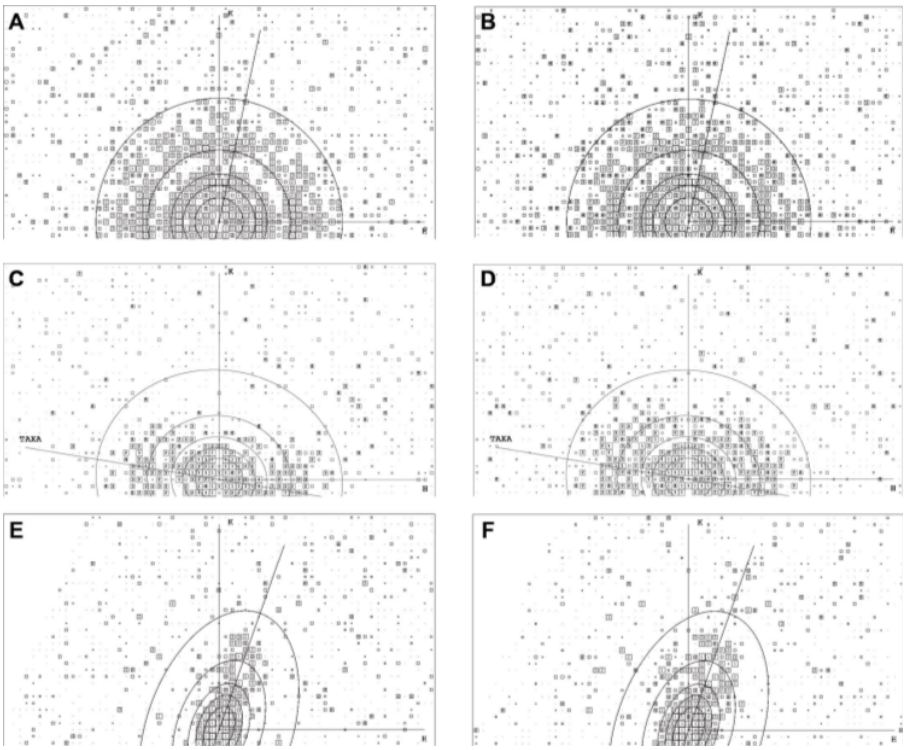

Figure 4.4: Canonical IQ-plot obtained without and with movie-mode unbending applied to crystals with different tilt angles. The Nyquist frequency of these plots is a 4Å; the resolution circles are at 36Å, 24Å, 18Å, 12Å and 7Å. (A) IQ-plot of a non-tilted 2D crystal without movie-mode unbending, i.e. unbending of the drift-corrected sum of all movie frames (QVal 650.7). (B) IQ-plot of the same exposure as in (A) but after movie-mode unbending (QVal 746.7). (C) IQ-plot of a 20° tilted crystal without movie-mode unbending (QVal 232.2) and (D) IQ-plot after movie-mode unbending (QVal 536.1). IQ-plot of a 50° tilted crystal (E) without movie-mode processing (QVal 370.0) and (F) IQ-plot generated by means of the new unbending procedure (QVal 523.4). Especially orthogonal to the tilt axis (labeled with "TAXA"), the IQ-plots after movie-mode processing show significantly more spots and also spots at higher resolution.

4.3.1 Data acquisition and classical image processing with 2dx

We recorded 237 dose-fractioned movie-sequences using a FEI Titan Krios equipped with a Gatan K2 summit detector (without energy filtration), operated in super-resolution electron counting mode, yielding movies of 8k × 8k images. Images of crystals with a nominal sample tilt of up to 50° were recorded. The microscope was operated in low-dose mode at a nominal magnification of 22,500x at screen level, resulting in an effective magnification of ∼37,000x on the detector. The resulting physical pixel size was 1.34Å on sample level, respectively 0.67Å for the super- resolution pixels. Movies consisting of 40 frames were recorded over 16 seconds total exposure, resulting in a 0.4 second exposure for each frame. The total electron dose per movie was ∼ 40 electrons/Å^2, respectively ∼ 1 electron/Å^2 per frame. Micrograph defoci varied between -0.6 and -4.5μm. To avoid undercounting [Li *et al.* 2013b], the electron dose rate was set to 5 electrons per physical pixel per second (respectively 1.25 electrons per super-resolution pixel per second). The images were corrected for frame-drift on-the-fly and automatically processed by 2dx_automator as detailed in [Scherer *et al.* 2014]. The best 97 movies were used for 3D merging, including tilt-geometry refinement by means of the MRC program Origtilt, resulting in a 3D reconstruction that we chose to limit to 5Å in plane and 7Å out of plane resolution (Figure 4.5A). The best movies were selected on the basis of quality of the computed diffraction patters, amount of visible structural features in the unit-cells and the phase-residual values. The resolution limit was set according to the calculated resolution dependent phase residuals and the plausibility of the fitted lattice lines.

4.3.2 Experimental procedure: Movie-mode unbending for 2D crystals

The 97 best movies were processed as described above. Frames 3-34 where movie-mode unbent. Based on our findings with the test data set (Figure 4.1 & Figure 4.3), batches of eight movie frames were averaged and the reference used for the unbending was 300 × 300 pixels in size. All generated distortion-vector ERROR plots were manually checked to avoid overfitting. In eight cases a larger reference (400 × 400 pixels) was used to prevent the accumulation of noise. For four movies, all of relatively small crystals, movie-mode unbending did not increase the QVal. Hence, for each of these we used the unit-cell obtained from the average of all 40 frames by classical methods in the final 3D merging step. An additional round of iterative phase-origin refinement was performed to obtain an improved 3D reconstruction of the 97 movie-mode unbent crystals.

4.3.3 Improvements due to movie-mode unbending

Figure 4.4 shows the quality improvement due to movie-mode unbending for different crystals featuring different tilt angles. The QVal of these particular crystals were significantly increased by the new unbending procedure. Over all test images we measured an average QVal increase of $\sim 30\%$. Particularly movie-mode unbending increased the completeness of observed reflections from 64.9% to 73.4% (Table C.1). The impact of sub-frame motion correction on a resulting crystallographic unit cell is shown in Figure C.1.

The correspondingly improved quality of the 3D density map in terms of visible details and resolved densities is shown in Figure 4.5. To avoid biasing the comparison of movie-mode and classical unbending on the same dataset, we did not change the applied resolution limits during processing and the same contour level was used for the visualization of both maps. Even with these restrictions, the intra-cellular region of the S6-helix and the S3-S4-linker (Figure 4.5B&E) as well as the S6-helix in the pore- region (Figure 4.5C&F) are much better resolved after movie-mode unbending and more easily interpretable when referring to the fitted atomic model. The real-space refinement program DireX [Schröder *et al.* 2007] was used to adapt the previously determined atomic model (PDB-4CHW, [Kowal *et al.* 2014]) to the improved cryo- EM map from movie-mode processing.

Finally an additional round of phase origin refinement including tilt geometry refinement was applied with higher resolution limits, yielding an again improved 3D reconstruction (Figure C.4, Table C.1) with significantly easier interpretable structural details (Figure 4.6).

4.3.4 Algorithmic alternatives

Averaging batches of sequential movie frames to reduce the risk of overfitting is the key point of the presented algorithm. The drawback is that processing these averages produced distortion-vector ERROR fields that are a compromise for the averaged frame. The distortion-vector ERROR plots calculated for different averaged frames differed significantly from one averaged frame to the next. Due to the discrete frames that contributed to each averaged frame, these differences did not reflect smooth underlying physical movements. An alternative would be to compute so-called “rolling averages” of the frames, where each frame would contribute to multiple average frames. The average built for frame (n) would then also include down- weighted contributions from frames $(n-2)$, $(n-1)$ and frames $(n+1)$, $(n+2)$, etc., which Deleted: sub-could be used with a Gaussian weighting scheme centered on frame (n). The determined ERROR field of each average could then be applied

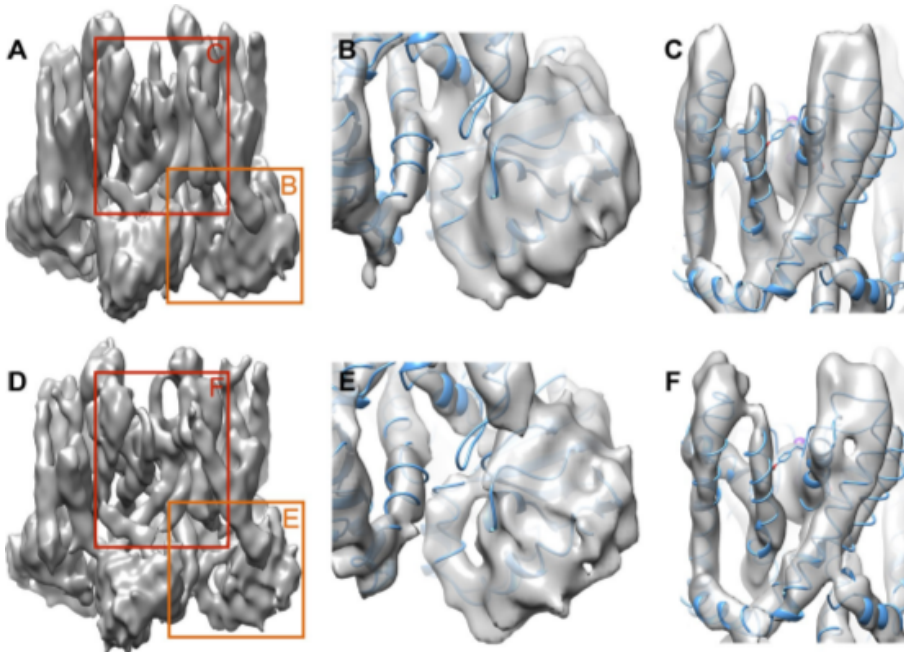

Figure 4.5: Comparison of classical and movie-mode unbending based on the 3D reconstruction of MloK1. Both maps are generated from the same dataset and visualized using the same contour level and resolution limits. (A-C) The map obtained by classical processing. (A) Structural overview. (B) Closer look at the cyclic nucleotide-binding domain (CNBD), the cytoplasmic part of the S6 helix and the S3-S4-linker (C) Membrane view of the selectivity filter. (D-F) The map obtained by movie-mode unbending applied to the same dataset. (D) Structural overview. (E) The region shown in (B); note that the linker, the cytoplasmic part of the S6 helix and the channel side-facing region of the CNBD are more clearly separated. (F) Membrane view on the selectivity filter as in (C); note that the S6-helix is more clearly separated within the trans-membrane part of the ion-channel.

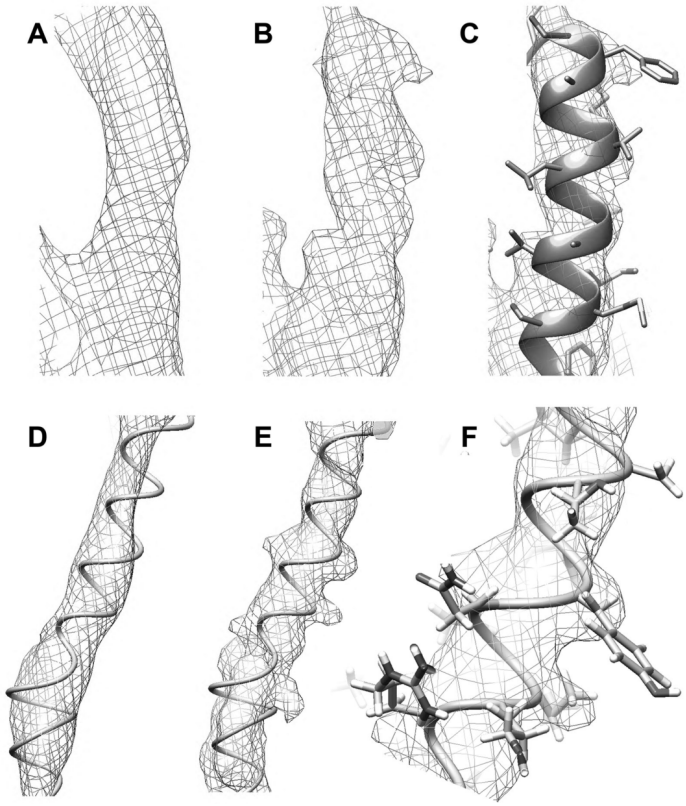

Figure 4.6: Structural improvements due to movie-mode unbending on selected protein structure areas. (A) Classically unbent reconstruction of the extracellular end of the S6-helix next to the selectivity filter filtered at 5/7Å (in plane / out of plane). (B) higher resolution (4.2/5.5Å) reconstruction of the same region based on movie-mode unbent crystallographic unit cells. (C) Movie mode reconstruction with atomic model flexibly fitted from PDB-4CHW. (D) Intracellular domain of the S6-helix reconstructed from classically unbent crystal images (resolution 5/7Å) and (B) reconstruction based on movie-mode unbending (resolution 4.2/5.5Å). (F) Atomic model fitted into the middle part of the helix shown in (E). The correct helical pitch is clearly visible in the movie-mode reconstruction.

to the single central frame of the rolling average only. Preliminary experiments with this method applied to the presented test dataset did not yield better results than the algorithm described before, possibly because rolling average images might not have been sufficiently dominated by the central frame to produce a valid ERROR field for that frame. At the extremely low SNR of the individual frames recorded here, computing ERROR fields that are valid for just one frame failed.

4.4 Conclusions

The described algorithm is the first movie-mode processing procedure for 2D crystals. It offers an efficient and generic approach to correct for beam-induced translational motion within the frame, which previously limited the quality of 3D reconstruction obtained. We prevent overfitting by averaging multiple drift-corrected movie frames and computationally optimize the electron dose steps and employ filters to ensure reliable processing. We carefully investigate the properties of the two key parameters of the new algorithm namely the "number of frames to average" and "reference size". Application of the new software to a real-world dataset significantly improved the results. However, contrary to our hopes, the improvement was not sufficient for amino acid side chain structures to be visible. The resolution in this project might be limited by, (i) limited sample flatness, (ii) inhomogeneous arrangement and conformations of individual proteins with the crystals, and (iii) presence of a large missing cone due to limited sample tilt (effect was analyzed in Supplementary Figure 4.5 and a solution to this was introduced in [Gipson *et al.* 2011], but was not included here).

Furthermore we applied the here introduced unbending method to a different type of 2D crystals featuring a higher amount of internal order and flatness. Remarkably the movie-mode unbent computed diffraction patterns were of the same quality then the ones produced by electron diffraction (Figure C.3). We previously presented a single particle approach for the analysis of 2D crystals to overcome resolution loss due to limited sample flatness [Scherer *et al.* 2013]. Implementation of a movie-mode single particle method is now needed to enable full 3D tracking of the 2D crystals and correction for the physical movements of the membrane proteins forming them.

Acknowledgments

We thank Shirley A. Müller and Nikhil Biyani for insightful discussions and critically reading of the manuscript, and Marcel Arheit for fruitful discussions in the early project stage. This work was supported by the Swiss National Science Foundation (grants 315230_146929, 205320_144427, and the NCCR TransCure). All the here-

described procedures are implemented in the 2dx package, which is available under the GNU Public License on www.2dx.org.

Supplementary data

Supplementary data associated with this article can be found in C.

5 Applications

The methods detailed in the three previous chapters were successfully applied to multiple current research projects in our lab. Brief image processing focused summaries of these projects are given in this chapter. First we show the impact and benefit of the novel automatic crystallographic image processing pipeline on the basis of three different cryo-electron crystallography studies. Second we detail the use of the real-time processing pipeline in context of two high-resolution helical image processing projects. Finally the implementation of an automated high-throughput single particle cryo-electron microscopy pipeline is illustrated. A high-resolution density map proves correct functioning of our equipment and our high-performance image processing pipeline.

Contents

5.1 Structure of the receptor-binding protein of phage T5 bound to its Escherichia coli receptor FhuA

Bacteriophages are viruses that infect bacteria and replicate within them. Structurally, the bacteriophage T5 is composed of a head, a collar and a tail. To attack the host cell, the tip of the bacteriophage's tail, termed the receptor binding protein (RBP) binds to a specific receptor in the bacteria's outer membrane. Upon receptor binding the viral DNA, originally stored in the phage's head, is transferred via the phage's tail into the host, where new phages are assembled. Finally, after being released by cell lysis, the newly built phages start to infect other bacteria.

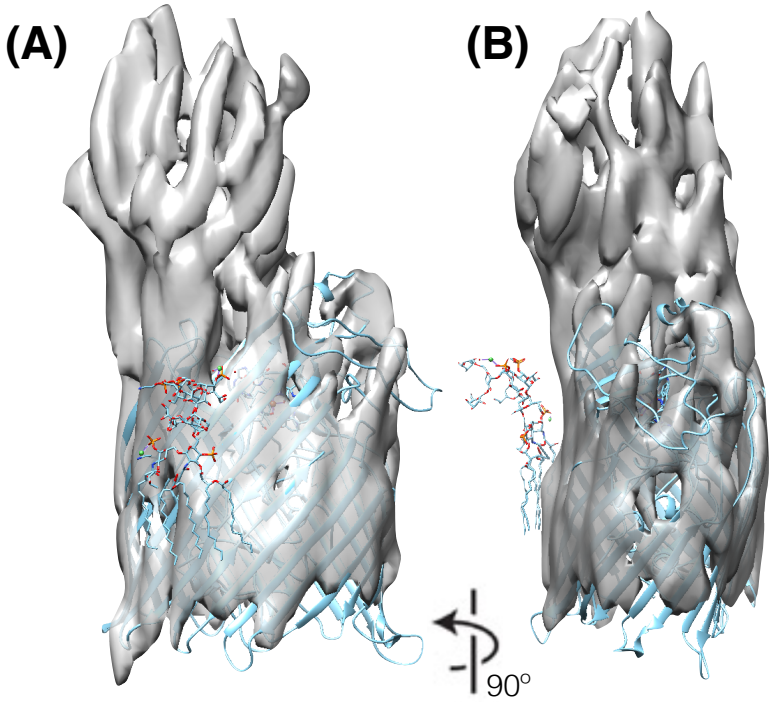

Figure 5.1: Preliminary 3D reconstruction of the FhuA-pb5 complex by electron crystallography. Resolution was estimated in 2dx to be $\sim$ 6Å in xy-plane respectively $\sim$ 12Å in z-direction. The atomic model of the membrane-embedded FhuA (PDB-1QJQ) was fitted in UCSF Chimera [Pettersen *et al.* 2004] by means of rigid-body fitting: (A) side-view of the FhuA-pb5 complex, (B) 90° rotated side-view.

In case of the bacteriophage T5, the receptor binding protein pb5 binds to the outer membrane protein FhuA of Escherichia coli. In general, little is known about the very first steps of infection of bacterial viruses, particularly the structure of the receptor binding protein pb5 is unknown. The conformational changes caused by receptor binding are of particular interest, because receptor binding is thought to trigger the release of viral DNA [Flayhan *et al.* 2012, Basit *et al.* 2012].

Our collaborators were able to crystallize the FhuA-pb5 complex in 2D. The obtained crystals feature a remarkably high quality in terms of their orderdness and flatness. On some Fourier transforms of raw images, which were not even corrected for crystal imperfection, diffraction spots at around 3Å were visible. However the particular upside down arrangement of the elongated FhuA-pb5 monomers in the 2D crystals made 3D merging a challenging task. Already minor tilt geometry changes resulted in completely different 2D projection maps, which made the manual initialization of the 3D refinement processes impossible for a long time. Previous attempts of structural investigations of the same crystals failed due to the image quality obtained on photographic film. Our new direct electron detector in combination with the automation pipeline introduced in Chapter 3 and careful manual 3D merging, allowed to obtain a reasonable 3D reconstruction at moderate resolution (Figure 5.1). Both, the possibility to record and process a large number of images in a short time and the higher resolution information captured by a direct electron detector were key to success in this project.

The observed overall shape of the FhuA-pb5 complex (Figure 5.1) confirms previous negative stain electron microscopy results [Breyton *et al.* 2013]. Acquisition of much more images, correction for beam-induced motion (Chapter 4) and computational restoring of information in the missing cone [Gipson *et al.* 2011] will significantly increase the resolution and quality of the 3D map. A high-resolution reconstruction of the FhuA-pb5 complex should be suited for de-novo model building of pb5 and thus will lead to new insights about receptor binding of the phage T5.

5.2 Assessment of different sample preparation methods for cryo-electron crystallography

In general there are two different widely used sample preparation methods for cryo-electron crystallography (Section 1.4.3): plunge-freezing and trehalose embedding. Careful optimization of all parameters of the sample preparation procedure for any new crystallography project is indispensable. When optimizing the sample preparation conditions for two previous cryo-electron crystallography projects in our lab [Kebbel *et al.* 2013, Kowal *et al.* 2014] we had the impression that trehalose embedding negatively impacts the quality of the projection maps and leads to 3D reconstructions that are hardly or not at all interpretable. To further understand these

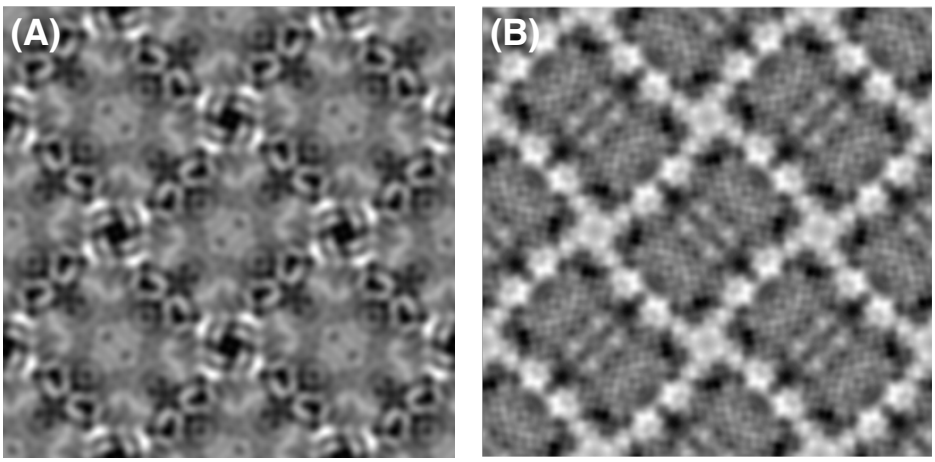

Figure 5.2: Assessing different sample preparation methods by reference to different non-tilted MloK1 crystal projection maps: (A) 7Å symmetrized projection map of a plunge-frozen MloK1-crystal. (B) Symmetrized projection map of a trehalose embedded MloK1-crystal at 7Å resolution.

observations, we investigated the effect of different sample preparation methods on the resulting maps in this side-project. Therefore potassium channel MloK1 crystals [Kowal *et al.* 2014] were cryo-prepared in different ways and imaged by means of our novel automation setup (Chapter 3). Before this automatic processing pipeline was available, the analysis of multiple different data sets was a time-consuming process. Based on our automation software 2dx_automator [Scherer *et al.* 2014] we were able to obtain medium-resolution projection maps of differently prepared samples within a few hours. Our initial 2D results suggest that trehalose degrades the cyclic nucleotide binding domains (CNBD) of MloK1 (Figure 5.2). In general we believe that trehalose embedding can easily introduce artifacts in the resulting 3D reconstruction, which is particularly severe for studies of structurally completely uncharacterized proteins. Thus we preliminary conclude that trehalose embedding should be used carefully or ideally even be avoided completely. Most likely such a systematic analysis of sample preparation methods within a reasonable amount of time would not be possible without the pipeline developed in Chapter 3.

5.3 pH-induced conformational changes in the E. coli chloride/proton antiporter ClC-ec1

Although the structure of the chloride/proton antiporter ClC-ec1 has been determined more than a decade ago [Dutzler *et al.* 2002], its molecular ion-transport mechanism remains unclear. In this study we used cryo-electron crystallography

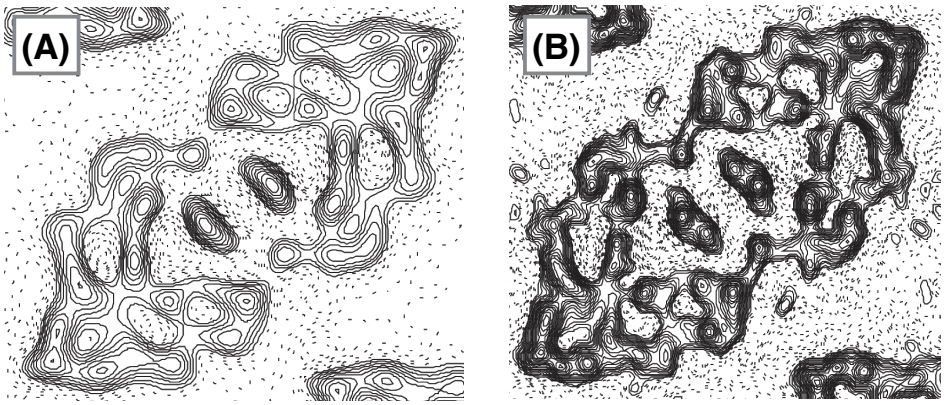

Figure 5.3: ClC-ec1 2D projection maps: (A) Projection map from 11 non-tilted ClC-ec1 crystals recorded on photographic film. Quality evaluation in 2dx reported a resolution of $\sim$ 8Å. (B) 4.2Å projection map from 7 non-tilted crystals recorded on a Gatan K2 summit direct electron detector. Image processing was done automatically with 2dx_automator.

to investigate pH-induced conformational changes in ClC-ec1 dimers that are artificially embedded into a lipid bilayer mimicking their native cellular environment. Therefore the ClC-ec1 2D crystals were imaged at different pH-values using film as recording medium. The resulting projection maps were analyzed with an in-house developed tool [Arheit *et al.* 2015] that reliably detects significant differences between two families of projection maps. The study was augmented by a 3D reconstruction of ClC-ec1 at moderate resolution and by a molecular dynamics simulation predicting the molecular function of the protein. However these film-based initial results did not answer the biological questions. Therefore data acquisition was repeated on a direct electron detector, which was installed in our lab in the meantime. Processing of the newly acquired images was successfully automated by means of 2dx_automator (Chapter 3). Once again we obtained a significantly better result (Figure 5.3) in a fraction of time. Due to the very tiny expected conformational changes, an unambiguous interpretation of the three-dimensional helix movements based on two-dimensional projection maps was not possible. Producing two 3D reconstructions of ClC-ec1 at two different pH-levels is indispensable for the successful continuation of this promising project. The limited crystal quality makes recording of high-quality images almost impossible, particularly for highly tilted crystals. The movie-mode unbending algorithm, presented in Chapter 4, is a promising novel approach to overcome heterogeneous beam-induced motion currently limiting the quality of the projection maps. Nevertheless this project illustrates another successful application of the automation pipeline developed in Chapter 3.

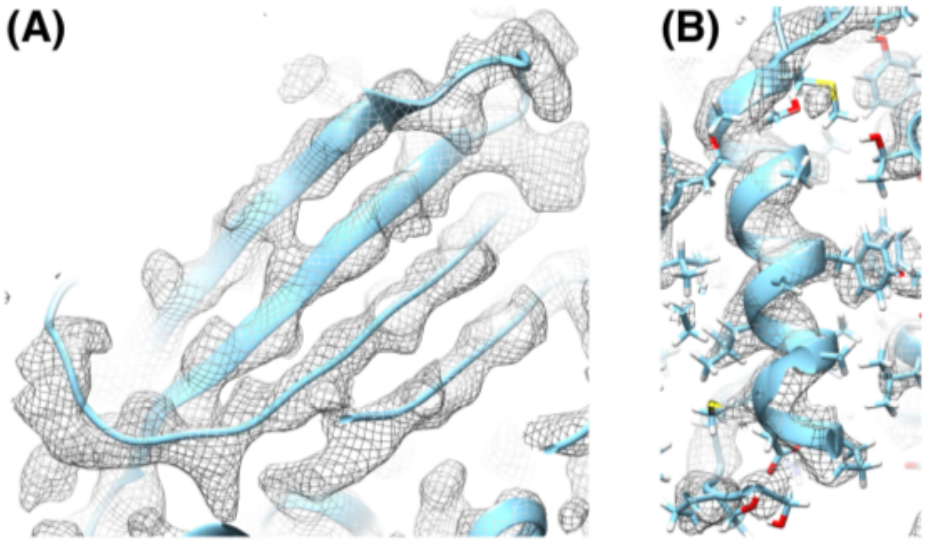

Figure 5.4: Type VI secretion system contractile sheath map highlights. The atomic model in light-blue (PDB-3J9G) was built into the experimental three-dimensional density map (EMDB-2699, resolution 3.4Å): (A) Clearly resolved collection of beta-sheets (hand-shake domain, see [Kudryashev *et al.* 2015]). (B) High-resolution helical region featuring clearly visible side-chains.

5.4 The structure of the type VI secretion system contractile sheath solved by cryo-EM

Bacterial secretion is the process of transport or translocation of effector molecules (e.g. toxins or enzymes) from the bacterial cell into its target cell. The type VI secretion system (T6SS) uses fast contraction of a long sheath to transport effectors accross biological membranes. Our study reports the atomic structure of the entire T6SS sheath for the first time in history. Previous X-ray studies uniquely revealed models for individual sheath associated proteins in their monomeric state, which did not explain the true biological assembly. Although a cryo-electron microscopy density map became available recently [Kube *et al.* 2014], the atomic model of the T6SS sheath remained unclear due to the limited resolution of the released map. By means of direct electron detector based helical image processing [Egelman 2007], we obtained a high-resolution reconstruction ($\sim$ 3.4Å, [Kudryashev *et al.* 2015]) of the sheath in its contracted conformation by means of helical image processing. The superior quality of the 3D density map (Figure 5.4) enabled atomic model building, which allowed understanding of sheath assembly at atomic level.

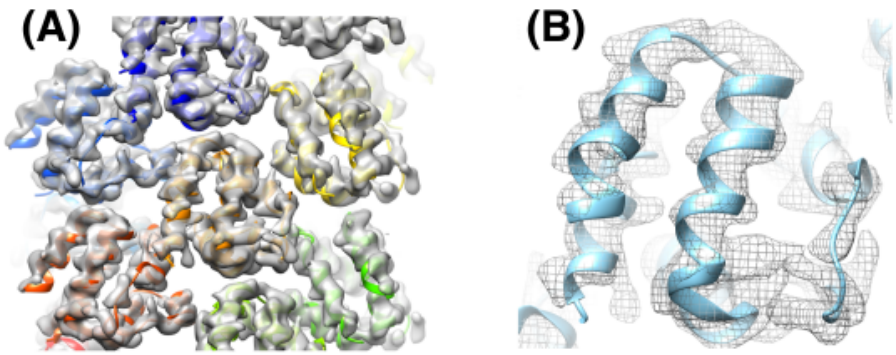

Figure 5.5: Cryo-electron microscopy map of the ASC-PYD filament at ∼ 3.6Å resolution: (A) Helically assembled ASC-PYD filament (PDB-3J63) rigid-body fitted into the experimental density map. (B) Overlay of the 3D reconstruction and the atomic model (ribon representation) for one monomer of the ASC-PYD filament.

The real-time drift-correction, which belongs to the automation pipeline described in Chapter 3, was used during data acquisition. In fact this study is the first - in a list of many - successful applications making use of 2dx_automator's drift-correction module for non-crystallographic samples. For this particular study I contributed to data acquisition and image processing. Our findings were published in:

[Kudryashev *et al.* 2015] *Kudryashev, M., Wang, R.Y-R., Brackmann, M., Scherer, S., Maier, T., Baker, D., DiMaio, F., Stahlberg, H., Egelman, E.H., and Basler, M., 2015.* ***The structure of the type six secretion system contractile sheath solved by cryo-electron microscopy.*** *Cell, vol. 160, no. 6, pages 952-962.*

5.5 Structure of the ASC inflammasome filament by combined NMR spectroscopy and cryo-electron microscopy

The real-time drift-correction was successfully used for another high-resolution helical image processing project - the structure determination of the ASC-PYD inflammasome filament. Inflammasomes are part of the nonspecific immune system. They are responsible for the immune response by activating the caspases cascade leading to cell pyroptosis, a process of programmed cell-death. The so-called ASC-PYD protein plays a key role in imflammasome assembly. A recent cryo-electron microscopy study [Lu *et al.* 2014a] revealed the structural template of the human ASC-PYD filament. In our collaborative project we studied the structure of the ASC-PYD filament from mice. Therefore a new protocol, combining solid-state

NMR and cryo-electron microscopy, was developed. Technically, the high-resolution cryo-electron microscopy reconstruction at $\sim$ 3.6Å resolution (Figure 5.5) was used to derive initial distance restraints for the NMR analysis [Sborgi *et al.* 2015]. This powerful new approach allowed studying the filament's morphology and its atomic architecture. The corresponding manuscript is currently under review:

[Sborgi *et al.* 2015] *Sborgi, L., Ravotti, F., Dandey, V., Dick, M., Mazur, A., Reckel S., Chami, M., Scherer, S., Böckmann, A., Egelman, E., Stahlberg, H., Broz P., Meier, B., Hiller, S,. 2015.* ***Structure and assembly of the mouse ASC filament by combined NMR spectroscopy and cryo-electron microscopy.*** *(submitted).*

This project shows that the real-time drift-correction (Chapter 3) is a powerful, lightweight and versatile tool for image acquisition on a direct electron detector. Beside the two exceptionally high-resolution projects detailed above, we use the software for all images that are acquired manually on our Gatan K2 summit camera.

5.6 Implementation of a high-throughput high-resolution automatic single particle pipeline

Indicated by the large number of near-atomic resolution reconstructions deposited into the electron microscopy data bank (Figure 1.5), single particle cryo-electron microscopy became the standard high-resolution electron microscopy tool. For example in 2014, single particle analysis revealed 24 out of 36 density maps at a resolution equal or below 4Å. Thus it is indispensable for a leading and innovative electron microscopy lab to have a high-throughput acquisition and processing pipeline at its disposal. In particular when dealing with flexible protein complexes or inhomogeneous samples, recording and processing a large number of particles (typically more than 200'000) is required, which makes automatic image acquisition indispensable.

There are multiple well-tested electron microscopy automation tools available, e.g. SerialEM [Mastronarde 2003], Leginon [Carragher *et al.* 2000] or vendor specific software solutions. After carefully exploiting the different options, we decided to setup a Leginon-based automation system for all our high-throughput microscopes. There were three major reasons why we preferred Leginon over the alternatives: (i) all microscopes share one common backend (i.e. Storage, Database and Webserver), which simplifies maintenance significantly, (ii) the seamless integration of Appion [Lander *et al.* 2009] naturally provides real-time image processing capabilities directly within Leginon and (iii) previous experience promised a short installation time.

Today, most of the high-resolution single particle cryo-electron microscopy projects make use of Relion [Scheres 2012b]. The statistical image processing approach, implemented in this software package [Scheres 2012a], prevents overfitting by construction, computationally deals with multiple conformations and ensures reliable structure determination. Recently the developers introduced statistical processing of dose-fractionated movie sequences in Relion-1.3 [Scheres 2014a]. This development further improved the achieved resolution by accounting for beam-induced particle motion and radiation damage. Due to the complex underlying optimization problem, Relion requires a lot of computational recourses. Typically the software runs on centralized high-performance computing clusters.

Below we give an overview about the different building blocks of our Leginon installation and high-resolution image processing pipeline. We mainly distinguish three different elements: (i) microscope controllers, (ii) backend services and (iii) high-performance computing image analysis tools.

Microscope controllers Usually two dedicated computers are used to control the electron microscope and the camera. Automated image acquisition with Leginon requires integration of an additional workstation that is connected to both, the microscope computer and the camera device controller. This additional workstation is used to plan, control and execute an automatic image acquisition session. A typical Leginon session consists of four phases: (i) recording of a grid atlas at low magnification to select optimal grid squares, (ii) selection of holes from a medium magnification image on square level, (iii) identification of optimal exposure and focus positions for each hole and finally (iv) high-magnification focusing and image acquisition. Depending on the used hardware and its acquisition mode, it is possible to record up to a few hundred images per hour with our setup.

Leginon backend The Leginon backend in our lab consists of multiple components such as storage servers for long term storage and caching of dose-fractionated exposures, a webserver for image quality assessment and remote supervision of acquisition sessions, powerful processing servers for real-time image processing, a GPU-server for real-time drift-correction and a database storing various meta-information about imaging sessions, microscopes and acquired images. Typically after optimizing the imaging condition for an acquisition session, the user launches multiple real-time processing jobs, such as CTF-estimation or automatic particle picking, in the background. If data is recorded on a direct electron detector in dose-fractionated mode, the recorded stacks have to be drift-corrected. Similar to our previous semi-automatic pipeline (Chapter 3), Leginon makes use of the motion-correction software developed by Li et al. [Li *et al.* 2013a]. After an imaging session, the webserver is used to select good images, which will be used for the subsequent image processing.

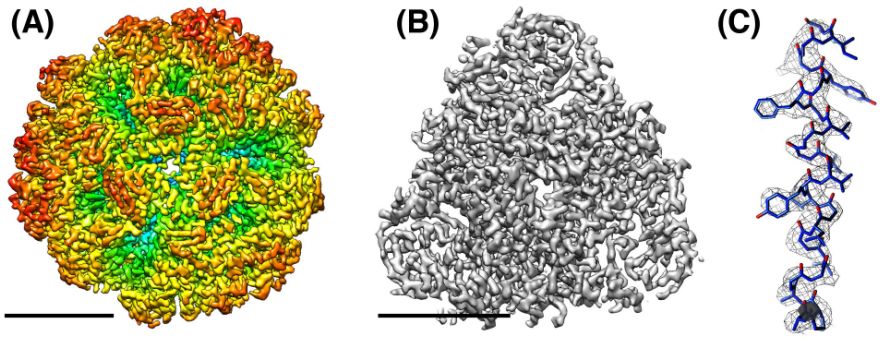

Figure 5.6: Urease density map at 3.43Å resolution from ∼7'900 particles (scale bar 50Å) : (A) Biological assembly featuring a tetrameric symmetry. Color scheme according to radial distance from the center of the complex. (B) Isolated three-fold symmetric building block of the overall assembly. (C) Exemplary high-resolution density map with fitted atomic model produced by *AutoBuild* from the *Phenix*-package [Adams *et al.* 2010].

High-performance computing image processing Today most of the near atomic single particle reconstruction are computed by means of Relion [Scheres 2012b]. In collaboration with the *Center for Scientific Computing at the University of Basel*[3] we installed Relion-1.3 on their high-performance computing infrastructure. Initial tests run successfully and future hardware extensions are planned.

In order to exploit the baseline of our automation setup, we manually recorded ∼ 60 images of cryo-prepared Urease from Yersinia enterocolitica[4]. The best 46 images (automatically drift-corrected with the previous automation setup (Chapter 3)) were used for image processing. ∼12'300 particles were semi-automatically picked using EMAN2's *swarm-picking* tool [Tang *et al.* 2007]. All subsequent image processing steps, including defocus estimation based on CTFFIND3 [Mindell & Grigorieff 2003], were done in Relion-1.3 [Scheres 2012b]. ∼8'100 particles were selected from "shiny" 2D classes. Initial 3D classification was performed with three different 3D models. The observation that most particles (∼7'900) clustered in one 3D class indicates a high homogeneity among the particles. A first high-resolution automatic 3D refinement revealed a reconstruction at 3.66Å resolution. The movie-mode polishing procedure, which is implemented in Relion-1.3 [Scheres 2014a], improved the final resolution of the 3D map to 3.43Å (Figure 5.6).

Although the images were recorded manually, these results are an important step towards the correct functioning of our automation pipeline as they prove correct

[3] http://scicore.unibas.ch
[4] generously provided by Torsten Schwede's lab at the Biozentrum, University of Basel

functioning of our sample preparation equipment, our microscope, our camera and our high-performance computing image processing pipeline. Additionally the obtained map will serve as playground for future learning about atomic model building and quality evaluation. However, final hardware calibrations are still required to enable automatic acquisition of images featuring the same quality than our manually recorded reference data set.

6 Conclusions and Outlook

Major aim of this thesis was to design and implement algorithms that increase the throughput and resolution of cryo-electron crystallography using direct imaging of the specimen. Therefore three novel image processing approaches were developed and successfully applied to various biological projects.

Conventional image processing of 2D crystals assumes perfectly flat crystals, which are almost impossible to grow in practice. We developed a single particle processing approach for 2D crystals that accounts for out-of-plane tilt on individual protein level. This study conceptionally shows that single particle refinement approaches applied to 2D crystals moderately improve the resolution of 3D reconstructions from images recorded on photographic film. However with the introduction of direct electron detectors, cryo-electron crystallography obtained a boost in resolution, similar to all other cryo-electron microscopy techniques. By correcting for beam-induced sample motion it was possible to overcome the resolution limit at around 7Å, which was previously observed on film. Thus we know today that out-of-plane tilt was likely not the most severe limiting factor for film-based electron crystallography. Nevertheless out-of-plane tilt is a severe limiting factor that prevents the generation of 3D reconstructions at even higher resolution based on today's new technology. High-resolution electron microscopy requires careful optimization of algorithmic details such as advanced interpolation schemes, accurate CTF-correction procedures or gridding techniques for back-projecting [Penczek *et al.* 2004]. As our initial proof of concept was not designed for today's resolution, a successful application to datasets recorded on direct electron detectors was not possible so far. However our initial study is an important cornerstone for future method developments in this direction. Connecting 2dx with standard single particle processing software such as FreAlign [Grigorieff 2007] or Relion [Scheres 2012b] seems to be the most sustainable and promising approach for the future.

Direct electron detectors, recently introduced to the field of cryo-electron microscopy, are superior in many different ways: (i) the signal-to-noise ratio of the recorded images, (ii) the possibility of dose-fractionated exposure series and (iii) the complete digital nature of the whole imaging and image processing workflow. After receiving our first direct electron detector, we realized the need for automatic image processing for 2D crystallographic data. The developed automation pipeline, termed 2dx_automator, features an automatic drift-correction procedure for dose-fractionated exposures based on the software developed by Li et al. [Li *et al.* 2013a], augmented with automatic 2D crystal image processing capabilities including an intuitive image quality assessment tool. As most of the required parameter tuning only depends on the used imaging system, setting up the automation for a new 2D crystal project is straightforward. However 2dx_automator is a generic automation

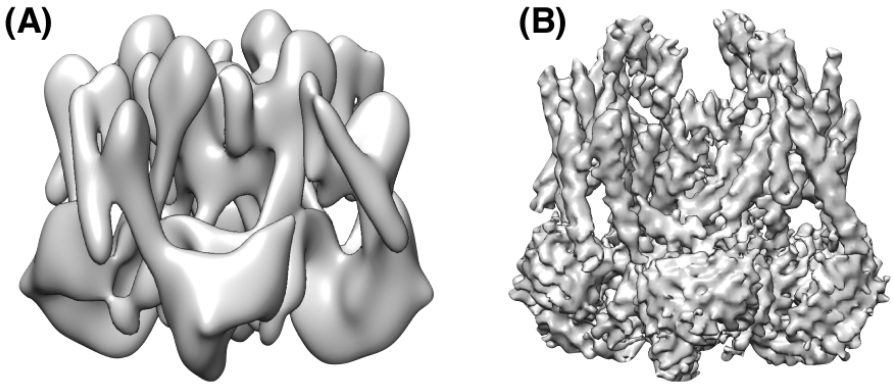

Figure 6.1: Resolution improvements in cryo-electron crystallography over the course of this thesis: (A) State of the art cryo-electron crystallography in 2012; MloK1 reconstruction (7/12Å resolution) from images recorded on photographic film [Kowal *et al.* 2014]. (B) Direct electron and novel image processing approaches based 3D reconstruction of MloK1 in 2015 featuring a resolution of 4/6Å.

tool that was successfully used for various internal and external electron crystallography projects. The real-time drift-correction module was used for many different projects in our lab as well as in a few other labs worldwide.

When processing images of 2D crystals recorded on a direct electron detector, we learnt that beam-induced sample movements are the major limiting factor in electron crystallography today. Impact of beam-induced motion is particularly dramatic for images of highly tilted crystals. The motion-correction approach introduced by Li et al. [Li *et al.* 2013a], does correct for beam-induced motion on the level of entire frames. The drift-correction module of our automation pipeline (Chapter 3) also uses the software by Li et al. Our observations on a large number of images suggested that different areas of a 2D crystal under the electron beam move differently in terms of absolute distance and direction. To correct for these non-linear locally varying motion patterns, we developed a method called movie-mode unbending that efficiently deals with translational crystal disorders and beam-induced locally varying deformations. The developed algorithm was applied to a dataset consisting of almost thousand images in the meantime. The resulting 3D reconstruction featured a resolution of approximately 4Å in-plane, respectively 6Å in z-direction (Figure 6.1). The fact that some amino acid side chains were resolved partially emphasizes the quality of the resulting 3D map.

Although electron crystallography is a powerfull method to determine the structure of small membrane proteins in their native environment, the vast majority

of today's cryo-electron microscopy projects use single particle analysis techniques. Thus it is indispensable for a leading cryo-electron microscopy lab to own a high-throughput high-resolution single particle cryo-electron microscopy pipeline. Based on an automation framework called Leginon [Carragher *et al.* 2000], a powerful and customized single particle cryo-electron microscopy pipeline was installed during this thesis. Initial results confirm correct functioning and benefits of our new state of the art setup for single particle samples. The developed pipeline is available for a wide range of collaborations.

Currently there are multiple factors limiting the resolution in electron crystallography. Even when correcting for beam-induced motion by movie-mode unbending, the maximal tilt angle is typically limited to $50-60°$. Thus a large portion of the Fourier space is not sampled at all, which leads to the formation of the so-called missing cone. Gipson et al. [Gipson *et al.* 2011] published a method that computationally restores the missing information in the missing cone. However the method by Gipson et al. requires a lot of parameter tuning. Thus this method was only applied successfully to one particular dataset. Future adaptions of the existing approach will allow us to restore the missing cone of any arbitrary electron crystallography reconstruction. Additionally this generic approach allows to elegantly by-pass the obsolete lattice line fitting step (Section 1.4.5), which does not perform appropriately on current high-resolution datasets.

Following Zhang and Zhou [Zhang & Hong Zhou 2011] an accurate defocus correction is a must for high-resolution cryo-electron microscopy. Accurate defocus correction for images of tilted crystals is another unsolved problem in electron crystallography. In practice such an image features a defocus gradient perpendicular to the tilt axis. Usually the reference used for unbending is extracted from the center of the image. Thus there are regions within an image of a tilted crystal featuring different defocus values compared to the reference that is cross-correlated with the image. The different locations of Thon rings (dependent on the actual z-height) will lead to an inaccurate cross-correlation profile and thus negatively impacts the processing results. Currently we are experimenting with an approach that corrects for CTF-effects in a stripe-wise fashion prior to unbending in order to increase image processing quality for tilted samples.

In general as cryo-electron microscopy is rapidly improving its power in terms of resolution, more and more previously ignored limiting factors become important [Zhang & Hong Zhou 2011]. For instance although the beam of an electron microscope is thought to be parallel in theory, there is a phase-shift induced by a slightly tilted beam. Beside inaccurate CTF-correction schemes for high frequencies, beam-tilt is a major limiting factor. Computationally correcting for beam-tilt and other resolution limiting effects is an efficient approach to further improve electron crystallography's resolution.

Even though 2D crystallization is often an extremely difficult and time-consuming task, electron crystallography will have its "raison d'être" in the future. Mainly because there is no other technique that can study small (<120kDa) membrane proteins in their native environment at high-resolution. Combined with the prediction that near-atomic reconstructions from 2D crystals will be available on a regular basis in the future, we are convinced that electron crystallography will survive in long terms in its niche.

7 Acknowledgments

First of all I want to thank my thesis supervisor **Henning Stahlberg** for giving me the opportunity to work on this thesis in his lab. Thank you for your guidance, support, motivation and freedom to pursue my own scientific ideas. I very much enjoyed our hours of discussion about science.

I would like to thank also my PhD advisory committee members and co-supervisors **Volker Roth** and **Torsten Schwede** for their support during the last three years. **Timm Maier** thank you for chairing my examination.

Thanks to **Bryant Gipson, Xiangyan Zeng** and **Marcel Arheit** for their initial work on 2dx. Your work was a great foundation for this thesis.

Julia Kowal, Fabian Kebbel, Priyanka Abeyrathne and **Mohamed Chami** taught me the essential basics of electron crystallography and structural biology. Thanks a lot for your patience.

I very much appreciated the guidance and senior support on image processing from **Misha Kudryashev** and **Daniel Castaño-Diez**. Thank you for sharing your knowledge and challenging my ideas.

Without **Ariane Fecteau-LeFebvre, Kenny Goldie** and **Bill Anderson**, keeping our microscopes up and running, this thesis would not have been possible.

Many thanks to **John Ryan** and **Christian Placzek** from the BSSE-IT for their technology assistance and support.

Without the support and computing power, provided by **Martin Jacquot, Pablo Lopez** and **Konstantin Arnold**, I would still be waiting for my results. Thank you for providing the essential computing power.

Thanks to **Shirley Müller** for carefully proofreading all scientific publications of this thesis.

I could additionally enjoy numerous inspiring discussions with **Philippe, Mark, Karen, Raphael, Stefan, Andrej, Benjamin, Simone, Thomas, Paul, Jarek, Sarah, Kushal, Venkat, Nikhil, Rosmarie, Nadja, Roger, Jan, Stefan, Claudio, Christel, Alexandra, Christopher, Cedric, Dominic, Christoph, Nora, Luc** and **Simon**. Thanks for been such nice colleagues.

Finally I would like to thank my **family and friends** for their support and understanding during my PhD-studies. Thank you for your moral encouragement.

Appendices

A Supplementary Material for Single Particle 3D Reconstruction for 2D Crystal Images of Membrane Proteins

A.1 Supplementary Methods

A.1.1 Local CTF correction

The CTF is corrected in Fourier space by means of phase flipping according to the CTF values given by Scherzer's formula [Scherzer 1949]:

$$\mathrm{CTF}(\vec{k}) = -\omega_1 \sin(\chi(\vec{k})) - \omega_2 \cos(\chi(\vec{k})) \tag{A.1}$$

with

$$\chi(\vec{k}) = \pi\lambda k(\Delta f(\vec{k}) - \frac{1}{2}\lambda k^2 C_s) \tag{A.2}$$

where Δf denotes the defocus value, $\vec{k}$ the scattering vector, λ the wave length and C_s the spherical aberration constant of the electron microscope. We follow the convention that a negative value for the defocus indicates an underfocus. ω_1 and ω_2 are the percentage contributions of phase contrast and amplitude contrast, respectively, to the total CTF. Phase- and amplitude contrast are related through:

$$\omega_1 = \sqrt{1 - (\omega_2)^2} \tag{A.3}$$

Following [Toyoshima & Unwin 1988] we use a phase contrast contribution ω_1 of 99.75%, thus an amplitude contrast contribution ω_2 of 7%. As images of tilted crystals lead to astigmatism in the CTF of the images, the total defocus Δf is given by:

$$\Delta f = \frac{1}{2}\Big(\Delta f_1 + \Delta f_2 + (\Delta f_1 - \Delta f_2)\cos(2(\alpha_l - \alpha_{\mathrm{ast}}))\Big) \tag{A.4}$$

where Δf_1 and Δf_2 are the respective defocus values in the two perpendicular directions, α_k denotes the angle between the first defocus direction Δf_1 and the x-axis on the micrograph, and α_{ast} describes the angle between the scattering vector $\vec{k}$ and the x-axis. For images of tilted 2D crystals, the defocus values vary across

the image depending on the specific position of the particle and the tilt geometry of the crystal according to:

$$\Delta F_i = \sqrt{x^2 + y^2}\sin(\beta - \theta)\tan(\alpha) + \overline{\Delta F_i}, \qquad i \in [1, 2] \tag{A.5}$$

where $\overline{\Delta F_i}$ is the defocus at the center of the image in one of the two defocus directions, x and y are the coordinates of the particle relative to the center of the micrograph, the tilt angle α, β is the angle between the line from the origin and the particle and the x-axis, and θ is the angle between the tilt axis and the x-axis. The two central defocus values and the amount of astigmatism are determined for each crystal by means of 2dx on the raw micrograph. 2dx internally uses the program CTFFIND3 [Mindell & Grigorieff 2003]. 2dx additionally allows interactive fine-tuning and verification of the values found by the automatic processing.

A.2 Supplementary Material

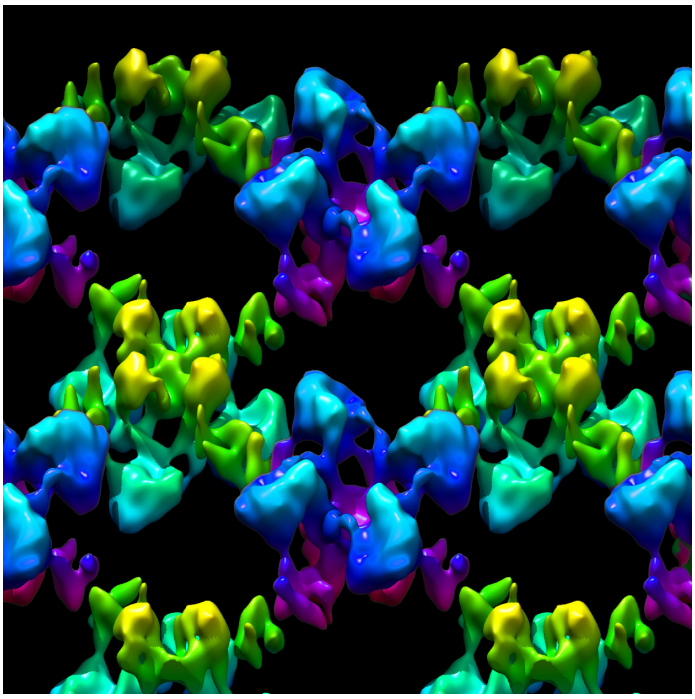

Figure A.1: The packing of MloK1 tetramers in the P_{4212} symmetry 2D crystal. This image shows 2×2 unit cells of the 3D reconstruction, obtained by the classic MRC-based processing within 2dx. Four-fold symmetric tetramers are arranged in an alternating head-to-tail packing, of which the upside-down oriented tetramers are shown in blue colors.

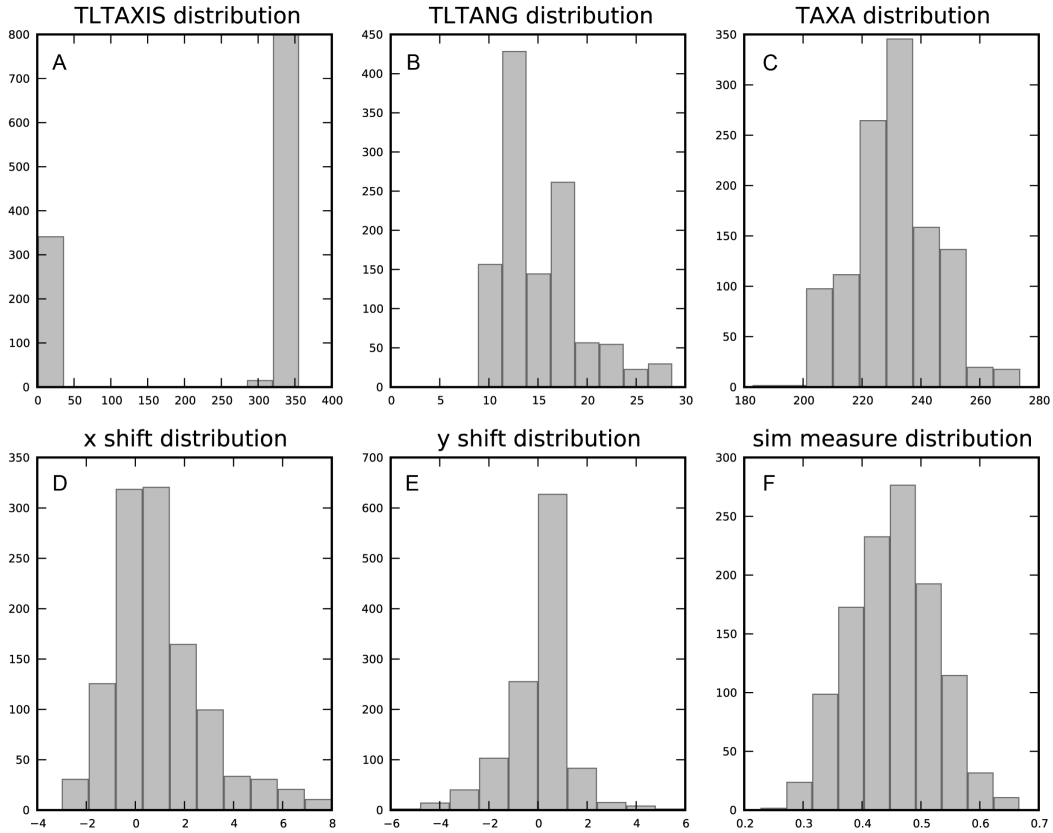

Figure A.2: Probability distributions of the six particle fingerprint parameters for one selected image of 15° nominal tilt: (A) Tilt axis, (B) Tilt angle, (C) TAXA (in-plane orientation of the crystal on the grid), (D) x-shift, (E) y-shift and (F) maximal cross-correlation value.

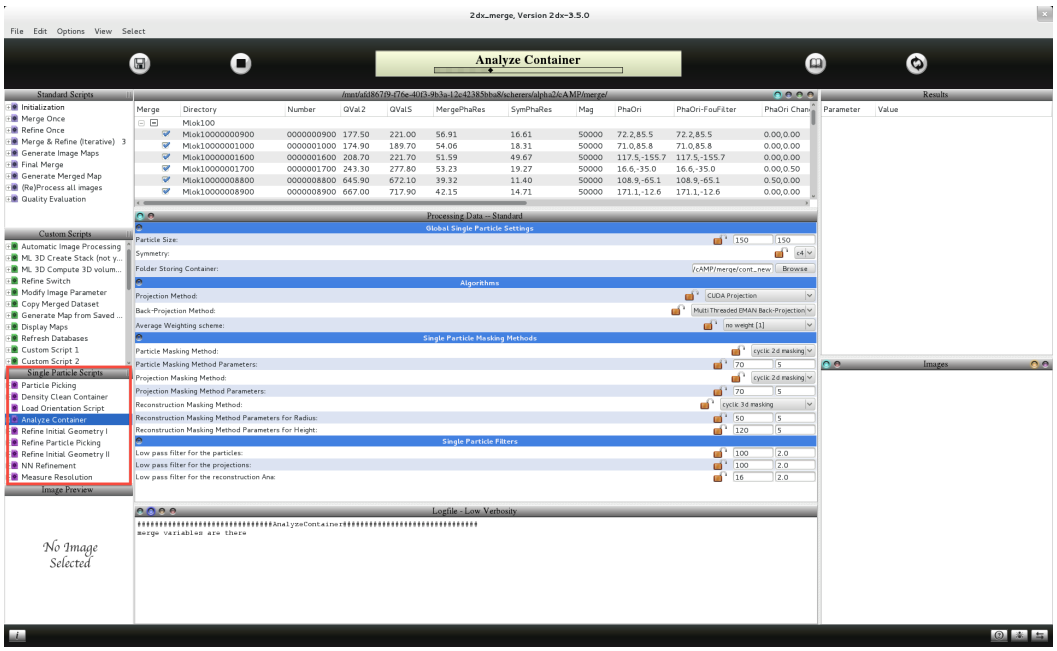

Figure A.3: Updated 2dx_merge graphical user interface with the new single particle script section (red box in the left column).

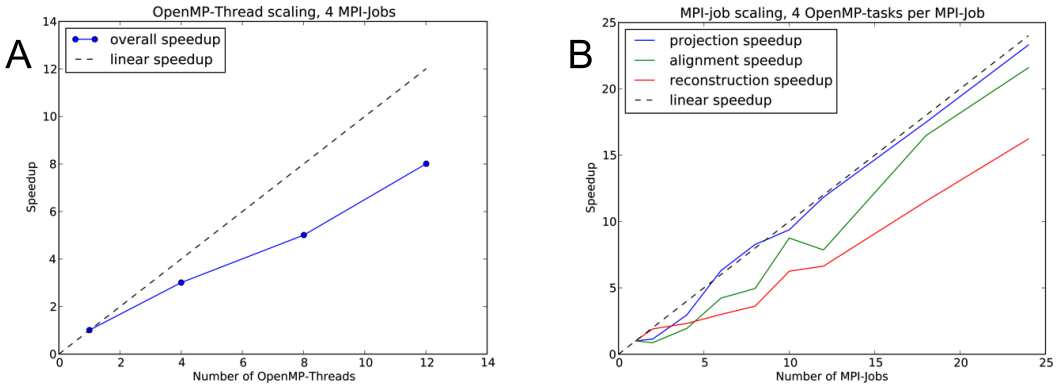

Figure A.4: Scaling analysis of the cluster implementation. (A) Number of shared memory tasks plotted against the observed speedup. Measurement of the one refinement step of 60'000 particles on four AMD Opteron 6174 12-core nodes. (B) Distributed memory parallelization scaling measured on AMD Opteron 8380 4-core processors.

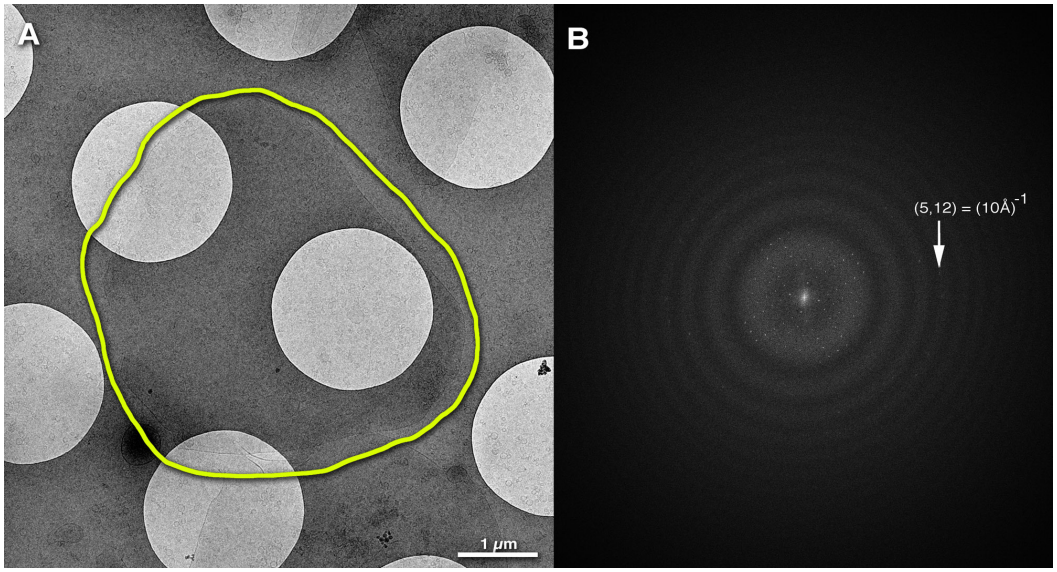

Figure A.5: 2D crystals of MloK1. (A) Representative overview image of a frozen hydrated 2D crystal, vitrified in buffer solution on a holey carbon film grid. The yellow circle indicates the borders of the faintly visible 2D membrane crystal. (B) Calculated Fourier transform (displayed as power spectrum) of an unprocessed high-resolution cryo-EM image recorded over a hole in the holey carbon film. The corresponding image was recorded on photographic film (Kodak SO-163), and digitized with a Heidelberg Primescan drum scanner. This unprocessed powerspectrum shows Thon rings and crystal lattice reflections out to order (5,12), which corresponds to a resolution of 1/(10Å).

B Supplementary Material for 2dx_automator: Implementation of a Semiautomatic High-Throughput High-Resolution Cryo-Electron Crystallography Pipeline

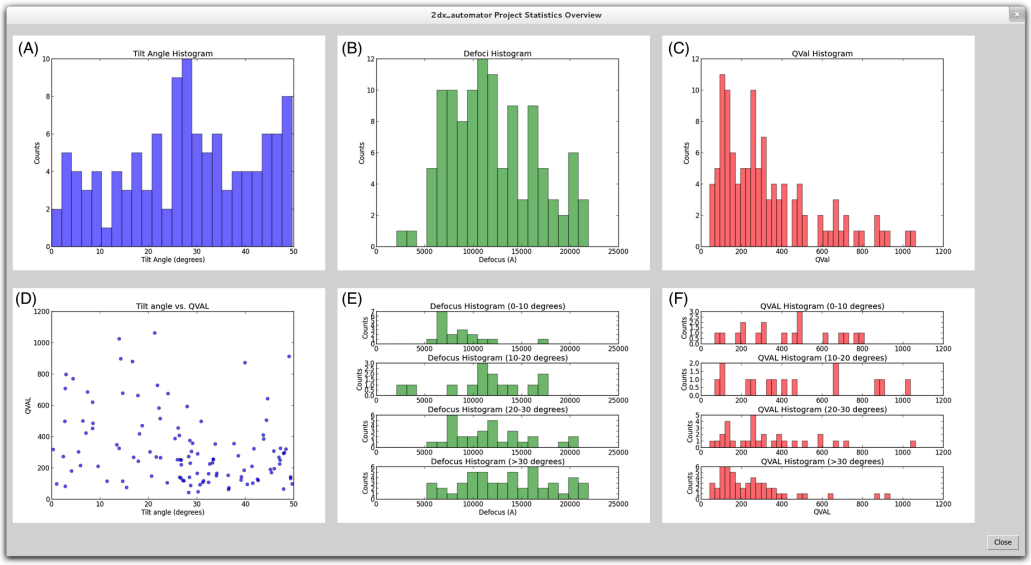

Figure B.1: Project statistics overview for 148 recorded images. (A) tilt angle histogram, (B) Defocus histogram, (C) QVal histogram, (D) QVal plotted against tilt angle, (E) tilt angle depended defocus histograms and (F) tilt angle depended QVal.

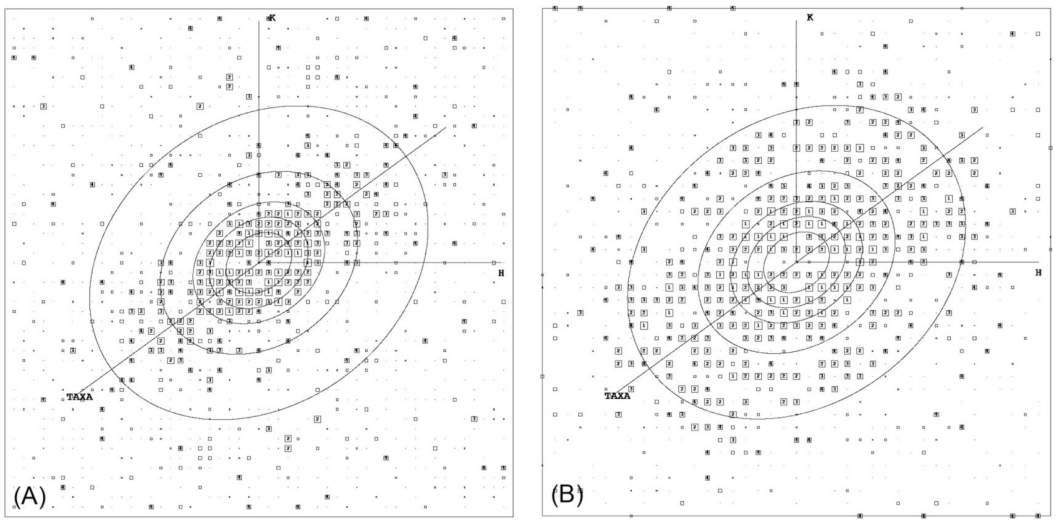

Figure B.2: Drift correction applied to a 40° tilted 2D crystal. (A) Resolution circle plot (circles correspond to 36Å, 24Å, 18Å, 12Å, and 7Å) from the canonical lattice calculated from a MloK1 2D crystal at 40° tilt without motion correction. (B) Resolution circle plot of the same crystal, but here the frames where previously aligned and the first two dose-fractioned images were removed.

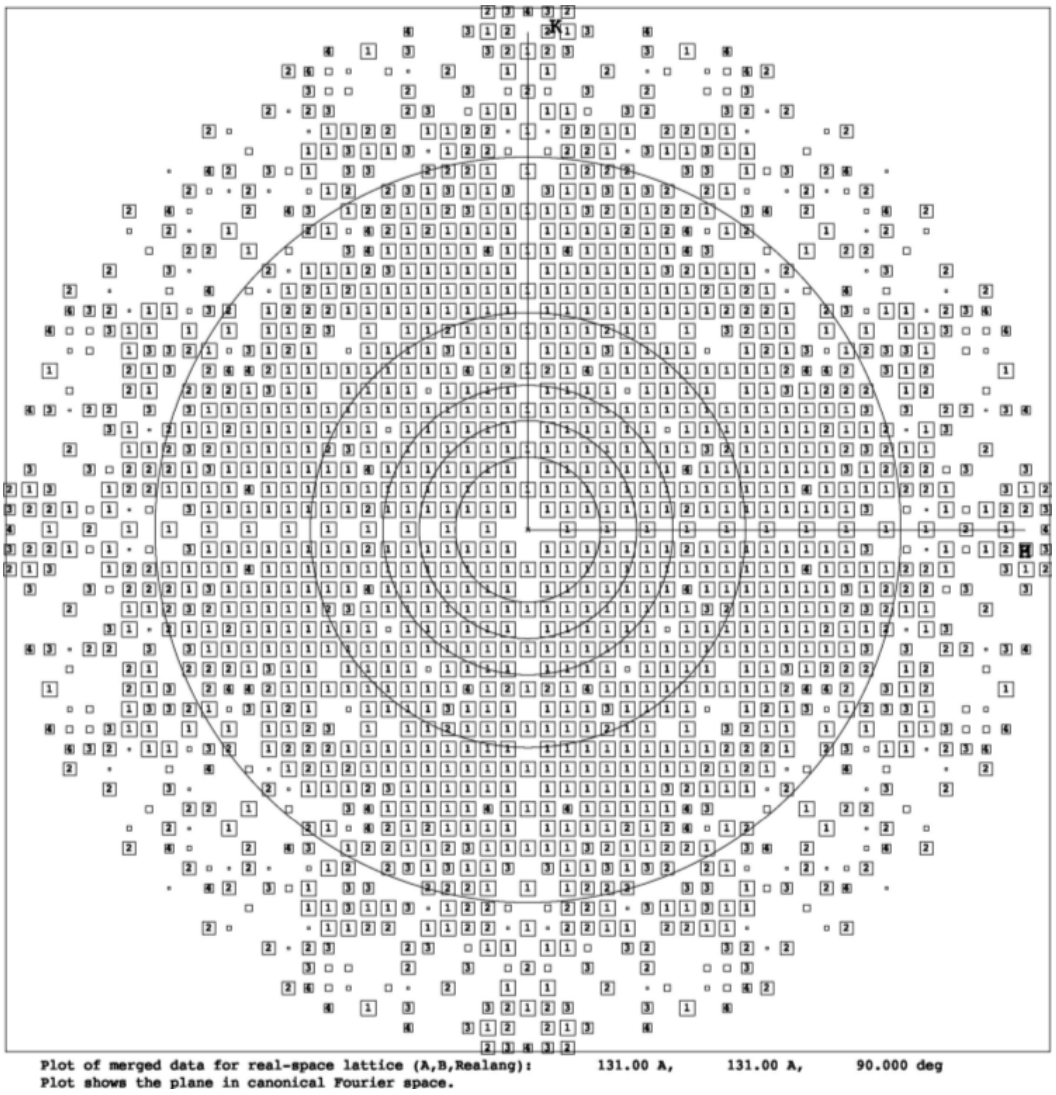

Figure B.3: PLOTRES resolution circle plot of non-tilted merged data. Images were recorded at 0° nominal tilt, automatically processed with 2dx_image (correcting for lattice imperfections and CTF-effects) and merged using 2dx_merge. Resolution circles correspond to 36Å, 24Å, 18Å, 12Å, and 7Å, the border is at 5Å. Symbols indicate figure of merit (FOM) categories of merged reflections, and are defined as 1:FOM > 95%; 2:FOM > 90%; 3:FOM > 85%; 4:FOM > 80%; 5:FOM > 75%; 6:FOM > 70%; 7:FOM > 65%; 8:FOM > 60%; 9:FOM < 60%.

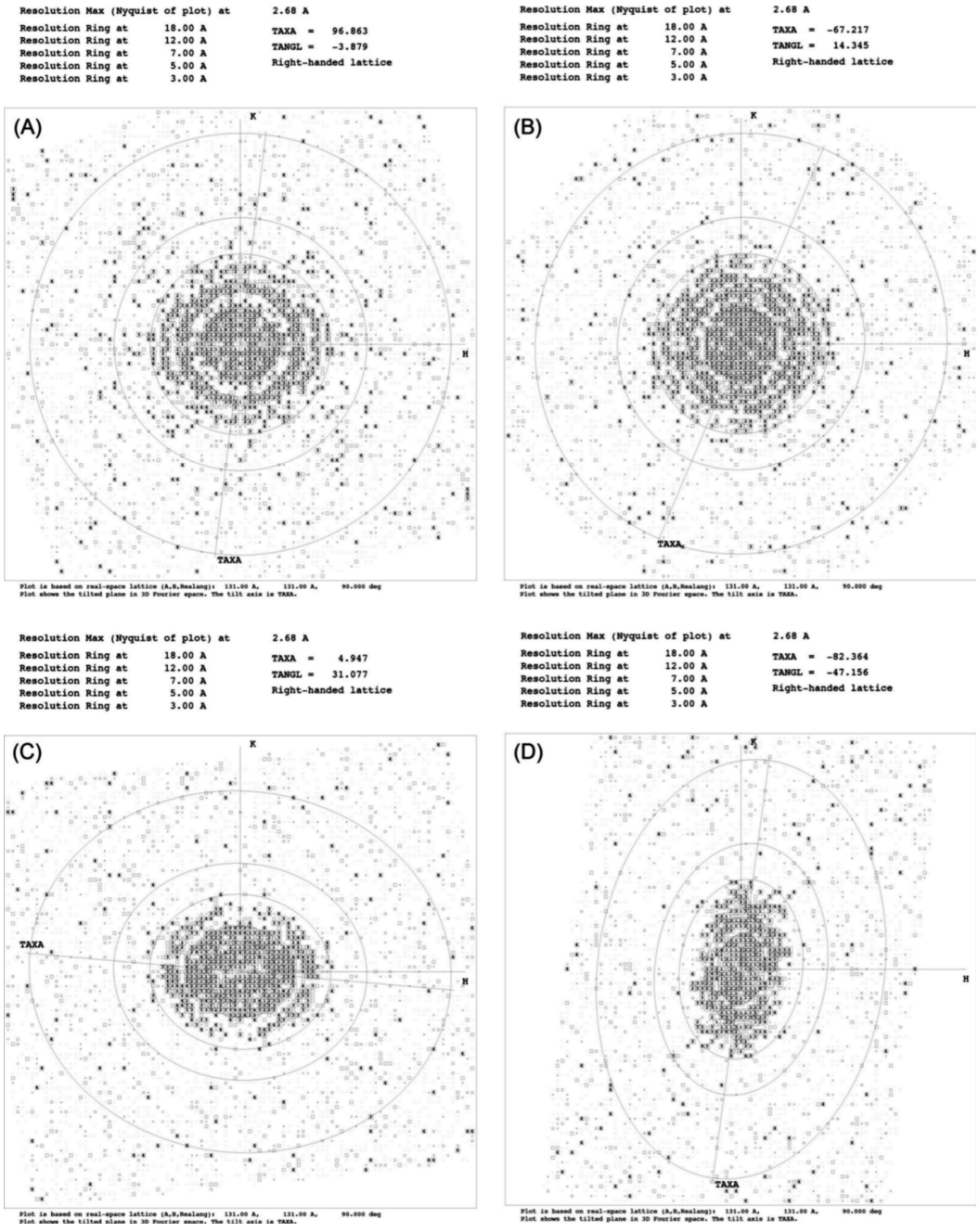

Figure B.4: IQ-statistics gallery up to Nyquist frequency (1/2.68Å) for titled and non-tilted images: (A) non-tilted sample, (B) $\sim 15°$ tilted sample, (C) $\sim 30°$ tilted sample, and (D) $\sim 45°$ tilted sample. Resolution circles correspond to 18Å, 12Å, 7Å, 5Å, and 3Å, the border is at 2.68Å. Plotted are reflections from single images, the symbols indicating the IQ values as defined by Henderson et al. [Henderson *et al.* 1990].

Table B.1: 3D Reconstruction parameters of the final 3D reconstruction consisting of 65 images.

Crystal plane group symmetry:	P_{4212}
Crystal unit cell parameters:	$a = 131.0$Å
	$b = 131.0$Å
	$\gamma = 90.0°$
	$c = 400.0$Å
Number of images:	65
Range of defocus [micrometers]:	0.319 ... 2.217
IQ range used:	1 ... 8
Tilt range used:	0.0° ... 51.9°
In-plane resolution cut-off [Å]:	5
Vertical resolution cut-off [Å]:	7
Number of observed reflections:	70'070
Number of observed unique reflections:	40'543
Number of possible unique reflections in asym. unit	12'134
Number obs. uniq. refl. in asym. unit with FOM>50%	6'228
Completeness to 50° tilt and 5Å (7Å vertical), counting reflections with FOM > 50%:	51.3%
Overall weighted phase residual:	13.6°
Overall weighted R-factor:	26.8%

C Supplementary Material for Movie-mode image procssing for 2D crystals

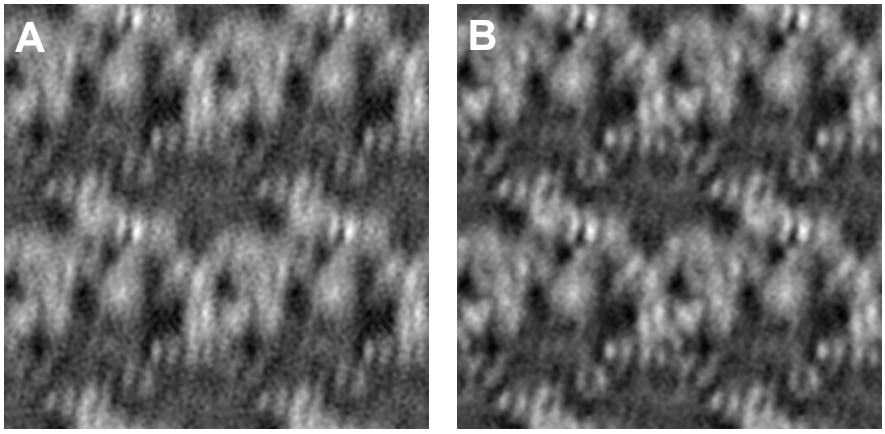

Figure C.1: Movie-mode unbending applied to a 20 degree tilted crystal compared to classical processing. (A) Projection map obtained by classical processing (QVal:, 232.2.2). (B) Projection map produced by movie-mode unbending (QVal,536.1).

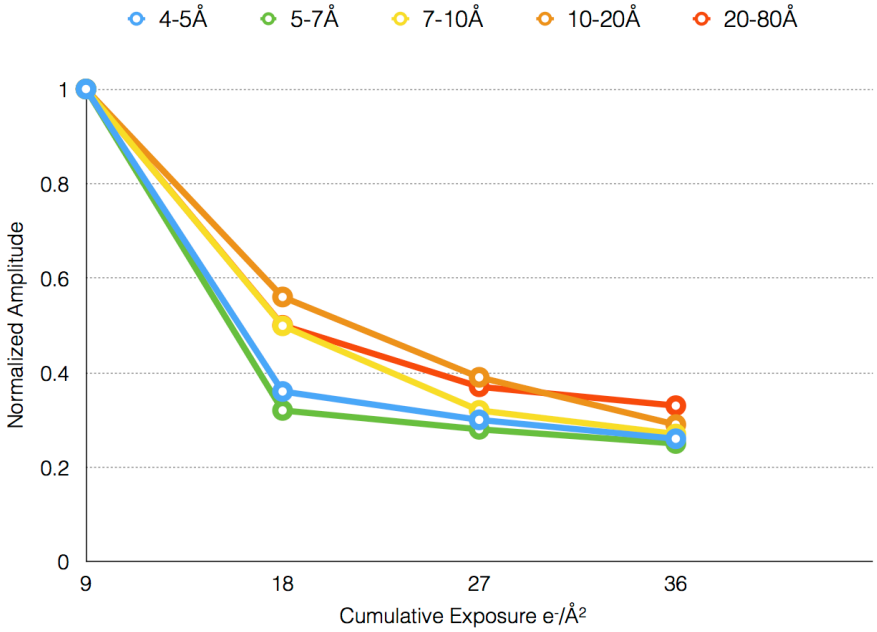

Figure C.2: Dose effects studied as described in [Bammes *et al.* 2010]: The Fourier space was arbitrary divided into 5 "resolution intervals": 4-5Å, 5-7Å, 7-10Å, 10-20Å, and 20-80Å. For each zone the mean of all non-zero normalized Fourier amplitudes of all peaks with IQ<7 was calculated. The normalized amplitude of the reflection with Miller indices (h, k) is defined as $NA_{(h,k),i} = \frac{AMP_{(h,k),i}}{AMP_{(h,k),0}}$, where $i = 0$ is the first image of an exposure series with cumulated electron dose of 9 $e^-/Å^2$, respectively for $i = 1$ the second image with a cumulated dose of 18 $e^-/Å^2$. Diffraction spots of three different non-tilted crystals were analyzed. As we only measured a slightly faster dose-dependent decay of diffraction spots as previously reported by [Baker *et al.* 2010, Bammes *et al.* 2010] the previous results on thin 3D crystals can be used as reasonable approximation. Additionally movie-mode unbending in 2dx provides the functionality to determine the dose-effects for any new project, which allows the future user to fine-tune the damage compensation and to further study the behavior of 2D lipid membrane crystals under the electron beam.

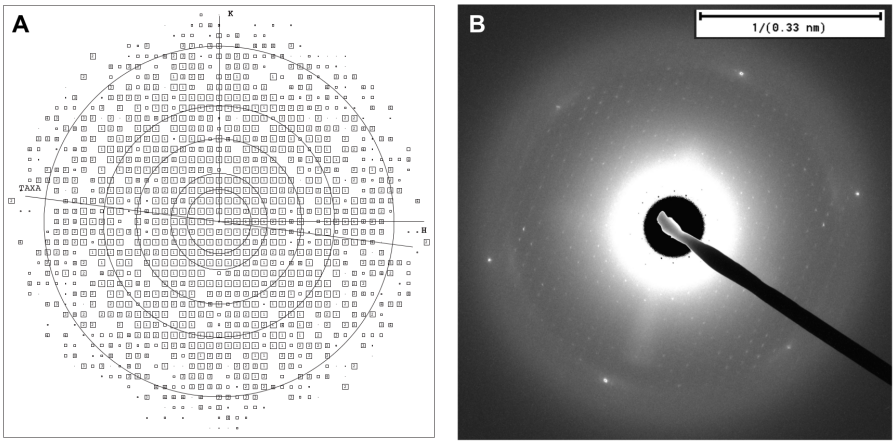

Figure C.3: Comparison of a computed and a recorded diffraction pattern of the same kind of remarkably well ordered 2D crystal. The crystals contain the bacteria outer membrane receptor FhuA with an additional compound bound to it (data unpublished). (A) Canonical IQ plot of a non-tilted movie-mode unbent image recorded on a Titan Krios equipped with a Gatan K2 summit (image conditions described in Section 4.3.1 of the main text), Nyquist frequency of the plot is a 2.68Å, the resolution circles are at 18Å, 12Å, 7Å, 5Å and 3.3Å. (B) Electron diffraction pattern of the same sample recorded CM200 equipped with a TVIPS F416 CMOS camera. Both the computed and the measured diffraction pattern show spots up to the same resolution range. Due to the different detectors used for data acquisition (movie-mode unbending required a DED, which is not suited for electron diffraction exposures) a quantitative comparison is difficult. Nevertheless we believe that DEDs combined with the latest drift-correction approaches are able to record high-quality data that was previously only attainable by electron diffraction.

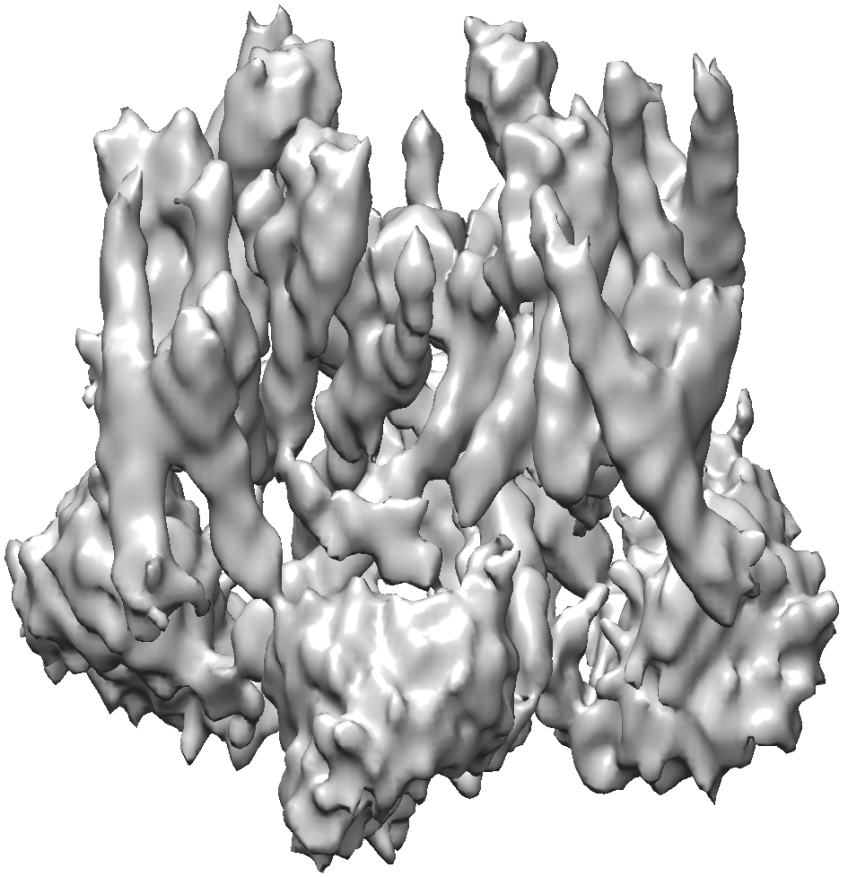

Figure C.4: 3D density map of MloK1 at 4.2/5.5Å resolution (in plane / out of plane) reconstructed from movie-mode unbent unit cells.

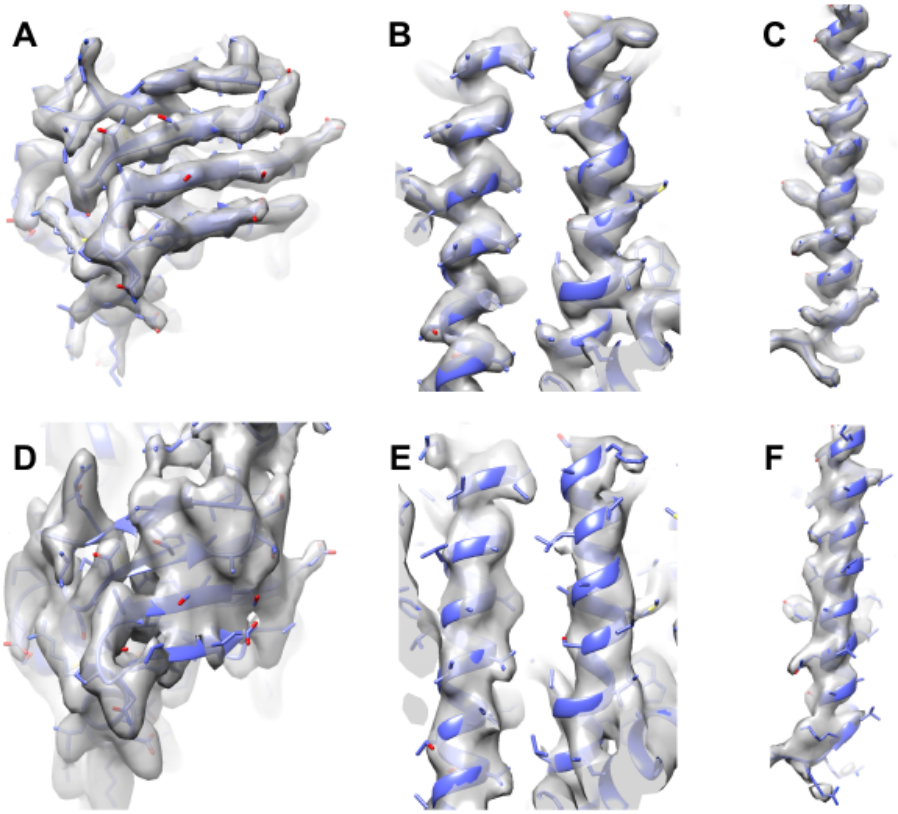

Figure C.5: Impact of the missing cone to different areas of the protein structure. We used e2pdb2mrc.py from EMAN2 [Tang *et al.* 2007] to convert the atomic model (PDB-4CHW) into a 5Å density map. A python script, which was augmented by image processing functions implemented in SPARX [Hohn *et al.* 2007], was written to simulate the effect the missing cone. Simulated density maps not affected by missing cone artifacts are shown in the top row: (A) CNDB of the ion-channel, (B) transmembrane region of the selectivity filter and (C) S2-helix in the voltage-sensing domain. Bottom row (D-F) shows the same regions affected by a missing cone of 40°, i.e. completely sampled Fourier space up to 50° tilt. Negative effects of higher sample tilt on image quality were neglected.

Table C.1: Quality measurements for all shown 3D reconstructions.

	Classical Unbending	**Movie-mode Unbending**	**Movie-mode unbending plus higher resolution refinement**
In-plane resolution cut-off [Å]:	5	5	4.2
Vertical resolution cut-off [Å]:	7	7	5.5
Number obs. uniq. refl. in asym. unit with FOM>1%	10'194	12'196	18'861
counting reflections with FOM > 1%:	78.4%	93.8%	74.7%
Number obs. uniq. refl. in asym. unit with FOM>50%	8'439	9'546	14'250
counting reflections with FOM > 1%:	64.9%	73.4%	56.4%
Overall weighted phase residual:	16.0°	15.9°	16.2°
Overall weighted R-factor:	39.6%	39.5%	36.3%

References

[Abeyrathne *et al.* 2010] Priyanka D Abeyrathne, Mohamed Chami, Radosav S Pantelic, Kenneth N Goldie and Henning Stahlberg. *Chapter One-Preparation of 2D Crystals of Membrane Proteins for High-Resolution Electron Crystallography Data Collection.* Methods in enzymology, vol. 481, pages 25–43, 2010. (Cited on pages 1 and 8.)

[Adams *et al.* 2010] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, L-W Hung, Gary J Kapral, Ralf W Grosse-Kunstleve*et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D: Biological Crystallography, vol. 66, no. 2, pages 213–221, 2010. (Cited on page 88.)

[Adrian *et al.* 1984] Marc Adrian, Jacques Dubochet, Jean Lepault and Alasdair W McDowall. *Cryo-electron microscopy of viruses.* Nature, vol. 308, no. 5954, pages 32–36, 1984. (Cited on page 4.)

[Allegretti *et al.* 2014] Matteo Allegretti, Deryck J Mills, Greg McMullan, Werner Kühlbrandt and Janet Vonck. *Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector.* Elife, vol. 3, 2014. (Cited on pages 60, 67 and 70.)

[Amunts *et al.* 2014] Alexey Amunts, Alan Brown, Xiao-chen Bai, Jose L Llácer, Tanweer Hussain, Paul Emsley, Fei Long, Garib Murshudov, Sjors HW Scheres and V Ramakrishnan. *Structure of the yeast mitochondrial large ribosomal subunit.* Science, vol. 343, no. 6178, pages 1485–1489, 2014. (Cited on pages 15, 16 and 60.)

[Anderson & Olkin 1985] Theodore Wilbur Anderson and Ingram Olkin. *Maximum-likelihood estimation of the parameters of a multivariate normal distribution.* Linear Algebra and its Applications, vol. 70, pages 147–171, 1985. (Cited on page 32.)

[Arheit *et al.* 2013a] Marcel Arheit, Daniel Castaño-Díez, Raphaël Thierry, Priyanka Abeyrathne, Bryant R Gipson and Henning Stahlberg. *Merging of image data in electron crystallography.* In Electron Crystallography of Soluble and Membrane Proteins, pages 195–209. Springer, 2013. (Cited on pages 12, 24 and 56.)

[Arheit *et al.* 2013b] Marcel Arheit, Daniel Castaño-Díez, Raphaël Thierry, Bryant R Gipson, Xiangyan Zeng and Henning Stahlberg. *Automation of*

image processing in electron crystallography. In Electron Crystallography of Soluble and Membrane Proteins, pages 313–330. Springer, 2013. (Cited on page 53.)

[Arheit *et al.* 2013c] Marcel Arheit, Daniel Castaño-Díez, Raphaël Thierry, Bryant R Gipson, Xiangyan Zeng and Henning Stahlberg. *Image processing of 2D Crystal images.* In Electron Crystallography of Soluble and Membrane Proteins, pages 171–194. Springer, 2013. (Cited on pages 12, 22, 24, 49, 53, 62, 63 and 66.)

[Arheit *et al.* 2015] Marcel Arheit, Paulino Paulino, Daniel Castano-Diez, Werner Kühlbrandt and Henning Stahlberg. *Conformational changes in difference maps of Cryo-EM 2D crystals.* Manuscript in preparation, 2015. (Cited on page 83.)

[Bai *et al.* 2013] Xiao-chen Bai, Israel S Fernandez, Greg McMullan and Sjors HW Scheres. *Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles.* Elife, vol. 2, 2013. (Cited on pages 34, 45, 61 and 70.)

[Baker *et al.* 2010] Lindsay A Baker, Eric A Smith, Stephanie A Bueler and John L Rubinstein. *The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals.* Journal of structural biology, vol. 169, no. 3, pages 431–437, 2010. (Cited on pages 62, 65, 67 and 114.)

[Bammes *et al.* 2010] Benjamin E Bammes, Joanita Jakana, Michael F Schmid and Wah Chiu. *Radiation damage effects at four specimen temperatures from 4 to 100K.* Journal of structural biology, vol. 169, no. 3, pages 331–341, 2010. (Cited on pages 67, 114 and 132.)

[Bammes *et al.* 2012] Benjamin E Bammes, Ryan H Rochat, Joanita Jakana, Dong-Hua Chen and Wah Chiu. *Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency.* Journal of structural biology, vol. 177, no. 3, pages 589–601, 2012. (Cited on pages 48 and 60.)

[Basit *et al.* 2012] Hajra Basit, K Shivaji Sharma, Angéline Van der Heyden, Chantal Gondran, Cécile Breyton, Pascal Dumy, Françoise M Winnik and Pierre Labbé. *Amphipol mediated surface immobilization of FhuA: a platform for label-free detection of the bacteriophage protein pb5.* Chemical Communications, vol. 48, no. 48, pages 6037–6039, 2012. (Cited on page 81.)

[Breyton *et al.* 2013] Cécile Breyton, Ali Flayhan, Frank Gabel, Mathilde Lethier, Grégory Durand, Pascale Boulanger, Mohamed Chami and Christine Ebel. *Assessing the conformational changes of pb5, the receptor-binding protein of*

phage T5, upon binding to its Escherichia coli receptor FhuA. Journal of Biological Chemistry, vol. 288, no. 42, pages 30763–30772, 2013. (Cited on page 81.)

[Brilot *et al.* 2012] Axel F Brilot, James Z Chen, Anchi Cheng, Junhua Pan, Stephen C Harrison, Clinton S Potter, Bridget Carragher, Richard Henderson and Nikolaus Grigorieff. *Beam-induced motion of vitrified specimen on holey carbon film*. Journal of structural biology, vol. 177, no. 3, pages 630–637, 2012. (Cited on pages 15, 49 and 61.)

[Brown *et al.* 2015] Alan Brown, Fei Long, Robert A Nicholls, Jaan Toots, Paul Emsley and Garib Murshudov. *Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions*. Acta Crystallographica Section D: Biological Crystallography, vol. 71, no. 1, pages 136–153, 2015. (Cited on page 13.)

[Campbell *et al.* 2012] Melody G Campbell, Anchi Cheng, Axel F Brilot, Arne Moeller, Dmitry Lyumkis, David Veesler, Junhua Pan, Stephen C Harrison, Clinton S Potter, Bridget Carragher*et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy*. Structure, vol. 20, no. 11, pages 1823–1828, 2012. (Cited on pages 15, 34, 45, 49 and 61.)

[Carragher *et al.* 2000] Bridget Carragher, Nick Kisseberth, David Kriegman, Ronald A Milligan, Clinton S Potter, James Pulokas and Amy Reilein. *Leginon: an automated system for acquisition of images from vitreous ice specimens*. Journal of structural biology, vol. 132, no. 1, pages 33–45, 2000. (Cited on pages 34, 86 and 93.)

[Castaño-Díez *et al.* 2008] Daniel Castaño-Díez, Dominik Moser, Andreas Schoenegger, Sabine Pruggnaller and Achilleas S Frangakis. *Performance evaluation of image processing algorithms on the GPU*. Journal of structural biology, vol. 164, no. 1, pages 153–160, 2008. (Cited on page 36.)

[Crowther *et al.* 1996] RA Crowther, R Henderson and JM Smith. *MRC image processing programs*. Journal of structural biology, vol. 116, no. 1, pages 9–16, 1996. (Cited on pages 9, 11, 22 and 26.)

[Dagum & Menon 1998] Leonardo Dagum and Ramesh Menon. *OpenMP: an industry standard API for shared-memory programming*. Computational Science & Engineering, IEEE, vol. 5, no. 1, pages 46–55, 1998. (Cited on page 34.)

[Diaz *et al.* 2010] Ruben Diaz, William Rice, David L Stokes*et al. Chapter Five-Fourier–Bessel Reconstruction of Helical Assemblies*. Methods in enzymology, vol. 482, pages 131–165, 2010. (Cited on page 6.)

[Drenth 2007] Jan Drenth. Principles of protein x-ray crystallography. Springer Science & Business Media, 2007. (Cited on page 3.)

[Dutzler *et al.* 2002] Raimund Dutzler, Ernest B Campbell, Martine Cadene, Brian T Chait and Roderick MacKinnon. *X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity.* Nature, vol. 415, no. 6869, pages 287–294, 2002. (Cited on page 82.)

[Egelman 2007] Edward H Egelman. *The iterative helical real space reconstruction method: surmounting the problems posed by real polymers.* Journal of structural biology, vol. 157, no. 1, pages 83–94, 2007. (Cited on pages 6 and 84.)

[Elmlund & Elmlund 2012] Dominika Elmlund and Hans Elmlund. *SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles.* Journal of structural biology, vol. 180, no. 3, pages 420–427, 2012. (Cited on page 23.)

[Flayhan *et al.* 2012] Ali Flayhan, Frank Wien, Maïté Paternostre, Pascale Boulanger and Cécile Breyton. *New insights into pb5, the receptor binding protein of bacteriophage T5, and its interaction with its< i> Escherichia coli</i> receptor FhuA.* Biochimie, vol. 94, no. 9, pages 1982–1989, 2012. (Cited on page 81.)

[Förster & Hegerl 2006] Friedrich Förster and Reiner Hegerl. *Structure determination in situ by averaging of tomograms.* Methods in cell biology, vol. 79, pages 741–767, 2006. (Cited on page 6.)

[Frank 1975] Joachim Frank. *Averaging of low exposure electron micrographs of non-periodic objects.* Ultramicroscopy, vol. 1, no. 2, pages 159–162, 1975. (Cited on pages 5, 23, 29 and 30.)

[Fromm *et al.* 2014] Simon A Fromm, Tanmay AM Bharat, Arjen J Jakobi, Wim JH Hagen and Carsten Sachse. *Seeing tobacco mosaic virus through direct electron detectors.* Journal of structural biology, 2014. (Cited on page 15.)

[Gipson *et al.* 2007a] Bryant Gipson, Xiangyan Zeng and Henning Stahlberg. *2dx_merge: Data management and merging for 2D crystal images.* Journal of structural biology, vol. 160, no. 3, pages 375–384, 2007. (Cited on page 54.)

[Gipson *et al.* 2007b] Bryant Gipson, Xiangyan Zeng, Zi Yan Zhang and Henning Stahlberg. *2dx - User-friendly image processing for 2D crystals.* Journal of structural biology, vol. 157, no. 1, pages 64–72, 2007. (Cited on pages 9, 23, 26, 54, 62 and 68.)

[Gipson *et al.* 2011] Bryant R Gipson, Daniel J Masiel, Nigel D Browning, John

Spence, Kaoru Mitsuoka and Henning Stahlberg. *Automatic recovery of missing amplitudes and phases in tilt-limited electron crystallography of two-dimensional crystals.* Physical Review E, vol. 84, no. 1, page 011916, 2011. (Cited on pages 13, 76, 81 and 93.)

[Glaeser *et al.* 2011] RM Glaeser, G McMullan, AR Faruqi and R Henderson. *Images of paraffin monolayer crystals with perfect contrast: minimization of beam-induced specimen motion.* Ultramicroscopy, vol. 111, no. 2, pages 90–100, 2011. (Cited on pages 56, 60 and 61.)

[Gonen *et al.* 2005] Tamir Gonen, Yifan Cheng, Piotr Sliz, Yoko Hiroaki, Yoshinori Fujiyoshi, Stephen C Harrison and Thomas Walz. *Lipid–protein interactions in double-layered two-dimensional AQP0 crystals.* Nature, vol. 438, no. 7068, pages 633–638, 2005. (Cited on page 7.)

[Grigorieff 2007] Nikolaus Grigorieff. *FREALIGN: high-resolution refinement of single particle structures.* Journal of structural biology, vol. 157, no. 1, pages 117–125, 2007. (Cited on pages 23 and 91.)

[Henderson *et al.* 1986] Richard Henderson, JM Baldwin, KH Downing, J Lepault and Fritz Zemlin. *Structure of purple membrane from Halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution.* Ultramicroscopy, vol. 19, no. 2, pages 147–178, 1986. (Cited on pages 12 and 58.)

[Henderson *et al.* 1990] R Henderson, J M_ Baldwin, TA Ceska, F Zemlin, E a Beckmann and KH Downing. *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy.* Journal of molecular biology, vol. 213, no. 4, pages 899–929, 1990. (Cited on page 110.)

[Henderson 1977] Richard Henderson. *The purple membrane from Halobacterium halobium.* Annual review of biophysics and bioengineering, vol. 6, no. 1, pages 87–109, 1977. (Cited on pages 7 and 9.)

[Heymann 2001] J Bernard Heymann. *Bsoft: image and molecular processing in electron microscopy.* Journal of structural biology, vol. 133, no. 2, pages 156–169, 2001. (Cited on page 23.)

[Hohn *et al.* 2007] Michael Hohn, Grant Tang, Grant Goodyear, PR Baldwin, Zhong Huang, Pawel A Penczek, Chao Yang, Robert M Glaeser, Paul D Adams and Steven J Ludtke. *SPARX, a new environment for Cryo-EM image processing.* Journal of structural biology, vol. 157, no. 1, pages 47–55, 2007. (Cited on pages 23 and 117.)

[Jap *et al.* 1992] BK Jap, M Zulauf, T Scheybani, A Hefti, W Baumeister, U Aebi

and A Engel. *2D crystallization: from art to science.* Ultramicroscopy, vol. 46, no. 1, pages 45–84, 1992. (Cited on page 8.)

[Jost *et al.* 2003] Gabriele Jost, Haoqiang Jin, Dieter an Mey and Ferhat F Hatay. *Comparing the OpenMP, MPI, and hybrid programming paradigms on an SMP cluster.* In Proceedings of EWOMP, volume 3, page 2003, 2003. (Cited on page 34.)

[Kebbel *et al.* 2013] Fabian Kebbel, Mareike Kurz, Marcel Arheit, Markus G Grütter and Henning Stahlberg. *Structure and substrate-induced conformational changes of the secondary citrate/sodium symporter CitS revealed by electron crystallography.* Structure, vol. 21, no. 7, pages 1243–1250, 2013. (Cited on page 81.)

[Kowal *et al.* 2014] Julia Kowal, Mohamed Chami, Paul Baumgartner, Marcel Arheit, Po-Lin Chiu, Martina Rangl, Simon Scheuring, Gunnar F Schröder, Crina M Nimigean and Henning Stahlberg. *Ligand-induced structural changes in the cyclic nucleotide-modulated potassium channel MloK1.* Nature communications, vol. 5, 2014. (Cited on pages 37, 54, 56, 70, 73, 81, 82 and 92.)

[Kube *et al.* 2014] Sebastian Kube, Nicole Kapitein, Tomasz Zimniak, Franz Herzog, Axel Mogk and Petra Wendler. *Structure of the VipA/B type VI secretion complex suggests a contraction-state-specific recycling mechanism.* Cell reports, vol. 8, no. 1, pages 20–30, 2014. (Cited on page 84.)

[Kudryashev *et al.* 2015] Mikhail Kudryashev, Ray Wang, Maximilian Brackmann, Sebastian Scherer, Timm Maier, David Baker, Frank DiMaio, Henning Stahlberg, Edward Egelman and Marek Basler. *The structure of the type six secretion system contractile sheath solved by cryo-EM.* Cell, vol. 160, no. 6, pages 952–962, 2015. (Cited on pages 16, 84 and 85.)

[Kühlbrandt 2014] Werner Kühlbrandt. *The Resolution Revolution.* Science, vol. 343, no. 6178, pages 1443–1444, 2014. (Cited on page 13.)

[Kunji *et al.* 2000] Edmund RS Kunji, Susanne von Gronau, Dieter Oesterhelt and Richard Henderson. *The three-dimensional structure of halorhodopsin to 5 Å by electron crystallography: A new unbending procedure for two-dimensional crystals by using a global reference structure.* Proceedings of the National Academy of Sciences, vol. 97, no. 9, pages 4637–4642, 2000. (Cited on page 30.)

[Lander *et al.* 2009] Gabriel C Lander, Scott M Stagg, Neil R Voss, Anchi Cheng, Denis Fellmann, James Pulokas, Craig Yoshioka, Christopher Irving, Anke

Mulder, Pick-Wei Lau*et al.* *Appion: an integrated, database-driven pipeline to facilitate EM image processing.* Journal of structural biology, vol. 166, no. 1, pages 95–102, 2009. (Cited on page 86.)

[Lepault 1985] J Lepault. *Cryo-electron microscopy of helical particles TMV and T4 polyheads.* Journal of microscopy, vol. 140, no. 1, pages 73–80, 1985. (Cited on page 6.)

[Li *et al.* 2013a] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard and Yifan Cheng. *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* Nature methods, vol. 10, no. 6, pages 584–590, 2013. (Cited on pages 15, 34, 50, 51, 60, 87, 91 and 92.)

[Li *et al.* 2013b] Xueming Li, Shawn Q Zheng, Kiyoshi Egami, David A Agard and Yifan Cheng. *Influence of electron dose rate on electron counting images recorded with the K2 camera.* Journal of Structural Biology, vol. 184, no. 2, pages 251–260, 2013. (Cited on pages 48, 50 and 72.)

[Liao *et al.* 2013] Maofu Liao, Erhu Cao, David Julius and Yifan Cheng. *Structure of the TRPV1 ion channel determined by electron cryo-microscopy.* Nature, vol. 504, no. 7478, pages 107–112, 2013. (Cited on pages 15, 16, 49, 51 and 60.)

[López-Blanco & Chacón 2015] José Ramón López-Blanco and Pablo Chacón. *Structural modeling from electron microscopy data.* Wiley Interdisciplinary Reviews: Computational Molecular Science, vol. 5, no. 1, pages 62–81, 2015. (Cited on page 13.)

[Lu *et al.* 2014a] Alvin Lu, Venkat Giri Magupalli, Jianbin Ruan, Qian Yin, Maninjay K Atianand, Matthijn R Vos, Gunnar F Schröder, Katherine A Fitzgerald, Hao Wu and Edward H Egelman. *Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes.* Cell, vol. 156, no. 6, pages 1193–1206, 2014. (Cited on page 85.)

[Lu *et al.* 2014b] Peilong Lu, Xiao-chen Bai, Dan Ma, Tian Xie, Chuangye Yan, Linfeng Sun, Guanghui Yang, Yanyu Zhao, Rui Zhou, Sjors HW Scheres*et al.* *Three-dimensional structure of human [ggr]-secretase.* Nature, vol. 512, no. 7513, pages 166–170, 2014. (Cited on pages 5 and 16.)

[Ludtke *et al.* 1999] Steven J Ludtke, Philip R Baldwin and Wah Chiu. *EMAN: semiautomated software for high-resolution single-particle reconstructions.* Journal of structural biology, vol. 128, no. 1, pages 82–97, 1999. (Cited on page 23.)

[Lyumkis *et al.* 2013] Dmitry Lyumkis, Axel F Brilot, Douglas L Theobald and Nikolaus Grigorieff. *Likelihood-based classification of cryo-EM images using FREALIGN.* Journal of structural biology, vol. 183, no. 3, pages 377–388, 2013. (Cited on page 45.)

[Mastronarde 2003] David Mastronarde. *SerialEM a program for automated tilt series acquisition on Tecnai microcopes using prediction of specimen position.* Microscopy and Microanalysis, vol. 9, no. S02, pages 1182–1183, 2003. (Cited on page 86.)

[McMullan *et al.* 2009] G McMullan, S Chen, R Henderson and AR Faruqi. *Detective quantum efficiency of electron area detectors in electron microscopy.* Ultramicroscopy, vol. 109, no. 9, pages 1126–1143, 2009. (Cited on pages 5, 15 and 48.)

[Mersereau & Oppenheim 1974] Russell M Mersereau and Alan V Oppenheim. *Digital reconstruction of multidimensional signals from their projections.* Proceedings of the IEEE, vol. 62, no. 10, pages 1319–1338, 1974. (Cited on page 4.)

[Milazzo *et al.* 2011] Anna-Clare Milazzo, Anchi Cheng, Arne Moeller, Dmitry Lyumkis, Erica Jacovetty, James Polukas, Mark H Ellisman, Nguyen-Huu Xuong, Bridget Carragher and Clinton S Potter. *Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy.* Journal of structural biology, vol. 176, no. 3, pages 404–408, 2011. (Cited on pages 48 and 60.)

[Mindell & Grigorieff 2003] Joseph A Mindell and Nikolaus Grigorieff. *Accurate determination of local defocus and specimen tilt in electron microscopy.* Journal of structural biology, vol. 142, no. 3, pages 334–347, 2003. (Cited on pages 10, 24, 28, 53, 88 and 100.)

[Penczek *et al.* 2004] Pawel A Penczek, Robert Renka and Hermann Schomberg. *Gridding-based direct Fourier inversion of the three-dimensional ray transform.* JOSA A, vol. 21, no. 4, pages 499–509, 2004. (Cited on page 91.)

[Penczek *et al.* 2014] Pawel A Penczek, Jia Fang, Xueming Li, Yifan Cheng, Justus Loerke and Christian MT Spahn. *CTER—Rapid estimation of CTF parameters with error assessment.* Ultramicroscopy, vol. 140, pages 9–19, 2014. (Cited on page 51.)

[Penczek 2010a] Pawel A Penczek. *Chapter three-resolution measures in molecular electron microscopy.* Methods in enzymology, vol. 482, pages 73–100, 2010. (Cited on page 40.)

[Penczek 2010b] Pawel A Penczek. *Chapter Two-Image Restoration in Cryo-Electron Microscopy.* Methods in enzymology, vol. 482, pages 35–72, 2010. (Cited on page 10.)

[Pettersen *et al.* 2004] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng and Thomas E Ferrin. *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, vol. 25, no. 13, pages 1605–1612, 2004. (Cited on pages 29 and 80.)

[Radermacher 1988] Michael Radermacher. *Three-Dimensional reconstruction of single particles from random and nonrandom tilt series.* Journal of electron microscopy technique, vol. 9, no. 4, pages 359–394, 1988. (Cited on page 23.)

[Rémigy *et al.* 2003] H-W Rémigy, D Caujolle-Bert, K Suda, A Schenk, M Chami and A Engel. *Membrane protein reconstitution and crystallization by controlled dilution.* FEBS letters, vol. 555, no. 1, pages 160–169, 2003. (Cited on page 8.)

[Rigaud *et al.* 1997] Jean-Louis Rigaud, Gervaise Mosser, Jean-Jacques Lacapere, Anders Olofsson, Daniel Levy and Jean-Luc Ranck. *Bio-Beads: an efficient strategy for two-dimensional crystallization of membrane proteins.* Journal of structural biology, vol. 118, no. 3, pages 226–235, 1997. (Cited on page 8.)

[Ruskin *et al.* 2013] Rachel S Ruskin, Zhiheng Yu and Nikolaus Grigorieff. *Quantitative characterization of electron detectors for transmission electron microscopy.* Journal of structural biology, vol. 184, no. 3, pages 385–393, 2013. (Cited on pages 15, 48, 51 and 60.)

[Sanders & Myers 2004] Charles R Sanders and Jeffrey K Myers. *Disease-related misassembly of membrane proteins.* Annu. Rev. Biophys. Biomol. Struct., vol. 33, pages 25–51, 2004. (Cited on page 1.)

[Sborgi *et al.* 2015] Lorenzo Sborgi, Francesco Ravotti, Venkata Dandey, Mathias Dick, Adam Mazur, Sina Reckel, Mohamed Chami, Sebastian Scherer, Anja Böckmann, Edward Egelman, Henning Stahlberg, Petr Broz, Beat Meier and Sebastian Hiller. *Structure and assembly of the mouse ASC filament by combined NMR spectroscopy and cryo-electron microscopy.* (manuscript in prepartation), 2015. (Cited on page 86.)

[Schenk *et al.* 2010] Andreas D Schenk, Daniel Castaño-Díez, Bryant Gipson, Marcel Arheit, Xiangyan Zeng and Henning Stahlberg. *Chapter Four-3D Reconstruction from 2D Crystal Image and Diffraction Data.* Methods in enzymology, vol. 482, pages 101–129, 2010. (Cited on pages 12 and 14.)

[Scherer *et al.* 2013] Sebastian Scherer, Marcel Arheit, Julia Kowal, Xiangyan Zeng and Henning Stahlberg. *Single particle 3D reconstruction for 2D crystal images of membrane proteins.* Journal of structural biology, 2013. (Cited on pages 51, 58 and 76.)

[Scherer *et al.* 2014] Sebastian Scherer, Julia Kowal, Mohamed Chami, Venkata Dandey, Marcel Arheit, Philippe Ringler and Henning Stahlberg. *2dx_automator: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline.* Journal of structural biology, vol. 186, no. 2, pages 302–307, 2014. (Cited on pages 61, 72 and 82.)

[Scheres & Chen 2012] Sjors HW Scheres and Shaoxia Chen. *Prevention of overfitting in cryo-EM structure determination.* Nature methods, vol. 9, no. 9, pages 853–854, 2012. (Cited on page 41.)

[Scheres 2012a] Sjors HW Scheres. *A Bayesian view on cryo-EM structure determination.* Journal of molecular biology, vol. 415, no. 2, pages 406–418, 2012. (Cited on page 87.)

[Scheres 2012b] Sjors HW Scheres. *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* Journal of structural biology, vol. 180, no. 3, pages 519–530, 2012. (Cited on pages 5, 23, 45, 87, 88 and 91.)

[Scheres 2014a] Sjors HW Scheres. *Beam-induced motion correction for sub-megadalton cryo-EM particles.* Elife, vol. 3, page e03665, 2014. (Cited on pages 15, 87 and 88.)

[Scheres 2014b] Sjors HW Scheres. *Semi-automated selection of cryo-EM particles in RELION-1.3.* Journal of structural biology, 2014. (Cited on page 5.)

[Scherzer 1949] O Scherzer. *The theoretical resolution limit of the electron microscope.* Journal of Applied Physics, vol. 20, no. 1, pages 20–29, 1949. (Cited on pages 9 and 99.)

[Schröder *et al.* 2007] Gunnar F Schröder, Axel T Brunger and Michael Levitt. *Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution.* Structure, vol. 15, no. 12, pages 1630–1641, 2007. (Cited on pages 13, 41 and 73.)

[Schur *et al.* 2013] Florian KM Schur, Wim JH Hagen, Alex de Marco and John AG Briggs. *Determination of protein structure at 8.5 Å resolution using cryo-electron tomography and sub-tomogram averaging.* Journal of structural biology, vol. 184, no. 3, pages 394–400, 2013. (Cited on page 6.)

[Shaw 1984] PJ Shaw. *Constrained least-squares fitting of the lattice lines in three-dimensional reconstruction of monolayer crystals.* Ultramicroscopy, vol. 14, no. 4, pages 363–365, 1984. (Cited on page 12.)

[Signorell *et al.* 2007] Gian A Signorell, Thomas C Kaufmann, Wanda Kukulski, Andreas Engel and Hervé-W Rémigy. *Controlled 2D crystallization of membrane proteins using methyl-β-cyclodextrin.* Journal of structural biology, vol. 157, no. 2, pages 321–328, 2007. (Cited on page 8.)

[Sorzano *et al.* 2004] COS Sorzano, Roberto Marabini, Javier Velázquez-Muriel, José Román Bilbao-Castro, Sjors HW Scheres, José M Carazo and Alberto Pascual-Montano. *XMIPP: a new generation of an open-source image processing package for electron microscopy.* Journal of structural biology, vol. 148, no. 2, pages 194–204, 2004. (Cited on page 23.)

[Stewart & Grigorieff 2004] Alex Stewart and Nikolaus Grigorieff. *Noise bias in the refinement of structures derived from single particles.* Ultramicroscopy, vol. 102, no. 1, pages 67–84, 2004. (Cited on page 41.)

[Tang *et al.* 2007] Guang Tang, Liwei Peng, Philip R Baldwin, Deepinder S Mann, Wen Jiang, Ian Rees and Steven J Ludtke. *EMAN2: an extensible image processing suite for electron microscopy.* Journal of structural biology, vol. 157, no. 1, pages 38–46, 2007. (Cited on pages 33, 88 and 117.)

[Toyoshima & Unwin 1988] Chikashi Toyoshima and Nigel Unwin. *Contrast transfer for frozen-hydrated specimens: determination from pairs of defocused images.* Ultramicroscopy, vol. 25, no. 4, pages 279–291, 1988. (Cited on page 99.)

[Trabuco *et al.* 2008] Leonardo G Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank and Klaus Schulten. *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.* Structure, vol. 16, no. 5, pages 673–683, 2008. (Cited on page 13.)

[Unger 2013] Vinzenz M Unger. *Evaluation of Electron Crystallographic Data from Images of Two-Dimensional Crystals.* In Electron Crystallography of Soluble and Membrane Proteins, pages 211–227. Springer, 2013. (Cited on page 40.)

[Veesler *et al.* 2013] David Veesler, Melody G Campbell, Anchi Cheng, Chi-yu Fu, Zachary Murez, John E Johnson, Clinton S Potter and Bridget Carragher. *Maximizing the potential of electron cryomicroscopy data collected using direct detectors.* Journal of structural biology, vol. 184, no. 2, pages 193–202, 2013. (Cited on pages 48 and 61.)

[Wang *et al.* 2015] Ray Yu-Ruei Wang, Mikhail Kudryashev, Xueming Li, Edward H Egelman, Marek Basler, Yifan Cheng, David Baker and Frank Di-

Maio. *De novo protein structure determination from near-atomic-resolution cryo-EM maps.* Nature Methods, 2015. (Cited on page 13.)

[Wilson 1970] Arthur James Cochran Wilson. Elements of x-ray crystallography. Addison-Wesley Reading, Massachusetts, 1970. (Cited on page 3.)

[Winkler 2007] Hanspeter Winkler. *3D reconstruction and processing of volumetric data in cryo-electron tomography.* Journal of structural biology, vol. 157, no. 1, pages 126–137, 2007. (Cited on page 6.)

[Zeng *et al.* 2007a] Xiangyan Zeng, Bryant Gipson, Zi Yan Zheng, Ludovic Renault and Henning Stahlberg. *Automatic lattice determination for two-dimensional crystal images.* Journal of structural biology, vol. 160, no. 3, pages 353–361, 2007. (Cited on pages 11, 53 and 54.)

[Zeng *et al.* 2007b] Xiangyan Zeng, Henning Stahlberg and Nikolaus Grigorieff. *A maximum likelihood approach to two-dimensional crystals.* Journal of structural biology, vol. 160, no. 3, pages 362–374, 2007. (Cited on page 27.)

[Zhang & Hong Zhou 2011] Xing Zhang and Z Hong Zhou. *Limiting factors in atomic resolution cryo electron microscopy: no simple tricks.* Journal of structural biology, vol. 175, no. 3, pages 253–263, 2011. (Cited on pages 28 and 93.)

[Zhang *et al.* 2010] Xing Zhang, Lei Jin, Qin Fang, Wong H Hui and Z Hong Zhou. *3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry.* Cell, vol. 141, no. 3, pages 472–482, 2010. (Cited on pages 5 and 15.)

List of Figures

List of Tables