Workflow Automation for Image Analysis of 2D Crystals of Membrane Proteins

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Marcel Andreas Arheit
aus Basel BS

Basel 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Henning Stahlberg
Biozentrum, Universität Basel

Prof. Dr. Volker Roth
Departement Mathematik und Informatik, Universität Basel

Basel, den 10.12.2013

Prof. Dr. Jörg Schibler
Dekan

Abstract

Membrane proteins carry out various functions essential to the survival of organisms. They transfer signals between the cell's internal and external environments, move molecules and ions across the membrane, act as enzymes, and allow cell adhesion. This is why membrane proteins represent more than half of all drug targets. A deeper insight into the functional mechanisms of a protein can be gained from structural information. And so far only a fraction of membrane protein structures has been determined.

The topic of this thesis is structure determination of membrane proteins through electron crystallography focusing on the image processing of 2D crystals. The thesis combines both method development and structure studies. In the Methods part, state of the art processing of 2D crystal images is presented. The workflow embedding all the processing steps from the initial micrographs of 2D crystals to the resulting 3D electron density map of the reconstituted membrane protein is described. The possibility of autonomous high-throughput processing is discussed as the ultimate goal of automation of this workflow. An additional processing step of the workflow that captures the variation of tilt geometry in the 2D crystal is introduced. This is implemented as an iterative refinement of the local tilt geometry using a Single Particle processing approach.

A great benefit of electron crystallography is the fact that through reconstitution the purified protein is embedded in a natural environment, a membrane. Biochemical manipulations of this environment can lead to structural changes, which yields insight into the functional states of the protein. An new method of analyzing these structural changes in 2D projection maps is presented here. The method identifies significant changes in the protein by distinguishing them from noise derived artifacts.

The second part of this thesis covers applications of these methods in structural studies of unknown membrane proteins. In the study of the Secondary Citrate/-Sodium Symporter CitS, the substrate binding domain was identified with help of the significant difference map method. The improvements of the image processing routines were directly applied in the analysis of the 2D crystals.

The structural studies of nucleotide-modulated potassium channel MloK1 also benefited from the automated image processing workflow and the significant difference map, while identifying structural changes through ligand binding. To gain a more detailed electron density map of MloK1, the local tilt geometry of the crystals were refined with the single particle 3D reconstruction for 2D crystal images method.

Acknowledgments

First of all I want to thank my thesis supervisor Henning Stahlberg for giving me the opportunity for this PhD program. Thank you for being such an motivational and optimistic person, it has helped me a lot during my PhD. I need to thank you for lending me all your knowledge in structural biology and especially electron crystallography.

I would also like to thank my co-supervisors Volker Roth and Torsten Schwede. During committee meetings both gave me not just scientific feedback but also shared different perspectives with me.

A big thank you goes to Daniel Castano, wihtout whom this thesis probably would not exist. He has been a great mentor and helped me throughout my PhD with countless discussions. Thank you for taking your time.

Misha Kudryashev made we want to be a better scientist. But first of all I have to thank him for being a friend.

Not just me, put the 2dx community should thank Cristina Paulino for exploring every detail of the software. Thank you for being a perfectionist, although it has cost me a lot of nerves.

Sebastian Scherer you make me want to be a better programmer. Thank you for your momentum.

Without Bryant Gipson 2dx would not exist. Thank you Bryant. Too bad you left Basel.

Fabian Kebbel, you made biology so easy. Thank you for being a great colleague.

Thanks Julia Kowal for being such an honest person. I always knew, when something wasn't to your liking in 2dx.

Priyanka Abeyrathne you are a lovely person and a never sleeping scientist. Thank you for that.

Unfortunately I cannot mention all of you, but thank you all my colleagues at C-CINA. I can honestly say, I liked working with all of you.

Finally I would like to thank my family for supporting me in whatever decision I made and giving me the courage to pursue my dreams. My biggest thanks goes to my love Johanna Hänggi. You are the warmest, most understanding person I know. Thank you for your moral support.

Contents

1. Introduction

Membrane proteins are of great interest because of their key role in controlling processes of life. They perform a variety of functions vital to the survival of organisms such as energy conversion, nutrient uptake, drug efflux, protein secretion, and signal transduction [Abeyrathne *et al.* 2012]. Not surprisingly membrane proteins comprise around 30 % of proteins encoded by the eukaryotic genome [Engel & Gaub 2008]. Membrane proteins can be associated (integral) or attached (peripheral) to membranes of cells or organelles. Due to their vital functions in the cell, a dysfunction of membrane proteins can cause severe diseases. Therefore membrane proteins represent the largest class of drug targets, composing the majority of pharmaceuticals on the market [Arinaminpathy *et al.* 2009].

1.1. Structural Analysis of Membrane Proteins

A detailed understanding of the mechanisms and regulation of membrane proteins at the molecular level requires structural insights. However, due to their amphipatic nature, this class of proteins is very challenging to express and handle. As a consequence, the number of determined structures of membrane proteins make only less than 0.1% of all known protein structures (`http://blanco.biomol.uci.edu/mpstruc`). Nevertheless with the right expression system, a sufficient amount of protein can be produced for structural studies. There are different methods for structural analysis of membrane proteins: X-Ray crystallography, nuclear magnetic resonance, and electron microscopy. Each method has its strength and benefits, but also has its specific requirements when it comes to sample preparation, data collection and structural analysis. Therefore, not every method is suited for membrane proteins.

X-Ray crystallography is the most advanced method to solve structures of membrane proteins and has also provided most of the 3D structures of membrane proteins known today. As the name states, the method uses X-rays to acquire structural information from a 3D crystal. The main challenge is to get a regular repeating structure of the protein, the so called 3D crystal. The general principle of protein crystallization involves the slow increase of the protein concentration in a droplet. Unfortunately there are numerous factors (precipitant, pH, additives etc.) that have to be varied in order to find the right conditions to induce a 3D crystal. Accordingly there is no general rule that applies for every protein. Once a crystal has been gained and it contains enough repeats, irradiation will produce a diffraction pattern. The resulting pattern echoes the repeating structures of the protein through a lattice of so called reflections. Mathematically speaking, the diffraction pattern reflects amplitudes of the Fourier transform of the crystal. To obtain an interpretable electron density map of the protein the phases of the Fourier transform are also needed. They

cannot be recorded directly through diffraction (phase problem), however there are different methods to obtain phase information e.g. molecular replacement, anomalous X-ray scattering or heavy atom methods.

In contrast to X-Ray crystallography, nuclear magnetic resonance (NMR) spectroscopy can determine the structure of membrane proteins in aqueous solution. The structure determination in NMR spectroscopy utilizes the fact that certain atomic nuclei e.g. ^{1}H, ^{13}C, ^{15}N, and ^{31}P, have a magnetic moment. Therefore the nuclei can be aligned by placing the protein in a strong magnetic field. This alignment equilibrium is then perturbed by an electromagnetic radio frequency (RF) pulse, which leads to oscillation of the atomic nuclei at their resonance frequency. The interruption of the external irradiation returns the nuclei to thermal equilibrium by emitting RF radiation. The emitted RF waves can be detected and converted into frequency spectrum from which the individual nuclear resonance frequencies can be detected. These magnetic forces describe the distances between pairs of atomic nuclei, which finally leads to the structure of the protein.

Electron microscopy (EM) also allows structural analysis of macromolecules. This method utilizes electrons to illuminate the sample, which in contrast to the visible photons in light microscopy ($\lambda = 3800\text{Å} - 7600\text{Å}$) can have a much shorter wavelength ($\lambda = 0.4\text{Å} - 0.009\text{Å}$ depending on their kinetic energy). But there are several limiting factors e.g. beam tilt, focus gradients, and non-elastic scattering electrons, contrast transfer function [Zhang & Zhou 2011]. The specimen can be either prepared by negative staining or by freezing it to a vitrified state. This is called cryo electron microscopy (cryo-EM). In cryo-EM there are different techniques, depending on the size of the specimen and the biological question one wants to answer: 1) *Electron tomography* images the object of interest under several angles and thereby gains its 3D structure. 2) In *Single Particle EM*, samples with multiple copies of the same particle are imaged, but in different orientations, hence different projections of the same object are collected. 3) *Electron Crystallography* utilizes crystals of proteins as a sample similar to X-ray crystallography to gain structural information. In contrast to X-ray crystallography the sample consists of a two dimensional crystal. Electron crystallography is the main technique used for structural analysis of membrane proteins in this thesis. Therefore the next section will go into more detail about this technique.

1.2. Electron Crystallography

Electron crystallography yields structures of membrane proteins with the highest resolution within EM (1.9Å Aquaporin; AQP0 [Gonen *et al.* 2004]). It also has the benefit that trough the crystallization the protein is embedded in lipid bilayer and can be examined in a near-native environment, which preserves its functional form. Electron crystallography became the first technique to solve the 3D structure of an

integral membrane protein by visualizing the seven α-helices of Bacteriorhodopsin [Henderson & Unwin 1975]. Although there have been numerous structures solved since then, in comparison to X-ray crystallography, electron crystallography only contributes a fraction to the structural analysis of macromolecules. One of the reasons is the lack of large well-ordered 2D crystals. There have been advances in different steps of the crystallization workflow due to growing knowledge, better instruments, but also due to automation of several processes. The electron crystallography workflow includes the following steps:

1. Protein expression and purification

2. 2D crystallization

3. Sample preparation

4. Data collection

5. Image processing

6. Model building

The following sections will sketch the main steps of this workflow.

1.2.1. Protein Expression and Purification

The first step in an electron crystallography project is to get sufficient amounts of the protein of interest in a pure and homogenous state. For this purpose the target membrane protein can be either extracted from native tissue or membrane. This works for several types of native membrane proteins e.g. porins. But usually the quantities of membrane proteins in native tissues or membranes are low. The most common methods is to heterologously express the target membrane protein in a given prokaryotic or eukaryotic expression system [Abeyrathne *et al.* 2012]. There are several expression vectors containing promoters to produce a significant amount of recombinant proteins. Alternatively, expression of a target membrane proteins can be achieved in vitro by the cell-free system.

The proteins are subsequently extracted from the membranes, solubilized and purified with the help of detergents. Finally, proteins are reconstituted into a lipid bilayer of choice.

1.2.2. 2D Crystallization

Two dimensional crystallization is the task of reconstituting the recombinantly expressed protein to a 2D crystal. Although there exist some natural occurring 2D crystals (e.g. Bacteriorhodopsin found in the purple membrane [Henderson & Unwin 1975]), usually they have to be produced via crystallization.

Crystallization consists of three components: 1) the membrane protein, 2) phospholipids of the bilayer and 3) detergent molecules. After purification the membrane protein is in a detergent-solubilized state (Figure 1, state 1). In order to form a lipid bilayer the protein is mixed with in detergent solubilized lipids at a defined lipid-to-protein ratio. The crystal formation is now induced by decreasing the detergent concentration of this ternary mixture. During detergent removal, the exposed hydrophobic surface areas of the membrane protein strongly favors to be embedded in a lipid environment, rather than being exposed to water. The lipids themselves join together their hydrophobic sites to form a phospholipid bilayer in which the protein is inserted either during or before the bilayer formation. The detergent removal can be established by different methods: Dilution, dialysis, addition of Biobeads, or addition of cyclodextrin. It is important to note, that the crystallization mechanism is triggered by intrinsic features of the membrane protein and the experimental conditions. The success of reconstituting the membrane protein into the lipid bilayer depends on the choice of lipids, the lipid-to-protein ratio, the choice of detergents, the buffer conditions, possible additives, temperature, and the method of detergent removal.

The resulting 2D crystals can either form planar sheets, vesicles or tubes. Their planar repeating protein structure can be exploited to determine the projection map of the protein as described in subsubsection 1.2.5.

1.2.3. Sample Preparation

Once highly ordered 2D crystals of membrane proteins have been obtained, they have to be prepared for high-resolution data collection by electron microscopy. It is crucial to minimize distortions of the 2D crystal during sample preparation. Essentially there are three problems to solve: 1) Dehydration of the biological specimen leads to a collapse of the structure in the electron microscope's vacuum. 2) Biological specimens mainly consist of light atoms, which are prone to beam damage by electron scattering [Reimer & Ross-Messemer 1990]. 3) Flatness of the sample. Fortunately the method of cryo-EM can solve these issues. Freezing the sample fixates the biological specimen and keeps the membrane proteins in a native-like state, since membranes are naturally in an aqueous environment. And imaging at cryogenic temperatures (< 123 K) in the transmission electron microscope (TEM) increases the protection of the biological specimen from beam damage.
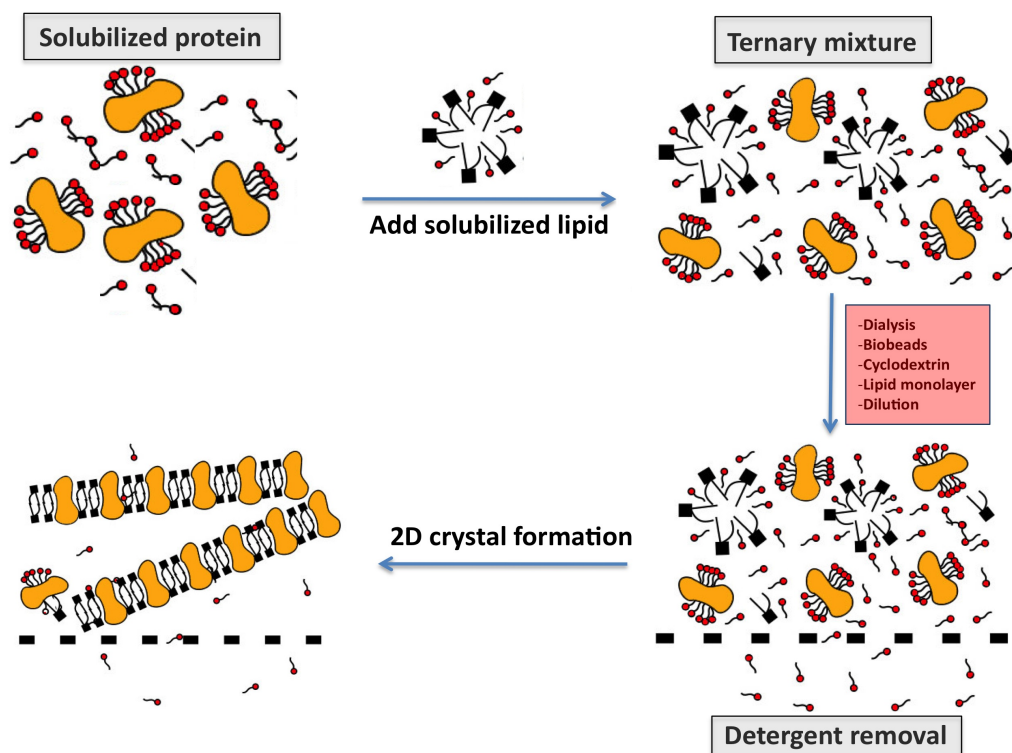
Figure 1: **Principle of 2D crystallization of membrane proteins.** First the purified protein (orange) is solubilized in detergent(red). Second, detergent solubilized lipids are added to the protein at a defined lipid-to-protein ratio. Third, the detergent is removed selectively. Fourth, this induces the formation of a lipid bilayer with the membrane protein embedded to form a 2D crystal.

5

For cryo-EM there are two different sample preparation methods for 2D-crystals: vitrification and sugar embedding. During vitrification a plunge-freezing device is utilized. This device plunges the fully hydrated sample in liquid ethane or liquid propane. Freezing rates faster than $10\,000\,\frac{K}{s}$ vitrifies the water of the sample, i.e. it solidifies without forming destructive ice-crystals. During sugar embedding, water molecules are replaced by less volatile compounds e.g. trehalose [Henderson & Unwin 1975], and afterwards frozen in liquid nitrogen.

To keep the 2D crystal as flat as possible the sample is adsorbed to a continuos carbon film, which also reduces beam induced movement. This can be achieved by the so-called back-injection [Wall *et al.* 1985] method or the carbon sandwich method [Gyobu *et al.* 2004].

1.2.4. Data Acquisition

Acquiring high resolution data of 2D crystals with an electron microscope is highly challenging. As mentioned before in section subsubsection 1.2.3 the image recording is also performed at cryogenic temperatures to protect the specimen from the beam damage. Additionally low-dose electron mode is used to further reduce radiation damage [Kuo & Glaeser 1975]. Modern electron microscopes are controlled via an acquisition software. During the data collection the specimen is routinely screened at low magnification to identify areas with good ice or sugar thickness that contain a 2D crystal. The image recording is then done at higher magnification with some defocus for image contrast enhancements after having focused on area within the vicinity of the crystal. Radiation damage prohibits from imaging a 2D crystal multiple times if high-resolution data should be acquired. Successive images are thus acquired from other 2D-crystals. To gain the 3D structure of the membrane protein that composes the 2D crystal, projections of the protein from different directions are needed. This is achieved by tilting the whole sample through the goniometer to a specified tilt angle. The in-plane tilt angle of the crystal cannot be specified and is therefore random. Hence, different crystals lead to different projection images even with the same specified tilt angle.

There are two modes to acquire high-resolution data of 2D crystals from a TEM: direct imaging and electron diffraction. Direct imaging has the benefit that it can record even small crystalline patches, whereas electron diffraction needs large well-ordered crystals. Diffraction mode utilizes one electromagnetic lens less than direct imaging and the resulting diffraction pattern can be understood as the power spectrum of the Fourier transform of the image. Therefore diffraction patterns only contain the amplitudes of the structure, but they are also shift invariant, hence specimen drift is not an issue. In direct imaging the spot-scanning mode [Downing 1991] can be used to reduce the amount of beam-induced image drift. Images and diffraction patterns can be recorded on photographic film or digital detectors. The re-
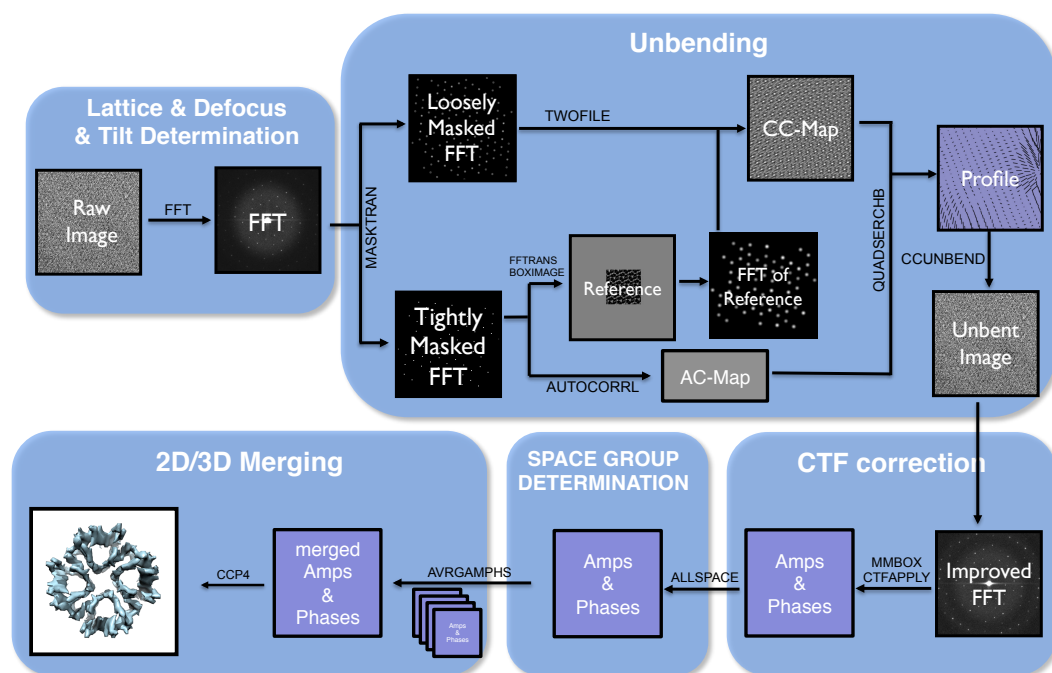
Figure 2: **Workflow of Image Processing for 2D Crystals**

cent advancements in digital recording have brought along *Direct Electron Detectors* [Bai *et al.* 2013, Campbell *et al.* 2012, Li *et al.* 2013] that even allow counting of the incident electrons. This has led to a new mode of data acquisition, where not one image but a sequence of images are recorded at low-dose, so called "movies". The images are then aligned to each other after acquisition to reduce beam induced movement of the specimen.

1.2.5. Image Processing for 2D Crystals

Cryo-EM images of membrane proteins suffer from a very low signal-to-noise ratio. This is not just due to the light atoms that build up the protein, but also they are small size. Therefore simple signal averaging is not sufficient to gain high resolution structures. But since a 2D crystal spans a regular array of proteins in a membrane, this systematic repetition can be exploited. The multiple copies of the same protein regularly distributed and equally oriented makes Fourier filtering and averaging possible. This means that one image of a well ordered crystal results in one averaged projection structure, i.e. one view of the membrane protein. To get further views, images from differently oriented crystals (tilt angle, lattice vector) have to be captured. With enough views a 3D reconstruction of the membrane can be calculated.

As simple as the method sounds, the actual image processing has to overcome several hurdles: deviation from perfect crystallinity, instrumental restrictions, limits and perturbations of the imaging systems. Specific image processing tools and methods have been developed to overcome these limitations. In the next section of this introduction the general image processing methods for 2D crystals will be presented (Figure Figure 2).

1.2.5.1   Processing of Individual 2D Crystal Images

Before projections of different orientation can be incorporated to a 3D reconstruction of the membrane protein, the structural information of every single crystal image has to be extracted. This processing step involves the following three tasks: 1) determination of crystal lattice, 2) correction of image perturbations introduced by the electron microscope, 3) identification and correction of crystal distortions.

Lattice Determination   When recording in diffraction mode the resulting pattern of the well ordered 2D crystals spans a regular grid of diffraction spots, the *lattice*. In direct space images of 2D crystal this lattice also becomes apparent when looking at the amplitudes of its Fourier transform. Determining the lattice means finding the two vectors that best span the regular diffraction spot pattern. From these vectors, the size of the unit cell and its orientation can be calculated. Through the lattice vectors one can index each diffraction spot by, what is known as, Miller index. It also allows to filter out noise by setting the amplitudes that are not in the vicinity of the spots to zero (masking). When dealing with images of a tilted crystal, the tilt angle and orientation can be calculated from the lattice distortion by comparison to the lattice of non tilted crystals. Multilayered crystals show multiple lattices. If the number of lattices is high i.e. the crystal has a tendency to stack, structure determination is nearly impossible. But if there are two or three crystal layers, the processing can even profit from that, because the crystallinity is usually stronger than in single layer crystals from the additional crystal contacts. However multiple layers also lead to overlapping diffraction spots, which have to be identified an discarded for further processing steps.

CTF Correction   The image recorded with an electron microscope is not a one-to-one projection of the specimen. The interactions of electrons with the specimen and the subsequent image contrast formation process in the microscope lead to modulations of the recorded image. This modulation causes coherent misrepresentation of frequencies of different ranges in the image formation model. In the Fourier domain this resembles a modulation of the original signal with a radially symmetrical,

oscillating function, the Contrast Transfer Function (CTF). To retrieve the original signal the image needs to be corrected for this modulation.

The CTF is defined by features of the imaging system: defocus, astigmatism, and spherical aberration of the objective lens, as well as the electron wavelength. Determining these features allows to some extent to computationally correct the image modulation. The spherical aberration constant is taken from the TEM as well as the specified accelerating voltage. The defocus and astigmatism are determined on the power spectrum (squared amplitudes) through identifying the zero crossing of the CTF, the Thon rings. The problem becomes even more difficult when the sample in the TEM is tilted, leading to a defocus gradient. For this case there is no efficient mathematical solution [Philippsen *et al.* 2007a].

Unbending   The concept of electron crystallography is to profit from the extensive repetition of the same aligned protein. Unfortunately a real 2D crystal of membrane proteins will deviate from perfect periodicity, and therefore limit the resolution of the computed projection structure. However, by correction of these deviations data can be refined and improved, resulting i.e. in higher resolved structures.

Correction of crystal distortions is a two step process: 1) identify crystal defects and 2) numerical correction on the image. Cross correlation of a reference with the crystal image detects the deviations of the crystal from the ideal lattice. The difference of the cross correlation peaks from the ideal lattice spots produces a distortion map. This vector map then specifies how the image is corrected by either shifting small patches of the image along the vectors or alternatively by pixel-wise interpolation. The reference for the cross correlation is initially calculated from the image itself by filtering out all the diffraction spots that do not lay on the perfect lattice. If the process is iterated a new reference can be calculated from the "unbent" image. Projection maps from already processed crystal data provide an even better reference (synthetic unbending).

1.2.5.2   3D reconstruction

Not just electron crystallography, but all EM imaging techniques (electron tomography and single particle analysis) share the same principle of 3D structure reconstruction by integrating 2D data sets recorded by the microscope. The central section theorem mathematically links the experimental 2D data with the computed 3D model. The theorem states that the projection of a 3D object holds the same information as the central (i.e., passing though the origin of reciprocal space) 2D plane cross-section of the 3D Fourier transform of the volume and is perpendicular to the projection direction [De Rosier & Klug 1968]. This implies that a 3D reconstruction of an object is possible by fully sampling the Fourier transform of

the object, which is equivalent to the full set of projections along all orientations. With the approximation of an image by a TEM corresponding to a projection of the sample, the sample's 3D density map can be estimated if sufficient projections of the sample with different orientations are captured.

We have seen in subparagraph 1.2.5.1 that the Fourier transform of a 2D crystal spans a lattice due to the repeating structure of the membrane protein. Hence Fourier space is the natural domain for processing 2D crystals. As the name states 2D crystals repeat in two dimensions, which means horizontally and not vertically. This reflects in the 3D Fourier transform of a 2D crystal by not having discrete points, but rather having continuous vertical lines. These so called "lattice lines" (h, k, z^*) occur at the Miller indices h and k defined by the lattice. Images of untilted 2D crystal fill the $z^* = 0$ plane due to central section theorem. Tilting the sample with respect to the electron beam results in a set of diffraction spots with the index h and k and the vertical height z^* defined by the tilt angle and the crystals in plane rotation. The goal is to collect enough images with varying tilt geometry to densely sample the lattice line data. Since the 2D crystal can only be imaged in a limited tilt range from $-70°$ to $+70°$, a cone in Fourier space can not be measured. This space of missing data is usually referred to as the *missing cone*.

Having acquired enough data points for each lattice line, we have to equidistantly sample the measurements before we can compute the 3D density map. The finite density of the crystal in vertical direction that manifests as lattice lines in the reciprocal space can be induced in the image domain by masking the crystal structure so that density above and below the protein are set to zero. This can be done by multiplication of the real space volume by a vertical rectangular mask, which in Fourier Space is a convolution with a *sinc* function along the lattice lines. Fitting lattice lines then becomes a linear combination of these *sinc* functions, which in turn allows equidistant sampling. Now reconstructing the 3D density map of the protein is just a Fourier back-transformation.

The sinc functions introduce vertical neighborhood correlation, that can be exploited to predict Fourier components in the missing cone, based on the oversampled Fourier components in the measured space [Gipson *et al.* 2011].

Single Particle 3D Reconstruction for 2D Crystals   The Central Section Theorem has one strong constraint, which is that a 2D crystal is planar. Unfortunately this assumption does not hold by the experimental data, since a 2D crystal is never perfectly flat. Therefore the protein orientation in a 2D crystal slightly varies. This variation in tilt geometry can be recovered with a Single Particle analysis method. The method refines the local varying tilt geometry of the crystal by comparing patches (particles) with projections of the 3D reconstruction in different orientations. This approach is presented in section 6.
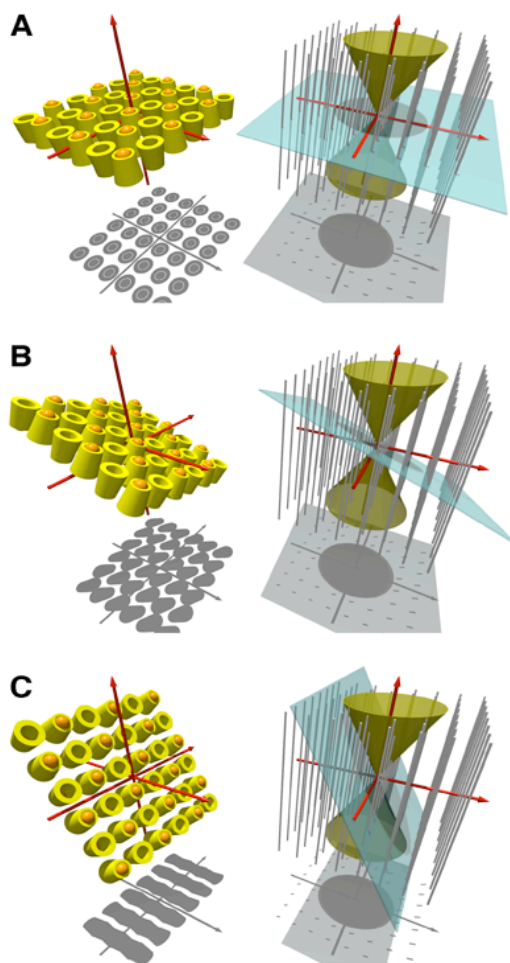
Figure 3: **A schematic view of a 2D crystal in the image domain (left) and the Fourier domain (right).** [Schenk *et al.* 2010] Three sample tilt angles are shown: (A) untilted; (B) 30° tilt; (C) 60° tilt.
The shadow on the left is the projection of the crystal along the *z*-axis. Ideally the projections resembles images gained from TEM (neglecting CTF). The corresponding Fourier transform of the image comes to lie in the central section depicted by a cyan plane (Central Section Theorem). Due to vertical limitation of 2D crystals, its Fourier representation extends the diffraction spots vertically to *lattice lines*. These continuos lines need to be densely sampled by the measurements of tilted samples. Since the samples cannot be imaged at tilt angles higher than ±70°, the lattice line values in the indicated missing cone region in Fourier space cannot be experimentally determined, resulting in the so-called missing cone.

1.2.6. Model Building

The 3D reconstruction of the image processing is an electron density map. To characterize the membrane protein fully, one would need an atomic model. Unfortunately cryo-EM does not achieve atomic structures. X-Ray crystallography on the other hand can reveal structure to atomic detail, but usually in a non functional state. Molecular dynamics flexible fitting (MDFF) now combines the atomic structures and the cryo-EM reconstruction to an atomic model in a functional state [Trabuco *et al.* 2008]. The method fits the atomic structure into the electron density map by molecular dynamics. The electron density map is used as a potential in the molecular dynamics simulation, driving the atomic structure is into the EM map. A common concern of the flexible fitting method is overfitting. In MDFF harmonic restraints are used to preserve the secondary structure of proteins and nucleic acids. Another hybrid method that fits atomic models into cryo-EM maps is implemented in real-space structure refinement program DireX [Schröder *et al.* 2007]. DireX prevents overfitting with means of deformable elastic network. The model building with this method will be illustrated in the structural study of the potassium channel Mlok1 in section 8.

1.3. Structure & Aims of the Dissertation

The aim of image processing for 2D crystals is to extract the structure of the repeating membrane protein from a large set of images. The main research focus of this thesis was to enhance image processing methods in order to determine the strucure of membrane proteins. The development of image processing methods in this thesis included to simplify and accelerate the process for the user and to yield more reliable structural information. Weaknesses of the existing image processing workflow were identified in several electron crystallography studies of unknown proteins structures. These studies also revealed the need for new methods to improve the resolution of the structural 3D data. Based on the image processing issues in these studies, new image processing methods could be developed and (directly) applied. Therefore this thesis is divided into two sections, a method part and an application part.

The methods part captures the whole image processing workflow, starting with the conventional algorithms of 2D crystal image processing and their implementation in the electron crystallography software `2dx` (section 2). The chapter depicts the steps to exploit the repeating structure within the 2D crystal to an average projection map of the inherent membrane protein. Subsequently 3D reconstruction of a membrane protein from projections of 2D crystals with varying orientation is described (section 3). At the same time the limitation of the 3D reconstruction due to physical restrictions on the imaging system and its discrepancy to the underlying mathematical model is shown. section 4 outlines how the image processing workflow

can be automated. The automation spares user interaction and thereby prevents incorrect processing due to human error, but also carries the risk that errors during automatic processing can go undetected. It results in an automated workflow that simplifies and accelerates the image processing of 2D crystals. Along with development of direct electron detectors automation complements the development to higher throughput processing. This progress has already been made in the field of X-ray crystallography and is certainly needed for electron crystallography. One advantage that 2D crystals in electron crystallography have in comparison to crystals in X-ray crystallography, is that the protein is embedded in a natural environment. This environment can be altered biochemically to induce conformational changes of the protein. The identification of these conformational changes with the help of 2D projection maps is the subject of section 5. It shows that the significance of the conformational changes has to be determined before they can be interpreted. section 6 depicts that the assumption of 2D crystals being perfectly flat does not hold. By using a single particle-based image processing approach the slightly varying tilt geometry within a 2D crystal can be determined. This in turn improves the resolution of the 3D reconstruction of the membrane protein.

The application part of the thesis consists of two structural studies of membrane proteins through electron crystallography. In both studies the here presented methods were applied. In section 7 the method to determine significant conformational changes was used to depict the substrate binding site of the citrate/sodium symporter CitS. The need of a reliable interpretation of the different CitS conformations is what led to the significant difference map method development. In the study of the cyclic nucleotide-modulated potassium channel MloK1 in section 8 we were able to compare the method for significant changes in projection maps to the 3D structural differences in density maps. The stucture of potassium channel MloK1 was also used to evaluate the improvement in resolution by the single particle based approach.

Part I.
Methods

2. Image Processing of 2D Crystal Images

This section was published as a book chapter in **Electron Crystallography of Soluble and Membrane Proteins: Methods and Protocols** [Arheit *et al.* 2013c].

Abstract

Electron crystallography of membrane proteins uses cryo-transmission electron microscopy to image frozen-hydrated 2D crystals. The processing of recorded images exploits the periodic arrangement of the structures in the images to extract the amplitudes and phases of diffraction spots in Fourier space. However, image imperfections require a crystal unbending procedure to be applied to the image before evaluation in Fourier space. We here describe the process of 2D crystal image unbending, using the `2dx` software system.

2.1. Introduction

Cryo-transmission electron microscopy (TEM) of biological specimens suffers from a very low signal-to-noise ratio in the recorded images, which makes it difficult to recognize the exact localization and orientation of smaller molecules in the images. Crystalline arrangement of the proteins allows overcoming this problem by exploiting the structural repetition and extracting the common signal from multiple noisy images. The crystallization of the proteins in two dimensions ensures that the proteins lie on a plane and are oriented in the same direction. Tilting this plane in the TEM results in different views of the proteins, which can be combined through image processing into a three-dimensional (3D) reconstruction of the protein.

Image processing of two-dimensional (2D) crystal data needs to cope with several problems: deviations from perfect crystallinity have to be recognized and corrected, as well as perturbations introduced by the imaging system. Specific image analysis concepts and algorithms have been developed over the years that accomplish these tasks. In this chapter we give an overview of the methods used to process 2D crystal images. We introduce the algorithms and software systems for image processing of 2D crystals, explain the methods involved, and present a guide for processing an image with the help of `2dx`.

2.2. Software Resources

The field of electron crystallography of membrane proteins was created through the work of Richard Henderson and Nigel Unwin (1, 2). Their developed algorithms were made available to the public in the so-called MRC programs for image processing (3-6). Electron crystallography has produced atomic models for seven transmembrane proteins and tubulin so far: BR (5), LHCII (7), AQP1 (8, 9), nAChR (10), AQP0 (11, 12), AQP4 (13, 14), MGST (15), and Tubulin (16). The 3D map from double-layered 2D membrane crystals of AQP0 from the Walz laboratory at 1.9 Åresolution not only showed the well-resolved lipids in the membrane around the membrane proteins, but even allowed localizing water molecules embedded in the hydrophobic core of the membrane (12).

The data for all these structures were processed with the MRC programs. In the majority of these projects, several years of work were needed to construct a high-resolution map from existing 2D crystals. AQP0 and AQP4 were solved from electron diffraction patterns alone, using molecular replacement methods. Their structure determination could be done significantly faster. It did not require the computer processing of recorded images, since the homologous structural model of AQP1 was used for homology modeling.

The MRC software is composed of a collection of standalone programs, most of them written in Fortran-77, which contain a rich repertoire of algorithms and solutions. The MRC programs are an implementation of brilliant algorithms and theory, so that it comes as no surprise that their usage is not simple. Detailed knowledge about these programs can best be gathered from other advanced users, or by "reading" the Fortran-77 code itself. The MRC programs also include a graphical image display program called Ximdisp (17).

The SPECTRA software from the ICE package was a program that generated Unix Shell scripts that launched an early version of the MRC programs (18, 19). SPECTRA contained its own image visualization program, and facilitated the usage of the MRC programs significantly. To our knowledge, SPECTRA is no longer supported. Similar functions are found in the GRIP system by Wilko Keegstra in the University of Groningen (unpublished), and Sven Hovmoeller's Calidris software package, which reproduces some of the MRC functionality in a commercially distributed PC software package (`http://www.calidris-em.com/`). A recent effort is the Image Processing Library and Toolbox (IPLT), a new soft- ware system that was introduced by Ansgar Philippsen in the laboratory of Andreas Engel in Basel. IPLT is written in C++ and Python (20, 21), and is available at http://iplt.org. Other program systems that can partly be employed for the image processing of 2D crystal images of membrane proteins are XMIPP (22, 23), Spider (24), bsoft (25), and others. A collection of links to the different systems can be found at `http://en.wiki-books.org/wiki/Software_Tools_For_Molecular_Microscopy`. We here describe the processing of 2D crystal images with the `2dx` software package (26, 27).
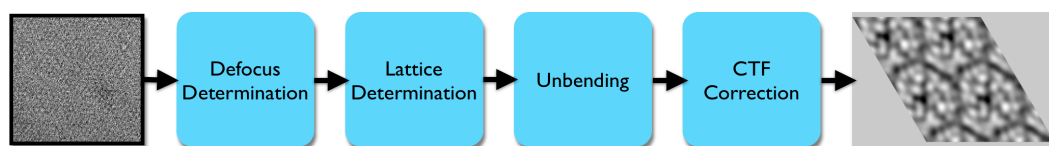
Figure 4: Block diagram of the processing flow for one image.

`2dx` is based on the MRC programs, which are run as background programs behind a graphical user interface (GUI). The installer for `2dx` includes an adapted ver- sion of the MRC programs, which have been modified to interface with the `2dx` GUI, and in some cases additional functions were added to the MRC programs. `2dx` features a user-friendly GUI that assists the user in the management of an image processing project, that provides user-guidance, streamlines respectively auto- mates parts of the image processing, and allows detailed monitoring of the processing progress and results. `2dx` maintains a data structure in form of simple text files, where image-processing parameters for each image, as well as for the entire processing project are stored. `2dx` also includes a single-particle processing module for 2D crystal images, which uses a maximum likelihood (ML)-based algorithm. This method may give a significantly better result, when the crystal quality does not allow high-resolution work through the unbending algorithm. This ML software is fully integrated into the `2dx` package, and produces results as Fourier-space amplitude and phase values, to allow merging with other MRC-based processing results in Fourier space. We distribute `2dx` as open-source software under the GPL at `http://2dx.org`.

2.3. Processing of Individual Images

The processing of 2D crystal images involves several steps, some of which are listed in the simplified block diagram in Fig. 4. The processing involves the following steps:

1. Defining basic processing parameters

2. Calculating the Fourier transform of the image

3. Measuring the defocus in the image center, and in different locations in the image

4. Calculating potential specimen tilt from the defocus gradient

5. Determining the 2D crystal lattice

6. Refining potential specimen tilt from distortions of the 2D crystal lattice

7. Determining a spotlist of significant Fourier reflections

8. Determine lattice distortion vectors and perform a first image unbending (Unbend I)

9. Iteratively refine the lattice distortion vectors, do refined unbending (Unbend II)

10. Extract amplitude and phase values for each Fourier reflection from the unbent image

11. Correct the list of amplitudes and phases for the instruments CTF

12. Calculate a final projection map from this image

These steps are now discussed in detail, using the `2dx` software package as an example.

2.3.1. Defining Basic Processing Parameters

Chapter 11 describes the management of a 2D crystal image-processing project. We refer to that chapter for the initialization of the processing environment, how to import images into `2dx`, and how to open them with `2dx_image`.
Upon opening an image with `2dx_image`, the user needs to define the basic processing parameters that are already known for this project. These concern the magnification under which the image was recorded on the microscope, and the pixel size of the recording medium. `2dx_image` will then calculate the image pixel size from these values. Electron microscope parameters also need to be defined correctly, concerning for example the acceleration voltage, the spherical aberration (Cs) value of the objective lens, and some other parameters. As described in Chapter 11, these values can be saved as project default values, so that for other freshly imported images they are already set correctly.

2.3.2. Calculating the Fourier Transform of the Image

The technical quality of an image should be verified before starting the image processing. Depending on the method by which the image was recorded, it may be available in byte (8-bit unsigned), short integer (16-bit unsigned), or real (32-bit floating point) modus. Conversion of an image from a recording medium (TIFF or MRC are currently supported by `2dx`) to the internal file format (`2dx` uses the MRC 2000 format) is trivial, once it is implemented, but may pose a problem the first time a new image source is used. Typical errors occur when images are saved as 16-bit signed integer, but are interpreted as 16-bit unsigned integer values, or vice versa. Such images then show the darkest pixels in the displayed maps as white

pixels. Another possible source of error is that images recorded on film might be scanned with a film scanner as positive or negative film transmission, instead of the film's optical density (OD). Only the latter is proportional to the protein density on the film so that the logarithm of the pixel values in the transmission files need to be calculated in order to produce OD values.

The first script "Calculate FFT" of `2dx_image` produces among other files a pixel histogram of the input image. The histogram should show a symmetric, Gaussian peak that is centered somewhere in the middle of the histogram and that is not truncated at either end. If the histogram is not symmetric but tilted to either side, then the data are most likely film transmission values that have not yet been transferred into OD values. If the histogram is truncated at either end, then the film digitization or file format transfer was not done correctly, or the image is over or under exposed. This script also calculates down-sampled versions of the input image, and Fourier transforms of the original and the down-sampled images.

2.3.3. Measuring the Defocus in the Image Center

The recorded image is not a 1:1 projection of the specimen. Instead, it rather represents a collection of measured numbers, which in their entirety contain the information about the high- resolution structure of our sample. The imaging process in the TEM is a complex phenomenon, which can be mathematically described by a contrast transfer function (CTF) (28). This function describes in Fourier space, how different resolution components of the structure were transferred onto the image: The amplitude of the Fourier transformed image underwent strong modulations, and the phases were shifted by 180° where the CTF had a negative value, and may have been affected by additional alterations due to beam-tilt (29-31). The resolution rings where the CTF is of value zero are the so-called Thon rings (32). The CTF is dependent on the defocus under which the image was recorded and therefore strongly varies across an image of a tilted specimen (5, 33).

For 2D membrane protein crystals that are weak phase scatterers to the electron beam, the CTF is an oscillating function that depends among other factors on the defocus and astigmatism of the objective lens, the electron wavelength (and therefore the acceleration voltage), and the spherical aberration constant (Cs) of the objective lens. The CTF is dampened by an envelope function, which depends on the chromatic and spatial coherence of the electron beam, and which is affected by specimen drift and vibration during the electron exposure. In addition, our micrograph is covered by a high amount of noise, which stems from the low number of electrons used to record the image (shot noise), and detector noise. In addition, the recording medium usually modulates the image with a further transfer function, the detector modulation transfer function (MTF), due to the extended point spread function of the electrons on the detector material.

For the processing of a 2D crystal image, correct definition of the acceleration voltage and Cs value are needed, and the defocus and astigmatism for each image have to be measured experimentally from the image. This, together with knowledge about the beam tilt (fitted during the merging process), allows defining the resolution rings in Fourier space (Thon rings), where the contrast reversal occurs.

As described in subsubsection 2.3.9 below, these contrast reversals will be corrected by phase-flipping, which adds 180° to the phase values of the affected zones in Fourier space. The effects of the CTF's amplitude oscillations can be computationally corrected, by multiplication of the amplitudes with a corrective factor. A correction for the dampening envelope function on the CTF and for the detector MTF is usually approximated by employing a negative temperature factor during calculation of the final map, which increases the amplitude of the higher-resolution components.

To determine the defocus and astigmatism one can use the program CTFFIND3 (34). The program's objective is to fit the oscillatory function of the measured power spectrum. It is also integrated in the `2dx` program suite and used in its "Determine Defocus" script. CTFFIND3 subtracts from the power spectrum the smooth background estimated from the image. It then fits a two-dimensional CTF to the background-free power spectrum to determine the defocus and astigmatism. In CTFFIND3 the fit is found by maximizing the correlation coefficient of the observed power spectrum and the estimated squared CTF. The maximum correlation is found by doing a grid search of defocus values and astigmatic angle. In `2dx` the resulting defocus and astigmatism can be examined and manually adjusted in the power spectrum of the image. The defocus is defined in Ångströms in two orthogonal directions **a** and **b** in the image, while a third parameter defines the angle in degrees between the horizontal X-axis and the direction **a**. As always, a positive angle describes a counterclockwise rotation.

2.3.4. Measuring Potential Specimen Tilt

The defocus is a measure for the distance between the sample and the focal plane of the objective lens. In the case where we record images of tilted samples in the TEM, the defocus varies across the image, due to the varying distance to the objective lens. This defocus variation can be used to determine the tilt geometry of the sample. The program CTFTILT (34) is used to determine tilt axis and angle by measuring the defocus at several locations of the image. The tilt axis is defined by a line with constant defocus in the image. In practice the tilt axis direction is determined by minimizing the variance of defoci between sub-tiles in the image. Once the tilt axis is estimated, the tiles along the tilt axis are used to determine the defocus and astigmatism for the axis. Having found the tilt axis allows determining the tilt angle. Therefore, CTFTILT performs a search on the tiles of the entire image, which are assumed to lie in a single (tilted) plane. Thus the defocus can be formulated as a

function of its image coordinates and the angle between tilt axis and X-axis and the tilt angle. The search range for the tilt angle is usually between $\pm 65°$. In the end, the estimated values for defocus, astigmatism, tilt axis direction and tilt angle are refined by maximizing the correlation between observed power spectra and the CTF as described earlier with CTFFIND3. An alternative way to determine the tilt geometry is also implemented in `2dx_image`, which cuts the image into 7×7 sub-images, on which the script uses CTFFIND3 to calculate the central defocus on each sub-image. A separate function then fits a tilted plane to those 49 defocus measurements, while excluding outliers from the fitted plane.

The tilt geometry for the specimen plane is defined by two variables, TLTAXIS and TLTANG, in degrees. TLTAXIS defines the angle between the horizontal X-axis and the tilt axis in the recorded image, where positive values describe counter-clockwise rotation. (A value of TLTAXIS=+30 defines a tilt axis that is "pointing at 2 o'clock.") TLTAXIS is defined between $-89.999°$ and $90.000°$. TLTANG defines the tilt angle around this axis. In the MRC programs and in 2dx, the origin of an image is considered bottom left. The image pixel in the bottom left corner has the coordinates (0,0). Pixels at the bottom of the image starting bottom left and walking right are defined as (1,0), (2,0), (3,0), ..., (x,0). If the tilt axis is horizontal (TLTAXIS = 0), and the underfocus gets stronger as you go up the image (towards higher y values), then TLTANG is positive. TLTANG is defined between $-89.999°$ and $89.999°$. Note that TLTAXIS and TLTANG are independent of any crystal lying on the sample plane.

2.3.5. Determining the 2D Crystal Lattice

Image processing for single particle electron microscopy and electron tomography requires localizing proteins or other image features in the images, which usually is done by cross-correlation of the image with a reference image (Fig. 5). This works for large protein complexes, but becomes difficult for small particles, if the image has a low signal-to-noise ratio (SNR). In electron crystallography, the alignment of the proteins is done biochemically, by inducing 2D crystal formation. Even with a very low SNR in the images, the averaging of the signal from several proteins is then possible, because the image processing can rely on the protein being approximately at the predicted position. This makes it possible to get high-resolution structural data from small, crystallized proteins, even though the images of the individual proteins have an SNR so low that they would be very difficult to detect individually in the image.

The periodic alignment of the projected proteins in the 2D crystal image is the cause of diffraction peaks in the calculated Fourier transform of the image. The basic unit cell repeat in the image defines the first order diffraction in the Fourier transform. Higher-resolution details of the protein structures cause the lattice in the Fourier
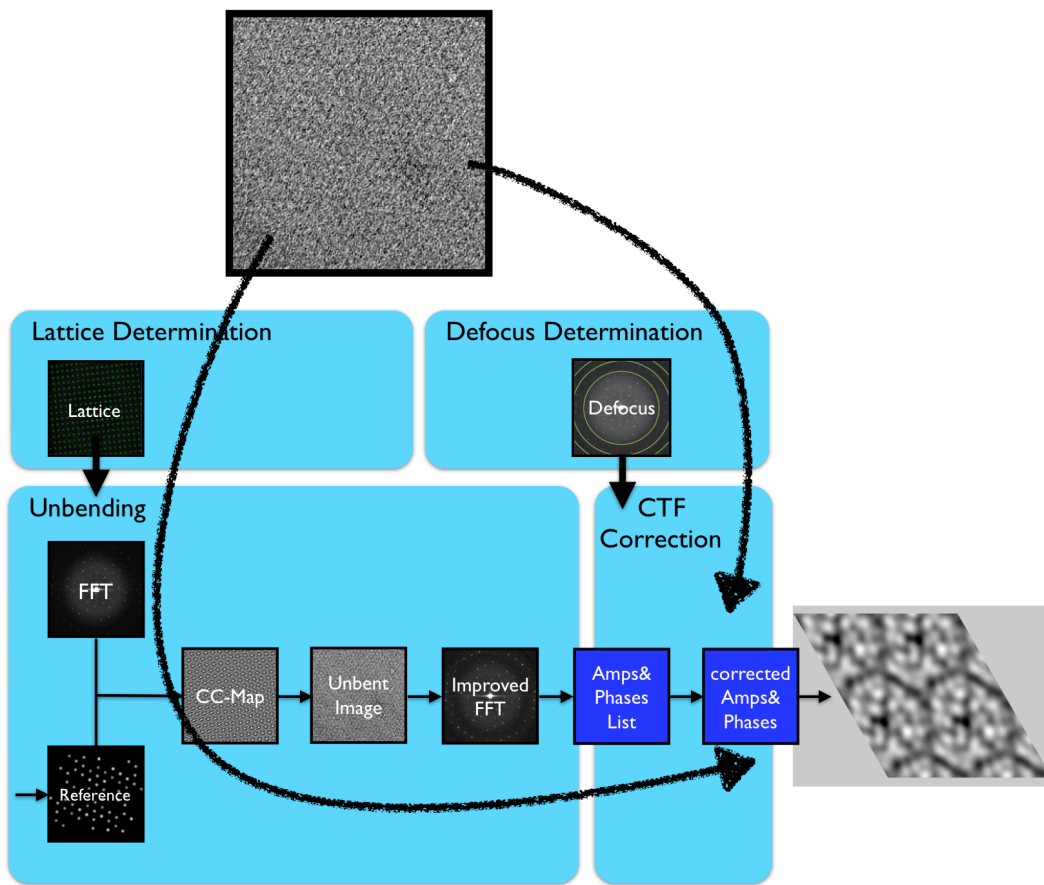
Figure 5: Flowchart of the image processing steps of a single 2d crystal image.

transform to show also higher-order diffraction spots. The Fourier transform of a good cryo-EM image of a perfectly ordered 2D crystal will show a lattice that reaches with sharp diffraction spots out to very high resolution.

In practice, however, the 2D crystals are not perfect. Crystals have defects, in form of dislocations in individual unit cells, or distortions from shear stress or other stress on the crystalline lattice. Crystals can be polycrystalline, so that several "2D crystal grain" form a patchwork. In addition, the electron microscope can introduce image distortions, often in form of a pincushion or spiral distortion, which then not only result in displacement of the proteins from the predicted lattice, but also induce slight variations of the image magnification. The Fourier transform of such distorted crystals then only shows low-resolution diffraction peaks, while the higher-resolution peaks are smeared out, and very high-resolution peaks are so strongly smeared out that they become invisible to our detection methods.

Richard Henderson et al. introduced the concept of 2D crystal lattice unbending. The lattice distortions in the image are recognized by software, and computationally corrected, by compiling a corrected output image that is constituted with segments from specific image locations in the input image. The Fourier transform of the resulting "unbent" image would then again show very high-resolution diffraction spots, which can be evaluated for their amplitude and phase information.

For this to be possible, the first step is to find and define the lattice. This is best done in reciprocal Fourier space, by a process called "Lattice Indexing." As can also be done with `Ximdisp` (17), `2dx_image` features a full-screen browser that allows manually indexing a reciprocal lattice. In a displayed Fourier transform in `2dx_image`, the user would manually click on an identifiable diffraction peak, enter its Miller index coordinates, repeat the same for a few more linearly independent peaks, and have the software calculate the best fitting reciprocal 2D crystal lattice. From then on, the user can add more diffraction peaks, and the software will propose the corresponding Miller index automatically.

`2dx_image` also includes two different algorithms to automatically determine the reciprocal lattice. Both algorithms start by calculating an origin-shifted averaged power spectrum, which shows any present lattice in the Fourier transform with much better clarity and with less absences in the low-resolution spots. This is done by the program `2dx_peaksearch`, which also identifies all peak positions in the resulting averaged power spectrum, as described in (35).

The first lattice identification algorithm is using a program `2dx_getlat`, which calculates difference vectors among all identified peak positions. The two linearly independent shortest difference vectors that occur with a higher frequency between this set of peak positions are assumed to be the base vectors $\mathbf{u}$ and $\mathbf{v}$ of the reciprocal lattice. `2dx_getlat` doesn't need any prior knowledge about the crystal project, but may fail in case of systematic absences in the reciprocal lattice, or in case of multiple overlapping lattices.

The second lattice identification algorithm implemented in `2dx_image` is using a

program `2dx_findlat`. Using prior knowledge about the (project-wide) real-space crystal lattice dimensions, the magnification, pixel size, and the tilt geometry (determined from the defocus gradient), this program calculates a predicted reciprocal lattice. The predicted lattice is then rotated around in the Fourier transform by 360°, taking distortions from the tilt geometry into account and varying the lattice parameters and magnification, while for each orientation the program compares the predicted peak positions with the actually measured diffraction peaks. If a theoretically predicted lattice fits to the experimental peaks, the lattice is refined to the identified peak locations by a least-square algorithm. These test lattices are then evaluated by a score value calculated from the sum of peaks that lie within the listed peak coordinates. The lattice with the best score is identified as *first reciprocal lattice*, and its diffraction peaks are removed from the list of peaks. The process is then repeated with the remaining diffraction peaks, to identify potential additional lattices, so that more than one reciprocal lattice can be identified. `2dx_findlat` requires prior knowledge about the crystal and approximate tilt geometry, but then promises a higher reliability to find lattices also in difficult cases.

Once the lattice is identified, and if the recognized tilt angle is at least 25°, the distortion of the reciprocal lattice from the lattice for a non-distorted lattice is used to refine the tilt geometry. At low tilt angles, this lattice distortion is minimal (proportional to $1/\cos(\text{TLTANG})$), so that the more reliable source of the tilt geometry is the defocus gradient. At higher tilt angles, an error in magnification for example results in a wrong calculation of the tilt angle, so that the lattice distortion is a more reliable basis to calculate the tilt geometry. In this case, the lattice distortion allows for calculating the tilt geometry, but for geometric reasons in the general case two different tilt geometries can explain any given lattice distortion. In addition, the sign of the tilt angle TLTANG is not defined by the lattice distortion, so that four different tilt geometries could explain any given lattice distortion. For this reason, the script takes from the four possibilities the tilt geometry that is closest to the one determined from the defocus gradient as the refined tilt geometry. This obviously bears a possibility for error, if two possible solutions lie closely together, if the defocus gradient was wrongly measured, or if the lattice was wrongly indexed.

Once the refined tilt geometry for the sample plane is known (in the variables TLTAXIS and TLTANG, as described above), it is possible to describe the orientation of the crystal with respect to that tilted plane. The variable TLTAXA describes how the crystal is seen in the image with respect to the tilt axis, by measuring the angle between the tilt axis and the first lattice vector $\mathbf{a}^*$ in the recorded image. The variables TAXA and TANGL finally describe the tilt geometry from the point of view of the crystal lattice: TAXA measures the angle between the tilt axis and the vector $\mathbf{a}^*$ in the 3D space of the crystal (this is similar but not the same as TLTAXA), and TANGL measures the tilt angle for the crystal (the sign may be different than that of TLTANG). Only the latter two variables, TAXA and TANGL, are important for the 3D merging.
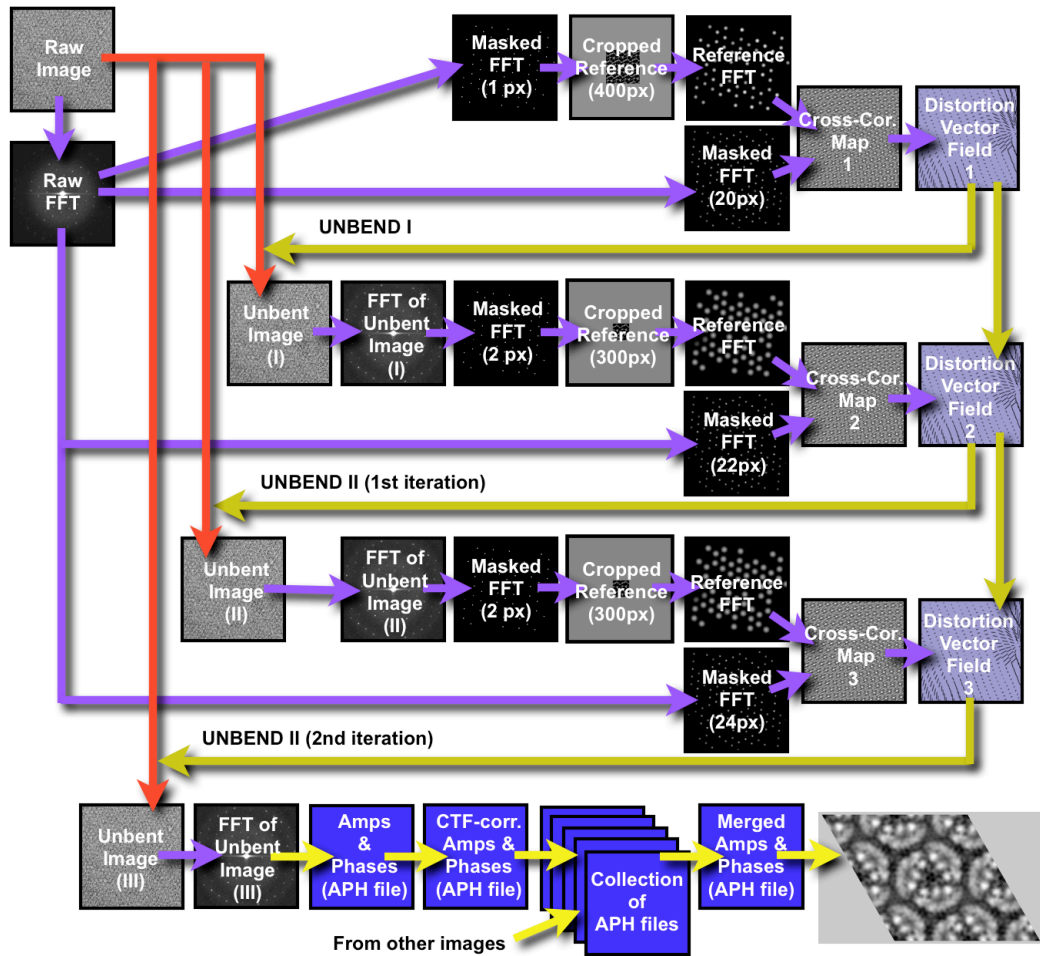
Figure 6: Detailed flowchart of the unbending process for a single 2D crystal image.

2.3.6. The Unbending Algorithm

Figure 6 introduces the detailed flow diagram of the image processing for one image, as it is implemented in `2dx_image`.

Let us start analyzing Fig. 6. The raw image (top left) is Fourier transformed (Raw FFT), and the resulting Fourier transform is masked on the lattice positions with a very narrow masking radius of 1 pixel, so that only the basic periodicity of the crystal lattice remains. The masked FFT is back-transformed into real space, and a central patch of for example 400 pixels width is masked, while the remainder of the image is grayed out. This gives our first Cropped Reference that shows a very rough approximation of the crystal unit cells. This is again Fourier transformed for later

25

use. In parallel to this, the Raw FFT is also masked with a wider masking radius (e.g., 20 pixels), so that the majority of the noise in the image is removed, without loosing too much of the structure signal. This masked Fourier transform is then multiplied with the complex conjugated Fourier transform of the Cropped Reference. The result is back-transformed into real space, to give a cross-correlation map (CC-map) (Cross-Cor. Map 1) between the Cropped Reference and our raw Fourier filtered Image. Not shown in Fig. 6 is a further refinement step for the CC-map: That map is cross-correlated with the autocorrelation function of the reference, to identify true unit-cell correlation peaks in the CC-map with sub-pixel precision. The result is a refined CC-map that serves the same purpose as before.

That CC-map (Cross-Cor. Map 1) shows peaks for each location of a unit cell. Those peaks can be quite weak and the CC-maps can be quite noisy, so that a simple peak search algorithm would not find the correct lattice node peaks in those maps. The MRC program QUADSERCH is evaluating those CC-maps, and is performing a sophisticated trick: Starting from the middle of the CC-map where the reference should still lie precisely on the crystal lattice, QUADSERCH uses the next for example seven peaks along the predicted lattice in the CC-map, to extrapolate, where the next 8th peak should be, and then only searches the CC-map in the immediate vicinity of the predicted position for the real presence of that 8th peak. This search is only done within a small radius of for example 9 by 9 pixels, where the parameter RADLIM can adjust if the search field should be round or elliptical. Once that predicted 8th lattice peak in the CC-map is found, it is used to predict peak number 9, and the search in the vicinity of the predicted location is started again. This process is repeated, whereby the last for example seven lattice nodes are always used to predict the next node, until the found lattice nodes cover the entire CC-map. The deviation of the found peak positions from the perfect lattice is stored as a vector field (Distortion Vector Field 1).

The MRC program CCUNBEND is then using the determined Distortion Vector Field, to perform an image unbending of the raw image (UNBEND I). This results in an Unbent Image (I), in which each 2D crystal unit cell is moved into a location that corresponds roughly to the predicted lattice position. The Fourier transform of such an unbent image now shows much sharper diffraction peaks (FFT of Unbent Image (I)), so that the narrow masking can be done much more efficiently (now with a mask radius of 2 pixels), to produce a reference map that shows at much better clarity the crystal unit cells. Of this clearer reference, a smaller Cropped Reference can be cut (Cropped Reference (300 px)), so that lattice distortions can be followed more easily. This Cropped Reference is then cross-correlated with the raw image, to produce a refined Cross-Cor. Map 2, which is evaluated by QUADSERCH to refine the Distortion Vector Field 1, giving the Distortion Vector Field 2. This refined distortion vector field is then used by CCUNBEND to unbend the raw image (UNBEND II, 1st iteration). This process can be repeated a few more times (UNBEND II, 2nd (and more) iterations), until a final unbent image is obtained.

The resulting Unbent Image (III) is then Fourier transformed (FFT of Unbent Image III), and this transform now shows sharp diffraction spots. These spots can be evaluated for the Amplitudes and Phases of the diffraction peaks (APH file), which can be CTF-corrected, merged with the APH files from other images, and a resulting final map can be computed from the merged APH files.

2.3.7. Determining a Spotlist

The above-described unbending process involves masking of Fourier transforms during the calculation of cross-correlation maps. This masking is done on the diffraction peaks in the Fourier transforms such that only diffraction spots remain that are listed in a so-called spotlist. These spots are masked with a circular mask of a given radius (e.g., 20 pixels), while the rest of the Fourier transform is set to zero. The spotlist allows identifying the spots with good signal-to-noise (SNR) ratio, which should be used for the determination of the lattice distortion. Diffraction lattice positions that do not show peaks would only contribute noise to the cross-correlation map and should therefore not be listed in the spotlist. Also, spots that lie not only on the reciprocal lattice of interest, but also lie on or close to a peak from a second interfering lattice in that Fourier transform, should not be part of that spotlist. This can occur, when two 2D crystals are overlaid in the image, so that the Fourier transformation would show two sets of peaks, forming two lattices.
`2dx_image` contains a script to automatically determine a spotlist. This is done by running the MRC program MMBOX on the Fourier transformation of either the raw image (script Get Spotlist for Unbend I) or of the unbent image (script Get Spotlist (complete)), and all diffraction peaks that have an SNR above a certain threshold (above a defined IQ value) are taken as good spots for the spotlist. `2dx_image` also contains a script to do some spotlist arithmetic: Spots that lie within a certain proximity (e.g., of 7 pixels) to peaks from a second lattice, can be deleted from a given spotlist with the script Refine Spotlist.
`2dx_image` also allows manually selecting, viewing, or editing spots in the full-screen browser, by clicking spots in the Fourier transform on or off.

2.3.8. Unbending

As described above, the 2D crystal image is unbent to remove crystal distortions. This unbending is done with a profile, which is refined in several iterations. However, during this process, the original image is usually only unbent once, but with an iteratively refined profile. Re-unbending of an already unbent image would be dangerous, in the sense that image noise might get aligned to any high-resolution feature in a given reference, after which the statistics of the processed image might

appear nice, but the result has little to do with the true protein structure.

If the distortion vector plot resembles a "fur" with smoothly flowing ordered lines, then the unbending process most likely picked up the true distorted lattice of the 2D crystal. However, if the distortion vector plot instead resembles chaotic vectors that point into random directions, then the unbending process might have just followed noise correlation instead of the protein. Given that the vector plot appears coherent, the quality of the unbending procedure can be evaluated by analyzing the signal-to-noise ratio (SNR) of the diffraction spots in the Fourier transform of the newly obtained image. The MRC programs (6) have introduced a quality measure of the spots in form of an Image Quality (IQ) value, which roughly corresponds to the inverse SNR of the diffraction spot: The ratio of a diffraction peak intensity and its local background is translated into an IQ value, such that an IQ=1 means that the peak is at least seven times above background. IQ=7 means that the peak signal is at least as high as the background (the peak height is calculated after background subtraction). A spot of IQ=8 still has a SNR that is one or slightly positive, and a spot of IQ = 9 has a negative signal, which means that the Fourier transform shows a dip instead of a peak in that location. In `2dx` (26, 27) we introduced a further extension of this concept, by calculating with an empirically defined formula a single quality value QVal for the entire image. The formula to calculate QVal includes among others a weighted sum of the IQ values of the image. This QVal value quantifies the quality gained through unbending, and can be used to judge the quality of an unbent image in general. This can be helpful later in the merging process, where good images can be recognized by their high QVal values, while bad images with low QVal values can be excluded from the merging process. The QVal value can also be useful to automatically refine parameter settings, as done in several "Custom Scripts" for parameter refinement in `2dx_image`.

An alternate reference for cross-correlation and unbending can be calculated from a merged data set. The merged set contains the information of all the individually processed images and therefore is usually capable of producing a better reference than if obtained from a single image. Given that the specimen tilt geometry and defocus were correctly determined, the script "Synthetic Unbend" can be used, to run the MRC program MAKETRAN (36) to create a synthetic reference from a merged dataset, and use that reference for the lattice unbending. Applying this so called "synthetical unbending" to every individual image will result in an improved merged map. Hence this should be repeated until the unbending of the individual image does not yield any quality improvements.

2.3.9. CTF Correction

The Fourier transform of a correctly unbent image shows sharp diffraction peaks that extend to high resolution. The amplitude and phase values of these peaks are then

evaluated by the MRC program MMBOX, which also records the local background signal around each diffraction peak. The ratio of the signal to the background is translated into an IQ value for each diffraction peak, and can also be used to calculate a Figure of Merit (FOM) for each peak.

This list of Amplitudes and Phases (APH file) is still modulated by the microscope's contrast transfer function (CTF). The effect of the CTF has to be deconvoluted from the image, which can be done in Fourier space with the MRC program CTFAPPLY, by dividing (in form of a Wiener filter to avoid division by zero) the evaluated APH file by the CTF.

This is done in `2dx_image` with the script "Correct CTF." The result is an APH file, where the phases of spots in the negative CTF regions have been flipped by 180°, and where the amplitudes have been corrected through Wiener filtration.

2.3.10. Calculation of a Projection Map

The result of the processing of one image is a final APH file that lists for each Miller index (H,K) an amplitude and phase value, that is accompanied by a figure of merit. This information can then be used to calculate a corresponding real-space projection map, which also can be symmetrized in case of crystal symmetry. This is done by the script "Generate Map," which will employ MRC and CCP4 programs for this purpose. This script can also run a `Spider` script (24), to produce larger multi-unit-cell final maps. This, however, requires that Spider be installed on the computer in addition to `2dx` and `CCP4` (`http://www.ccp4.ac.uk`, (37)).

2.4. Application Notes for `2dx_image`

This section describes the implementation of the presented image processing methods in `2dx` (26, 27) for Version 3.3.0. The processing of an individual image in `2dx` is handled by the program `2dx_image`. In `2dx_image` the whole processing workflow can be automated, which is shown in section 4. Here the focus lies on how `2dx_image` deals with the processing steps of defocus determination, lattice determination, unbending, and CTF correction.

1. At first one would have to choose the image to be processed. Usually in `2dx` the images are imported in the program `2dx_merge` (see section 4) and `2dx_image` is launched for that image by double clicking a listed image in the GUI of `2dx_merge`.

2. The GUI of `2dx_image` consists of different panels displaying data, process-
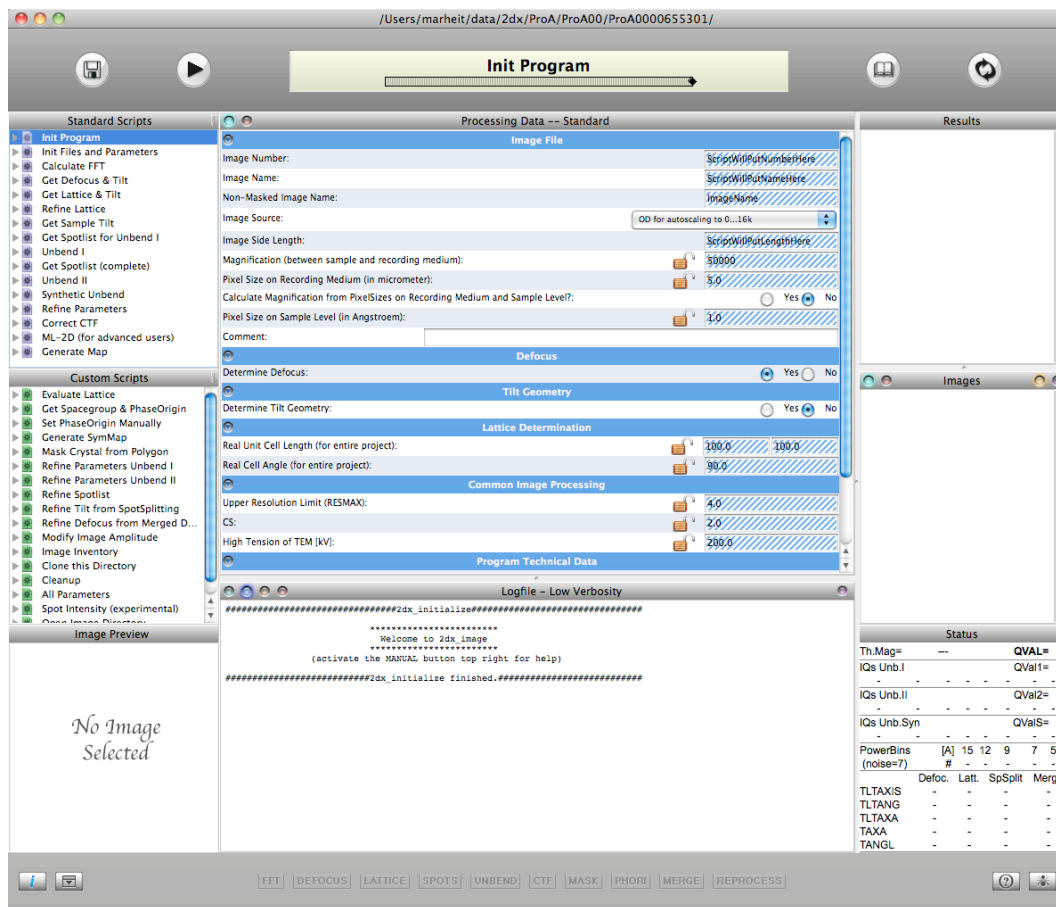
Figure 7: The GUI of `2dx_image` that holds processing data, parameters, and processing routines.

ing parameters and image-processing routines, as shown in Fig. 7. When `2dx_image` is launched it automatically runs the "Initialization" script as displayed by the progress bar in the top section of the GUI.

3. The script "Init Files and Parameters" fills in the "Image File" parameters in the "Processing Data" panel, e.g., the image side length is taken from the image header displayed in the "Image Header" panel.

4. The image parameters that need the user's attention have striped blue background. It is for example essential to set the correct values for "Magnification" and the "Pixel Size on Recording Medium."

5. If an image of a non-tilted sample is processed the switch "Determine Tilt Geometry" should be set to "No" in the "Tilt Geometry Section" of the "Processing Data" panel. This will prompt the program to not try to determine any specimen tilt, which would be time consuming and might lead to wrong results.

6. There is a help function for the parameters in the "Processing Data" section. A right click on any parameter will display its description and also give a link to the online documentation. Each script also has a help text, which is shown when the user clicks on the "Manual" button in the top section on the right side of the progress bar.

7. The "Images" panel on the right will show the resulting images and other files that are produced by an executed script. These files will be displayed under their nicknames. The full file names can be seen when placing the mouse pointer over the nickname, or by changing the display mode of that panel. Double-clicking a nickname (e.g., "Non-Masked Image") will show that image in the full screen viewer, where the image display can be adjusted, and additional information through the "Navigator" pull-down menu can be displayed. For example under "Selection based FFT" the user can activate a life FFT of the selected region, which can then be moved over the image to search for a well-diffracting image area. We call this function also a "virtual laser diffractometer." The displayed life FFT can be adjusted in brightness with the keys 1...9, and zoomed with the keys - and + (or =).

8. On the left of the GUI is the "Standard Scripts" panel. As the name states, this holds the scripts that should normally be executed sequentially. Each of these scripts represents an individual processing routine. When double-clicking a script entry, the corresponding CShell script template will open in a text editor. When launched, `2dx_image` will incorporate the parameters from the GUI into the script template, create a valid CShell script, save it in the local image directory under the sub-directory "proc," and execute it there.

`2dx_image` will then monitor the running of the script, and display the output (logfiles, resulting image files, and determined parameter values) in the GUI.

9. `2dx_image` allows automation in form of consecutive execution of the selected scripts in the "Standard Scripts" Panel. To launch all "Standard Scripts," one can double-click on the top bar of that "Standard Scripts" panel, and press the "Play" Button. The progress bar then depicts, which script is running momentarily and the "Logfile" panel displays the text output of the running script. The last "Standard Script" is the "Generate Map" script. The result can be examined by double-clicking the "Non-symmetrized Map" in the "Images" panel. This will display 2×2 unit cells.

10. The automatic processing will only succeed if all the processing parameters are chosen carefully. One can then improve the processing result by manually refining the parameters of each step.

11. First step to improve the final map would be to look at the determined defocus and astigmatism for the CTF correction. This can be examined by looking at the power spectrum of the image. Therefore, one needs to double click on "FFT of Downsampled Image" in the "Images" panel of the "Calculate FFT" section. This then opens up in the full screen browser. Pressing the C key or selecting "View CTF" in the "Navigator" menu displays the Thon rings as overlays that correspond to the current setting for defocus. The defocus and astigmatism can be altered in the dialog window with the page up/down and cursor up/down buttons. It can help to adjust the contrast and brightness in the "Navigator" menu to see the Thon rings (keys "b" and "n"). The values for defocus and astigmatism are then used in the "Correct CTF" script.

12. The next step would be to examine the lattice that was found. Again this can be done by looking at the power spectrum in the full screen browser. The determined lattice is displayed by pressing the L key. Helpful is the "Display Coordinate Info" (key "i") dialog window that displays amplitude, phase, Miller index, etc. of the pixel that the mouse pointer hovers over.

13. The lattice can be refined manually in "Lattice Refinement Mode" through the "Lattice Refinement" in the "Navigator" menu. This enables the user to add spots to the lattice to refine it. These would ideally be high resolution spots. `2dx` helps the user with a smart-mouse feature: A double click in the surroundings of a diffraction spot will select the pixel with the maximum intensity, or it will fit a Gaussian blob to the pixels in the vicinity of the double-clicked position, and take the center of that Gaussian as peak coordinate. The behavior of this "smart mouse" function can be adjusted in the menu "Display Parameters" reached by "d" as shown in Fig. 8. In the "Lattice Refinement"
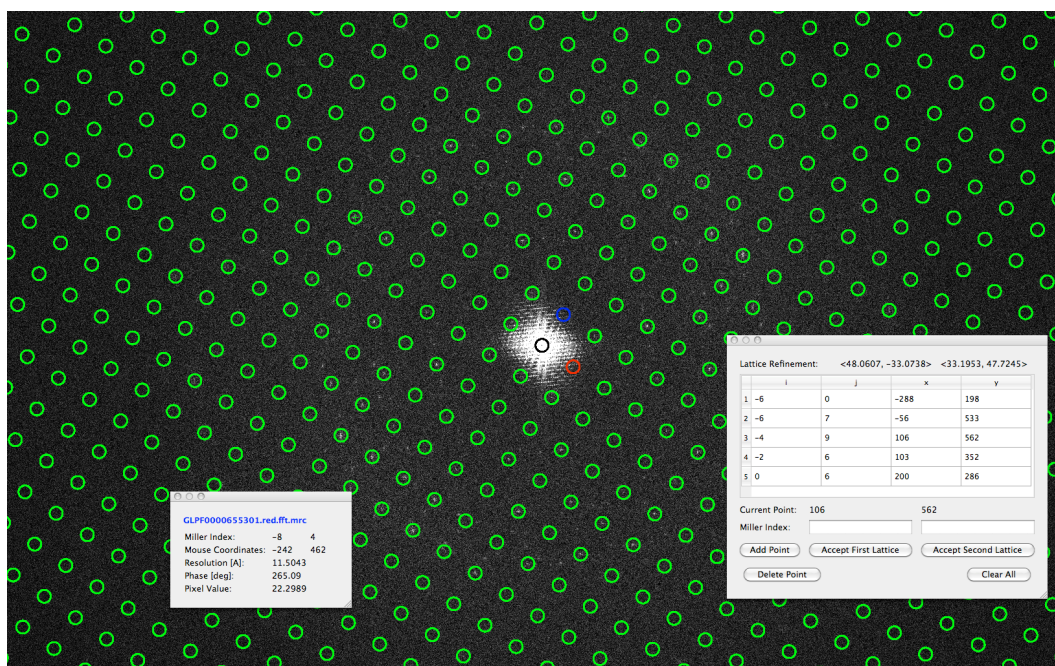
Figure 8: The "Lattice Refinement Mode" allows manually refining the lattice by picking spots on the power spectrum of the image.

dialog window an identified spot is added to the lattice refinement table by pressing the Enter key or the "Add Point" button. The Miller index is proposed according to the existing lattice, but can be altered in the dialog window. The "Accept First Lattice" button will apply the lattice refinements according to the picked spots and display the refined lattice directly in the full screen viewer.

14. In many cases, when vesicular 2D crystals were formed and adsorbed to carbon-coated grids, the two surfaces of the vesicles will give two overlapping 2D crystals in the image. The Fourier transform will then always show two lattices. If there are two similarly well ordered 2D crystal lattices in the image, then both lattices should be indexed, and stored as "First Lattice" and "Second Lattice." Process the image for one lattice first, then copy the image directory to another sub-directory under a different name (e.g., a different sub-frame number), and process it for the other "Second Lattice." In each processing, use the alternative lattice to filter out conflicting overlapping spots from the spotlist, using the script "Refine Spotlist."

15. `2dx` uses a minimum of two unbending rounds as depicted in the "Standard Scripts" panel by the "Unbend I" and "Unbend II" scripts. In the unbending

Status							
Th.Mag=	67712.3		**QVAL=**			218.2	
IQs Unb.I				QVal1=		143.3	
13	21	12 25 44	48	49	286	544	
IQs Unb.II				QVal2=		230.9	
19	30	22 31 50	50	50	264	526	
IQs Unb.Syn				QValS=		218.2	
19	20	13 35 43	50	48	281	533	
PowerBins	[A]	15 12	9	7	5	3	
(noise=7)	#	317 54	56	10	8	7	

	Defoc.	Latt.	SpSplit	Merge
TLTAXIS	-5.7	0.0	-	-
TLTANG	43.5	44.2	-	-
TLTAXA	-74.7	-68.9	-	-
TAXA	-69.3	-61.7	-	-69.3
TANGL	43.5	44.2	-	43.5

Figure 9: The Status panel of the GUI of `2dx_ image`, here for an image of a 45-degree tilted crystal.

scripts one can change the mask radius (holea, holeb) for generating the reference as well as its diameter in real space (boxa, boxb). The Fourier filtering mask of the original image is defined by the radius maska and maskb. The suffix depicts for which round the parameters are used. The optimal settings for these box and mask values can be found automatically with the custom script "Refine Parameters Unbend I" and "Refine Parameters Unbend II." As described above, the validity of a found set of unbending parameters can be verified by the coherence of the unbending distortion vector plot: It should look like a "well-brushed fur," and not like a chaotic pile of jumping lines.

16. The automatic refinement of box size and mask size bears a trap: If you use an unreasonably small reference area (e.g., a box size of 3 pixels), an extremely large mask that effectively doesn't do any Fourier filtering, and allow the QUADSERCH search to find any peaks in the cross-correlation map with great flexibility (very large quadrada value of for example 20 pixels, and very low number of lattice nodes for predicting the next lattice node, i.e., quadpreda of only three nodes), then QUADSERCH will produce a lattice distortion vector plot that only follows random noise peaks. The unbending routine then will produce an unbent image that shows in the Fourier transform strong diffraction spots out to the highest possible Nyquist frequency, and even your processing statistics look too good to be true, with IQ=1 spots out to Nyquist. You could get similar results when unbending an image that was already unbent. In such a case, the resulting structure doesn't bear any valuable content, and cannot be merged with results from other images. A situation like this can be easiest recognized by inspecting the unbending distortion vector plot (Does it look like a "fur"?).

17. The progress of the image processing can be monitored in the Status panel in the bottom right corner of the GUI of `2dx_ image`. This panel shows a collection of values from the processing, as shown in Fig. 9.

18. The top line of the Status panel shows the Theoretical Magnification, which is calculated from the given pixel size, the indexed reciprocal lattice, and the real-space lattice dimensions. That calculated Th.Mag. should correspond to the given Magnification under which the image was recorded. If the Th.Mag. deviates from the nominal magnification, then most likely the lattice indexing is wrong.

19. The Status panel also shows the IQ value statistics for each unbending step: The Unbend I script produces APH files with reflection values, in which each reflection has an IQ value. The number of IQ=1, IQ=2, IQ=3,..., IQ=9 spots is listed for each Unbending step Unbend I, Unbend II, and Unbend Synthetic in that Status panel. For each unbending script this panel also shows the resulting QVal (QVal1, QVal2, QValS), which summarizes the image quality and the unbending performance into one single number. The last unbending step defines the final QVal, which is shown in the top right corner of this panel. In this example, QValS is lower than QVal2, indicating that the better result was obtained from the conventional Fourier-filter based unbending (Unbend II), rather than using the synthetically generated reference (Synthetic Unbend). This could indicate problems in the available 3D merged data-set, or in the tilt geometry assignment for this image.

20. The Status panel also displays the summed intensity of the diffraction peaks in a certain resolution band, where the background intensity is normalized to the value "7." This is displayed in the lines "PowerBins." The example in Fig. 9 shows strong signal out to 15, 12 and even 9 Åresolution. The range up to 7 Åresolution still has some signal (PowerBin is 10), and this image has almost no signal in the range from 7 Åto 5 Å(PowerBin is 8), and no signal at all beyond 5 Å(PowerBin is 7).

21. The Status panel finally shows the determined tilt geometry parameters (TL-TAXIS, TLTANG, TLTAXA, TAXA, and TANGL). These are shown in different columns as determined from the defocus gradient (first column), from the lattice distortions (second column), from the spot-splitting of higher-resolution spots (SpSplit, not calculated in this example since TTF correction wasn't applied yet), and the tilt geometry can also be refined from the merged dataset (last column Merge). The determined tilt geometries should agree with each other, which is the case in the example shown in Fig. 9.

22. The GUI of `2dx_image` shows on the bottom of the window a list of stamps,

Figure 10: The Progress Stamps at the bottom of the GUI of `2dx_image`, here for a partly processed image.

as shown in Fig. 10. The scripts that have already been executed appear in solid dark, while the steps that still need to be done are pale. These stamps indicate to the user, which steps may need to be repeated: In the above example, the crystal area in the image was not yet masked (MASK is grey). If that were done, then the stamps FFT, SPOTS, UNBEND, CTF would become grey, because those steps would have to be repeated. The steps DEFOCUS determination and LATTICE determination would not have to be repeated and therefore would remain solid.

`2dx_image` contains additional functions (e.g. found in custom scripts), and also contains a tool for single particle processing of the unit cells of a 2D crystal image, which is done with a Maximum Likelihood method. This is described in (38), and may significantly improve the resolution for badly ordered 2D crystals.

2.5. Conclusion

This chapter has focused on the image processing steps to get from a micrograph of a 2D crystal to the projection map. The key concept of extracting the periodicity of the structure by Fourier analysis has been shown. An introduction to the Contrast Transfer Function has been given and methods to correct for it have been presented. `2dx` aims at making image processing for 2D crystals user-friendly, while still offering to the expert user the possibility to adjust the algorithms through own scripts, and to fine-tune every detail of the processing. Although the processing is simplified through the software it still needs an experienced and trained operator to get the best results.

The next step in processing would be to merge the results from all the individual processed images together. This is described in the next chapter.

3. Merging of Image Data in Electron Crystallography

This section was published as a book chapter in **Electron Crystallography of Soluble and Membrane Proteins: Methods and Protocols** [Arheit *et al.* 2013a].

Abstract

Electron crystallography of membrane proteins uses cryo-transmission electron microscopy to record images and diffraction patterns of frozen-hydrated 2D crystals. Each two-dimensional (2D) crystal is only imaged once, at one specific tilt angle, and the recorded images are then processed with dedicated image processing software packages. Processed image data from non-tilted and tilted 2D crystals can then be merged into a 3D reconstruction of the membrane protein structure. We here describe the process of the 3D merging, using the 2dx software system.

3.1. Introduction

In electron crystallography of 2D crystals of membrane proteins, frozen-hydrated 2D crystals are either directly imaged with the transmission electron microscope, or they are subjected to electron diffraction. Due to the beam-sensitive nature of the samples, each crystal is imaged only once, at a given tilt angle, at a typical dose in the range of 10 electrons per square Ångström, which is submitted to the sample in one single exposure. Images and diffraction patterns are then processed individually. Processed images result in a set of Fourier amplitudes and phases from each image, and evaluated diffraction patterns result in a set of amplitudes only. These measurements in Fourier space are accompanied with an approximate knowledge of the tilt geometry under which that 2D crystal was positioned in the microscope during data collection.

Here, we discuss the merging of evaluated amplitude and phase data from images. We first discuss how data from images of non- tilted specimens are merged into a 2D average projection map. We then discuss how image data from tilted specimens are added to that 2D map, to form a 3D reconstruction.

The merging or inclusion of data from electron diffraction is not discussed here.

3.2. The General Concept

The general goal of electron crystallography data processing is to gain a computer simulated 3D model by integrating sets of 2D data recorded by the electron micro-

scope.

The structure reconstruction typically utilizes Fourier space and exploits the Central Section Theorem for the 3D reconstruction of the protein structures. The Central Section Theorem states that a 2D Fourier transform of a projection image along a certain direction is equivalent to the central section in the reciprocal space of the three-dimensional Fourier transform in the plane perpendicular to the projection direction. This implies that a full sampling of the Fourier transform of the specimen can be gained through the projections along all directions. Through the reversibility of the Fourier transform we can then reconstruct the imaged object.

Application of this theorem to electron crystallography exploits the fact that the images created by a transmission electron microscope correspond (approximately) to projections of the sample. This means that with sufficient projections of samples imaged in different orientation to the electron beam, one can create an approximation to its three-dimensional density distribution. (Effects of nonparallel illumination, multiple electron scattering in the sample, and the curvature of the Ewald sphere are ignored here).

The Fourier transform is a suitable tool for processing of 2D crystal data, since it exploits its repeating structure. The repetitive arrangement of membrane proteins in a 2D crystal results in projection images that can well be studied with a 2D Fourier transform. However, that periodicity only holds true in the X/Y plane of the crystal. In the vertical (Z) direction the crystal is only formed by one layer, so that there are no repetitions in this direction. The Fourier components in the vertical direction are therefore not restricted to discrete diffraction points, but are continuously distributed. The 3D Fourier transform of a three-dimensional 2D crystal structure can therefore be seen as a set of "lattice lines". (h, k, z^*), where the indices h and k identify a discrete lattice in the Fourier space, while the z^* coordinate defines the (continuously sampled) position along that lattice line. The Central Section Theorem now ensures that all the information gained from projections of non-tilted crystals will contribute to the $(h, k, 0)$ Fourier components. To fill the central section $z^* = 0$ is the task of 2D merging, which is described in more depth in subsection 3.3.

For projections of tilted crystals, each diffraction spot in Fourier space corresponds to the value of the lattice line at height z^*. The height depicts where the tilted plane intersects with the lattice line, hence the z^* value can be determined from the tilt angle under which the crystal was oriented during data collection. Each evaluated 2D crystal image produces a set of (h, k, z^*) data, which are used to populate 3D Fourier space. The more tilted 2D crystal data are processed, the denser the lattice lines are sampled. The incorporation of the data gained from tilted 2D crystals is discussed in subsection 3.4.

The 2dx software package (1, 2) is available at http://2dx.org. The software package 2dx includes programs for different purposes, such as the program 2dx_image for processing individual images as described in the previous chapter. 2dx also contains

the program `2dx_merge`, which fulfills two functions: it facilitates the management of a 2D crystal project, and it enables 2D and 3D merging of processed crystal data. The texttt2dx package is based on the MRC image processing software (3-8). Most underlying Fortran programs were developed with and for the excellent MRC software. While the original MRC programs can be obtained from Richard Henderson and Jude Smith at the MRC, Cambridge, UK, the `2dx` package includes a slightly adapted version of those MRC programs.

The project management with `2dx_merge` is described in section 4 of this book. In short, you should start by setting up a suitable directory structure with `2dx_merge` (e.g., by following the default structure created by `2dx_merge`). Import one image into `2dx_merge`, process it manually with `2dx_image`, save those parameters as defaults for future image processing tasks, import the remaining images from that tilt angle category, and automatically process these newly imported images. If images of tilted specimens (here called "tilted images") are available, the same should be done with those images, until all available images have been processed with `2dx_image`.

After successful processing of all images with `2dx_image`, each image directory contains so-called "APH files". These are text files that list for each Miller index h and k the evaluated values for Amplitudes and PHases. These files also list the background amplitude around each h, k reflection, and may contain information about the contrast transfer function (CTF) value that was valid for that h, j reflection. These APH-files then have to be merged into a 2D or 3D merged dataset.

3.2.1. Comparison with Single Particle Image Processing

The merging of electron crystallography data is similar to the merging in single particle electron microscopy: In classical single particle electron microscopy, particle images that show the particles from the same viewing direction and that are all rotated into the same orientation are compared with a reference image by cross-correlation. The x, y peak in the cross-correlation map then corresponds to the x, y shift, under which a certain particle image would best agree with the current reference. Particle images are then shifted to those locations and are averaged by adding those aligned images up and dividing the result by the number of images.

3.2.2. Averaging in Electron Crystallography

In electron crystallography, a similar process is implemented to merge different APH files. Data from those APH files in Fourier space are compared with a reference dataset. Instead of shifting images by a certain number of pixels, however, the phase values in the APH files are modified. A phase shift of $360°$ applied to the reflection with h, k indices $(1, 0)$ thereby would correspond to shifting an image by

one full unit cell in h direction. A phase shift of 360° applied to the $(1, 0)$ reflection would correspond to a displacement by one unit cell length in k direction. Higher-order reflections require more phase shift, to result in the same amount of image shift: To shift a 10 nm unit cell image by 1 nm in h direction would require 36° phase shift applied to the reflection $(1, 0)$, a 72° phase shift applied to the reflection $(2, 0)$, a 108° phase shift added to reflection $(3, 0)$, etc. Arbitrary phase shifts can be applied to the phase variables in the APH files, which bypasses any limitation from discrete pixel steps. Aligned APH files are then averaged. In contrast to single particle microscopy, however, here we have a signal to noise ratio (SNR, encoded by so-called IQ values (4)) measurement for each individual h, k reflection from each image dataset. That IQ value has been calculated by the MRC programs under `2dx_image` from the ratio of the reflection peak amplitude and the amplitude of the background signal around each peak. When averaging APH reflection data from different images, Richard Henderson has devised a method, whereby each h, k reflection is contributing to the calculated average with its IQ-related weight (7).

When a low-resolution particle image is added to a good average in single particle EM, the average is usually made worse. In contrast, the reflection-based weighting in electron crystallography means that adding data from a mostly bad image does not necessarily make the average worse. Instead, that bad image may still have at least a few good reflection values, which then have a high SNR (low IQ value) and benefit the calculated average positively, while the bad reflections from that dataset are down-weighted by their low signal to noise ratios (high IQ values), so that they degrade the average only with minor contributions.

Alignment of data from different image datasets to a current reference is done in iterations, whereby all datasets from non-tilted images are aligned to the reference, and the aligned datasets are then averaged with the above described weighting. The resulting average can be refined through further rounds of alignment against the most recent average structure.

3.2.3. Extending 2D into 3D

The Fourier space representation of a 2D crystal is a reciprocal space, where the protein structure is encoded in Amplitude and Phase values on Fourier pixels along vertically oriented so-called lattice lines. Due to the real-space limits on the 2D crystal (above and below the crystal is emptiness), each lattice line is continuous in the vertical or z^*-direction, as its values are convoluted with a profile that is the Fourier transformation of the vertical structure mask function (e.g., a block function in real space, corresponding to a sinc function in Fourier space). Fourier pixels along one lattice line therefore have neighborhood correlation. In contrast, Fourier pixels on different lattice lines in the general case have no neighborhood correlation. This is also described in (9), and can be exploited to reconstruct the data in the so-called

missing cone (10-12).

To create a 3D reconstruction of the 2D crystal protein structure, the available APH files from the images of tilted samples have to be aligned with each other, and with the non-tilted dataset. Once all tilted and non-tilted APH files are correctly aligned, continuous lattice line functions are fitted through the amount of measured values. Those functions (e.g., sinc functions) are then equidistantly interpolated, to create a Cartesian (h, k, l) 3D Fourier space representation for our protein structure.

This Fourier space dataset can then be used to calculate a real-space 3D map for our protein. All this is implemented in the scripts in `2dx_merge`. We will now describe the different steps to perform this 2D and 3D merging in `2dx`.

3.3. Merging in 2D

To perform 2D merging, switch the parameter "Modus of Merging" in `2dx_merge` to "2D."

3.3.1. Creating a First Reference

1. The first step is to create a reference, to which the non-tilted images can be aligned. Choose the best non-tilted image in the graphical user interface (GUI) of `2dx_merge`. You can usually recognize this by it having the highest QVal value or the lowest symmetrization phase residual value, as listed in the columns for QVal2 or SymPhaRes in the top central panel of `2dx_merge`.

2. Double-click this image, to open it in `2dx_image`, and determine its correct phase origin for the determined symmetry. This can be done in `2dx_image` either with the Custom Script "Get Spacegroup & PhaseOrigin," or manually with the Custom Script "Set PhaseOrigin Manually."

3. After having correctly centered this image, save the database in `2dx_image`, quit `2dx_image`, and return to `2dx_merge`. In the top central panel of `2dx_merge`, select only this one image, by checking the square checkbox in front of that image, and uncheck all other images.

4. Now in `2dx_merge` use the Standard Script "Merge Once," to merge the available data from this single image into a reference file. For this, you should specify the resolution range, up to which you believe the data in this image can be reliably used for the alignment (parameter "Resolution of merged dataset for reference"). This script will run several MRC programs to create an amplitude and phase file called merge.aph. The script also eventually creates a MTZ format file with the amplitude and phase values for this image, after

phase origin correction, CTF correction, symmetrization, possibly removal of symmetry-forbidden reflections, and correction of symmetry-restricted phase values. This final reference MTZ file is listed as a red entry with the nickname "MTZ: Merged full reciproc. space 2D data" in the "Images" panel bottom right in the GUI of `2dx_merge`.

5. If you double-click on that MTZ file, its binary contents will be translated into a temporary HTML text file, which will be displayed in a local web browser. There you can see that the MTZ file contains not only values for amplitudes, phases and FOMs for each reflection, but that it also contains information about the crystal unit cell dimensions, the symmetry, the resolution range, and more.

3.3.2. Aligning the Non-tilted Image Data onto the Reference

1. After having created a first reference, uncheck that first image, and select the other processed non-tilted images. In a first round you may restrict this to only the reliably and solidly processed images.

2. In `2dx_merge`, use the Standard Script "Refine Once" to align the selected images onto the reference file merge.aph. You need to specify the phase origin search space, over which the alignment search should be performed. To perform a search over the entire crystal unit cells, you should scan an area of $360°$ phase range, which will be searched in the h and k direction by the MRC program origtilt. This search will be performed in steps of a certain increment size, which you need to specify under the parameter "Stepsize of the phase origin search." In a first round, you could specify a search in $6°$ step size. To cover the entire unit cells, you would need 60 steps of this $6°$ step size, so that the range from $[-180°, 180°)$ is covered. In later rounds, once the rough phase origins have been found for each image, you could perform local refinements of the phase origins by restricting the phase origin search range to a smaller area of for example a $30°$ range, by specifying 6' steps of $0.5°$ step size, giving a range of $\pm15°$ for the phase origin search.

3. The alignment of all image datasets against the specified reference is done by the MRC program ORIGTILT. This MRC program is a powerful tool that has grown over the years through historic modifications to a complex Fortran program. It requires to be called with a long parameter list, whereby for each image dataset to align, ORIGTILT requires several lines of input parameters. In `2dx_merge` we created a separate Fortran program called 2DX_MERGE_COMPILEB, which receives from `2dx_merge` the list of image directories to be included in the merging. This program then collects from all `2dx_image` configuration files

(2dx_image.cfg) in those directories the required information, and creates as output a C-Shell script file, which will later be launched by `2dx_merge` to call the actual ORIGTILT Fortran program. This is an awkward, but functional way to utilize the mathematical implementations in origtilt from the graphical user interface of `2dx_merge`.

4. Launching the script "Refine Once" in `2dx_merge` will therefore create the refinement script, run it, thereby run ORIGTILT, and evaluate the output of ORIGTILT. The refinement script can be inspected in the "Images" panel of the `2dx_merge` GUI, where it is listed under "CSH: refinement script." The output of ORIGTILT is listed under the nick-name "LOG: origtilt B output." You should inspect that file carefully: In this file, ORIGTILT will document for each aligned image what it did. It will specify the number of h, k values that were used to find the phase origin for a certain image. It will also list for the best identified phase origin position the summed-up phase residual values for the h, k reflections in a certain resolution range, which allows verification of up to which resolution that determined phase origin value is trustworthy for a certain image. This file will also list a larger densely populated field of numbers, whereby each digit corresponds to the normalized cross-correlation value between the reference and the image dataset aligned to this position. This field corresponds to a cross-correlation map, and it should indicate a clear peak in the position identified. In later rounds, this peak should be centered in the middle of the map, indicating that specific image dataset was already correctly aligned.

5. In `2dx` we modified the original MRC program origtilt slightly, so that it writes the determined optimal phase origin values for the images into a separate file, which is later used to update the configuration files 2dx_image.cfg in each image directory.

6. After successful completion of the Standard Script "Refine Once," you should inspect the column for "PhaOri Change" in the top central panel of `2dx_merge`, to see by how many degrees the phase origins of each image were changed.

7. Now include the first image in the list of selected image directories, and merge all selected images into a new reference data-set merge.aph, using the "Merge Once" script. If the symmetry is not P1 or P2, you will see in the logfile a table entitled "Phase Residuals in Resolution Ranges," which indicates, up to which resolution the merged data are trust-worthy. Limit the resolution of the reference file with the parameter "Resolution of the merged dataset for reference" to that value.

8. Now refine the phase origin alignment of the selected images, by again using the

script "Refine Once," this time with a smaller phase origin search range. You could, for example, perform a search over 6' steps of 0.5° step size (parameters "Number of steps in phase origin search" and "Stepsize of the phase origin search"), to search over a range of $\pm 15°$. Run this script, and inspect the output of ORIGTILT in the file "LOG: origtilt B output." You should see clear cross-correlation peaks in the correlation maps for each image.

9. You can use the following script "Merge & Refine (Iterative)" in 2dx_merge to perform three more rounds of merging and refinement. The iteration counter after that script name can be changed from 3 to higher numbers, by double-clicking it.

10. During those iterations, inspect the column "PhaOri Change" in the GUI of 2dx_merge. The indicated values for the phase origin change should be converging towards zero. The image phase origin values under the column "PhaRes" should also converge towards smaller values. If the Phase Origin value for a certain image remains high (90° = random), you could double-click that line in 2dx_merge, to open that image in 2dx_image, to manually see, where the phase origin was placed, and why this image was not correctly aligned.

11. If you activate the parameter "List Reflections into logfile (ILIST)," a separate logfile "LOG: reflections after origtiltk" will be created during the merging step. This file lists for every h, k index the available values from the contributing image datasets. For each h, k reflection, the amplitude, phase, micrograph number, IQ, film weight, background amplitude, and CTF value are listed. A dashed line appends these values, its length corresponds to the reflection's IQ value (lower IQ value = longer line). For reflections with IQ values 1-4, the values for amplitude and phase are repeated at the end of the lines. Perusal of this file allows you to efficiently recognize suspicious reflections, where despite strong IQ values the measurement for amplitude or phase deviates from the values for the same reflection from other images. In such cases, you might have identified a problem in the defocus assignment for that specific image, so that the Thon rings of the CTF estimation were not correctly placed, and you could manually correct this with 2dx_image.

3.3.3. Creating Maps

1. After finishing the alignment phase, you can re-create the projection maps for all aligned images with the script "Generate Image Maps" in 2dx_merge. Only after running this script, the maps for all images correspond to the new phase-shift corrected APH files. These maps can be displayed in the GUI of
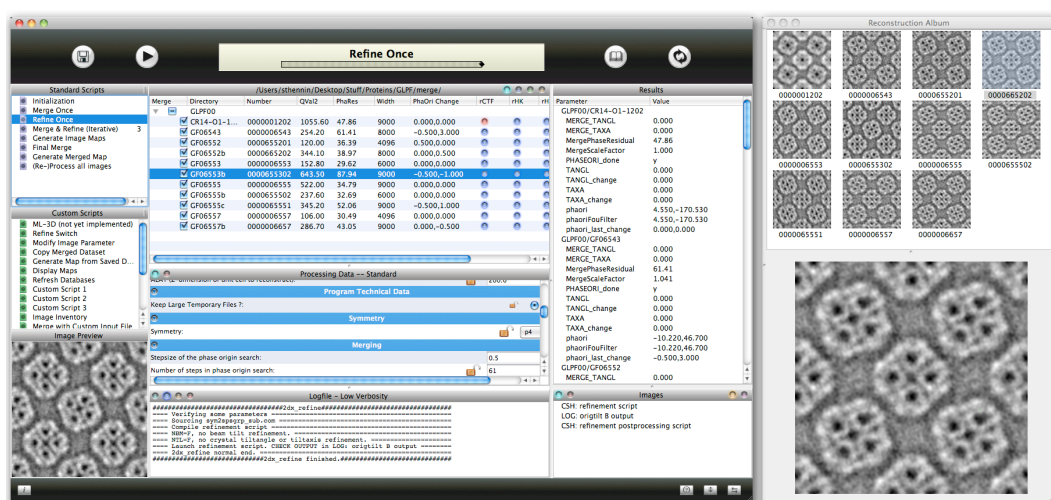
Figure 11: The GUI of `2dx_merge` (left), and the Reconstruction Album (right). In this example, image number 665202 is highlighted in the Reconstruction Album (right), it is correctly centered, and displayed as 2×2 unit cells. The alignment has not yet converged: PhaOri Change values and phase origins (PhaRes) are too high. Image number 1202 (the first in the list) has a phase residual of $47.86°$, but the Reconstruction Album shows that its handedness is wrong with respect to the other images. Image number 655302 (highlighted in blue in the GUI of `2dx_merge`) has an unreasonably high phase residual of $87.94°$, and the thumbnail preview in the GUI of `2dx_merge`, as well as the Reconstruction Album shows that its phase origin is misplaced by $(+180°, +180°)$.

`2dx_merge` by "walking" with the cursor line over all image directories in the top central panel. This is facilitated by activating the menu pull-down option "Show Only Selected Directories" (Fig. 11). Misaligned phase origins can often be recognized here, and manually corrected.

2. After having manually corrected or excluded all miss-aligned images, you should now merge the aligned images with a reasonable resolution limitation (script "Final Merge"), and create a merged projection map (script "Generate Merged Map"). This map can then be used to create synthetic references to reprocess all images, or it can be used to refine the defocus and astigmatism for all images. This is described in section 4.

3.4. Merging in 3D

To perform 3D merging, switch the parameter "Modus of Merging" in `2dx_merge` to "3D."

3.4.1. Extending from 2D into 3D

1. Extending the current merged dataset from 2D into 3D is not trivial. So far, our calculated 2D average corresponds only to the horizontal plane in h, k direction in a 3D Fourier space, where all values have a coordinate of zero in the vertical direction (l or z^*). Our 3D Fourier space is mostly empty or undefined, with the exception of this infinitely thin horizontal slab at $z^\circ = 0$. This Fourier space needs to be filled with continuous reciprocal values for amplitudes and phases along the above introduced lattice lines.

2. As before, we need to align the APH files from the processed images to the current average. Also the APH file from an image of a tilted sample contains amplitude and phase values for each h, k index. These measurements, however, belong onto a tilted plane in 3D Fourier space, where the tilt geometry variables TAXA and TANGL define the orientation of that tilted plane. With the exception of the few measurements on the tilt axis, most of the reflections from a tilted image have a z^* value that is not zero.

3. If we now would try to align by phase residual optimization the APH file from a tilted image with our currently available merged dataset, then the tilted plane of that APH file would intersect with our available reference (the horizontal plane) in only very few data points, that lie on the tilt axis. Those few measurements would by far not be sufficient to define a correct alignment of that image dataset. Especially in the direction perpendicular to the tilt axis
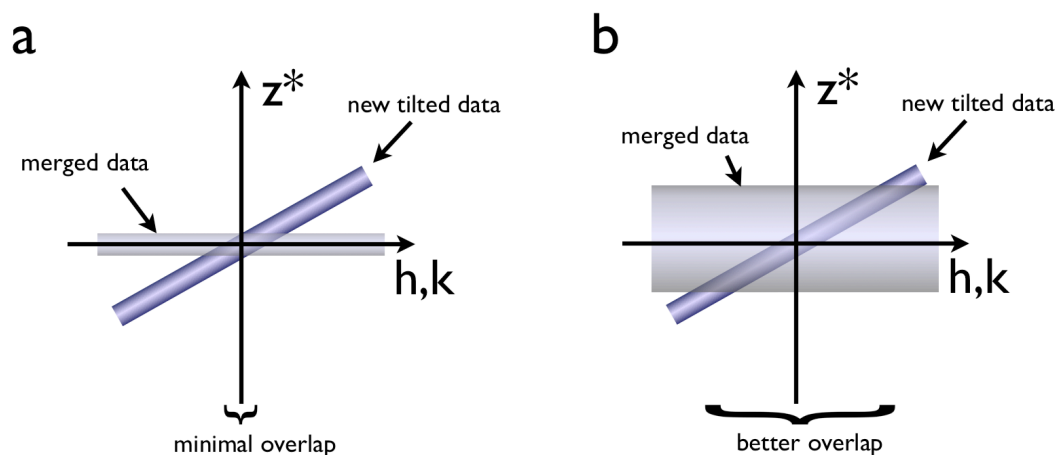
Figure 12: Phase origin determination for an APH file from a tilted crystal to a reference dataset. (a) If only data from non-tilted images are available as a reference ("merged data" slab), then only very few reflections from the "new tilted data" find reference partners in the merged data, resulting in "minimal overlap." The vertical tolerance zone for the merged data is defined by the variable zstarwin. (b) If the merged data has already an extension into z^* from successful inclusion of other tilted images, then the phase origin determination for another new tilted dataset is easier.

the alignment would most likely fail, see Fig. 12.

4. In order to successfully align the first tilted image dataset to the (non-tilted) average, we need to allow the programs to treat the z^* dimension with greater flexibility, so that a wider region of overlap between the tilted image and the non-tilted ($z^* = 0$ plane) reference can be created. In 2dx_merge, the parameter zstarwin defines the vertical z* tolerance along a lattice line in the search for reference points for the phase origin determination. To start a 3D reconstruction, set this parameter to 0.2 or larger.

5. Once a 3D reconstruction is available, you should set this parameter back to the reciprocal of twice the vertical lattice size, or $\frac{1}{2 \cdot ALAT}$. If for example you reconstruct a unit cell of 200Å vertical dimension, then you should eventually set this value to 0.0025.

3.4.2. Aligning Data from Tilted Crystals to the Merged Dataset

1. Start the 3D alignment with backing up your project to a safe location, using the custom script "Synchronize Project with Backup." In case your align-

ment parameters are lost during the following processing steps, you can open `2dx_merge` in your backup location and use that same script to write the saved configuration back to the original data.

2. Also save the current merging result (MTZ and APH files) to one of the ten storage registers in `2dx_merge`, using the Custom Script "Copy Merged Dataset."

3. To align tilted APH files to the merged dataset, set `2dx_merge` to 3D Modus of Merging.

4. Select the script "Refine Once," and set zstarwin to 0.5 to increase the z^* tolerance during the phase origin search.

5. We already have a non-tilted reference. Select in the GUI of `2dx_merge` only a few of the best images of very low tilt angle, and unselect all other images, as well as all non-tilted images.

6. Set the step size of the phase origin search back to 6.0° and define the search to run over 60 steps.

7. Limit the resolution of the alignment search to a reliable conservative range.

8. Then run the script "Refine Once."

9. As before, inspect the logfile of the ORIGTILT run "LOG: origtilt B output."

10. Re-create the final maps for all images with the script "Generate Image Maps," and visually verify correct alignment.

11. Uncheck the wrongly aligned tilted images, and include the correctly aligned non-tilted images (use the Save Selection and Load Selection pull-down menu dialogues for this).

12. Merge the non-tilted and tilted datasets into a 3D dataset with the script "Merge Once."

13. Then refine the alignment of the tilted images, by selecting the script "Refine Once" with smaller phase origin search step size, while reducing the z^* tolerance parameter zstarwin (e.g., to 0.1).

14. Repeat the merge and refine cycles with finer step sizes a few more times.

15. Then include successively the datasets from higher tilted crystals. Since you now already have a 3D volume as a reference, there is no further need to set zstarwin to a large value.

16. Use the script "Generate Image Maps" to re-create the aligned final maps for all images.

17. Inspect the Reconstruction Album with the non-tilted and tilted images, to identify wrongly aligned images. As before, these can either be manually corrected with `2dx_image`, or removed from the merging process by unchecking their selection boxes.

3.4.3. Creating a 3D Reconstruction

The phase-origin aligned amplitude and phase datasets from non-tilted and tilted crystals now need to be merged into a 3D reconstruction. This involves transfer of data through different stages. The data flow during this process is illustrated in files on the web server `http://2dx.org/documentation/2dx-software/manual/data-flow`, where the involved programs are listed together with their input and output data files, and the formats of these files is shown in examples and with explanations.

The program ORIGTILT produces as an output a file called merge.aph, which contains a combined listing of all (h, k) reflections from all images. This file lists in columns the (h, k, z^*) coordinates that ORIGTILT calculated for the reflection, followed by a measurement of the amplitude and the phase of this reflection. The phase values in this file are already phase-flipped (180° added) where needed, and the phase shifts due to phase origin correction have also been applied to these phase values. The file further lists the micrograph number that this reflection came from, and the IQ value, a reflection weight factor, the background signal, and CTF multiplication factor for each reflection. Even though the phases for the reflections have already been corrected, the amplitudes in this file are still not CTF corrected.

The program LATLINPRESCAL uses this file merge.aph to correct the amplitude values. It scales reflections that originate from the same micrograph to equivalent average intensities, and applies the amplitude CTF correction to the amplitude values. The output is a file called latlines.dat. So far, the measured h, k reflections from non-tilted and tilted datasets were assigned to a z^* coordinate, which specifies its vertical height in 3D Fourier space along the h, k lattice line. These reflections contribute measurements along one lattice line; however, these measurements are not equidistantly but randomly spaced in z^* (5). The program LATLINE is now fitting sinc functions to the measured data in a least-square approximation. Equidistantly sampled amplitude and phase values along the fitted sinc functions are then written out into the file latfitteds.dat. Precise fine-tuning of parameters for LATLINE is an important part of successful 3D merging. The file "PS: latline.ps" should be inspected, and the lattice line carefully parameters fine-tuned (13). The produced lattice line data are then subjected to symmetrization by further programs, and finally transferred into the binary file format MTZ. The program PLTILT produces

a plot to verify the completeness of the 3D dataset (14). Inspection of its plot "PS: TLTPLOT file" allows to recognize at which tilt angle more data is needed to approach an even sampling of 3D Fourier space.

The final 3D reconstruction map is a volume in CCP4 format, which can be inspected with UCSF Chimera or another suitable display program (15).

3.4.4. Projective Constraint Optimization

An alternative to lattice line fitting is projective constraint optimization (PCO) (12). This method uses an iterative algorithm to alternately apply known constraints to the dataset between real space and Fourier space, while applying constraints in each space. Gipson also introduced a linear algebra solution to avoid interpolation onto the randomly sampled lattice line measurements in each iteration, by setting up large matrices that directly transform real-space representations of the lattice lines to the measured z^* coordinates of the experimentally determined randomly sampled reciprocal lattice line measurements. After a singular-value decomposition based inversion of these matrices, a method was developed that allows oscillating many times between real space and randomly-sampled Fourier space, with- out the need to interpolate sinc functions in each iteration.

Fine-tuning of the choice of constraints, their weighting factors, and the way the constraints are applied, allowed Gipson et al. to not only perform a 3D reconstruction of the protein without lattice line interpolation, but allowed for a reconstruction that showed a full coverage of the data in the so-called missing cone. The reconstruction contained correctly determined amplitude and phase values for high z^* values for low-resolution h, k lattice lines, including for the lattice line at $(h = 0, k = 0)$. These missing cone data had not been experimentally available from the electron microscopy images and diffraction patterns, but instead were computationally derived from the oversampled data collected at lower tilt angles. An important finding of that work by Gipson et al. is also that this algorithm is capable of eliminating the need for highly tilted image data by strongly oversampling at lower tilt angles: In the specific case of Bacteriorhodopsin (16) analyzed in that work, the data in the tilt angle range up to 45° sample tilt were sufficient to reconstruct the entire protein structure, so that the data collection in the tilt angle range from 45° to 60° tilt would not have been necessary.

As a consequence, focused efforts into the automation of electron crystallography data collection is now needed, so that very large numbers of high-resolution cryo-EM images and diffraction patterns of 2D crystals at lower tilt angles can be recorded automatically. These can then be used to computationally produce the data for higher tilt angles, respectively high z^* coordinates, to fill the missing cone region, and to produce complete 3D reconstructions of the proteins. Cryo-EM 2D crystal data can be collected with much greater success rate at lower sample tilt angles, so that au-

tomation of this process appears feasible. This will enable efficient high-throughput structure determination of membrane proteins by electron crystallography.

3.5. Conclusion

Electron crystallography of membrane proteins entails the collection of cryo-EM 2D crystal data in form of images and diffraction patterns. These are processed individually, and the processed amplitude and phase data from different crystals are merged into a 3D dataset. While the processing of individual images in `2dx` is automated, the merging process still requires manual control. The program `2dx_merge` assists in the management of an electron crystallography project, and provides user-guidance and user-friendly merging capabilities. Further automation of the data processing for merging is required.

3D reconstruction of membrane proteins from 2D crystal data can strongly benefit from application of projective constraint optimization, to use available oversampled data from 2D crystals imaged at lower tilt angles, to complement the missing data at higher tilt angles, so that eventually the entire so-called missing cone region can be filled. Automation of the cryo-EM data collection of 2D crystal images is therefore now needed to enable automated high-throughput structure determination of membrane proteins by electron crystallography.

4. Automation of Image Processing in Electron Crystallography

This first part of this section was published as a book chapter in **Electron Crystallography of Soluble and Membrane Proteins: Methods and Protocols** [Arheit *et al.* 2013b]. The Further Development section (subsection 4.6) is unpublished work.

Abstract

Electron crystallography of membrane proteins records images and diffraction patterns of frozen-hydrated two-dimensional (2D) crystals. To reconstruct the high-resolution three-dimensional (3D) structure of a membrane protein, a multitude of images of 2D crystals have to be processed. Certain processing steps are thereby similar for batches of images that were recorded under similar conditions. Here we describe how the 2dx software package can be used to automate the processing of 2D crystal images, and how the 2D and 3D merging results can be used to iteratively reprocess the images. While the processing of 2D crystal images has been fully automated, the merging process is still semi-manual.

4.1. Introduction

Electron crystallography of two-dimensional (2D) membrane protein crystals involves several steps. Sufficient quantities of pure, homogeneous, and stable membrane protein in a suitable detergent have to be produced, which then have to be two-dimensionally crystallized. Grid preparation for cryo-electron microscopy (cryo-EM) is a difficult task that requires persistence and experience. Data in form of images and diffraction patterns on the cryo-electron microscope has to be collected. Lastly, the data have to be processed and merged into a high-resolution 3D structure.

Significant progress has been made in all these steps, not only in terms of quality but also in speed. With the exception of sample preparation for cryo-EM, all steps have been streamlined, made more user-friendly, and are partly automated. Here, we discuss the automation of the processing of a larger number of 2D crystal images with the 2dx software package, version 3.3.0.

4.2. The General Concept

The `2dx` software package [Gipson *et al.* 2007a, Gipson *et al.* 2007b] is available at
`http://2dx.org`. It is based on the MRC programs for image processing [Crowther *et al.* 1996],
and includes additional programs for automation [Zeng *et al.* 2007a] and other func-
tions [Zeng *et al.* 2007b]. The `2dx` software suite contains a module `2dx_image`, which
allows the processing of one single image of a 2D crystal, and `2dx_logbrowser`, which
facilitates the analysis of the log files. `2dx` also includes the module `2dx_merge`, which
is used for the management of a 2D crystal structure determination project, and to
perform 2D or 3D merging of the data from different 2D crystal images.
For a new project, the user should first setup a suitable file structure via `2dx_merge`,
then process one image of a non-tilted 2D crystal with `2dx_image` manually, then save
the determined optimal processing parameters as project default values, followed by
launching the automatic processing for the remainder of the images of non-tilted
specimens via `2dx_merge`.
For including images of tilted specimens into the processing pipeline, the user should
again process the first image of a certain tilt angle category manually with `2dx_image`,
save the optimal processing parameters as default values for that tilt angle category,
and launch the automated processing of the remainder of the images in that category
with `2dx_merge`.
The 2D or 3D merging of the processed image data still has to be done manually,
but `2dx_merge` is designed to provide the user with guidance and feedback.
Automatic reprocessing of all non-tilted and tilted images can then again be launched
from `2dx_merge`, after which the reprocessed data again have to be merged manually.
These steps will now be explained in detail.
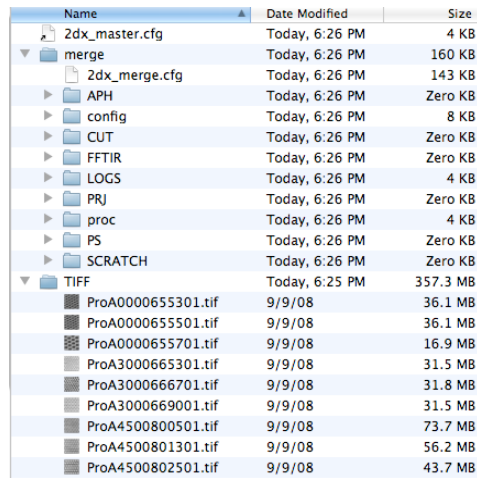
4.2.1. The Project File Structure

`2dx` is designed for a certain arrangement of the files on the hard drive. It is strongly
recommended to the users to adhere to this arrangement, so that the automated
processing finds the files and data where it expects them.
A new membrane protein project is to be initiated in a certain directory. We here call
that protein ProA, and assume it was initiated by user "John", who placed it into a
project base directory called /usr/people/john/proteins/ProA. For convenience we
abbreviate this directory with ~/proteins/ProA. If images were recorded on film and
digitized, or if images were recorded on a CMOS or CCD camera, the most suitable
image file format for the `2dx` software is TIFF, and image names should follow
a certain convention. By default, texttt2dx assumes a file naming in the form of
`ProA0012345601.tif`, where the first four letters are the protein name (here "ProA"),
the next two digits are the tilt angle (here "00"), the next six digits are the image
number (here "123456"), and the last two digits are a possible subframe number

Figure 13: The file structure created under the `2dx` project base directory upon
project initialization.

from that image (here "01"). Subframes can, for example, be used, if several images
were cropped from one large digitized film, or if one image should be processed
for different crystal layers separately. The image import tool of `2dx_merge` will
later extract this information (e.g., the nominal tilt angle) from the file name. If a
different file name convention is used, one can adjust the filter setting in `2dx_merge`
by defining a different regular expression (see below).

If `2dx` is launched for the first time with a base directory ~/proteins/ProA, `2dx_merge`
will initiate a file and directory structure under this base directory. An example is
shown in Fig. 13.

4.2.2. `2dx_merge` The Project Manager

1. `2dx_merge` will open its graphical user interface (GUI) with the Initialization
   screen (Fig. 14). This GUI is composed of several panels. Each panel has
   a certain function, indicated by the title on its top bar. The top left panel
   contains a list of the available "Standard Scripts," which should be executed
   sequentially for the merging of processed images. Below it is a panel for "Cus-
   tom Scripts," which may be executed if needed. The central top panel will
   list the processed or non-processed images, once they are available. Below it
   is a panel to define the "Processing Data," which by default displays only the
   "Standard" set of parameters. Additional "Advanced" parameters can be ac-
   cessed with the buttons in the top left corner of that panel. The other panels
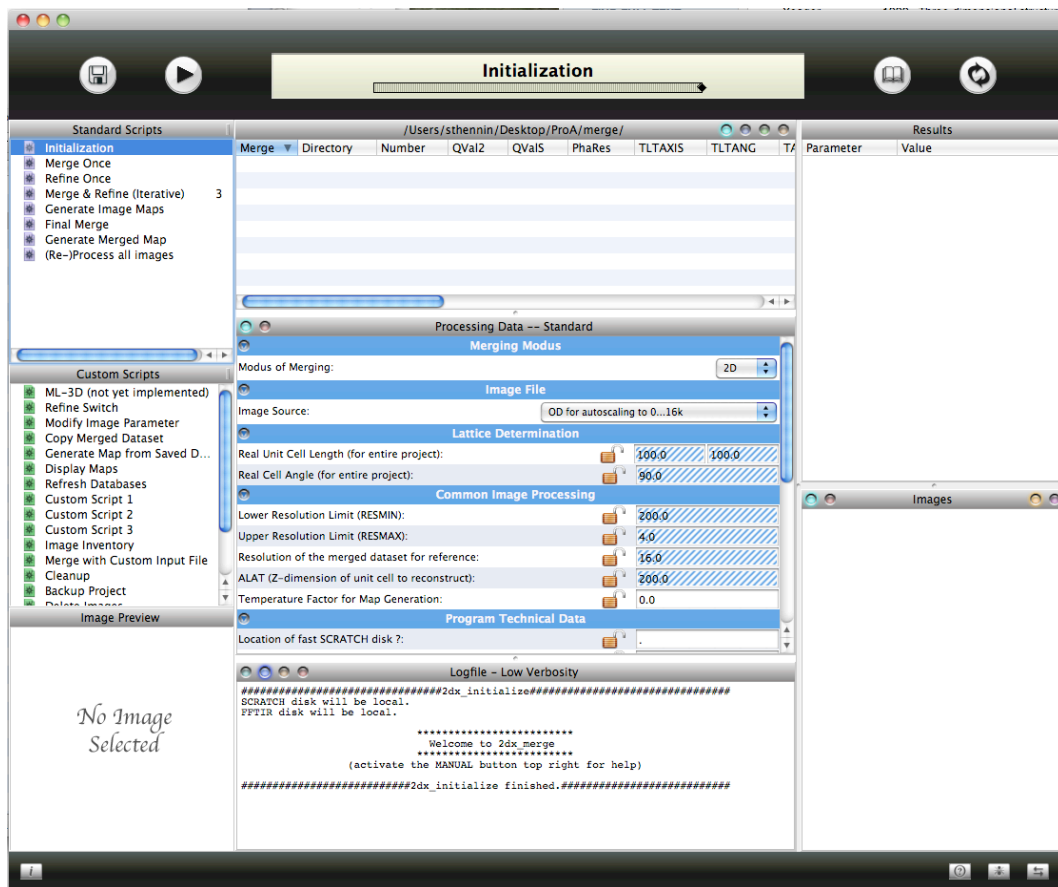   will be explained below.

Figure 14: The GUI of `2dx_merge` upon project initialization.

2. The user should now look through the "Processing Data," and see if any of
those image processing parameters can already be defined for this project. For
example, the user may have an estimate for the upper and lower boundaries
for the resolution range to be considered. For a typical negative stain project,
these would be set as RESMIN to 200.0Å, and RESMAX to 16.0Å. For a
beginning cryo-EM project, a RESMIN of 12.0Å may be adequate, depending
on the quality of crystal images.

3. The "SAVE" button on the top left in the GUI of `2dx_merge` allows to save
the current set of parameters in a file called 2dx_merge.cfg. This file is a
text file, which can be opened with a conventional text editor for inspection.
It contains the definition of various parameters, and the current settings for
those parameters. The entry for the upper resolution limit RESMAX, for
example, reads:

```
1  # LABEL:   Upper Resolution Limit (RESMAX)
2  # LEGEND:  The upper limitation in resolution, in Angstroem. This one
              should be high resolution (e.g. 4.0 A), while the other RESMIN should
              be the lower resolution cutoff (e.g. 200A).
3  # EXAMPLE:  RESMAX = "4.0"
4  # HELP:  http://2dx.org/documentation/2dx-software/parameters/Resolution
5  # TYPE:  Float "MIN=0.1;MAX=10000000.0;DEFAULT=16.0"
6  # LOCKED:  NO
7  # INHERITABLE_UPON_INIT:  YES
8  # SYNC_WITH_UPPER_LEVEL:  NO
9  # ISWRONG:  YES
10 set RESMAX = "4.0"
```

4. The meaning of the lines trailed by a # is as follows: LABEL defines the name
under which this parameter will appear in the GUI. LEGEND, EXAMPLE,
and HELP define the contents of the right-mouse-click help panel for that pa-
rameter in the GUI. TYPE defines the parameter type. LOCKED defines if
that parameter value can be changed manually or through the programs, or if
it is currently protected. INHERITABLE_UPON_INIT defines if this param-
eter setting should be inherited, when initializing a new image for processing,
see below. SYNC_WITH_UPPER_LEVEL defines those parameters that
should be synchronized throughout the project. The crystallographic space
group symmetry for a crystal project, for example, should be the same for
all images in this project: If a crystal has a fourfold symmetry P4, then all
images should be processed under this assumption. ISWRONG defines if the
current setting for this parameter is still possibly wrong. If so, it will show up
in the GUI with a blue-white striped underlay, to remind the user that this
parameter still might have to be modified. And finally, "set RESMAX" defines
the actual current value for this parameter. This configuration parameter file

2dx_merge.cfg is used by 2dx_merge to store a set of processing parameters within the ~/proteins/ProA/merge directory under the project base directory.

4.3. Workflow for Automated Image Processing

4.3.1. Importing a First Image into `2dx_merge`

1. Make sure the local database in `2dx_merge` is saved, by clicking the top left "Save" icon. Now use the pull-down menu "Import Images..." under the "File" menu, to select one TIFF image of a good non-tilted 2D crystal. Navigate into the TIFF folder, and select only one image file, which you open. `2dx_merge` will now display the "Import Images" window, displaying on the left side the filename of your TIFF file, and on the right side the interpretation of parameter strings for the protein "Name," the nominal tilt "Angle," the image "Number," and the "Sub-image" number (Fig. 3). If the TIFF files are named following the above described convention, then `2dx_merge`'s default regular expression "^(\w{3,4})(\d{2})(\d{6})(\d{2})?$" will correctly populate the fields for those values. Regular expressions are defined at http://www.regular-expressions.info and are a versatile way to define a parameter lookup scheme. The regular expression in this example will search for a word of 3 or 4 letters, then a number of 2 digits, then a number of 6 digits, and finally a number of 2 digits. Clicking on "OK" will cause `2dx_merge` to create a directory for the tilt angle (~/ProA/ProA00), create an image directory for the image processing of this image under this tilt angle directory (~/ProA/ProA00/ProA00006553101), copy the TIFF file into this directory, create an image processing parameter file 2dx_image.cfg in this directory (see below), and launch `2dx_image` with its first "Init Program" script for this image.

2. The here-created image processing configuration file 2dx_image.cfg results from a hierarchical processing of several configuration files that `2dx_merge` traverses sequentially:

    a) For this, `2dx_merge` starts with the system-wide default configuration file, which is usually stored in /usr/local/2dx/kernel/config/2dx_master.cfg.

    b) It then updates all parameters with settings from the personal choice of the users, usually stored in ~/.2dx/2dx_master.cfg.

    c) It then updates all parameters with settings from the project database, here in ~/proteins/ProA/2dx_master.cfg. (Note that by default this file is a symbolic link to the configuration file in the merge directory, here in

~/proteins/ProA/merge/2dx_merge.cfg, which then is the actual master configuration file for this project.)

  d) It then updates all parameters with the tilt-angle-specific databases, here in ~/proteins/ProA/ProA00/2dx_master.cfg, but for each parameter it only does this last update, if the parameter "INHERIT_UPON_INIT" in the last 2dx_master.cfg file was set to "YES." Note that by default this file is another symbolic link, pointing the 2dx_master.cfg file one level higher up (~/proteins/ProA/2dx_master.cfg), which by itself is a link to the 2dx_merge.cfg file in the merge directory. By default, all reference 2dx_master.cfg files are therefore symbolically linked to the 2dx_merge.cfg file in the merge directory. This assures that initially only one definition of the project-wide parameters is available.

3. `2dx_merge` then stores the resulting database in a configuration file for the newly imported image, here under
~/proteins/ProA/ProA00/ProA00006553101/2dx_image.cfg.

4. `2dx_merge` also creates several sub-directories in this image directory, which will be needed later for the processing of this image. This concludes the import of this image.

4.3.2. Processing the First Non-tilted Image with `2dx_image`

1. This first imported image of a non-tilted 2D crystal is processed with `2dx_image` by double-clicking in `2dx_merge` onto the line entry for this image in the central top panel of the GUI of `2dx_merge`. This will open `2dx_image` with this image. Proceed to process this image carefully, as described in section 2. Make sure you correctly define the acceleration voltage, Cs value, magnification, pixel step size, etc. Perform the processing, and fine-tune the unbending parameters. Determine the real lattice unit cell with the specific script "Evaluate Lattice," and save the newly determined real-space lattice as project-wide real-space lattice.

2. Once these parameters are all well known for this image, save the local configuration file (this will be stored in
~/proteins/ProA/ProA00/ProA00006553101/2dx_image.cfg).

3. This configuration should now also be saved as "Project Default" database, by using the pull-down menu option "Save as Project Default" under the "File" menu. This will overwrite the file ~/proteins/ProA/2dx_master.cfg, thereby overwriting the merge database ~/proteins/ProA/ merge/2dx_merge.cfg. You can now close this instance of `2dx_image`.
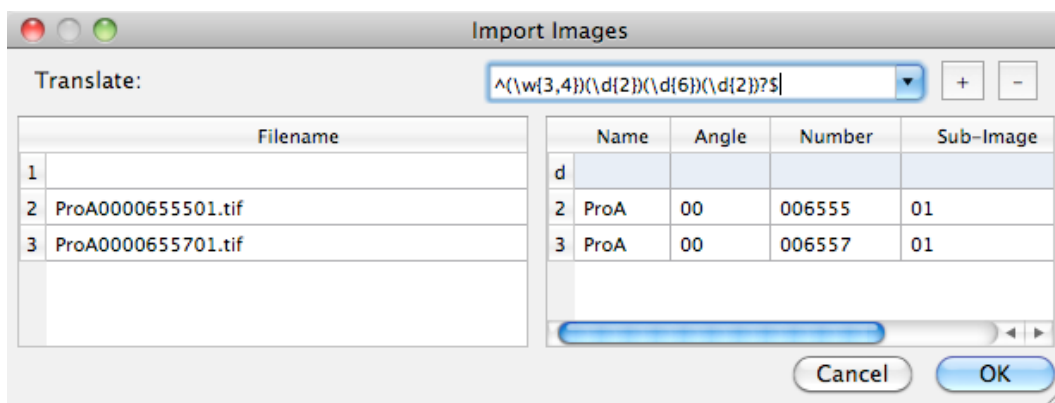
Figure 15: The file import dialogue of `2dx_merge`.

4. Now it is important to close also the program `2dx_merge`, without saving its local configuration. This way, the local database for `2dx_merge` in /protein-s/ProA/merge/2dx_merge.cfg remains as `2dx_image` has just saved it. Now reopen `2dx_merge`, so that it has a fresh working copy of the current database. This now represents the default values for this project.

4.3.3. Processing the Remaining Non-tilted Images

1. You can now proceed to import the remaining non-tilted images. In `2dx_merge` launch the pull-down menu option "Import Images." Specify the other non-tilted images in the file section dialogue. With suitable naming convention, the import dialogue will have already determined the nominal tilt category (00), and the image numbers, see Fig. 15.

2. The first line in the import dialogue headed by "d" can be used to manually define default values, which are then populating the fields for all images to import. For the image number, the given default value will be incremented for each of the following images.

3. This import function will copy the TIFF files into newly created sub-directories, and establish local 2dx_image.cfg configuration files for each image in those sub-directories.

4. To automatically process these imported images, select them in the central top panel in the GUI of `2dx_merge`, by checking the box in front of each line for those images. Then launch the script "(Re)process all images." This will prompt `2dx_merge` to launch 2dx_image with all standard scripts on all selected images. For this, `2dx_merge` launches a GUI-less instance of 2dx_image,

and gives it as parameters the directory of the image to work on, and a "*"
as symbol to process all standard scripts in sequential order, see also ref.
[Gipson *et al.* 2007a]. For traps and pitfalls in this process, see Note 4.4.1.

5. After all images have finished processing, update the database in `2dx_merge`
   by clicking the top right update icon, and examine the result in the central
   top panel of the GUI of `2dx_merge`.

4.3.4. Verifying the Automatic Processing of the Non-tilted Images

1. The result of the processing of all selected images can be inspected in the
   top central panel of the GUI of `2dx_merge`. As described in section 2 and
   also in [Gipson *et al.* 2007a, Gipson *et al.* 2007b], the `2dx` software calculates
   a singular value termed "QVal" to quantify the quality of an image and the
   performance of the image processing. A higher QVal value means a better-
   processed image. The available images in the top central panel in `2dx_merge`
   can be sorted by the column QVal2, which is the QVal after the second un-
   bending script UnbendII in `2dx_image`. This will list the best images on the
   top of the image list, while the weaker images move to the bottom of the list.

2. The weaker images, where the automatic processing didn't produce satisfactory
   results, could simply be excluded from further processing by unclicking their
   selection boxes. This could be an option, if a very large number of 2D crystal
   images were available through automatic data collection on an automated cryo-
   EM system (e.g., driven by Leginon [Carragher *et al.* 2000] or other systems).
   In most cases, however, it is better to double-click on the images in that
   central panel of `2dx_merge`, to open them in `2dx_image`, and inspect or correct
   the processing steps.

3. To process different image batches with different default settings, see Note 4.4.2.

4. To implement the project-wide synchronization of parameters, see Note 4.4.3.

4.3.5. Processing of Images of Tilted Specimens

The processing of images of tilted crystals is done analogously to the method de-
scribed above. Import one good image of a tilted crystal with `2dx_merge`, which
for a 30 o tilted crystal would in the above example create a sub-directory ProA30.
Process that image manually with `2dx_image`, and save the determined optimal pro-
cessing parameters as (temporary) project default values. Then continue importing
the remaining images from this (tilt) batch and try the automatic processing via the

`2dx_merge` script "(Re)Process all images". See also Note 4.4.4.

4.3.6. Merging

The image processing of the individual images results in text files with amplitude and phase values for the individual Miller indexed diffraction spots of each image. For images of non-tilted crystals, these can be merged in 2D into a merged projection map. If images of tilted specimens of sufficient quality are available, these can be merged into a 3D merged dataset. `2dx_merge` will produce a final 2D projection map, or respectively a 3D volume. The amplitudes and phases of these reconstructions are available in so-called APH and MTZ files in `2dx_merge`, as described in section 3.

4.3.7. Re-unbending with a Synthetic Reference

The merged reconstructions can be used to create references to reprocess all available images. As described in [Gipson *et al.* 2007a], the MRC program MAKETRAN [Kunji *et al.* 2000] can be used by `2dx_image`, to create a synthetic reference for the unbending of the individual images. For this, `2dx_image` would not use the conventional unbending scripts Unbend I and Unbend II, but rather use the script Synthetic Unbend. The parameter SYN_UNBENDING (label "Use Synthetical Reference if possible?") determines, if the Unbend I and II scripts, or rather the Synthetic Unbend scripts should be called. Also the remaining standard scripts in `2dx_image` change their behavior depending on the setting of this switch. For example, if the Synthetical Reference is to be used, then the scripts to determine the defocus or lattice or spotlist are not running, in order to not loose previously determined settings. The unbending algorithms for the scripts Unbend I and II are different than for the script Synthetic Unbend: Unbend I and Unbend II use a Fourier-filtered version of the original image, from which a box of a certain width (e.g., 500×500 pixels, corresponding to, e.g., 5×5 crystal unit cells) is cropped. While Unbend I uses a very narrow Fourier filter mask radius (e.g., 1 pixel) from the raw image, Unbend II uses multiple iterations of Fourier masking with larger masks on Fourier transforms of already unbent images, to produce increasingly improved references from this one image. In Unbend I and Unbend II, the raw image is always unbent only once, but each time with an increasingly improved reference structure. Optimization processes here concern two parameters: the Fourier masks for the filtering and the diameter of the boxed reference structure.

In contrast, the script Synthetic Unbend does not require iterative reference refinement, because its synthetically created reference is based on the entire dataset and therefore is likely a near-perfect reference. The optimization parameter here is now only one parameter: The diameter of Gaussian-shaped blobs placed into a synthetic

Fourier transform, which is then used directly as a reference structure.

Even though we assume that the first synthetically generated reference structure is already near-perfect, the implementation in `2dx_image` for the Synthetic Unbend nevertheless creates two different reference structures. These are created with blob diameters SYN_maska and SYN_maskb, where typically the first is a smaller value than the second. The reason for this is that the first reference with the more narrow Fourier blobs corresponds to a larger-area reference pattern in real-space, which is used in a first round to determine the coarse lattice distortion parameters. These are stored by the MRC program QUADSERCH in an ERROR field. The second synthetic reference, created with the larger masking value SYN_maskb, shows a smaller-area reference pattern in real space. This second synthetic reference is used to refine the ERROR field, while only allowing minor deviations from the originally determined lattice distortions (ERROR field).

The image unbending with the synthetic reference should result in an improved reconstruction over the conventional unbend- ing. However, the availability of a correct synthetic reference depends on the presence of a true and sufficiently sampled 2D or 3D reference dataset, and for tilted images sensibly depends on a correct assignment of the tilt geometry for this image.

The image unbending with the synthetic reference should result in an improved reconstruction over the conventional unbending. However, the availability of a correct synthetic reference depends on the presence of a true and sufficiently sampled 2D or 3D reference dataset, and for tilted images sensibly depends on a correct assignment of the tilt geometry for this image.

4.3.8. Automating Custom Scripts and Procedures

`2dx_merge` also offers the possibility to apply the user's own custom scripts or procedures to a larger number of images. This can be done by adapting and launching one of the Custom Scripts "CustomScript1" to "CustomScript3". Double-clicking such a script name will open it with a text editor. These scripts contain code to apply specific (custom) commands to all selected image directories. The commands to adapt appear after the text section "Modify only the section below." You can use such a script for example to set a specific variable to a certain value for all selected image directories. This can be done by un-commenting and adapting a command like

```
echo "set TANGL = 0.0" >> LOGS/${scriptname}.results
```

This would reset the tilt angle variables TANGL to zero for all selected image directories. Another useful command is

```
echo "set SYN_Unbending = 1" >> LOGS/${scriptname}.results
```

which would switch all selected images to processing with the synthetically generated reference.

These custom scripts can also be used to execute one or several standard or custom scripts of `2dx_image` for all selected images. This makes use of the GUI-free command-line modus of `2dx_image`. If `2dx_image` is called from the command line with the image directory and the script identifier as parameters, it will execute the script as background process. This could for example be done with

```
${app_2dx_image} ${newdir} "+2dx_refinedefocus"
```

to launch the script 2dx_refinedefocus. You could for example use a Custom Script in `2dx_merge` to apply the `2dx_image` custom script "Refine Defocus from Merged Data" to all selected images. In this example, this script would use the MRC program CTFSEARCH to refine the defocus and astigmatism for all selected images, using the comparison between the image amplitude and phases and the merged reference dataset.

4.3.9. Optimizing the Performance of `2dx` on Your Computer

The `2dx` software package launches MRC programs, which were originally written in the 1970 in Fortran-77, when computers had little amounts of RAM and were still using tape drives. Consequently, these programs rarely store images in the RAM, but permanently read and write files from and to the local hard drive. Some of these operations read and write the image files in stripe sections of the images, which further intensifies the disk access. The performance of the `2dx` image processing workflow on a modern multi-core computer therefore depends mostly on the speed of your hard drive. To allow acceleration of disk access, `2dx` arranges the majority of the larger image files in sub-directories called FFTIR and SCRATCH.

The disk access can be significantly accelerated, if those FFTIR and SCRATCH sub-directories are located on a separate disk with lower latency, e.g., a Solid State Drive (SSD). Alternatively, a smaller performance improvement can already be achieved by placing those sub-directories on separate hard drives, so that the data input and output flow are spread over three drives: The main data drive, another drive for the FFTIR directory, and a third drive for the SCRATCH directory. Big speed increases are also gained when those FFTIR and SCRATCH directories are located on a solid state drive (SSD).

1. To allow comfortable management of such sub-directories on separate disk, use the `2dx_merge` parameters "Location of fast FFTIR disk" and "Location of fast SCRATCH disk" to define a path to a fast disk. If a fast SSD drive is mounted under /Volumes/SSD, then these two parameters could be set to "/Volumes/SSD." This would then prompt `2dx` to replace all FFTIR and

SCRATCH sub-directories in all image directories with links to appropriate sub-directories on that SSD drive. Especially when processing several images simultaneously on one multi-core computer, speed gains by a factor of 10 or more can be gained with this setting.

4.3.10. The Overall Workflow for a 3D Structure Determination

The 3D structure of a 2D crystal membrane protein project can be determined as described above: A first non-tilted image is processed with `2dx_image` (semi-) manually, and all image processing parameters are refined and saved as project default values. These are then used to automatically process the remaining non-tilted images. Results can be manually inspected and manually optimized, if desired. Additional tilted images are then added, by processing the first image of each tilt angle category manually, followed by automatic processing of the remaining images from that tilt angle category.

After successful 2D and/or 3D merging of the data, all available images can be reprocessed from `2dx_merge`, using the merged dataset to produce synthetic reference data.

Iterative 2D/3D merging and reprocessing of the entire data- set can be used to improve the final resolution.

4.4. Notes

4.4.1. Comments on the Automatic Processing via `2dx_image`

`2dx_image` optionally allows fully automatic processing of one 2D crystal image, by double-clicking in the GUI of `2dx_image` on the top bar of the Standard Scripts panel. This will select all standard scripts, which are then sequentially executed by clicking on the "Run" button. `2dx_image` includes programs and scripts to automatically determine the processing parameters, and then perform the processing. For this, we extended the MRC programs [Crowther *et al.* 1996] with additional programs and functions and also combined it with other programs, e.g., CTFFIND [Mindell & Grigorieff 2003]. The following steps are performed in the fully automatic mode in `2dx_image`.

- The defocus in the center of the image is determined via CTFFIND.

- The approximate tilt geometry is determined via CTFFIND or CTFTILT.

- One or several lattices are determined, taking a potential tilt geometry into account [Zeng *et al.* 2007a].

- If the project-wide crystal lattice dimensions are known, they are used to refine the tilt geometry, based on the distortion of the identified lattice in this image, and provided that the tilt angle is higher than 25°.

- The list of spots in the Fourier transform is determined via MMBOX and refined.

- Unbending in step I and step II is done, and the FFT of the unbent image is evaluated to give measurements for amplitudes, phases, background values, and IQ values.

- These measurements are then subjected to correction for the CTF.

- A final map is calculated.

This process can go wrong at several points.

If on an image of a tilted crystal the defocus determination or the tilt geometry determination is failing, e.g., because the image was recorded very close to focus so that no Thon rings are discernible, then a wrong tilt geometry determination will make it unlikely that the correct lattice is found. Without lattice identification, all following processing steps fail.

If the lattice identification fails, for example because a SpotScan image has interfering low-resolution patterns in the Fourier transform, then an adjustment of the lattice search parameters to ignore low-resolution diffraction spots, may help. If the lattice identification fails due to the presence of multiple closely overlapping lattices, or due to other reasons, then manual lattice identification at this single step may rescue this image, and the subsequent image processing steps may still be executed automatically.

2dx_image allows locking specific parameters, e.g., the lattice parameters, so that their values will be protected from overwriting. This would be advisable in the above example: The manually determined lattice parameters should be locked, to protect them from accidental deletion or changes through other (automatic) processing steps.

4.4.2. Processing Different Image Batches with Different Default Settings

The here described automatic processing assumes that all imported images can be processed with the same default parameter settings. This would be the case, if the images stem from the same microscope, operated under the same acceleration voltage, recorded from specimens on the same grid or prepared by the same grid preparation method, and imaged under the same magnification and defocus settings.

If, however, images from different sessions on different microscopes are present, the

above-described approach would not work. In order to enable automatic processing
of different image batches, the user can create more than one super-directory for
this tilt angle image class. In the above example, we called the protein ProA, and
the non-tilted images were imported into a sub-directory called ProA00. The user
can manually rename that directory name ProA00 to ProA00_batch1. (Please note
that no spaces are allowed in file or directory names.)

Additionally imported non-tilted images will now be imported into a new sub-
directory ProA00, so that they can be processed and handled in the `2dx_merge` GUI
separately from the previously imported images. At a later stage, the new directory
ProA00 can be renamed to ProA00_batch2.

In the above example, the master configuration files 2dx_master.cfg are still all sym-
bolically linked to the configuration file in the merge directory, called 2dx_merge.cfg.
Because the initialization of the 2dx_image.cfg files happens during import of the
new images via 2dx_merge, the default settings for acceleration voltage and magni-
fication, etc., are inherited from 2dx_merge.cfg at the time of import. This feature
can be exploited to import a larger number of images recorded at one magnifica-
tion, then rename the super-directory ProA00 to ProA00_batchX, then import one
additional image from the new batch, which creates a fresh ProA00, define in it
the new magnification, save that as new project default, and then go on to import
the remaining new images, which then will be all imported with the new (other)
magnification setting into the new ProA00.

If the user wants to permanently uncouple the default settings for one batch of images
from another batch of images, those sym-bolic links 2dx_master.cfg can be deleted
and replaced with local copies of the file 2dx_merge.cfg. This allows for processing
and reprocessing of large numbers of images with different default parameters. For
this, see also the comments on parameter synchronization.

4.4.3. Project-Wide Synchronization of Parameters

Certain image processing parameters need to be synchronized throughout the entire
2D crystal processing project. These concern, for example, the crystal symmetry
space group (e.g., P4), and the real-space dimensions of the crystal lattice (e.g.,
$103.0\text{Å} \times 103.0\text{Å}$ with an included angle of $90°$). Such parameters should be valid
for all images that belong to one project.

The configuration files are 2dx_image.cfg for images and 2dx_merge.cfg for the
merging. They all have identical structure and syntax, as described above.

If `2dx_image` is launched for a certain image directory, it will read the local configu-
ration file 2dx_image.cfg for this image. If in this configuration file the parameter

```
# SYNC_WITH_UPPER_LEVEL: YES
```

were set to YES, `2dx_image` will ignore the local value setting for this parameter, and

instead keep the setting from the upper-level configuration file. If for example the configuration file 2dx_merge.cfg has a symmetry definition of P4, and the upper-level files 2dx_master.cfg are all symbolic links pointing to this one 2dx_merge. cfg file, and if for the symmetry the parameter SYNC_WITH_UPPER_LEVEL is set to YES, then that P4 setting will overwrite the local setting in each image as soon as 2dx_image is opened on them, so that the entire processing for all images assumes the same symmetry.

This allows the user to change a central parameter only once, and all images in this project will then be (re)processed with this coherent setting. If instead you wish to locally try different settings for a certain parameter, i.e., not to have a parameter sync with the upper level configuration file, you can change that SYNC_WITH_ UPPER_LEVEL parameter in the GUI of `2dx_image`, via right-clicking on its parameter, and then clicking on the current value (yes or no) of the SYNC_WITH_UPPEL_LEVEL setting.

4.4.4. Processing Tilted Images

In general, processing up to 30° tilt can be done with CTF correction algorithms, while the processing of high-resolution images at tilt angles of 40° or higher require TTF (titled transfer function) correction algorithms.

Since the processing of images of specimens at higher tilt angles becomes increasingly more complicated due to the anisotropic distortions in the lattice and other image parameters, we recommend to first start with lower tilt angle images, and then include the higher tilt angles later in the project.

4.4.5. Available Sources of Information

The `2dx` image processing software is based on the so-called MRC image processing programs [Crowther *et al.* 1996], which are in their original form available from Richard Henderson and Jude Short at the MRC, Cambridge. These MRC programs are also included in the distribution of `2dx`, though in a modified form, allowing interaction with the `2dx` environment, and we added in few cases additional functions to the MRC programs, mostly related to automation. The detailed flow of data from one MRC program to the next and the involved file formats are documented at http://2dx.org/documentation/2dx-software/manual/data-flow. `2dx` also includes additional programs for a Maximum Likelihood (ML) based single particle processing, which for non-perfect crystals may significantly improve the resulting resolution [Zeng *et al.* 2007b]. The resulting map of the ML processing is thereby translated into amplitudes and phases, so that ML results for individual images can be used for 3D merging with lattice line interpolation via the remainder of the MRC-based

`2dx` suite.

`2dx` is available as open-source under the GPL software at `http://2dx.org`. This website also contains a documentation of the `2dx` software, the included MRC programs, the parameters, and the general algorithms. A general introduction to electron crystallography on this web server, as well as detailed instructions to specific MRC programs were contributed by Anchi Cheng and Vinzenz Unger. The usage of `2dx` and also the image processing packages IPLT [Philippsen *et al.* 2007b] and MRC was lectured on at several workshops, including the EMBO Cryo-EM workshops, and the biennial workshops on Electron Crystallography of Membrane Proteins, described at `http://2dx.org/workshop`.

4.4.6. Other Image Processing Systems to Facilitate Processing of 2D Crystal Data

For the processing of electron diffraction data, the MRC software package offers robust programs that automatically index diffraction patterns and evaluate the data from those patterns. The XDP software tool [Mitsuoka *et al.* 1999] can be used to facilitate the evaluation of diffraction pattern.

A new development for electron crystallography data processing for images and diffraction patterns is IPLT, [Philippsen *et al.* 2007b, Philippsen *et al.* 2003] a package that does not make use of the MRC software suite, but provides a reimplementation of existing algorithms as well as newly developed algorithms for 2D crystal processing in a modular way.

These software systems were recently comparatively evaluated by Schenk et al. [Schenk *et al.* 2010].

Another image processing system for 2D crystal data is the Calidris system available at `http://www.calidris-em.com` [Hovmöller 1992, Zou *et al.* 1993, Zou *et al.* 2004].

4.5. Conclusion

Automation in the processing of images of 2D crystals of membrane proteins is implemented in the `2dx` software package for the processing of individual non-tilted and tilted images. Successful automation of the image processing depends on the correct setting of suitable default parameters and on the availability of images of comparable image and crystal quality.

While `2dx` offers user guidance and streamlines the 2D and 3D merging process, further software development is needed to also automate this merging process. A first step into this direction is the projective constraint optimization (PCO) method that was recently developed for 2D crystal data [Gipson *et al.* 2011]. As is described in more detail in section 3 of this volume, with sufficiently oversampled data at

lower tilt angles PCO was shown to be able to fully automatically calculate a 3D reconstruction from the available dataset from non-tilted and tilted 2D crystals, while bypassing lattice line interpolation, and on the same creating correct amplitude and phase values also in the so-called missing cone. `2dx_merge` is also prepared for the processing of large numbers of images of 2D crystals, so that future efforts can now be extended to the automation of the collection of high-resolution electron crystallography data.

Acknowledgment

4.6. Further Development

Since writing this book chapter for **Electron Crystallography of Soluble and Membrane Proteins** [Arheit *et al.* 2013b] time has passed and the development in the automation of image processing has continued. Using the described automation workflow in different structural studies has revealed common sources of errors. It has also revealed missing steps in the pursuit of full automation. There have been improvements to the automatic workflow from my side to avoid the identified errors. In addition the automation of merging has been facilitated. This section is an add-on to the published chapter on **Automation of Image Processing in Electron Crystallography**.

4.6.1. Improvements in Automatic Processing of Images

Common errors in automatic processing unfortunately occur in the most crucial steps of the processing, in *Lattice Determination* and *Defocus Determination*.

Lattice Determination The lattice determination is crucial in the automatic processing of 2D crystal images. If the lattice is not found, this means that the protein structure is not found. To determine the lattice two algorithms are available: "Get Lattice" and "Find Lattice". The first has to be used on the first micrograph, when the unit cell dimensions of the 2D crystal are not known. "Get Lattice" has been changed so that the resulting vector pair always includes the smaller angle if the lattice is not rectangular and that the first vector has positive coordinates. This is just a convention that helps to merge the data later in the workflow. Additionally we defined a larger exclusion radius for spot scan images to exclude the low frequency spots from the scan. The processing of the first image is usually supervised by the user, hence the outcome of "Get Lattice" is validated. It can also be evaluated by checking if the QVal of the subsequent unbending is above a certain threshold.
The automatic image processing workflow now always uses "Find Lattice" after the unit cell dimensions have been determined through the first image. This is the first step to use the outcome from prior images. The other hand of lattice vector pairs is not checked anymore, also to allow merging of the data later.

Defocus Determination We have shown that adapting the search range to the nominal defocus of the microscope improves the result. In the Section 4.6.3 we will discuss reading the acquisition data e.g. defocus from the image header. For images of tilted crystals we introduced a new more robust algorithm where the defocus gradient on 7 by 7 patches is measured to determine the tilt geometry. This

method still uses CTFFIND [Mindell & Grigorieff 2003] to fit the Thon rings. There is an international CTF estimation challenge in progress, where we provided a test data set (`http://i2pc.cnb.csic.es/3dembenchmark`). The resulting benchmarks will define the further development.

Additionally, smaller fixes have been made for the image processing workflow to be less error prone. A new a custom script was implemented in `2dx_merge` which is called "Automatic Image Processing" and can be run on a selection of images. This script runs, similar to the "(Re)process all images" script, all the steps needed to get from a raw micrograph to a unit cell of the membrane protein. But in contrast to the former script, it subsequently automatically masks the image based on the cross-correlation map of the second unbending and then runs the unbending and CTF correction again for the masked image. This has been proven to improve the quality of the resulting projection maps, since non crystalline areas where masked out. This workflow was used and validated on images from crystals of the human copper transporter hCTR1 from Vinzenz Unger's group [De Feo *et al.* 2009]. Still the initial processing parameters were chosen manually. But one can make use of the synchronization of the configurations file. The behavior of the SYNC_WITH_UPPER_LEVEL attribute described in Section 4.4.3 has been changed in a way, that when changing a parameter that is synchronized with the upper level on the image level it is propagated to the global configuration file. This is a simple way to carry over processing results from one image to the next one.

Another new step in the automation workflow is the space group determination and subsequent symmetrization of the projection map. Again this functionality was already provided by `2dx_image`, but not used in the automation. The phase group and symmetry determination not only gives multiple measurements for the same diffraction spots, but also a measure of quality with the phase residual. The phase residual can be used as alternative to the QVal to exclude low quality images in the merging procedure. Another benefit is that the symmetrized projection maps are centered in their phase origin, which already aligns them for merging. Unfortunately the symmetrization can only be used for non-tilted crystal images.

4.6.2. Automatic 3D Structure Determination

3D Structure determination involves merging, scaling and aligning the structural data from images of different oriented crystal together. This procedure has been described in section 3. In the conclusion of this chapter (subsection 4.5) it was stated that the automation of the merging is not yet provided. We have made advancements in this area in order to facilitate automatic merging. The main contribution towards this automation was the implementation of a command line version of `2dx_merge`. Like in the case of `2dx_image` this allows to define and write workflows for merging
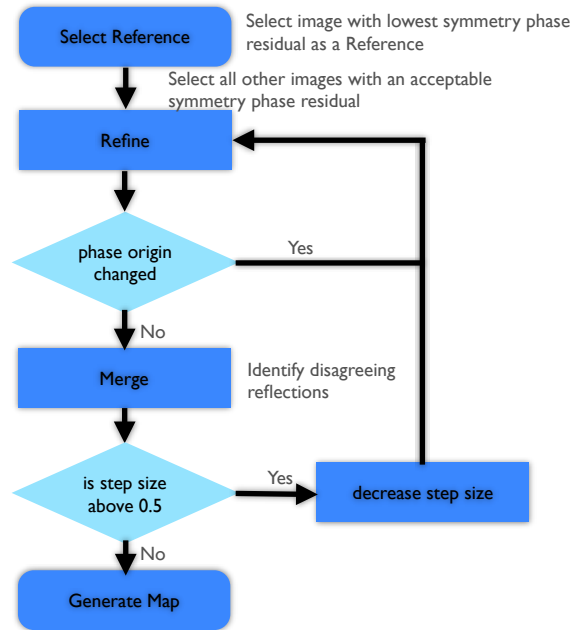
Figure 16: Flowchart that depicts the workflow of 2D merging in `2dx_merge` for al-
ready with `2dx_image` processed crystal images.

in form of scripts. For example to merge images in a given selection file one would
have to write

```
${app_2dx_merge} ${projectdir} 2dx_merge ${selection_file}
```

where `${app_2dx_merge}` points to the `2dx_merge` executable, `${projectdir}` depicts the
project directory and the `2dx_merge` script is executed with the files specified in the
`${selection_file}`. If the selection file is not given the script will be executed on the
last specified file selection. One can also append a comma separated list of image
files that should be processed to the command. With this scheme every `2dx_merge`
processing step can also be called from the command line. To be able to control the
processing further we have implemented functions that allow us directly to get and
set the processing parameters from the configuration. This allows for example to
get the overall phase residual of the 2D merging step

```
get_processing_param overall_phase_residual_2D
```

or one can set the step size for the refinement in the following manner

```
set_processing_param MergeStepSize "1.0"
```

With the command line version of `2dx_merge` and the ability to control the processing
parameters automatic merging can now be performed through a script. We were able
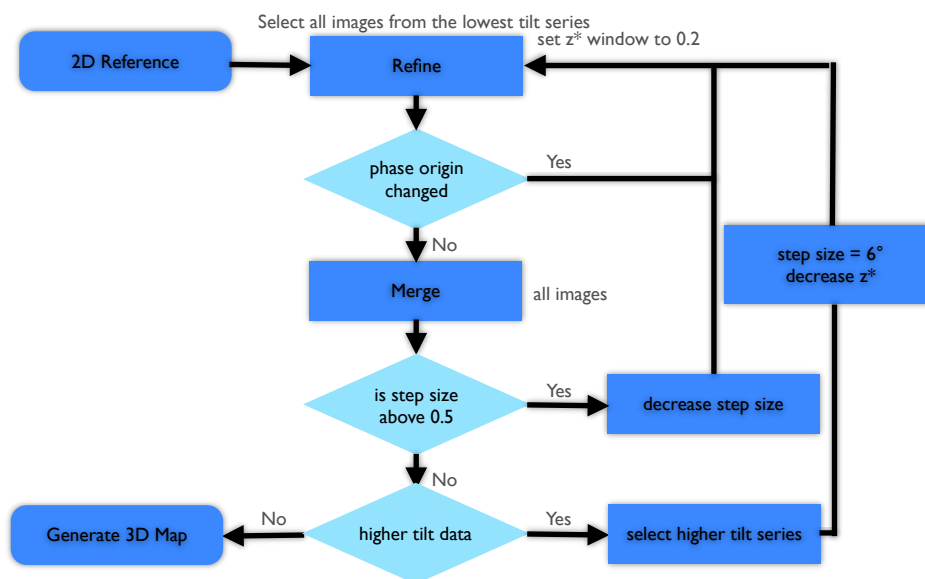
Figure 17: Flowchart for the 3D merging workflow in `2dx_merge` that leads to a 3D density map.

to perform 2D merging automatically by following the flowchart in Fig. 16.

After 2D merging one ends up with a merge projection map. This map is the initial reference for the 3D reconstruction. As described in section 3, 3D merging starts by aligning (*Refine*) images of tilted crystal to the given reference. For this images with the lowest tilt series are chosen. The parameter restricting the search range of the alignment are the *stepsize* (vertically) and z^* (horizontally). For each tilt series the images are iteratively aligned while decreasing the step size until the phase origin of each image to a certain extent does not change anymore. Then the images can be merged together, which results in a novel 3D reference. This procedure is then repeated for images of higher tilt series while reducing the vertical search range z^*. The whole procedure of 3D merging is depicted int the flowchart in Fig. 17. This is the automated workflow of 3D merging.

4.6.3. Future Work

While we now theoretically provide an automated workflow to get from 2D crystal images to a 3D reconstruction, there is still much room for improvement. There are still steps in the workflow that are error prone and could be optimized. The workflow is fault tolerant in the sense that it continues the processing even if a single step has not produced the right outcome. It is not the goal to have a fully

autonomous system, but what is missing is a clear signaling system that gives the user a summary of where errors occurred. It exists to some extent in form of the `2dx_logbrowser` but just for each local processing step. But the signaling mechanism should supervise the processing outcomes of every step and stop the processing whenever it thinks that the output of a program is not convincing, because it either deviates too much from prior knowledge or it outputs values that are generally not realistic. This involves defining such criteria for every processing step outcome.

The measure that defines the outcome of the unbending procedure is the QVal. We have implemented scripts to optimize the two essential unbending parameters: the Fourier filtering mask and the size of the reference. Unfortunately the way the QVal is defined, the optimization always favors the smallest reference with the largest mask. Hence there is a need for a better measure of the unbending.

The automatic processing can be improved by incorporating prior knowledge either from acquisition or previously processed images. There is ongoing work to automatically fill the image configuration file with the acquisition parameters that are stored in the header of the image file. This has only become relevant with the recent developments of direct detectors that have a detective quantum efficiency (DQE) that allow imaging of 2D crystals with good resolution. Previous generations of digital detectors (i.e. CCD, CMOS) where not suitable for 2D crystals, hence micrographs were still recorded on film and subsequently scanned. This development of digital recording has also revitalized the use of a dedicated image acquisition folder, that is scanned frequently for new images and then triggers their processing.

Prior knowledge from previously image processing is for now only implemented insufficiently with the synchronization of parameters. Adapting a processing parameter based on the outcome of all the other images could already improve the processing. One could also think of applying a probabilistic design to optimize the processing by implementing e.g. a bayesian approach [Scheres 2012].

5. Conformational Changes of Proteins Detected by Electron Crystallography: Difference Map Calculation and Analysis

The content of this chapter has been formatted for submission to Journal of Structural Biology. It has not been published at the time of the thesis submission. [Arheit *et al.* 2016].

Abstract

Membrane proteins can be studied by electron microscopy when organized in 2D crystals. Conformational changes undergone by membrane proteins of 2D are optimally visualized by difference maps. However, the correct interpretation of such maps, requires a reliable method to identify significant differences, e.g., a statistical hypothesis test or a more general method that empirically determines the background noise level. The performances of two existing and two newly introduced methods are compared by using synthetically generated 2D crystal data sets with the membrane protein in different conformations at various signal-to-noise levels. Statistical methods assume a specific probability distribution and the two tested both had drawbacks: The Student's t-Test carried out in real space does not hold when more than straightforward averaging is required to merge projection maps, and the new statistical method based on Welch's t-test and carried out in Fourier space does not translate well to the image domain. The two empirical approaches tested performed better, whereby the new mixed-merge method was more reliable than the classic method. Further, both empirical methods became more robust when locally varying thresholds were introduced to account for image heterogeneity. The mixed-merge method considers data heterogeneity within and across conformations and can cope with varying unit cell size. The improvement brought by this new approach and the use of local thresholds is demonstrated using experimental data documenting the effect of ligand binding to the secondary citrate/sodium symporter CitS and the potassium channel MloK1.

5.1. Introduction

The relationship between protein function, sequence and structure is complex and accurately predicting function from knowledge of the latter two is still difficult [Radivojac *et al.* 2013] Indeed, protein function is frequently dependent on specific conformational changes, which can be induced by many factors including change

in pH, voltage, ion concentration, phosphorylation, temperature or ligand binding. For meaningful conclusions to be drawn it is, therefore, important to obtain structural information under native conditions, or conditions similar to those employed for the biochemical and kinetic characterization of the purified protein. This can be achieved by cryo-electron microscopy (cryo-EM), where protein samples are shock-frozen in a thin layer of vitrified ice and imaged. Importantly, membrane proteins can be embedded in a lipid bilayer, crystallized in two dimensions, and examined vitrified in this close-to-native environment before and after a pH change or the addition of a ligand [Abe *et al.* 2011, Kebbel *et al.* 2013]. The images are projection maps of the crystal structure, and differences between them evident in so-called difference maps indicate the regions where conformational changes occur. Resolutions of 7 to 6 Å have proved sufficient for this [Kebbel *et al.* 2013, Kowal *et al.* 2014]. Further, if the two dimensional (2D) crystals are well-ordered and highly planar, electron crystallography can yield high resolution 3D structures; a 1.9 Å resolution was obtained for aquaporin-0 [Gonen *et al.* 2005]. The fact that this resolution is not routinely obtained indicates the need for improved experimental protocols and image processing. The recent advent of direct electron detectors and movie-mode imaging [Campbell *et al.* 2012] will facilitate the latter; these detectors already allow resolutions of 3 to 4 Å to be achieved by the analysis of single protein complexes [Kühlbrandt 2014]. Nevertheless, methods to reliably determine which differences between projection maps are 'true and which arise from sample heterogeneity and imaging noise are still required.

Difference maps reveal changes in protein density, and are obtained by simply subtracting one projection map of a 2D crystal. They visually document the net structural change that takes place when a protein shifts from one conformational state to another, and identify the positions at which changes occur. To ensure that only significant differences are interpreted, one has to consider how the compared projection maps were computed. Although measures are generally taken to detect and eliminate data heterogeneity when projection maps are calculated, it is still difficult to be certain that the differences between two independent maps arise solely from conformational changes. Difference maps were first used by Nogales et al. to visualize the conformational changes occurring when taxol binds to tubulin [Nogales *et al.* 1995]. They calculated the variance of the data for each conformational state and used the Student's t-test[Student 1908] to determine which differences were significant. Later, to detect pH-induced structural changes in a sodium/proton antiporter Vinothkumar et al. used contour plots to assess the difference map obtained by electron crystallography[Vinothkumar *et al.* 2005]. First, the data for one protein conformation was randomly divided into two sets that were independently analyzed; the standard deviation of the corresponding difference map was taken as a measure of background noise. Second the difference map of the two conformations was calculated and contoured by a multiple of the determined background noise.

The present paper focuses on electron crystallography and the use of difference maps to define the conformational changes occurring in membrane proteins. Two new methods to identify significant changes are introduced: (i) A statistical hypothesis test and (ii) a non-probabilistic way to account for data heterogeneity due to imaging noise and sample imperfections. In extended versions of the latter methods, a locally adaptive threshold is used to account for local data variations. New and existing methods are evaluated using series of synthetic 2D crystals generated with varying signal-to-noise levels. The performances of the best new method and the best classic approach are then compared using experimental data, i.e., difference maps generated to determine conformational changes undergone by the secondary citrate/sodium symporter CitS [Kebbel *et al.* 2013] and by the cyclic nucleotide-modulated potassium channel MloK1 are assessed [Kowal *et al.* 2014].

5.2. Approach

The analysis steps required to obtain merged projection sets, to calculate a difference map and to determine the significant differences are implemented in the `2dx`software (version 3.3.0; `www.2dx.org`).

The first step is to merge images recorded from cryo-EM grids prepared from 2D protein crystals stored under the same conditions (i.e., buffer, pH, presence/absence of ligands etc.). The merging procedure ("Final Merge" `2dx` script) is performed in the Fourier domain, the information from every crystal for each lattice reflection (amplitude, phase, background, CTF) being combined, weighted and scaled. The merged data set is then transformed to the real domain, resulting in a merged projection map. At this stage it is important to ensure, that the same phase origin is used for the different data sets. This is straightforward as long as the same plane group is applied to all projection maps ("Generate SymMap" `2dx` script). However, if a conformational change results in a different plane group, it might be necessary to set the phase origin manually ("Set PhaseOrigin Manually" `2dx` script). As different data sets are merged separately, in the "Difference Map" script the amplitudes of all conformations are scaled so that they have the same mean. Possible changes in the unit cell dimensions also have to be considered prior to difference map calculation. If necessary the dimensions are adjusted by cropping the larger unit cell to the width and height of the smaller. Finally, a raw difference map is calculated by subtracting corresponding pixels of one projection map from those of the other. Densities that are stronger in conformation 1 (minuend) are assigned positive values in the resulting map and densities that are stronger in conformation 2 (subtrahend) negative values. The difference map can then be visualized by mapping the values to colors. The values are then classified as significant, i.e. related to changes in protein

Table 1: Characteristics of the different methods to determine significant changes in difference maps

Method	Domain	Probability Distribution	Threshold
Student's t-Test	Real	defined	local
Welch's t-Test	Fourier	defined	local
Classic	Real	non-defined	global/local
Mixed Merge	Real	non-defined	global/local

conformation, or as background noise arising from variations within the individual data sets. In practice, a threshold is imposed so that the difference map only displays significant changes. This threshold can be determined in various ways as detailed below.

5.2.1. Determination of Significant Differences

Four different ways to determine the significance of differences were investigated. As summarized in Table 1, these were, (1) a classic statistical approach including a Student's t-test method in real space, (2) a new statistical method developed here that uses Welch's t-test in Fourier space, (3) the classic non-statistical approach that relies on an empirical determination of background noise, and (4) a new 'mixed merged' method developed here that merges all data sets compared and empirically determines a local threshold to discard background noise. Further, the local thresholding variant was examined for the classical non-statistical method (3) as well as for (4). The required scripts were implemented in the `2dx` software [Gipson *et al.* 2007a, Gipson *et al.* 2007b] and are available for use; see custom script "Diffmap" in `2dx merge`).

5.2.1.1 Statistical hypothesis test in Real domain: Student's t-test method

The Student's t-test [Student 1908] was first used in 1987 by, Trachtenberg and DeRosier to analyze variations in cryo-EM data and determine which features in single particle reconstructions were reliable [Trachtenberg & DeRosier 1987]. It was later adapted for use with projection maps obtained from images of 2D crystals by [Nogales *et al.* 1995]. The approach calculates the probability that the mean pixel intensities of the compared projection maps are the same (null hypothesis; Fig. 25). This is done by applying Student's t-test pixel by pixel in real space. The result is a binary classifier that determines which differences are significant (Fig. 25, Classification). The classifier is applied to the raw difference map (Fig. 25,

A-B). Only the differences considered significant are retained, the others are set to zero. The Student's t-test assumes that the set of intensities obtained for a specified pixel across different images of the same conformation form a normal distribution, that each intensity value contributes equally to the final merged map and that the variance of the compared data sets is the same.

5.2.1.2   Statistical hypothesis test in Fourier domain: Welch's t-test method

In order to determine the significance of differences between two projection maps, one has to consider how the single image data was merged, the dimensions of the unit cells and the plane group symmetry. The following can only be applied if both of the latter parameters are the same.

In the Fourier domain the image data of every 2D crystal is a list of reflections belonging to a lattice depicted by Miller indices. Merging image data of different crystals means primarily grouping reflections with the same Miller index together. This means that if n crystals are in conformation 1, there will be a maximum of n data points for every Miller index (not every reflection is present in the Fourier transform of every image). The plane group symmetry of the 2D crystals is then considered, and the number of data points associated with a reflection is increased by the number of symmetry-related reflections, i.e., the so called symmetry mates; e.g, if the plane group symmetry is P4, 4n data points are associated with each reflection. Comparing merged projection maps is now a matter of comparing reflections with the same Miller index (Fig. 18).

If the merging step simply consisted of averaging the data points associated with each reflection, then the standard deviation of the amplitudes of each reflection would suffice to determine the significance of the differences. However, in electron crystallography merging all the data points of one reflection to one amplitude and phase, involves weighting by the signal-to-noise ratio (SNR) and CTF correction of each image. The merged amplitude, $\bar{A}$, is calculated from the n data points for each Miller index (h,k) according to the expression,

$$\bar{A}(h,k) = \frac{\sum_{i=1}^{N} A_i(h,k)\frac{CTF_i(h,k)}{B_i^2(h,k)}}{\sum_{i=1}^{N} \frac{CTF_i^2(h,k)}{B_i^2(h,k)}} \tag{1}$$

where B is the background signal of the reflection peak. The merged phase is then calculated as angular mean

$$\bar{\phi}(h,k) = \arctan \frac{\sum_{i=1}^{N} w_i(h,k)\sin\phi_i(h,k)}{\sum_{i=1}^{N} w_i(h,k)\cos\phi_i(h,k)} \tag{2}$$

where each phase of each reflection is again weighted by a factor $w_i(h,k)$ proportional to its SNR.

Calculating the merged amplitude for each reflection for both 2D crystal conformations individually yields values for $\bar{A}_1(h,k)$ and $\bar{A}_2(h,k)$. The weighted mean amplitudes for conformations 1 and 2 can now be compared using a statistical hypothesis test. Although the assumption that the reflection amplitudes form a normal distribution is valid, the variance of the compared data sets might not be the same. Thus, Welch's t-test [Welch 1947] is employed. This adaptation of the Student's t-test is intended for use with two samples that possibly have unequal variances. It leads to the following expression,

$$t(h,k) = \frac{\bar{A}_1(h,k) - \bar{A}_2(h,k)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (3)$$

where s_i^2 are the sample variance and n_i the number of samples. With help of the Student's t-distribution, $t(h,k)$ can then be translated to the probability of $\bar{A}_1(h,k)$ and $\bar{A}_2(h,k)$ being the same. We chose a two-sided confidence interval of 90% as significance level.

Reflections that are not significantly different in the compared maps must be discarded when the difference map is calculated. In the unbending procedure, for example, the reference image is created by discarding high order reflections by setting their amplitudes to zero [Crowther *et al.* 1996]. However, removing Fourier components will just increase the Gibbs phenomenon [Gibbs 1898]. This overshoot of the Fourier series is negligible when only a few weak Fourier components are removed, but if more are removed can itself cause differences in the Real domain. Therefore, reflection amplitudes are not removed in Fourier space, but replaced by the average calculated from the corresponding amplitudes of the two conformations.

5.2.1.3 Classic method (Heterogeneity within conformations)

This method empirically determines which differences are linked to structural changes and which are caused by heterogeneity of the compared data sets. Unlike the above methods, probability distributions are not considered. The approach was introduced by Kühlbrandt et al.[Vinothkumar *et al.* 2005, Appel *et al.* 2009a]. The heterogeneity within one conformation, i.e., the background fluctuation, is determined by randomly splitting the data into two sets, calculating two independent maps and subtracting one from the other (Fig. 26). The resulting root-mean-square deviation (RMSD) of the difference map gives a good estimate of the fluctuations. The random splitting is repeated multiple times, giving various RMSDs. This is done
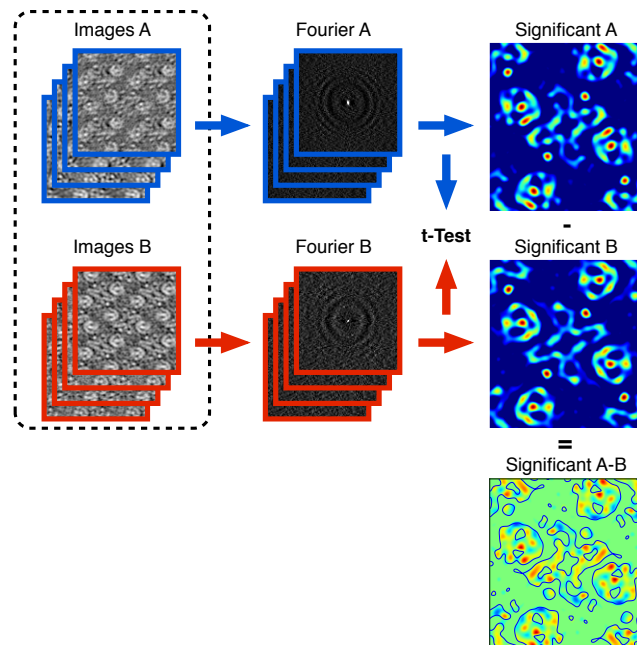
Figure 18: Welch's t-Test: The significance of each Fourier reflection is determined by comparing the mean amplitude of both conformations with the Welch's t-Test. The synthesized data sets were used.

independently for both conformations. The overall maximum RMSD (or a multiple of it) defines the contour level to employ when the differences between the two conformations are mapped using the CCP4 program NPO. In the resulting map every difference above the contour level is considered to be significant.

This classic way to determine significant differences calculates one global parameter, and uses this to define the contour threshold of the whole difference map. By contrast the statistical methods described in Sections 5.2.1.1 and 5.2.1.2 determine the significance of every difference by considering the local data variations. The latter idea is included in a modification of the classical approach also introduced here, where the variations of the data at each position of the final difference map are calculated (Fig. 26). This is done by comparing the difference maps of the split data sets used to determine the intrinsic data heterogeneity for each conformation. Since the random splitting is carried out multiple times, multiple local threshold values are obtained for each position. The maximum differences are selected in each case and displayed in a so-called variation map, which defines the local thresholds to be applied. Significant differences remain when this map is subtracted from the raw difference map (Fig. 26).

5.2.1.4   Mixed Merge method (Heterogeneity within and across conformations)

The new mixed-merge method determines a common background noise level for all of the compared data sets. This new approach is similar to the classic and modified classic methods described in section 5.2.1.3. However, instead of calculating an empirical background level for each conformation independently, the overall background variation is determined using all of the data. To do this the data for each conformation is randomly split into two sets of equal size and merged with the sub sets of the other conformation as illustrated in Fig. 19. This gives two independent projection maps, each based on data from both of the observed conformations. These maps are subtracted from one another, resulting in a variation map for all of the data being compared. To eliminate any effect of how the two data sets are split and merged, this procedure is repeated multiple times and the maximum local variation values obtained are stored. The latter values are used to generate the final variation map, which indicates the overall heterogeneity of the data decoupled from the conformational state of the protein. Subtracting this variation map from the raw difference map obtained for the two conformations, shows which changes are significant.
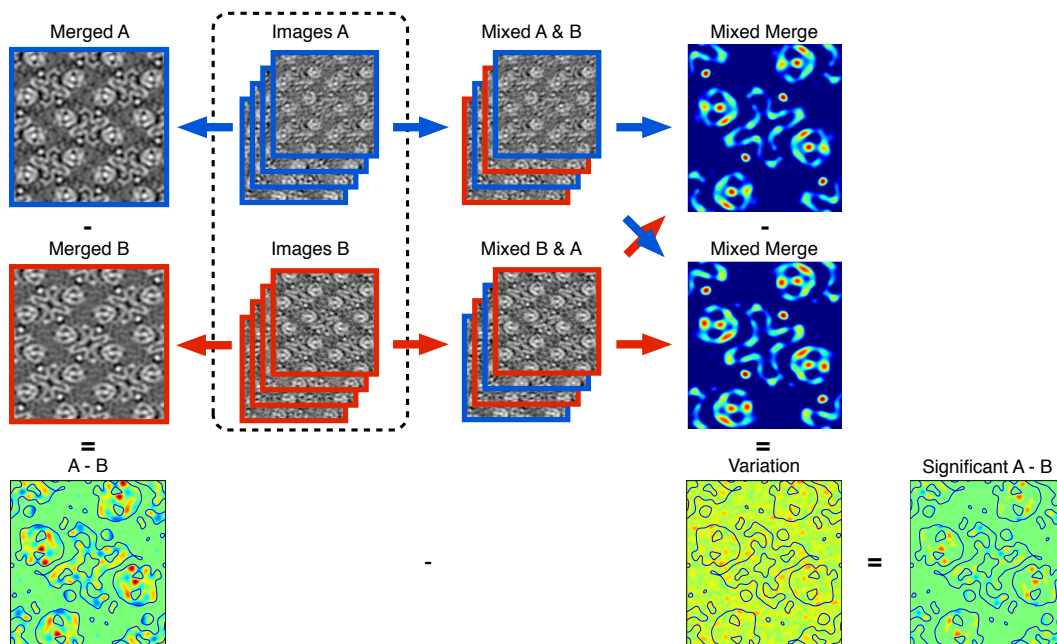
Figure 19: **Mixed Merge method:** Determination of data heterogeneity within and across conformations by the mixed-merge method. These variations yield in a lower threshold to determine significant changes in difference maps. The synthesized data sets were used.

5.3. Materials and Methods

5.3.1. Test data sets

The various methods were validated and compared using synthetic image data. To assess their performance in a 'real-case' scenario, the best two were then applied to experimental data.

5.3.1.1   Synthetical Data

The two final 2D projection maps obtained by separately merging data obtained for the two conformations of the secondary citrate/sodium symporter CitS of *Kleb-*

siella pneumoniae [Kebbel *et al.* 2013] served as starting point. Both were limited to 6 Å and symmetrized with the $P22_12_1$ plane group. The Dynamo software was then employed to generate synthetic crystals based on their unit cells [Castaño-Díez *et al.* 2012]. Ten synthetic crystals with random lattice orientation were generated for each conformation. In order to control the source of data heterogeneity, lattice distortions were not introduced and the crystal images were not modulated by a point spread function. Various degrees of Gaussian noise ($\mu = 0, \sigma = [0 - 150]$) were applied as required. The effect of this noise was determined by calculating merged projection maps and difference maps from the 'noisy' synthetic image data. The maps obtained from the noise-free synthetic crystals were taken as ground truth. The term noise-free is used to indicate the absence of added Gaussian noise; it is recognized that they were not completely devoid of noise, as the unit cell templates were taken from real data.

5.3.1.2 Experimental Data

Experimental data sets collected for the membrane proteins CitS [Kebbel *et al.* 2013] and MloK1[Kowal *et al.* 2014] were used. MloK1 is a cyclic nucleotide-modulated potassium channel from the bacterium *Mesorhizobium loti.* The 2D crystals of CitS had been incubated in situ in the presence of sodium acetate or sodium citrate, and those of MloK1 in the presence or absence of ligand, to induce conformational changes.

5.3.1.3 Computation

All four methods employed to determine significant changes in difference maps were implemented in `2dx` [Gipson *et al.* 2007a, Gipson *et al.* 2007b] version 3.3.0 and are available as a Custom Script "Diffmap" in `2dx merge`.

5.4. Results and discussion

5.4.1. Method assessment using synthetic data sets

The ability of the methods outlined above to reliably determine significant differences between projection maps was assessed using two sets of synthetic crystals based on two different conformations (here termed 1 and 2) of the citrate/sodium symporter CitS of *Klebsiella pneumoniae* [Kebbel *et al.* 2013] (see Materials and Methods). The quality of the synthetic crystals was varied by adding different amounts of

Gaussian noise to give SNRs ranging from of 1 to $\frac{1}{150}$ at standard deviation increments of 10. The advantage of this approach was that the ground truth, i.e., the difference map calculated from the two corresponding 'noise-free' merged projection maps was unambiguously defined (Fig. 21. The quality of the merged crystal data was determined for each conformation and noise level, by calculating the overall phase residual of the corresponding 6Å projection map (see Section 5.3.1.1; Fig. 27. The phase residual is the sum of the difference between the phase of each reflection and its average phase. The smaller the overall phase residual the better the data agrees and the better the quality of the merged projection map. A phase residual of 90° and higher indicates no agreement [Unger 2013]. As shown in Fig. 27, the synthetic crystals employed were of acceptable quality up to a SNR of $\frac{1}{150}$ after which the overall phase residual rapidly increased. Next, the synthetic data pairs, i.e, synthetic crystal conformations 1 and 2 at the same SNR, were analyzed. As illustrated in Fig 20, difference maps calculated by simply subtracting merged projection maps derived from noisy images, deviate from the ground truth (compare Fig. 20c and f). The robustness of the various methods to the amount of added noise, was assessed by comparing the resulting difference maps to the 'ground truth. A representative example where Gaussian noise was added to the synthetic data to give a SNR ratio of $\frac{1}{70}$, is shown in Fig. 20d-e and Fig. 21; To aid discussion, the positions of three true differences that occur at the substrate binding site[Kebbel *et al.* 2013] are indicated (black boxes in the protomer region surrounded by a dashed white line (Figs. 20 and 21); differences at the central dimerization interface (Fig 20c) are relatively minor.

As shown in Fig. 21, the Student's t-test is able to discriminate the effect of added noise to a certain extent (compare a, b and c). However although the difference map obtained is similar to the ground truth map, they are by no means identical (Fig. 21a, c). In particular, the Student's t-test tends to produce false positives, e.g., the dark blue densities at the dimer interface arising from noise added to conformation 2 (Fig. 20e). Although it considers local variations, the Student's t-test fails as the assumption that the merged map is the mean of the image data does not hold. In electron crystallography the image data is transformed into the Fourier domain to determine the contributing frequencies, which for a crystal span a lattice of reflections. In the merging step, each reflection is weighted according to its SNR and its corresponding CTF correction. Consequently, pixel intensities do not form the required normal distribution and do not contribute equally to the final merged map. Further, the variance of the compared image data sets might not be the same. The partial loss of symmetry evident in (Fig. 21c), e.g., in the dimer interface region (non-symmetric yellow/red densities), is not a result of the analysis but because non-symmetrized synthetic projection maps were compared in this particular example. The newly developed Welch's t-test aims to correct the invalid assumptions made in the above method. Rather than validating differences in the real domain,

this method determines the significance of differences in the Fourier domain (Section 5.2.1.2). The benefit of using the Fourier domain is that the contribution of each reflection is known and that the statistical hypothesis test can be applied to determine whether the weighted means of the two conformations differ significantly. Reflections with the same Miller indices are compared, hence both conformations must belong to the same plane group and have the same unit cell size. The disadvantage of the method is that determining significant differences between reflections in Fourier space means that you are determining frequency differences which, when transformed back to the real domain, do not occur in local regions but throughout the unit cell. (Fig. 21c). This so-called Gibbs phenomenon [Gibbs 1898] comes from discontinuities that are introduced by removing or modifying reflections and cause oscillations in the difference map. Notably, differences at the dimer interface are greatly overestimated by this new statistical approach.

The output of the Classic method shown in Fig. 21e indicates highly significant differences in the protomer region. However, visual comparison with the ground truth (Fig. 21a) shows that not all differences are detected. The global threshold determined by the Classic method is too high, which means that some differences are falsely erased. This is particularly clear at the dimerization interface where almost all differences evident in the ground truth are missing (compare Figs. 21a and e). On the other hand, the approach ensures that the remaining differences are indeed significant and must derive from conformational changes.

The mixed-merge method was developed to overcome the above threshold problem. Whereas in the classic method the background level is determined using the higher of the two separately determined thresholds, one for each conformation, the mixed-merge method allows the overall heterogeneity of multiple data sets to be assessed. This results in a more accurate, lower threshold. Although the results (Fig. 21f) are similar to those obtained by the classic method, they are closer to the ground truth (Fig. 21a). In particular, the mixed-merge method is able to determine significant changes correctly at the dimer interface. To conclude, the mixed-merge method proved to be the most accurate of the four methods tested, and the classic method the second best.

5.4.2. Consideration of local variations

In contrast to the statistical methods, the classic and mixed-merge methods use a global threshold to determine significant changes, with the disadvantage that local changes in the data variance are not considered. To overcome this, the calculation of a so-called *variation map* is now proposed as a refinement. The effect this has in a real-case scenario is demonstrated in Fig. 22; where various thresholding methods are applied to raw difference maps calculated by the classic (Fig. 22a) and mixed-

merge (Fig. 22a) methods from data recorded for the secondary citrate/sodium symporter CitS with either Na citrate or Na acetate substrate. Real experimental data was required since due to the simplified random noise model applied, variation maps calculated from the synthesized data did not show any distinct variation peaks.

The global threshold used in the existing methods has been defined as the maximum variation, the mean variation or a multiple of the RMSD of the raw difference map. Use of the maximum raw difference map variation is the most conservative method. However, this threshold is too high, since almost no differences are left in the significant difference map (Fig. 22b). Employing this high threshold has the advantage that non-significant differences will never be classified as significant. Using the mean of all raw difference map variations indicates many more significant changes (Fig. 22c), and these occur in all protein regions. This suggests that the whole CitS protein rearranges when it changes conformation, which is biochemically implausible and does not agree with [Kebbel *et al.* 2013]. Thus, the mean of all variations is not a sufficiently strict global threshold. Vinothkumar et al. [Vinothkumar *et al.* 2005] used the classic method to determine significant differences. They set the contour level to the RMSD when plotting the difference map and assumed that every difference above two contour levels was significant. Similarly, a threshold of twice the RMSD (2·RMSD) was used in the present work to obtain the significant difference maps shown in Fig. 22d. Comparing (a) to (c) within each figure, shows that using 2·RMSD as the threshold is not as conservative as using the maximum raw difference map variation, but stricter than using the mean of all raw difference map variations. The strongest differences in the raw maps (Fig. 22a) are classified as significant, but most of the differences at the dimerization interface are discarded.

In contrast to the above, a variation map (Fig. 22e) yields a local threshold for every pixel in the raw difference map, allowing regional variation changes to be accounted for when the significant difference map (Fig. 22f) is determined. A variation map is generated by comparing the raw difference maps obtained for split data sets as outlined in sections 5.2.1.3 and 5.2.1.4. The boxes in Fig. 22 mark the position of three specific regions of the variation maps. Comparing the signals within them clearly demonstrates how important it is to consider local fluctuations when classifying differences. If this is not done, peaks might be incorrectly classified as significant (a - d of Fig. 22, left box) or be incorrectly weighted, appearing to be stronger/weaker than they actually are (a - d of Fig. 22, center box) or even to have the opposite sign (a - d of Fig. 22, right box). Although the result of local thresholding (Fig. 22f) is similar to that obtained using 2·RMSD as a global threshold, there are small but important disparities. Further, the mixed-merge method classifies slightly more differences as significant, and these are in the same regions as those indicated by the classic method (compare Fig. 23b and c). The trend was the same when a global threshold was employed to analyze synthetic data (Fig. 21).

5.4.3. Conformational changes in CitS

Both the classic and the mixed-merge methods indicate that significant conformational changes occur in the distal helix cluster of both CitS monomers (Fig. 23 and [Kebbel *et al.* 2013] where the mixed-merge method was also applied). In contrast the maps indicate only minor differences in protein regions towards the center of the crystal unit cell, implying that these domains are not involved in substrate binding but form the dimerization interface.

5.4.4. Conformational changes in the potassium channel MloK1

Electron crystallography data sets recorded from 2D crystals of MloK1 in the presence and absence of cyclic AMP (cAMP) [Kowal *et al.* 2014] were examined by the mixed-merge and classic methods using local thresholds. This second method comparison gave further insights into the conformational changes that occur on ligand binding [Kowal *et al.* 2014]. MloK1 is a tetramer and the 2D crystals are members of the $P42_12$ plane group. Like other hyperpolarization-activated and cyclic nucleotide-modulated channel (HCN) proteins, the monomer has six transmembrane helices. The first four (S1-S4) form a voltage sensor domain (VSD) and the remaining two (S5, S6) form part of the single ion-conduction pore at the center of the tetramer. There is a cyclic nucleotide binding domain (CNBD) at the C-terminus of each MloK1 monomer.

From the 3D structures of both MloK1 conformations, the intracellular CNBDs move vertically towards the membrane when cAMP binds [Kowal *et al.* 2014]. Although this vertical movement cannot be shown in a 2D projection along the z-axis the resulting changes in the density distribution are evident and can be documented by difference maps. The raw difference map obtained by subtracting the structure of MloK1 without cAMP from the structure with bound ligand is shown in Figure 24 together with the significant difference maps obtained by the classic and mixed-merge methods using local thresholds. In agreement with Kowal et al [Kowal *et al.* 2014], all maps indicate a clear counterclockwise rotation of the whole tetramer on ligand binding, most of the movement occurring in the peripheral monomer region where the CNBDs and the VSD domains are situated. Comparison (Fig. 24b and c) shows that the mixed-merge method classifies a few more differences as significant than the classical method. The strong differences at the protein interface in all maps (Fig. 24) shows that Mlok1 interacts with adjacent tetramers via the putative VSDs. There are no significant differences in the pore region. This confirms the theory of [Kowal *et al.* 2014], that the pore diameter does not drastically change, and therefore the MloK1 channels gate at the selectivity filter and not at the helix bundle crossing.

5.5. Conclusion

Comparison of existing and newly developed methods to determine significant changes in projection maps of 2D crystals, showed that methods not attached to an assumed probability distribution model perform better i.e., the classic and mixed-merge methods. The use of synthesized 2D crystal data sets for the comparison meant that the ground truth was known and proved that the new mixed-merge method determined the significant changes in difference map most reliably. In contrast to the classical method it considers the background variations across all data sets employed i.e., across both conformations as well as within the single data sets. As demonstrated using experimental data, the mixed-merged method is also the most robust when varying noise levels are accounted for by the use of local thresholds rather than a single global threshold. Local thresholds are more adaptive and sensitive to data heterogeneity.

Both non-statistical methods were used with local thresholds on real data: 1) CitS: a secondary citrate/sodium symporter CitS with the substrate Na citrate and Na acetate. 2) Mlok1: A cyclic nucleotide-modulated potassium channel with and without the ligand cAMP. For both proteins, significant conformational changes were shown. Therefore these methods help to better understand the protein function.

All methods are implemented and available in 2dx [Gipson *et al.* 2007a, Gipson *et al.* 2007b] version 3.3.0.

(a) Na Citrate   (b) Na Acetate   (c) Ground Truth

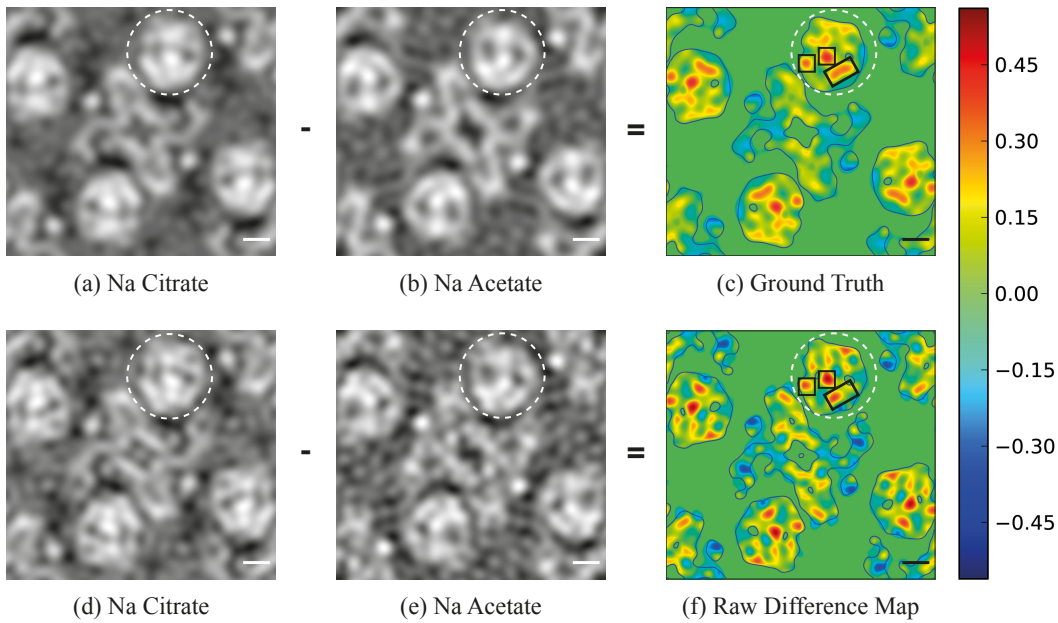(d) Na Citrate   (e) Na Acetate   (f) Raw Difference Map

Figure 20: **Raw Difference Map:** Raw difference maps calculated by simple sub-
traction of the image intensities; The difference was calculated after the
foregoing steps listed in Section 5.2 had been carried out on the synthe-
sized data sets. (a) Conformation 1: from synthetic data based on the
merged projection map of CitS in the presence of Na citrate (b) Con-
formation 2: from synthetic data based on the merged projection map
of CitS in the presence of Na acetate. (c) The difference map obtained
by subtracting (b) from (a); since no noise was added to these data sets
employed this difference map is the ground truth. (d) and (e) represent
Conformations 1 and 2, respectively where Gaussian noise with a vari-
ance of 70 had been added to both synthetic crystal data sets. (f) The
raw difference map obtained by subtracting (e) from (d); therefore noise
is present and the significance of the differences has not been determined.

(a) Ground Truth (c) Student's t-Test (e) Classic

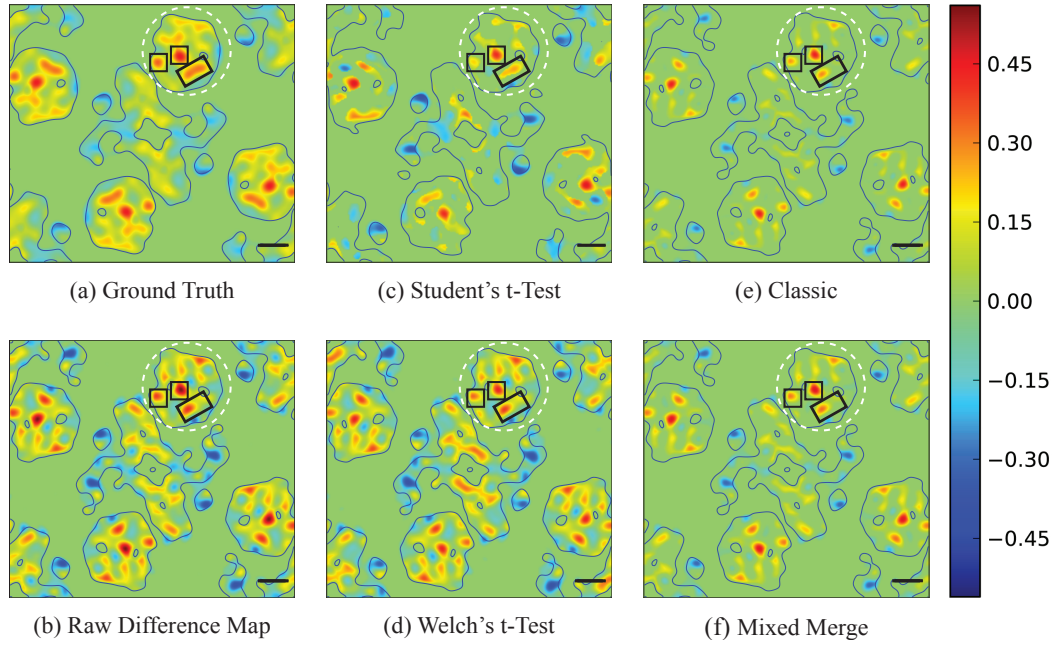(b) Raw Difference Map (d) Welch's t-Test (f) Mixed Merge

Figure 21: **Significant Difference Maps:** Comparison of the four methods used to determine significant differences in the raw difference map calculated from the merged projection maps generated after Gaussian noise with a variance of 70 had been added to the synthetic crystal data sets . (a) The ground truth, as shown in Fig. 20c. (b) The raw difference map, as shown in Fig. 20f. (c-f) Significance difference maps obtained by (c) the Students t-test (real domain), (d) the new Welch's t-test method (Fourier domain). (e) the classic method (real domain), (f) the new mixed-merge method (real domain). Local variations are accounted for in the two statistical methods. A global threshold (see Section 5.2.1) was used in both the classic and the mixed-merge methods.

(a) Max
(b) Mean
(c) 2·RMSD

(d) Raw Difference Map
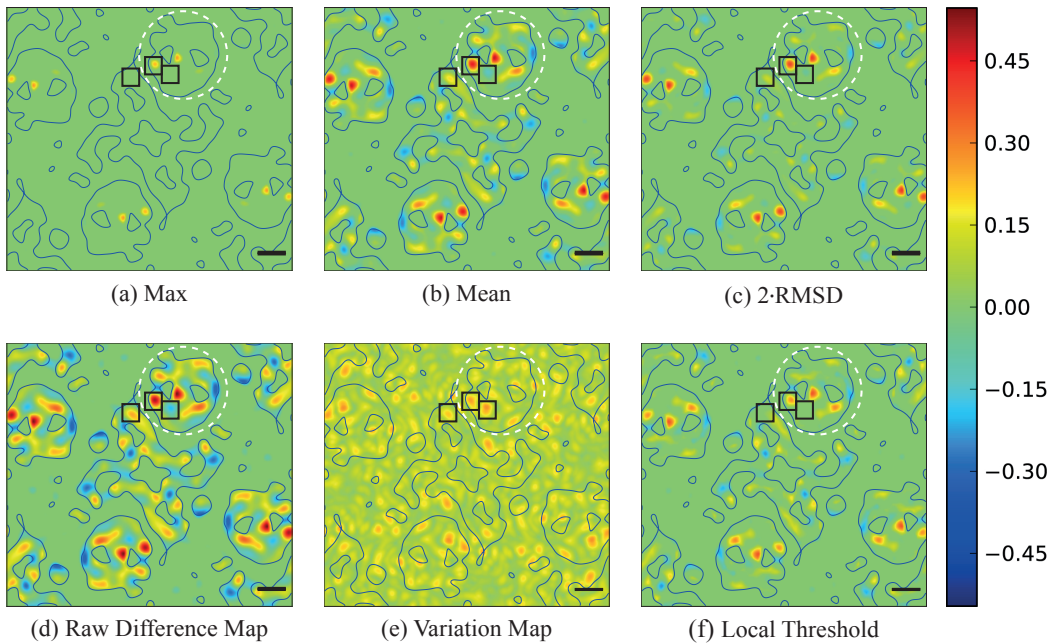(e) Variation Map
(f) Local Threshold

Figure 22: **Global vs. Local Thresholds:** Experimental data obtained for the secondary citrate/sodium symporter CitS [Kebbel *et al.* 2013] were employed for this test; CitS with Na citrate and Na acetate. The classic method to determine significant differences is to compute a global threshold from the split data sets as described in Section 5.2.1.3. This threshold can be defined in different ways. (a-c) The significant difference maps obtained using (a) the maximum, (b) the arithmetic mean, (c) or, as outlined in section 5.2.1.3, the RMSD of the empirically determined variations. Alternatively, as suggested in the present paper, significant differences in the raw difference map (d) can be detected using the local thresholds indicated by a variation map (e). The significant difference map (f) is simply obtained by subtracting the variation map (e) from the raw difference map (d).

(a) Raw Difference Map   (b) Classic   (c) Mixed Merge

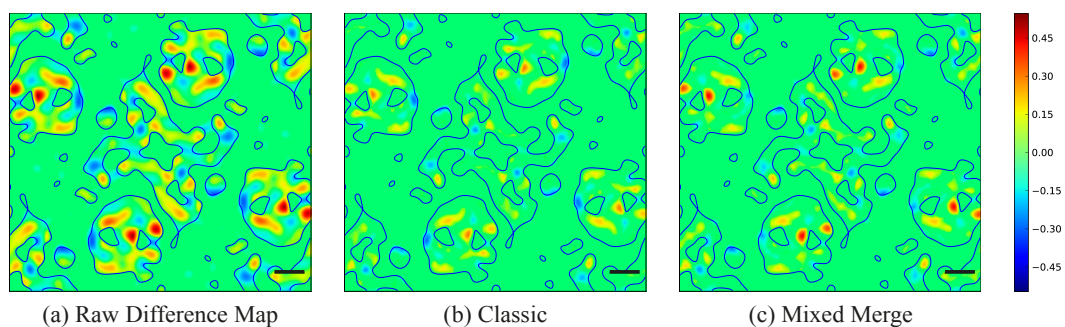Figure 23: **CitS Difference Maps:** The conformational change of secondary cit-
rate/sodium symporter CitS when bound to Na citrate and Na acetate
illustrated through difference maps.

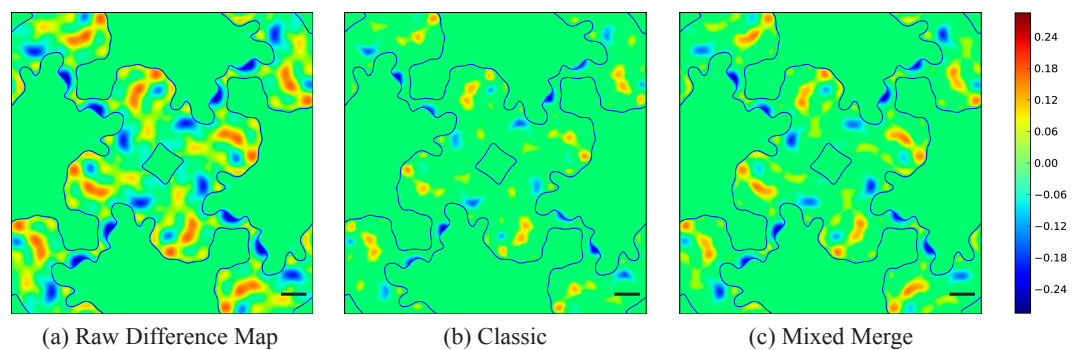(a) Raw Difference Map   (b) Classic   (c) Mixed Merge

Figure 24: **MloK1 Difference Maps:** The conformational change of the cyclic
nucleotide-modulated potassium channel MloK1 due to binding of the
ligand cAMP.
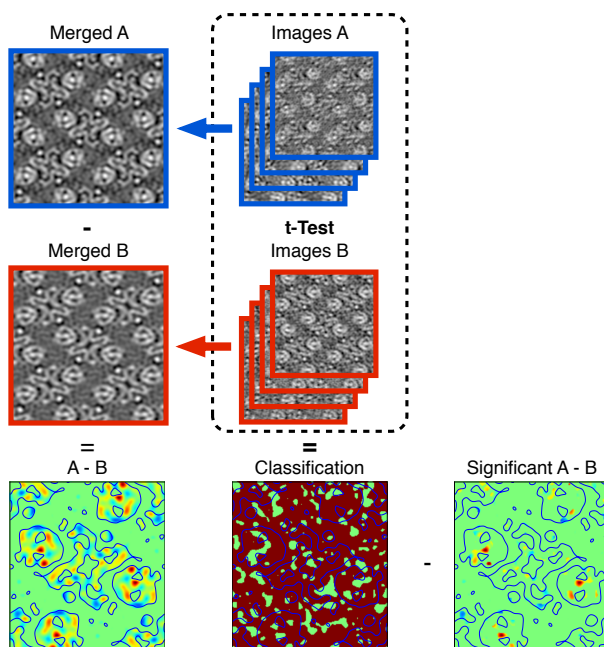
Figure 25: Student's t-Test: Significant differences are determined by comparing the mean pixel intensity of both conformations with the Student's t-Test.
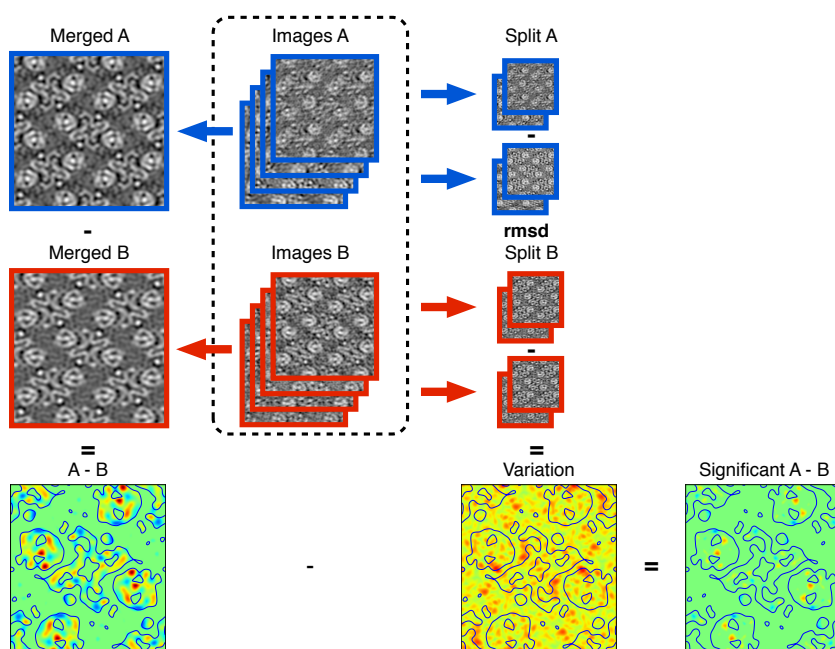
Figure 26: Classic: The classic method to determine data heterogeneity by ran-
domly splitting the conformation images in a set of two and calculating
the difference map. Traditionally the RMSD was calculated for both con-
formations and all differences above two times their mean was considered
significant. We took the maximum of both conformations at each position
and used it as a local threshold.

Figure 27: The relation of the overall phase residual during merging to the noise level
applied to the synthetic crystals of CitS with Na-Acetate and Na-Citrate.

6. Single Particle 3D Reconstruction for 2D Crystal Images

The content of this chapter has been published in [Scherer *et al.* 2014]. My contribution to this manuscript was the analysis and reimplementation of the original work by Xiangyan Zeng along with some minor additions. The main work was done by Sebastian Scherer.

Abstract

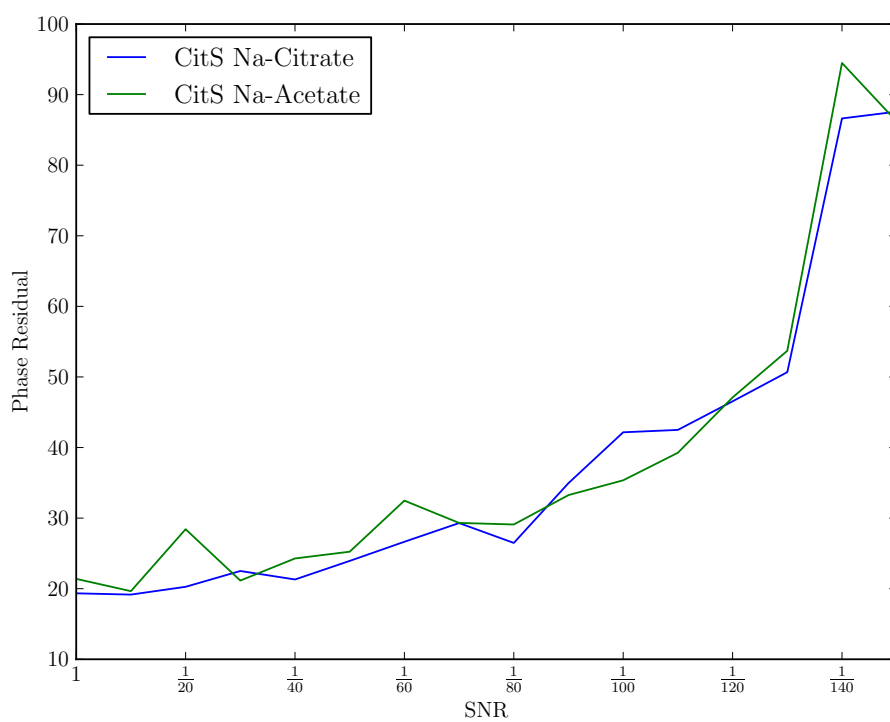I n cases where ultra-flat cryo-preparations of well-ordered two-dimensional (2D) crystals are available, electron crystallography is a powerful method for the determination of the high-resolution structures of membrane and soluble proteins. However, crystal unbending and Fourier-filtering methods in 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.

6.1. 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 2012], 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.

6.2. 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 28. 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. Our new processing approach (Figure 29) is a refinement of the conventional crystallographic image processing pipeline described by Arheit et al. [Arheit *et al.* 2013c, Arheit *et al.* 2013a] (Figure 29A), and is implemented as an option in the open source 2dx software.

The initial stages of the classical approach (Figure 29A) include determination of image defocus and astigmatism 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 are merge into one 3D reconstruction (MRC programs *Origtilt*, *Latline*, and others). Our new refinement procedure (Figure 29B 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 29B). Subsequently the 3D orientation of each protein of each crystal is iteratively optimized using our new local refinement tool (Figure 29C), taking the correlation between neighboring particles into account.
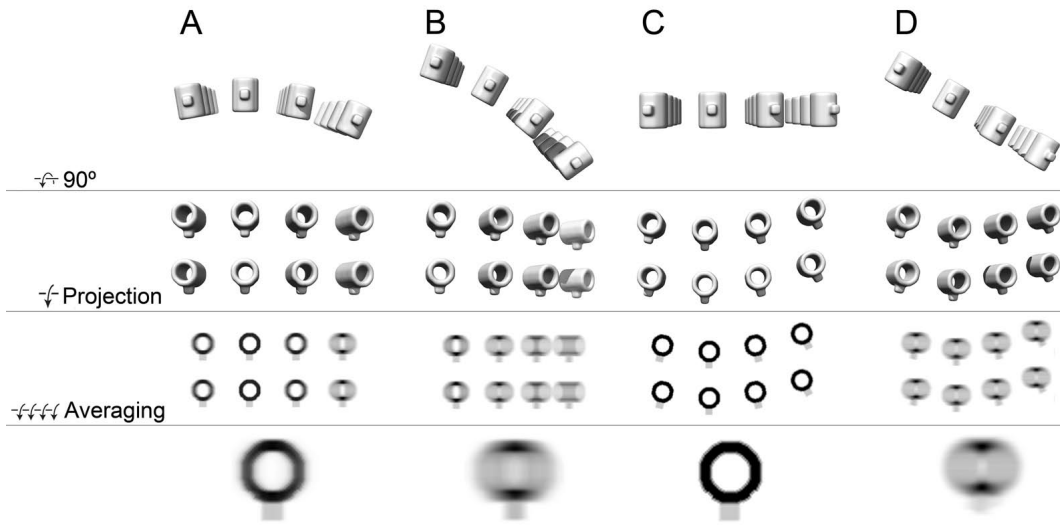
Figure 28: Effect of sample flatness and local disorder on resolution in crystallographic processing. Artificial crystals composed of multiple copies of a tube with a protrusion are considered. From top to bottom: Perspective side-view of the crystal; top-view of the crystal (only two protein rows shown 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.
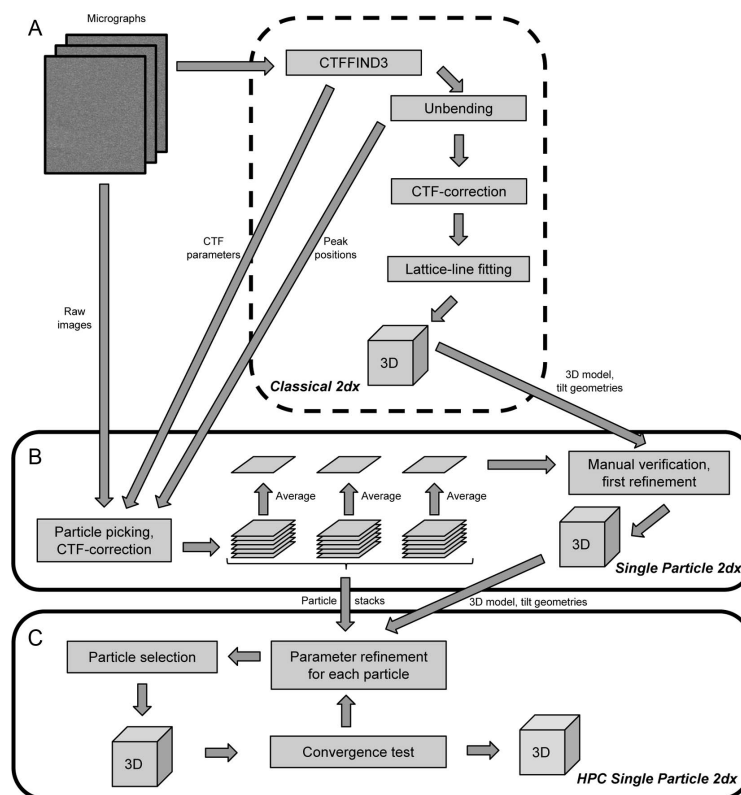
Figure 29: Single particle refinement for 2D crystals. The presented approach refines classically processed cryo-electron crystallographic datasets. (A) All micrographs are processed and merged classically with 2dx [Gipson *et al.* 2007b]. The unit-cell positions determined while unbending the crystals, 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.

6.2.1. 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 (Fig. 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 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.

6.2.2. 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]. A detailed explanation of the local CTF correction can be found in the supplementary material 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 28A and B). However defocus values are not refined in the current implementation. Following [Zhang & Zhou 2011] the resolution limit imposed by a defocus determination that
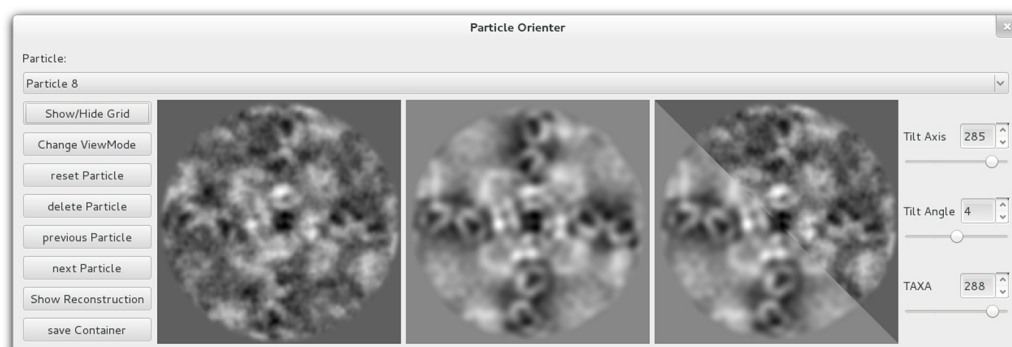
Figure 30: Interactive graphical user interface to verify and refine the particle orientation. The goal is to find the three Euler angles that lead to the projection of the 3D reconstruction (central window) that matches best the per-crystal average (left window). Changing the sliders on the right immediately 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.

is inaccurate by 124 nm at 300kV acceleration voltage is at 7Å. This resolution limit is beyond the resolution of the test dataset used in the results section. 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 $1\mu m$ 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.

6.2.3. 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 30) 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 30) enables interactive refinement of the initial orientations 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. 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 $\pm 5°$.

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.

6.2.4. 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:

$$(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s}_x, \hat{s}_y)_i = \operatorname*{argmax}_{(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s}_x, \hat{s}_y) \in \mathcal{T}_i} \{\operatorname{sim}(P_i, \mathcal{P}(A, \theta, \psi, \phi))\} \tag{4}$$

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, ϕ 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. In practice, the optimization of Equation 6.2.4 is carried out over a finite set of trial projections $\mathcal{T}$ 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 6.2.4.

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 free single particle refinement-based version of Equation 6.2.4 to individual particles picked from cryo-EM images of unstained 2D membrane protein crystals was not successful 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 28). 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 6.2.4) by the following expression:

$$(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s}_x, \hat{s}_y)_i = \underset{(\hat{\theta}, \hat{\psi}, \hat{\phi}, \hat{s}_x, \hat{s}_y) \in \mathcal{T}_i}{\mathrm{argmax}} \{\mathrm{sim}(P_i, + \sum_{j \in \langle \mathcal{N}_i \rangle} \lambda_j P_j, \mathcal{P}(A, \theta, \psi, \phi))\} \qquad (5)$$

where the similarity measure, $\mathrm{sim}(\cdot, \cdot)$, is the same as in Equation 6.2.4. 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 $\langle \mathcal{N}_i \rangle$ 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:

$$\lambda_j = \exp\left(-\frac{1}{2}\left(\frac{R_j}{\sigma}\right)^2\right) \qquad (6)$$

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.

6.2.5. 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:

$$\vec{x}_i = (\theta_i, \psi_i, \phi_i, (s_x)_i, (s_y)_i, cc_i)^T \tag{7}$$

consisting of the three Euler angles θ_i, ψ_i, ϕ_i, the two shift parameters $s_x)_i, s_y$ and the corresponding highest cross-correlation value cc_i determined during the refinement process. As suggested by supplementary Figure 35 the components of the particle fingerprint vectors follow a 6-fold multivariate normal distribution. Based on this assumption the log-likelihood function $\log f(\vec{x}_i | \mu_{\mathbb{C}(i)}, \sum_{\mathbb{C}(i)})$, which describes the likelihood (or "probability") of finding a single observation $\vec{x}_i$ given the estimated mean vector $\mu_{\mathbb{C}(i)}$ and covariance matrix $\sum_{\mathbb{C}(i)}$ (with $\sum_{mn} = \langle (x_m - \mu_m)(x_n - \mu_n) \rangle$), is given by [Anderson & Olkin 1985]:

$$\log f(\vec{x}_i | \mu_{\mathbb{C}(i)}, \sum_{\mathbb{C}(i)}) = -\frac{1}{2}\log(|\sum_{\mathbb{C}(i)}|) - \frac{1}{2}(\vec{x}_i - \mu_{\mathbb{C}(i)})^T \sum_{\mathbb{C}(i)}^{-1}(\vec{x}_i - \mu_{\mathbb{C}(i)}) \tag{8}$$

Here, $\mathbb{C}(i)$ indicates that the mean vector and the covariance matrix depend on the crystal from which the particle P_i was originally picked. $|\sum_{\mathbb{C}(i)}|$ 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.

6.3. General implementation details

6.3.1. 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 36). 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 30) 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.

6.3.2. 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.* 2013] 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 (~3000 particles/crystal) 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 6.2.1 to Section 6.2.3) 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 6.2.4. The parallelization approach is visualized in Fig. 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 6.2.4). After disabling inconsistent particles based on the procedure detailed in Section 6.2.5 all particles stored on a compute node are back-projected 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.

6.3.3. GPGPU accelerated projection method

The performance bottleneck of the initial model refinement (Section 6.2.3) 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
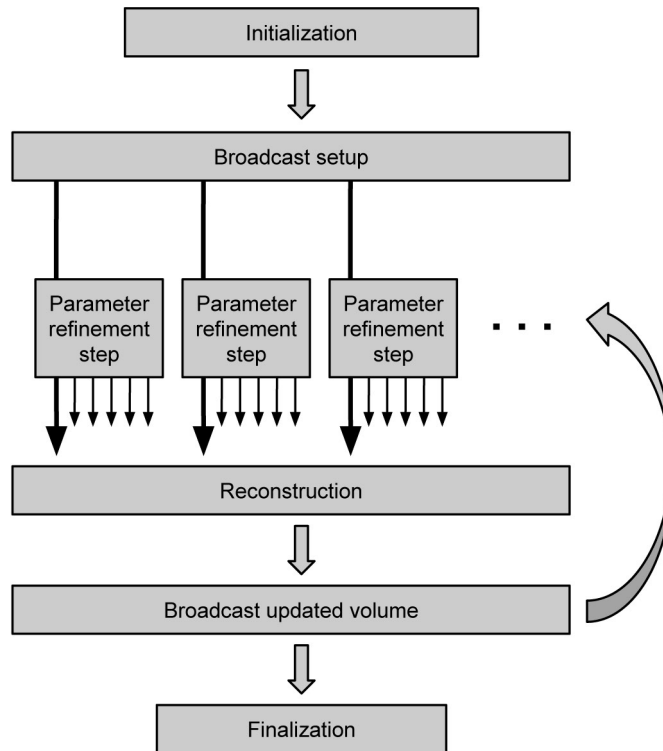
Figure 31: Hybrid parallelization approach. The application is launched on a large number of multi-core compute nodes connected by a high-bandwidth network. Each of the node runs only one MPI rank on one of its cores and keeps the rest of the cores initially idle. In order to setup the refinement, one 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.

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 [Castano-Diez *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.

6.4. Results and discussion

6.4.1. 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 CM200FEG cryo-transmission electron microscope, operated at an acceleration voltage of 200kV and a nominal magnification of 50'000x, 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 $5\mu m$, resulting in a pixel size of 1Å at the specimen level. All crystal images were processed and merged into one 3D structure (Figure 32A, mesh, red) with `2dx`, yielding a final resolution at 7Å resolution in xy-direction and 12Å orthogonal to the imaging plane [Kowal *et al.* 2014]. These images will be made available in the EMDB database.

6.4.2. Experimental procedure: Single particle refinement of an 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 orienta-
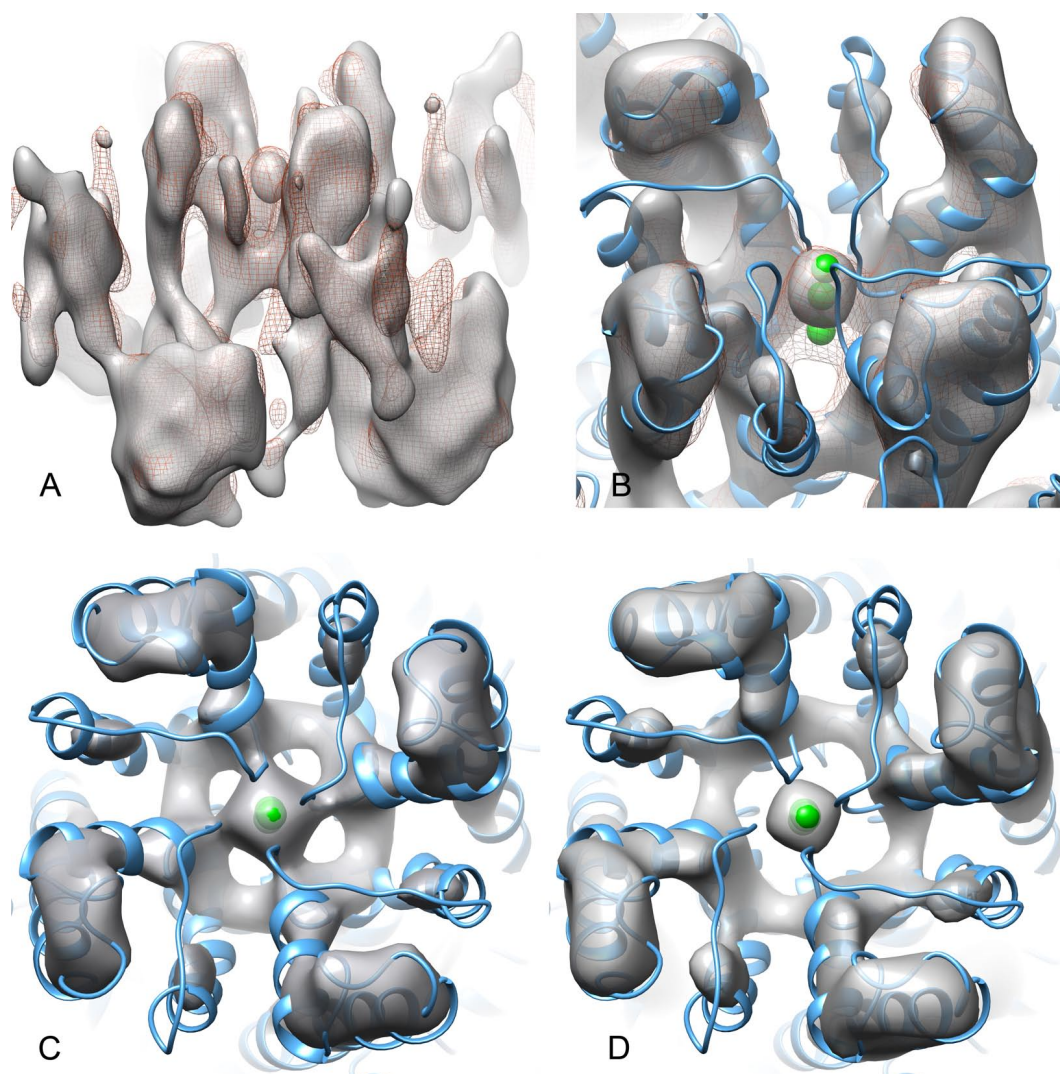
Figure 32: Results obtained with the 3D local tilt geometry refinement procedure on a cryo-EM data set. (A) Reconstruction obtained by classical unbending processing within `2dx` (mesh, red) compared to the locally refined structure obtained using our new algorithm (solid, grey). (B) Closer look at the 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,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 the per-crystal averages was manually refined against the initial 3D structure by means of the new interactive particle alignment tool (Figure 30).

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 6.2.1). 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 6.2.3 was executed within a $\pm 5°$ 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 position 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 6.2.4 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. Another round of iterative initial model refinement was then 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 $\pm 12°$ 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 $\pm 3°$ range in $1°$ steps. For this initial coarse refinement, information in the resolution range from 12-75Å 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 $\pm 3°$ cone with $1°$ separated trial directions. The in-plane rotation was optimized over $\pm 1.5°$ 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 com-
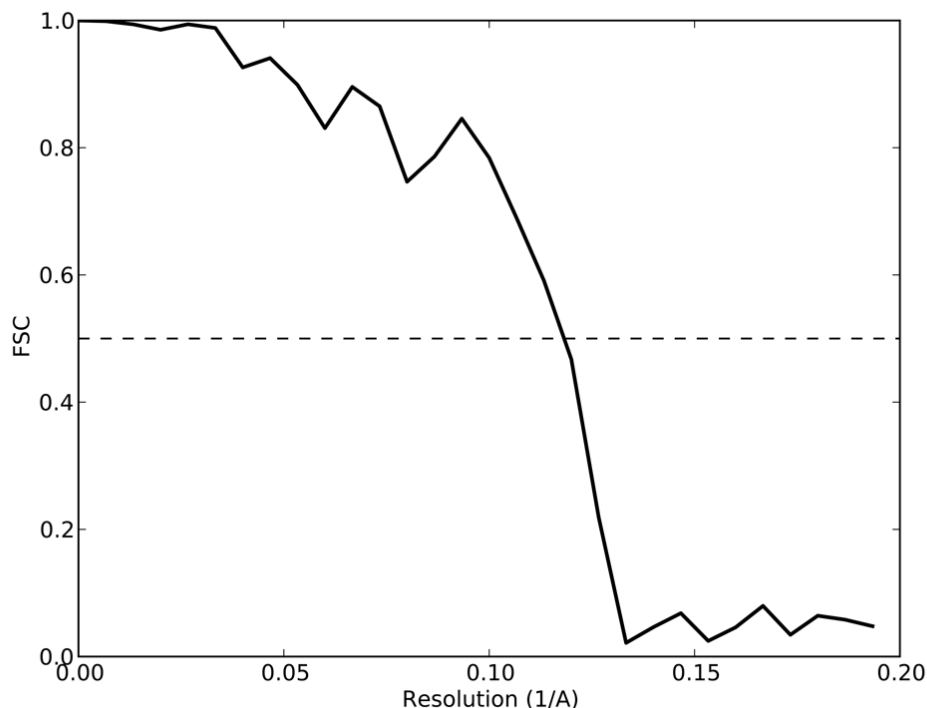
Figure 33: Fourier shell correlation of the final single particle based refined 3D structure indicating a resolution of 8.3Å (0.5 FSC-criterion, pixel size 1Å).

bine 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 2.5) 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 show in Figure 32A. The probability 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 36A-E), suggesting that our method resolves the true variations in membrane flatness of the 2D crystals in 3D.

6.4.3. Resolution measure, prevention of overfitting

From the Fourier shell correlation (FSC) shown in Figure 33 we conclude (0.5 criterion) that the final reconstruction obtained by means of the presented algorithm is 8.3Å. 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 2010]. 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 32).

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 foreword 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.

6.4.4. 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 (Fig. 5). From Figure 32A, where the classically MRC-based merged structure (mesh, red) 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 33B-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 33C 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.

6.4.5. Performance evaluation of the likelihood-based particle selection approach

Figure 34 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 34 shows regions of strong inconsistency (Figure 34A; black ellipse). Based on the procedure described in Section 6.2.5 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 Fig-
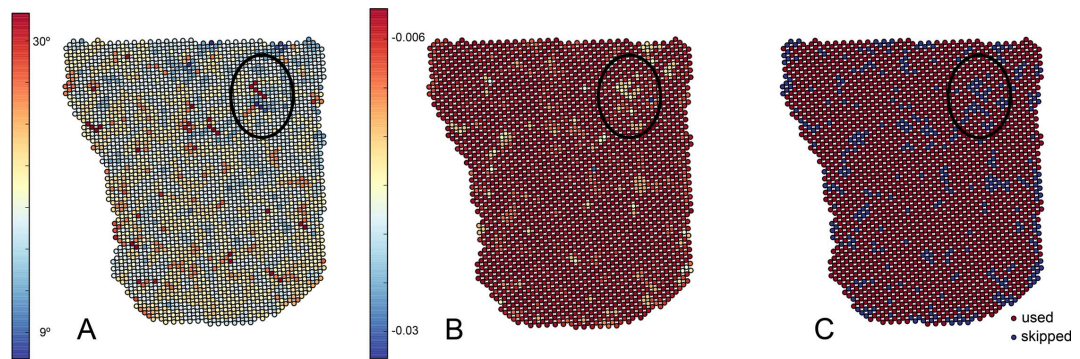
Figure 34: Likelihood-based particle selection within one 2D crystal. (A) Tilt angle distribution within the crystal. The area marked with a black ellipse shows inconsistently aligned particles as the tilt geometry varies more than 10 degrees within the distance of two particles. (B) The likelihood values of the particles. Based on the particle finger print vector (Equation 6.2.5), 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.

ure 34B. The 5-15% of the particles with the lowest consistency values were not used for the subsequent reconstruction (Figure 34C), which led to a further improvement of the reconstruction.

6.4.6. Computing performance of the refinement program

The processing presented here was done on a desktop machine featuring two Intel Xeon E5-2600 3.0 GHz 8-core CPUs, 64Gb memory and two Nvidia GeForce GTX Titan with 2688 CUDA cores and 6Gb DDR5 RAM each. Solid-state hard drives were used for faster writing and loading of the particle stacks.

As mentioned in Section 6.3.3 the bottleneck of the initial model refinement is the calculation of the projection gallery. Table 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 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 1). The calculation of one single projection in the "full parallel GPU-setup" took 0.14ms on average.

One iteration step of the first phase (coarse refinement level, Section 6.4.2) of the local tilt geometry refinement of the 210'538 particles with a box size of $150 \in 150$ pixels took about 100 minutes 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 minutes 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 37).

6.5. 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 2012]. 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.

6.6. Acknowledgments

6.7. Supplementary Material

Figure 35: 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, (F) maximal cross-correlation value
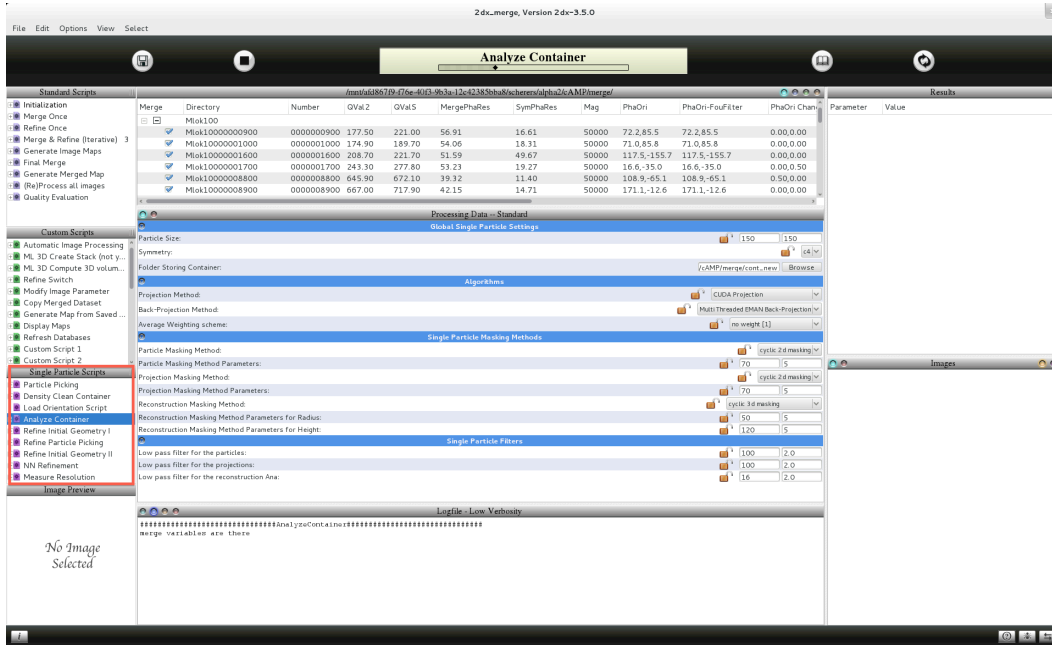
Figure 36: Updated `2dx_merge` graphical user interface with the new single particle script section (red box in the left column).
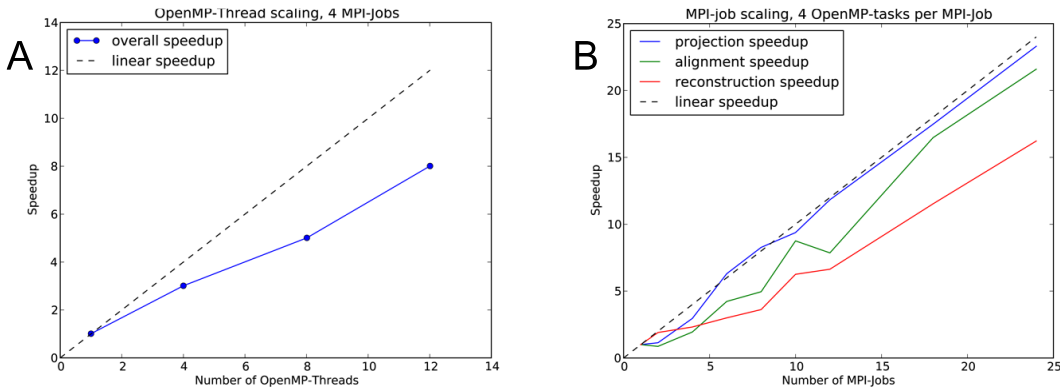
Figure 37: Scaling analysis of the cluster implementation. (A) Number of shared memory tasks plotted against the observed speedup. Measurement of the one refinement step of 60'000 particles on four AMD Opteron 6174 12-core nodes. (B) Distributed memory parallelization scaling measured on AMD Opteron 8380 4-core processors.

Part II.
Applications

7. Structure and Substrate-Induced Conformational Changes of the Secondary Citrate/Sodium Symporter CitS

The following section has been published in [Kebbel *et al.* 2013]. My contribution in this study was in assisting the image processing and implementing the significant difference map method as presented in section 5.

Abstract

The secondary Na+/citrate symporter CitS of *Klebsiella pneumoniae* is the best-characterized member of the 2-hydroxycarboxylate transporter family. The recent projection structure gave insight into its overall structural organization. Here, we present the three-dimensional map of dimeric CitS obtained with electron crystallography. Each monomer has 13 a-helical transmembrane segments; six are organized in a distal helix cluster and seven in the central dimer interface domain. Based on structural analyses and comparison to VcINDY, we propose a molecular model for CitS, assign the helices, and demonstrate the internal structural symmetry. We also present projections of CitS in several conformational states induced by the presence and absence of sodium and citrate as substrates. Citrate binding induces a defined movement of a helices within the distal helical cluster. Based on this, we propose a substrate translocation site and conformational changes that are in agreement with the transport model of "alternating access".

7.1. Introduction

Two available classification systems group secondary transport proteins according to their functionality and sequence homology (TC classification [Saier 2000]) or their hydropathy profiles (ST[1-4], MemGen classification [Lolkema & Slotboom 1998, Lolkema & Slotboom 2003, ter Horst & Lolkema 2012]). Both systems underline the enormous phylogenetic, functional and structural diversity among secondary transporters. The number of available high-resolution 3D structures for these proteins is growing rapidly, providing unexpected structural insights. So far, all structures reveal 4-14 α-helical transmembrane segments (TMS), the majority of protein monomers being comprised of 11-13 [Tsai & Ziegler 2010]. Surprisingly, numerous unrelated secondary transporters seem to share common global structural folds, e.g., the fold of LeuT [Forrest & Rudnick 2009, Yamashita *et al.* 2005] and the fold of the major facilitator superfamily (MFS) [Sun *et al.* 2012]. Most secondary transporters occur as dimers or trimers [Forrest *et al.* 2011] and most of

their structures reveal internal structural symmetry within the single monomers based on (inverted) repeats of a defined number of helices. More importantly, the growing number of atomic structures within a common fold provides an unprecedented insight into the molecular mechanism of secondary active transport [Abramson & Wright 2009, Forrest *et al.* 2011, Krishnamurthy *et al.* 2009]. In the original model proposing "alternating access", the secondary transporter alternately exposes its substrate binding sites to both sides of the membrane, which facilitates a unique framework for substrate translocation [Jardetzky 1966]. This model has been refined and extended by numerous crystallographic and biochemical breakthroughs [Forrest *et al.* 2011]. According to current knowledge, secondary transporters cycle through defined structural states. Initial substrate binding, e.g., to the empty "outward open" transporter, induces the closure of outer molecular gates, resulting in the closed "occluded" conformation as a transition state. A further conformational switch opens the inner gates, leading to the "inward open" state, where the substrate(s) can be released. The free energy barriers of these sometimes substantial conformational changes are overcome by utilizing the binding energy of both substrates (main- and co-substrate) to the transporter [Forrest *et al.* 2011]. Different crystallographically-captured conformational states have allowed the transport cycle to be studied and visualized in detail, leading to three mechanistic models referred to as "rocker-switch", "gating" and "rocking bundle" [Law *et al.* 2007, Reyes *et al.* 2009]. All of these models accentuate the necessity of internal structural symmetry and each is in good agreement with the original "alternating access" model.

The secondary citrate/Na^+ symporter CitS of extit*Klebsiella pneumoniae* is the best characterized member of the 2-hydroxycarboxylate transporters (2-HCTs), a subclass of bacterial transport proteins within ST[3] of the MemGen system for which three-dimensional (3D) structural information is still not available. CitS is postulated to couple the import of two Na^+ ions and one bivalent citrate ion for anaerobic metabolism [Lolkema *et al.* 1994, Pos & Dimroth 1996]. Models predict an inverted topology of 2×5 helices organized in two domains plus one N-terminal helix [Krupnik *et al.* 2011, Lolkema 2006, Lolkema *et al.* 2005, Sobczak & Lolkema 2005c, van Geest & Lolkema 2000]. In confirmation, the two-dimensional (2D) projection structure of CitS at 6 Åresolution published recently [Kebbel *et al.* 2012] reveals 11 α-helical TMS organized in a distal helix cluster and a central dimerization interface. CitS thereby closely resembles the Na^+/H^+ antiporters NhaA and NhaP1 [Appel *et al.* 2009a, Goswami *et al.* 2011]. A detailed insight into the structural and functional properties of a member of the ST[3] family, although not of a 2-HCT, was delivered by the recent crystal structure of the divalent anion/Na^+ symporter (DASS) VcINDY [Mancusso *et al.* 2012, ter Horst & Lolkema 2012].

Here, we present the three-dimensional (3D) map of the 2-HCT CitS at a resolution of 6 Åobtained by electron crystallography of 2D crystals. Each monomer of the dimer reveals 13 rod-shaped densities, representing 11 single α-helices and two

putative helical reentrant loops. Resemblance to VcINDY enables us to refine our model with respect to the membrane orientation and dimer interface. Helices are assigned and the internal structural symmetry of the N- and C-terminal domain is documented. In addition, projection structures of CitS in different substrate combinations indicate a rearrangement of α-helices within the distal helix clusters after citrate exposure, particularly in the presence of Na^+. The data highlight the co-dependence of these two substrates, support our 3D model and demonstrate that the substrate binding site is part of the distal helix cluster. The observed helix movements are in agreement with those expected for molecular gates.

7.2. Results & Discussion

7.2.1. Electron crystallography

2D crystals of CitS were grown as described previously [Kebbel *et al.* 2012]; use of a temperature controlled dialysis machine with an optimized sodium acetate buffer allowed the size and quality of the crystals to be improved. Tubular/vesicular 2D crystals with diameters up to 600 nm were obtained (Figure S1). Calculated Fourier transforms of cryo-electron microscopy (cryo-EM) images of the flattened tubes usually showed two lattices resulting from the two crystalline layers. On image processing in 2dx [Gipson *et al.* 2007b, Gipson *et al.* 2007a], these lattices were treated as two independent datasets. The unit cell parameters ($96.0\,\text{Å} \times 106.0\,\text{Å}$ with an angle of $90.0°$) and the $p22_12_1$ plane group are the same as reported previously [Kebbel *et al.* 2012], and were unchanged for crystals that had been soaked in various substrates (Table 1). To extract 3D information, the sample was tilted up to $45°$ in the microscope for data collection. 3D merging of all 79 lattices enabled us to continuously sample amplitudes and phases along the lattice lines up to a vertical resolution of 15 Å(Table 1, Figure S2).

7.2.2. Three-dimensional map and structural model of CitS

Figure 38 shows the 3D map of CitS in the presence of Na^+ acetate at a resolution of 6 Åin the membrane plane and 15 Åin the vertical direction. As found in the earlier projection structure, viewed from the top dimeric CitS is oval measuring $52 \times 96\,\text{Å}$ (Figure 38A; [Kebbel *et al.* 2012]). In the z-dimension, the CitS dimer spans 40-60 Åthrough the lipid bilayer. The central part of the dimer is mostly buried in the membrane, while densities towards the end of the long axis extend 10-20 Åout of the bilayer. The resulting "M-shape" of the side-view (Figure 1B) confirms the previous low-resolution single particle structure [Moscicka *et al.* 2009]. Dimeric CitS has three cavities, i.e., one at the center of the dimer and one at the center of each
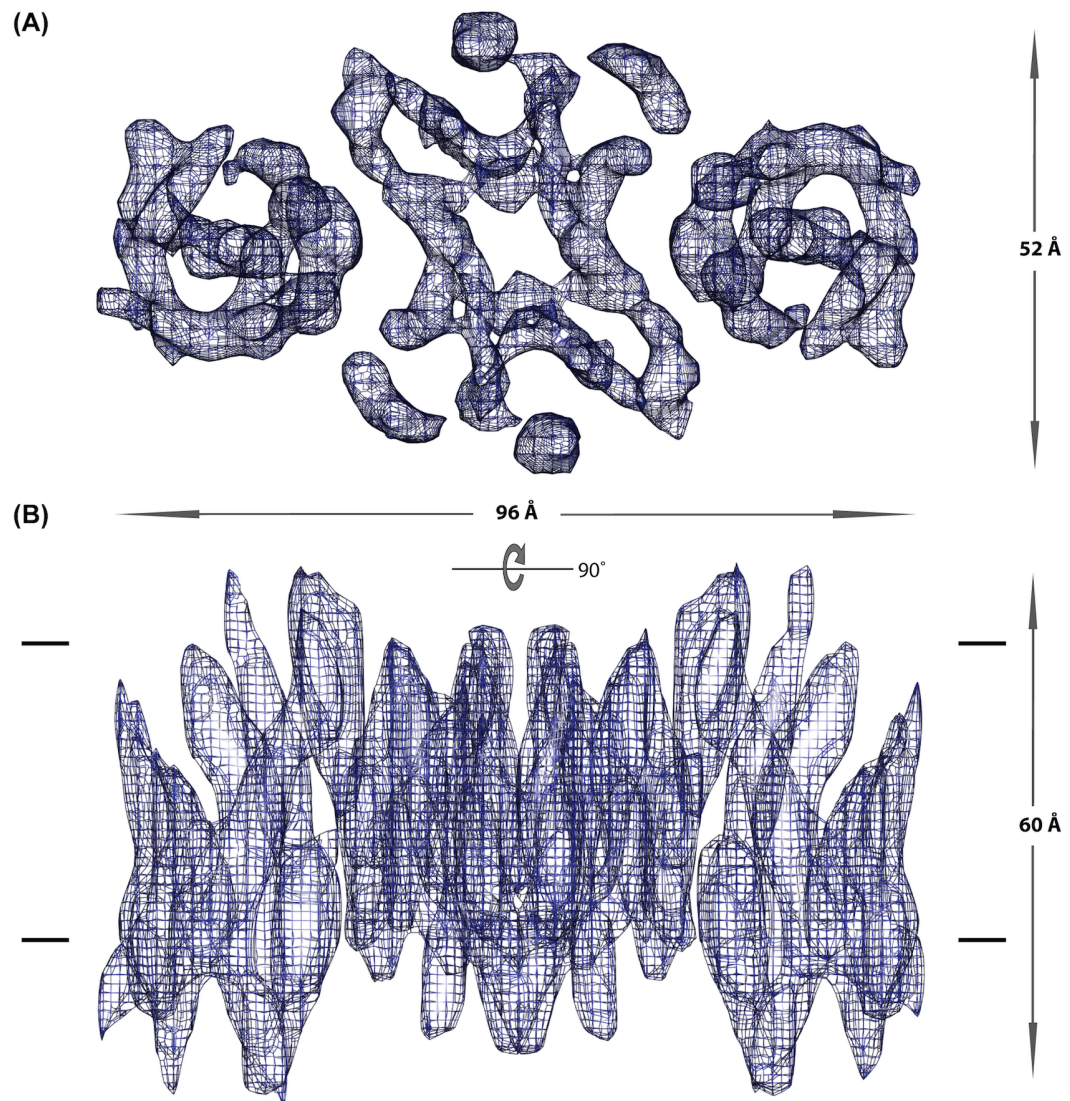
Figure 38: **Three-dimensional map of CitS.** The 3D map of CitS viewed from the top (A) and the side (B). One dimer measures 52×96 Å within the membrane plane, spans $40 - 60$ Å in the vertical direction and has a central dimer interface domain and two distal helix clusters.

monomer. These cavities are probably filled with lipids or water, respectively.
Manual placement of α-helical poly-A chains into the 3D map of CitS led to the
model shown in Figure 39A/B. The two identical monomers of the dimer are high-
lighted in red and blue. The assignment of helices to each monomer is based on their
location and proximity to other helices, and on the comparison to the structurally
related VcINDY (see below). Each CitS monomer has 13 α-helical TMS organized
in two domains. The domain involved in dimerization, the interface domain, is com-
prised of seven helices that are partially tilted or kinked up to 45° relative to the
membrane plane; contact between the two monomers is mainly provided by four
helices. The second characteristic domain of each monomer, the distal domain, is
formed by a dense cluster of six helical elements and is located at the distal tips of
the dimer. Besides one kinked helix, most of these elements are nearly perpendicular
in the membrane. Two central helices within the cluster are split into two shorter
parts.

Of the 13 α-helices of each monomeric CitS molecule, 11 are single membrane-
spanning helices while two are probably helical re-entrant loops. This extends and
refines the model based on our previous projection structure [Kebbel et al. 2012],
which exhibited 11 strong densities and four less dense regions. In the present
model (Figure 39A), two of the light regions within the helix bundle and the cen-
tral dimerization interface, respectively, appear to arise from single α-helical seg-
ments. Furthermore, the new model confirms that the monomer-monomer interface
is formed by the short axis of the dimer [Kebbel et al. 2012]. Due to the lower ver-
tical resolution of 15 Å, our dataset did not allow the visualization of amphipathic
surface helices; these are predicted for CitS and present in other secondary trans-
porters [Hunte et al. 2005, Mancusso et al. 2012, Sobczak & Lolkema 2005b].

Hydropathy profiles of 2-HCTs and DASSs predict CitS and VcINDY to share
a very similar 3D structure with 10+1 TMSs plus two α-helical reentrant loops
[Mancusso et al. 2012, ter Horst & Lolkema 2012]. However, the low sequence ho-
mology of 14 % (Figure S3) does not allow homology modeling. Comparison of
the CitS and VcINDY (pdb 4F35) structures reveals numerous common features
(Figure 39A-D). Both dimeric transporters have the same overall shape and archi-
tecture; the interface domain of each monomer contains seven partially tilted helices
and there is a second distal helical cluster. Further, in both cases, the interface
and distal domains of the monomers are separated by an aqueous basin (asterisks,
Figure 39A/C). Viewed from the side, the CitS and VcINDY dimers have a charac-
teristic M-shape, and the position of the dimerization interface is almost identical.
Assuming the same orientation of both proteins, CitS would protrude into the cyto-
plasmic space (Figure 39B). Another common salient feature is a vertically oriented
helix at both ends of the dimer's short axis (TMS1 on the VcINDY structure in
Figure 2C). In VcINDY, however, this is further away from the protein's main body.
Although the global structures of VcINDY and CitS dimers look similar, there are
significant differences in both the interface and distal helix clusters, and the indi-

vidual monomers superimpose poorly (not shown). In particular, there are major differences in the helix positions and orientations at the dimer interfaces (TMSs 1-4 and 7-9; Figure 39A-D). Separate superposition of corresponding helix clusters reveals major structural matches (Figure 39E), but there are still substantial differences in the helical architecture. In VcINDY the distal helix cluster is composed of four partially unwound TMSs (green) and four shorter helical reentrant loops, $HP_{in/out}$, each spanning half of the membrane (yellow; Figure 39F). This is also true for CitS (Figure 39G/Figure 40A), but rather than flanking the TMSs as in VcINDY, the four short helices are adjacent to each other and at the very center of the cluster (cyan). Nevertheless, their length and proximity allow us to speculate that these four short helical elements represent the reentrant loops Vb/Xa. The significantly different positions of Vb/Xa and $HP_{in/out}$ in CitS and VcINDY was to be expected since VcINDY's $HP_{in/out}$ are found between TMSs 4/5 and 9/10, CitS's are predicted to be between helices 5/6 and 10/11 [Mancusso *et al.* 2012, Dobrowolski *et al.* 2010]. Furthermore, in VcINDY helix 11 sits at the outer border of the helical bundle and does not contribute to substrate binding [Mancusso *et al.* 2012], while TMS11 of CitS is postulated to be directly involved in citrate binding via R428 [Sobczak & Lolkema 2005a]. Together, these differences underline the different molecular details of CitS and VcINDY although both proteins are found within the same subclass of ST[3]. As a consequence, further structural analysis is essential to reliably assign the helices of CitS (see below).

7.2.3. Molecular model and internal symmetry of CitS

Cross-linking studies on CitS showed that helices 5/6 and 10/11 plus the reentrant loops Vb/Xa form the translocation site [Dobrowolski *et al.* 2010]. The number of TMSs corresponds well to our model of the distal helical cluster, in which case the remaining seven TMSs 1-4 and 7-9 constitute the dimer interface as in VcINDY [Mancusso *et al.* 2012, ter Horst & Lolkema 2012]. Further biochemical studies [Krupnik *et al.* 2011] and the VcINDY structure [Mancusso *et al.* 2012] allow us to assign the perpendicular helix at the outer end of the dimerization interface as TMS 1. Additional consideration of inter-helical distances led to the detailed molecular model of CitS shown in Figure 40A. Helix 1 is depicted in yellow, TMSs belonging to the N-terminal domains in green (2-6) and TMSs belonging to the C-terminal in blue (7-11). Interestingly, helices of the C- and N-terminal domains intertwine much more than proposed in previous models [Krupnik *et al.* 2011, Kebbel *et al.* 2012]. In our new model, helix 11 and the helical reentrant loops Vb/Xa are adjacent to each other at the inner edge of the distal cluster. Thus, all known functionally important and highly conserved elements are positioned at the inner edge and center of the distal helical cluster, close to the aqueous basin. This includes Arg428 of TMS11 and the GGxG motifs at the tips of Vb/Xa, which are directly involved in substrate
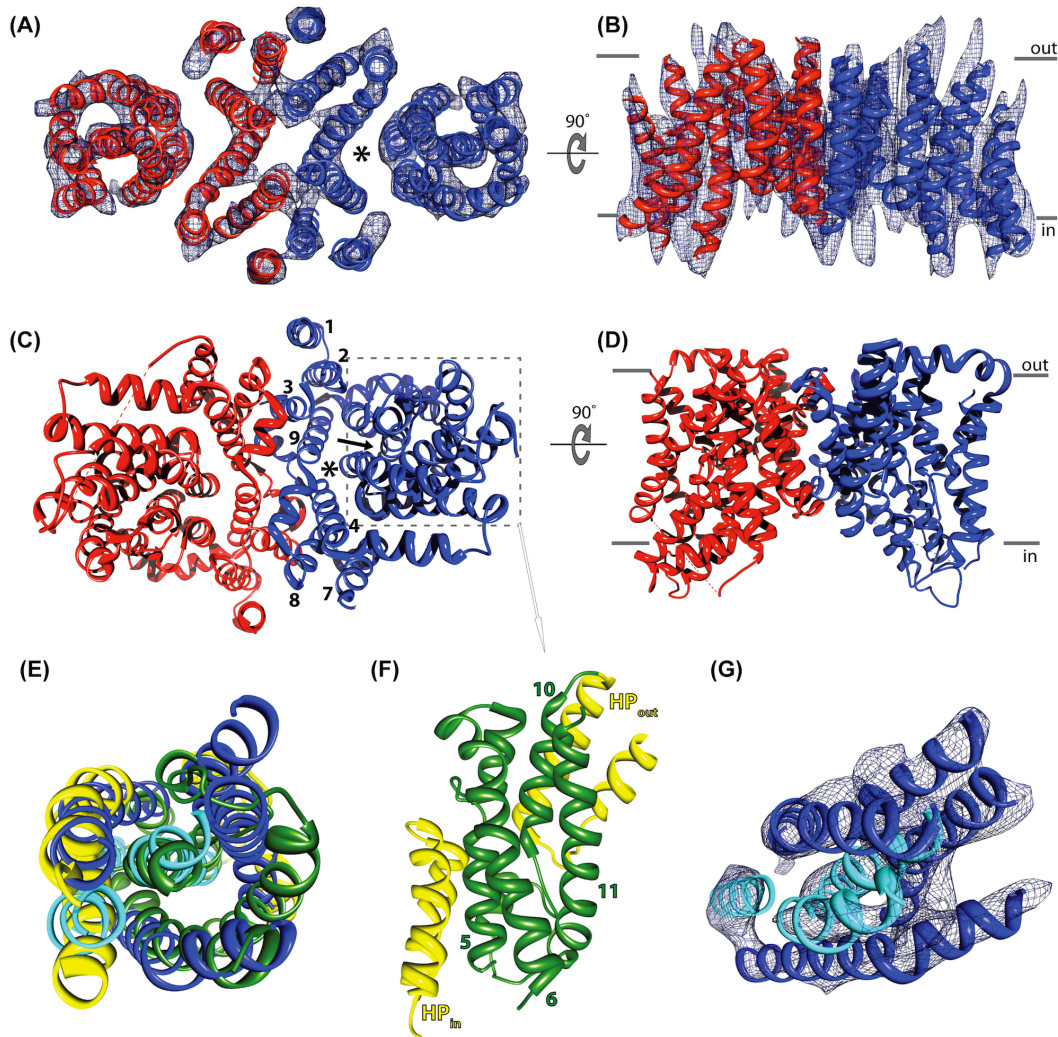
Figure 39: **Structural model of CitS and comparison to VcINDY.** Dimeric
CitS viewed from the cytosol (A) and the side (B). VcINDY (pdb 4F35)
viewed from the cytosol (C) and the side (D). Single monomers are colored
in red/blue. The two proteins have a similar global architecture. The
central dimerization domain and the distal helix clusters are separated
by an aqueous basin ($*$). The substrate binding sites in VcINDY ($\rightarrow$)
are in the center of a monomer. (E) Superposition of the helix clusters
from CitS (blue/cyan) and VcINDY viewed from the cytosol. (F) Distal
helix cluster of VcINDY. TMS 5/6/10/11 (green) and $HP_{in/out}$ (yellow)
are shown. The surface helices 4c/9c have been removed. (G) Distal
helix cluster of CitS. Broken TMSs (cyan) may represent helical reentrant
loops.

binding [Dobrowolski *et al.* 2010, Sobczak & Lolkema 2005a].

Analysis for the expected internal structural symmetry of each monomer validated the presented molecular model and the helix assignment. Rotating helices 1-4 by 180° along the dimer's long axis gives a good match to TMSs 7/8/9 (Figure 40B), but the putative symmetry-related elements of the distal helix cluster do not fit each other. Similarly, a superposition focusing on the distal cluster, with TMSs 5/6 and Vb corresponding to 10/11 and Xa (Figure 40C), leads to a bad fit within the interface domain. Alignment of the two domains individually emphasizes the symmetry observed for each (Figure 40D). Other helix assignments do not fulfill the symmetry correlations expected for a CitS dimer and none result in a symmetry mate for TMS1, which again supports our model.

The transport cycle of VcINDY is thought to be accomplished by a defined movement of the N- and C-terminal halves of the distal helix cluster relative to each other using the dimer interface as a static anchor point [Mancusso *et al.* 2012]. This would lead to the internal symmetry relationship being valid either for the central dimerization domain or for the helix cluster but not for both simultaneously, in agreement with our CitS model. Thus, the transport mechanisms employed by CitS and VcINDY, and possibly by 2-HCTs and DASSs in general, are probably very similar. Overall, the presented molecular model confirms, refines and extends most previous findings for CitS.

7.2.4. Substrate induced conformational changes

2D crystals of CitS grown in buffer containing Na^+ acetate, were soaked in selected substrate combinations before cryo-EM sample preparation. This led to four different projection structures in (1) Na^+ acetate, (2) Na^+ citrate, (3) K^+ acetate and (4) K^+ citrate, each at 6 Åresolution. All four projections have the shape and dimensions of unsoaked dimeric CitS, and all datasets exhibit low phase residuals (Table 1) and reliable calculated diffraction spots (Figures S4-5). Difference maps were calculated from these four maps to examine the influence of the different substrates on the conformation of CitS.

The first difference map (Figure 41A) was calculated from the projections obtained in Na^+-/ and K^+-acetate. Structural differences are negligible across the whole dimer as indicated by bluish and reddish areas with a maximum intensity of +/- 0.2. In the presence of K^+ and absence of Na^+, citrate induced slight density changes (red peaks) towards the center of the distal helix cluster (Figure 41B). The largest structural changes were found when the Na^+-citrate projection was compared to its citrate free counterparts (K^+-/Na^+-acetate). Both difference maps (Figure 41C/D) exhibit strong peaks ($\pm 0.3/\pm 0.5$) at central and inner regions of the distal helix cluster, while the dimer interface domain is almost unaffected. The positions of the observed citrate-induced density changes are the same in both cases although the
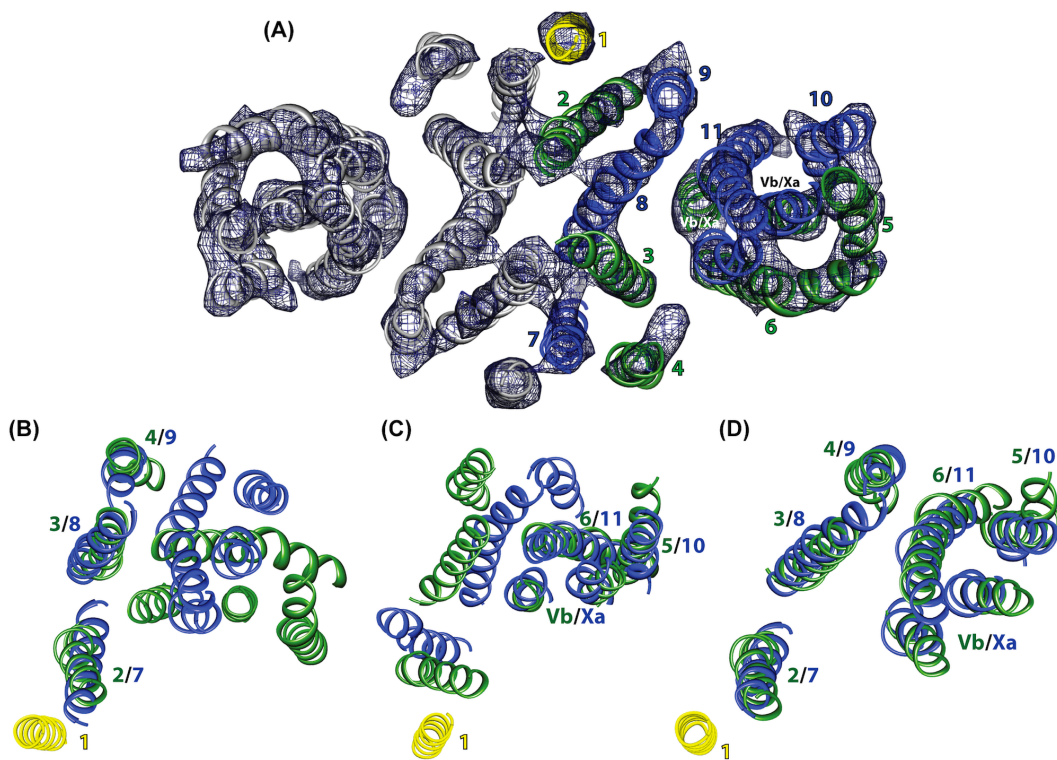
Figure 40: **Molecular model and internal structural symmetry of CitS.** (A) Molecular model of CitS. Helices belonging to one monomer are depicted in yellow (TMS1), green (TMS2-6) and blue (TMS7-11). Superposition of the N- and C-domains focusing on (B) the dimer interface and (C) the distal helix cluster. (D) Independent alignment of both domains.

intensities are slightly different, and confirm the weaker peaks found in the absence of Na$^+$ (Figure 41B). Two additional difference maps shown in Figure S5E/F confirm these results. In all difference maps, the background (lipid bilayer) is without noteworthy changes.

The density changes observed demonstrate the rearrangement of α-helices within the distal helix cluster of each CitS monomer induced by the binding of citrate. This primarily occurs in the presence of Na$^+$; K$^+$ supports minor changes at similar locations. The central dimerization interface remains unaffected by substrate exchanges. These findings suggest (1) that substrate binding occurs within the distal helix cluster of each monomer, (2) that the central helices primarily provide the dimer interface and are not involved in substrate binding, and (3) that citrate induced conformational changes require Na$^+$ ions as co-substrate. Overall, these findings are in agreement with available models for secondary symport. With the exception of EmrE [Ubarretxena-Belandia *et al.* 2003], the monomeric protein is the functional unit of the secondary transporter, oligomerization may regulate transport activity and enhance stability [Herz *et al.* 2009, Perez *et al.* 2011]. This also applies to CitS. None of the 14 helices at the center of the dimer, seven from each monomer, respond to substrate exchange, emphasizing their primary role as a static anchor point with little or no functional role. From the citrate-induced α-helix movements observed at highly reproducible positions, the substrate binding site is located close to the inner edge of the distal helix cluster of each monomer (Figure 41D, black cross). In agreement, according to our model (Figure 40A) this site is formed by helices 5, Vb, Xa and 11, which harbor all known functionally important residues. Citrate would bind close to the aqueous basin at the center of the monomer, providing a structural framework for effective substrate translocation. The proposed binding site closely resembles that of the structurally related transporter VcINDY (arrow, Figure 39C; [Mancusso *et al.* 2012]).

As there was Na$^+$ acetate but no citrate in the buffer, our model probably corresponds to an empty "epmty inward" or "outward open" CitS conformation, "Ci" or "Ce" respectively. The conformational change occurring in the presence of citrate probably relates to the Na$^+$ and citrate induced closure of inner/outer molecular gates at the start of the transport cycle [Forrest *et al.* 2011]. In the resulting inward or outward facing occluded state "CSic" or "CSec", the substrates are inaccessibly buried within the membrane. The movements of helices or hairpins during such gate closures are usually relatively small [Shimamura *et al.* 2010]. In good agreement, the observed shifts were in the range of 5 Å.

Like numerous other Na$^+$ coupled transporters, CitS was previously shown to be inactive in the presence of K$^+$, and less active in the presence of Li$^+$ [Lolkema *et al.* 1994, Mancusso *et al.* 2012, Boudker *et al.* 2007]. Na$^+$ usually binds to the protein first providing a suitable structural and electrostatic framework for the main substrate by direct or indirect interaction [Yamashita *et al.* 2005, Mancusso *et al.* 2012]. This is also valid for CitS [van der Rest *et al.* 1992]. Consequently, the weak citrate in-

duced density shifts observed in the presence of K^+ ((Figure 41B) were unexpected. Although the 2D crystals were extensively soaked in K^+ buffer, CitS might have retained minute amounts of previously bound Na^+, which then enabled the slight conformational change observed. Additional biochemical and higher resolution structural data are required to further elucidate ion coupling in CitS.

7.3. Conclusion

In this study we present the 3D model of the dimeric citrate/sodium symporter CitS of *Klebsiella pneumoniae* based on electron crystallography of two-dimensional crystals. Each CitS monomer is comprised of 13 helices. These are organized in two characteristic domains, seven being in a central cluster forming the dimerization interface and six in a dense distal cluster. Considering previous models, we developed a detailed molecular model of CitS in which we assigned 11 transmembrane helices, two helical reentrant loops and a substrate binding site. The global architecture of CitS resembles that of VcINDY with substantial differences in the helix orientations and positions of reentrant loops. The helical assignments proposed in our model are validated by the internal structural symmetry within each CitS monomer. Additional structural analyses revealed conformational changes induced by the binding of Na^+ and citrate. The observed helix shifts are spatially limited to the distal helix cluster, and in agreement with gate movements predicted to take place during the transport cycle by the "alternating access" model.

7.4. Materials and Methods

See the Supplemental experimental procedures of [Kebbel *et al.* 2013] for a detailed description.

7.4.1. 2D crystallization

2D crystals were produced by reconstitution of detergent solubilized CitS into phospholipid bilayers.

7.4.2. Electron microscopy and image processing

Cryo-EM of plunge-frozen 2D crystals was performed on a Philips CM200 with a field emission gun. Micrographs were recorded at an acceleration voltage of 200 kV and processed using the `2dx` software package [Gipson *et al.* 2007b, Gipson *et al.* 2007a].
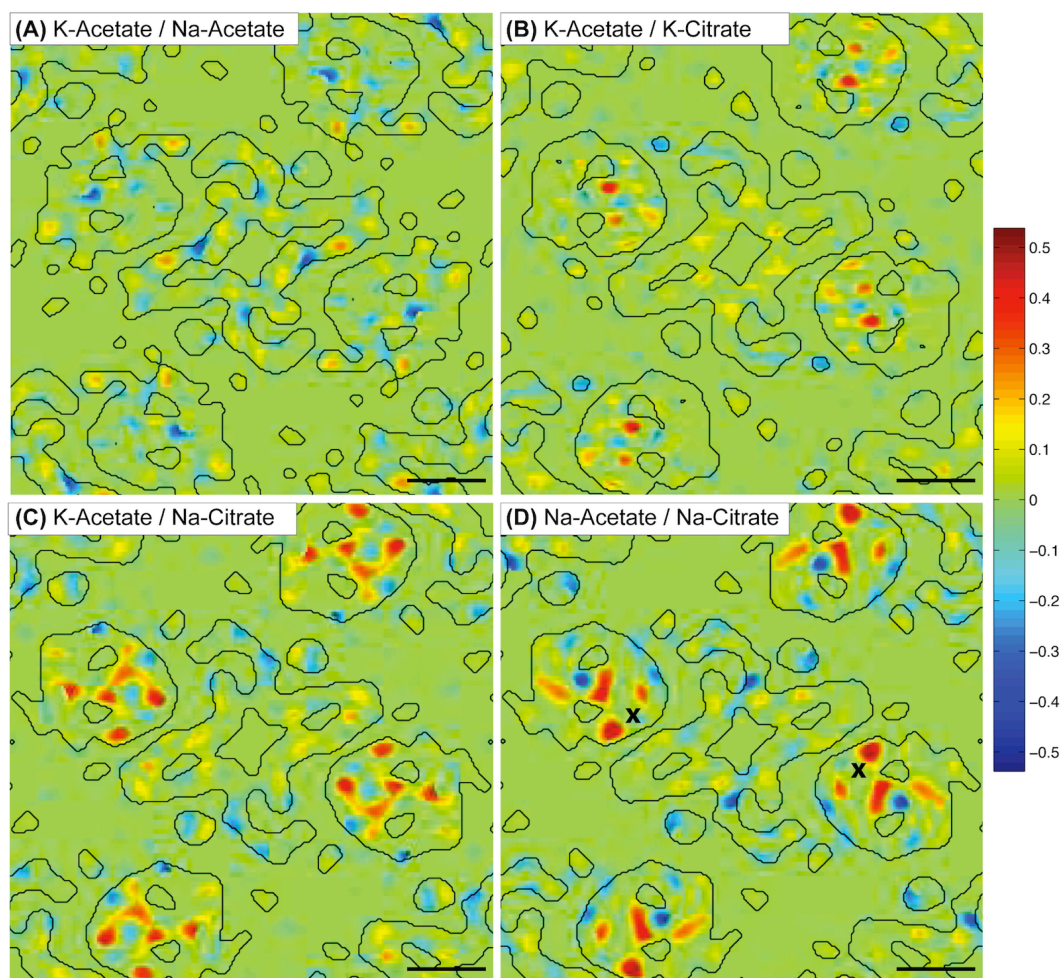
Figure 41: **Substrate induced conformational changes.**   Difference maps of CitS in the presence of different substrates: (A) K-Acetate/Na-Acetate, (B) K-Citrate/K-Acetate, (C) K-Acetate/Na-Citrate, (D) Na-Acetate/Na-Citrate. Cation exchange in acetate (Na^+/K^+) only causes minor structural changes. Binding of citrate causes density shifts within the distal helix cluster, particularly in the presence of Na^+. The proposed substrate translocation site is indicated (x). The contour of the minuend is plotted. Scalebar, 2 nm.

7.4.3. Model building and difference maps

The 3D model of CitS was generated in Chimera [Pettersen *et al.* 2004] by manually placing α-helices into the 3D volume. Difference maps were calculated by subtraction of two corresponding merged projection maps in real space, taking statistical analyses into account.

7.5. Accession Numbers

The EMDataBank accession number for the density map reported in this paper is 2387. The PDB accession number for the molecular model reported in this paper is 4BPQ.

7.6. Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at `http://dx.doi.org/10.1016/j.str.2013.05.011`.

7.7. Acknowledgements & Author Contributions

8. Ligand-induced structural changes in the cyclic nucleotide-modulated potassium channel MloK1

The following study has been published in [Kowal *et al.* 2014]. I was involved in the processing of the 2D crystal images, which also lead to development of automation scripts described in section 4. The significant difference map method from section 5 was also applied.

Abstract

Cyclic nucleotide-modulated ion channels are important for signal transduction and pacemaking in eukaryotes. The molecular determinants of ligand gating in these channels are still unknown, mainly due to a lack of direct structural information. We report here ligand-induced conformational changes in full-length MloK1, a cyclic nucleotide-modulated potassium channel from the bacterium *Mesorhizobium loti*, analyzed by electron crystallography and corroborated by atomic force microscopy. Upon cAMP binding, the cyclic nucleotide binding domains move vertically towards the membrane, and directly contact the S1-S4 voltage sensor domains of their own and adjacent subunits. This conformational change is accompanied by a significant shift-and-tilt of the voltage sensor domain helices. In both states, the inner pore-lining helices are in an "open" conformation. We propose a mechanism for channel gating by cAMP, where ligand binding can favor pore opening via a direct interaction between the cyclic nucleotide binding domains and voltage sensors. This offers a simple mechanistic hypothesis for the coupling between ligand-gating and voltage-sensing in eukaryotic HCN channels.

8.1. Introduction

Cyclic nucleotide-modulated ion channels play a crucial role in signal transduction and pacemaking in eukaryotes, by coupling changes in the levels of the intracellular signaling molecules, cAMP or cGMP, to changes in the ionic flux through the pore of the channel [Kaupp & Seifert 2001, Kaupp & Seifert 2002, Robinson & Siegelbaum 2003]. Members of this family include the cyclic-nucleotide activated channels (CNG), which are transducers in the olfactory and visual sensory system [Kaupp & Seifert 2001, Craven & Zagotta 2006], and the hyperpolarization-activated and cyclic nucleotide-modulated channels (HCN) involved in pacemaker activity and resting potential regulation in heart and brain [Kaupp & Seifert 2001, Robinson & Siegelbaum 2003, Kim *et al.* 2012]. Topologically, HCN channels are part of the voltage-gated ion channel family. They are tetrameric, with each monomer containing six transmem-

brane helices (S1-S6), of which the first four helices (S1-S4) form a voltage sensor domain (VSD). Helices S5-S6 from the four subunits form a single ion-conducting pore in the center of the tetrameric complex. In addition, CNG and HCN channels have a cyclic nucleotide-binding domain (CNBD) at the C-terminus of each monomer. These domains are thought to undergo conformational changes that activate the channel upon cAMP or cGMP binding [Craven & Zagotta 2006]. The CNBDs are connected to the pore via conserved C-linker domains, which are believed to provide the functional coupling between ligand binding and channel opening [Paoletti *et al.* 1999]. The structures of the isolated CNBD/C-linker regions in liganded form from several HCN channels have been determined using X-ray crystallography [Taraska *et al.* 2009, Xu *et al.* 2010, Flynn *et al.* 2007, Lolicato *et al.* 2011, Zagotta *et al.* 2003]. The isolated CNBDs in the 3D crystals were arranged in tetramers connected via the C-linker regions, which seemed to establish all contacts between the subunits [Craven & Zagotta 2006, Zhou *et al.* 2004]. The structures of CNBDs (without C-linkers) in both cAMP-bound and cAMP-free states were determined by X-ray and NMR [Clayton *et al.* 2004, Altieri *et al.* 2008, Schünke *et al.* 2011] only from a prokaryotic homolog of these channels, MloK1 (also called MlotiK1) from *Mesorhizobium loti* [Nimigean *et al.* 2004]. The isolated MloK1 CNBDs did not form tetramers in these studies, likely because of the absence of C-linkers and transmembrane domains [Schünke *et al.* 2011, Schünke *et al.* 2009].

Of all cyclic nucleotide-modulated channels, the only available X-ray structure is that of the MloK1 transmembrane domain alone [Clayton *et al.* 2008]. MloK1 functions as a cyclic nucleotide-modulated K^+ channel despite the presence of an unusually short C-linker containing only ~20 instead of the typical ~80 amino acids [Clayton *et al.* 2004, Nimigean *et al.* 2004]. As expected, the VSDs resolved in the MloK1 transmembrane structure were found in a similar conformation as the voltage-gated K^+ channel Kv1.2, and the pore region was very similar to that of all other K^+ channels [Clayton *et al.* 2008]. The helix bundle gate appeared to be closed and the CNBDs were not resolved in this structure. At present, it is not known whether the MloK1 S1-S4 domain is a traditional voltage sensor, since the S4 helix of MloK1 has only two of the 6 to 7 positively charged residues found in other voltage-sensitive channels, and voltage-sensing properties of MloK1 have not yet been reported.

The intact, full-length cyclic nucleotide-modulated channel MloK1, has been studied by transmission electron microscopy (EM) as single particles and reconstituted into 2D crystals, resulting in a 16 Å-resolution 3D model of the structure in the presence of cAMP [Chiu *et al.* 2007]. Despite its low resolution, this EM analysis showed four CNBDs that make no contacts with each other, arranged with four-fold symmetry in the intact channel. In support of this finding, an atomic force microscopy (AFM) study on 2D crystals of an MloK1 mutant channel also resolved four separate CNBDs in the presence of cAMP. In the absence of cAMP however, the four CNBDs were coalesced into an unstructured blob that extended about 17 Å away from the membrane [Mari *et al.* 2011]. The low resolution and surface

topography restriction made it impossible to interpret the molecular mechanisms underlying these observations. Thus, a higher resolution and real 3D structure of a cyclic nucleotide-modulated channel in the presence and absence of ligand is needed to gain insight into the mechanism of channel opening upon ligand binding.

Here, we present the 3D structures of lipid membrane-reconstituted and 2D-crystallized full-length MloK1 in the presence and absence of cAMP, determined by cryo-EM of membranes vitrified in buffer solution. For both structures, in presence and absence of cAMP, the densities of the voltage sensor helices (S1-S4) are in good agreement with the X-ray structure of the transmembrane region [Clayton *et al.* 2008] while the pore helices of our structures suggest a wider opening at the S6 helix bundle. The cAMP-free CNBDs are positioned away from the membrane, but upon ligand binding, these domains approach the membrane and form an intra-subunit interaction with the S1 helix and an inter-subunit interaction with the VSD of the adjacent subunit. Comparison of the two structures indicates that upon cAMP binding the CNBDs approach the membrane vertically by 3 Å, which is accompanied by a rearrangement in the S1-S4 helices and only a small twist of the S6 helices at the bundle opening. The movement of the CNBDs towards the membrane upon ligand binding within the full-length channel was further confirmed by time-resolved atomic force microscopy (AFM) topography measurements. A mechanism for the gating of cyclic nucleotide modulated channels upon ligand binding is proposed based on the documented conformational changes and the S6 helix bundle conformation.

8.2. Results

8.2.1. Projection maps of MloK1 2D crystals show the molecular packing in the membrane

Well-ordered, large 2D crystals of MloK1 were obtained in the presence and absence of cAMP (Figure 48a,b), by mixing detergent-solubilized MloK1 protein with lipids, followed by slow removal of the detergent by dialysis. Liganded MloK1 crystals were grown in 200 µM cAMP conditions, and ligand-free MloK1 crystals were grown by continuously dialyzing for 10 days in cAMP-free buffer. The successful removal of cAMP during the dialysis was verified by analysis of the cAMP content of the ligand-free crystals, reporting that over 97% of the MloK1 molecules were ligand-free (see Methods). The cryo-EM images of the crystals vitrified in buffer solution produced projection maps in good agreement with the Kv1.2 and MloK1 X-ray models [Clayton *et al.* 2008, Long *et al.* 2007], in conditions similar to those employed by Clayton et al. [Clayton *et al.* 2009]. (Figure 48c,d.) For additional discussion regarding analysis of trehalose-embedded crystals, see Figure 48l e,f. In the 2D maps from vitrified crystals grown in both presence and absence of cAMP, MloK1

tetramers are arranged with p4212 crystal symmetry. The screw-axis symmetry of this crystal form places adjacent MloK1 tetramers in opposite orientations relative to the membrane plane. These maps showed the MloK1 tetramers to interact with adjacent tetramers via the putative VSDs, while the densities corresponding to the CNBD are outside of the membrane plane. Differences in the 2D projection maps between the cAMP-free and cAMP-bound conformations indicate rearrangements of the CNBDs, as well as movements of the peripheral regions of the channel, the VSDs (Figure 48h).

8.2.2. Architecture of MloK1 from 3D maps

To gain insight into the conformational changes of the entire channel upon cAMP binding, we determined the 3D structures of the membrane-embedded full-length MloK1 protein in the presence and absence of cAMP by electron crystallography. For this, we performed cryo-EM imaging of up to 45° tilted 2D crystals followed by 3D merging with 2dx [Arheit *et al.* 2013c, Gipson *et al.* 2007b]. This resulted in two 3D datasets that have comparable resolution (7 Å resolution in the membrane plane, and 12 Å resolution in the vertical direction) and completeness statistics (Figure 43, Figure 42a, and Figure 49). The MloK1 tetramers in both maps have similar architecture, the expected pore on the fourfold symmetry axis, long densities corresponding to the expected 6 transmembrane helices per monomer, and globular densities corresponding to the CNBDs located below the helices. An immediately obvious difference between the two maps is their height difference: the liganded MloK1 is more compact and shorter than the ligand-free MloK1 by about 3 Å (Figure 42).

Precise alignment is required to compare the two cryo-EM maps with one another and with the X-ray structure fragments available for MloK1. To exclude a bias in the z-direction due to crystal contacts, we aligned the two maps to each other based on cross-correlation maximization of the transmembrane parts alone (Figure 42a and (Figure 44a), and docked the X-ray structure of the transmembrane part of MloK1 (PDB ID 2ZD9, [Clayton *et al.* 2008]) into them using Chimera [Pettersen *et al.* 2004] (Figure 44b-e). In both maps the disposition and packing of the S1-S4 helices as well as the selectivity filter and top portion of the pore region correspond very well to the X-ray data of MloK1 and other voltage-gated channel structures [Clayton *et al.* 2008, Long *et al.* 2007, Payandeh *et al.* 2011], though the VSDs of the X-ray model fit slightly better to the ligand-free MloK1 cryo-EM data (Figure 44b). The crystal contacts of adjacent oppositely oriented tetramers are established by helices S2 and S3 of each subunit (Figure 50).

The long S6 helices in both EM maps displayed a wider pore opening at the presumed S6 helix crossing than seen in the MloK1 X-ray structure (Figure 44d). At
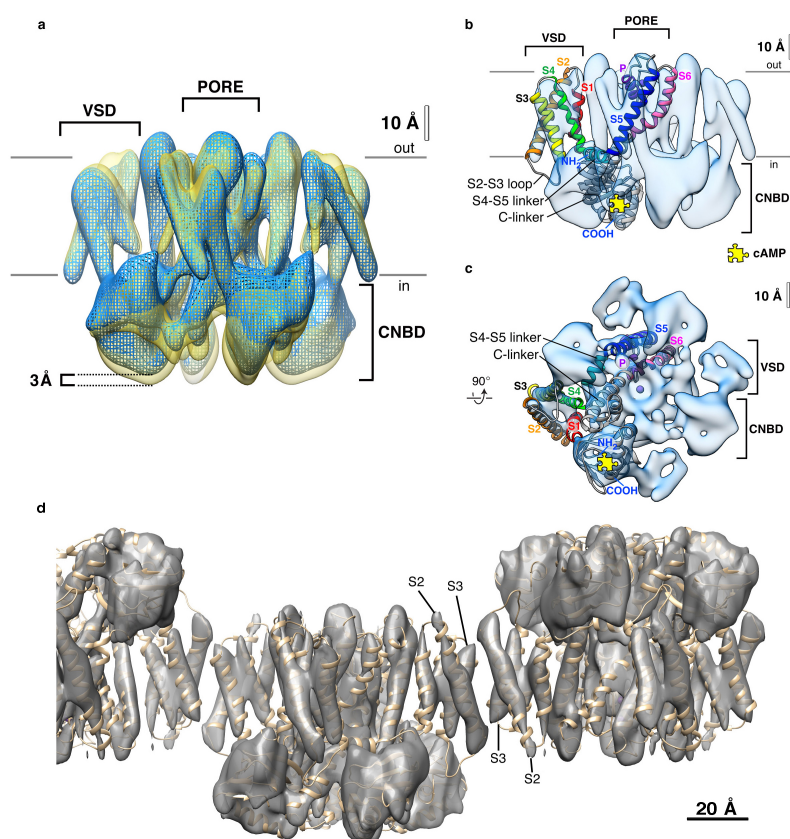
Figure 42: **3D cryo-EM maps reveal the MloK1 conformations in the presence and absence of cAMP.** (a) Superimposed density maps of cAMP-bound (mesh, blue) and cAMP-free (solid, yellow) forms of MloK1. The membrane region as well as the intra- and extra- cellular sides (labels in and out) are indicated. The channel without ligand is ~3Å longer than its cAMP-bound counterpart. An equivalent isocontour level was used for both structures. (b) Density map of MloK1 in the presence of cAMP (blue) shown from the membrane plane, and (c) shown from the intracellular side. A fitted atomistic structure is threaded through one of the subunits. The yellow jigsaw puzzle piece marks the location of the cAMP-binding site; K^+ and/or Ba^+ ions are depicted in purple. Atomistic fitting of the 3D maps was performed using a homology model for MloK1 based on the X-ray structure of the transmembrane region (PDB ID 2ZD9, [Clayton *et al.* 2008]), and the X-ray structure of CNBD in the presence of cAMP (PDB ID 3CL1, [Altieri *et al.* 2008]). (d) The protein-protein contact between oppositely oriented adjacent MloK1 tetramers (here in the absence of cAMP) in the 2D membrane protein crystals is established via the S2 and S3 helices of the VSDs. The determined model of MloK1 without cAMP (brown) is shown together with the cryo-EM map, which is shown here as raw data without the resolution limitation to $7 \times 7 \times 12$Å.

3D Reconstruction Parameters

	With cAMP	Without cAMP
Crystal plane group symmetry	p42$_1$2	p42$_1$2
Crystal unit cell parameters		
a, b, c [Å]	131.0, 131.0, 400.0	130.0, 130.0, 400.0
γ[°]	90.0	90.0
Number of images	78	67
Range of defocus [µm]	0.655 ... 3.077	0.498 ... 2.600
IQ range used	1 ... 8	1 ... 8
Tilt range used [°]	0.0 ... 46.0	0.0 ... 44.6
In-plane resolution cut-off [Å]	7.0	7.0
Vertical resolution cut-off [Å]	12.0	12.0
Number of observed reflections	42588	36314
Number of observed unique reflections	24242	20574
Number of possible unique reflections in asymmetric unit	4025	3911
Number observed unique reflections in asymmetric unit with FOM > 1%	3409	3337
Completeness to 45° tilt and 7.0 Å (12 Å vertical), counting reflections with FOM > 1% [%]	84.7	85.3
Number observed unique reflections in asymmetric unit with FOM > 50%	3083	2836
Completeness to 45° tilt and 7.0 Å (12 Å vertical), counting reflections with FOM > 50% [%]	76.6	72.5
Overall weighted phase residual [°]	15.7	15.4
Overall weighted R-factor [%]	33.5	32.3

Figure 43: Crystallographic data and statistics.

the level of the presumed helix bundle closure (A207), each S6 helix is 1.5-2 Å further away from the central axis than found in the X-ray structure of MloK1. The conformation of the S6 helices in both our EM maps corresponds more closely to that of the X-ray structure of Kv1.2 (at V406), an open pore (Figure 51). The central location of the selectivity filter region showed a large elliptic density, which we attributed to a partial occupancy with K^+ or other ions (e.g., Ba^{2+}) present in the buffer (Figure 44c.)

In both maps, the long S6 helices are connected to four individually resolved CNBD densities, which can be well docked with the X-ray-determined CNBD structure [Altieri *et al.* 2008]. In the presence of cAMP, the CNBDs were found in direct contact with the VSD domains and thus abutting the surface of the lipid membrane, while in the absence of cAMP, the CNBDs were seen approximately 3 Å below this surface. The cAMP-bound CNBD thereby interacts not only with the VSD of its own subunit via the S1 helix, but also with the VSD of the adjacent subunit in the same tetramer via the S3 helix (Figure 44e).

8.2.3. Structural differences between cAMP-bound and cAMP-free MloK1 channels

The alignment of the two maps highlighted significant differences in the CNBDs and the VSD helices between the ligand-free and the liganded MloK1 structures. To better visualize these differences induced in the MloK1 channel by cAMP binding, we constructed atomic models of the full-length channel in the presence and absence of cAMP by flexible model fitting to the cryo-EM maps, starting with the X-ray structures of MloK1 (PDB ID 2ZD9) and isolated CNBDs (PDB ID 3CL1) (Figure 42b,c Figure 45, Movies 4-6).

The most prominent change upon cAMP binding is a movement of the MloK1 CNBDs towards the membrane surface that vertically shortens the protein by ~3 Å (Figure 42a, Figure 44e). The CNBD structure determined by X-ray crystallography (PDB ID 3CL1) [Altieri *et al.* 2008] could be docked into the obtained EM maps (Figure 45a). Our structures therefore allow detailed conclusions about the disposition of the CNBDs relative to the membrane parts. However, a higher resolution cryo-EM analysis will be required to discern conformational changes within the CNBDs upon ligand binding.

Comparison of the VSDs from the ligand-free and liganded MloK1 structures and models showed differences in the arrangement of helices S1-S4 (Figure 45b). In the presence of cAMP, and thus while in contact with the CNBDs, all four helices were tilted relative to their more vertical orientation in the absence of ligand. In the presence of cAMP, the largest relative tilts with respect to the central axis of the channel were observed for helix S1 (9°) and helix S2 (6.5°). The linkers connecting
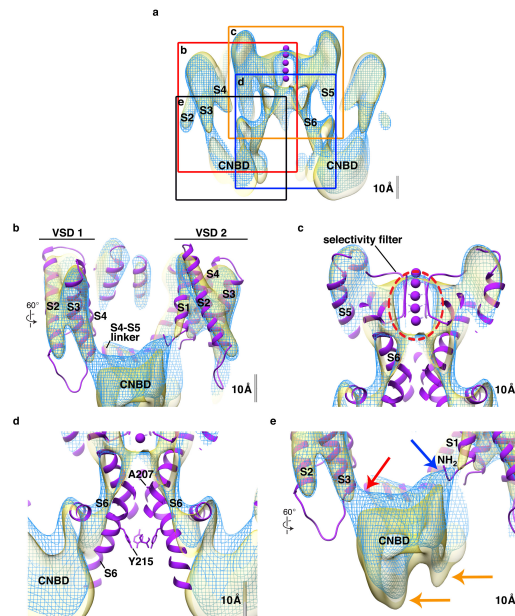
Figure 44: **The architecture of MloK1.** (a) Cross-section through the superimposed cryo-EM maps; the regions of interest (ROI) indicated by the frames are shown in (b-e). MloK1 with cAMP: mesh, blue; MloK1 without cAMP: solid, yellow. The K^+ ions (purple) in *a* as well as the ribbon models shown in *b-e* are from the X-ray structure (PDB ID 2ZD9, purple) and fitted to the cryo-EM map of MloK1 without ligand. (b) Detailed view of the VSDs shows that the X-ray structure agrees better with the densities of the cryo-EM map of cAMP-free than cAMP-bound MloK1. The tilt of helix S2 with respect to the X-ray structure is evident. (c) Detailed view of the selectivity filter region. The ion densities determined in both cryo-EM maps (red dashed ellipse) are in the selectivity filter region, their positions correspond well to the X-ray structure. (d) Detailed view of the intracellular pore permeation pathway region. Both maps of MloK1 show a significantly wider opening at the helix bundle crossing than the X-ray structure. Residues that form the inner helix bundle crossing (A207), and a constriction region from the X-ray structure (Y215)19 are shown as sticks for the X-ray structure. These are clearly outside of the cryo-EM maps. (e) Detailed view of the CNBD-VSD interaction region. Upon binding of cAMP the CNBD undergoes conformational changes resulting in a 3 Å-shift towards the lipid bilayer (orange arrows), establishing an interaction of the CNBDs with helix S1 of the same monomer (blue arrow) and helix S3 (red arrow) of the adjacent monomer.

helices S4 and S5, believed to play a role in channel gating in other voltage-gated channels, are slightly displaced toward the extracellular side by 1-2 Å in liganded MloK1 (Movie 1). The observation of conformational changes of the VSDs following liganding of the CNBDs provides direct evidence of coupling between ligand binding and voltage gating in HCN channels, as further discussed below.

We next examined whether there is a change in the pore opening diameter between the ligand-free and liganded MloK1 structures and models. Inspection of the protein cross-section from the extracellular to the intracellular side shows a small extracellular funnel, the selectivity filter, the S6 helix bundle crossing, the central cavity, and an intracellular chamber (Figure 51). The pore-helices S6, which are connected to the CNBDs by short (~20-aa) C-linkers, are found in very similar positions in both models (Figure 44). The pore opening at the lower end of the S6 helices (Y215) located between the central cavity and the intracellular chamber has a diameter of at least 8.5-10 Å (Figure 45c and Figure 51b). Thus, for a hydrated potassium ion of a diameter of ~5.6 Å [Mahler & Persson 2012], the opening at the lower end of the S6 helices is always sufficiently wide to allow the passage into the central cavity, regardless of whether cAMP is bound or not. Comparison of the two maps and models further shows that upon ligand binding helices S5 and S6 are twisted slightly counter-clockwise from their original positions (Figure 45d, Movies 1-3). At the present resolution, the cryo-EM maps do not resolve the state of the S4 helix (3_{10} vs. α-helical [Clayton *et al.* 2008]), nor the changes in the selectivity filter conformation necessary to stop ionic conduction through the pore.

8.2.4. Conformational changes in MloK1 imaged with AFM

To obtain an independent assessment of the ligand-induced conformational changes in MloK1, we imaged both the liganded and the ligand-free MloK1 2D crystals with AFM. Surface topography recordings by AFM showed the MloK1 2D crystals to be mono-layered with an average height in the presence of cAMP of 11.2 ± 0.3 nm and in the absence of cAMP of 11.1 ± 0.4 nm (Figure 46). Individual CNBDs were resolved on the surface of the wild-type MloK1 in topographs recorded in the presence of cAMP showing a left-handed windmill structure in raw data and in average topography (Figure 46a, left), while they assembled into a single featureless protrusion in cAMP-free conditions (Figure 46a, right). This rearrangement of the CNBDs is accompanied by a height increase of 1.5 nm (Figure 46b). Qualitatively, the increase in height in the ligand-free MloK1 molecules is consistent with the cryo-EM observations.

Furthermore, to ascertain that the changes in CNBDs we observed in the ligand-free crystals are indeed due to ligand removal, we performed time-lapse imaging experiments on the same membrane region while removing cAMP from the AFM
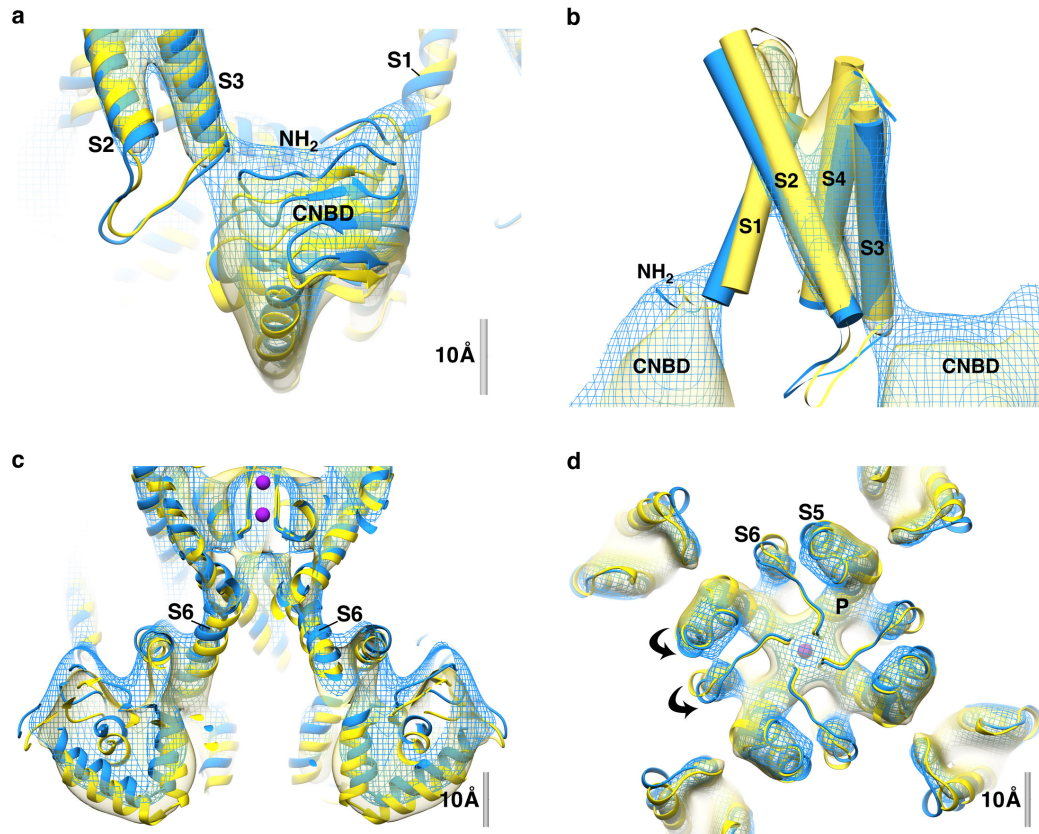
Figure 45: **Structural changes in MloK1 upon cAMP binding.** 3D atom-
istic models are fitted to the maps of the cAMP-bound (mesh, blue) and
cAMP-free (solid, yellow) MloK1 structures. (a) Side view of the CNBD
region. Upon ligand binding, CNBDs interacts with the N-terminal re-
gion of its own monomer's S1 helix, and the S3 helix of its neighbor.
(b) The VSD rearrangement viewed from the side. Helices are shown as
cylinders. (c) Cross-section through the transmembrane pore and CNBD
regions. (d) Extracellular view on inner helices S5 and S6, which twist
slightly counter-clockwise upon ligand binding.

fluid cell. Individual MloK1 tetramers underwent conformational changes within a few hours, after which CNBDs revealed an increased height as observed under static conditions (Figure 46a), and could no longer be individually resolved by AFM (Figure 46c). Thus, the occurrence of the conformational change is slow, consistent with previous reports of difficulty in removing the ligand from the binding pocket and obtaining a ligand-free CNBD [Cukkemane *et al.* 2007, Peuker *et al.* 2013]. It is important to note that even after 221 minutes of imaging in the absence of ligand, there are still some well-resolved tetramers that presumably have not yet lost all the ligands, a good internal control that the conformational changes we observe happen at the level of the individual molecules and are not due to AFM-tip/sample interactions.

8.3. Discussion

We 2D-crystallized MloK1, a cyclic nucleotide-modulated potassium channel, in the presence and absence of its ligand, cAMP. Cryo-EM imaging of fully hydrated and plunge-frozen samples produced 3D maps at 7 Å by 12 Å resolution, to which atomic structure models based on the X-ray structures of the isolated transmembrane regions and CNBDs of MloK1 were fitted. A mechanism of channel gating upon cyclic nucleotide binding can be formulated from the observed differences between the two models in the presence and absence of cAMP. By comparing the two models (Figure 47a) aligned by their selectivity filter as a point of reference, we observed that upon cAMP binding to the CNBDs, the CNBDs approach the lipid membrane and establish contact with the voltage sensor domains that shift slightly towards the CNBD and its helices tilt in the membrane (Figure 45b and Movie 1-3). These conformational changes should favor the opening of the channel pore. However, as indicated by the cryo-EM maps, these domain movements do not lead to any detectable diameter changes either at the level of the Y215 in the cavity, or at the lower end of the S6 helices (A207, commonly referred to as the helix bundle crossing). At both levels, our map shows sufficiently wide openings to allow ion permeation, with significantly larger diameters than in the crystal structure of MloK1.
Our data suggest only a very small twist of helices S5 and S6 (Figure 45d, Movie 2). Such small movements near the C-terminus of S6 with no pore diameter changes are not in agreement with channel gating at the helix bundle-crossing, as observed for KcsA [Perozo *et al.* 1999, Kelly & Gross 2003, Doyle *et al.* 1998], and proposed for voltage-gated channels such as Shaker and Kv1.2 [Long *et al.* 2007, del Camino & Yellen 2001, Jensen *et al.* 2012], including HCN-type channels [Rothberg *et al.* 2002, Wahl-Schott & Biel 2009]. Furthermore, the central cavity and lower intracellular chamber of both our MloK1 cryo-EM structures are wider than in the X-ray structure of MloK1 [Clayton *et al.* 2009], and have a diameter of at least 8.5-10 Å, easily allowing the passage of a hydrated K^+ ion ((Figure 51b). The bundle-crossing gate is thus "open" in both ligand-free
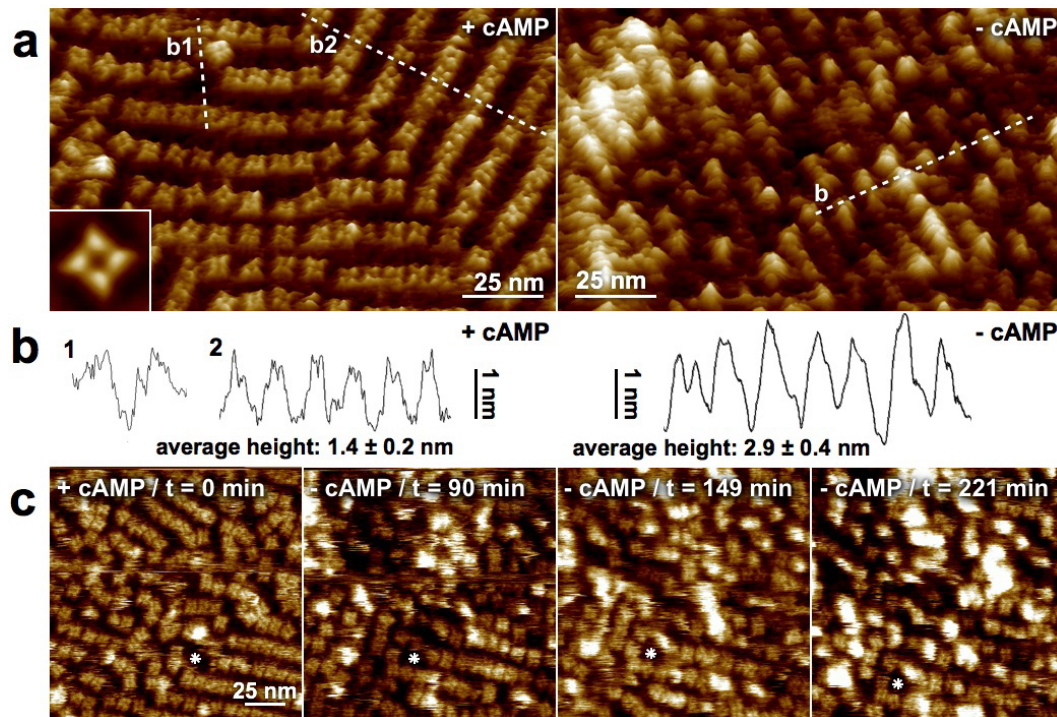
Figure 46: **AFM analysis of MloK1 in presence and absence of cAMP.** (a)
High-resolution 3D AFM topographs of MloK1. In the cAMP-bound
state (left) each subunit of the tetrameric channels is well resolved on each
single molecule. The four CNBDs were arranged in a left-handed windmill
(inset: average). In the cAMP-free state (right) sub-molecular details
could not be resolved and the tetramer is contoured as a single protrusion
of variable height and appearance. (b) Height profile analysis of MloK1
molecules along the dashed lines in (a). In the +cAMP conformation
(left) the channels protruded 1.4 ± 0.2 nm (n=50) from the membrane.
In absence of ligand (right) MloK1 protruded 2.9 ± 0.4 nm (n=50). A
lattice defect was exploited to ensure that the lower protrusion height
of the +cAMP conformation was not underestimated due to the dense
molecule packing (height profile b1). (c) Time-lapse AFM of MloK1
channels undergoing the conformational change from cAMP-bound to
unbound state. MloK1 channels were imaged in the presence of $50\,\mu$M
cAMP (t=0), after which the AFM fluid cell was extensively rinsed with
cAMP-free buffer. After incubation time of 90 minutes first single MloK1
channels lost their characteristic four-fold symmetric windmill structure
and changed into a shapeless protrusion of increased height, and their
number increased with time. The asterisks in the four image panels
indicate the same membrane position.

and liganded conformations, and the helix bundle-crossing opening for both channel forms is larger than in X-ray model, suggesting that these channels do not gate at the bundle crossing.

An alternative is that the MloK1 channels gate at the selectivity filter. It is possible that the conformational changes occurring in the channel following cAMP binding lead to subtle changes in the selectivity filter (sufficient to restrict or allow permeation). Such changes, as for example a transition from a KcsA-like collapsed selectivity filter that does not allow K^+ permeation to a conductive selectivity filter, is unlikely to be captured in our cryo-EM analysis at the current resolution [Zhou *et al.* 2001]. This type of selectivity filter gating was shown to be involved in C-type inactivation [Yellen 1998, Cuello *et al.* 2010] and also be the major gating mechanism in several ligand-gated ion channels, such as CNG, SK, BK, and MthK channels [Flynn *et al.* 2001, Wilkens & Aldrich 2006, Bruening-Wright *et al.* 2002, Posson *et al.* 2013]. It appears unlikely that the crystal packing would allow conformational changes at the VSDs, while restricting changes near the channel pore, keeping it in the same state, either open or closed, regardless of whether the ligand is bound or not [Tsai *et al.* 2013]. Thus, higher-resolution structural data of both channel forms are necessary to pinpoint the location of the gate in MloK1.

The cryo-EM structures presented here provide new insight into the packing of the CNBDs in the context of the full-length channel in the presence and absence of ligand. In the non-liganded state, these domains are separated from the transmembrane-spanning portions by 3 Å, and make no contact with each other, in agreement with earlier reports [Zagotta *et al.* 2003, Chiu *et al.* 2007]. In the presence of cAMP, however, the CNBDs showed to make direct contact with the membrane and the voltage sensors from both the adjacent and the same subunit, suggesting a simple mechanism for the previously observed coupling between ligand gating and voltage-gating in eukaryotic HCN channels [Robinson & Siegelbaum 2003, Wang *et al.* 2001, Kusch *et al.* 2010]. The contact of one liganded CNBD to two VSDs (one from the same and the other from an adjacent subunit within the same channel tetramer, as shown in (Figure 44e) also suggests a way to achieve cooperativity in ligand gating. Eukaryotic CNG and HCN channels are known to display cooperativity in ligand gating [Biskup *et al.* 2007, Chow *et al.* 2012, Dekker & Yellen 2006, Ruiz & Karpen 1999], while the jury is still out for MloK1 [Cukkemane *et al.* 2007, Peuker *et al.* 2013].

To the best of our knowledge, our study provides the first structural indication that CNBDs and voltage sensors interact directly, although this was proposed recently for HCN channels based on functional assays [Kusch *et al.* 2012]. Our structures show that in the presence of cAMP the VSDs are slightly lower in the membrane and their helices are tilted. This finding is in agreement with the observation that cAMP binding in HCN channels lowers the activation voltage. Different mechanistic models

have been proposed to describe the movement of the S4 helix and its relation to the other helices in the VSDs of voltage-gated potassium channels [Clayton *et al.* 2008, Williams 1984, Catterall 1986, Papazian & Bezanilla 1997, Jiang *et al.* 2003].   Of these models, our findings correlate best with the S4 motion in the "transporter"-like model, which comprises a tilt and rotation of the S4 helix, accompanied by a 2-3 Å movement [Papazian & Bezanilla 1997].

A direct interaction between CNBDs and VSDs also explains why the CNBDs of the tetrameric MloK1 were only resolved in AFM images when cAMP was present, in our work on the wild type MloK1 as well as in a mutant with lower apparent affinity for cAMP [Mari *et al.* 2011]. Under conditions where CNBDs are coupled to the VSDs and thus restricted in mobility, the AFM is able to contour the CNBDs in nice detail. In the absence of cAMP, the CNBDs lose contact with the VSDs, as shown by cryo-EM, and are thus more mobile so that individual CNBDs could no longer be resolved by AFM. Thus, the protrusion height difference of 1.5 nm (1.7 nm in the mutant [Mari *et al.* 2011]) between the ligand-free and liganded conformations observed by AFM may not reflect a conformational change of that amplitude but corroborates the uncoupling of the CNBDs from the transmembrane portions. Moreover, the in situ time-lapse AFM experiment allowed us to observe the coexistence of both morphologies, clearly indicating that they are two realistic conformations that the channel achieves during gating.

In summary, we report the 3D structures of a cyclic nucleotide-modulated potassium channel in the presence and absence of ligand, which allowed us to formulate a mechanism of channel gating upon ligand binding. We provide structural evidence for direct interaction between the ligand binding domains and the voltage sensor domains during channel gating. Furthermore, based on similar inner pore entrance conformations, we suggest that the channel may not gate at the bundle-crossing.

8.4. Online Methods

8.4.1. MloK1 expression and 2D crystallization

Intact, full-length, cyclic nucleotide-modulated potassium channel MloK1 was expressed and purified as described [Nimigean *et al.* 2004, Chiu *et al.* 2007], with minor modifications. Transformed E. coli cells containing a C-terminally hexahistidine-tagged MloK1 construct were grown in L-broth at 37° C. Expression was induced with 0.2 mg/ml anhydrotetracycline for 2 h at an OD600 of 0.7. Cells were pelleted and lysed by sonication. Membrane proteins were solubilized by adding 1.2% n-decyl-β-D-maltopyranoside (DM; Anatrace) to the solubilization buffer (295 mM NaCl, 5 mM KCl, 20 mM Tris-HCl pH 8.0, 10% glycerol, 1 mM PMSF, 0.2 mM cAMP) and incubating for 2.5 h at 4° C. Extracted MloK1 was purified in 295 mM
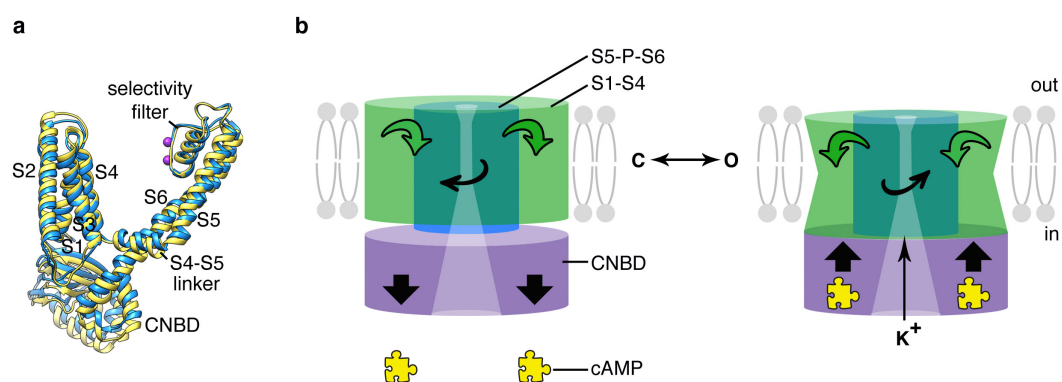
Figure 47: **Mechanism of MloK1 gating with cAMP.** (a) The modeled individual subunits of MloK1 in the cAMP-bound (blue) and cAMP-free (yellow) structures are shown, aligned to the selectivity filters. (b) Proposed mechanism based on the alignment shown in (a). Left: without cAMP, right: with bound cAMP (symbolized by the yellow jigsaw puzzle piece). cAMP binding leads to CNBDs (purple cylinder) to move 3 Å towards the membrane, while VSDs (green cylinder) move towards the CNBDs and tilt in lipid bilayer. In parallel, the S5-P-S6 core (blue cylinder) twists slightly in a counter-clockwise (CCW) direction when seen from the extracellular side. This is reversed when cAMP dissociates (left). C/O indicate the closed/open channel configurations.

NaCl, 5 mM KCl, 20 mM Tris-HCl pH 8.0, 10% glycerol, 1 mM PMSF, 40/500 mM (wash/elution) imidazole, 0.2% DM, 0.2 mM cAMP (purification buffer) by Co^{2+}-affinity chromatography. To maintain the integrity and function of MloK1, 0.2 mM cAMP (Fluka) was present throughout the purification process. Purified MloK1 was quite monodisperse as judged by TEM imaging. The purified protein in detergent had a tendency to aggregate in the absence of ligand [Nimigean *et al.* 2004].
Pure, detergent-solubilized MloK1 in purification buffer was mixed with E. coli polar lipid extract (Avanti Polar Lipids) at a lipid-to-protein ratio of 0.8-1.0, and dialyzed against detergent-free buffer (20 mM KCl, 20 mM Tris-HCl pH 7.6, 1 mM $BaCl_2$, 1 mM EDTA) with and without 0.2 mM cAMP. Protein was 2D-crystallized in dialysis buttons for 5 to 10 days (in buffer with cAMP ligand) or 10 days (in buffer without cAMP ligand), at 37° C, whereby the dialysis buffer was exchanged every other day.

8.4.2. Determination of the bound cAMP

After extensive dialysis under crystallization conditions at 37°C for up to 12 days (buffer-dialysis-volume ratio of ~3000, several buffer changes), 30 µl of dialysed MloK1 2D crystal sample was digested with 100 µg/mL proteinase K (Sigma-Aldrich P2308) with a supplement of 2 mM CaCl2 at 37°C for 15h and centrifuged at 12000 g at 4°C for 15 minutes. SDS-PAGE analysis confirmed the absence of any Mlok1 protein after proteolysis. Liquid Chromatography / Mass Spectrometry (LC/MS) quantification of the total content of cAMP in the digested sample reported a concentration of less than 0.5 µM cAMP after dialysis.
Considering the MloK1 input protein concentration for 2D crystallization of 1 mg/ml or 20 µM MloK1, and the LC/MS-determined final cAMP concentration of less than 0.5 µM, the calculated concentration of unliganded MloK1 protein at chemical equilibrium was therefore determined to be higher than 19.5 µM, corresponding to more than 97% of the MloK1 protein being unliganded.

8.4.3. Cryo-EM and structure determination

MloK1 2D crystal samples in buffer solution with or without cAMP were applied to glow-discharged, thin carbon film-coated copper EM grids. Samples were adsorbed for 1 min in a humid atmosphere. Afterwards, the grids were blotted for 3.5 sec and vitrified, without the addition of cryoprotectant, by plunging them into liquid nitrogen-cooled liquid ethane using an FEI Vitrobot MK4 (Vitrobot, Maastricht Instruments). Alternatively, MloK1 2D crystal solution was adsorbed to glow-discharged carbon film, exposed to varying concentrations (2% - 15%) of

trehalose [Hirai *et al.* 1999], blotted, and frozen by plunging into liquid nitrogen. The frozen grids were transferred into a Philips CM200 FEG or a FEI Polara microscope, operated at an acceleration voltage of 200 respectively 300 kV and a nominal magnification of 50,000x to 60,000x. Low electron dose images were recorded on Kodak SO163 film using a defocus range of 500-2500 nm. Micrographs were checked by optical diffraction, and promising micrographs were digitized on a Heidelberg Primescan D7100 drum scanner with a step size of 5 % m (1 or 0.83 Å/pixel at the specimen level).

Digitized images were processed with the `2dx` software package (`http://2dx.org`); [Arheit *et al.* 2013c, Gipson *et al.* 2007b, Arheit *et al.* 2013a, Gipson *et al.* 2007a]), which is based on the MRC software [Crowther *et al.* 1996]. Within `2dx`, images were corrected for crystal disorder by three rounds of unbending, without the use of any external reference such as generated by MAKETRAN. Defocus and astigmatism were determined by CTFFIND3 [Mindell & Grigorieff 2003]. The tilt geometry of individual crystals was determined from CTF variations across the image, refined based on lattice distortions, and further refined by comparing the single images to the merged 3D dataset. 3D merging was done with `2dx_merge`, using ORIGTILT [Crowther *et al.* 1996], lattice lines were interpolated with LATLINEK [Agard 1983]. The Quality Evaluation statistics was reported by `2dx` for the 7×12Å resolution-limited dataset. A final 3D volume was calculated within `2dx` with the CCP4 software package [CCP4 1994], and visualized using Chimera [Pettersen *et al.* 2004].

8.4.4. Flexible Model Fitting to Cryo-EM Maps

A starting model was built for the transmembrane part of MloK1 (residues 7-228) from the crystal structure PDB ID 2ZD9 [Clayton *et al.* 2008], and for the CNBD (residues 229-349) from the crystal structure PDB ID 3CL1 [Altieri *et al.* 2008]. The two structures were placed in the EM density and combined using the program Coot [Emsley *et al.* 2010]. A first tetrameric model was produced by rigid-body fitting of four copies of this monomer into the density using Chimera [Pettersen *et al.* 2004]. The real-space refinement program DireX [Schröder *et al.* 2007] was used to fit the structure into the cryo-EM map, applying a B-factor of 200 Å^2 parallel, and 450 Å^2 perpendicular to the membrane plane. These B-factors were optimized to best match the point-spread function of the EM density map as calculated by `2dx`. A simple Babinet bulk solvent model with ksol=0.6 and Bsol=350 Å^2 was used. A mask was applied to avoid artifacts from the periodicity of the 2D protein crystal. The mask was generated as a 30 Å-resolution density map computed from the starting model, and its edges were then steepened with the apply-cos-mapping tool, which is part of DireX, using 0.25 and 0.15 as upper and lower threshold values, respectively. The four monomers were restrained by NCS restraints, which keep the monomer

structures similar but not identical to each other, with an average RMSD of 0.4 Å. We modeled two barium ions; their positions within the channel were maintained by 14 distance restraints to nearby amino acid side chains.

The model was refined into the EM map using 500 steps of DireX refinement. Deformable elastic network (DEN) restraints were defined between randomly chosen atoms that were within 3 to 15 Å in the starting structure and separated by no more than 30 amino acids in the sequence, yielding 20 686 restraints, which is twice the number of atoms in the model. DEN restraints within α-helices and β-sheets were set to be twice as strong as within loop regions.

A cross-validation approach was used to identify the optimally fitted model and to prevent overfitting [Falkner & Schroder 2013]. In brief, only Fourier components of the cryo-EM maps up to a resolution of 7 Å were used for fitting, while reflections from the 'free' interval 5-7 Å were used for validation. For this, the cross-correlation coefficient, C_{free}, was calculated between the model density map and the cryo-EM density map, both of which were band-pass filtered to the 'free' interval, thus containing information that had not been used for fitting.

The same starting model was fitted into both density maps, i.e., for the cAMP-bound and cAMP-free states, and the same set of DireX parameters were used. During the refinement, the C_{free} value for the cAMP-free case increased from 0.081 to 0.142 and the C_{work} value increased from 0.794 to 0.833. For the cAMP-bound case, the C_{free} value increased from 0.048 to 0.154, and the C_{work} value increased from 0.688 to 0.827.

The refined models were further optimized using CNS [Brunger *et al.* 1998] with 1,000 steps of Cartesian dynamics at a temperature of 40 K followed by 300 minimization steps. For both dynamics and minimization, all hydrogen atoms were added to the model and electrostatic interactions were only used between backbone atoms to improve the hydrogen-bonding network. In addition, statistical Ramachandran restraints were used to improve the backbone dihedral angles.

As the conformation of the CNBDs in the final models for both conformations was quite similar to each other with an RMSD of 0.8 Å, we generated a second starting model as a control. This was built from the same transmembrane part (residues 7-228) from crystal structure PDB ID 2ZD9 [Clayton *et al.* 2008] as before, but the CNBDs (residues 229-349) were replaced by the open NMR CNBD structure PDB ID 2KXL [Schünke *et al.* 2011] determined in absence of cAMP. Refinements from this second starting model, using the same DireX parameters as the previous refinements, resulted in final refined structures with rather large RMSD from the starting model, of 7.8 Å and 7.3 Å in the presence and absence of cAMP, respectively. The large conformational changes occurring during the refinements and the lower C_{free} values of 0.114 and 0.087 in presence and absence of cAMP, respectively, indicated that these final models from the second starting model were of much lower quality than with the first starting model. An analysis of the internal conformations of the CNBDs was therefore not feasible with the current cryo-EM data.

8.4.5. Difference map

Difference maps were calculated for the 2D projection maps of MloK1 with and without cAMP, limited to 7 Å resolution. The alignment of the individual projection maps was achieved using the common phase origin of the $p42_12$ symmetry gained from the MRC program ALLSPACE [Valpuesta *et al.* 1994] and refined by MRC program ORIGTITLT [Crowther *et al.* 1996], all within the 2dx software. The merged projection maps were scaled to $[0, 1]$, and a raw difference map was calculated by subtraction of one map from the other.

To determine the significance of the differences, we calculated the variations within the individual projection maps, in a similar manner as done before [Appel *et al.* 2009b]. To this end, we empirically determined the variations across the two conformations, by randomly splitting images from both conformations into two sets, averaging the images from both sets, and calculating the corresponding difference map. Repeated application of this procedure resulted in a large number of difference maps, which were then averaged to produce a general variation map. Differences between conformations with and without cAMP were only considered to be significant if they exceeded those present in the variation map. The final result was visualized as a heat map using Matplotlib.

8.4.6. Atomic Force Microscopy (AFM)

MloK1 2D crystal preparations (as above) were adsorbed on freshly cleaved mica at a protein concentration of 0.1 mg/mL in 10 mM Tris, pH 7.4, 150 mM KCl, 3mM EDTA, and in case of cAMP-bound crystals containing additional 50 M cAMP (Sigma-Aldrich). After 20 min incubation the sample was rinsed carefully with 10 mM Tris, pH 7.4, 150 mM KCl buffer cAMP (as indicated) to remove non-adsorbed membrane crystals. Membranes were imaged in contact-mode AFM using a Nanoscope E system (Bruker, Santa Barbara, CA, USA) equipped with a 160 m scanner (J-scanner), a fluid cell and OMCL TR400PSA cantilevers (Olympus, Tokyo, Japan) with a nominal spring constants of 80 pN/nm. The measurements were performed in buffer solution and at ambient temperature and pressure. Imaging forces were minimized to ~100 pN. Images were acquired at scanning frequencies of 4-7 Hz and optimized proportional and integral gain parameters used. A more detailed description of AFM data acquisition can be found in [Scheuring *et al.* 2005]. For time-lapse AFM experiments the sample was washed thoroughly by exchanging 10 times the volume of the fluid cell with cAMP free buffer, and the very same

MloK1 membrane patch was continuously imaged.

8.4.6.1 Accession codes

The cryo-EM volumes will be deposited before publication of this manuscript at the EM Data Bank, accession number XXX. The atomic models will be deposited before publication of this manuscript at the Protein Data Bank, accession number YYY. The raw cryo-EM image data used for calculating the maps reported here, together with the employed `2dx` processing parameters, will be made available as public dataset at the (`http://2dx.org`) website before publication of this manuscript.

8.4.6.2 Acknowledgements

8.4.6.3 Author contributions

H.S. and C.M.N. conceived and designed the research, C.M.N. provided construct and bacterial strain, P.B. expressed and purified protein, J.K. performed 2D crystallization with contributions from P.L.C., J.K. and M.C. collected cryo-EM data, J.K., M.A., and H.S. performed image processing, M.R. recorded AFM data, H.S., J.K., S.S. and G.S. analyzed the data. P.B. performed ligand-binding assay. M.A. generated difference maps. J.K., H.S., and C.M.N. wrote the manuscript.

8.4.6.4 Competing financial interest

The authors declare no competing financial interests.

8.5. Supplementary Material
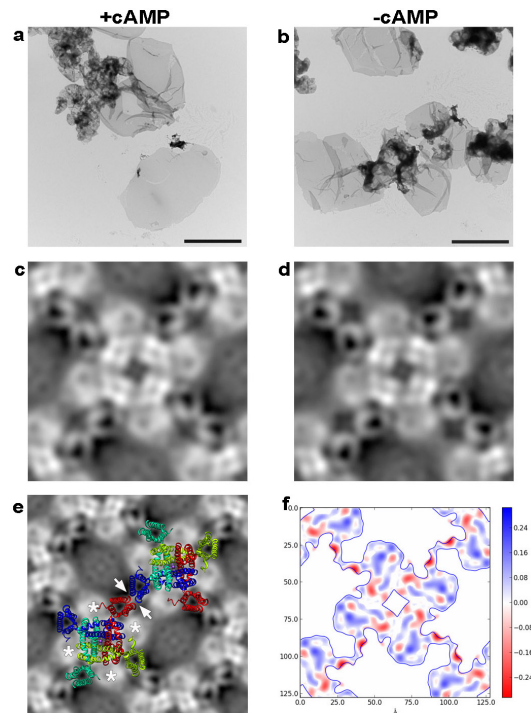
Figure 48: 2D crystallization and 2D projection maps of MloK1. MloK1 was crystallized in the presence (left panels) and absence (right panels) of cAMP ligand.

(a, b) Large, well-ordered, vesicular 2D crystals with diameters over 5 µm and low mosaicity were obtained. Scale bars correspond to 5 µm. (c, d) The projection maps of plunge-frozen 2D crystals vitrified in buffer solution. Shown is one crystal unit cell $P42_12$ symmetry, where one white MloK1 protein tetramer is located at the center of each map, and a second tetramer is located in the corners of the map in opposite orientation. The unit cell dimensions were $a = b = 131$ Å, $\gamma = 90°$ for crystals grown in the presence of ligand (c), and $a = b = 130$ Å, $\gamma = 90°$ for ligand-free channels (d). (e) Projection of the X-ray structure of the MloK1 transmembrane region (PDB ID 2ZD9) superimposed on the projection map of MloK1 with ligand, (c). Each monomer is depicted in a different color. Crystal contacts are indicated by white arrows. The CNBD densities of one tetramer are indicated by white stars. (f) Significant differences of the map shown in (c) minus the map from (d) (see Methods section). Densities corresponding to MloK1 with ligand are in blue, and without in red. Maps (c), (d), and (f) show one crystal unit cell, map (e) shows 1.5×1.5 unit cells.
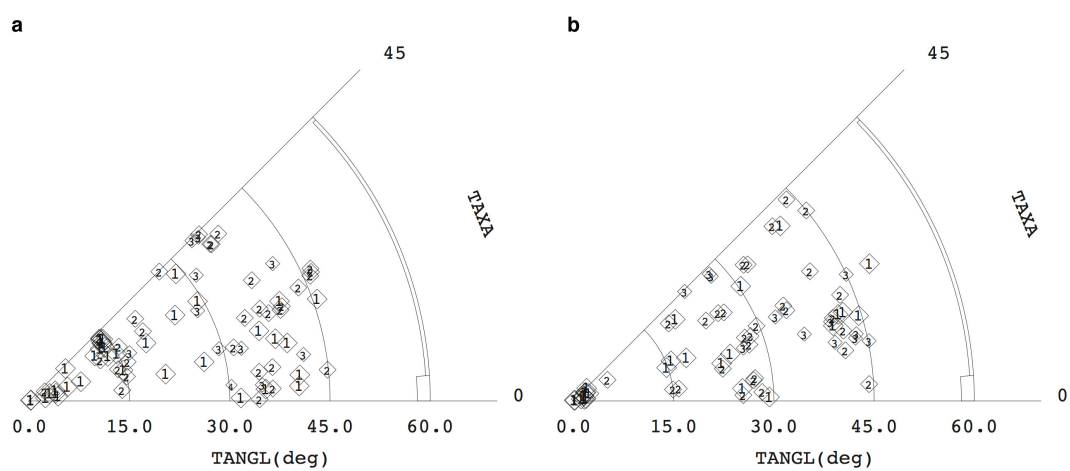
Figure 49: The tilt angle distribution plots from the 3D merging process in 2dx, for MloK1 cryo-EM data in the (a) presence and (b) absence of cAMP. The plots are produced within 2dx and show the image distributions according to their tilt angles (TANGL) in the asymmetric triangle. Each symbol represents one image; the number in the symbols corresponds to the significance of the image in the 3D dataset [Cheng & Yeager 2004].
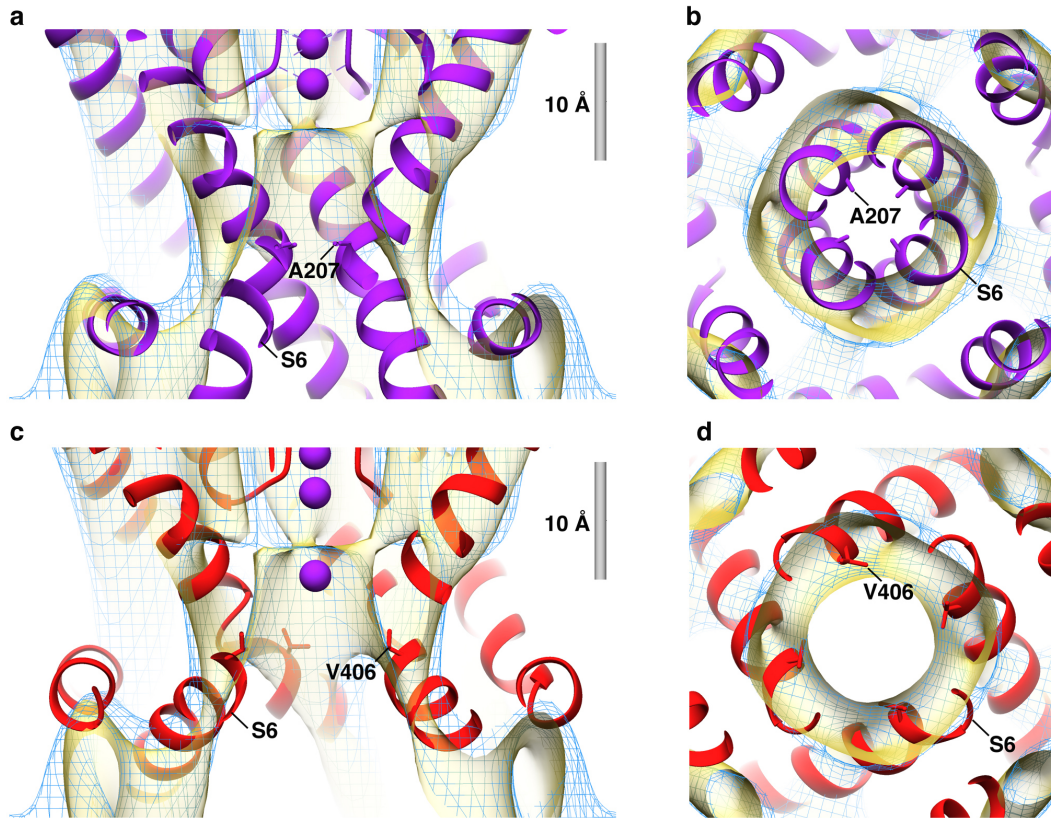
Figure 50: : X-ray structures of MloK1 (purple, PDB ID 2ZD9) in the closed conformation, and Shaker Kv1.2 (red, PDB ID 2A79) in the open conformation were fitted to the cryo-EM maps of cAMP-bound (mesh, blue) and cAMP-free (solid, yellow) MloK1.
(a) Helix S6 of the MloK1 X-ray structure is much more narrow at the helix bundle crossing area than the cryo-EM map.
(b) Cross-section through the helix bundle crossing region (A207) of MloK1.
(c) Helix S6 of Kv1.2 is wide open below the helix bundle crossing, in agreement with the cryo-EM map.
(d) Slice through constriction region of Kv1.2 (V406) corresponding to the A207 area in MloK1. For both MloK1 cryo-EM maps, this region has a similar diameter as in Kv1.2.
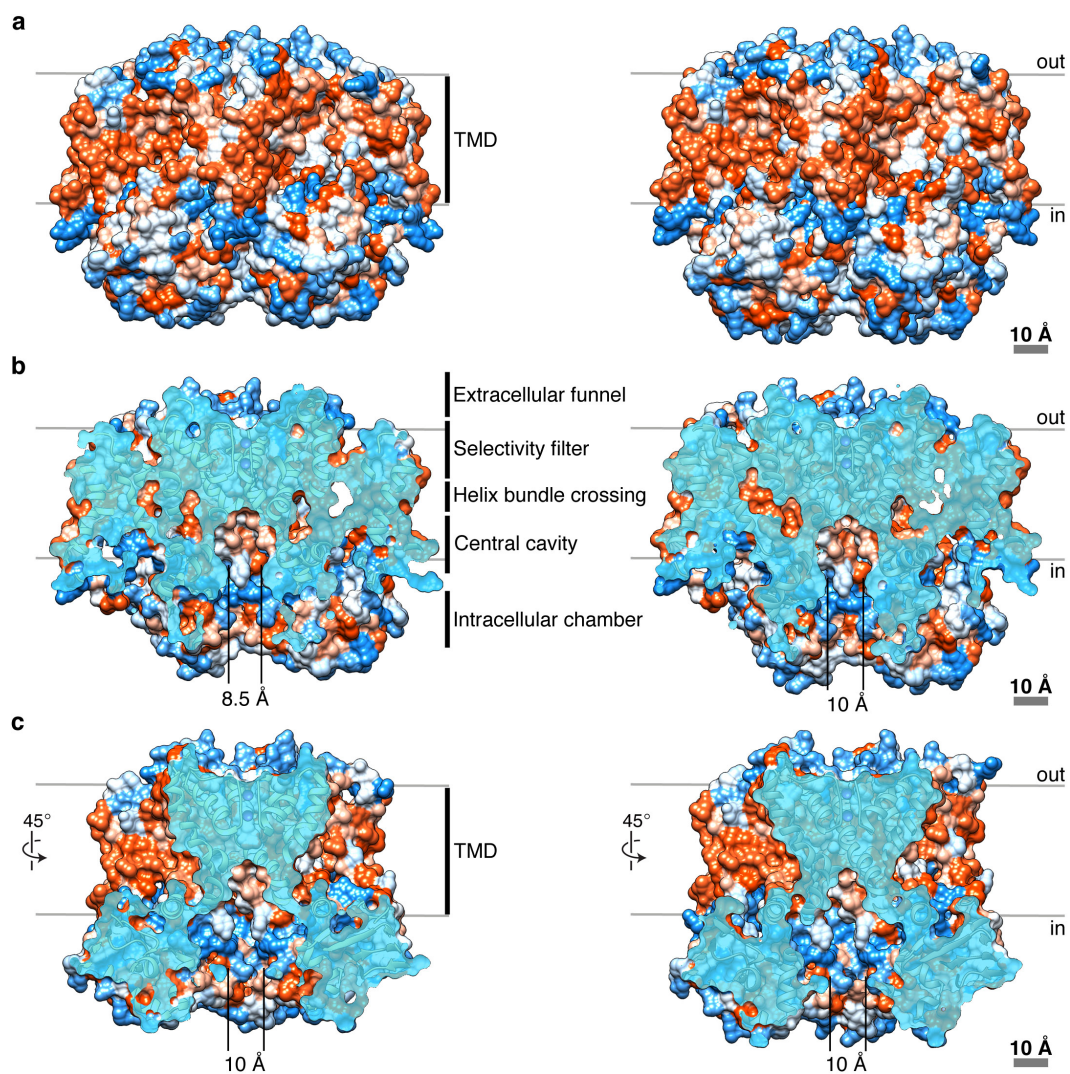
Figure 51: Electrostatic surface representation of liganded (left) and ligand-free (right) MloK1 models.
(a) Surface representation of fitted MloK1 models.
(b) Cut-open view, revealing the central cavity, which is accessible for K+ in both forms of the channel. All main pore elements are indicated.
(c) Structures from (b) rotated by 45° around the y-axis.
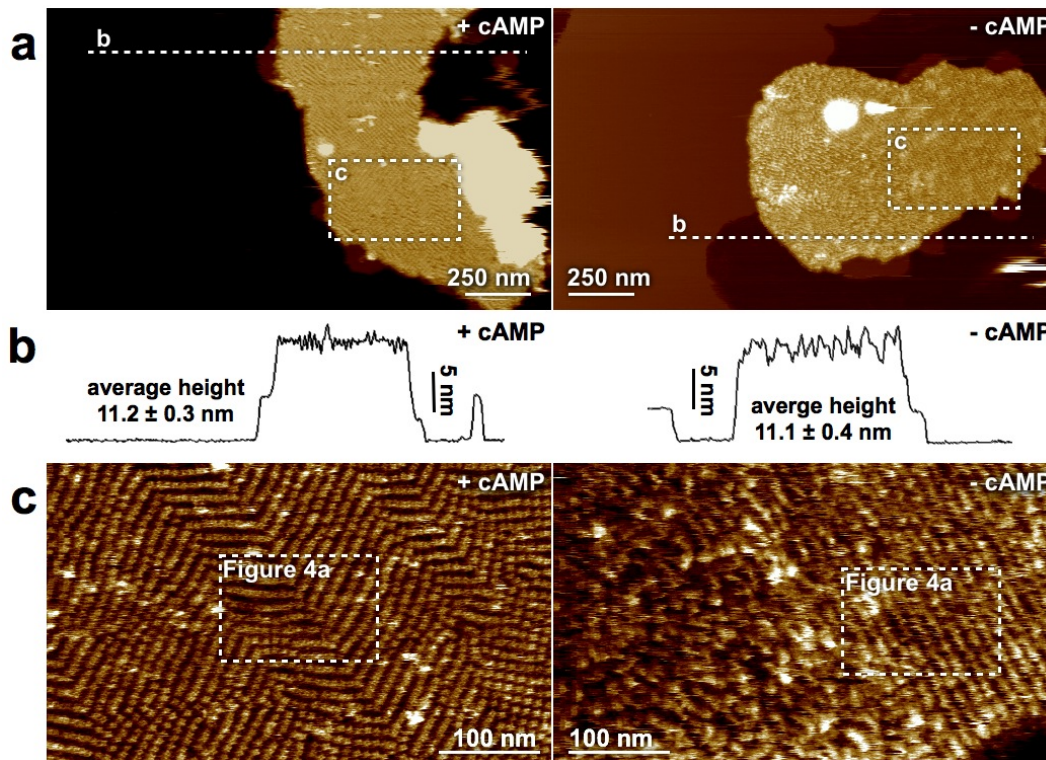
Figure 52: (a) Overview topographs of MloK1 channels reconstituted in lipid membranes with (left) and without (right) bound cAMP. The freshly adsorbed membranes showed MloK1 crystals of diameters up to several micrometers.

(b) Height profile of MloK1 membranes along the dashed lines in (a). The Cross-section analysis of MloK1 reconstitution membranes had very similar average height for the cAMP-bound 11.2 ± 0.3 nm (n = 50) (left) and cAMP-free 11.1 ± 0.4 nm (n = 50) (right) MloK1 membranes.

(c) Medium-resolution AFM topographs of Mlok1 in presence (left) and absence (right) of cAMP. In the +cAMP conformation single MloK1 channels could readily be observed, whereas the -cAMP conformation (right) the molecules appear as rather featureless protrusions. Different crystal forms were observed by AFM, including apparent P121 (shown here) and $P42_12$ symmetries.

9. Conclusions

This dissertation is an article thesis, where the image processing of 2D crystals plays a key role. Hence all publications have made use of the image processing software `2dx`. They have either influenced its development (methods part) or profited from it (application part).

The three chapters following the Introduction (section 2-section 4) sketch the complete workflow of image processing of 2D crystals. In Image Processing of 2D Crystal Images the key concept of extracting the repeating structure from a crystal by Fourier analysis were shown. This includes the image corrections needed due to physical distortions of the crystal (unbending) or due to the imaging system (CTF correction). Merging of Image Data in Electron Crystallography links the experimental image acquisition and the Central Section Theorem. This mathematical concept allows to combine different projections to a 3D reconstruction. Additionally it provides the `2dx` users a description how to compute a 3D electron density map. These two chapters both act as a introduction to the image processing concepts as well as a user guide for `2dx`. Automation of Image Processing in Electron Crystallography then tries to take the user out of the processing loop. The book chapter describes the automatic processing of individual images and how the merging of this data only requires minimal user interactions. Further development in automation produced a processing workflow, where the sole user interaction is the selection of the images and starting a script. The presented automation of the merging the structural data gained from individual images to a 3D reconstruction is a big step towards high throughput processing. Whilst automation has simplified and speed up the processing of 2D crystal images, it also increased the risk of errors.

Electron crystallography is a time-consuming technique, especially the crystallization, where countless conditions have to be tested. We have now shown that we can speed up the image processing step of the workflow. Nevertheless electron crystallography cannot compete with X-Ray crystallography in terms of time. The duration from protein expression to finally structure determination takes much longer. But as has been pointed out, 2D crystals are reconstituted in a membrane, hence a natural environment. Therefore the protein is observed in a functional state. We have shown two examples in the application part of this thesis, where different conformations in 2D crystals were analyzed. In the study of the Secondary Citrate/Sodium Symporter CitS, structural changes induced by the substrates Na^+ and citrate were revealed. The other example is the structural analysis of the cyclic nucleotide-modulated potassium channel MloK1. This study showed the conformational changes due to binding to the ligand cAMP and provided insight into the gating mechanism of the channel. Both structural analysis made use of the "significant difference map" method to confirm conformational changes. In the case of CitS, the substrate binding site could be identified. In the Mlok1 study this method

was more an aid until the 3D structures of both states, with or without ligand, were solved. However the significant difference map supported the changes seen in the 3D density maps. This project also benefited from parts of the automated image processing.

The last publication in the methods parts shows how single particle processing on top of the conventional 2D crystal workflow can yield higher resolved 3D structures. The local tilt geometry of a crystal was refined iteratively by using small crystalline patches. In each iteration local averages were compared to projections of the previous 3D reconstruction with varying orientation. This local averaging made use of the neighborhood correlation that exists in 2D crystals and thereby overcame the low SNR issue, that prohibited classical single particle processing. The processing of over 200'000 particle images, required a high performance cluster computer, where the particles were equally distributed among the nodes. The method was used to refine the 3D structure of Mlok1 and has improved its structural details. Single particle 3D reconstruction has only been applied to this 2D crystal so far. We hope this method could help deriving high-resolution structures from 2D crystal in future studies.

There has been a stagnation in methods development for 2D crystals the past years. Mainly because of the improvements in the single particle EM processing which have increased the resolution of the resulting structures. In addition the laborious and time intensive task of crys-tallization, where there is not guarantee, that the protein will form a 2D crystal hampered the development of methods. Nevertheless electron crystallography has its strengths. It still yields the structures with the highest resol-tuion in cryo-EM and captures the membrane protein in the membrane. There just needs to be a speed up of the whole electron crystallography pipeline. We have shown how automation can do this for image processing. Automation in other steps is also being developed. Ioan Iacovache et al. [Iacovache *et al.* 2010] have introduced a fully automated robot for detergent removal. Coudray et al.[Coudray *et al.* 2011] have introduced automated screening of 2D crystallization trials at room temperature. There exist also automated acquisition systems for electron microscopy with Leginon [Carragher *et al.* 2000] as one of the most promising, but they are still working on the 2D crystal determination. The recent development in complementary metal-oxide semiconductor (CMOS) cameras allow counting of single electrons and have made photographic film as recording medium obsolete. This not only speeds up data collection it also facilitates direct 2D crystal screening. These new direct detectors have rapid data read outs, which allow to record consecutive frames instead of one image. By aligning these frames to each other one can compensate for beam induced motion of the sample [Bai *et al.* 2013, Campbell *et al.* 2012, Li *et al.* 2013]. Future studies using direct detector with 2D crystal will likely provide high-resolution 3D structures. This will help to solve the numerous unknown membrane protein structures.

References

[Abe *et al.* 2011] Kazuhiro Abe, Kazutoshi Tani and Yoshinori Fujiyoshi. *Conformational rearrangement of gastric H+, K+-ATPase induced by an acid suppressant.* Nature communications, vol. 2, page 155, 2011. (Cited on page 76.)

[Abeyrathne *et al.* 2012] P. D. Abeyrathne, M. Arheit, F. Kebbel, D. Castano-Diez, K. N. Goldie, M. Chami, H. Stahlberg, L. Renault and W. Kühlbrandt. *Analysis of 2-D Crystals of Membrane Proteins by Electron Microscopy.* In H. Egelman Editor-in Chief: Edward, editeur, Comprehensive Biophysics, volume 1, pages 277–310. Elsevier, Amsterdam, 1 édition, 2012. (Cited on pages 1 and 3.)

[Abramson & Wright 2009] J. Abramson and E. M. Wright. *Structure and function of Na(+)-symporters with inverted repeats.* Curr Opin Struct Biol, vol. 19, no. 4, pages 425–32, 2009. (Cited on page 125.)

[Agard 1983] D. A. Agard. *A least-squares method for determining structure factors in three-dimensional tilted-view reconstructions.* Journal of Molecular Biology, vol. 167, no. 4, pages 849–52, 1983. (Cited on page 153.)

[Altieri *et al.* 2008] S. L. Altieri, G. M. Clayton, W. R. Silverman, A. O. Olivares, E. M. De la Cruz, L. R. Thomas and J. H. Morais-Cabral. *Structural and energetic analysis of activation by a cyclic nucleotide binding domain.* Journal of Molecular Biology, vol. 381, no. 3, pages 655–69, 2008. (Cited on pages 138, 141, 143 and 153.)

[Anderson & Olkin 1985] T. W. Anderson and I. Olkin. *Maximum-likelihood estimation of the parameters of a multivariate normal distribution Merging of Image Data in Electron Crystallography.* In Electron Crystallography of Soluble and Membrane Proteins, volume 70, pages 147–171. Springer, 1985. (Cited on page 107.)

[Appel *et al.* 2009a] M. Appel, D. Hizlan, K. R. Vinothkumar, C. Ziegler and W. Kuhlbrandt. *Conformations of NhaA, the Na+/H+ exchanger from Escherichia coli, in the pH-activated and ion-translocating states.* Journal of Molecular Biology, vol. 388, no. 3, pages 659–72, 2009. (Cited on pages 80 and 125.)

[Appel *et al.* 2009b] M. Appel, D. Hizlan, K. R. Vinothkumar, C. Ziegler and W. Kühlbrandt. *Conformations of NhaA, the Na/H exchanger from Escherichia coli, in the pH-activated and ion-translocating states.* Journal of Molecular Biology, vol. 386, no. 2, pages 351–65, 2009. (Cited on page 155.)

[Arheit *et al.* 2013a] M. Arheit, D. Castano-Diez, R. Thierry, P. Abeyrathne, B. R. Gipson and H. Stahlberg. *Merging of image data in electron crystallography.* Methods Mol Biol, vol. 955, pages 195–209, 2013. (Cited on pages 37, 99 and 153.)

[Arheit *et al.* 2013b] M. Arheit, D. Castano-Diez, R. Thierry, B. R. Gipson, X. Zeng and H. Stahlberg. *Automation of image processing in electron crystallography.* Methods Mol Biol, vol. 955, pages 313–30, 2013. (Cited on pages 52 and 70.)

[Arheit *et al.* 2013c] M. Arheit, D. Castano-Diez, R. Thierry, B. R. Gipson, X. Zeng and H. Stahlberg. *Image processing of 2D crystal images.* Methods Mol Biol, vol. 955, pages 171–94, 2013. (Cited on pages 15, 98, 99, 140 and 153.)

[Arheit *et al.* 2016] M. Arheit, C. Paulino, D. Castano-Diez, W. Kühlbrandt and H. Stahlberg. *Conformational Changes of Proteins Detected byElectron Crystallography: Difference Map Calculation and Analysis*, 2016. Manuscript in preparation. (Cited on page 75.)

[Arinaminpathy *et al.* 2009] Y. Arinaminpathy, E. Khurana, D. M. Engelman and M. B. Gerstein. *Computational analysis of membrane proteins: the largest class of drug targets.* Drug discovery today, vol. 14, no. 23-24, pages 1130–1135, 2009. (Cited on page 1.)

[Bai *et al.* 2013] X.-C. Bai, I.S. Fernandez, G. McMullan and S. H. Scheres. *Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles.* Elife, vol. in press, 2013. (Cited on pages 7, 108, 119 and 164.)

[Biskup *et al.* 2007] C. Biskup, J. Kusch, E. Schulz, V. Nache, F. Schwede, F. Lehmann, V. Hagen and K. Benndorf. *Relating ligand binding to activation gating in CNGA2 channels.* Nature, vol. 446, no. 7134, pages 440–3, 2007. (Cited on page 149.)

[Boudker *et al.* 2007] O. Boudker, R. M. Ryan, D. Yernool, K. Shimamoto and E. Gouaux. *Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter.* Nature, vol. 445, no. 7126, pages 387–93, 2007. (Cited on page 133.)

[Bruening-Wright *et al.* 2002] A. Bruening-Wright, M. A. Schumacher, J. P. Adelman and J. Maylie. *Localization of the activation gate for small conductance Ca2+-activated K+ channels.* Journal of Neuroscience, vol. 22, no. 15, pages 6499–506, 2002. (Cited on page 149.)

[Brunger *et al.* 1998] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S.

Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren. *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallographica. Section D, Biological Crystallography, vol. 54, no. Pt 5, pages 905–21, 1998. (Cited on page 154.)

[Campbell *et al.* 2012] M. G. Campbell, A. Cheng, A. F. Brilot, A. Moeller, D. Lyumkis, D. Veesler, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher and N. Grigorieff. *Movies of ice-embedded particles enhance resolution in electron cryo-microscopy.* Structure, vol. 20, no. 11, pages 1823–8, 2012. (Cited on pages 7, 76, 108, 119 and 164.)

[Carragher *et al.* 2000] B. Carragher, N. Kisseberth, D. Kriegman, R. A. Milligan, C. S. Potter, J. Pulokas and A. 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 60, 108 and 164.)

[Castano-Diez *et al.* 2008] D. Castano-Diez, D. Moser, A. Schoenegger, S. Pruggnaller and A. S. Frangakis. *Performance evaluation of image processing algorithms on the GPU.* Journal of Structural Biology, vol. 164, no. 1, pages 153–60, 2008. (Cited on page 111.)

[Castaño-Díez *et al.* 2012] D. Castaño-Díez, M. Kudryashev, M. Arheit and H. Stahlberg. *Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments.* Journal of structural biology, 2012. (Cited on page 84.)

[Catterall 1986] W. A. Catterall. *Voltage dependent gating of sodium channels: Correlating structure and function.* Trends in Neurosciences, vol. 9, pages 7–10, 1986. (Cited on page 150.)

[CCP4 1994] CCP4. *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, vol. 50, no. Pt 5, pages 760–3, 1994. (Cited on page 153.)

[Cheng & Yeager 2004] A. Cheng and M. Yeager. *A graphical representation of image quality for three-dimensional structure analysis of two-dimensional crystals.* Acta Crystallogr A, vol. 60, no. Pt 4, pages 351–4, 2004. (Cited on page 159.)

[Chiu *et al.* 2007] P. L. Chiu, M. D. Pagel, J. Evans, H. T. Chou, X. Zeng, B. Gipson, H. Stahlberg and C. M. Nimigean. *The structure of the prokaryotic cyclic nucleotide-modulated potassium channel MloK1 at 16 A resolution.* Structure, vol. 15, no. 9, pages 1053–64, 2007. (Cited on pages 138, 149 and 150.)

[Chow *et al.* 2012] S. S. Chow, F. Van Petegem and E. A. Accili. *Energetics of cyclic AMP binding to HCN channel C terminus reveal negative cooperativity.* Journal of Biological Chemistry, vol. 287, no. 1, pages 600–6, 2012. (Cited on page 149.)

[Clayton *et al.* 2004] G. M. Clayton, W. R. Silverman, L. Heginbotham and J. H. Morais-Cabral. *Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel.* Cell, vol. 119, no. 5, pages 615–27, 2004. (Cited on page 138.)

[Clayton *et al.* 2008] G. M. Clayton, S. Altieri, L. Heginbotham, V. M. Unger and J. H. Morais-Cabral. *Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel.* Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 5, pages 1511–5, 2008. (Cited on pages 138, 139, 140, 141, 145, 150, 153 and 154.)

[Clayton *et al.* 2009] G. M. Clayton, S. G. Aller, J. Wang, V. Unger and J. H. Morais-Cabral. *Combining electron crystallography and X-ray crystallography to study the MlotiK1 cyclic nucleotide-regulated potassium channel.* Journal of Structural Biology, vol. 167, no. 3, pages 220–6, 2009. (Cited on pages 139 and 147.)

[Coudray *et al.* 2011] N. Coudray, G. Hermann, D. Caujolle-Bert, A. Karathanou, F. Erne-Brand, J. L. Buessler, P. Daum, J. M. Plitzko, M. Chami, U. Mueller, H. Kihl, J. P. Urban, A. Engel and H. W. Remigy. *Automated screening of 2D crystallization trials using transmission electron microscopy: a high-throughput tool-chain for sample preparation and microscopic analysis.* Journal of Structural Biology, vol. 173, no. 2, pages 365–74, 2011. Coudray, Nicolas Hermann, Gilles Caujolle-Bert, Daniel Karathanou, Argyro Erne-Brand, Francoise Buessler, Jean-Luc Daum, Pamela Plitzko, Juergen M Chami, Mohamed Mueller, Urs Kihl, Hubert Urban, Jean-Philippe Engel, Andreas Remigy, Herve-W Research Support, Non-U.S. Gov't United States Journal of structural biology J Struct Biol. 2011 Feb;173(2):365-74. doi: 10.1016/j.jsb.2010.09.019. Epub 2010 Sep 22. (Cited on page 164.)

[Craven & Zagotta 2006] K. B. Craven and W. N. Zagotta. *CNG and HCN channels: two peas, one pod.* Annual Review of Physiology, vol. 68, pages 375–401, 2006. (Cited on pages 137 and 138.)

[Crowther *et al.* 1996] R. A. Crowther, R. Henderson and J. M. Smith. *MRC image processing programs.* Journal of Structural Biology, vol. 116, no. 1, pages 9–16, 1996. (Cited on pages 53, 64, 67, 80, 98, 101, 153 and 155.)

[Cuello *et al.* 2010] L. G. Cuello, V. Jogini, D. M. Cortes and E. Perozo. *Struc-*

tural mechanism of C-type inactivation in K(+) channels. Nature, vol. 466, no. 7303, pages 203–8, 2010. (Cited on page 149.)

[Cukkemane *et al.* 2007] A. Cukkemane, B. Gruter, K. Novak, T. Gensch, W. Bonigk, T. Gerharz, U. B. Kaupp and R. Seifert. *Subunits act independently in a cyclic nucleotide-activated K(+) channel.* EMBO Reports, vol. 8, no. 8, pages 749–55, 2007. (Cited on pages 147 and 149.)

[Dagum & Menon 1998] L. Dagum and R. Menon. *OpenMP: an industry standard API for shared-memory programming.* IEEE Comput Sci Eng, vol. 5, no. 1, pages 46–55, 1998. (Cited on page 108.)

[De Feo *et al.* 2009] C. J. De Feo, S. G. Aller, G. S. Siluvai, N. J. Blackburn and V. M. Unger. *Three-dimensional structure of the human copper transporter hCTR1.* Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 11, pages 4237–4242, 2009. 420HD Times Cited:81 Cited References Count:48. (Cited on page 71.)

[De Rosier & Klug 1968] DJ De Rosier and A Klug. *Reconstruction of three dimensional structures from electron micrographs.* Nature, vol. 217, no. 5124, pages 130–134, 1968. (Cited on page 9.)

[Dekker & Yellen 2006] J. P. Dekker and G. Yellen. *Cooperative gating between single HCN pacemaker channels.* Journal of General Physiology, vol. 128, no. 5, pages 561–7, 2006. (Cited on page 149.)

[del Camino & Yellen 2001] D. del Camino and G. Yellen. *Tight steric closure at the intracellular activation gate of a voltage-gated K(+) channel.* Neuron, vol. 32, no. 4, pages 649–56, 2001. (Cited on page 147.)

[Dobrowolski *et al.* 2010] A. Dobrowolski, F. Fusetti and J. S. Lolkema. *Cross-linking of trans reentrant loops in the Na(+)-citrate transporter CitS of Klebsiella pneumoniae.* Biochemistry, vol. 49, no. 21, pages 4509–15, 2010. (Cited on pages 129 and 131.)

[Downing 1991] Kenneth H Downing. *Spot-scan imaging in transmission electron microscopy.* Science, vol. 251, no. 4989, pages 53–59, 1991. (Cited on page 6.)

[Doyle *et al.* 1998] D.A. Doyle, J.M. Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait and R. MacKinnon. *The structure of the potassium channel: molecular basis of K+ conduction and selectivity.* Science, vol. 280, no. 5360, pages 69–77, 1998. (Cited on page 147.)

[Elmlund & Elmlund 2012] D. Elmlund and H. Elmlund. *SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles.* Journal of Structural

Biology, vol. 180, no. 3, pages 420–7, 2012. (Cited on page 98.)

[Emsley *et al.* 2010] P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan. *Features and development of Coot.* Acta Crystallographica. Section D, Biological Crystallography, vol. 66, no. Pt 4, pages 486–501, 2010. (Cited on page 153.)

[Engel & Gaub 2008] A. Engel and H. E. Gaub. *Structure and Mechanics of Membrane Proteins.* Annual Review of Biochemistry, vol. 77, no. 1, pages 127–148, 2008. (Cited on page 1.)

[Falkner & Schroder 2013] B. Falkner and G. F. Schroder. *Cross-validation in cryo-EM-based structural modeling.* Proceedings of the National Academy of Sciences of the United States of America, vol. 110, no. 22, pages 8930–5, 2013. (Cited on page 154.)

[Flynn *et al.* 2001] G. E. Flynn, Jr. Johnson J. P. and W. N. Zagotta. *Cyclic nucleotide-gated channels: shedding light on the opening of a channel pore.* Nature Reviews. Neuroscience, vol. 2, no. 9, pages 643–51, 2001. (Cited on page 149.)

[Flynn *et al.* 2007] G. E. Flynn, K. D. Black, L. D. Islas, B. Sankaran and W. N. Zagotta. *Structure and rearrangements in the carboxy-terminal region of SpIH channels.* Structure, vol. 15, no. 6, pages 671–82, 2007. (Cited on page 138.)

[Forrest & Rudnick 2009] L. R. Forrest and G. Rudnick. *The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters.* Physiology (Bethesda), vol. 24, pages 377–86, 2009. (Cited on page 124.)

[Forrest *et al.* 2011] L. R. Forrest, R. Kramer and C. Ziegler. *The structural basis of secondary active transport mechanisms.* Biochim Biophys Acta, vol. 1807, no. 2, pages 167–88, 2011. (Cited on pages 124, 125 and 133.)

[Frank 1975] J. Frank. *Averaging of low exposure electron micrographs of non-periodic objects.* Ultramicroscopy, vol. 1, no. 2, pages 159–62, 1975. (Cited on pages 98 and 104.)

[Gibbs 1898] J Willard Gibbs. *Fourier's series.* Nature, vol. 59, page 200, 1898. (Cited on pages 80 and 86.)

[Gipson *et al.* 2007a] B. Gipson, X. Zeng and H. Stahlberg. *2dx_merge: Data management and merging for 2D crystal images.* Journal of Structural Biology, vol. 160, no. 3, pages 375–84, 2007. (Cited on pages 53, 60, 61, 78, 84, 89, 126, 134 and 153.)

References

[Gipson *et al.* 2007b] B. Gipson, X. Zeng, Z. Zhang and H. Stahlberg. *2dx–User-friendly image processing for 2D crystals.* Journal of Structural Biology, vol. 157, no. 1, pages 64–72, 2007. (Cited on pages 53, 60, 78, 84, 89, 98, 101, 126, 134, 140 and 153.)

[Gipson *et al.* 2011] B. R. Gipson, D. J. Masiel, N. D. Browning, J. Spence, K. Mitsuoka and H. Stahlberg. *Automatic recovery of missing amplitudes and phases in tilt-limited electron crystallography of two-dimensional crystals.* Phys Rev E Stat Nonlin Soft Matter Phys, vol. 84, no. 1 Pt 1, page 011916, 2011. (Cited on pages 10 and 68.)

[Gonen *et al.* 2004] T. Gonen, P. Sliz, J. Kistler, Y. Cheng and T. Walz. *Aquaporin-0 membrane junctions reveal the structure of a closed water pore.* Nature, vol. 429, no. 6988, pages 193–7, 2004. (Cited on page 2.)

[Gonen *et al.* 2005] T. Gonen, Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison and T. Walz. *Lipid-protein interactions in double-layered two-dimensional AQP0 crystals.* Nature, vol. 438, no. 7068, pages 633–8, 2005. (Cited on page 76.)

[Goswami *et al.* 2011] P. Goswami, C. Paulino, D. Hizlan, J. Vonck, O. Yildiz and W. Kühlbrandt. *Structure of the archaeal Na(+)/H(+) antiporter NhaP1 and functional role of transmembrane helix 1.* The EMBO journal, vol. 30, no. 2, pages 439–449, 2011. (Cited on page 125.)

[Grigorieff 2007] N. Grigorieff. *FREALIGN: high-resolution refinement of single particle structures.* Journal of Structural Biology, vol. 157, no. 1, pages 117–25, 2007. (Cited on page 98.)

[Gyobu *et al.* 2004] N. Gyobu, K. Tani, Y. Hiroaki, A. Kamegawa, K. Mitsuoka and Y. Fujiyoshi. *Improved specimen preparation for cryo-electron microscopy using a symmetric carbon sandwich technique.* Journal of structural biology, vol. 146, no. 3, pages 325–333, 2004. (Cited on page 6.)

[Henderson & Unwin 1975] R. Henderson and P. N. Unwin. *Three-dimensional model of purple membrane obtained by electron microscopy.* Nature, vol. 257, no. 5521, pages 28–32, 1975. (Cited on pages 3, 4 and 6.)

[Herz *et al.* 2009] K. Herz, A. Rimon, G. Jeschke and E. Padan. *Beta-sheet-dependent dimerization is essential for the stability of NhaA Na+/H+ antiporter.* The Journal of Biological Chemistry, vol. 284, no. 10, pages 6337–47, 2009. (Cited on page 133.)

[Heymann 2001] J. B. Heymann. *Bsoft: image and molecular processing in electron microscopy.* Journal of Structural Biology, vol. 133, no. 2-3, pages 156–69,

2001. (Cited on page 98.)

[Hirai *et al.* 1999] T. Hirai, K. Murata, K. Mitsuoka, Y. Kimura and Y. Fujiyoshi. *Trehalose embedding technique for high-resolution electron crystallography: application to structural study on bacteriorhodopsin.* Journal of Electron Microscopy, vol. 48, no. 5, pages 653–8, 1999. (Cited on page 153.)

[Hohn *et al.* 2007] M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams and S. 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 page 98.)

[Hovmöller 1992] S. Hovmöller. *CRISP: crystallographic image processing on a personal computer.* Ultramicroscopy, vol. 41, pages 121–135, 1992. (Cited on page 68.)

[Hunte *et al.* 2005] C. Hunte, E. Screpanti, M. Venturi, A. Rimon, E. Padan and H. Michel. *Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH.* Nature, vol. 435, no. 7046, pages 1197–202, 2005. (Cited on page 128.)

[Iacovache *et al.* 2010] I. Iacovache, M. Biasini, J. Kowal, W. Kukulski, M. Chami, F. G. van der Goot, A. Engel and H. W. Remigy. *The 2DX robot: a membrane protein 2D crystallization Swiss Army knife.* Journal of Structural Biology, vol. 169, no. 3, pages 370–8, 2010. Iacovache, Ioan Biasini, Marco Kowal, Julia Kukulski, Wanda Chami, Mohamed van der Goot, F Gisou Engel, Andreas Remigy, Herve-W Research Support, Non-U.S. Gov't United States Journal of structural biology J Struct Biol. 2010 Mar;169(3):370-8. doi: 10.1016/j.jsb.2009.12.001. Epub 2009 Dec 4. (Cited on page 164.)

[Jardetzky 1966] O. Jardetzky. *Simple allosteric model for membrane pumps.* Nature, vol. 211, no. 5052, pages 969–70, 1966. (Cited on page 125.)

[Jensen *et al.* 2012] M. O. Jensen, V. Jogini, D. W. Borhani, A. E. Leffler, R. O. Dror and D. E. Shaw. *Mechanism of voltage gating in potassium channels.* Science, vol. 336, no. 6078, pages 229–33, 2012. (Cited on page 147.)

[Jiang *et al.* 2003] Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B. T. Chait and R. MacKinnon. *X-ray structure of a voltage-dependent K+ channel.* Nature, vol. 423, no. 6935, pages 33–41, 2003. (Cited on page 150.)

[Jost *et al.* 2003] G. Jost, H. Jin, D. an Mey and F. 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 109.)

References

[Kaupp & Seifert 2001] U. B. Kaupp and R. Seifert. *Molecular diversity of pacemaker ion channels.* Annual Review of Physiology, vol. 63, pages 235–57, 2001. (Cited on page 137.)

[Kaupp & Seifert 2002] U. B. Kaupp and R. Seifert. *Cyclic nucleotide-gated ion channels.* Physiological Reviews, vol. 82, no. 3, pages 769–824, 2002. (Cited on page 137.)

[Kebbel *et al.* 2012] F. Kebbel, M. Kurz, M. G. Grutter and H. Stahlberg. *Projection structure of the secondary citrate/sodium symporter CitS at 6 A resolution by electron crystallography.* Journal of Molecular Biology, vol. 418, no. 1-2, pages 117–26, 2012. (Cited on pages 125, 126, 128 and 129.)

[Kebbel *et al.* 2013] F. Kebbel, M. Kurz, M. Arheit, M. G. Grütter and H. 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 pages 76, 77, 84, 85, 87, 88, 92, 124 and 134.)

[Kelly & Gross 2003] B. L. Kelly and A. Gross. *Potassium channel gating observed with site-directed mass tagging.* Nature Structural Biology, vol. 10, no. 4, pages 280–4, 2003. (Cited on page 147.)

[Kim *et al.* 2012] C. S. Kim, P. Y. Chang and D. Johnston. *Enhancement of dorsal hippocampal activity by knockdown of HCN1 channels leads to anxiolytic- and antidepressant-like behaviors.* Neuron, vol. 75, no. 3, pages 503–16, 2012. (Cited on page 137.)

[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 76, 77, 84, 88, 111 and 137.)

[Krishnamurthy *et al.* 2009] H. Krishnamurthy, C. L. Piscitelli and E. Gouaux. *Unlocking the molecular secrets of sodium-coupled transporters.* Nature, vol. 459, no. 7245, pages 347–55, 2009. (Cited on page 125.)

[Krupnik *et al.* 2011] T. Krupnik, A. Dobrowolski and J. S. Lolkema. *Cross-linking of dimeric CitS and GltS transport proteins.* Mol Membr Biol, vol. 28, no. 5, pages 243–53, 2011. (Cited on pages 125 and 129.)

[Kühlbrandt 2014] Werner Kühlbrandt. *The resolution revolution.* Science, vol. 343, no. 6178, pages 1443–1444, 2014. (Cited on page 76.)

[Kunji *et al.* 2000] E. R. Kunji, S. von Gronau, D. Oesterhelt and R. 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.* Proc. Natl. Acad. Sci. USA, vol. 97, no. 9, pages 4637–4642, 2000. (Cited on pages 61 and 104.)

[Kuo & Glaeser 1975] I. A. M. Kuo and R. M. Glaeser. *Development of methodology for low exposure, high resolution electron microscopy of biological specimens.* Ultramicroscopy, vol. 1, no. 1, pages 53–66, 1975. (Cited on page 6.)

[Kusch *et al.* 2010] J. Kusch, C. Biskup, S. Thon, E. Schulz, V. Nache, T. Zimmer, F. Schwede and K. Benndorf. *Interdependence of receptor activation and ligand binding in HCN2 pacemaker channels.* Neuron, vol. 67, no. 1, pages 75–85, 2010. (Cited on page 149.)

[Kusch *et al.* 2012] J. Kusch, S. Thon, E. Schulz, C. Biskup, V. Nache, T. Zimmer, R. Seifert, F. Schwede and K. Benndorf. *How subunits cooperate in cAMP-induced activation of homotetrameric HCN2 channels.* Nature Chemical Biology, vol. 8, no. 2, pages 162–9, 2012. (Cited on page 149.)

[Law *et al.* 2007] C. J. Law, Q. Yang, C. Soudant, P. C. Maloney and D. N. Wang. *Kinetic evidence is consistent with the rocker-switch mechanism of membrane transport by GlpT.* Biochemistry, vol. 46, no. 43, pages 12190–7, 2007. (Cited on page 125.)

[Li *et al.* 2013] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard and Y. Cheng. *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM.* Nature Methods, vol. 10, no. 6, pages 584–90, 2013. (Cited on pages 7, 108 and 164.)

[Lolicato *et al.* 2011] M. Lolicato, M. Nardini, S. Gazzarrini, S. Moller, D. Bertinetti, F. W. Herberg, M. Bolognesi, H. Martin, M. Fasolini, J. A. Bertrand, C. Arrigoni, G. Thiel and A. Moroni. *Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels.* Journal of Biological Chemistry, vol. 286, no. 52, pages 44811–20, 2011. (Cited on page 138.)

[Lolkema & Slotboom 1998] J. S. Lolkema and D. J. Slotboom. *Estimation of structural similarity of membrane proteins by hydropathy profile alignment.* Mol Membr Biol, vol. 15, no. 1, pages 33–42, 1998. (Cited on page 124.)

[Lolkema & Slotboom 2003] J. S. Lolkema and D. J. Slotboom. *Classification of 29 families of secondary transport proteins into a single structural class using*

hydropathy profile analysis. Journal of Molecular Biology, vol. 327, no. 5, pages 901–9, 2003. (Cited on page 124.)

[Lolkema *et al.* 1994] J. S. Lolkema, H. Enequist and M. E. van der Rest. *Transport of citrate catalyzed by the sodium-dependent citrate carrier of Klebsiella pneumoniae is obligatorily coupled to the transport of two sodium ions.* Eur J Biochem, vol. 220, no. 2, pages 469–75, 1994. (Cited on pages 125 and 133.)

[Lolkema *et al.* 2005] J. S. Lolkema, I. Sobczak and D. J. Slotboom. *Secondary transporters of the 2HCT family contain two homologous domains with inverted membrane topology and trans re-entrant loops.* FEBS Journal, vol. 272, no. 9, pages 2334–44, 2005. (Cited on page 125.)

[Lolkema 2006] J. S. Lolkema. *Domain structure and pore loops in the 2-hydroxycarboxylate transporter family.* Journal of Molecular Microbiology and Biotechnology, vol. 11, no. 6, pages 318–25, 2006. (Cited on page 125.)

[Long *et al.* 2007] S. B. Long, X. Tao, E. B. Campbell and R. MacKinnon. *Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment.* Nature, vol. 450, no. 7168, pages 376–82, 2007. (Cited on pages 139, 140 and 147.)

[Ludtke *et al.* 1999] S. J. Ludtke, P. R. Baldwin and W. 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 98.)

[Lyumkis *et al.* 2013] D. Lyumkis, A. F. Brilot, D. L. Theobald and N. Grigorieff. *Likelihood-based classification of cryo-EM images using FREALIGN.* Journal of Structural Biology, 2013. (Cited on page 119.)

[Mahler & Persson 2012] J. Mahler and I. Persson. *A study of the hydration of the alkali metal ions in aqueous solution.* Inorganic Chemistry, vol. 51, no. 1, pages 425–38, 2012. (Cited on page 145.)

[Mancusso *et al.* 2012] R. Mancusso, G. G. Gregorio, Q. Liu and D. N. Wang. *Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter.* Nature, 2012. (Cited on pages 125, 128, 129, 131 and 133.)

[Mari *et al.* 2011] S. A. Mari, J. Pessoa, S. Altieri, U. Hensen, L. Thomas, J. H. Morais-Cabral and D. J. Muller. *Gating of the MlotiK1 potassium channel involves large rearrangements of the cyclic nucleotide-binding domains.* Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 51, pages 20802–7, 2011. (Cited on pages 138 and 150.)

[Mindell & Grigorieff 2003] J. A. Mindell and N. Grigorieff. *Accurate determination*

of local defocus and specimen tilt in electron microscopy. Journal of Structural Biology, vol. 142, no. 3, pages 334–47, 2003. (Cited on pages 64, 71, 99, 102 and 153.)

[Mitsuoka *et al.* 1999] K. Mitsuoka, T. Hirai, K. Murata, A. Miyazawa, A. Kidera, Y. Kimura and Y. Fujiyoshi. *The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution.* Journal of Molecular Biology, vol. 286, no. 3, pages 861–882, 1999. (Cited on page 68.)

[Moscicka *et al.* 2009] K. B. Moscicka, T. Krupnik, E. J. Boekema and J. S. Lolkema. *Projection structure by single-particle electron microscopy of secondary transport proteins GltT, CitS, and GltS.* Biochemistry, vol. 48, no. 28, pages 6618–23, 2009. (Cited on page 126.)

[Nimigean *et al.* 2004] C. M. Nimigean, T. Shane and C. Miller. *A cyclic nucleotide modulated prokaryotic K+ channel.* Journal of General Physiology, vol. 124, no. 3, pages 203–10, 2004. (Cited on pages 138, 150 and 152.)

[Nogales *et al.* 1995] E. Nogales, S. G. Wolf, I. A. Khan, R. F. Ludueña and K. H. Downing. *Structure of tubulin at 6.5 A and location of the taxol-binding site.* Nature, vol. 375, no. 6530, pages 424–427, 1995. (Cited on pages 76 and 78.)

[Paoletti *et al.* 1999] P. Paoletti, E. C. Young and S. A. Siegelbaum. *C-Linker of cyclic nucleotide-gated channels controls coupling of ligand binding to channel gating.* Journal of General Physiology, vol. 113, no. 1, pages 17–34, 1999. (Cited on page 138.)

[Papazian & Bezanilla 1997] D. M. Papazian and F. Bezanilla. *How does an ion channel sense voltage?* News in Physiological Sciences, vol. 12, pages 203–210, 1997. (Cited on page 150.)

[Payandeh *et al.* 2011] J. Payandeh, T. Scheuer, N. Zheng and W. A. Catterall. *The crystal structure of a voltage-gated sodium channel.* Nature, vol. 475, no. 7356, pages 353–8, 2011. (Cited on page 140.)

[Penczek 2010] P. A. Penczek. *Resolution measures in molecular electron microscopy.* Methods Enzymol, vol. 482, pages 73–100, 2010. (Cited on page 115.)

[Perez *et al.* 2011] C. Perez, K. Khafizov, L. R. Forrest, R. Kramer and C. Ziegler. *The role of trimerization in the osmoregulated betaine transporter BetP.* EMBO Rep, vol. 12, no. 8, pages 804–10, 2011. (Cited on page 133.)

[Perozo *et al.* 1999] E. Perozo, D. M. Cortes and L. G. Cuello. *Structural rearrange-*

ments underlying K+-channel activation gating. Science, vol. 285, no. 5424, pages 73–8, 1999. (Cited on page 147.)

[Pettersen *et al.* 2004] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin. *UCSF Chimera–a visualization system for exploratory research and analysis.* Journal of Computational Chemistry, vol. 25, no. 13, pages 1605–12, 2004. Pettersen, Eric F Goddard, Thomas D Huang, Conrad C Couch, Gregory S Greenblatt, Daniel M Meng, Elaine C Ferrin, Thomas E P41-RR01081/RR/NCRR NIH HHS/United States Research Support, U.S. Gov't, P.H.S. United States Journal of computational chemistry J Comput Chem. 2004 Oct;25(13):1605-12. (Cited on pages 103, 136, 140 and 153.)

[Peuker *et al.* 2013] S. Peuker, A. Cukkemane, M. Held, F. Noe, U. B. Kaupp and R. Seifert. *Kinetics of ligand-receptor interaction reveals an induced-fit mode of binding in a cyclic nucleotide-activated protein.* Biophysical Journal, vol. 104, no. 1, pages 63–74, 2013. (Cited on pages 147 and 149.)

[Philippsen *et al.* 2003] A. Philippsen, A. D. Schenk, H. Stahlberg and A. Engel. *Iplt–image processing library and toolkit for the electron microscopy community.* Journal of Structural Biology, vol. 144, no. 1-2, pages 4–12, 2003. (Cited on page 68.)

[Philippsen *et al.* 2007a] A. Philippsen, H. A. Engel and A. Engel. *The contrast-imaging function for tilted specimens.* Ultramicroscopy, vol. 107, no. 2-3, pages 202–12, 2007. (Cited on page 9.)

[Philippsen *et al.* 2007b] A. Philippsen, A. D. Schenk, G. A. Signorell, V. Mariani, S. Berneche and A. Engel. *Collaborative EM image processing with the IPLT image processing library and toolbox.* Journal of Structural Biology, vol. 157, no. 1, pages 28–37, 2007. (Cited on page 68.)

[Pos & Dimroth 1996] K. M. Pos and P. Dimroth. *Functional properties of the purified Na(+)-dependent citrate carrier of Klebsiella pneumoniae: evidence for asymmetric orientation of the carrier protein in proteoliposomes.* Biochemistry, vol. 35, no. 3, pages 1018–26, 1996. (Cited on page 125.)

[Posson *et al.* 2013] D. J. Posson, J. G. McCoy and C. M. Nimigean. *The voltage-dependent gate in MthK potassium channels is located at the selectivity filter.* Nature Structural and Molecular Biology, vol. 20, no. 2, pages 159–66, 2013. (Cited on page 149.)

[Radermacher 1988] M. Radermacher. *Three-dimensional reconstruction of single particles from random and nonrandom tilt series.* Journal of Electron Mi-

croscopy Technique, vol. 9, no. 4, pages 359–94, 1988. (Cited on page 98.)

[Radivojac *et al.* 2013] Predrag Radivojac, Wyatt T Clark, Tal Ronnen Oron, Alexandra M Schnoes, Tobias Wittkop, Artem Sokolov, Kiley Graim, Christopher Funk, Karin Verspoor, Asa Ben-Hur*et al. A large-scale evaluation of computational protein function prediction.* Nature methods, vol. 10, no. 3, pages 221–227, 2013. (Cited on page 75.)

[Reimer & Ross-Messemer 1990] L. Reimer and M. Ross-Messemer. *Contrast in the electron spectroscopic imaging mode of a TEM.* Journal of Microscopy, vol. 159, no. 2, pages 143–160, 1990. (Cited on page 4.)

[Reyes *et al.* 2009] N. Reyes, C. Ginter and O. Boudker. *Transport mechanism of a bacterial homologue of glutamate transporters.* Nature, vol. 462, no. 7275, pages 880–5, 2009. (Cited on page 125.)

[Robinson & Siegelbaum 2003] R. B. Robinson and S. A. Siegelbaum. *Hyperpolarization-activated cation currents: from molecules to physiological function.* Annual Review of Physiology, vol. 65, pages 453–80, 2003. (Cited on pages 137 and 149.)

[Rothberg *et al.* 2002] B. S. Rothberg, K. S. Shin, P. S. Phale and G. Yellen. *Voltage-controlled gating at the intracellular entrance to a hyperpolarization-activated cation channel.* Journal of General Physiology, vol. 119, no. 1, pages 83–91, 2002. (Cited on page 147.)

[Ruiz & Karpen 1999] M. Ruiz and J. W. Karpen. *Opening mechanism of a cyclic nucleotide-gated channel based on analysis of single channels locked in each liganded state.* Journal of General Physiology, vol. 113, no. 6, pages 873–95, 1999. (Cited on page 149.)

[Saier 2000] Jr. Saier M. H. *A functional-phylogenetic classification system for transmembrane solute transporters.* Microbiol Mol Biol Rev, vol. 64, no. 2, pages 354–411, 2000. (Cited on page 124.)

[Schenk *et al.* 2010] A. D. Schenk, D. Castano-Diez, B. Gipson, M. Arheit, X. Zeng and H. Stahlberg. *3D reconstruction from 2D crystal image and diffraction data.* Methods Enzymol, vol. 482, pages 101–29, 2010. (Cited on pages 11 and 68.)

[Scherer *et al.* 2014] 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, vol. 185, no. 3, pages 267–277, 2014. (Cited on page 97.)

References

[Scheres & Chen 2012] S. H. Scheres and S. Chen. *Prevention of overfitting in cryo-EM structure determination.* Nat Methods, vol. 9, no. 9, pages 853–4, 2012. (Cited on page 115.)

[Scheres 2012] S. H. Scheres. *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* Journal of Structural Biology, vol. 180, no. 3, pages 519–30, 2012. (Cited on pages 74, 98 and 119.)

[Scheuring *et al.* 2005] S. Scheuring, D. Levy and J. L. Rigaud. *Watching the components of photosynthetic bacterial membranes and their in situ organisation by atomic force microscopy.* Biochimica et Biophysica Acta, vol. 1712, no. 2, pages 109–27, 2005. (Cited on page 155.)

[Schröder *et al.* 2007] G. F. Schröder, A. T. Brunger and M. Levitt. *Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution.* Structure, vol. 15, no. 12, pages 1630–41, 2007. (Cited on pages 12, 116 and 153.)

[Schünke *et al.* 2009] S. Schünke, M. Stoldt, K. Novak, U. B. Kaupp and D. Willbold. *Solution structure of the Mesorhizobium loti K1 channel cyclic nucleotide-binding domain in complex with cAMP.* EMBO Reports, vol. 10, no. 7, pages 729–35, 2009. (Cited on page 138.)

[Schünke *et al.* 2011] S. Schünke, M. Stoldt, J. Lecher, U. B. Kaupp and D. Willbold. *Structural insights into conformational changes of a cyclic nucleotide-binding domain in solution from Mesorhizobium loti K1 channel.* Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 15, pages 6121–6, 2011. (Cited on pages 138 and 154.)

[Shimamura *et al.* 2010] T. Shimamura, S. Weyand, O. Beckstein, N. G. Rutherford, J. M. Hadden, D. Sharples, M. S. Sansom, S. Iwata, P. J. Henderson and A. D. Cameron. *Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1.* Science, vol. 328, no. 5977, pages 470–3, 2010. (Cited on page 133.)

[Sobczak & Lolkema 2005a] I. Sobczak and J. S. Lolkema. *The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism.* Microbiol Mol Biol Rev, vol. 69, no. 4, pages 665–95, 2005. (Cited on pages 129 and 131.)

[Sobczak & Lolkema 2005b] I. Sobczak and J. S. Lolkema. *Loop VIII/IX of the Na+-citrate transporter CitS of Klebsiella pneumoniae folds into an amphipathic surface helix.* Biochemistry, vol. 44, no. 14, pages 5461–70, 2005. (Cited on page 128.)

[Sobczak & Lolkema 2005c] I. Sobczak and J. S. Lolkema. *Structural and mecha-*

nistic diversity of secondary transporters. Curr Opin Microbiol, vol. 8, no. 2, pages 161–7, 2005. (Cited on page 125.)

[Sorzano *et al.* 2004] C. O. Sorzano, R. Marabini, J. Velazquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo and A. 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 98.)

[Stewart & Grigorieff 2004] A. Stewart and N. Grigorieff. *Noise bias in the refinement of structures derived from single particles.* Ultramicroscopy, vol. 102, no. 1, pages 67–84, 2004. (Cited on page 115.)

[Student 1908] Student. *The probable error of a mean.* Biometrika, 1908. (Cited on pages 76 and 78.)

[Sun *et al.* 2012] L. Sun, X. Zeng, C. Yan, X. Sun, X. Gong, Y. Rao and N. Yan. *Crystal structure of a bacterial homologue of glucose transporters GLUT1-4.* Nature, vol. 490, no. 7420, pages 361–6, 2012. (Cited on page 124.)

[Tang *et al.* 2007] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. 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 page 108.)

[Taraska *et al.* 2009] J. W. Taraska, M. C. Puljung, N. B. Olivier, G. E. Flynn and W. N. Zagotta. *Mapping the structure and conformational movements of proteins with transition metal ion FRET.* Nature Methods, vol. 6, no. 7, pages 532–7, 2009. (Cited on page 138.)

[ter Horst & Lolkema 2012] R. ter Horst and J. S. Lolkema. *Membrane topology screen of secondary transport proteins in structural class ST[3] of the MemGen classification. Confirmation and structural diversity.* Biochim Biophys Acta, vol. 1818, no. 1, pages 72–81, 2012. (Cited on pages 124, 125, 128 and 129.)

[Trabuco *et al.* 2008] L. G. Trabuco, E. Villa, K. Mitra, J. Frank and K. Schulten. *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.* Structure, vol. 16, no. 5, pages 673–83, 2008. (Cited on page 12.)

[Trachtenberg & DeRosier 1987] S. Trachtenberg and D. J. DeRosier. *Three-dimensional structure of the frozen-hydrated flagellar filament. The left-handed filament of Salmonella typhimurium.* Journal of molecular biology, vol. 195, no. 3, pages 581–601, 1987. (Cited on page 78.)

References

[Tsai & Ziegler 2010] C. J. Tsai and C. Ziegler. *Coupling electron cryomicroscopy and X-ray crystallography to understand secondary active transport.* Curr Opin Struct Biol, vol. 20, no. 4, pages 448–55, 2010. (Cited on page 124.)

[Tsai *et al.* 2013] C. J. Tsai, K. Tani, K. Irie, Y. Hiroaki, T. Shimomura, D. G. McMillan, G. M. Cook, G. F. Schertler, Y. Fujiyoshi and X. D. Li. *Two Alternative Conformations of a Voltage-Gated Sodium Channel.* Journal of Molecular Biology, 2013. (Cited on page 149.)

[Ubarretxena-Belandia *et al.* 2003] I. Ubarretxena-Belandia, J. M. Baldwin, S. Schuldiner and C. G. Tate. *Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer.* The EMBO journal, vol. 22, no. 23, pages 6175–81, 2003. (Cited on page 133.)

[Unger 2013] V. M. Unger. *Evaluation of electron crystallographic data from images of two-dimensional crystals.* Methods in Molecular Biology, vol. 955, pages 211–227, 2013. (Cited on pages 85 and 115.)

[Valpuesta *et al.* 1994] J. M. Valpuesta, J. L. Carrascosa and R. Henderson. *Analysis of electron microscope images and electron diffraction patterns of thin crystals of phi 29 connectors in ice.* Journal of molecular biology, vol. 240, no. 4, pages 281–287, 1994. (Cited on page 155.)

[van der Rest *et al.* 1992] M. E. van der Rest, D. Molenaar and W. N. Konings. *Mechanism of Na(+)-dependent citrate transport in Klebsiella pneumoniae.* Journal of Bacteriology, vol. 174, no. 15, pages 4893–8, 1992. (Cited on page 133.)

[van Geest & Lolkema 2000] M. van Geest and J. S. Lolkema. *Membrane topology of the Na(+)/citrate transporter CitS of Klebsiella pneumoniae by insertion mutagenesis.* Biochim Biophys Acta, vol. 1466, no. 1-2, pages 328–38, 2000. (Cited on page 125.)

[Vinothkumar *et al.* 2005] K. R. Vinothkumar, S. H. J. Smits and W. Kühlbrandt. *pH-induced structural change in a sodium/proton antiporter from Methanococcus jannaschii.* The EMBO journal, vol. 24, no. 15, pages 2720–2729, 2005. (Cited on pages 76, 80 and 87.)

[Wahl-Schott & Biel 2009] C. Wahl-Schott and M. Biel. *HCN channels: structure, cellular regulation and physiological function.* Cellular and Molecular Life Sciences, vol. 66, no. 3, pages 470–94, 2009. (Cited on page 147.)

[Wall *et al.* 1985] J. S. Wall, J. F. Hainfeld and K. D. Chung. *Films that wet without glow discharge.* In Proceedings of the 43rd Meeting of the Electron Microscopy Society of America. San Francisco Press, Inc., San Francisco, pages

716–717, 1985. (Cited on page 6.)

[Wang *et al.* 2001] J. Wang, S. Chen and S. A. Siegelbaum. *Regulation of hyperpolarization-activated HCN channel gating and cAMP modulation due to interactions of COOH terminus and core transmembrane regions.* Journal of General Physiology, vol. 118, no. 3, pages 237–50, 2001. (Cited on page 149.)

[Welch 1947] B. L. Welch. *The generalization of student's problem when several different population variances are involved.* Biometrika, 1947. (Cited on page 80.)

[Wilkens & Aldrich 2006] C. M. Wilkens and R. W. Aldrich. *State-independent block of BK channels by an intracellular quaternary ammonium.* Journal of General Physiology, vol. 128, no. 3, pages 347–64, 2006. (Cited on page 149.)

[Williams 1984] M. Williams. *Molecular aspects of the action of benzodiazepine and non-benzodiazepine anxiolytics: a hypothetical allosteric model of the benzodiazepine receptor complex.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, vol. 8, no. 2, pages 209–47, 1984. (Cited on page 150.)

[Xu *et al.* 2010] X. Xu, Z. V. Vysotskaya, Q. Liu and L. Zhou. *Structural basis for the cAMP-dependent gating in the human HCN4 channel.* Journal of Biological Chemistry, vol. 285, no. 47, pages 37082–91, 2010. (Cited on page 138.)

[Yamashita *et al.* 2005] A. Yamashita, S. K. Singh, T. Kawate, Y. Jin and E. Gouaux. *Crystal structure of a bacterial homologue of Na+/Cl–dependent neurotransmitter transporters.* Nature, vol. 437, no. 7056, pages 215–23, 2005. (Cited on pages 124 and 133.)

[Yellen 1998] G. Yellen. *The moving parts of voltage-gated ion channels.* Quarterly Reviews of Biophysics, vol. 31, no. 3, pages 239–95, 1998. (Cited on page 149.)

[Zagotta *et al.* 2003] W. N. Zagotta, N. B. Olivier, K. D. Black, E. C. Young, R. Olson and E. Gouaux. *Structural basis for modulation and agonist specificity of HCN pacemaker channels.* Nature, vol. 425, no. 6954, pages 200–5, 2003. (Cited on pages 138 and 149.)

[Zeng *et al.* 2007a] X. Zeng, B. Gipson, Z. Y. Zheng, L. Renault and H. Stahlberg. *Automatic lattice determination for two-dimensional crystal images.* Journal of Structural Biology, vol. 160, no. 3, pages 353–61, 2007. (Cited on pages 53 and 64.)

[Zeng *et al.* 2007b] X. Zeng, H. Stahlberg and N. Grigorieff. *A maximum likelihood*

approach to two-dimensional crystals. Journal of Structural Biology, vol. 160, no. 3, pages 362–74, 2007. (Cited on pages 53, 67 and 102.)

[Zhang & Zhou 2011] X. Zhang and Z. H. Zhou. *Limiting factors in atomic resolution cryo electron microscopy: no simple tricks.* Journal of Structural Biology, vol. 175, no. 3, pages 253–63, 2011. (Cited on pages 2 and 102.)

[Zhou *et al.* 2001] Y. Zhou, J. H. Morais-Cabral, A. Kaufman and R. MacKinnon. *Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution.* Nature, vol. 414, no. 6859, pages 43–8, 2001. (Cited on page 149.)

[Zhou *et al.* 2004] L. Zhou, N. B. Olivier, H. Yao, E. C. Young and S. A. Siegelbaum. *A conserved tripeptide in CNG and HCN channels regulates ligand gating by controlling C-terminal oligomerization.* Neuron, vol. 44, no. 5, pages 823–34, 2004. (Cited on page 138.)

[Zou *et al.* 1993] X. D. Zou, Y. Sukharev and S. Hovmöller. *Quantitative electron diffraction - new features in the program system ELD.* Ultramicroscopy, vol. 52, pages 436–444, 1993. (Cited on page 68.)

[Zou *et al.* 2004] X. D. Zou, A. Hovmöller and S. Hovmöller. *TRICE - A program for reconstructing 3D reciprocal space and determining unit-cell parameters.* Ultramicroscopy, vol. 98, no. 2-4, pages 187–93, 2004. (Cited on page 68.)

Marcel Arheit
Blauensteinerstrasse 10
4053 Basel
Tel: +41 78 858 70 16

marcel@arheit.com
11.09.81
Swiss/Swedish

Curriculum Vitae

Education

UNIVERSITY OF BASEL, BASEL, SWITZERLAND

2009-now	• **Ph.D. in Bioinformatics:** Research on image processing for the 3D reconstruction of membrane protein structures from electron microscopy images of 2D crystal. Member of the Prof. H. Stahlberg group, the Center for Cellular Imaging and NanoAnalytics (C-CINA) of the Biozentrum at the University of Basel.
2006-2009	• **Master of Science in Computer Science:** Master's thesis in Image Analysis "Separating Shading from Reflectance in Face Images" (supervised by Prof. Dr. Vetter and Reinhard Knothe)
2003-2006	• **Bachelor of Science, Major in Computer Science:** Bachelor's thesis in Computer Graphics "An Audiovisual Text-To-Speech System using the Morphable Model" (supervised by Prof. Dr. Vetter and Reinhard Knothe)

UNIVERSITY OF KARLSTAD, KARLSTAD, SWEDEN

WS 2007	• Erasmus Exchange Program for one Master semester.

ETH, ZÜRICH, SWITZERLAND

2002-2003	• Studied my first year of Computer Science at the ETH. Switched to the University of Basel because of financial reasons.

GYMNASIUM MUTTENZ, MUTTENZ, SWITZERLAND

1998-2001	• Matura with focus on mathematics and physics.

HIGHSCHOOL DOLGEVILLE, DOLGEVILLE, USA

1997-1998	• Exchange year through Rotary at the Highschool of Dolgeville in the state of New York.

Professional Experience

VARIAN MEDICAL SYSTEMS IMAGING LABORATORY, DÄTTWIL, SWITZERLAND

20013-now	• **Scientific Software Engineer** Respiratory motion signal analysis and marker detection for radiation therapy.

X8X PROCESS SOLUTIONS AG, ZÜRICH, SWITZERLAND

2008-2009	• **Software Engineer** Web application development with JSF, GWT and XML especially for Novartis.

UNIVERSITY OF BASEL, BASEL, SWITZERLAND

2007-2009	• **Scientific Assistant** Software development for the computer graphics and vision group of Prof. Dr. Vetter. Sole projects realization in C++ and QT. Mediator in between different research interests.

Teaching Experience

2012	• **NCCR TransCure Hands-On Workshop on Image Processing with 2dx** Organizer, Speaker and Tutor of the Workshop held 19.8.-24.8 2012 at the Biozentrum, University of Basel.
2010	• **Workshop on Electron Crystallography of Membrane Proteins** Speaker and Tutor of the Workshop held 1.8.-7.8 2010 at C-CINA, University of Basel.

Teaching Experience (continued)

2012-2010	• **Blockcourse Structural Biology and Biophysics** Tutor for practical in Image Processing of electron microscopy data at University of Basel.
2007	• **Computer Graphics** Tutor for exercises accompanying the computer graphics lecture at University of Basel.
2006-2007	• **Programing I & II** Tutor for exercises accompanying the lectures Programming I and II at University of Basel.

Publication List

- **Conformational changes in difference maps of Cryo-EM 2D crystals** (M. Arheit, C. Paulino, C., D. Castaño-Díez, W. Kühlbrandt, H. Stahlberg), *Manuscript in preparation*, 2016.
- **2dx_automator: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline** (S. Scherer, J. Kowal, M. Chami, V. Dandey, M. Arheit, P. Ringler, H. Stahlberg), *In Journal of Structural Biology*, Vol. 186, Issue 2, 2014.
- **Single Particle 3D Reconstruction for 2D Crystal Images of Membrane Proteins** (S. Scherer, M. Arheit, J. Kowal, X. Zeng and H. Stahlberg), *In Journal of Structural Biology*, Vol. 185, Issue 3, 2014.
- **Ligand-induced structural changes in the cyclic nucleotide-modulated potassium channel MloK1** (J. Kowal, M. Chami, P. Baumgartner, M. Arheit, P.-L. Chiu, G. F. Schröder, C. M. Nimigean and H. Stahlberg.), *In Nature Communications,* Vol. 5, 2014.
- **Structure and substrate-induced conformational changes of the secondary citrate/sodium symporter CitS revealed by electron crystallography** (F. Kebbel, M. Kurz, M. Arheit, M. G. Grütter, H. Stahlberg), *In Structure,* Vol. 21, 2013.
- **Automation of Image Processing in Electron Crystallography** (M. Arheit, D. Castaño-Díez , R. Thierry, B. Gipson, X. Zeng, H. Stahlberg), *In Electron Crystallography of Soluble and Membrane Proteins, Methods in Molecular Biology,* Vol. 955, Chapter 18, 2012.
- **Merging of Image Data in Electron Crystallography** (M. Arheit, D. Castaño-Díez , R .Thierry, P. Abeyrathne, B. Gipson, H. Stahlberg), *In Electron Crystallography of Soluble and Membrane Proteins, Methods in Molecular Biology, Vol. 955, Chapter 11,* 2012.
- **Image Processing for 2D Crystal Images** (M. Arheit, D. Castaño-Díez , R. Thierry, B. Gipson, X. Zeng, H. Stahlberg), *In Electron Crystallography of Soluble and Membrane Proteins, Methods in Molecular Biology, Vol. 955, Chapter 10,* 2012.
- **Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments** (D. Castao-Dez, M. Kudryashev, M. Arheit, H. Stahlberg), *In Journal of Structural Biology*, Elsevier, 2012.
- **Electron Microscopy Analysis of 2D Crystals of Membrane Proteins** (P.D. Abeyrathne, M. Arheit, F. Kebbel, D. Castaño-Díez , K.N. Goldie, M. Chami, L. Renault, W. Khlbrandt, H. Stahlberg), *In Comprehensive Biophysics,*Elsevier, volume 1.19, 2012.
- **3D Reconstruction from 2D Crystal Image and Diffraction Data** (A.D. Schenk, D. Castaño-Díez , B. Gipson, M. Arheit, X. Zeng, H. Stahlberg), *In Methods in enzymology*, Elsevier, volume 482, 2010.