Developing Electron Crystallography for Structure Elucidation of Organics and Macromolecules

Inauguraldissertation

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

von

Max Theo Ben Clabbers

Delft, die Niederlande

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Jan Pieter Abrahams

Prof. Dr. Henning Stahlberg

Basel, den 18. September 2018

_____

Prof. Dr. Martin Spiess, Dekan

CONTENTS

CHAPTER 1

Introduction to electron crystallography

1.1   Challenges in structural biology

The past decades, macromolecular X-ray crystallography (MX) has been the standard for protein structure determination from single three-dimensional (3D) crystals. Diffraction experiments require a crystalline sample of sufficient size and quality, which is usually the most critical step in successful structure elucidation. A crystalline sample consists of an orderly and repeated stack of the same molecule, where individual molecules are related by translational and rotational symmetry operators. Constructive interference occurs when waves that are diffracted by the crystal lattice are in phase, and because of the many identical copies of the same molecule being present in the crystal, the signal gets amplified giving a Bragg boost to the total signal (Fig. 1.1). In diffraction, this gives rise to the reciprocal lattice that is the Fourier transform of the real space lattice of the crystal.

X-rays can be scattered either elastically or inelastically by the electron clouds surrounding the nuclei (see Fig. 2.1)[1]. Elastic scattering causes the incoming X-rays to diffract at a scattering angle 2θ, with only a negligible loss of energy (Fig. 1.1). However, in case of an inelastic scattering event, all the energy of the X-ray photon is deposited into the sample, resulting in the complete absorption of the X-ray photon (see Fig 2.1). On average, 12 keV X-rays are expected to scatter elastically only 1 out of 20 scattering events on an average protein atom, thus coming at a substantial cost of about 200 keV of collateral damage per useful diffracted quantum (Helliwell, 1988; Henderson, 1995; Nave & Hill, 2005). Radiation damage is a major concern in crystallography, as it will deteriorate the crystal, and subsequently decrease the diffracting intensity. Half the diffraction intensity is lost when the total exposure dose received by the crystal exceeds the critical dose limit of 10.0 $e^-.Å^{-2}$ (Henderson, 1995; Owen *et al.*, 2006). Owing to the considerable amounts of radiation damage caused by inelastic X-ray scattering, crystals need to be approximately a few micrometre in size for successful structure determination. Size restrictions on the minimum required crystal volume can be relieved to a certain extent by using micro-focused beam lines (Hedman *et al.*, 1985; Holton & Frankel, 2010; Sanishvili *et al.*, 2011; Evans *et al.*, 2011).

Unfortunately, obtaining sufficiently sized 3D crystals remains a major limiting factor in X-ray crystallography as proteins, especially membrane proteins, are often difficult to crystallise (Terwilliger *et al.*, 2009; Grimes *et al.*, 2018). However, many seemingly unsuccessful crystallisation trials, failing to yield crystals of sufficient size and quality for X-ray diffraction, do actually contain smaller nanometre-sized crystals (nanocrystals) (Stevenson *et al.*, 2014, 2016). These small crystals can be obtained using crystallisation methods that are standard in X-ray crystallography (Georgieva *et al.*, 2007; Nederlof *et al.*, 2011; Calero *et al.*, 2014), and small crystals may even have fewer defects and better order than macro-sized crystals (Cusack *et al.*, 1998). Although these nanocrystals are too small for

---

[1] The introduction was written with the intent of providing the necessary basis for the forthcoming chapters and to highlight the coherency between the different topics that are discussed. Certain key concepts are therefore already briefly introduced here, while a more detailed discussion is featured in the forthcoming chapters. Each chapter is presented as reprint of a peer-reviewed manuscript that has been previously published, or is accepted for publication. Therefore several concepts touched upon in the introduction are reintroduced in the forthcoming paragraphs of this work.

6

synchrotron experiments, they can be used for structure elucidation in time-resolved serial crystallography using an X-ray free-electron laser (XFEL) (Chapman *et al.*, 2011; Schlichting, 2015; Spence, 2017). However, determining the protein structure from snapshots of individual crystals requires large quantities of sample that are often not available, and require access to large-scale XFEL facilities.
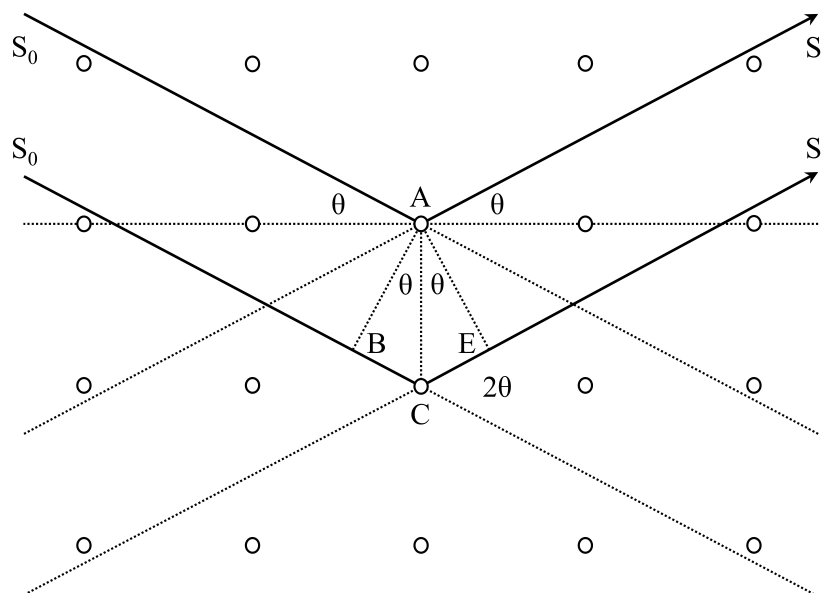

Figure 1.1 Electrons (or X-rays) are scattered by the crystal lattice with a scattering angle 2θ, which is defined as the angle between the incident beam S_0 and the scattered wave S_1. Constructive interference occurs when the path length difference between two outgoing waves is equal to an integer number of wavelengths, which holds true when $n\lambda = BC + CE$, where n is an integer number and λ is the wavelength. From the trigonometry, we can then derive that $BC = CE = d \sin \theta$, where d is the lattice spacing in real space that is equal to the distance AC, and θ is half the scattering angle that is equal for the incident beam S_0 and the outgoing wave S_1. After substitution, we then derive what is known as Bragg's law, defined as $n\lambda = 2d \sin \theta$ (see also Equation 2).

1.2 Electrons as alternative to X-rays

Alternatively, electrons can be used to obtain structural information from nanometre-sized 3D crystals. Electrons are negatively charged particles that interact with the electrostatic potential from the electron clouds and the atomic nuclei. On average, 200 keV electrons are expected to scatter elastically 1 out of 5 times on an average protein atom, and per inelastic scattering event only 40 eV of energy is deposited on average into the sample as collateral damage (see Fig 2.1). Each useful elastically diffracted quantum therefore comes at the cost of about 200 eV of collateral damage. Thus, electrons are several orders of magnitude less damaging compared to X-rays (Henderson, 1995). As a result, even very small 3D nanocrystals can successfully be used for structure determination.

Another major benefit of electron diffraction is an improved contrast for hydrogen atoms, which are difficult but not impossible to resolve in X-ray crystallography (Woińska *et al.*, 2016). The atomic scattering cross sections in X-ray crystallography are proportional to Z^2, whereas in electron diffraction the cross

sections are proportional to approximately $\sqrt[3]{Z^4}$ (Egerton, 2011). The contribution of the lighter atoms to the overall signal is thus stronger in electron diffraction, implying an increased contrast for the hydrogen atoms (Cowley, 1953*a,b*; Vainshtein, 1964; Dorset, 1995, 2007). Hydrogen atom positions in organic and inorganic samples could recently be refined by reducing the effects of dynamical scattering using precession electron diffraction combined with dynamical refinement (Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017).

Furthermore, high-energy electrons are generated in a transmission electron microscope (TEM), and no large-scale facilities such as synchrotrons or XFELs are required. Another advantage the TEM offers is that it can be operated in both diffraction as well as imaging mode (Williams & Carter, 2009; Zou *et al.*, 2011). Whereas the phases are lost in diffraction (see §1.10), the spatial phase information is retained in imaging (see §2.2.3). The combination of imaging and diffraction has successfully been applied in two-dimensional electron crystallography of membrane proteins (Henderson & Unwin, 1975; Grigorieff *et al.*, 1996; Gonen *et al.*, 2005; Abeyrathne *et al.*, 2010; Schenk *et al.*, 2010; Stahlberg *et al.*, 2015).

More recently, single-particle cryo-EM imaging emerged as a powerful method for structure determination of macromolecules, owing to the introduction of highly sensitive direct electron detectors (DEDs) and novel processing algorithms (Kuehlbrandt, 2014; Cheng *et al.*, 2015; Fischer *et al.*, 2015; Bartesaghi *et al.*, 2015; McMullan *et al.*, 2016). The phases do however come at a cost, and an improvement in signal-to-noise ratio (*SNR*) of several orders of magnitude can be expected when moving from imaging to diffraction (see §2.3.4), *e.g.* in diffraction the diffracting waves are separated from the direct beam, thus most of the noise is left out and not spread across the entire image, diffraction is shift-invariant and therefore only slightly affected by sample drift, and the signal in diffraction is not affected by dampening from the contrast transfer function (CTF).

1.3 Signal-to-noise ratio and radiation damage

Electron diffraction makes 3D nanocrystals of macromolecules accessible for structure determination. However, the reduced crystal volume makes data acquisition more challenging as it decreases the intensities $I_{exp}(hkl)$ of the reflections that are measured experimentally during data acquisition. The measured signal is proportional to the volume of the crystal and inversely proportional to the volume of the individual unit cells (Holton & Frankel, 2010). Thus in macromolecular electron crystallography, substantially reduced intensities can be expected because of the small crystal volume and the relative large unit cells commonly found for proteins. This has a negative effect on the *SNR* that is substantially worse than that of macro-sized crystals. The *SNR* can be improved by increasing the intensity of the incident beam, which will subsequently increase the reflection intensities.

However, increasing the intensity of the incident electron beam comes with a considerable increase in radiation damage. Although electrons are several orders of magnitude less damaging than X-rays, they can still cause major damage to nanometre-sized crystals of beam-sensitive organic samples (Henderson, 1995). The smaller the crystalline sample, the more sensitive it will be to exposure from the

incident beam. The critical dose limit, were half the intensity from the crystal is lost, is reached when the total exposure dose exceeds 10.0 $e^-.Å^{-2}$ (Henderson, 1995; Owen *et al.*, 2006). It is thus important to outrun the radiation damage, before reflection intensities reduce even further and high-resolution Bragg peaks start to fade. To reduce the effects of radiation damage sand preserve them in the vacuum of the TEM, crystals are flash-cooled in liquid ethane to preserve them in a thin layer of vitreous ice, and are kept cryo-cooled with liquid nitrogen throughout data acquisition (Adrian *et al.*, 1984; Dubochet *et al.*, 1988; Garman & Owen, 2006).

1.4   Inelastic and multiple scattering

Inelastic electron scattering occurs predominantly on the electron clouds and is characterised by only very low scattering angles (see Fig. 2.1, 2.2). For an average protein atom, the atomic scattering cross section for inelastic electron scattering is approximately 4.2 times larger than that of elastic electron scattering (see §2.2.1). Inelastic scattering has thus a much higher probability and therefore dominates over elastic scattering. Furthermore, multiple scattering events are likely to occur frequently as high-energy electrons are scattered very efficiently by the crystal. These multiple scattering events can be elastic, inelastic, or a combination of both (see Fig. 2.3), and the probability of multiple scattering events increases with crystal thickness (see Fig. 2.4). In contrast to X-ray diffraction, inelastically scattered electrons are not absorbed by the crystal and thus interfere with the kinematic signal (§1.5). Inelastic scattering is negatively affecting the *SNR* and leads to an increased diffuse background and a broadening of the Bragg peaks (see §2.3.2). Zero-loss energy filtering can experimentally remove most of the inelastically scattered electrons on the basis of their reduced energy, resulting in a decrease of the diffuse background and a sharpening of the Bragg peaks, thereby improving the *SNR* and revealing high-resolution Bragg peaks that previously were hidden in the background (Yonekura *et al.*, 2002; Egerton, 2011; Gemmi & Oleynikov, 2013).

1.5   Dynamical scattering

Elastic scattering predominalty occurs on the nuclei and is characterised by a relatively high scattering angle (see Fig. 2.2). When the sample is sufficiently thin, the probability that an elastically scattered electron will scatter a second time is small, resulting in essentially kinematical diffraction. However, as high-energy electrons are scattered very efficiently, electrons can scatter elastically more than once, known as dynamical scattering (Fig. 1.2, see also §2.2.3). Factors influencing the probability of dynamical scattering include the accelerating voltage of the microscope, crystal alignment in the beam, sample composition (*e.g.* presence of light or heavy atoms, organic or inorganic material, solvent content), and thickness of the crystal. Structure determination in crystallography is based on the accurate measurement of the Bragg peak intensities that result from coherent kinematic scattering that is focused by the crystal lattice. The electrostatic scattering potential map, basis for model building, is calculated by a Fourier transform of the phased structure factor amplitudes, using a purely kinematic approximation where the amplitudes are proportional to the measured and scaled intensities:

$$|F_{ideal}(hkl)| \propto \sqrt{I_{ideal}(hkl)} \qquad (1)$$

In presence of dynamical scattering, this kinematic approximation is no longer valid. Dynamical scattering affects the reflection intensities as illustrated in Figure 1.2 (Cowley & Moodie, 1957; Glaeser & Downing, 1993; Dorset, 1995; Weirich *et al.*, 2000). Electrons that scatter elastically more than once will appear to have originated from a different lattice point then where they originally scattered first, and as a result the electron will end up being measured at a different Bragg peak. As the scattering angles in electron diffraction are relatively low (§1.9), dynamical scattering is expected to mainly affect neighbouring peaks in close proximity. Since strong peaks have much higher intensities than weak peaks, the overall effect is that electrons leak over from strong peaks into weaker adjacent peaks that thus become more intense (Weirich *et al.*, 2000). As dynamical scattering is coherent and focused within the Bragg peaks, it is indistinguishable from the kinematic signal. Since there is also no loss of energy, it is not possible to separate dynamical from kinematical scattering experimentally by using an in-column energy filter (Yonekura *et al.*, 2002; Egerton, 2011).
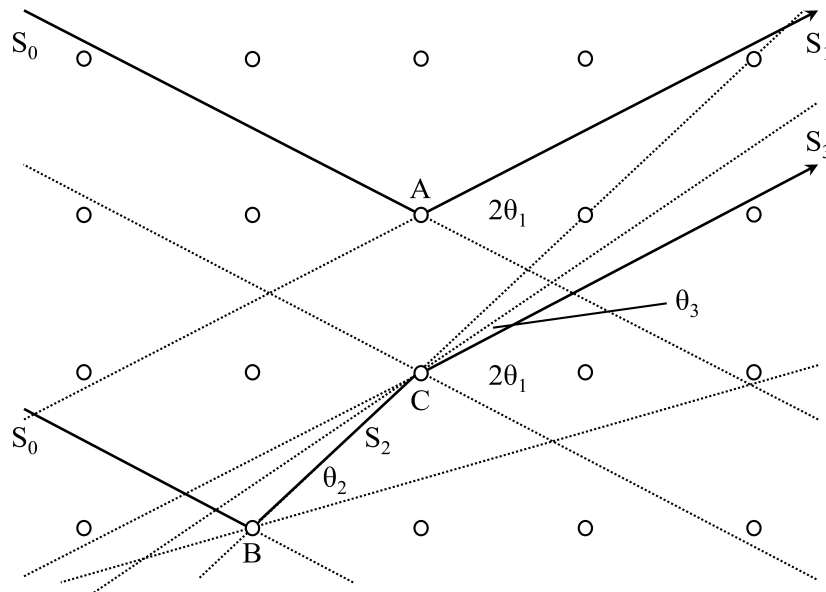

Figure 1.2 Kinematical scattering are single elastic scattering events, depicted here by the incident beam S_0 that is scatted at lattice point A in the direction of S_1 at a scattering angle of $2\theta_1$. Dynamical scattering are multiple elastic scattering events, illustrated here by the incident beam S_0 that is initially scattered at lattice point B along the direction S_2 at an angle $2\theta_2$, and then scatters a second time at lattice point C along the direction S_3 at $2\theta_3$ (the scattering angle between S_2 and S_3). As a result, the electron that originally scattered at lattice point B actually appears to originate from lattice point C where the scattering angle between the actual incident wave S_0 and S_3 is equal to $2\theta_1$, and thus S_3 is equal to the outgoing diffracted wave S_1 (Fig. 1.1). In this scenario, S_2 acts as an apparent secondary incident beam that is then scattered at C. As the scattering angles in electron diffraction are generally very small, dynamical scattering only tends to affect neighbouring peaks. Thus on average, dynamical scattering will result in leaking of electrons from strong peaks into weaker adjacent peaks (see §2.3.3).

For a typical protein crystal of 100 nm, 300 keV electrons are expected to scatter dynamically about 1 out of 10 elastic scattering events (see Fig. 2.4). Although dynamical scattering affects structure determination, structures can still be solved successfully when the crystalline sample is sufficiently thin, thereby minimising the contribution of dynamical scattering (Cowley & Moodie, 1959*a*; Dorset *et al.*, 1992; Dorset, 2007; Subramanian *et al.*, 2015). However, as discussed in §1.3, moving to thinner samples means that the diffracting intensity of the crystal is reduced, and offsetting the reduced diffracting power due to the small crystal size by increasing the electron dose can result in reduced resolution due to an increase of radiation damage. Modelling of dynamical scattering is far more complicated than a purely kinematic approximation as given in Equation 1 (Cowley & Moodie, 1959*a*; Dorset *et al.*, 1992; Jansen *et al.*, 1998; Egerton, 2011; Subramanian *et al.*, 2015), but initial results from dynamical refinement of inorganic and organic small molecules look very promising (Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017).

1.6   Small molecule electron crystallography

The potential of electron crystallography for structure determination was recognised several decades ago (Cowley, 1953*a,b*, Cowley & Moodie, 1957, 1959*a,b*; Vainshtein, 1964; Dorset, 1995). As radiation damage is a major issue in electron crystallography of biological samples, early electron diffraction studies focused mainly on crystals of more radiation-hard materials and inorganic compounds. Electron diffraction data were collected on photographic film from carefully aligned crystals oriented along a zone axis with selected-area electron diffraction (SAED) (Vainshtein, 1964; Weirich *et al.*, 1996, 2000; Hovmöller *et al.*, 2002; Zou *et al.*, 2011). Alternatively, non-aligned crystals were tilted and multiple consecutive diffraction patterns at discrete tilt angles were collected as a tilt series (Vainshtein, 1964; Zou *et al.*, 2004, 2011).

Structure determination benefits from collecting highly complete data and reducing dynamical scattering. This was achieved more recently by using precession electron diffraction (PED) which increases sampling of reciprocal space, and decreases the contribution of dynamical scattering by tilting the beam off-axis at a fixed angle which is then rapidly precessed around the optical axis (Vincent & Midgley, 1994; Midgley & Eggeman, 2015). In electron diffraction tomography, rapid beam precession is combined with tomography where the sample is tilted with discrete goniometer tilt steps (Kolb *et al.*, 2007, 2008; Mugnaioli *et al.*, 2009; Gemmi *et al.*, 2013, 2015). Alternatively, in rotation electron diffraction discrete goniometer tilt steps are combined with fine steps of beam tilt (Zhang *et al.*, 2010; Zou *et al.*, 2011; Wan *et al.*, 2013; Yun *et al.*, 2015). Rotating the crystal in a random orientation combined with precessing and/or tilting the electron beam increases data completeness by sampling the missing information between discrete goniometer tilt steps, and reduces dynamical scattering from merging symmetry related reflections averaged over multiple random orientations. Advances in data collection and the introduction of charged coupled device (CCD) detectors for faster and automated data collection led to *ab initio* structure determination of inorganic samples, metal-organic frameworks, and even small organic compounds (Dorset *et al.*, 1992; Dorset, 1992, 2007; Kolb *et al.*, 2010; Gorelik *et al.*, 2012; Mugnaioli *et al.*, 2011; Guo *et al.*, 2015).

1.7   Macromolecular electron crystallography

Nanometre-sized 3D crystals of macromolecules are highly beam-sensitive, making data acquisition more challenging. To outrun the radiation damage, crystals need to be cryo-cooled, and data should be acquired fast and under low-dose conditions. However, low-dose conditions will result in a poor *SNR*, which is rapidly getting worse from exposing the crystal to the electron beam. The *SNR* can get a boost from recording diffraction data using novel hybrid pixel detectors (HPDs) that are highly sensitive, have fast pixel-based read-out with minimal dead-time, a high dynamic range, are radiation-hard, and have reduced background as they can discriminate in energy of incoming particles and count only single electron hits for each individual pixel (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013; van Genderen, Clabbers *et al.*, 2016; Casanas *et al.*, 2016; Tinti *et al.*, 2018).

The preferred strategy for collecting diffraction data of macromolecular crystals is the rotation method (Arndt & Wonacott, 1977), which is suitable for large unit cells and has become standard in X-ray crystallography (Dauter, 1999). Data are acquired of non-aligned crystals that are continuously rotated about a single rotation axis. The rotation method is the most efficient way of measuring (near) complete data by continuous sampling of reciprocal space (Dauter, 1999). These considerations also apply to electron diffraction. Furthermore, a reduction of dynamical scattering can be expected as the crystal in not aligned and captured in random orientations. The rotation method can be adapted for electron crystallography when data are acquired using a TEM that is aligned for parallel beam illumination. The crystal should be centred on the rotation axis such that the sample is at the correct eucentric height and will stay in the beam during continuous crystal rotation.

Data acquisition using the rotation method on the TEM benefits from a stable goniometer with reliable rotation movement (Yonekura *et al.*, 2015; Shi *et al.*, 2016), and from fast and highly sensitive detectors (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013). Continuous rotation data are preferably acquired by fine-slicing through reciprocal space with an angular increment ($\Delta\varphi$) of half the crystal mosaicity[2] (Pflugrath, 1999). Hybrid pixel detectors have fast read-out, minimal dead-time and are operated in shutterless mode, thus making it feasible to record fine-sliced data (Mueller *et al.*, 2012). By collecting data with a small $\Delta\varphi$, partial reflections are recorded thus increasing the sampling frequency of the spot profiles, thereby separating the background and improving accuracy of the intensity estimations from data integration (Dauter, 2010; Mueller *et al.*, 2012).

---

[2] Crystals are often imperfect and have disorder and defects, a crystal can be described as a combination of perfect mosaic blocks in slightly different orientations. Mosaicity (in degrees) is the distribution size of the orientations of the mosaic blocks and high mosaicity leads to broadening of Bragg peaks as spots of highly mosaic crystals stay longer in Bragg condition and are smeared out over a wider rotation range. Mosaicity is defined in *XDS* and *DIALS* as the standard deviation of the Gaussian intensity distribution used to model the reflecting range. The reflecting range is also known as the rocking curve, and is defined as the total rotation range required for a single reflection to fully pass through the Ewald sphere.

The rotation method in combination with cryo-cooling the sample makes it an attractive tool for structure determination of 3D nanocrystals from beam-sensitive organics and macromolecules. As data from crystals are acquired in random orientation there is no need for crystal alignment, and by rotating the crystal continuously at slow speed near complete data can be collected. Recently, the rotation method was successfully applied in macromolecular electron crystallography and led to structure determination of proteins from micrometre- and nanometre-sized 3D crystals (Nederlof, van Genderen *et al.*, 2013; Shi *et al.*, 2013; Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015; Clabbers *et al.*, 2017; De La Cruz *et al.*, 2017). The same considerations that make the rotation method attractive for protein crystallography also apply to other (beam-sensitive) samples such as organic pharmaceutical and peptide filaments that were successfully solved by electron diffraction using continuous sample rotation (van Genderen, Clabbers *et al.*, 2016; Gemmi *et al.*, 2015, 2016; Rodriguez *et al.*, 2015; Sawaya *et al.*, 2016; Gruene *et al.*, 2018).

1.8 Diffraction geometry and data integration

In data integration the reflections are indexed and the experimental intensities $I_{exp}(hkl)$ are extracted from the recorded frames by counting of individual electron hits after background subtraction. Integrating electron diffraction data is very similar to routines that were originally developed for processing X-ray diffraction data (Kabsch, 1988*a,b*, 2010*b*; Leslie, 1999; Waterman *et al.*, 2016; Winter *et al.*, 2018). Hybrid pixel detectors count single electron hits and because there is no dark-current they have a reduced background. This required modifying the error model that has a Poisson distribution instead of a Gaussian distribution as is used for data recorded with CCD detectors, and a background estimation that is optimised for weak data with low-background (Parkhurst *et al.*, 2016; Kabsch, 2010*b*).

Data integration programs like *XDS* (Kabsch, 2010*b*) and *DIALS* (Waterman *et al.*, 2016; Winter *et al.*, 2018) can be used to successfully integrate electron data diffraction recorded with the rotation method with only some minor adaptations (see §2.5.1, §3.2.2, §4.2). These programs allow for scaling the intensities based on a error model to put all reflections on the same scale and obtain $I_{ideal}(hkl)$ (Evans, 2006, 2011; Kabsch, 2010*a*). They also apply 3D profile fitting to allow extraction of weak intensities at or even below the noise level (French & Wilson, 1978; Oatley & French, 1982; Kabsch, 2010*b*).

The relative short wavelength of electrons affects the diffraction geometry. In general, constructive interference of diffracted waves occurs when Bragg's law is satisfied (see Figure 1.1 for a derivation)[3]:

$$n\lambda = 2d_{hkl}\sin\theta \qquad (2)$$

From Equation 2 we can derive that the maximum scattering angle $2\theta_{max}$ is only 1.44° for a 200 keV electron at 1.0Å resolution. The low opening angle for even

---

[3] Here n is an integer number, λ is the electron wavelength, d is the lattice spacing [Å], and θ is the diffraction angle.

high-resolution reflections requires a relatively long detector distance, which is defined as the path length between the sample and the detector (see Fig. 2.6). Another representation of Bragg's law in reciprocal space is given by constructing the Ewald sphere with a radius of $|S_0| = |S_1| = 1/\lambda$. The Bragg condition for constructive interference is satisfied when the incident beam vector S_0 is equal to the scattered wave vector S_1 as sketched in Figure 4.2. Thus, the difference in reciprocal space between the two vectors $|S_1 - S_0| = 2\sin\theta/\lambda$ should be equal to the length of the reciprocal lattice vector[4] $|g_{hkl}| = 1/d_{hkl}$, resulting in Bragg's law as given by Equation 2.

The Ewald sphere is thus a geometrical representation of Bragg's law in reciprocal space, describing which reflections are in Bragg condition. Because of the short wavelength in electron diffraction, the Ewald sphere is virtually flat (see Fig. 4.2). In electron diffraction, we therefore see mostly only reflections from the plane perpendicular to the incident beam and reflections from higher-order Laue zones are rarely observed at low resolution. Furthermore, the detector distance is highly correlated with the unit cell dimensions because of the flat Ewald sphere[5]. Therefore the detector distance should not be refined simultaneously with the unit cell dimensions (see also §2.5.1, §4.2.6, §5.2.3.2).

1.9 Structure solution and refinement

Reconstructing the electrostatic potential map of the crystal is the basis for structure determination. The electrostatic scattering potential $\varphi(xyz)$ of the crystal can be described as a sum of cosine wave functions[6]:

$$\varphi(xyz) = \frac{\lambda}{\sigma V_{cell}} \sum_{hkl} |F(hkl)| \cos[2\pi(hx + ky + lz) - \phi(hkl)] \qquad (3)$$

In electron diffraction we measure the intensities of the reflections, which are proportional to the structure factor amplitude $|F(hkl)|$ as defined in Equation 1. However, the structure factor phases $\phi(hkl)$ are lost in diffraction and need to be retrieved before the structure can be solved.

Provided highly complete data up to (sub-)atomic resolution are available, and the molecule of interest is not too complex, the structure can likely be solved *ab initio* using direct methods (Sheldrick, 1990, 2008, 2015*b*; Morris & Bricogne, 2003). The

---

[4] The reciprocal lattice vector is described as $g_{hkl} = ha^* + kb^* + lc^*$, here h, k, l are the Miller indices and a^*, b^*, c^* are the reciprocal unit cell vectors.

[5] As the diffraction angle $\theta \to 0$, $\arctan(\theta) \simeq \arcsin(\theta) \simeq \theta$. Substituting this into Bragg's law (Equation 2) and the geometry of the diffraction (see Figure 2.6 where D is the detector distance and r is the distance between the central beam position and a Bragg spot with resolution d_{hkl}) results in a linear correlation where $D/r = d/\lambda$.

[6] Here $\varphi(xyz)$ is the electrostatic scattering potential in real space, λ is the electron wavelength, V_{cell} is the unit cell volume, σ is the interaction constant $2\pi me\lambda/h^2$, $|F(hkl)|$ is the crystallographic structure factor amplitude, and $\phi(hkl)$ are the structure factor phases in reciprocal space, see also Zou *et al.*, 2011.

only assumption made by direct methods is the atomicity of the sample, *i.e.* that a molecule is composed of atoms. With an electrostatic scattering potential that is positive everywhere, atom positions are described as sharp peaks in an otherwise empty unit cell that has a close to zero potential. The atoms are described by three positional coordinates, and one additional parameter describing a single isotropic atomic displacement parameter, totalling a number of four unknowns to solve per atom. When the number of measured reflections is substantially higher than the number of independent atoms, interdependency between the phases of strong reflections due to the atomicity restraint can be exploited by direct methods. Since strong reflections are *not* independent, their phases can be improved iteratively using statistical correlations between structure factor amplitudes, starting with an initial set of random phases. Direct methods have shown to be successful in solving structures of inorganic and organic crystals (Dorset *et al.*, 1992; Dorset, 1992, 1995, Weirich *et al.*, 1996, 2000; Zou *et al.*, 2011), including model building and refinement (Sheldrick, 2008, 2015*a*; Mueller, 2009; Thorn *et al.*, 2012). Because of the high-resolution information, individual atoms can be visualised and structure refinement is largely unrestrained.

As macromolecules are generally much more complex than (in-)organic small molecules and diffract to low-resolution, direct methods are insufficient to solve the phases. In these cases, molecular replacement (MR) can be used to solve the phase problem as experimental phasing of macromolecular 3D crystals is not yet feasible (Hovmöller *et al.*, 2002; Zou *et al.*, 2011; Wan *et al.*, 2012; Nederlof, Li *et al.*, 2013; van Genderen, Li *et al.*, 2016). Molecular replacement does however require prior structural information from a related model to preform a rotational and translational search to get the best fit of the potential map to the target structure (Vagin & Teplyakov, 1997; Evans & McCoy, 2007; McCoy *et al.*, 2007; Read & McCoy, 2016). Routines that were originally developed for X-ray crystallography can be used with minor adaptations for all subsequent steps of structure determination, including data reduction (French & Wilson, 1978; Evans, 2006; Evans & Murshudov, 2013), automated and manual model building (Cowtan, 2006; Emsley *et al.*, 2010), refinement (usually heavily constrained) (Murshudov *et al.*, 2011; Kovalevskiy *et al.*, 2018), and validation (Brunger, 1997; Luebben & Gruene, 2015; Joosten *et al.*, 2014).

1.10   Aim and scope of the thesis

The rotation method is in principle well suited for acquiring electron diffraction data of three-dimensional nanometre-sized crystals. However, the small crystal volume severely limits the *SNR*, and increases sensitivity to beam-induced radiation damage that reduces the *SNR* even further. Therefore it is necessary to boost the *SNR* and outrun the radiation damage by cryo-cooling the crystal and using fast and highly sensitive detectors.

Here I review the advantages and disadvantages of electron diffraction for structural biology. I illustrate how electron crystallography can successfully be applied to solve protein structures from 3D nanocrystals, yielding high quality data and structural model with a geometry comparable to that of X-ray diffraction. I show that many routines originally developed for X-ray crystallography can also be applied

in electron diffraction, and discuss the different aspects of the data processing that needed to be modified for processing electron diffraction data.

Electron crystallography is also of interest for related beam-sensitive samples such as small organic pharmaceuticals. These have the advantage of routinely diffracting to (sub-)atomic resolution, allowing visualisation of individual hydrogen atom positions after solving the structure by direct methods, even before manual model building. Furthermore, the data allowed for unrestrained positional refinement of the hydrogen atoms. Although the routines illustrated here can successfully be applied for electron crystallography, certain abnormalities remain that are discussed in this work. Most notably, electron diffraction data are affected by dynamical scattering. I discuss the problems related to dynamical scattering, and introduce a straightforward method to computationally reduce the negative effects of dynamical scattering on structure elucidation.

Electron diffraction and three-dimensional crystallography for structural biology

Electron crystallography is increasingly becoming a viable alternative for structure elucidation of three-dimensional, multi-nanometre sized crystals of beam-sensitive organics and macromolecules. Because electrons interact with matter strongly, crystals cannot be much more than 200 nm thick. Diffracted volumes are therefore small, leading to a poor signal-to-noise ratio (SNR) as beam damage limits the total electron dose. Data can be collected in diffraction – and imaging mode. Imaging has the advantage of providing spatial phase information, but comes at a substantial cost in SNR. Highly sensitive hybrid pixel detectors push the limits of high-quality diffraction data acquisition even further. Data integration, structure solution and refinement are feasible with existing software after minor adaptations. We review the current state of electron diffraction for structural biology, including instrumentation, data acquisition, and structure determination.

Max T.B. Clabbers[1] and Jan Pieter Abrahams[1,2,3]

[1]Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

[2]Paul Scherrer Institute (PSI), Department of Biology and Chemistry, CH-5232 Villigen PSI, Switzerland

[3]Institute of Biology Leiden (IBL), Sylviusweg 72, 2333 BE Leiden, The Netherlands

2.1 The limits of X-ray crystallography

Single crystal X-ray crystallography is the dominant method in structural biology for structure determination of macromolecules. Since the average diffracted Bragg intensity is a function of the internal order and diffracted volume of the crystal, the method requires crystals of sufficient quality and size. Obtaining such crystals is often the most critical step in successful structure elucidation. Unfortunately, growing protein crystals that are of sufficient size can be challenging, especially for membrane proteins (Terwilliger *et al.*, 2009). But nanocrystals may be far more common than expected, and many seemingly failed crystallization experiments may contain nanocrystals (Stevenson *et al.*, 2014, 2016). Furthermore, smaller crystals may be better ordered, have fewer defects and lower mosaicity that macro-sized crystals (Cusack *et al.*, 1998).

X-ray radiation damage determines the minimum size of crystals that can be studied by synchrotron radiation. Advances in data collection, including data collection at cryogenic temperatures and the development of micro-focused beam lines, allowed study of smaller crystals (Holton & Frankel, 2010; Sanishvili *et al.*, 2011; Evans *et al.*, 2011). The introduction of hybrid pixel detectors (HPDs) also had a major positive impact due to their high speed, increased sensitivity and wide dynamic range (Broennimann *et al.*, 2006). Only X-ray photons that are scattered elastically[7] provide high-resolution information. However, this useful diffraction cannot be measured without substantial collateral damage due to unrelated, but inevitable inelastic events (Helliwell, 1988; Henderson, 1995). This collateral damage is mainly caused by the photoelectric effect, in which the energy of an X-ray photon is transferred to a photo-electron. Typically, X-rays with a wavelength of 1.0Å are used for protein crystallography, which dislodge 12 keV photo-electrons (Fig. 2.1). Such a photo-electron then gradually loses its energy through multiple inelastic collisions with other atoms. These collisions increase in frequency as the photo-electron slows down, and in a protein crystal, most of the energy of a 12 keV photo-electron is deposited within a micrometre from where is was emitted (Nave & Hill, 2005). The average amount of collateral damage per elastic scattering event is determined by the average elastic and photoelectric scattering cross section of the atoms of the sample.

In practice, high-quality synchrotron X-ray data acquisition is only feasible when crystals are at least a few micrometres in size. For sub-micron sized macromolecular crystals, serial femtosecond crystallography using free-electron lasers is a viable alternative (Chapman *et al.*, 2011; Schlichting, 2015; Spence, 2017). In serial crystallography, randomly oriented crystals are each illuminated to destruction with a single, short and intense X-ray burst. Thus, each crystal produces a single diffraction pattern with the highest resolution allowed by its internal order and size. Serial crystallography requires merging diffraction data from thousands of individual nanocrystals (White *et al.*, 2013).

---

[7] Here we use the term 'scattering' for both elastic and inelastic events. We reserve the term 'diffraction' exclusively for elastic events in which the particle does not lose energy and causes no radiation damage. Diffraction does not require the sample to be crystalline.
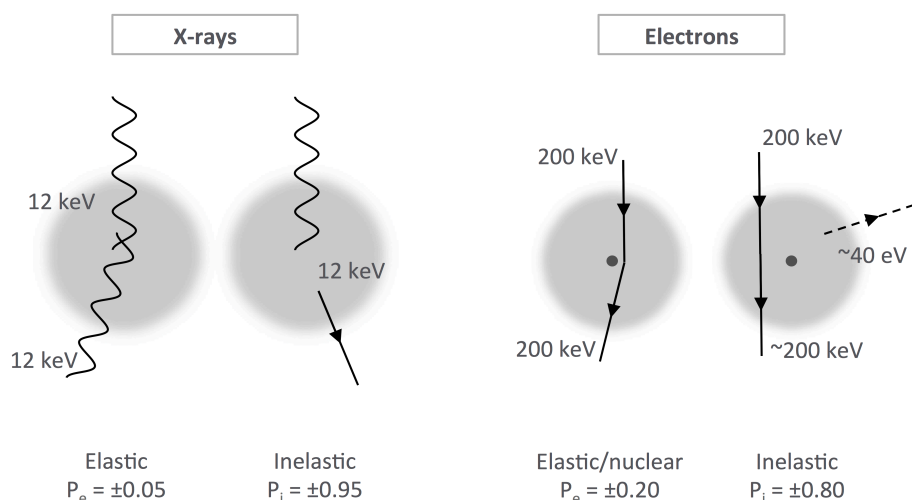
X-rays **Electrons**

12 keV

200 keV 200 keV

12 keV

~40 eV

12 keV

200 keV 200 keV ~200 keV

Elastic	Inelastic	Elastic/nuclear	Inelastic
$P_e = \pm0.05$	$P_i = \pm0.95$	$P_e = \pm0.20$	$P_i = \pm0.80$

Figure 2.1 On average, X-ray photons of 12 keV (λ=1.0332Å) are scattered elastically by a hydrated protein crystal once every 20 scattering events, resulting in an average damage of about 200 keV per elastic event. High-energy electrons of 200 keV (λ=0.0251Å) scatter elastically once every five scattering events, depositing on average only 200 eV energy into the hydrated protein crystal. Electrons are therefore several orders of magnitude less damaging than X-rays (Henderson, 1995).

2.2 The case for electrons

2.2.1 Reduced collateral radiation damage

Diffraction of 200 to 300 keV electrons provides an alternative approach to X-ray diffraction. These high-energy electrons can be scattered inelastically by the atomic electron clouds (Fig. 2.1). Inelastic electron scattering is characterised by a minute scattering angle and a small, but measurable loss in energy averaging 40 eV for typical protein atoms. Elastic electron scattering occurs mainly on the positively charged atomic nuclei, has a much higher scattering angle and an energy loss that is too small to be measured experimentally, but which can be calculated from the electron's change in momentum. The angular distributions of elastic and inelastic electron scattering can be approximated by Lorentzian functions. Especially for the lighter atoms ($Z < 40$), the Lenz equations for elastic and inelastic scattering allow calculating these angular distributions with reasonable accuracy (Egerton, 2011).

The average angular distributions of elastic and inelastic scattering of a hydrated protein can be estimated by a weighted summation of Lenz equations over its atomic composition. The average atomic composition of protein[8] and the average protein density of 1.35 g.ml^{-1} (Fischer *et al.*, 2004), imply that a protein has a density of approximately 106 atoms.nm^{-3}. Similar considerations indicate water to contain about 100 atoms/nm^{-3}. The fractional atomic composition of a typical protein crystal containing 50% (v/v) solvent in the form of H_2O, is therefore $H_{0.57}C_{0.165}N_{0.045}O_{0.21}S_{0.01}$. This allows calculating the average angular distributions of elastic and inelastic scattering of a hydrated protein by weighted summation of their respective Lenz equations. Within the energy range of 80 to 1000 keV, the full width

---

[8] We assume a typical protein to have an atomic composition of 48% H, 32% C, 9% N, 9% O and 2% S, and a protein crystal to contain ±50% solvent.

half maxima of elastic ($FWHM_e$) and inelastic ($FWHM_i$) scattering of a typical hydrated protein turn out to be linearly related[9]:

$$FWHM_e \approx 166 \times FWHM_i \frac{c}{v} \qquad (4)$$

At 300 keV (where v/c = 0.78), the angular distribution of electrons scattered elastically by a hydrated protein is typically more than two orders of magnitude wider than that of the inelastically scattered electrons (Fig. 2.2).

The overall probabilities of elastic and inelastic scattering by an atom are proportional to its scattering cross sections σ_e and σ_i, respectively. These can be calculated by integrating the Lenz equations over the scattering angle, resulting in:

$$\sigma_e \approx \sqrt[3]{Z^4} \left(\frac{c}{v}\right)^2 \ 1.87 \ 10^{-4} \ \text{Å}^2 \qquad (5)$$

Summation over Z weighted by fractional atomic composition f_Z, implies the elastic scattering cross section of an average atom of a hydrated protein:

$$\sigma_{pr,e} \approx \sum_{Z=1}^{Z=16} f_Z \sqrt[3]{Z^4} \left(\frac{c}{v}\right)^2 \ 1.87 \ 10^{-4} \ \text{Å}^2 \approx 6.7 \left(\frac{c}{v}\right)^2 \ 1.87 \ 10^{-4} \ \text{Å}^2 \qquad (6)$$

The virtual, average atomic number Z_{pr} for calculating the elastic atomic cross section of a hydrated protein is therefore about $\sqrt[4]{6.7^3} \approx 4.16$. For energy ranges between 50 to 1000 keV, and atoms with $Z < 40$, the ratio between the elastic and inelastic scattering cross sections is reasonably approximated by $0.05Z$ (Egerton, 2011). For a typical protein crystal, the average inelastic scattering cross section is therefore:

$$\sigma_{pr,i} \approx 20 \sum_{Z=1}^{Z=16} f_Z \sqrt[3]{Z} \left(\frac{c}{v}\right)^2 \ 1.87 \ 10^{-4} \ \text{Å}^2 \approx 28 \left(\frac{c}{v}\right)^2 \ 1.87 \ 10^{-4} \ \text{Å}^2 \qquad (7)$$

These calculations indicate that the average inelastic cross section of a protein atom is about 4.2 times larger than its elastic scattering cross section within an energy range of 80-1000 keV, so within this range about 20% of all scattering events are elastic. An earlier analysis based on a carbon-only sample suggested about 30% of events are elastic (Henderson, 1995). Clearly, the increased inelastic scattering by hydrogen atoms relative to carbon cannot be ignored. Given the average amount of damage per inelastic event of about 40 eV, every elastically scattered electron is therefore accompanied on average by about 170 eV of collateral radiation damage.

Similar calculations using X-ray scattering cross sections[10] indicate that elastic scattering is 18 times less prevalent in hydrated proteins than inelastic scattering for

---

[9] c is the speed of light, v is the relativistic speed of the electron.

X-rays with a wavelength of 1.0Å. Since X-rays of this wavelength deposit 12 keV per inelastic scattering event, every elastically diffracted X-ray photon is accompanied on average by about 200 keV of collateral damage. Including the contributions of all atoms, and not just carbon, confirms the observation that X-rays are about three orders of magnitude more damaging than electrons to biological samples (Henderson, 1995).
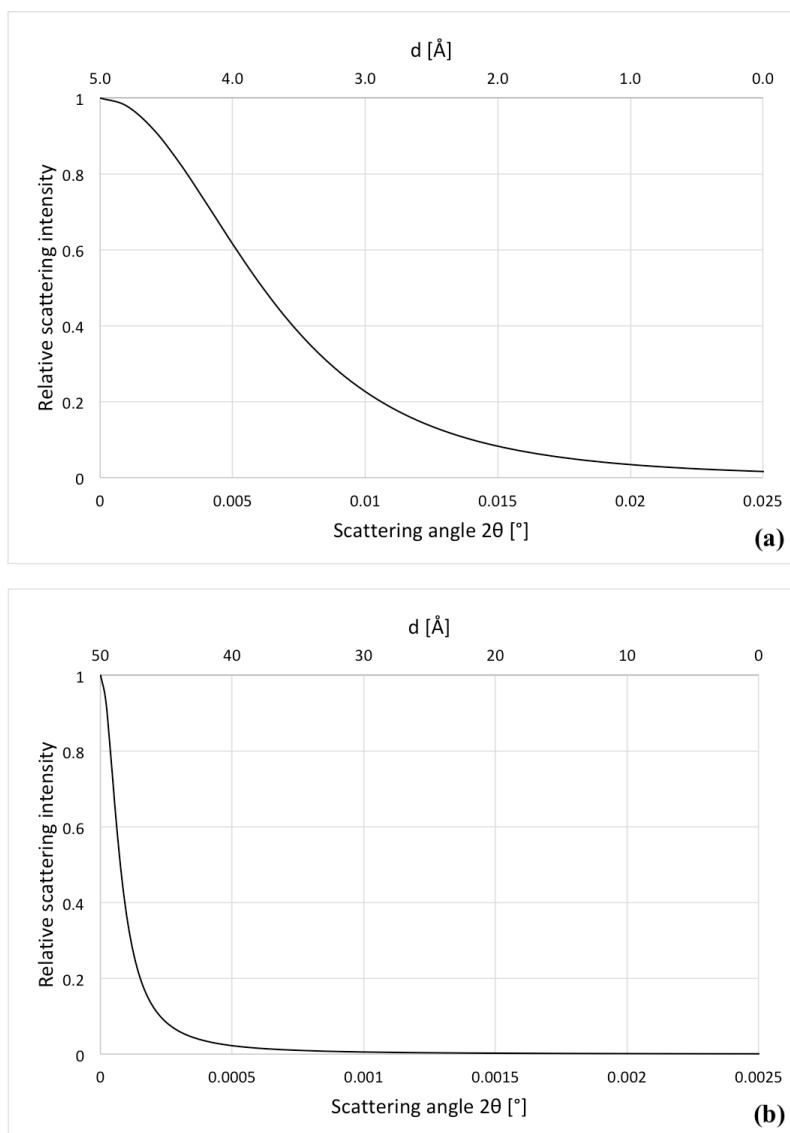

Figure 2.2 The angular spread for an average hydrated protein crystal atom ($Z = 4.16$) showing the relative scattering intensity as a function of the scattering angle 2θ for 300 keV electrons considering (a) elastic scattering, and (b) inelastic scattering. Note the order of magnitude difference in scale between graphs (a) and (b).

---

[10] Atomic scattering cross sections were taken from W.H. McMaster N.K. Del Grande, J.H. Mallett and J.H. Hubbell, "Compilation of X-ray cross sections", Lawrence Livermore National Laboratory Report UCRL-50174 (section I 1970, section II 1969, section III 1969 and section IV 1969)

2.2.2 Increased low-Z contrast

Another potential advantage of electrons is that their atomic scattering cross sections are proportional to $\sqrt[3]{Z^4}$, instead of being proportional to Z^2 as for X-rays (Egerton, 2011). This implies that the reduced scattering potential of the lighter atoms is less pronounced in the case of electrons (Cowley, 1953*a,b*; Vainshtein, 1964; Dorset, 1995, 2007). This difference should contribute to increased contrast of H-atoms when observed with electrons, as their diffraction signal represents a larger fraction of the total signal, compared X-ray diffraction. When dynamic scattering is modelled by Bloch waves (Palatinus *et al.*, 2017), but even in straightforward structure determination hydrogen positions can be made visible (see also §5.3.1).

Neutron diffraction can deliver even higher contrast for lighter elements. Neutrons scatter on the atomic nuclei and the neutron scattering potential of hydrogen atoms does not differ greatly form heavier atoms. Compared to X-ray diffraction, this implies an increased signal coming from hydrogen atoms. Therefore, even at moderate to low resolution, hydrogen positions can be visible in neutron crystallography (Blakeley, 2009). Even more so, exchanging hydrogen atoms for deuterium can further enhance contrast. Because of the relatively low flux of even the most intense neutron beams, neutron crystallography requires large crystals (~0.03 mm^3) and long exposure times (~24h) (Henderson, 1995; Blakeley, 2009).

2.2.3 Phasing electron diffraction data by imaging

Electron lenses are much more efficient than X-ray lenses, and phase information of diffracted electrons is mostly retrieved by real space imaging (Williams & Carter, 2009). These properties allow single-particle cryo-EM imaging in which thousands of individual particles are imaged in random orientations in a thin layer of vitreous ice. Structures can be solved by a 3D reconstruction from the captured views, after classification and averaging to increase their inherently poor contrast (De Rosier & Klug, 1968). High-resolution imaging requires macromolecules to be of sufficient size (Merk *et al.*, 2016; Glaeser, 2016), but new developments including phase plates are relaxing these restrictions (Danev *et al.*, 2014). The high detective quantum efficiency of direct electron detectors significantly boosts image contrast (McMullan *et al.*, 2016), which benefits particle classification and 3D reconstruction, and therefore the maximum resolution that can be obtained. Their fast readout also enables data acquisition in movie-mode, allowing the correction of beam-induced specimen motion (Campbell *et al.*, 2012; Li *et al.*, 2013). These developments are now allowing single-particle cryo-EM to become a very powerful alternative for studying macromolecular complexes to near-atomic resolution (Bartesaghi *et al.*, 2015; Fischer *et al.*, 2015; Kuehlbrandt, 2014; Cheng *et al.*, 2015).

Imaging of crystals can be used for determining their crystallographic phases. For proteins, this has so far only succeeded with two-dimensional crystals (Henderson & Unwin, 1975; Henderson *et al.*, 1990; Gonen *et al.*, 2005; Stahlberg *et al.*, 2015), but three-dimensional crystals of radiation-hard materials could also be phased by imaging (Zou *et al.*, 2011; Hovmöller *et al.*, 2002; Wan *et al.*, 2012). Phasing protein crystals by electron imaging seems an attractive option when neither a molecular replacement model, anomalous dispersion data, nor isomorphous heavy atom derivatives are available. However, for 3D protein crystals, so far only projection structures could be phased by imaging (Valpuesta *et al.*, 1994). One of the unresolved

problems is the determination of a common origin of the various projection views. Finding a common origin is greatly facilitated by the presence of high crystal symmetry within a projection image. Radiation resistant crystals with small, highly symmetric unit cells allow careful orienting and selecting such high-symmetry projections, but proteins do not survive such a procedure due to their radiation sensitivity. This problem is compounded by the lower symmetry found in protein crystals. We assume that this problem is not insurmountable: instead of having to determine the phase of each structure factor, imaging reduces the problem to having to determine only two phases (shifts in x and y) for each projection image.

However, experimental evidence indicates that the phase information that imaging can provide comes at a cost. Data collected in imaging mode demonstrate a significantly poorer *SNR* than corresponding data collected in diffraction mode. Electron diffraction patterns of 2D bacteriorhodopsin crystals at 100 to 120 keV show substantially higher resolution than corresponding images (Henderson *et al.*, 1990). Using 200 keV electrons and a dose of 0.017 e$^-$.Å^{-2}.s^{-1}, allowed measuring diffraction data of 3D protein crystals with a substantially higher resolution than images of the same crystals, measured at Scherzer focus using a dose that was 100 times higher (Nederlof, Li *et al.*, 2013; Clabbers *et al.*, 2017).

2.3   The caveats of electrons

2.3.1   The strong interaction of electrons with matter
Even the lightest atoms have a relatively large electron scattering cross section. This implies that the sample must be very thin and must be placed in a high vacuum, to prevent unwanted absorption. This affects the experimental design considerably. For instance, hydrated samples must be frozen to prevent evaporation in the vacuum of the microscope. Increasing the energy of the electrons decreases the atomic scattering cross sections, and this allows thicker samples to be studied. It also enables environmental microscopy, in which a non-frozen sample is separated from the microscope's vacuum by a cell with thin silicon nitrate windows. Biological samples that can be studied with high-end 300 keV transmission electron microscopes (TEMs) are up to 500 nm thick. But for such thick samples, it is often essential to also employ an energy filter to remove inelastically scattered electrons, which do not contribute useful spatial information, yet compromise the data by increasing the background noise.

2.3.2   Inelastic and multiple scattering
Inelastically scattered electrons continue through the sample and can reach the detector. So, unlike X-rays – which are absorbed by the sample when interacting inelastically – the inelastic signal contaminates the coherent signal that provides information on the spatial structure of the sample. Furthermore, as a typical protein sample scatters high-energy electrons very effectively, multiple scattering interactions must be taken into account. The path length travelled by the electron through the sample, determines the prominence of such multiple scattering. For calculating the probability distribution of diffraction angle and energy of an electron as a function of sample thickness, we distinguish (see also Fig. 2.3):

i.   the coherent direct beam that is unaffected by the sample,

ii.   coherent kinematic diffraction: an electron scattered elastically a single time,

iii. incoherent kinematic scattering of electrons that first scatter elastically and then one or more times inelastically,

iv.   diffuse kinematic scattering of electrons that first scatter inelastically and then elastically,

v.   coherent dynamic diffraction: an electron scattered elastically multiple times[11],

vi.   mixed multiple scattering where an electron scatters multiple times both elastically and inelastically.

Note that in scenarios *iii.* and *iv.*, electrons are scattered at an angle that is determined by the elastic scattering event. These electrons have lost a small fraction of their energy compared to electrons in scenario *ii.* and are no longer coherent. In protein crystals, inelastic electron scattering is five times more likely than elastic electron diffraction, causing mixed multiple scattering (*vi.*) to dominate over dynamical diffraction (*v.*).
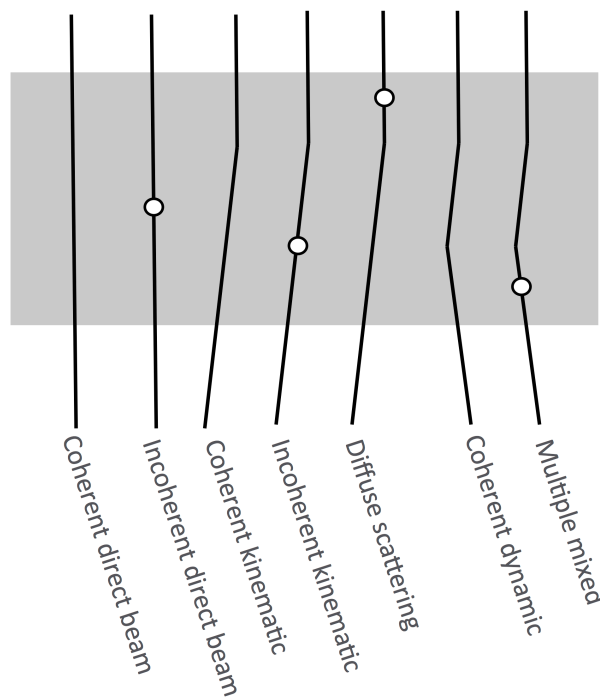

Figure 2.3 Electron scattering by a typical hydrated protein crystal can not solely be described in terms as single elastic (depicted as a large scattering angle) and inelastic (depicted as white dots with only a very small change in direction) scattering events, but also multiple scattering events have to be taken into account. Here, we describe the following scattering interactions: *i.* the coherent and incoherent direct beam, *ii.* coherent kinematic scattering, *iii.* incoherent focused kinematic scattering, *iv.* incoherent diffuse scattering, *v.* coherent dynamic scattering and *vi.* mixed multiple scattering.

---

[11] It can be argued that all elastic electron scattering is essentially dynamical to a varying degree. However, we make a distinction between scenarios *ii.* and *v.* for practical purposes, and define the kinematic signal as that fraction which is linearly related by a Fourier transform to the scattering potential of the sample.

Provided scattering probabilities are independent, Poisson statistics allow calculating the prevalence of these various types of (multiple) scattering as a function of sample thickness, composition and electron energy. This approach is not valid for oriented, highly ordered crystals with small unit cells. In such crystals, the increased probability of electron scattering within columns of atoms that line up with the incident beam causes electron channelling (van Aert *et al.*, 2007). However, for a typical protein crystal, this is not the case. So, the probability of an electron being scattered either elastically or inelastically (p_e and p_i, respectively) by a 1 nm^3 cube of a hydrated protein crystal, is the product of its expected projected atomic density (in nm^{-2}) and its average atomic scattering cross section (in nm^2). For instance, assuming a total electron dose D_{in} and a sample thickness τ (in nm), the number of electrons that pass through the sample without interacting equals $D_{in}e^{-(p_e+p_i)\tau}$. Similarly, we can calculate the probabilities of other (multiple) scattering events as a function of sample thickness and energy (Fig. 2.4).

Zero-loss filtering removes most of the inelastically scattered electrons, and results in a reduced diffuse background and sharpened Bragg spots (Yonekura *et al.*, 2002; Gemmi & Oleynikov, 2013). At first sight, this is surprising given the minute effect of inelastic scattering on the scattering angle: why do inelastically scattered electrons spread diffusely and either end up very close to the direct beam, or very close to a Bragg spot? The reason is that, when a multiple scattered electron scatters inelastically, its wave function collapses at the site of the inelastic event. This causes the electron to become incoherent and its new wave function now emanates from the site of the collision. In this fashion, an inelastic event turns the electron wave function from a coherent planar wave into a narrow ray that continues in almost the same direction, but with a coherence width that is many times smaller. Therefore, if such a localised electron ray scatters a second time, but now elastically at a higher angle, it will no longer be focussed by the crystal lattice into a Bragg spot, because its coherence width is many times smaller than the unit cell of the crystal. It is these multiple, inelastically *and* elastically scattered electrons that cause the diffuse, radially symmetric background in electron diffraction patterns that energy filtering can remove.

When an electron first scatters elastically and then inelastically, it also collapses from a plane wave function into a narrow ray, but now travelling in the direction determined by the preceding elastic event. Thus, it is still focussed in a Bragg spot, but has only interacted coherently with the top part of the sample, since its wave function interacted with the sample as a plane wave up to the location of the inelastic event. It is as if it was diffracted by a thinner crystal.

Energy filtering cannot remove diffuse diffraction caused by the crystal disorder or phonon scattering. Phonon scattering in proteins is not anticipated to contribute significantly to the scattering power. For 100 keV electrons, the phonon mean free path in a sample of an averaged hydrated protein at 300 K, can be extrapolated to be more than 1 μm, whereas it is only about 30 nm for all other types of scattering combined. Plasmon scattering has been reported to be more significant for the lighter atoms and also causes a diffuse background, whilst it is not necessarily radially symmetric (like inelastic scattering followed by elastic scattering). Since plasmon scattering is accompanied with an appreciable energy loss, it can be filtered out.
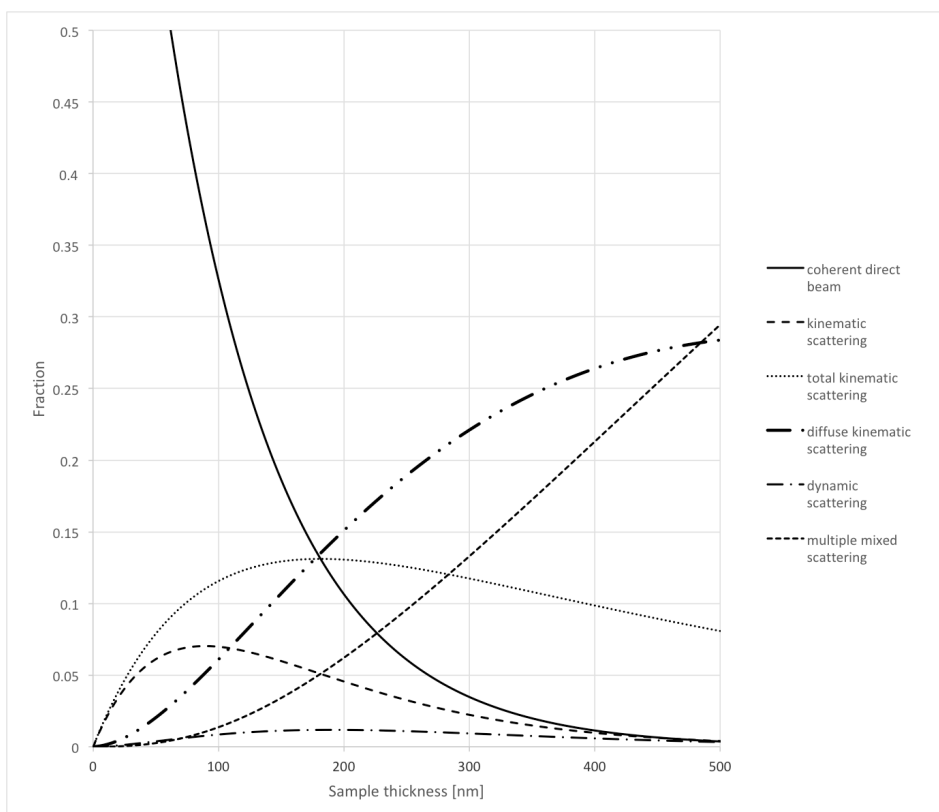
Figure 2.4 Fraction of each scattering event expected to occur for a typical hydrated protein crystal as a function of the crystal thickness [nm] at 300 keV. The total kinematic scattering curve corresponds to electrons that have scattered elastically once and subsequently may or may not have scattered inelastically (so scenarios *ii.* and *iii.* as in Figure 2.3 combined). A similar graph at other electron energies (100-1000 keV) can be obtained by scaling the horizontal axis by $(v/v_{300})^2$, where v and v_{300} are the relativistic speed of the electron with he desired energy, and of a 300 keV electron.

2.3.3   Dynamical diffraction

When all incoherent electrons are removed by energy filtering, only kinematical and dynamical scattering are observed (scenarios *ii.* and *v.*). Factors that affect dynamical diffraction include the accelerating voltage of the microscope, crystal thickness, sample composition (heavy atoms scatter more than light atoms), solvent content, and crystal alignment in the beam. Dynamical diffraction affects the structure factor amplitudes and phases, but does not affect the Bragg peak positions or energy of diffracted electrons. Hence, dynamically diffracted electrons cannot be separated experimentally from kinematically diffracted electrons, which have scattered elastically and only once. The dynamic and kinematic components of the diffraction signal can be separated computationally by multi-slice or Bloch wave approximations, but these methods require prior knowledge of the sample (Cowley & Moodie, 1959*a*; Dorset *et al.*, 1992; Subramanian *et al.*, 2015; Egerton, 2011), so these methods are useful in structure refinement, but less so for structure solution. Dynamical diffraction affects the Bragg peak intensities of the diffraction pattern, causing strong peaks to become weaker, and weak peaks to become stronger (Weirich *et al.*, 2000). This leads to a consistent overestimation of the lower intensity peaks, affecting the observed structure factor amplitudes (Clabbers *et al.*, 2017).

The electrostatic scattering potential map correlates to the Fourier transform of the structure factors of the kinematically diffracted electron wave function. Current phasing methods rely on this equivalence. Because the signal of dynamically scattered electrons reduces this correlation, structure solution is compromised. Dynamic refinement of diffraction data is promising, but far more complicated than a purely kinematic approximation (Jansen *et al.*, 1998; Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017), and there are currently no practical implementations for protein crystals. So, for initial phasing, it is important to reduce dynamical scattering. For the electrons to predominantly scatter kinematically, the crystal needs to be very thin.

It has been argued from multi-slice simulations that structure solution is still feasible for protein crystals of about 100 nm thickness with 200 keV electrons (Subramanian *et al.*, 2015). However, these simulations did not take into account inelastic scattering. The calculations presented previously in §2.3.2 imply that for 300 keV electrons, the thickness of a protein crystal should not exceed much more than 200 nm (about 20% of all elastic events is dynamic), but ideally no more than 100 nm (about 10% of all elastic events is dynamic). Increasing the energy of the electrons extends this size limitation and would allow structure determination from thicker crystals. Often, dynamical scattering is thought to inhibit any structure determination by electron diffraction. However, numerous structures have been solved by electron diffraction from 3D crystals of small compounds, materials and proteins despite the presence of dynamic scattering, in the case of proteins even for crystals that were more than 200 nm thick (Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015; Hattne *et al.*, 2015; De La Cruz *et al.*, 2017; Clabbers *et al.*, 2017).

2.3.4   Low imaging contrast

2.3.4.1 Signal-to-noise in diffraction and imaging
The major advtage of measuring data in imaging mode is that it allows retrieving phase information. Yet experimental evidence suggests that the signal-to-noise ratio[12] (SNR) in imaging mode is worse than in diffraction mode (see §2.2.3). In order to understand and quantize these observations, we introduce the detective quantum efficiency (DQE) of imaging. The DQE of a measuring device is an effective, established measure for the contrast that can be achieved. It is defined as the SNR of the output signal, divided by the SNR of the input signal[13]. The input signal is determined by the scattered electrons, and the output signal is the image:

---

[12] The signal-to-noise ratio is defined here as the ratio of the power of the signal and the power of the background noise:

$$SNR = \frac{\sum signal^2}{\sum noise^2}$$

Note that this is the signal-to-noise ratio can also be defined as the square root of this value (equivalent to the average signal divided by its standard deviation). Both definitions are found in the literature. The advantage of the definition used here, is that also applies to signals that can become negative and to complex-valued signals. When the SNR is determined by counting statistics, it is linear with the number of counts (*i.e.* doubling the dose also doubles the SNR).

[13] If the SNR is alternatively defined as the expected signal divided by its standard deviation (see footnote 12), the DQE is defined as SNR_{out}^2 / SNR_{in}^2.

$$DQE_{img}(\pmb{r}) = \frac{SNR_i(\pmb{r})}{SNR_d(\pmb{r})} \tag{8}$$

Here, $SNR_i(\pmb{r})$ and $SNR_d(\pmb{r})$ are the signal-to-noise ratios of imaging and diffraction, respectively. We express both in Fourier space, as they both depend on the scattering angle and hence of resolution. We also ensure they are both expressed on the same scale: that of the kinematically scattered structure factor $F(\pmb{r})$. We do not consider as a source of noise any real features of the sample that are not included in the model. For instance, diffuse scatter due to crystal imperfections is not considered to be noise. It would affect $SNR_i(\pmb{r})$ and $SNR_d(\pmb{r})$ in a similar fashion. Furthermore, dependent on the research question, such real features can also be treated as signal rather than noise.

Kinematic diffraction decomposes the scattering potential $\rho(\pmb{x})$ of a sample into complex-valued structure factors $F(\pmb{r})$ describing the wave functions of elastically, singly scattered electrons. The corresponding mathematical operation is a Fourier transform, indicated below by the symbol $\mathfrak{F}()$. The spatial phase distribution of $F(\pmb{r})$ equals the phase difference between the diffracted electron wave $\psi(\pmb{x}')$, and the undiffracted wave function $\psi(0,0,|\pmb{x}'|)$. The probability of detecting an electron scattered in the direction defined by $\pmb{r}$, is determined by the squared amplitude of $\psi(\pmb{x}')$ or its equivalent $F(\pmb{r})$. In the Fraunhofer diffraction geometry, the phase of $F(\pmb{r})$ is lost when measured in the far field; only the kinematic real-valued intensity $I_{far}(\pmb{r})$ is observed:

$$I_{far}(\pmb{r}) = |F(\pmb{r})|^2 + N(\pmb{r}) \tag{9}$$

If we consider counting statistics as the only source of noise (here denoted as $N(\pmb{r})$), error propagation theory implies that when we measure $I_{far}(\pmb{r})$ in diffraction, the variance of $|F(\pmb{r})|$ (which is the square root of $I_{far}(\pmb{r})$) equals ¼. The *SNR* of $|F(\pmb{r})|$ in diffraction is therefore:

$$SNR_d = 4|F(\pmb{r})|^2 \tag{10}$$

Any TEM can be switched at the flick of a button from diffraction and imaging mode, which yields near-field intensities $I_{near}(\pmb{x}')$. A Fourier transform of the measured, background-corrected intensities allows calculating $F(\pmb{r})$, the complex structure factor[14]:

$$\mathfrak{F}\left(I_{near}(\pmb{x}') - E\left(I_{near}(\pmb{x}')\right)\right)$$
$$\approx |F(\pmb{0})|^2 + 2F(\pmb{0})F(\pmb{r})CTF(\pmb{r}) + \mathfrak{F}\left(N(\pmb{x}')\right) \tag{11}$$

---

[14] $E\left(I_{near}(\pmb{x}')\right)$ is the expected, average near-field intensity, representing the constant background of the measurement; $F(\pmb{0})$ is the structure factor of the coherent, undiffracted electrons, which has a *zero* phase by definition; $CTF(\pmb{r})$ is the contrast transfer function, which equals *zero* at $\pmb{r} = \pmb{0}$ and oscillates with increasing frequency between -1 and +1 as a function of $|\pmb{r}|$: $CTF(\pmb{r}) = -\sin\left(\pi|\pmb{r}|^2\lambda\left(C_s|\pmb{r}|^2\frac{\lambda}{2} + \Delta_f\right)\right)$, where C_s is the spherical aberration coefficient; $\mathfrak{F}\left(N(\pmb{x}')\right)$ is the Fourier transform of the noise due to counting statistics.

This approximation is conditional on the following assumptions:

i. The diffracted wave functions only interfere with the undiffracted wave, and mutual interference between diffracted electron wave functions originating from different atoms is insignificant, so dynamic diffraction is excluded.

ii. The sample is an ideal weak phase object: $\psi_{re}(x) = 0$.

iii. The resolution is sufficiently low to allow Ewald sphere curvature and defocus variation to be ignored: $\lambda \left| \Delta_{f,max} - \Delta_{f,min} \right| \ll d^2$ (where d is the resolution and $\Delta_{f,max}$ and $\Delta_{f,min}$ are the maximum and minimum defocus)[15].

If we consider counting statistics to be the only source of noise $N(x')$, the variance of the noise in imaging equals the expected, average near field intensity $E\left(I_{near}(x') \right)$ by definition[14]. Parseval's theorem implies that the variance of the noise in Fourier space therefore also equals $E\left(I_{near}(x') \right)$. If only kinematically scattered electrons are considered, $E\left(I_{near}(x') \right) = |F(\mathbf{0})|^2$. Disregarding inelastic, dynamical and multiple scattering, the SNR of the Fourier transform of the electron image of a phase object is therefore:

$$SNR_i = \frac{4|F(\mathbf{0})|^2|F(r)|^2 CTF^2(r)}{|F(\mathbf{0})|^2} = 4|F(r)|^2 CTF^2(r) \qquad (12)$$

This equation suggests that $SNR_i(r)$ and $SNR_d(r)$ are equivalent if the $CTF^2(r)$-term in Equation 12 is ignored. This assumes an ideal phase plate. However, also other factors affect the SNR_i, and these will be discussed later. Each of their effects (quantised below as $Q(r)$-factors) scales down $DQE_{img}(r)$.

2.3.4.2 Elastic momentum transfer

The energy loss of a 300 keV electron ($\lambda \approx 1.97\times10^{-12}$ m) upon elastic nuclear scattering can be calculated. It depends on the scattering angle; at 3 Å resolution, the energy loss, which is absorbed as a phonon, is about 0.01 eV. This far too small to be measured directly. Yet this partial coherency loss is real as it follows from first principles. It increases the wavelength of the diffracted electron by $\Delta\lambda \approx 0.44\times10^{-19}$ m. This effect does not affect diffraction mode. But in imaging mode, the diffracted electron wave interferes with the undiffracted wave. Here, the difference in wavelength must cause a beating pattern. The period of this beating pattern can be calculated: it equals $\lambda^2/\Delta\lambda$, so for 300 keV electrons scattered at 3 Å resolution, it

---

[15] The thickness of the sample (D) determines the maximum defocus variation: $D = \left| \Delta_{f,max} - \Delta_{f,min} \right|$ and defines the rectangular envelope function of the sample in the direction of the beam. The Fourier transform of this rectangular envelope (a normalized sinc()-function) determines delocalization of diffraction in reciprocal space and has its first zeros at $\sim D^{-1}$. Ewald sphere curvature can therefore only be ignored if ($d^{-1}\sin(\theta) \ll D^{-1}$). Combining this limiting condition with the Bragg equation and substituting D with the focus variation, results in the boundary condition given in the text.

would be about $10\,\mu m$. All current electron detectors are thicker than this beat frequency, so no electrons are lost by coincidence of the detector with one of the beat's zero nodes. However, the beating interference does attenuate the average squared amplitude of the wave function by $\left(\pi^{-1}\int_0^\pi \sin^2(x)\,\partial x = \pi^{-1}\right)$, so scales down the $SNR_i(\boldsymbol{r})$ and hence the $DQE_{img}(\boldsymbol{r})$ by this same factor, defined here as $Q_m(\boldsymbol{r}) = \pi^{-1}$.

2.3.4.3 Friedel symmetry in diffraction vs. Hermitian symmetry in imaging
In diffraction mode, the two reflections that make up a Friedel pair are measured independently. This is not the case in the Fourier transform of an image that shows perfect Hermitian symmetry: $F(\boldsymbol{r})$ is the complex complement of $F(-\boldsymbol{r})$. This is a fundamental property of the Fourier transform of any real-valued function and therefore these symmetry equivalents are not independent. If we assume a flat Ewald sphere (required for SNR equivalence between imaging and diffraction), there are two independent real-valued measurements in diffraction for every single, complex-valued measurement in imaging. By summing the independent Friedel mates in diffraction, we double $SNR_d(\boldsymbol{r})$, thus scaling down $DQE_{img}(\boldsymbol{r})$ by a factor $Q_H(\boldsymbol{r}) = 1/2$.

2.3.4.4 Focal gradient & ambiguity
Either defocus or a phase plate is required to achieve contrast in imaging. An ideal phase plate changes the phase of $F(\boldsymbol{0})$ by exactly $-\pi/2$. Combined with a perfect Cs corrector, which removes spherical aberration of the electron optics, the ideal phase plate modifies the $CTF(\boldsymbol{r})$ in equation (11) into $CTF_{\pi/2}(\boldsymbol{r})$ according to:

$$CTF_{\pi/2}(\boldsymbol{r}) = \cos\left(\pi|\boldsymbol{r}|^2\lambda\Delta_f\right) \qquad (13)$$

At $\Delta_f = 0$ an object is in ideal focus and $CTF_{\pi/2}(\boldsymbol{r}) = 1$ for all $\boldsymbol{r}$ in Equation 13, resulting in maximal contrast. However, when $2\lambda\Delta_f = d^2$ all contrast is obliterated (where the resolution $d = |\boldsymbol{r}|^{-1}$). Any object with a thickness that is larger than $d^2/2\lambda$ will therefore have at least one plane that has resolution range without contrast in imaging. For instance, for a protein complex with a diameter of 20 nm of which the core is in focus, top and bottom will be +10 and -10 nm out of focus, respectively. At these defocus levels, $CTF_{\pi/2}(\boldsymbol{r})$ drops to zero at 2 Å resolution for 300 keV electrons, reducing the $DQE_{img}(\boldsymbol{r})$ for top and bottom of the particle also to zero at 2 Å resolution.

For a crystal with a thickness of 100 nm and 300 keV electrons, the effect occurs at 4.5Å resolution. However, when dealing with crystals, the situation is somewhat different. For instance, when Friedel mates no longer appear on the same diffraction pattern, only half their signal would be measured in the corresponding image, compared to what was assumed in the derivation of Equation 10, which reduces the $DQE_{img}(\boldsymbol{r})$ for such crystals only by a factor of four beyond 4.5Å resolution.

For biological samples, it may not be possible to determine the defocus level with an accuracy that was assumed for the above calculations. For single particle

analysis, the molecular complexes are embedded in a layer of ice that is thicker than the size of the complex. When the film of ice is too thin, this can for instance lead to protein denaturation at the air-water interface (Glaeser & Han, 2017). Typically, good data are collected from vitrified aqueous films with a thickness between 50 and 100 nm. High resolution requires establishing the exact defocus level of each complex that can reside at any height within this sheet of vitreous ice. But it is difficult to determine the defocus of a complex on the basis of a single, partially defocussed, low dose projection that may be moving due to charging effects in the direction of the electron beam. This problem causes many views of particles that have the same orientation, but a potentially different z-height, to be averaged in order to improve contrast. As a result, differences in defocus are also averaged out, which deteriorates the resolution. However, averaging over a range of defocus levels also corrects for the defocus gradient within each particle. So, for cyro-EM of single molecular complexes, resolution loss due to uncertainty of defocus levels will in practice dominate errors due caused by ignoring Ewald sphere curvature.

Since diffraction is translationally invariant, also in the z-direction, these considerations do not apply to measurements in diffraction mode. Defocus uncertainty therefore reduces the $\mathrm{DQE}_{\mathrm{img}}(r)$ of a sample embedded in an ice film with a thickness or focal ambiguity τ by:

$$Q_{\Delta f}(\boldsymbol{r}) = \left(\frac{\int_{\Delta_f}^{\Delta_f + \tau} \cos(\pi\lambda|\boldsymbol{r}|^2\tau')\partial\tau'}{\pi\lambda|\boldsymbol{r}|^2\tau} \right)^2 = E_{\Delta f}(\boldsymbol{r})\cos^2\left(\frac{\pi}{2}\lambda|\boldsymbol{r}|^2\big(\tau + 2\Delta_f\big)\right) \quad (14)$$

Regardless of the defocus, $Q_{\Delta f}(\boldsymbol{r})$ can never be larger than its envelope function $E_{\Delta f}(\boldsymbol{r}) = \mathrm{sinc}^2(0.5\lambda|\boldsymbol{r}|^2\tau)$, where sinc() is the normalized sinc-function[16]. This optimal contrast is achieved by focusing on the middle of the ice layer, where $\tau + 2\Delta_f = 0$, and therefore the $\cos^2()$ term is unity over all $|\boldsymbol{r}|^2$:

$$Q_{\Delta f,max}(\boldsymbol{r}) = \mathrm{sinc}^2(0.5\lambda|\boldsymbol{r}|^2\tau) \quad (15)$$

For 300 keV electrons and a focal ambiguity of 70 nm, the envelope function $E_{\Delta f}(\boldsymbol{r}) \approx 0.1$ at 3Å resolution. As none of these defocus effects apply to measurements in diffraction mode, even for a Cs corrected 300 keV microscope with an ideal image plate and a sample embedded in ice of virtually ideal thickness, focal ambiguity can cause $Q_{\Delta f}(\boldsymbol{r}) \leq 0.1$ beyond 3Å.

2.3.4.5 Loss of coherency
Inelastically scattered electrons continue through the sample and can reach the detector. This inelastic signal will increase noise due to counting statistics if it is not filtered out. But even when all inelastically scattered electrons are filtered out, the reduction of the signal corresponding to coherent, unscattered electrons still deteriorates the measured electron image. Imaging relies on coherent interference

---

[16] $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$

between all the electrons that passed through the sample (Equation 11). For a weak phase object, the signal that is measured in imaging is proportional to the product of the wave functions of the coherent, undiffracted electron and the diffracted electron. In imaging, loss of coherency of the undiffracted electrons reduces $F(\mathbf{0})$, and thereby somewhat paradoxically also reduces the measured signal of the diffracted electrons, which is proportional to $2F(\mathbf{0})F(\mathbf{r})$. This effect does not occur when measuring in diffraction mode.

We can quantise this effect using the analysis of scattering probabilities discussed in §2.3.2. These scattering probabilities refer to *detected* electrons, hence to intensities and not structure factor amplitudes. The reduction of the overall SNR_i (and DQE_{img}) caused by coherency losses within the sample, therefore is determined by the product of the curves corresponding to coherent direct beam (scenario *i.*) and the coherent kinematically diffracted electrons (scenario *ii.*), divided by total number of kinematically scattered electrons (scenarios *ii.* and *iii.*). The signal of coherent undiffracted electrons decays exponentially as it passes through the sample, for 300 keV electrons halving with every additional 60 nm of sample thickness. The signal of kinematically diffracted electrons is likewise negatively affected by subsequent partial coherence loss through secondary inelastic scattering. These two effects combined, halve the $DQE_{img}(\mathbf{r})$ every 40 nm for 300 keV electrons. At other electron energies, the halving distance turns out to be given by $70\,v^2/c^2$. This results in an attenuation factor of $DQE_{img}(\mathbf{r})$ defined by:

$$Q_i(\mathbf{r}) = \exp\left(-\frac{\ln(2)\,c^2}{70v^2}\tau\right) = \exp\left(-\frac{\tau c^2}{100v^2}\right) \tag{16}$$

For a sample that is 100 nm thick, the coherency losses of 300 keV electrons within the sample attenuate the DQE_{img} by a factor of 5.

2.3.4.6 The cost of the phase
In the previous paragraphs, we discussed fundamental limitations to the $SNR_i(\mathbf{r})$ under optimal imaging conditions. Not all of these conditions can be realized in practice yet, and some may never be achievable. Nevertheless, the analysis reveals specific limitations to imaging that do not apply to measuring diffraction data. As these limitations of imaging are mutually independent, their effects must be multiplied:

$$DQE_{img}(\mathbf{r}) \leq Q_m(\mathbf{r})Q_H(\mathbf{r})Q_{\Delta f}(\mathbf{r})Q_i(\mathbf{r}) \tag{17}$$

After substituting, assuming an ideal phase plate, the following equation results:

$$DQE_{img}(\mathbf{r}) \leq \frac{\text{sinc}^2(0.5\lambda|\mathbf{r}|^2\tau)}{2\pi}e^{-\frac{\tau c^2}{100v^2}} \tag{18}$$

Figure 2.5 shows how the DQE_{img} at 300 keV is affected by sample thickness as a function of resolution. Clearly, the phase information of electron imaging comes

at a cost. When data are collected by an electron detector with a DQE_{det} that is unity over the entire relevant resolution range, using a 300 keV Cs corrected microscope with an ideal phase plate, with a fully coherent, parallel beam and with an energy filter with 100% efficiency, the signal at 3Å resolution of a molecular complex in an ice layer that is only 50 nm thick, has a 50 times lower SNR in imaging mode compared to data collected in diffraction mode. Yet in diffraction mode, neither phase plate, energy filter or Cs correction are required. We would need to measure the data with a parallel beam that has a diameter not larger than the particle, as any additional scattering interferes with the diffraction signal of the particle. For crystals the latter limitation does not apply. We conclude that there is agreement between first principles and experimental evidence obtained from protein crystals that show imaging to have a significantly lower SNR than measuring in diffraction mode.
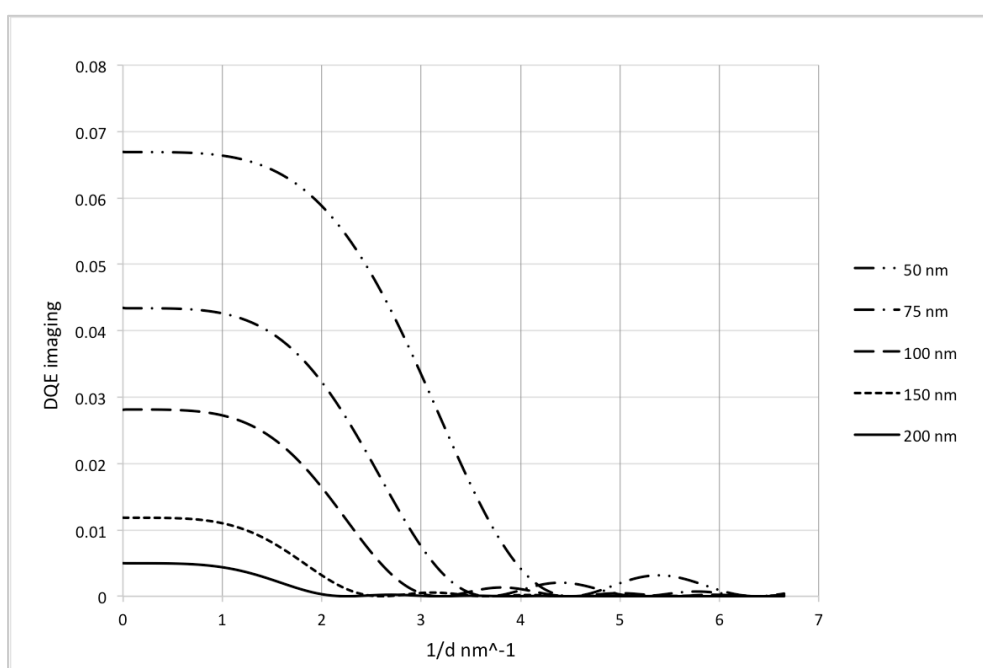

Figure 2.5 *DQE* of imaging at 300 keV is affected by the sample thickness as function of resolution; factors that were taken into account influencing the image contrast are elastic momentum transfer, Fourier symmetry, focal ambiguity, and loss of coherency due to inelastic scattering.

2.4. Data acquisition

2.4.1 Crystallisation

When sufficiently thin crystals cannot be produced by fracturing large crystals (De La Cruz *et al.*, 2017), nanocrystals need to be grown and identified specifically for the purpose of electron diffraction studies. One obvious approach is to limit the volume of the mother liquor but this has the disadvantage that fewer crystals grow, as the probability of homogeneous crystal nucleation is linearly related to the volume of the mother liquor (Bodenstaff *et al.*, 2002). To grow nanocrystals, similar strategies for crystallization trials that are standard in X-ray crystallography can be adapted (Georgieva *et al.*, 2007; Nederlof *et al.*, 2011). Confirming the presence of nanocrystals is less straightforward, as they fall beyond the resolution limit of light

microscopy used for screening. This can be overcome by additional detecting methods such as second harmonic generation (Wampler *et al.*, 2008), UV-fluorescence of predominantly tryptophan residues that can distinguish protein crystals from salt (Judge *et al.*, 2005), dynamic light scattering, in situ diffraction analysis, and electron microscopy (Calero *et al.*, 2014; Stevenson *et al.*, 2014, 2016)

2.4.2   Crystal mounting and cryo-preservation

For macromolecular electron diffraction, the 3D crystals need to be transferred to an electron microscope (EM) grid and vitrified in liquid ethane in order to reduce radiation damage and prevent their disintegration in the vacuum of the electron microscope (Nederlof, van Genderen *et al.*, 2013; Shi *et al.*, 2016). An EM grid is coated with a thin layer of carbon, and is glow-discharged prior to sample preparation to make it hydrophilic with a plasma cleaning system. The nanocrystals are in their buffer solution, of which 2-3µl is then deposited onto an EM grid. Excess liquid is blotted away from one or both sides of the grid with blotting paper, followed by plunging the grid into liquid ethane to vitrify the sample and prevent ice crystal formation, a common procedure in single-particle cryo-EM (Adrian *et al.*, 1984). The sample is then stored in liquid nitrogen prior to loading into the microscope. Cryo-protectants are often used in X-ray crystallography to prevent crystal damage through the formation of crystalline water ice (Garman & Owen, 2006). In EM imaging, cryo-protectants are usually not added since they are often viscous, which can interfere with blotting and lead to samples that are too thick. Cryo-protectants can also decrease contrast. In diffraction mode, such a loss of contrast would not be a problem and the use of cryo-protectants may be beneficial for diffraction studies, but this has not systematically been investigated so far.

2.4.3   Diffraction data acquisition

The rotation method has been the standard mode of data acquisition in macromolecular X-ray crystallography for over four decades (Arndt & Wonacott, 1977). The method involves continuously rotating a randomly oriented crystal about a single rotation axis that is normal to the direction of the X-ray beam. The crystal is centred on the rotation axis to prevent it from rotating out of the beam during data collection, so no crystal re-alignment during data acquisition is required. The rotation method is the most efficient way of measuring (near) complete data of crystals with large unit cells (Dauter, 1999). These considerations equally apply to electron diffraction. No specific adaptations to the electron microscope are required when applying the rotation method to electron diffraction (Nederlof, van Genderen *et al.*, 2013; Nannenga, Shi, Leslie *et al.*, 2014; Yonekura *et al.*, 2015; Shi *et al.*, 2016; Clabbers *et al.*, 2017). It benefits from a sensitive, radiation hard electron detector with a high dynamic range (see also §2.4.4) (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013), and a stable goniometer with a reliable tilt movement (Yonekura *et al.*, 2015; Shi *et al.*, 2016). However, unlike in X-ray diffraction, the goniometer usually only has a -70° to +70° tilt range, which for lower symmetry space groups can lead to incomplete data. In those cases, data from multiple crystals in multiple orientations need to be merged. However, protein 3D nanocrystals regularly show a preferred orientation on the grid, compromising data completeness, especially for low symmetry space groups (Clabbers *et al.*, 2017; Nannenga, Shi,

Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014), in which case experimental persistence is essential for acquiring a complete data set.

2.4.4   Measuring high-energy electrons

Bragg spot intensity is a linear function of the variance of the correlated electron scattering potential of coherently illuminated unit cells. Radiation damage reduces the correlation between unit cells and half of the diffracted intensity is lost when the total exposure exceeds the critical dose limit of 10.0 $e^-.Å^{-2}$ (Henderson, 1995; Owen *et al.*, 2006). Since the crystal size must be limited to avoid excessive dynamical and multiple scattering, and since the total dose that the crystal can take is also is limited to 10.0 $e^-.Å^{-2}$, electron diffraction data are inherently weak for protein crystals. Here, preventing inelastically scattered electrons from hitting the detector by zero-loss energy filtering improves the signal (Yonekura *et al.*, 2002; Gemmi & Oleynikov, 2013; Gemmi *et al.*, 2016; Mugnaioli *et al.*, 2016). However, the high background of X-ray photons inside the microscope compromises the accuracy of data acquisition. *DQE* measurements of electron detectors that do not discriminate between the high X-ray background that exists in any electron microscope, and high-energy electrons, tend to overestimate the accuracy of the detector if no corrections are applied.

Hybrid pixel detectors (HPDs) can discriminate energies of impacting quanta (Heijne, 2001) A hybrid pixel detector operated in counting mode will only count an incoming quantum as a hit when its energy is above a user-defined threshold (Llopart *et al.*, 2002, 2007). This has the advantage of separating the signal of highly energetic, monochromatic electrons from the background X-ray photons, which predominantly have a much lower, broad-spectrum energy distribution. Thus, HPDs offer a distinct advantage over photographic film, CCD and CMOS detectors

Top-end electron detectors measure electrons directly, discriminating and localising the signal of each single high-energy electron hitting its surface (Faruqi *et al.*, 2015; McMullan *et al.*, 2016). This concept allowed a revolution in cryo-EM structure determination that started a few years back (Kuehlbrandt, 2014), which culminated in the Nobel prize for Chemistry in 2017. These monolithic pixel detectors (MPDs; like the Gatan K2/3, or the Thermo Fisher Falcon detector family) have small pixels (5 to 15 μm). Separating and locating the individual signal of each high-energy electron ('electron counting'), requires a low electron flux, whilst reading out frames at a high rate that is so high that the total number of electrons per frame does not exceed about $1/20^{th}$ of the total number of pixels. Because of this, high frame rates are required, with each frame capturing the signals of relatively few electrons. Because the individual frames capture only so few electrons, the number of accumulated exposures that can be acquired by the very fastest MPDs in single electron counting mode, is only about 100 images per hour (Biyani *et al.*, 2017). There is no technology foreseen for MPDs that will significantly alter this situation in a critical manner.

Hybrid pixel detectors operate on a different principle (Broennimann *et al.*, 2006). Each pixel has an internal logic circuitry that quantifies the number of electrons landing at its location, *without* having to read out the whole frame. This allows frames to accumulate a much higher number of electrons prior to read-out. Fast HPDs have minimal dead-time since the counting is done before writing out the frame. Timepix3 can reach frame rates of up 1300 frames per second, whilst EIGER2

(which has 50% larger pixels) can reach frame rates of up to 4500 frames per second, with an 8- to 16-bit counting depth. Effectively, this allows an increase in data collection speed by multiple orders of magnitude, compared to monolithic detectors. HPDs have been proven to be highly effective for X-ray and electron diffraction data collection, reaching very high data collection throughputs, orders of magnitude higher than what can be achieved by MPDs (Llopart *et al.*, 2002, 2007; Johnson *et al.*, 2012).

The sophisticated logic on each pixel of a HPD requires more space than the more straightforward integrating logic of a MPD, hence the pixel size of the former is substantially larger than that of the latter. Because the resolution of an image is determined by the effective pixel size of the detector, HPDs have so far not been used for imaging: in current setups, the resolution offered by HPDs is insufficient.

HPDs have a high dynamic range and are radiation hard, making them ideal for measuring diffraction from high intensity electrons (Nederlof, van Genderen *et al.*, 2013; van Genderen, Clabbers *et al.*, 2016; Clabbers *et al.*, 2017). The high dynamic range is a direct result of having a counting detector that has a high linear detection range; accurate as well for low intensity Bragg peaks at high resolution with only a few counts as for high intensity Bragg peaks at low resolution. Because the logic circuitry of the detector ASIC is not exposed to the direct electron beam, HPDs are much more radiation hard than MPDs. This allows efficiently collecting frames, without the need for a beam-stop, with a small angular increment to finely sample 3D reciprocal space (Mueller *et al.*, 2012; Casanas *et al.*, 2016). Rotation data collected from a single protein nanocrystal can be sufficient for structure determination (Clabbers *et al.*, 2017). Merging of multiple rotation datasets can increase the data completeness, and – depending on crystal dimensions – multiple datasets may be recorded at different locations on the same crystal (Nederlof, van Genderen *et al.*, 2013; Clabbers *et al.*, 2017). However, the limited tilt range of the goniometer, and an increasing ice thickness at higher tilt angles are complicating factors.

For data acquisition, the optimal virtual detector distance is mainly determined by the desired experimental resolution, *i.e.* the maximum expected resolution of the diffraction, and the minimal separation needed between adjacent Bragg peaks to prevent overlapping. For macromolecular crystallography, distances in reciprocal space are relatively small and sufficient spacing between peaks is needed for accurate intensity measurements. For an optimal SNR ratio, the virtual detector distance should always be chosen to be as long as feasible, with the highest resolution allowed by the sample at the edge of the detector (Fig. 2.6). Then, the diffuse background will be spread over the largest possible area, while the diffraction peak remains focussed, which leads to a better SNR. Note that in a TEM, the physical sample-to-detector distance remains fixed, it is the virtual detector distance that is changed by adjusting the magnetic lenses (Williams & Carter, 2009). Calibration of the virtual detector distance is commonly performed with well-characterised standard samples that have known and well defined unit cell parameters.
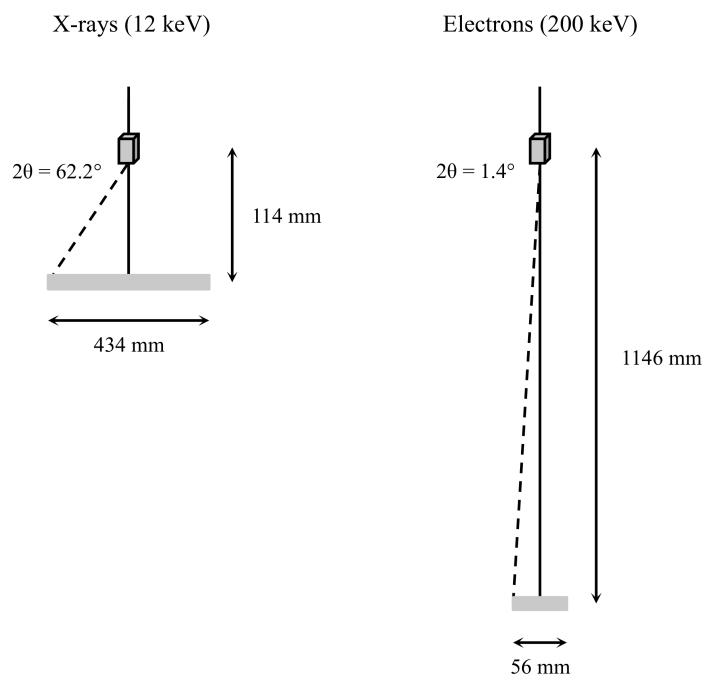
X-rays (12 keV)   Electrons (200 keV)

2θ = 62.2°   2θ = 1.4°

114 mm

1146 mm

434 mm

56 mm

Figure 2.6. The short wavelength in electron diffraction implies a low scattering angle that affects the diffraction geometry (see also §4.2.3, Fig 4.2), X-ray photons of 12 keV (λ = 1.0332Å) have a scattering angle 2θ of 62.2° for a 1.0Å reflection, requiring a short detector distance and benefiting from a large detector (*e.g.* PILATUS3 X 6M, 2463x2527 pixels, 172 μm^2 pixel size), in contrast, 200 keV electrons (λ = 0.0251Å) have a scattering angle 2θ of only 1.44° for the same 1.0Å reflection, and thus require a longer detector distance and can do with a smaller sized detector (*e.g.* Timepix 1MP, 1024x1024 pixels, 55 μm^2 pixel size).

2.5. Data processing

2.5.1 Data integration

Data integration of randomly oriented, single-shot diffraction patterns can be challenging (Jiang *et al.*, 2009). When diffraction patterns are acquired by continuous rotation with a small angular increment, reconstruction in reciprocal space is much more straightforward since all frames share a common rotation axis. Data integration profits from 3D profile fitting algorithms specifically for the weaker reflections (Kabsch, 1988*a*, 2010*b*; Leslie, 2006; Battye *et al.*, 2011). Determining the rotation axis can be less straightforward in electron diffraction, as the apparent rotation axis of the diffraction pattern is pivoted about the central beam by electron lenses, and the pivot angle depends on the lens settings. The rotation axis can be estimated by identifying reflections close to the rotation axis from their wider rocking curve (Dorset, 1976). Beam drift during data collection can be observed. When the sample has accumulated charge, the electron beam can be deflected upon sample rotation, often in a direction perpendicular to the rotation axis. Centring of diffraction frames prior to data integration can correct for beam drift (Wan *et al.*, 2013). Alternatively, data can be integrated whilst simultaneously refining the direct beam position (Kabsch, 2010*b*).

The diffraction geometry in electron crystallography is rather different from X-ray diffraction. The wavelength of a 200 keV electron, typically used in electron diffraction, is much smaller (λ=0.0251Å). This affects the Ewald sphere

reconstruction, which now is virtually flat (see also §4.2.3, Fig 4.2). An almost flat Ewald sphere implies that we only have a very low scattering angle, *e.g.* $2\theta \approx 1.44°$ for a high-resolution Bragg peak at 1.0Å. Therefore, the effective detector distance required in electron diffraction has to be much longer than what is common in X-ray crystallography to still have sufficient separation between adjacent spots (Fig. 2.6). Because the scattering angles are very small, the dimensions of the detector surface area can also be much smaller without losing high-resolution information.

The simultaneous refinement of unit cell parameters and detector distance during data integration can be difficult for lower symmetry space groups. The flatness of the Ewald sphere causes a strong correlation between detector distance and unit cell parameters, and generally both cannot be reliably refined at the same time (van Genderen, Clabbers *et al.*, 2016). Furthermore, lens distortions may interfere with determining the correct detector distance in diffraction, and can introduce spatial distortions (Capitani *et al.*, 2006; Clabbers *et al.*, 2017). Nevertheless, these problems can be surmounted once one is aware of them by careful calibration and data integration. In that case, software originally developed for X-ray data integration can be used successfully for electron diffraction data too (Kabsch, 2010*b*).

2.5.2   Phasing by direct methods

When samples are highly ordered and relatively resistant against radiation damage, they allow high resolution data collection, sufficient for phasing by direct methods (Dorset, 1991). Direct methods are based on constraints and probability, and the knowledge that when structure factors are normalised, the scattering potential consists of discrete point-like atoms that is always positive. This universal feature that is shared by all well-ordered crystals, implies that strong reflections are not independent, and relations between strong reflections can be used to estimate initial phases. This enables the complete workflow from structure solution to model building and refinement purely from the measured intensities (Zou *et al.*, 2011).

Direct methods do require high-resolution diffraction data, and generally require highly complete data to succeed (Sheldrick, 1990; Morris & Bricogne, 2003). Furthermore, dynamic scattering contributions should be reduced to minimise their effect on the measured structure factor amplitudes. Data collection by precession electron diffraction (PED), where the central beam is tilted off-axis at a fixed angle and then is rapidly precessed around the optical axis, increases sampling of reciprocal space and decreases the contribution of dynamically scattered electrons (Vincent & Midgley, 1994; Midgley & Eggeman, 2015). In electron diffraction tomography, precession is often combined with sample rotation by tilting it with discrete goniometer tilt steps (Kolb *et al.*, 2007, 2008; Mugnaioli *et al.*, 2009; Gemmi *et al.*, 2013, 2015). At each angle diffraction patterns are collected with rapid beam precession that fills in the information of the missing wedges between the tilt steps and decreases the contribution of dynamic scattering. Alternatively, coarse goniometer tilt steps are combined with fine steps of beam tilt in rotation electron diffraction (Zhang *et al.*, 2010; Zou *et al.*, 2011; Wan *et al.*, 2013; Yun *et al.*, 2015). These new strategies in data collection made it feasible to solve inorganic and organic structures *ab initio* from nanocrystals (Weirich *et al.*, 1996; Kolb *et al.*, 2010; Gorelik *et al.*, 2012; Guo *et al.*, 2015), and were accompanied by various improvements in hardware such as the advent of hybrid pixel detectors (van Genderen, Clabbers *et al.*,

2016; Wang *et al.*, 2017). These approaches appear not to be essential for protein crystals where the rotation method is the preferred strategy for data collection (see §2.4.3) (Arndt & Wonacott, 1977; Dauter, 1999; Nederlof, van Genderen *et al.*, 2013; Nannenga, Shi, Leslie *et al.*, 2014).

2.5.3 Molecular replacement

For macromolecular structures, molecular replacement can effectively be used to retrieve phase information. Indeed, all macromolecular structures solved so far by electron diffraction of 3D crystals have been solved using this method (Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015; Hattne *et al.*, 2015; De La Cruz *et al.*, 2017; Clabbers *et al.*, 2017). Molecular replacement and refinement routines that are standard in macromolecular X-ray crystallography are also relevant for electron diffraction data (Vagin & Teplyakov, 1997; McCoy *et al.*, 2007; Adams *et al.*, 2002; Winn *et al.*, 2011). Existing software does not necessarily require any major modifications as electron atomic scattering factors have already been implemented for structure solution and refinement (Adams *et al.*, 2002; McCoy *et al.*, 2007; Murshudov *et al.*, 2011; Joosten *et al.*, 2014).

2.5.4 Heavy atom derivatives and anomalous scattering

These techniques that allow experimental phasing in macromolecular X-ray crystallography, do not apply to electron diffraction. The main reason is that the scattered intensity of electrons scales with $Z^{4/3}$, instead of scaling with Z^2, which is the case for X-ray diffraction. Therefore, the signal of (for instance) an Au atom in electron diffraction is only about 5% of the signal of the same Au atom in X-ray diffraction. High-energy electrons do not scatter anomalously, so X-ray methods based on this phenomenon do not work either.

2.6   Conclusions

Electron crystallography of beam-sensitive samples is a relatively novel alternative to established methods in structural biology. Since the crystal (and unit cell) volume is inversely proportional to the diffracted intensity, protein 3D nanocrystals require sensitive electron detection. Furthermore, even at low dose conditions the crystals deteriorate quickly owing to radiation damage. Recent hardware improvements, *i.e.* faster and more sensitive detectors, made it feasible to collect diffraction data using the rotation method. This enabled structure determination from very thin protein crystals that provided sufficient high-resolution information. These detectors are available on standard TEMs that can be operated in diffraction mode and require no additional adaptations. However, data acquisition could further benefit from improving goniometer stability, using zero-loss filtering to get rid of inelastic scattering, and automation of the data collection that is currently mostly manual.

Processing electron diffraction data is very similar to routines that are standard in single crystal X-ray crystallography. While structure determination can be accomplished with only minor adaptations, there are still potential improvements in the way we treat our data. For example, data integration could be optimised by a better characterisation of the experiment, more accurate treatment of low intensity

spots, and treating the effects of multiple scattering that electron diffraction data inevitably suffer from. Electrons are charged particles and thus interact differently with the sample than X-rays, resulting in an electrostatic scattering potential map. Therefore, we would also need to take charge effects into account, next to multiple scattering. Anisotropy of electron clouds is probably less important, as electron scattering of these electron clouds is inelastic anyway, whilst its effect on anisotropic shielding of the atomic nuclei is probably only observable at very high resolution. When such changes are incorporated into existing software we could expect to get models that better represent the data.

Another key part of the structure determination is solving the phase problem. At present, all protein structures determined form 3D crystals by electron diffraction were solved using molecular replacement. Unfortunately, *de novo* structure determination from experimental phasing is currently not yet feasible. High-resolution EM images however retain both amplitude and phase information. Retrieving phase information from EM images is the basis for single-particle cryo-EM and 2D crystallography (Henderson & Unwin, 1975; Cheng *et al.*, 2015). This has also successfully been applied to electron diffraction data from nanocrystals of relatively radiation hard inorganics (Hovmöller *et al.*, 2002; Wan *et al.*, 2012). For three-dimensional beam-sensitive macromolecular crystals, phasing through imaging is less straightforward. Contrast of randomly oriented images of three-dimensional protein crystals can be improved with a lattice filter (van Genderen, Li *et al.*, 2016), while averaging highly correlated patches can enhance contrast for different mosaic blocks within the crystal (Nederlof, Li *et al.*, 2013). Initial phase information can be extended to higher resolution with density modification software (Bricogne *et al.*, 2003; Thorn & Sheldrick, 2013; Wisedchaisri & Gonen, 2011).

We conclude that electron diffraction is a promising technology that opens new applications in crystallography. So far, procedures and software from X-ray crystallography have been applied successfully on electron diffraction data. We anticipate that with the development of data analysis packages that also model effects that are specific for electron diffraction, the quality of the structure determinations by electron diffraction will improve. We also predict that phasing diffraction data of 3D (nano-)crystals by electron imaging will become more prominent for the cases where no molecular replacement model is available.

The significant improvement in *SNR* that is offered by acquiring data in electron diffraction, rather than imaging mode, opens the way towards more powerful methods for structure determination of non-crystalline samples by electron diffraction.

CHAPTER 3

Protein structure determination by electron diffraction using a single three-dimensional nanocrystal

Three-dimensional nanometre-sized crystals of macromolecules currently resist structure elucidation by single crystal X-ray crystallography. Here, a single nanocrystal with a diffracting volume of only 0.14 μm^3, *i.e.* no more than 6×10^5 unit cells, provided sufficient information for determining the structure of a rare dimeric polymorph of hen egg-white lysozyme by electron crystallography. This is at least an order of magnitude smaller than previously was possible. The molecular replacement solution, based on a monomeric poly-Ala model, provided sufficient phasing power to show side chain density, and automated model building was used to reconstruct the side chains. Diffraction data were acquired using the rotation method with parallel beam diffraction on a Titan Krios transmission electron microscope, equipped with an in-house designed 1024×1024 pixel Timepix hybrid pixel detector for low-dose diffraction data collection. Favourable detector characteristics include the ability to accurately discriminate single high-energy electrons from X-rays and count them, fast readout to finely sample reciprocal space, and a high dynamic range. This work, with other recent milestones, suggests that electron crystallography can provide an attractive alternative in determining biological structures.

Max T.B. Clabbers[1], Eric van Genderen[2], Wei Wan[3], Emiel L. Wiegers[4], Tim Gruene[2] and Jan Pieter Abrahams[1,2,5]

[1]Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, Basel University, Mattenstrasse 26, CH-4058 Basel, Switzerland

[2]Paul Scherrer Institut (PSI), Department of Biology and Chemistry, CH-5232 Villigen PSI, Switzerland

[3]Department of Materials and Environmental Chemistry, Stockholm University, SE 10691 Stockholm, Sweden

[4]Leiden Institute of Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, the Netherlands

[5]Leiden Institute of Biology, Sylviusweg 72, 2333 BE Leiden, The Netherlands

3.1 Introduction

Electron crystallography can be used for structure determination of macromolecules from crystalline samples. Originally the method concentrated on diffracting and imaging 2D crystals (Raunser & Walz, 2009; Stahlberg *et al.*, 2015), and resulted in important structures of membrane proteins (Henderson & Unwin, 1975; Gonen *et al.*, 2005). Electron diffraction of 3D crystals allowed structure solution of organic and inorganic samples (Vainshtein, 1964; Dorset, 1995; Weirich *et al.*, 1996; Mugnaioli *et al.*, 2009; Kolb *et al.*, 2010; Gorelik *et al.*, 2012; Zou *et al.*, 2011; Guo *et al.*, 2015). Crystallographic data are collected most efficiently by continuously rotating the crystal (Dauter, 1999). The rotation method has been the standard approach for data collection in protein crystallography for the last four decades (Arndt & Wonacott, 1977). In electron crystallography, alignment of the crystal with the rotation axis is not always straightforward and the rotation stages are not always as accurate as desired, which prompted enhancing the method, with either conical beam precession (Vincent & Midgley, 1994; Kolb *et al.*, 2007, 2008; Gemmi *et al.*, 2013) or beam tilt (Zhang *et al.*, 2010; Wan *et al.*, 2013; Yun *et al.*, 2015). Recently, continuous 3D data collection from protein nanocrystals was accomplished (Nederlof, van Genderen *et al.*, 2013). The first protein structure of a micron-sized crystal was determined soon after, using discrete rotation steps (Shi *et al.*, 2013). More recently, continuous rotation became the preferred method in protein electron crystallography (Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015). The attractiveness of electron crystallography for macromolecular samples is further encouraged by the observation that a large fraction of seemingly failed crystallization attempts contains nanocrystals (Stevenson *et al.*, 2014, 2016). Nanocrystals may also contain fewer defects than micrometre-sized crystals and lead to better data quality (Cusack *et al.*, 1998; De La Cruz *et al.*, 2017).

The electrostatic scattering potential map, basis for model building, is calculated by a Fourier transform of the phased structure factor amplitudes and assumes kinematic scattering. Dynamic scattering affects (Cowley & Moodie, 1957; Dorset *et al.*, 1992; Glaeser & Downing, 1993), but does not prevent structure solution using electron diffraction data (Dorset, 1995; Glaeser & Downing, 1993; Palatinus *et al.*, 2017). In presence of multiple scattering, the diffraction data can no longer be interpreted with a purely kinematic approximation where $I(hkl) \alpha |F(hkl)|^2$. Structure refinement against electron diffraction data using dynamical scattering theory (Jansen *et al.*, 1998; Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017), is not yet available for protein crystals. However, if the crystalline sample is sufficiently thin, this ensures that the measured data are predominantly kinematic, and should not hamper the structure solution too severely (Cowley & Moodie, 1957). The small crystal volume directly affects data acquisition; smaller crystals require longer exposure to get the same signal-to-noise ratio (SNR) as larger crystals, which results in more radiation damage. Radiation damage is a major limiting factor in the study of macromolecules (Henderson, 1995; Owen *et al.*, 2006), thus diffraction data need to be collected under low dose cryo-conditions, and sensitive, low-noise electron detection is imperative.

Previously, we used a single quad Medipix detector (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013), and Timepix detector (van Genderen, Clabbers *et al.*, 2016) of 512×512 pixels (55×55 μm^2 pixel size). For very well ordered crystals

this detector size is sufficient for resolving up to 50 orders of diffraction. However, for protein crystals with larger unit cells, preventing overlap between adjacent Bragg spots may impose a (virtual) detector distance that limits the resolution of the diffraction patterns[17]. Tiling of multiple Timepix quad detectors to bigger arrays can overcome these difficulties. Therefore, we developed a novel in-house designed 1024×1024 pixel Timepix hybrid pixel detector (55×55 μm^2 pixel size).

Detector features that are of particular interest for electron diffraction are the absence of read-out noise, a high dynamic range and the ability to distinguish between the signal of diffracted electrons and that of the high X-ray background that is inherently present in any TEM (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013; van Genderen, Clabbers *et al.*, 2016) These features require a counting detector, a concept that has recently also been introduced in monolithic and CMOS detectors. The hybrid pixel detectors (like the one employed here) only count high-energy electron hits in counting mode if the energy deposited in the silicon sensor layer for a single pixel is higher than a user-defined threshold during a clock cycle (Llopart *et al.*, 2002, 2007). This allows a linear detection range of more than 10^6 electrons per pixel per second in counting mode. Monolithic and CMOS detectors count after the frame has been read out. So, for these detectors, the dynamic range per pixel in counting mode cannot exceed about 1/10th of the number of frames that can be read out per second. This dynamic range is many orders of magnitude smaller than the dynamic range of hybrid pixel detectors.

Monolithic detectors are also more radiation sensitive than hybrid pixel detectors, because the electrons directly hit the integrating read-out electronics of the detector. Since electron diffraction data can have spikes of high intensity at low resolution and in Bragg peaks, monolithic detectors are currently not used for measuring electron diffraction data. However, in hybrid pixel detectors, the high-energy electrons are stopped by the silicon senor layer that is bump-bonded to the counting and integration electronics (McMullan *et al.*, 2007, 2009; Faruqi & McMullan, 2011). The integrating electronics of CMOS detectors can be shielded by a phosphor, at the expense of an increased point spread. So, the hybrid pixel detectors sacrifice pixel size to achieve radiation hardness, a high dynamic range and MHz counting mode. Pixel size is less important in diffraction data acquisition than in imaging, since the resolution of the data is not determined by the level of detail on the detector, but by the number of diffraction orders that can be resolved (Nederlof, van Genderen *et al.*, 2013).

Here we present structure determination from a very thin single protein nanocrystal with a diffracting volume of only 0.14 μm^3. Diffraction data were acquired using the rotation method on a novel Timepix hybrid pixel detector electron

---

[17] Just as an example: assuming a unit cell with 100Å axes in a well ordered protein crystal, and a preferred minimum distance of 5 pixels between adjacent Bragg peaks, the maximum attainable resolution for a single quad is approximately 3.5Å, if the direct beam is centered on the detector. For the same crystal, tiling four quads as presented here would increase the maximum resolution at which Bragg spots can be identified to beyond 1.0Å, again assuming a central direct beam.

diffraction camera, designed for electron crystallography. Standard data processing procedures and software as commonly used in macromolecular X-ray crystallography were adopted for electron diffraction data with minor adaptations. We discuss instrumentation and data acquisition throughout structure solution, model building and refinement.

3.2   Methods

3.2.1   Data acquisition

Electron diffraction data were acquired on a FEI Talos TEM (Center for Cellular Imaging and NanoAnalytics, Basel, Switzerland) and a FEI Titan Krios TEM (NeCEN, Leiden, the Netherlands). Both microscopes were equipped with a Timepix hybrid pixel detector (1024×1024 pixels, 55×55 μm^2 pixel size). We developed a prototype of such a tiled detector camera of 2 by 2 Timepix quad detectors (see Appendix A1), which gave an effective array of 1024x1204 pixels (Fig. 3.1). The Timepix quad cannot be abutted without gaps of ~35 pixels (horizontal) and of ~175 pixels (vertical). The former gap is imposed by the sensitive silicon layer being slightly larger than the pixel array, and the latter imposed by the presence of the readout wire-bonds on opposing sides of the detector chip.

Because high electron fluxes may be focused in Bragg spots, the energy of the incident electron should be completely deposited in the sensor layer to prevent any damage to the readout ASIC that is underneath. For 200 keV and 300 keV electrons the potential scattering distance is approximately 225 µm and 450 µm respectively (McMullan *et al.*, 2007, 2009; Faruqi & McMullan, 2011). For the prototype, we used a 300 µm sensitive silicon layer. A thicker sensitive layer was considered which would allow the use of 300 keV electrons. However, because of the perpendicular impact of a 300 keV incident electron with the detector, on average the first pixel and the last pixel of its track get the highest deposited dose. This means that at the energy threshold used for each pixel (~60keV), the electron is counted one time (70%) or two times (30%) (McMullan *et al.*, 2007, 2009; Faruqi & McMullan, 2011). This means that the Brag spot is spread out over a larger area. To reduce this effect, we opted for 200 keV electrons.

Hen egg-white lysozyme nanocrystals were prepared as described earlier (Nederlof, van Genderen *et al.*, 2013). The microscope was operated at 200 kV and aligned for diffraction with a parallel beam that had a diameter of 2.0 µm and 1.7 µm in microprobe mode for respectively the Talos and the Titan Krios TEMs. EM grids were scanned for nanocrystals in imaging mode at 4k-10k magnification. Once a suitable crystal was found, the crystal was centred on the rotation axis and the beam was centred on the crystal. Diffraction data were collected with the rotation method (Arndt & Wonacott, 1977), with continuous crystal rotation and shutterless data acquisition (Hasegawa *et al.*, 2009). A constant rotation of the goniometer was set using the *TADui* (FEI) and *TEMspy* (FEI) interfaces of respectively the Talos and the Titan Krios. Independently, a fixed frame exposure time was set with the *SoPhy* software (Amsterdam Scientific Instruments) for controlling the detector readout. Hence each frame received the same electron dose and captured a constant rotation increment, like in the rotation method for X-ray crystallography. Datasets were collected with different fixed frame exposure times (see Appendix A2). The dead time

44

of the detector during read-out amounted to 4-10% of the exposure time. During data acquisition, the dose rate on the Talos was ~0.017 $e^-.Å^{-2}.s^{-1}$. The electron flux on the Titan Krios was approximately 20 million electrons per second, amounting to a dose rate of ~0.08 $e^-.Å^{-2}.s^{-1}$ on the crystal (see Appendix A2).
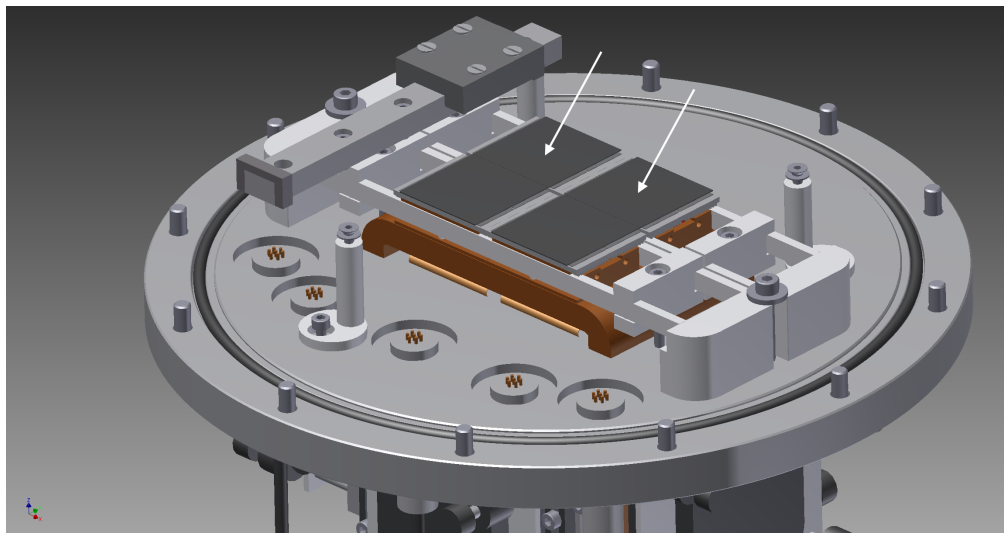

Figure 3.1 The flange design of the camera housing, including the Timepix hybrid pixel detector in the centre (see Appendix A1). The tiled detector assembly holds four Timepix-quads (512×512 pixels each). The dark grey top layers pointed out by the arrows, represent the sensitive silicon layers of a pair of Timepix quads and the light grey slabs below represent the chip board. The gaps between the chips are necessary to accommodate the wire bonds to the read-out boards.

3.2.2 Data processing

Output frames from the tiled detector were interpolated on an orthogonal grid and converted into *PCK* format (Abrahams, 1993) based on positioning and orientation of the four individual Timepix-quads (Fig. 3.1, 3.3). We observed a small but significant elliptical distortion from powder diffraction patterns of an aluminium diffraction standard, both before and after acquiring data. The distortion could not be modelled by a detector tilt. We determined the magnitude and orientation of the distortion (Fig. 3.3A). Correction tables for *XDS* were generated by first creating a fake brass plate pattern based on the distortion parameters, using the program *geocorr.f90* kindly provided by Dr. Wolfgang Kabsch. The calculated geometric correction tables were used with the PILATUS template from *XDS*, with keywords X-GEO_CORR and Y-GEO_CORR (Kabsch, 2010*b*).

The effective detector distance was calibrated using aluminium powder diffraction patterns, after correcting for the elliptical distortion (Fig. 3.3A). The orientation of the rotation axis was initially estimated by identifying reflections close to the rotation axis, which have a wider rocking curve. The angular frame width was assumed to be constant and was determined by dividing the total rotation range by the number of frames. Data were processed with *XDS* (Kabsch, 2010*b*). Since the unit cell parameters are unusual for lysozyme, and quality indicators of electron diffraction data are very different to X-ray diffraction, we confirmed the experimental parameters with *RED* (Wan *et al.*, 2013) that enables quick, routine inspection of electron diffraction patterns in 3D reciprocal space. After applying corrections for the

elliptical distortion, *XDS* found the unit cell dimensions with sufficient accuracy for data processing. Without applying these corrections, the elliptical distortion was too large for *XDS* to home in on the correct unit cell. The rotation axis parameters were refined during data integration. The angular frame width was refined by minimising the deviation of the unit cell angles from an orthorhombic cell (see Appendix A2). With this, *XDS* suggested Laue group mmm, consistent with space group $P2_12_12$ (see Appendix A2).

3.2.3   Structure solution

Data sets were scaled with *XSCALE* (Kabsch, 2010*b*)*,* converted to *MTZ* format with *POINTLESS* (Evans, 2006) and merged with *AIMLESS (Evans & Murshudov, 2013)*. Structure factor amplitudes were obtained with *TRUNCATE* (Winn *et al.*, 2011). A poly-Ala model of tetragonal lysozyme (2ybl) (De La Mora *et al.*, 2011) was created using *CHAINSAW (Winn et al., 2011)*. The poly-Ala monomer was used for a search in all orthorhombic primitive Sohncke groups in molecular replacement with *PHASER* (McCoy *et al.*, 2007). The Matthews coefficient suggested that the crystal contained two monomers per asymmetric unit and *PHASER* unequivocally identified the rotation and translation parameters of both monomers and confirmed the space group $P2_12_12$. Side chains were placed by automated model building with *BUCCANEER/REFMAC5* (Cowtan, 2006; Murshudov *et al.*, 2011). For the merged data three side chains were missing after autobuilding, although in all three instances clear difference potential was observed in the map (see Appendix A4). Thus after inspecting the model and map, these three missing residues where fitted using *COOT* (Emsley *et al.*, 2010). We did not further enhance the models by manual rebuilding, to evaluate to what extent refinement was able to correct errors in the model.

3.2.4   Refinement

The model was optimized by *PDB_REDO* (Joosten *et al.*, 2014), where electron scattering factors were set by placing 'EXPDTA ELECTRON CRYSTALLOGRAPHY' into the PDB header. The model was then refined with (Murshudov *et al.*, 2011) using NCS restraints. To ensure convergence, the input model was refined for 1,000 cycles by *REFMAC5* (see Appendix A5). Electron scattering factors were set in *REFMAC5* with the keyword 'SOURCE ELECTRON MB'. To calculate the map coefficients, *REFMAC5* was set to not restore unobserved reflections with the keyword 'MAPC FREE EXCLUDE'.

We validated refinement in *REFMAC5* with $R_{complete}$ instead of R_{free}. When considering data sets with less than about 10,000 unique reflections, which is the case for our data, calculating $R_{complete}$ is preferred (Brunger, 1997). The $R_{complete}$ validation method allows all reflections to be used for refinement, and thus our R_{work} is equivalent to R1. R1 defines how well the model explains all observed reflections. Like R_{work}, it is likely to be affected by model bias. The $R_{complete}$ was calculated afterwards according to standard procedures with a 0.2% test set size (Luebben & Gruene, 2015). Briefly, first all non-measured observations were removed from the reflection file with *SFTOOLS* (Winn *et al.*, 2011). Then 500 separate, non-overlapping and unique test sets were randomly created with *FREERFLAG* (Winn *et al.*, 2011), each containing 0.2% of the observed structure factor amplitudes. So, when combined, these test sets represent all data. Then, 500 independent refinements

were performed until convergence, each time omitting a different test set. Each refinement started with the same (final) model from which R1 had been calculated. After each of the 500 validation refinement cycles had converged, the values of F_c were calculated from the resulting model. Only $F_c(h)$ values corresponding to reflections that had been omitted from that particular cycle (and thus were not biased by that cycle) were extracted. Then, all these extracted reflections from each of the 500 independent refinement cycles were combined into a single reflection file representing the unbiased $F_c(h)$ values corresponding to all observed structure factor amplitudes. Finally, R$_{complete}$ was calculated by comparing these excluded data with the observed structure factor amplitudes. R$_{complete}$ is therefore not biased by the model, just like in standard R$_{free}$ calculations, yet it is a more robust measure of model bias, especially for incomplete and/or sparse data, because all reflections contribute to its value.

3.3   Results

3.3.1   Data integration

Data were acquired from a single cryo-cooled lysozyme nanocrystal with dimensions $200{\times}500{\times}1400$ nm^3 (Fig. 3.2). The crystal was found in a thin layer of vitreous ice over a hole in the carbon support film of the EM grid. The crystal was continuously rotated for 38.2° with an angular increment of 0.076° per frame in a 2 μm diameter beam. The central beam was positioned such that during data collection only the tip of the crystal over the hole was illuminated, thus eliminating any background noise from the amorphous carbon in the support film. In our experience, it was favourable to collect data from crystals that were still attached on one end to the carbon support. Crystal bending upon exposure to the beam was observed in cases where the crystals were suspended in vitreous ice, but not attached to the carbon, probably due to charging effects. The total dose received by the crystal did not exceed ~4.4 e$^-$.Å^{-2}. Data of the single crystal were integrated to 2.1Å resolution (Table 3.1, Fig. 3.3). The single crystal data had a completeness of only ~50%, but were sufficient for full structure solution (Table 3.1). To investigate the inter-crystal consistency of the data with that of other nanocrystals, we collected additional diffraction data (Table 3.1). After merging with diffraction data from six other nanocrystals (Appendix A2), diffracting to 2.5-3.0Å rather than 2.1Å, overall completeness increased to ~60% (Appendix A3). The limiting factors were radiation damage and preferred orientation of the crystals, combined with the limited rotation range of the goniometer holding the EM grid. At higher angles, the distance that the electrons have to travel through the surrounding amorphous ice and the protein crystal can become too large for accurate data acquisition. These limitations are inherent to current implementations of electron diffraction: others have collected up to ~44 degrees (Nannenga, Shi, Leslie *et al.*, 2014), ~61 degrees (Nannenga, Shi, Hattne *et al.*, 2014), and ~40 degrees (Yonekura *et al.*, 2015). These data were collected on crystals that were significantly larger than our nanocrystals, and in case of Yonekura *et al.* needed merging from 58 and 99 crystals (Table 3.2). Further, we compared the differences in measured intensities of Friedel pairs after scaling but before merging of the single crystal dataset (Fig. 3.4). The variation in Friedel pair intensities for the single crystal data is low, even when compared with X-ray data from a small macrocyclic depsipeptide crystal that could be solved by direct methods.
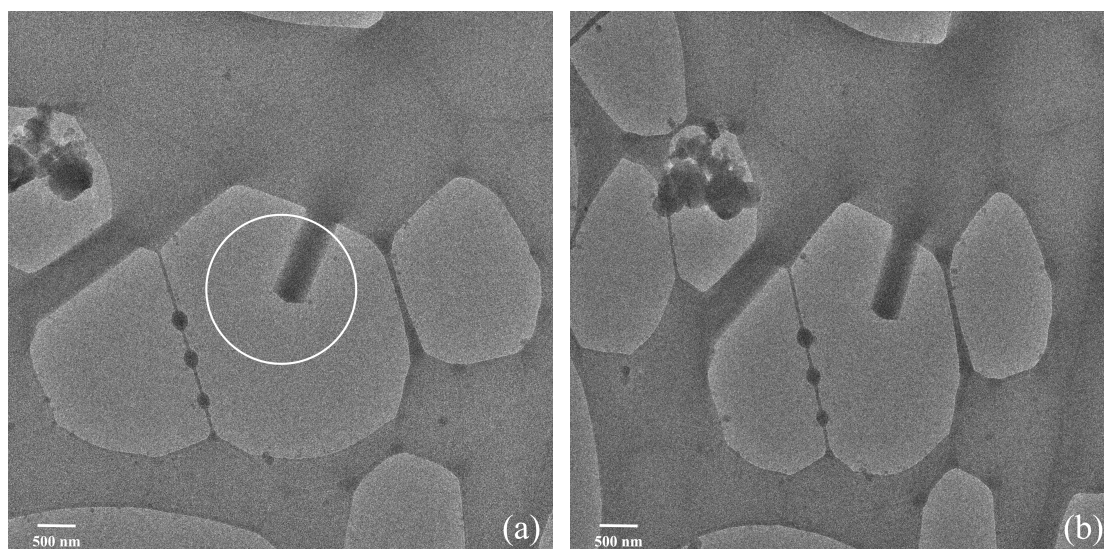
Figure 3.2 Micrographs of a single three-dimensional lysozyme crystal ($200 \times 500 \times 1400 nm^3$) in a thin layer of vitreous ice across a hole on the Lacey carbon EM grid at (a) +20° tilt angle, and (b) +50° tilt angle. Diffraction data were acquired with a 2.0 μm diameter parallel beam in microprobe mode indicated by a circle in (a). During data collection only the tip of the crystal was kept in the central beam to limit noise from the carbon support. The width of the crystal at both tilt angles was used to derive its dimensions, the length was measured from the tip of the crystal to the edge of the carbon that was the maximum size of the crystal within the central beam at any point during rotation.

3.3.2 Structure determination

Molecular replacement with a monomeric poly-Ala lysozyme model derived from a different, tetragonal space group, successfully located a single monomer in the asymmetric unit. Then it also placed the second one. A Z-score of 22.5 is sufficiently high above the threshold of 8.0, indicating a successful structure solution (McCoy *et al.*, 2007). Automated model building with *BUCCANEER/REFMAC5* (Cowtan, 2006; Murshudov *et al.*, 2011) was used for reconstructing the side chains (Fig. 3.6A,B). The densities that are shown were not refined. Hence, they look poor. Yet they show that the molecular replacement was successful, as they demonstrate that the phases from a poly-Ala MR solution, allow placing side-chain density of atoms that were not included in the MR model. Subsequent refinement using only the observed reflections improved the quality of the map, *e.g.* the refined density suggests that residue Ala9 is a *cis*-peptide, different from the tetragonal MR model (Fig 3.6C). However, the 1.9Å resolution X-ray structure of the same orthorhombic polymorph confirms that in this crystal form, the peptide is *cis*. This strongly validates the quality of our structure solution. Density that was refined according to standard, default protocols shows continuous, high-resolution density (Fig. 3.6D). At 2.1Å resolution, and in particular with incomplete data, maps are prone to model bias. To estimate how much information our data contain, we calculated rmsd values between an X-ray model of orthorhombic lysozyme in the same space group and with a similar unit cell (4r0f) (Sharma *et al.*, 2016) and *i.* our refined model with autobuilt side chains (rmsd=0.7Å), and *ii.* our model with side chain rotamers that are statistically preferred in proteins (rmsd=1.1Å) (Appendix A6). This indicates that placing of side chain residues is based on real information contained in the single crystal data. These results demonstrate the validity of the diffraction data, despite relatively poor merging and model statistics compared to complete X-ray data (Table 3.1).
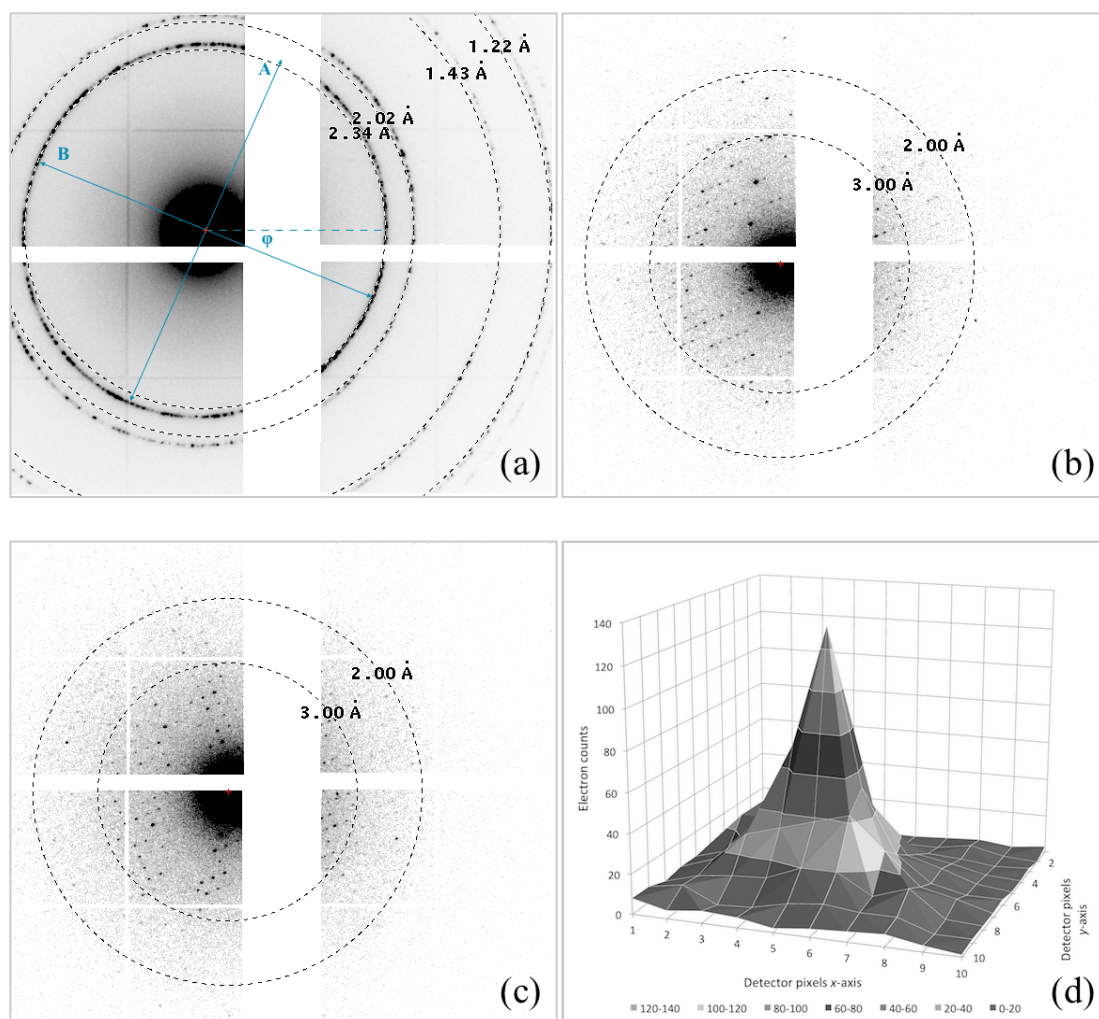
48

Figure 3.3 Electron diffraction data acquisition; (a) measured powder pattern of an aluminium diffraction standard after correcting for the tiling offsets of the Timepix quad ASICs. An elliptical distortion can be observed with a deviation of 1.043 (=A/B) at an angle of φ=21.3°. Diffraction from the single lysozyme crystal summed over 1.0° of rotation (b) from -17.0° to -16.0°, and (c) from -6.0° to -5.0°. Crosses on individual quads are due to corrections for larger border pixels as described (Nederlof, van Genderen *et al.*, 2013; van Genderen, Clabbers *et al.*, 2016), those pixels were not taken into account for processing of the protein diffraction data. Note that due to radiation hardness of the detector, no backstop was required. Resolution rings were plotted with *ADXV* (http://www.scripps.edu/tainer/arvai/adxv.html), (d) a typical spot profile of a high intensity peak at 16.33Å resolution recorded on a single frame with an angular increment of 0.076° per frame, at a dose rate of ~0.01 e⁻.Å⁻² per frame, shown in a 10×10 pixel array with 0.055×0.055 mm² pixel size.

3.4   Discussion and conclusions

Here we show the structure determination from electron diffraction data of a single continuously rotated cryo-preserved protein 3D nanocrystal with a diffracted volume of at least an order of magnitude smaller than was previously possible. For all steps of the structure elucidation, we used standard procedures and software that were originally developed for X-ray protein crystallography. Completeness of the data is low, but because there are two molecules in the asymmetric unit we could apply non-crystallographic symmetry restraints. This NCS was exploited during refinement, and the deleterious effects of data incompleteness could be mitigated. Completeness is

also determined by crystallographic symmetry. For instance, if the lysozyme nanocrystal would have had tetragonal symmetry, instead of orthorhombic symmetry, completeness with the same rotation range would have been 84% or more.

Dynamical scattering has been a longstanding objection against electron crystallography of 3D protein crystals. It causes the intensity of each Bragg peak to be affected by the structure factors of the other Bragg peaks that are recorded in the same exposure. When recorded in a different crystal orientation, its measured intensity will therefore be different even after scaling and Lorentz corrections. This effect also causes differences between measured intensities of symmetry equivalent reflections (Glaeser & Downing, 1993). Dynamical scattering can compromise structure solution of crystals of macromolecules, since current phasing methods and refinement procedures do not account for its effects. Thin crystals minimize the effects of dynamic scattering and on the basis of multi-slice simulations, it has been suggested that the maximal thickness of a protein crystal still allowing structure solution, is about 100 nm for 200 keV electrons (Subramanian et al., 2015), but these calculations ignore inelastic scattering, which for organic samples is three times more prevalent than elastic diffraction.

X-ray data where the intensities of Friedel pairs correlated as poorly as in our electron diffraction data have been solved and refined using standard procedures (Fig. 3.4), indicating that the noise that our data suffered due to dynamically scattering, was tolerable. Furthermore, we show a F_o vs. F_c graph of our electron diffraction data after model refinement (Fig. 3.5). It shows a linear correlation for the higher intensity, but at lower intensity the value of F_o is overestimated. On average, dynamical diffraction is anticipated to affect weaker reflections more than strong reflections. So on average, weak spots close to intense spots will become more intense, whereas intense spots close to weak spots will hardly be affected (Weirich et al., 2000). Assuming an expected complex-valued error $E(h)$, that is uncorrelated to $F(h)$, we can infer a hyperbolic relationship between the expected value of $\langle |F_o| \rangle$ and $|F_c|$:

$$\langle |F_o| \rangle = \sqrt{|F_c|^2 + \langle |E(h)| \rangle^2} \tag{19}$$

Out data indeed show such a relationship (Fig. 5.6). Merging reduces the random errors of the data, and should also reduce some of the dynamical effect, provided the merged crystals have different orientations. However, the fitting parameters from the F_o vs. F_c graph for the merged data are similar compared to the single crystal data (Appendix A7). The expected error increases at lower resolution (Appendix A7), indicating an increased dynamic effect within this resolution range. These observations suggest weak spots to become relatively more affected by other sources of noise with increasing resolution. Nevertheless, although the data were very weak and compromised by dynamical scattering, they were of sufficient quality for a realistic molecular replacement solution.

Radiation damage and the small volume of the crystal presented here severely limit the SNR and make data acquisition more challenging. We could improve the SNR substantially with a more accurate and sensitive detector. Previously, we measured 3D nanocrystals, similar to the polymorph presented here, using CCD detectors and image plates (Georgieva et al., 2007, 2011). For protein crystals that

were of similar diffracting volume as reported here, we could never measure more than a few diffraction patterns of high resolution data with a CCD detector or image plate before radiation damage became too severe. A quantitative comparison between image plates and a Medipix hybrid pixel detector indicated a substantial improvement offered by the latter (Georgieva *et al.*, 2011; Nederlof, van Genderen *et al.*, 2013). Hybrid pixel detectors like Timepix and Eiger (Llopart *et al.*, 2002, 2007; Johnson *et al.*, 2012), are well suited for measuring high-energy electrons (McMullan *et al.*, 2007), and can overcome difficulties in detecting weak peaks, *e.g.* for CCD and CMOS detectors (Hattne *et al.*, 2016; Rodriguez & Gonen, 2016). An inherent drawback of the detector design is the loss of information in the gaps between individual tiles, each tile being a 512×512 quad Timepix. Because quads are connected by wire bonds to their readout electronics, these gaps are unavoidable. Without the gaps the data would have been more accurate, but not much more complete, as the geometry of the experiment allowed collecting the data of the Friedel equivalent of most of the missing reflections (Fig. 3.3). The deleterious effect of the gap on data completeness can be further mitigated by aligning the rotation axis with the large gap. This would mainly loose reflections with Lorentz factors that are so high that they would be discarded by the data processing software anyway (Fig. 3.3).

The total illuminated volume of the single nanocrystal we used for data acquisition described here was only ~0.14 μm^3 (Fig. 3.2). The data provided sufficient information for structure solution, model building and refinement (Table 3.1, Fig. 3.6). The total diffracting volume of the crystal is no more than 6×10^5 unit cells (Table 3.2). A comparison with previously solved structures of macromolecules by electron diffraction recorded on CCD and CMOS detectors show that these used significantly larger crystals (Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015; Hattne *et al.*, 2015). Since the quality of diffraction data of protein crystals is in the limiting case determined by the crystallinity of the sample, these data need to be interpreted with great care and should only be used to infer trends. To correct for differences in unit cell volumes, we determined the number of unit cells used for structure solution. Resolution and crystal symmetry will also affect the amount of unique data within a data set. After correcting for these effects, the hybrid pixel detector allowed structure solution using at the very least an order of magnitude less unique diffracted intensity than obtained previously with other detectors (Table 3.2).

Additional hardware modifications may further benefit electron diffraction studies of macromolecular compounds, *e.g.* reliable and well-integrated goniometer tilt (Yonekura *et al.*, 2015), and using an in-column energy filter (Yonekura *et al.*, 2015). The data presented here show that with a highly sensitive and accurate hybrid pixel detector, now also nanometre-sized crystals of macromolecules are possible targets for 3D protein electron crystallography, which has the advantage of reducing the effects of dynamical diffraction. Perhaps also data from micron-sized crystals can be measured more accurately, although it needs to be investigated further if for such crystals data accuracy is limited by detector sensitivity or the amount of dynamical diffraction. The introduction of hybrid pixel detectors had a major positive impact on protein X-ray crystallography due to their high speed, increased sensitivity and high dynamic range (Broennimann *et al.*, 2006). Based on the results we present here, we suggest that specialized hybrid pixel detectors may have a similar impact on electron diffraction studies of protein crystals.
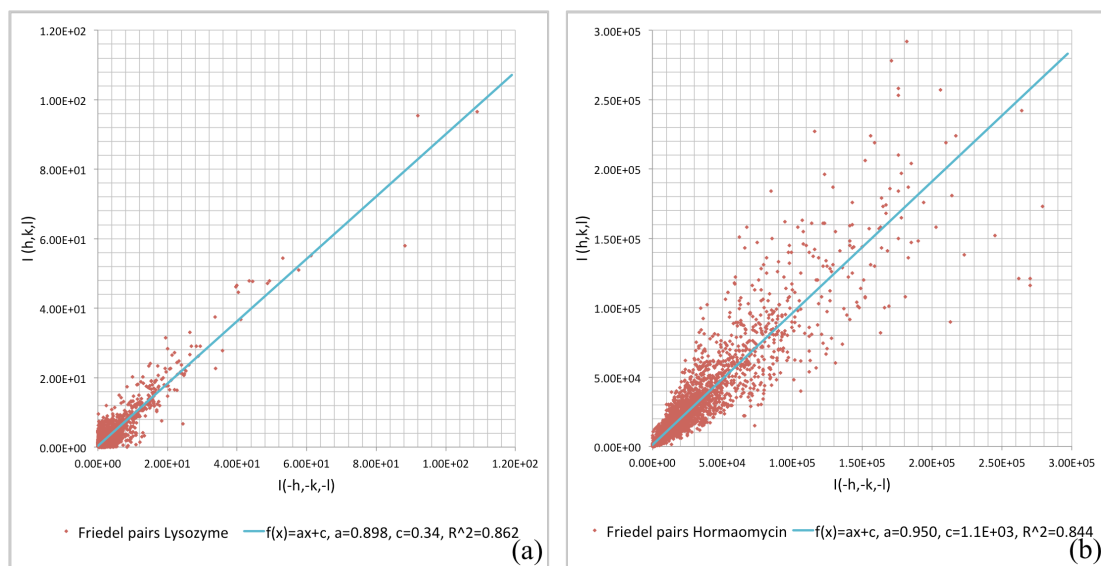
Figure 3.4 Differences in intensities of Friedel pairs after scaling plotted for: (a) single lysozyme crystal used for structure solution with $R_{Friedel} = 0.329$, and (b) X-ray data of hormaomycin, a macrocyclic depsipeptide in space group P1 with $R_{Friedel} = 0.151$ (Gruene, Sheldrick *et al.*, 2014).
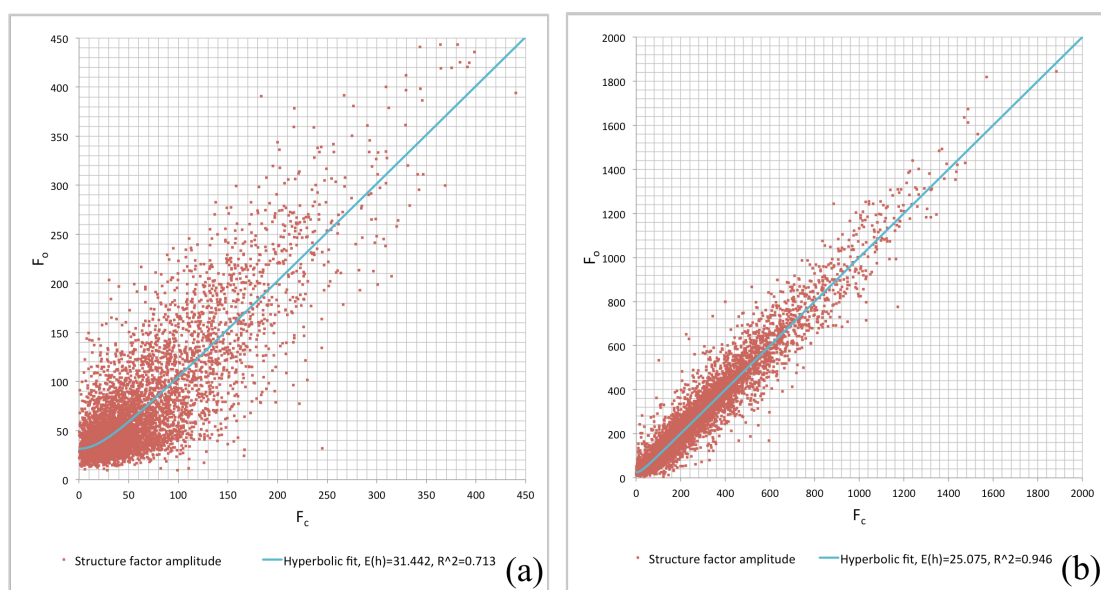
Figure 3.5 F_o vs. F_c graphs for (a) electron diffraction of a single lysozyme nanocrystal and (b) an X-ray dataset at 1.6Å of cubic (bovine) insulin. The data were LS fitted with a hyperbolic function described by $\langle |F_o| \rangle = \sqrt{|F_c|^2 + \langle |E(h)| \rangle^2}$. F_o vs. F_c graphs for only the low resolution part of the single crystal data, and for the merged crystal data are shown in Appendix A7.
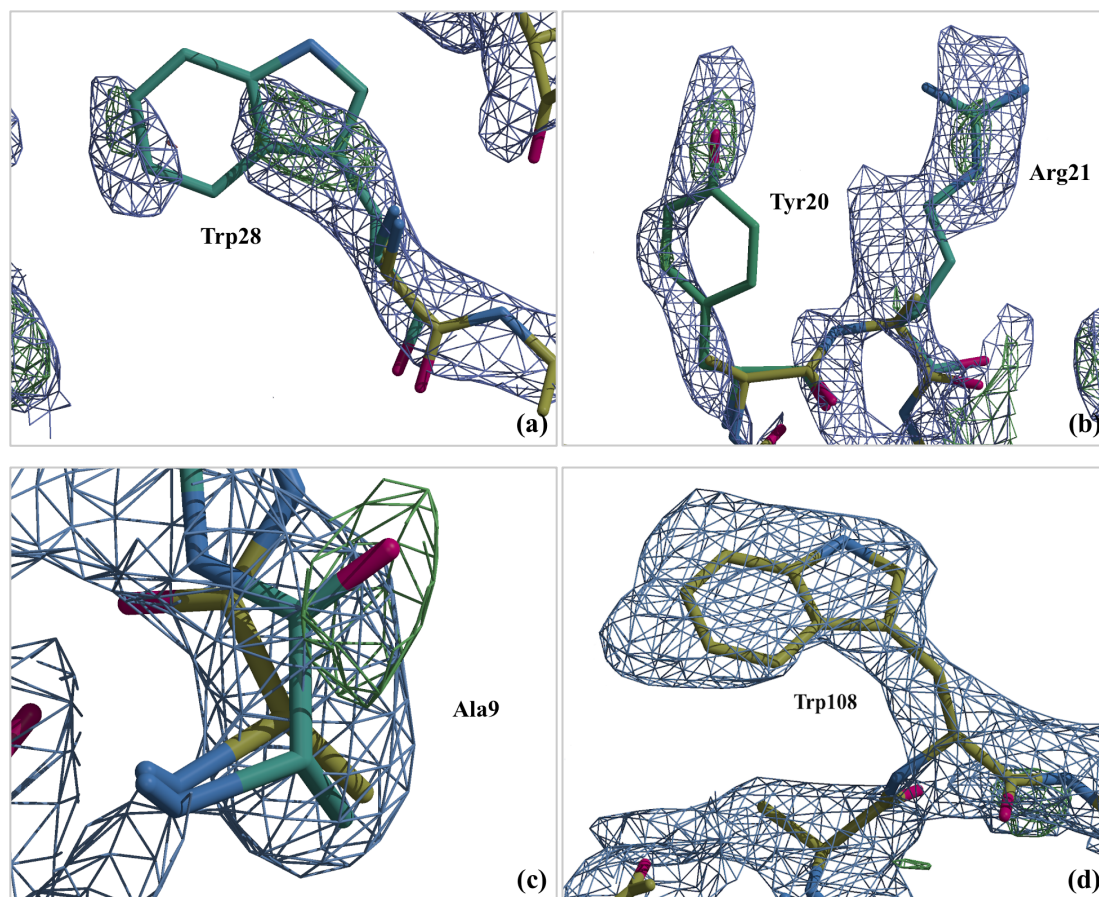
Figure 3.6 Automated model building using the single crystal data; (a) after molecular replacement with the poly-Alanine monomer (yellow carbon), the difference map shows the position of bulky side chain residues as Trp28 as placed during autobuilding by *BUCCANEER* (turquoise carbon), and (b) for Tyr20 and Arg21. The map is stretched, which is typical for incomplete data; as always with poor map quality, careful interpretation of the region is required. The map improves after side chain reconstruction with *BUCCANEER* and refinement with *REFMAC5*, (c) the refined density suggests that Ala9 (yellow carbon) is a cis-peptide, this is confirmed by the X-ray structure of the same polymorph (turquoise carbon, PDB ID 4r0f) the peptide is cis. Refinement using standard protocols can further improve the map and shows continuous density, (d) for a Trp108 side chain residue in the A chain of the single crystal model. All density is shown at a standard contour level of 1.2σ.

Table 3.1 Data integration and refinement statistics

	Single crystal (*5o4w*)	Merged data (*5o4x*)[18]
Data integration		
Space group	*P2₁2₁2*	
Unit cell dimensions		
a, b, c [Å]	104.56, 68.05, 32.05	
α, β, γ [°]	90.0, 90.0, 90.0	
Number of crystals	1	7
Resolution [Å][19]	41.46-2.11 (2.17-2.11)	57.03-2.11 (2.17-2.11)
R_{merge} [%]	26.3 (56.6)	39.8 (64.0)
I/σI	2.6 (1.0)	2.7 (1.0)
Completeness [%]	49.5 (49.8)	61.7 (49.8)
Reflections	12601 (1462)	41191 (1462)
Unique reflections	6749 (545)	8560 (545)
Structure solution		
TF Z-score	22.5	26.7
LLG score	395	535
Refinement		
Reflections	6717	8503
R1 [%][20]	33.5	26.4
$R_{complete}$ [%][21]	35.0	27.9
<B> [Å²]	24.0	27.0
RmsZ bonds	0.92	0.85
RmsZ angles	1.27	0.97
Ramachandran		
Favoured, allowed, outliers [%]	93.7, 5.9, 0.4	98.4, 1.6, 0.0

---

[18] Data integration statistics for individual crystals used for merging are shown in Appendix A2, data merging statistics are presented in Appendix A3.

[19] Values in parentheses correspond to the highest resolution shell, the data were truncated at I/σI > 1.0 (Diederichs & Karplus, 2013).

[20] $R1 = \sum_{hkl}||F_o(hkl)| - |F_c(hkl)||/\sum_{hkl}|F_o(hkl)|$

[21] We present R1 and $R_{complete}$ instead of R_{work} and R_{free}. With less than 10,000 unique reflections $R_{complete}$ is preferred over R_{free} since it is calculated from all reflections (Brunger, 1997; Luebben & Gruene, 2015). Since all structure factors are used in turn this leads to a more robust calculation than R_{free}. With this validation method, the actual refinement uses all reflections; hence R_{work} is equivalent to R1.

Table 3.2 Relative crystal volume used for structure determination in recent macromolecular electron diffraction studies

	PDB ID	Detector	d [Å]	Space group	Unit cell dimensions [Å]	No. of xtals	Individual crystal size and total diffracted volume [μm³][22]	No. of unit cells[23] (×10⁶)	Relative unique diffracted intensity[24] (×10⁶)
Lysozyme	5o4w	Hybrid pixel	2.1	P2₁2₁2	105x68x32	1	0.2x0.5x1.4 (0.14 μm³)	0.6	1.4
Lysozyme (Nannenga, Shi, Leslie *et al.*, 2014)	3j6k	CMOS	2.5	P4₃2₁2	76x76x37	1	0.5x2.0x2.0 (2 μm³)	9.4	18
Catalase (Nannenga, Shi, Hattne *et al.*, 2014)	3j7b	CMOS	3.2	P2₁2₁2₁	68x172x182	1	0.15x4.0x6.0 (3.6 μm³)	1.7	14
Catalase (Yonekura *et al.*, 2015)	3j7u	CCD	3.2	P2₁2₁2₁	69x173x206	58	0.1x2.0x2.0 (23 μm³)	9.4	77
Ca²⁺ATPase (Yonekura *et al.*, 2015)	3j7t	CCD	3.4	C2	166x64x147 (β=98)	99	0.1x2.0x2.0 (40 μm³)	25	490

[22] The illuminated crystal size used for data acquisition is estimated from reported crystal dimensions and used aperture sizes; for the structures with PDB accession code 3j7u and 3j7t (Yonekura *et al.*, 2015) we assumed that the plate-like crystals had a surface area of 2×2 micron. The total diffracted volume (indicated by the number between brackets) takes the number of crystals required for the 3D dataset into account.

[23] The required number of unit cells was calculated by dividing the total diffracted volume by the unit cell volume.

[24] We calculated the relative unique diffracted intensity by dividing the number of required unit cells (given in the previous column), by the number of asymmetric units in the unit cell and multiplying the result with the cube of the resolution of the dataset.

Electron diffraction data processing with *DIALS*

Electron diffraction is a relatively novel alternative to X-ray crystallography for the structure determination of macromolecules from three-dimensional nanometre-sized crystals. The continuous-rotation method of data collection has been adapted for the electron microscope. However, there are important differences in geometry that must be considered for successful data integration. The wavelength of electrons in a TEM is typically around 40 times shorter than that of X-rays, implying a nearly flat Ewald sphere, and consequently low diffraction angles and a high effective sample-to-detector distance. Nevertheless, the *DIALS* software package can, with specific adaptations, successfully process continuous-rotation electron diffraction data. Pathologies encountered specifically in electron diffraction make data integration more challenging. Errors can arise from instrumentation, such as beam drift or distorted diffraction patterns from lens imperfections. The diffraction geometry brings additional challenges such as strong correlation between lattice parameters and detector distance. These issues are compounded if calibration is incomplete, leading to uncertainty in experimental geometry, such as the effective detector distance and the rotation rate or direction. Dynamic scattering, absorption, radiation damage and incomplete wedges of data are additional factors that complicate data processing. Here, recent features of *DIALS* as adapted to electron diffraction processing are shown, including diagnostics for problematic diffraction geometry refinement, refinement of a smoothly varying beam model and corrections for distorted diffraction images. These novel features, combined with the existing tools in *DIALS*, make data integration and refinement feasible for electron crystallography, even in difficult cases.

Max T.B. Clabbers[1], Tim Gruene[2], James Parkhurst[3], Jan Pieter Abrahams[1,2] and David G. Waterman[4,5]

[1]Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, Basel University, Mattenstrasse 26, CH-4058 Basel, Switzerland

[2]Paul Scherrer Institut (PSI), Department of Biology and Chemistry, CH-5232 Villigen PSI, Switzerland

[3]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot OX11 0DE, England

[4]STFC, Rutherford Appleton Laboratory, Didcot OX11 0FA, England

[5]CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, England

4.1 Introduction

Electron diffraction (ED) allows the structural analysis of nanometre-sized samples of crystalline material. Since the maximal radiation dose is proportional to the sample volume, electron diffraction of organic and macromolecular compounds was long limited to two-dimensional samples (Henderson & Unwin, 1975). In contrast to X-ray crystallography, the three domains, inorganic, organic and macromolecular electron crystallography, developed rather independently of each other (Vainshtein, 1964; Dorset, 1995; Glaeser *et al.*, 2007; Zou *et al.*, 2011). Physical and instrumental limitations, such as miniature sample size or dynamic scattering effects and lens distortions affect data precision. However, several studies show that the model accuracy compares with that of X-ray structures (Weirich *et al.*, 1996; Dorset, 1995; Palatinus *et al.*, 2017). Only about one and a half decades ago, electron diffraction of three-dimensional crystals was pioneered with automated diffraction tomography (ADT) and was further refined with rotation electron diffraction (RED) (Kolb *et al.*, 2007, 2008; Zhang *et al.*, 2010; Gemmi *et al.*, 2015). Recently, single-crystal three-dimensional electron diffraction has also been applied to protein crystals by using the standard rotation method (Arndt & Wonacott, 1977; Nederlof, van Genderen *et al.*, 2013; Hattne *et al.*, 2015; Yonekura *et al.*, 2015; Clabbers *et al.*, 2017). The only very recent use of integration software with profile fitting and scaling is indicative of the independent development of electron diffraction. These methods have been in use for decades in X-ray crystallography, improving the quality of diffraction intensities and their standard uncertainties, whilst enabling heuristic correction for systematic errors (Pflugrath, 1999; Leslie, 1999).

DIALS is a relatively new package for diffraction integration (Winter *et al.*, 2018), designed as an extensible toolkit for the implementation of algorithms relevant to diffraction data analysis. The core set of algorithms is presented as a suite of command-line programs that can be used following simple protocols to integrate data sets collected using the rotation method (Arndt & Wonacott, 1977). Many of these algorithms are implementations of tried and tested methods described in numerous publications over the past three decades (Leslie, 1999; Bricogne, 1986*a,b*, Kabsch, 2010*a,b*). However, the toolkit design of *DIALS* facilitates the construction of new algorithms (Gildea *et al.*, 2014; Parkhurst *et al.*, 2016, 2017). *DIALS* is an open-source project, allowing scientists from outside the core collaboration to contribute software or to use DIALS within their own projects.

To date, *DIALS* development has focused on macromolecular (MX) and chemical crystallography data sets and has been optimized for continuous-rotation data collected in fine slices using photon-counting detectors at synchrotron light sources. Despite this emphasis, with suitable modification of the parameters at certain steps, high-quality results have also been obtained for wide-sliced X-ray data sets recorded on CCD detectors (Keegan *et al.*, 2016; Khasnis *et al.*, 2016). The common fundamental assumption is that reciprocal-lattice points pass through the Ewald sphere by constant-velocity rotation around a single axis. No artificial restrictions on the diffraction geometry are imposed, allowing the modelling of diffraction experiments using a generic vectorial description (Waterman *et al.*, 2016). By default, two measurements, summation integration and three-dimensional profile fitting, are made for each reflection along with estimated errors (Winter *et al.*, 2018). The simplicity of this approach, avoiding the assumptions inherent in the details of any

particular technique, means that *DIALS* is readily adapted for analysis beyond the original scope of its design.

A common feature shared between *DIALS* programs is the global modelling of an experiment, in which data are assumed to be complete before analysis begins. This has some advantages over the traditional approach of processing data by means of a moving window that passes over the complete data set in blocks of a local range of images. One is that the expensive step of integration can be performed with a high level of parallelism, as the experimental model is determined completely ahead of time. A second is that the programs can consider multiple experiments simultaneously without losing track of the connections between them. This feature has particular relevance to the global refinement of diffraction geometry, for which experiments may share some models (Waterman *et al.*, 2016), certain parameters may be constrained to shift together, or restraints may be applied between multiple crystal models. These features can be important for the analysis of electron diffraction data sets, for which determining accurate diffraction geometry may be challenging (Yun *et al.*, 2015), and current technology usually imposes the collection of incomplete wedges of data for each crystal. Here, we discuss the use of *DIALS* for the analysis of electron diffraction data that have been collected using the rotation method. As a motivational example, we describe the stages of data processing with reference to seven data sets collected at 200 keV from orthorhombic crystals of a dimeric form of hen egg-white lysozyme, as previously reported in Clabbers *et al.* 2017 (see §3.3)[25].

4.2   Methods and results

4.2.1   Image formats

The first stage in processing rotation data with *DIALS* is to import the images constituting the data set to form a DataBlock using the *dxtbx* library (Parkhurst *et al.*, 2014). This library contains format-reading classes for the majority of common file formats used in X-ray crystallography. The classes are arranged in a hierarchy from generic classes that contain code to read image data and construct an experimental model solely from metadata contained in the image headers to specific classes that may recognize a particular instrument and can override for incorrect or missing metadata. This feature is important for reading the file formats used in electron microscopy because current instruments usually do not transfer all of the information that is required to reconstruct the experimental geometry. There are three main approaches that can be taken to import electron diffraction data into *DIALS*:

  i.   Externally convert the native format into a format more common for MX. This is the usual approach adopted for data processing with other programs such as *MOSFLM* (Leslie & Powell, 2007) and *XDS* (Kabsch, 2010*b*). For example, data sets have been converted to SMV (Hattne *et al.*, 2015), PCK (Clabbers *et al.*, 2017), or CBF images (Gruene *et al.*, 2018). Where external conversion programs exist, this has the advantage that no coding or understanding of the original file format is required by the user. Often, missing metadata can be

---

[25] A detailed tutorial on processing electron diffraction data using *DIALS* is available online at https://dials.github.io/documentation/tutorials/dials_for_ed.html and can be used to reproduce the results presented here in §4.2 using the experimental data which are available online at https://dx.doi.org/10.5281/zenodo.1250447

supplied during the conversion so that the resulting images contain a proper description of the experiment and no additional overrides are required when importing the data set into *DIALS*. The same set of images can then also be used with other data-processing packages. However, the reliance on an external conversion tool has some drawbacks. There is the scope for errors when metadata are introduced manually during the conversion. The proliferation of conversion tools adds complication for the user and the fidelity of the conversion process must be checked. For example, image-export functions within microscope vendor-supplied software to common formats such as TIFF might not preserve the real pixel intensities, and this fact may not be clear to the user. Even when data are properly converted, the generic readers for standard MX formats may contain assumptions that are not appropriate for electron diffraction, such as the creation of a polarized beam model. Generic readers might also not allow the desired interpretation for sophisticated cases, such as splitting a data array for a multiple-panel detector model or defining masks for certain regions of images.

ii. Extend the *dxtbx* library to recognize native data formats. This approach entails writing a format class (typically a single, small Python module) to contribute to *dxtbx*, following the published description (Parkhurst *et al.*, 2014), and existing examples. This requires knowledge of the native data format and conventions used by *dxtbx*, as well as co-ordination with the *DIALS* developers. The advantage of investing this effort is that once included in the library, the native data format will be supported for all users with no additional conversion steps. In practice, however, where native formats lack the metadata describing the diffraction experiment, this will have to be supplied each time during data import, either by providing parameters at the command line or in a file in the PHIL format, a simple data-interchange format used within *cctbx* (Grosse-Kunstleve *et al.*, 2002). Appendix B1 contains an example of such a file. Format classes for native file types that have now been added to *dxtbx* include image stacks in the TIA Series Data (ESD) format used by software provided with Thermo Fisher (FEI) microscopes and image stacks in Gatan DM4 format.

iii. For local installations, testing or one-off developments for a particular data-processing problem it may be more appropriate to create a format class as a plugin rather than contributing to the *dxtbx* library. There is no difference in the procedure required to implement the class; the resulting Python module should simply be placed in a *.dxtbx* directory in the user's home area and this will automatically be picked up at runtime when required. Various plugins for electron diffraction are collected and can be downloaded and modified freely[26].

---

[26] Plugins are available online at https://github.com/dials/dxtbx_ED_formats
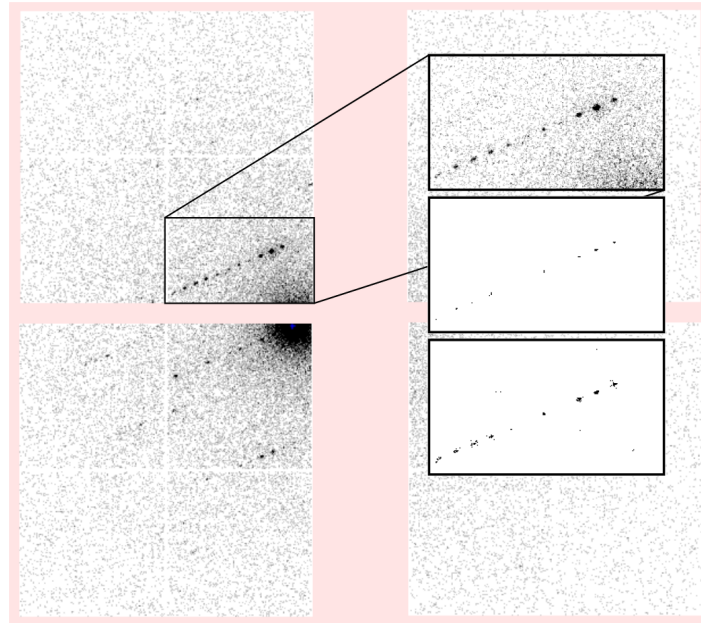
Figure 4.1 A diffraction image from data set 1 using *dials.image_viewer*. The four quads have independent geometry, such that hey are not forced to align on a single pixel grid. The upper inset panel shows a zoomed region of the upper left quad where a clear row of diffraction spots is visible. The middle inset panel shows the 'threshold' image with the default spot-finding settings, which indicates which pixels well be marked as strong during the spot-finding procedure. The lower inset panel shows the same region after spot-finding settings were adjusted for this data set. In this case, this amounted to setting `gain=0.833`, `sigma_strong=1.0`, and `global_threshold=1` as command-line options for the *dials.find_spots* program. The detector gain of 3.0 determined by the format class is already applied before the spot-finding operation; hence the spot-finding gain acts as a multiplier for this value.

The seven lysozyme data sets discussed here consist of diffraction images from a 1024×1024 pixel detector composed of a 2×2 array of Timepix quad detectors (Clabbers *et al.*, 2017). Large gaps between the Timepix quads are imposed by the form factor of each quad. For the original processing of these data by *XDS*, the images were converted into PCK format, in which pixel values were interpolated onto an orthogonal grid, with the gaps forming 'dead' areas of the image array. For processing with *DIALS* we chose a multiple-panel description instead (Parkhurst *et al.*, 2014). The images were converted to CBF without interpretation of the gaps[27]. We created a *dxtbx* format class specific for these images, which represents each quad as a separate panel of a composite detector. In this way, no interpolation is required because each panel has an independent position and orientation; thus, sub-pixel shifts and rotations can be represented precisely. The *dials.image_viewer* takes account of the relative position and orientation of independent panels and displays a composite image projected onto a viewing plane, as shown in Figure 4.1.

A 512×512 pixel Timepix quad is an assembly of four abutting Timepix ASICs, each with 256×256 55 μm pixels. However, the distance between two abutting Timepix ASICs is 350 μm, corresponding to a pitch for the abutting pixels that is about three times that of the other pixels. Since these pixels have a larger surface, they also have a higher probability of collecting more electrons. To correct

---

[27] https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Timepix2cbf

for this non-uniformity, the conversion to CBF splits pixels with an x (and/or y) coordinate that equals 256 or 257 into three pixels that are 55 µm wide (or high). This results in 516×516 pixel frames with a discernible, six-pixel wide cross, in which the pixels have a gain that is about three times higher than that of the other pixels outside the cross. This was corrected by multiplying the counts of the unaffected pixels by a factor of three. As the Timepix detector is operated in electron-counting mode, the converted images therefore model a detector with Poisson response and a multiplicative gain of 3.0. This was recorded in the *dxtbx* format class so that the correct gain value would be used automatically, for example in the calculation of error estimates for integrated intensities.

4.2.2 Spot finding

The spot-finding algorithm used in *DIALS* is rather sensitive to the detector gain. No automatic evaluation of the gain is performed prior to spot finding, although a value can be determined using the program *dials.estimate_gain*. This uses the mean and variance of pixels within a region of interest (Leslie, 2006), and may significantly underestimate the true gain for detectors that have a non-negligible point spread or corrections applied that reapportion signal between neighbouring pixels (Waterman & Evans, 2010). If the correct gain is known it is usual for this to be set by the format class used to import images. Otherwise, a suitable value should be passed to *dials.find_spots* for use by the spot-finding algorithm. In difficult cases it may be necessary to optimize the gain and other spot-finding parameters, the effects of which can be explored interactively using *dials.image_viewer*. For the seven example data sets discussed here we typically found that it was necessary to increase the sensitivity of spot finding and then reduce additional noise by using a global threshold. Appropriate spot-finding settings were determined manually for each data set separately. The effect of these settings for data set 1 is shown in Figure 4.1.

4.2.3 Experiment geometry

The most substantial difference between the processing of rotation data from electron diffraction compared with X-ray diffraction lies in the modelling of the diffraction geometry. The short wavelength of an electron beam (0.02508 Å for 200 keV electrons compared with 1.0332 Å for 12 keV X-rays) implies a correspondingly large Ewald sphere with a small 2θ scattering angle even for the highest resolution reflections.

The low diffraction angles imply that a large effective sample-to-detector distance is needed to magnify the diffraction pattern and achieve sufficient spatial separation between peaks. Large detectors are advantageous for crystallography because they allow the sample-to-detector distance to be increased, which both reduces diffuse background and improves the spatial separation of the peaks (Stanton, 1993). However, the detector distance is limited in a transmission electron microscope (TEM) by the largest possible magnification and the relatively small size of the detectors. Whilst the true camera position underneath the TEM column is always at a fixed distance, the effective detector distance is set by the projector lens system and does not correspond directly to a quantity that can be measured mechanically. Similar to an X-ray beamline, the sample-to-detector distance in a TEM is easily calibrated with reliable test crystals. However, inaccuracy in the recorded effective distance may

be difficult to correct by the usual process of diffraction geometry refinement owing to the high correlation between unit-cell parameters and the detector distance when 2θmax is small, in which case the Ewald sphere is almost invariant with respect to linear scale (see §4.2.6). In addition, imperfections in the lens system may introduce distortions in the recorded diffraction images. By disregarding such defects, which are discussed further in §4.2.4, the processing software ignores the lens system and models the experiment with an effective detector distance.

The relatively extreme geometry of electron diffraction is unfamiliar to many X-ray crystallographers. It is instructive to compare graphical schematics, such as Figure 2.6 for the real-space geometry of the instruments, and Figure 4.2 for a comparison of the Ewald construction in reciprocal space for the two cases.

Another potential source of inaccuracy in the initial model for the diffraction geometry arises because of the relatively poor characteristics of the sample-positioning stage of electron microscopes compared with X-ray goniometers for the purpose of rotation-method experiments. Improved setups are possible, but are not widely available (Yonekura *et al.*, 2015; Shi *et al.*, 2016). The rotation range per image is generally assumed to be constant and accurate. Instruments used for electron diffraction should therefore be well calibrated (Gemmi *et al.*, 2015). Small, smooth deviations from the expected rotation angle can then be modelled in *DIALS* as part of the scan-varying refinement of the crystal.

Generally, there may be uncertainty regarding the orientation of the rotation axis, the direction of rotation and the rotation range per image. Procedures have been developed to identify rotation-axis orientation for electron diffraction studies (Dorset, 1976; Kolb *et al.*, 2009); however, there is no implementation of an automated algorithm for this in *DIALS*. Nevertheless, for macromolecular samples there are a relatively large number of spots found throughout a data set and these can be used to obtain a reasonable estimate of the rotation-axis orientation in the plane of the detector by inspecting the images. This axis forms a line through the beam centre along which reflections have the widest reflecting range, and few reflections are found. As long as the initial estimate is good enough for successful indexing, the remaining error may be corrected by the geometry refinement procedure described in §4.2.6. The direction of rotation around the axis is more difficult to determine. For an X-ray experiment the curvature of the Ewald sphere makes the incorrect choice obvious, for example using a visual tool such as *dials.reciprocal_lattice_viewer* (Winter *et al.*, 2018). By contrast, the flatness of the Ewald sphere in electron diffraction ensures that either choice of handedness of rotation will produce regular reciprocal-lattice positions, as shown in Figure 4.3. If indexing is successful, it is likely to work either way. For any case where there is ambiguity, the inverse direction should also be tested and the results compared. The correct solution will have a lower rmsd for the angular residual between the predicted and observed positions of the reflections.
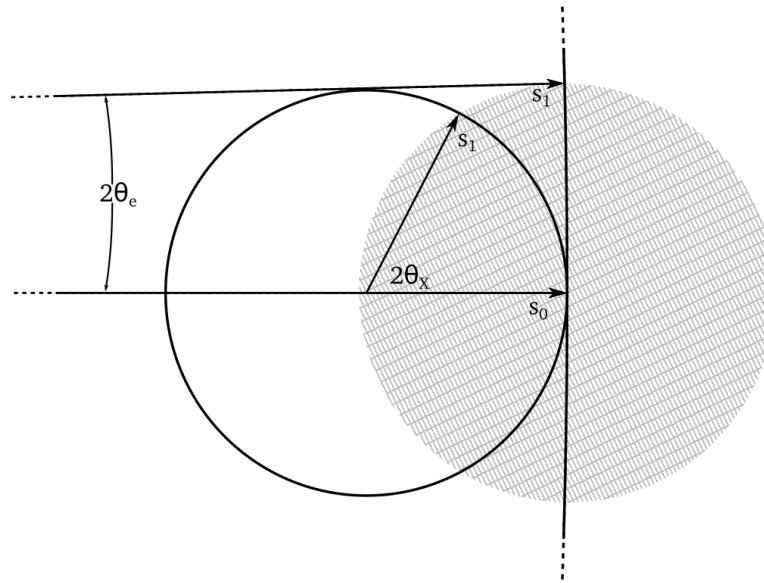
Figure 4.2 The Ewald constructions for the electron diffraction and X-ray cases are compared. The crosshatched circle represents a reciprocal lattice within a limiting sphere of 1Å resolution. The Ewald sphere for 12 keV X-rays with a wavelength of 1.0332Å is represented as a complete circle, with the scattering vector S_1 drawn at the 1Å limit, forming an angle of $2\theta_X = 62.2°$ from the incident beam direction along S_0. At this scale, the Ewald sphere for 200 keV electrons, with a wavelength of 0.02508Å cannot be shown as a complete circle as it has a radius over 40 times greater. The equivalent scattering vector S_1 for 1Å diffraction forms an angle of only $2\theta_e = 1.44°$ from the incident beam direction. It is worth noting that the reciprocal lattice is sampled along an almost planar surface, implying that data from a single image contain no information about the reciprocal lattice dimension in the direction along the incident beam.

4.2.4 Image distortion owing to lens effects

Image distortion is not unique to electron crystallography. In X-ray crystallography, geometrical distortions may be present owing to components of the detector system. A familiar example of these are spatial distortions introduced by the fibre-optic taper in a phosphor-taper CCD area detector (Stanton *et al.*, 1992). In this case, the distortion is a fixed property of the detector and it is usual for images to be corrected by manufacturer-supplied routines prior to analysis. Nevertheless, data-processing packages such as *XDS* have facilities for applying a distortion correction in the form of look-up tables. Even with the advent of hybrid pixel-array detectors, which have a direct coupling between the detector surface and the counting electronics, geometrical distortion may be used to correct for subpixel shifts and misorientations between the modules of the detector array. In electron crystallography, geometrical distortions of the detector are no less relevant, while there is the additional factor of the possibility of distortion of the diffraction pattern itself owing to effects of the electron optical system. Possible distortions include anisotropic magnification, where the diffraction pattern is elongated in one direction, transforming a circular powder pattern into an ellipse (Capitani *et al.*, 2006; Clabbers *et al.*, 2017). Care must be taken to investigate the presence of these effects in electron diffraction data sets and, as they are not mechanical properties of the instrument, it is necessary to recalibrate when instrument settings are changed.
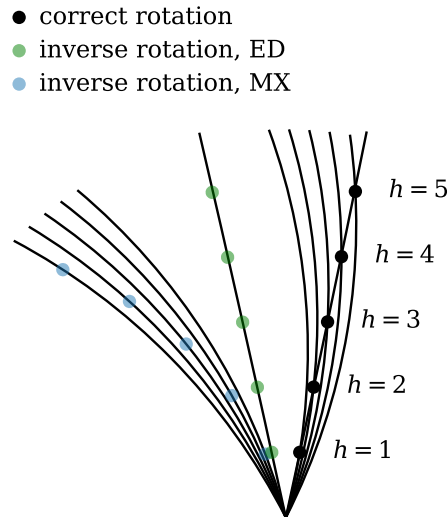
Figure 4.3 Five reciprocal lattice points are shown (in black and labelled) along the a^* axis for a crystal with unit-cell dimension a = 10Å. Arcs representing the surface of the Ewald sphere with a typical X-ray wavelength of 1.0332Å intersect these points at rotation angles between 15.0° for *h*=1 and 27.0° for *h*=5, where rotations are assumed to be clockwise from vertical in the plane of the figure. If the modelled rotation axis is inverted then φ centroids of observed spots would be mapped onto Ewald spheres rotated between -15.0° and -27.0°, resulting in a distinct curvature to the reconstructed reciprocal lattice (points shown in blue). In the case of electron diffraction at a wavelength of 0.02508Å the spots are observed almost simultaneously at rotation angles between -12.1° and -12.4°. For clarity a single Ewald arc is shown for *h*=3. If the assumed axis is inverted then φ centroids between -12.1° and -12.4° still result in almost a straight line (points shown in green). It is therefore difficult to determine the correct direction of rotation from the appearance of the reconstructed reciprocal lattice alone.

Despite the fact that the distortion occurs in the direction of the scattered rays rather than as a property of the detector, it is reasonable to correct images using the same means as for other sources of distortion. Within *DIALS*, we implemented a similar mode for distortion correction as used in *XDS*. A pair of distortion maps encode the pixel offset across the detector for both the fast and slow directions. These maps are equal in size to the pixel array of the detector (for a multiple-panel detector the correction files encode a list of separate maps for each panel). No interpolation is performed during the application of the distortion maps. In principle, sharp changes to correct for shear defects would be possible; however, for the case of lens aberration the offset varies slowly over the face of the detector so that neighbouring values in the look-up table are similar. The distortion maps are applied during the conversion between detector pixel coordinates and virtual detector millimetre coordinates. During the transformation from millimetre coordinates to pixel coordinates, the uncorrected pixel coordinate is first calculated and the correction is applied to obtain the distortion-corrected pixel coordinate. Likewise, during the transformation from pixel coordinates to millimetre coordinates the reverse correction is first applied and the millimetre coordinate is calculated from the reverse corrected pixel coordinate.

Data sets 2-7 in our examples all showed a significant elliptical distortion, which was constant across these data sets. The parameters of this distortion were determined using a well known diffraction standard, as described previously

(Clabbers *et al.*, 2017). The use of an independent standard for calibration is good practice that would become essential in the case where the sample of interest has an unknown unit cell. We extended the program *dials.generate_distortion_maps* to produce *X* and *Y* distortion maps for the four-panel detector model based on the known parameters. These maps were registered for each relevant data set during the *dials.import* step, after which they were loaded and applied automatically whenever required by *DIALS* programs.

4.2.5   Indexing

Provided that a sufficient number of strong spots have been collected (§4.2.2), indexing of electron diffraction works with similar reliability as for X-ray diffraction data. Difficulties mostly arise from systematic errors such as the stability of the rotation axis and the often large variation in the oscillation width $\Delta\varphi$. The default method for determining the unit-cell basis vectors in the *dials.index* program is based on the three-dimensional FFT of found spot positions, which works well even when the scan consists of a relatively narrow wedge, as is typical for an electron diffraction data set. The program *dials.index* performs refinement of the initial solution; therefore the guidance listed in §4.2.6 for refinement of ED geometry is also relevant and it is possible to pass options for the *dials.refine* program into *dials.index* where required.

Unless a model space group was chosen by the user, the indexing results are presented with triclinic symmetry. The compatibility of other choices of Bravais lattice with the triclinic solution can be tested using the program *dials.refine_bravais_lattice* (Winter *et al.*, 2018; Sauter *et al.*, 2006). There is no difference in usage compared with X-ray data; however, for electron diffraction the results might be more difficult to interpret. In particular, the metric fit reported for each trial solution (Le Page, 1982) may be large (for example greater than 1°) even for a correct solution, whereas much smaller values are expected for good-quality X-ray data. The correlation coefficients between intensities related by symmetry operations of the lattice are affected by low multiplicity of the data and by factors that cause deviation from expected intensities such as dynamic diffraction. As a result, these are not as useful in deciding on the correct lattice as they are in X-ray experiments. The key criterion is then the rmsd between predictions and observations. A pool of solutions with rmsd's similar to the original triclinic solution are good candidates. Any solution resulting in a significant increase in rmsd is suggestive of an over-constrained lattice and should be discarded.

For six of the seven example data sets, indexing followed by the selection of an orthorhombic lattice was successful with default options apart from fixing some detector parameters, as described in §4.2.6. For data set 6 we additionally fixed the beam orientation parameters and provided the expected unit cell and a restraint to this target cell during refinement. This data set shows relatively poor diffraction. Rather few spots were successfully indexed and rmsd's between the predicted and observed rotation angles remained high after refinement (see Table 4.1). The action of both constraints and restraints help to stabilize and guide refinement in such difficult cases.

4.2.6 Global refinement of the unit cell and instrument parameters

Following indexing, the model for the diffraction experiment geometry is further refined. This consists of the joint refinement of global parameters, including the beam direction, the unit-cell parameters, the cell orientation and the detector position and orientation. The choice of refined parameters is left to the user, with the default set being appropriate for typical X-ray data sets (for details, see Waterman *et al.* 2016). The flexible geometry description and refinement procedures of software such as *XDS* or *DIALS* is of great importance in electron diffraction studies, where the initial geometry may be quite poor. The radius of convergence of these procedures is high enough to correct large errors, as long as the indexing of spots is correct. In common with X-ray data processing with *DIALS*, it is usual to first refine a 'static' model for the whole data set, in which parameters such as the crystal unit cell and orientation angles are not allowed to vary across the scan. The global refinement of a data set improves the stability of the refinement procedure. However, the geometry of an electron diffraction experiment raises particular issues that should be taken into account, especially if the data quality is limited by low-resolution diffraction for some or all of the scan or poor-quality spot centroids, or if the scan is an especially narrow wedge. In this section, we offer some practical advice for *DIALS* refinement tasks with challenging electron diffraction data.

It is more difficult to refine unit-cell parameters using electron diffraction data than using X-ray data. This is mainly caused by the weaker signal and the much smaller diffraction angles $2\theta_{max}$ in electron diffraction[28]. A weak diffraction signal implies fewer diffraction spots and lower accuracies in determining their centroids, compromising the accuracy of the refinement. The small diffraction angle implies a low Ewald sphere curvature and a very high correlation between detector distance and a uniform unit-cell scale factor. In the limiting case the relative accuracy of the unit cell scales linearly with the relative accuracy of the detector-distance calibration. In cases where unit-cell imprecision does not prevent structure solution, the parameters can be adjusted during model refinement (Gruene *et al.*, 2018). Automatic options for performing this have recently been implemented in *REFMAC*5 (see §4.2.10).

The high level of correlation between parameters in diffraction geometry refinement problems has long been recognized. The method of eigenvalue filtering was proposed to allow refinement to proceed in such cases (Reeke, 1984; Bricogne, 1986*b*) by automatically selecting only those parameters, or linear combinations of parameters, that have the greatest effect at each step of refinement. This was deemed to be necessary at the time to refine crystal parameters using data from a single oscillation film. Within *DIALS*, all available data are used for a global refinement. This reduces correlations and provides a better determination for parameters when the scan range is wide; thus, the default behaviour is to refine the beam, crystal and detector parameters simultaneously, which works well for X-ray data. We have seen that when limited to a narrow wedge of data recorded with the geometry of the electron diffraction experiment, high correlations are again problematic. *DIALS* refinement does not use the eigenvalue-filtering method, but by default uses a Levenberg–Marquardt algorithm, which provides an alternative approach for dealing

---

[28] As the diffraction angle $\theta \rightarrow 0$, $\arctan(\theta) \simeq \arcsin(\theta) \simeq \theta$. Substituting this into Bragg's law (Equation 2) and the geometry of the diffraction (see Figure 2.6 where D is the detector distance and r is the distance between the central beam position and a Bragg spot with resolution d_{hkl}) results in a linear correlation where $D/r = d/\lambda$.

with near-singular least-squares problems. In practice, we find that this algorithm is robust even in the presence of very high parameter correlations. However, experience shows that the most challenging problems with electron diffraction geometry may need many steps before convergence is achieved, where this is defined as a negligible further reduction in rmsd's For this reason, from *DIALS* v.1.8 the maximum number of iterations before refinement terminates has been raised to 100 from 20 for the Levenberg–Marquardt algorithm (the limit can always be adjusted by the user *via* the *max_iterations* parameter).

If a good estimate for the unit cell is available as prior knowledge, this can be incorporated into refinement by the use of restraints, tying the unit-cell model to an external target. Unit-cell restraints are currently available for static refinement of unit-cell models but not scan-varying refinement, as they were originally developed for XFEL serial crystallography where scan-varying refinement is irrelevant. The unit-cell parameterization in *DIALS* is expressed with reciprocal metrical matrix elements as parameters (Waterman *et al.*, 2016). However, for ease of use, restraints are specified in terms of the real-space cell, as shown by the example given in Appendix B1. Each crystal included in refinement can add up to six restraint terms (for the triclinic case). Irrelevant restraints for unit-cell parameters that are already constrained by lattice symmetry are automatically excluded. Every restraint term adds a pseudo-observation to refinement. Taking the unit-cell parameter *a* as an example, the pseudo-observation term R_a consists of the squared residual between this parameter and its target value *a* with a weighting factor. In common with the real observations, the first derivatives of the pseudo-observations with respect to the refinable parameters (here arbitrarily denoted *p*) are also required for refinement by nonlinear least-squares methods:

$$R_a = \frac{(a - a_t)^2}{a_a^2} \tag{20}$$

$$\frac{\partial R_a}{\partial} = 2\frac{\partial a}{\partial p}\frac{(a - a_t)^2}{a_a^2} \tag{21}$$

In principle, statistical weighting could be achieved by setting the weights equal to the inverse variance of the target unit-cell parameter values. However, numerical uncertainties from refinement are known to be underestimated (Dauter & Wlodawer, 2015). For X-ray diffraction refinement we usually try values between $\sigma \approx$ 0.001 for qualitatively 'strong' restraints and $\sigma \approx 0.1$ for 'weak' restraints, monitoring the effect on the refined rmsd's. In the electron diffraction case setting even very weak restraints to a target cell can avoid issues with the unit cell and detector distance drifting when these are refined simultaneously. Nevertheless, the high correlation between these parameters means that the problem of distinguishing between cell volume and detector distance remains salient, and indeed the unit cell can be driven towards a target cell of incorrect volume with a minimal increase in refined rmsd's if the detector distance is also refined. It is generally advisable to accurately calibrate the effective detector distance prior to ED data collection and then to fix this during data processing. Other parameters that it may be prudent to fix include the detector $\tau2$ and $\tau3$ values, which describe rotations around axes in the

plane of the detector, similar to *MOSFLM*'s TILT and TWIST. Joint refinement of these parameters along with the beam direction and detector translations within the detector plane can be unstable.

For six of the example data sets, fixing the detector distance, $\tau 2$ and $\tau 3$ gave acceptable results for joint refinement of the beam, crystal and detector in-plane translation and rotation parameters. For the more difficult case, data set 6, no additional parameters were fixed, but a restraint to the target cell as given in Appendix B1 was used. Only 139 reflections were available for refinement in this case after outlier rejection. The use of the restraint ensured that the refined cell remained reasonable. In particular, without the restraint the long axis dimension drifted to above 108 Å. Including the restraint increased the rmsd's in X and Y by less than 0.07 and 0.14 pixels, respectively, and had a negligible effect on the rmsd in the rotation angle, demonstrating a case in which this feature can be used to guide refinement without resulting in a model that stands in dispute with the centroid data.

4.2.7   Scan-varying refinement of crystal and beam parameters

In a typical use of *DIALS*, the global static model for a data set is used as a starting point for scan-varying refinement. As originally implemented (Waterman *et al.*, 2016), this was intended to capture changes to the crystal unit-cell and orientation parameters during data collection. These parameters were allowed to vary in a smooth manner by evenly distributing sample points across the scan and interpolating values at any one position using a Gaussian smoother. The beam and detector parameters could be jointly refined to global, static values alongside the scan-varying crystal.

The analysis of electron diffraction images raises a new issue in that instrument stability during the course of data collection cannot be simply assumed, as it is for MX data. In some cases, there is significant drift of the beam centre during data collection caused by instability of the alignment or charging effects. Previous methods to handle this involve procedures to identify the shift for each image and write out corrected images in which the beam centre remains constant, effectively describing the drift in terms of shifts of the detector (Wan *et al.*, 2013; Nederlof, van Genderen *et al.*, 2013; Hattne *et al.*, 2015). The procedures differ in the way that the beam centre is determined for each image. In the simplest case, the high scattering cross-section for electrons allows, for some instrumentation, the direct beam to be recorded simultaneously with diffraction spots, avoiding the need for a beam stop. When images are not corrected, software such as *MOSFLM* or *XDS* can be set to independently refine the beam centre for each image or within small blocks of images. The focus on global refinement in *DIALS* means that an alternative approach was sought. Beam drift in electron diffraction experiments, at least those collected by a continuous rotation protocol, appears to occur gradually. Therefore, it seems reasonable to assume that a smoothly varying model for the beam-direction vector would suffice to represent this effect. For small magnitudes of the total drift, the difference between correction by implicit detector shifts and modelling of a drifting beam will be negligible. For the purposes of ED data processing, we extended the scan-varying refinement methodology from crystal parameters to optionally also apply to the beam parameters; this is available from *DIALS* v.1.9 onwards.

The difficulties with refinement inherent to electron diffraction geometry are exacerbated during scan-varying refinement. Like static refinement, scan-varying refinement in *DIALS* is also global, in that data from the full rotation scan are used in a single optimization procedure. However, at any point in the scan the local values for the crystal unit cell, angular misset and potentially the beam direction parameters are dominated by the data close to that point. Spot centroids at rotation angles further from that point have a diminishing effect on the local model, controlled by a Gaussian smoother. While this allows the model to express genuine smooth changes, it reduces the stability of the refinement procedure. This has been seen in cases where a static crystal model allows global refinement of both the detector and crystal parameters to reasonable values, but scan-varying refinement of the crystal results in a drift of the average unit-cell volume and detector distance. Despite these observations, scan-varying refinement is still preferable to static refinement of the beam, crystal and detector models within local narrow wedges, which suffers even more from high parameter correlations. To stabilize a problematic scan-varying refinement task we must either restrain or constrain (fix) some parameters of the model. There is no automatic determination of a suitable parameterization for refinement in *dials.refine*. Diagnostics (see §4.2.8) may help to understand the details of a particular case and guide choices; however, ultimately the user must inspect the resulting models for reasonable geometry as well as the final rmsd values.

We performed scan-varying refinement prior to integration for the seven example data sets. A variety of protocols was tested, and the best was chosen for each data set according to merging statistics after scaling of that data set in isolation by *AIMLESS* (Evans & Murshudov, 2013). In each case, we fixed all detector parameters so that the detector maintained the geometry from the static refinement step. For data set 1 a significant drift of the beam centre was observed. We enabled scan-varying refinement of both beam direction angles μ_1 and μ_2 in the nomenclature of Waterman *et al.* 2016 Remarkably, the simplest model consisting of two refineable sub-parameters for each angle resulted in the best merged data set, rather than models with more subparameters that are smoothed less in order to track higher frequency changes to the beam drift. Scan-varying refinement of the beam was tested for each of the other data sets. For two cases, data set 4 and data set 5, merging statistics favoured static refinement of the beam direction. In the other cases, the simple two-subparameter model for each beam angle was used. For each data set, the three crystal orientation 'misset' angles were refined in a scan-varying manner, using default smoother parameters. A scan-varying unit cell was refined for each case, except for data sets 3, 4 and 6, for which refining a global, static cell stabilized refinement and produced better merging statistics. Further details of the diffraction geometry modelling for each data set are given in Table 4.1.

4.2.8 Diagnostics for problematic diffraction geometry refinement

§4.2.6 and §4.2.7 describe parameters that need to be adjusted in difficult cases. To date, even electron diffraction data sets from standard proteins may be found to be difficult (Clabbers *et al.*, 2017; Hattne *et al.*, 2015). At this early development stage, diagnostic tools are important for fine-tuning parameters. The program *dials.refine* provides some facilities for investigating the main issue that we have identified, namely the high level of correlation between the effects of different parameters on the model. This information is contained within the Jacobian matrix built up as part of

each step taken by the nonlinear least-squares optimization algorithm. In this section, we present two diagnostics based on analysis of the Jacobian matrix and pick out the salient differences that occur simply as a feature of the refinement of geometry at the very short wavelength typical for electron diffraction.

Each step of the nonlinear least-squares problem is expressed as a linearized subproblem of the form:

$$\boldsymbol{J}\Delta\boldsymbol{p} = \Delta\boldsymbol{r} \qquad (22)$$

By convention, the three-dimensional observations are split so that $\Delta\mathbf{r}$, the vector of residuals, contains first the $(X - X_o)$ components, followed by the $(Y - Y_o)$ components and finally the $(\varphi - \varphi_o)$ values. $\mathbf{J}$, the Jacobian matrix of first partial derivatives of the residuals with respect to each parameter of the problem, is thus similarly formed in blocks, with the upper third of the matrix corresponding to $\partial X/\partial p$ values, the second to $\partial Y/\partial p$ and the lower third to $\partial\varphi/\partial p$. The vector $\Delta\mathbf{p}$ is the parameter shift vector to be determined for the step.

The first diagnostic consists of graphical 'corrgrams', which are a way of rapidly assessing correlations between the parameters of refinement in a visual manner. The data represented by a corrgram consist of the matrix of pairwise correlation values calculated between columns of the Jacobian. Since its introduction, described in (Waterman *et al.*, 2016), this diagnostic has been improved. Rather than calculating a single corrgram using correlation between each full column of the Jacobian, the three-dimensional nature of the centroid data is respected and three corrgrams are produced: one for each of the blocks of the Jacobian, corresponding to the dimensions *X*, *Y* and φ. These separate figures are more appropriate for assessing the levels of correlations between parameters implied by the data, whereas a single corrgram can obscure these features. This is because the derivatives of calculated centroid positions with respect to some parameter $\partial X/\partial p$, $\partial Y/\partial p$ and $\partial\varphi/\partial p$ come from different distributions and thus should not be combined in a meaningful calculation of correlation.

While the corrgram diagnostic qualitatively identifies which parameters are the least distinguishable from each other, it might still not give a clear indication of which refinement cases will actually cause problems. Certain correlations are high anyway even in unproblematic cases. For this reason we also investigated an alternative, quantitative, diagnostic with a simpler interpretation, namely the condition number of the Jacobian matrix $\mathbf{J}$. This provides a measure of how well posed the subproblem given by (22) is, but does not pick out which parameters are culpable. A condition number $\kappa(\mathbf{J})$ of infinity means that $\mathbf{J}$ is singular, while a finite value of $\kappa(\mathbf{J})$ gives a bound on the accuracy of the solution to Equation 22.

The Jacobian used to calculate both the corrgram and the condition-number diagnostics does not include any additional blocks related to pseudo-observations that may be used as restraints in refinement. For this reason, it should be noted that the diagnostics give information about the underlying degeneracy of parameters determined only by the geometry of the problem, not including the effects of modifications to the problem that may have been introduced to improve the robustness of the procedure. Similarly, the diagnostics inform us directly about

properties of the normal equations of the Gauss–Newton problem implied by Equation 22 rather than the modified normal equations of the Levenberg–Marquardt algorithm that is typically in fact used to find the solution. This ensures that these diagnostics can be used to warn us of problems with the setup of the diffraction geometry refinement itself, without conflation with factors relating to implementation details of the algorithm used to perform the optimization.

To investigate the difficulties faced with refinement problems that are solely a result of the electron diffraction geometry, we elected to perform refinement against simulated data. In this way, we could compare two refinement procedures using an identical crystal model, beam direction and rotation axis, while altering the wavelength and detector distance to match typical values for electron diffraction in one case and X-ray diffraction in the other. Details of how the simulated data were constructed are presented in the Supporting Information. Refinement was performed for the same sets of reflections with both versions of the geometry, using default settings in *dials.refine*. In each case 13 parameters were refined in total: six to describe the detector position and orientation, one beam orientation angle, three crystal orientation angles and three reciprocal metrical matrix elements for the unit cell. For the final step of refinement prior to termination at rmsd convergence, corrgrams were produced and the condition number calculated for comparisons.

The complete two sets of three corrgrams (see Appendix B2) The pattern of high correlations between parameters that affect the predicted reflection positions (X, Y) on the detector plane are similar in the cases of electron and X-ray diffraction geometries. However, in general, the absolute values of correlations are higher for the electron diffraction geometry. The most striking difference between the two cases is shown on the corrgram for the parameters that affect the predicted rotation angle φ_c. None of the detector parameters affect φ_c, so only the beam and crystal parameters are of interest. The relevant subset of the corrgram is reproduced in Figure 4.4. This figure shows that absolute correlations between certain parameters are high in either case, but that the electron diffraction geometry shows increased absolute correlations between φ_3, the crystal orientation around the Z axis, and other parameters. In general, absolute correlations are smallest between the parameter $g^{\cdot}_{11}$, here corresponding to the short axis of the cell, and other parameters for either version of the geometry. For this data set, the short cell axis was aligned closest to the rotation axis. As a result, this dimension is relatively well determined by centroid data from images throughout the data set. However, even for this parameter the electron diffraction geometry produces larger absolute correlations with other parameters, except one, $g^{\cdot}_{33}$, which parameterizes the long axis of the cell. Detailed interpretation of these plots is difficult and requires complete knowledge of the definitions of each of the parameters, including the directions about which they are defined and the order in which they act to compose the final model. Broadly, however, we can immediately see a pattern of greater magnitude correlations for the electron diffraction case and would expect a correspondingly more challenging refinement problem.
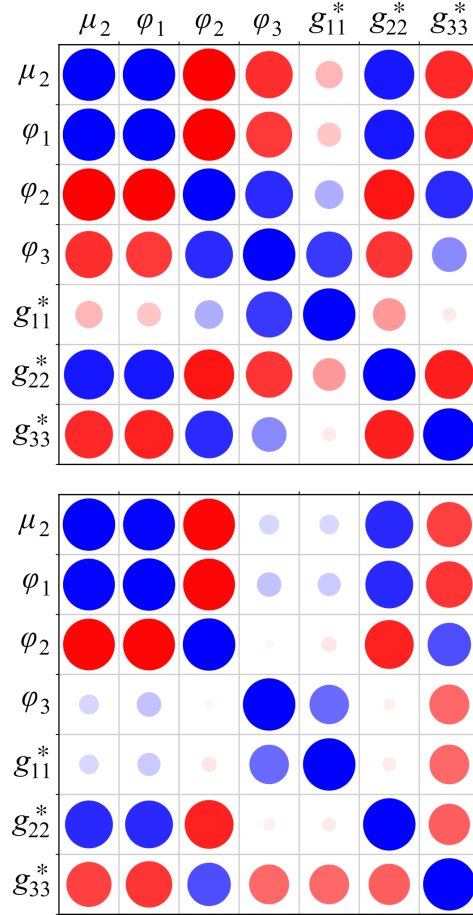
Figure 4.4 Geometry refinement against simulated data was performed assuming either typical electron diffraction geometry or X-ray diffraction geometry, as described in the text (see also Appendix B2). Corrgrams were produced for the final step of refinement to provide immediate visual feedback regarding correlations between the effects of refined parameters on the model. The colours and areas of the circles are related tot the values of the correlation coefficient, with large blue circles indicating strong correlation and large red circles indicating strong anticorrelation. This plot shows the correlation between the effects of different parameters on the angular residuals ($\varphi - \varphi_0$), with the refined detector parameters excluded from the plots as they have no effect on the φ residuals. Here parameter label μ_2 defines the beam orientation angle, φ_i define the crystal orientation angles along x, y, and z, and $g_{i,j}^*$ describes the crystal unit cell by reciprocal vectors a^*, b^* and c^* (Waterman *et al.*, 2016). The upper panel shows the corrgram for the electron diffraction geometry and the lower panel shows the equivalent corrgram for the X-ray diffraction geometry.

The second diagnostic provides a measure to quantify this effect. The condition number at the final step of refinement for the electron diffraction geometry $\kappa(\mathbf{J}_{ED}) \approx 8\times10^5$, while for the X-ray diffraction geometry $\kappa(\mathbf{J}_{MX}) \approx 2\times10^3$. This clearly indicates that the electron diffraction geometry presents a considerably less well posed problem for refinement. With the simulated data we jointly refined 13 parameters simultaneously; however, for the processing of the seven real example data sets we fixed the detector distance, τ_2 and τ_3 parameters to stabilize refinement and avoid the cell volume drifting away from reasonable values. The condition number quantifies this stabilization. When the same parameters are fixed during refinement of the electron diffraction geometry against simulated data this reduces to $\kappa(\mathbf{J}_{ED}) \approx 8\times10^3$, a two order-of-magnitude improvement of the problem condition.

The diagnostics presented here can help to design protocols for successful diffraction geometry modelling in difficult cases. However, there is much variation between data sets, and not yet enough experience to allow generalization or automated selection of an optimal protocol. For the example data sets, the best procedure we found was to rely on careful, independent calibration of distance and fix that during refinement. Other parameters were additionally fixed for individual cases. Small errors in the cell can be tolerated for the purposes of integration and improved later at the stage of model refinement.

4.2.9   Integration and data reduction

Following global modelling of an experiment using *dials.refine*, data are integrated with *dials.integrate*. No special options are required for integrating electron diffraction data. For the examples we specified options only to run multiple processes in parallel and to specify resolution limits within which all spots will be predicted for integration. The integrated data sets were then exported to MTZ format using *dials.export* and combined by *POINTLESS* (Evans, 2006) for scaling and merging together with *AIMLESS* (Evans & Murshudov, 2013). This procedure included a reindexing step to convert the orthorhombic cell used for integration, with $a < b < c$, to the conventional space group $P2_12_12$. Merging statistics are summarized in Table 4.2, while Appendix B3 summarizes statistics for data sets scaled individually.

4.2.10   Structure solution and refinement

The structure was determined as described in Clabbers *et al.* 2017 with the exception of using the intensities for molecular replacement in *PHASER* (McCoy *et al.*, 2007; Read & McCoy, 2016), and with the additional step of refining the lattice in *REFMAC*5 (Murshudov *et al.*, 2011). Lattice refinement allows the unit cell to be refined by a single scaling factor independent of the sample-to-detector distance, thus removing the ambiguity between detector distance and lattice parameters. The newly found unit cell was then used for subsequent structure refinement and validation (Murshudov *et al.*, 2011; Joosten *et al.*, 2014; Luebben & Gruene, 2015).

4.3   Discussion and conclusions

Electron diffraction from three-dimensional crystals has recently been used to solve the structures of macromolecules such as proteins. Previous authors have shown that where data are collected using the rotation method, as is standard in X-ray crystallography, data-processing software such as *MOSFLM* and *XDS* can be employed to successfully integrate the Bragg peaks. Here, we show that the *DIALS* package, with appropriate adaptations, is also a viable alternative, even for difficult data sets with problematic features such as distortions caused by microscope lens systems and drift of the direct beam position. A set of seven example data sets was successfully processed using *DIALS*, and the specific decisions required at each step are described in detail. The quality of data integrated with *DIALS* is very similar to what could be achieved with *XDS* (Table 3.1) (Clabbers *et al.*, 2017).

A major focus of the *DIALS* software is the global modelling of an experiment. The experimental geometry is optimized using all available data. Where components such as the crystal or the beam are expected to change during the course of the experiment, these changes are described using smoothly varying parameterizations, avoiding discontinuities in the model and stabilizing the refinement procedure. Other aspects of interest in global experiment modelling include unit-cell restraints and refinement diagnostics, which enable an exploration of the effects of different parameterizations on refinement stability. A demonstration using simulated data shows that problematic refinement is caused to a significant level simply by the short wavelength and large effective detector distance of electron diffraction experiments, even before additional factors such as instrumental instabilities are considered.

Besides errors occurring from instrumentation, there are additional issues specific to electron crystallography that will need to be addressed. The measured kinematic signal in ED is obscured by inelastic, dynamic and mixed multiple scattering events (Dorset, 1995; Zou *et al.*, 2011; Clabbers & Abrahams, 2018). Zero-loss energy filtering is an instrumental solution to this problem that can filter out most of the inelastically scattered electrons, reducing the diffuse background and sharpening Bragg peaks (Yonekura *et al.*, 2002). This should improve the accuracy of the intensity estimations from the recorded three-dimensional spot profiles. However, it is not possible to discriminate between kinematic and dynamic scattering energetically. On average, dynamic scattering increases the intensity of weak spots, which become stronger, and the stronger spots become weaker (Weirich *et al.*, 2000). This directly affects the measured intensities, which form the basis for any further structure determination. These adverse effects are currently not taken into account during data integration.

Electron diffraction of macromolecular crystals is still developing and is confronting crystallographers with new and sometimes unexpected problems. The extensive diagnostics offered by *DIALS*, in terms of corrgrams and its user-friendly, interactive tools for visual inspection of data and parameters, should help in identifying and solving the new challenges specific to optimally integrating electron diffraction data. The toolkit design philosophy of the software, including an extensible image-format reading system and permissive open-source licensing, lowers the barrier to entry for use and future development by scientists interested in this technique.

Table 4.1 Details relevant to the modelling of diffraction geometry are collated here for the seven example datasets.

	1	2	3	4	5	6	7
Detector distance [mm]	1890	2055	2055	2055	2055	2055	2055
Distortion correction	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of images	503	263	587	419	422	421	421
Image width [°]	0.076	0.1615	0.0344	0.0481	0.0481	0.0481	0.0481
No. of indexed spots	1624	1239	218	598	634	174	211
Condition no. $\kappa(\mathbf{J})$[29]	1.1×10^4	9.2×10^3	2.0×10^4	2.7×10^4	2.8×10^4	1.1×10^4	2.3×10^4
Static cell[30] [Å]							
a	31.967(7)	31.217(4)	31.56(5)	32.36(2)	31.841(11)	31.70(15)	31.63(2)
b	69.41(3)	68.59(2)	65.0(2)	67.25(6)	65.81(3)	65.6(4)	69.08(5)
c	104.62(3)	104.88(2)	106.4(3)	105.71(7)	103.2(3)	106.7(6)	104.07(4)
Average varying cell [Å]							
a	32.0	32.2	31.5	-	31.8	-	31.7
b	68.3	68.5	67.1	-	64.7	-	68.8
c	105.1	104.9	104.2	-	103.5	-	104.1
Beam centre							
Panel ID	2	0	0	0	0	0	0
X [pixels]	485.4-487.7	420.4-420.9	400.5-400.6	428.0	406.3	405.5-406.0	399.1-399.2
Y [pixels]	1.7-2.5	478.8-478.9	475.1-477.0	478.3	479.0	480.0-480.6	490.9-491.6
Final rmsd							
X [pixels]	0.93	0.42	0.94	0.67	0.53	0.65	0.51
Y [pixels]	0.83	0.59	0.95	0.85	0.66	0.58	0.63
Z [°]	0.06	0.08	0.09	0.11	0.04	0.48	0.04

---

[29] The condition number diagnostic is shown for the final step of static refinement.

[30] Errors as reported by *dials.refine* refer to precision estimated by the least-squares refinement procedure and are not indicative of the accuracy of the unit cell. The unit cell for data set 6 was refined with a restraint to an external target.

Table 4.2 Data processing and refinement statistics for seven merged datasets.

Data processing	
Space group	$P2_12_12$
Unit cell parameters	
a, b, c [Å]	104.57, 67.62, 31.87
α, β, γ [°]	90.00, 90.00, 90.00
Resolution [Å][31]	56.78-2.10 (2.16-2.10)
R_{merge}	0.313 (0.460)
R_{meas}	0.356 (0.574)
$R_{p.i.m.}$	0.160 (0.337)
No. of observations	31650 (1504)
Completeness [%]	59.2 (51.5)
Multiplicity	3.9 (2.7)
I/σI	3.0 (1.9)
$CC_{1/2}$ [%]	90.9 (62.6)
Refinement	
Space group	$P2_12_12$
Unit cell parameters[32]	
a, b, c [Å]	104.45, 67.54, 31.84
α, β, γ [°]	90.00, 90.00, 90.00
Resolution [Å]	56.72-2.10
No. of reflections	8143
R1 [%][33]	25.2
$R_{complete}$ [%][34]	29.2
<B> [Å^2]	18.3
RmsZ bond lengths	0.48
RmsZ bond angles	0.72
Ramachandran	
Favoured, allowed, outliers [%]	97.6, 2.0, 0.4

---

[31] Individual data sets 1-7 were truncated at $CC_{1/2} \geq 50\%$ and I/σI ≥ 1.0 (Diederichs & Karplus, 2013); the merged data set was limited to 2.1Å resolution based on the model refinement results. For processing statistics of individual dataset see Appendix B3.

[32] Unit cell dimensions after lattice parameter refinement in *REFMAC5* (Murshudov *et al.*, 2011).

[33] $R1 = \sum_{hkl} ||F_o(hkl)| - |F_c(hkl)|| / \sum_{hkl} |F_o(hkl)|$, where the sum includes all data

[34] $R_{complete}$ is a robust validation method, especially in cases where the data completeness is limited, making use of all reflections (Brunger, 1997), and R_{work} is thus equivalent to R1. $R_{complete}$ was calculated with a 0.2% test size as described in Luebben & Gruene (2015) and §3.2.4.

CHAPTER 5

Experimental and computational reduction of dynamical electron scattering allows visualising hydrogen atoms

Compared to X-rays, electron diffraction faces a crucial challenge: dynamical electron scattering compromises structure solution and its effects can only be modelled in specific cases. Dynamical scattering can be reduced experimentally by decreasing crystal size – but not without a penalty, as it also reduces the overall diffracted intensity. Here we show that nanometre-sized crystals from organic pharmaceuticals allow positional refinement of the hydrogen atoms, even whilst ignoring the effects of dynamical scattering during refinement. To boost the very weak diffraction data, we employed a highly sensitive hybrid pixel detector. We also introduce a general likelihood-based computational approach for further reducing the adverse effects of dynamic scattering, which significantly improved model accuracy – even for protein crystal data at substantially lower resolution.

Max T.B. Clabbers[1], Tim Gruene[2], Eric van Genderen[2] and Jan Pieter Abrahams[1,2,3]

[1]Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, Basel University, Mattenstrasse 26, CH-4058 Basel, Switzerland

[2]Paul Scherrer Institut (PSI), Department of Biology and Chemistry, CH-5232 Villigen PSI, Switzerland

[3]Leiden Institute of Biology, Sylviusweg 72, 2333 BE Leiden, The Netherlands

5.1   Introduction

Electron crystallography allows solving structures of beam-sensitive macromolecules and organic compounds using sub-micron sized three-dimensional crystals. The minute sample volumes used in electron diffraction are limiting the maximum radiation dose and the diffracted intensity (Henderson, 1995). Electrons interact with matter so strongly, that the majority of 300 keV electrons will have scattered either elastically or inelastically after having passed through only 50 nm of organic matter (Clabbers & Abrahams, 2018). Because this was considered to prevent successful analysis of three-dimensional protein crystals, macromolecular electron crystallography was limited until recently to two-dimensional crystals (Henderson & Unwin, 1975; Gonen *et al.*, 2005). Chemical electron crystallography using inorganic and organic crystals developed independently from protein crystallography (Cowley, 1953*a*,*b*; Vainshtein, 1964; Dorset, 1995; Kolb *et al.*, 2007; Mugnaioli *et al.*, 2009; Zhang *et al.*, 2010; Zou *et al.*, 2011). Recently, the rotation method that has been the standard for decades in X-ray protein crystallography (Arndt & Wonacott, 1977; Dauter, 1999), was adapted to electron crystallography for determining the structure of beam-sensitive macromolecules (Nederlof, van Genderen *et al.*, 2013; Nannenga, Shi, Leslie *et al.*, 2014; Nannenga, Shi, Hattne *et al.*, 2014; Yonekura *et al.*, 2015; Clabbers *et al.*, 2017), organics and inorganics (Gemmi *et al.*, 2015; van Genderen, Clabbers *et al.*, 2016; Gruene *et al.*, 2018).

Analogous to similar developments in X-ray crystallography and single-particle cryo-EM (Broennimann *et al.*, 2006; Kuehlbrandt, 2014; McMullan *et al.*, 2016), the introduction of new camera technologies, such as the advent of hybrid pixel detectors, were of vital importance for the development of electron diffraction of beam-sensitive crystals, as they offer high sensitivity, fast read-out and reduced background (van Genderen, Clabbers *et al.*, 2016; Tinti *et al.*, 2018). Electron diffraction data integration is feasible with existing X-ray crystallography packages, including profile fitting and scaling (Leslie, 1999; Kabsch, 2010*b*; Winter *et al.*, 2018), and require only minor adaptations (Clabbers *et al.*, 2018). The introduction of electron counting detectors required a Poisson error model and background estimation optimised for weak data with low-background, which were implemented for integrating very weak, few photon X-ray diffraction data as measured by hybrid pixel detectors (Kabsch, 2010*b*; Parkhurst *et al.*, 2016).

In general, crystallographic structure determination assumes single, kinematical scattering. Given the strong interaction of electrons with matter, multiple elastic or dynamical scattering is a major concern in electron crystallography, as it changes the observed structure factor amplitudes (Cowley & Moodie, 1957; Glaeser & Downing, 1993; Dorset, 1995; Weirich *et al.*, 2000). Dynamical scattering cannot be removed by zero-loss energy filtering, and coincides with at the kinematic Bragg scattering angles. On average, dynamical scattering increases the intensity of weaker reflections, whereas the strong reflections become less intense. Typically, dynamical scattering out of intense Bragg reflections into the weaker ones predominantly affects adjacent reflections when low resolution terms dominate (Weirich *et al.*, 2000). Thus, it depends on the diffraction geometry and can therefore be reduced by averaging corresponding intensities obtained from multiple crystal orientations. This explains why precessing or tilting the electron beam and/or rotating the crystal in a random

orientation, combined with averaging symmetry related reflections, reduces the effects of dynamical scattering (*e.g.* Vincent & Midgley, 1994).

Crystal size is an important factor affecting dynamical scattering as the probability of multiple scattering events increases with sample thickness (Subramanian *et al.*, 2015; Clabbers & Abrahams, 2018). A small crystal size, combined with data acquisition using the rotation (or precession) method reduces dynamical scattering. This approach allows full integration of the Bragg reflections resulting in a better estimate of the kinematic scattering intensity (Arndt & Wonacott, 1977; Dauter, 1999; Vincent & Midgley, 1994). It benefits from an electron counting quantum area detector with a high dynamic range to optimize the weak diffraction signal. Dynamical refinement of electron diffraction data is an effective, independent approach for dealing with dynamical diffraction (Jansen *et al.*, 1998; Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017). It requires knowledge of the atomic crystal structure, the crystal shape or thickness, and the three-dimensional orientation of the crystal for each unmerged reflection. Because it models the electron wave function travelling through the entire crystal, dynamical refinement is computationally demanding and has so far only been implemented for relatively straightforward cases with small unit cells.

A potential benefit of electron diffraction is the improved contrast of hydrogen atoms. For the lighter elements (up to about sulphur) the atomic scattering cross sections are approximately proportional to $\sqrt[3]{Z^4}$, instead of Z^2 as for X-ray diffraction (Egerton, 2011). The increased contrast of light atoms relative to heavier ones implies a larger contribution from the H-atoms to the overall signal (Cowley, 1953*a*,*b*; Vainshtein, 1964; Dorset, 1995; Clabbers & Abrahams, 2018). Localising hydrogen atoms can be notoriously difficult in X-ray crystallography, even with high-resolution data. However, recently hydrogen positions in organic and inorganic samples could be refined by reducing the effects of dynamical scattering by precession electron diffraction combined with dynamical refinement (Palatinus *et al.*, 2017).

Here, we present electron diffraction data of three-dimensional nanocrystals of two pharmaceutical organic compounds, recorded at cryogenic temperature using the rotation method and a Timepix hybrid pixel detector. The small size of the crystals reduced dynamical scattering, whilst the highly sensitive hybrid pixel detector boosted the weak diffraction signal. This combination allowed localising the positions of individual hydrogen atoms already at a very early stage of the refinement, and allowed unconstrained refinement of the hydrogen atoms, without any modelling of dynamical scattering. Furthermore, we introduced a likelihood-based approach for correcting dynamical scattering that down-weights overestimated reflection intensities as function of intensity and resolution in an approach similar to Wiener filtering. These corrections significantly improved the quality of the data and the model accuracy, and even work for protein crystal data with a substantially lower resolution.

## 5.2	Methods

### 5.2.1	Data acquisition

Diffraction data were acquired of two pharmaceutical organic compounds [35] (Appendix C1), kindly provided by Novartis, corresponding to CCDC entries IRELOH ($C_{16}O_5H_{18}$) (Dai *et al.*, 2010) and EPICZA ($C_{18}O_6N_2S_2H_{16}$) (Deffieux *et al.*, 1977) in the Cambridge Structural Database (CSD). Data were acquired by continuous rotation at cryogenic temperature with a ~2.0 μm diameter parallel beam using a FEI Polara TEM, operated at 200 kV and equipped with a 512×512 pixel Timepix hybrid pixel detector (van Genderen, Clabbers *et al.*, 2016; Gruene *et al.*, 2018). The effective detector distance was determined with an aluminium powder standard (Ted Pella).

### 5.2.2	Data processing

The raw data recorded from the detector were corrected for the pixel offsets between individual chips and converted to CBF format as described[36] (Gruene *et al.*, 2018). Data were integrated and scaled using *XDS* (Kabsch, 2010*b*). The Laue group was constrained to *mmm*. Multiple crystal datasets were merged, scaled, and converted to *SHELX* format within *XDS* (Kabsch, 2010*b*). Data were truncated where there was still significant correlation at approximately I/σI ≥ 1.0 and $CC_{1/2}$ ≥ 50% (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

### 5.2.3	Structure determination

5.2.3.1 Structure solution and model refinement

Structures of both organic compounds were solved using *SHELXT* with default settings (Sheldrick, 2015*b*). The models were refined using *SHELXL* (Sheldrick, 2015*a*), and built in *SHELXLE* (Hübschle *et al.*, 2011). Hydrogen atoms were placed automatically when possible using HFIX. For anisotropic refinement against incomplete data, mild restraints were applied in *SHELXL* to ensure similarity (RIGU) (Thorn *et al.*, 2012) and positivity (XNPD 0.001) of the atomic displacement parameters. Electron atomic scattering factors were fitted to the nine Cromer-Mann coefficients as used by *SHELXL* (Appendix C6) (Peng, 1999; Gruene *et al.*, 2018).

5.2.3.2 Lattice parameter refinement

In electron diffraction, the unit cell dimensions and sample to detector distance are highly correlated because of the short electron wavelength. Therefore, both cannot be refined reliably at the same time from the diffraction data alone. However, the lattice can also be refined independently from the detector distance by minimising the deviation from ideal model geometry (Gruene *et al.*, 2018). Geometrical restraints on bond lengths (DFIX) and bond angles (DANG) for IRELOH and EPICZA were generated using the *GRADE* server[37] (Appendix C5). The unit cell parameters were then refined against the idealised interatomic distances using *SHELXL* and the

---

[35] Experimental data are available online at https://doi.org/10.5281/zenodo.1407682

[36] https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Timepix2cbf

[37] http://grade.globalphasing.org/cgi-bin/grade/server.cgi

82

program *CellOpt*[38]. The geometrical restraints were only used for lattice refinement. After the geometrical restraints were removed from the instruction file, the model was refined in *SHELXL* against the new unit cell parameters.

5.2.3.3 Refining hydrogen positions

To illustrate the quality of the data and the stability of the hydrogen positions, we refined their positions by subsequently removing the constraints on the hydrogen atoms (AFIX), which are normally refined using the riding-model in *SHELX*. The hydrogen positions were then refined isotropically in *SHELXL* until convergence.

5.2.3.4 Validation

The models were validated using $R_{complete}$ (Luebben & Gruene, 2015), as a more robust alterative to R_{free} (Brunger, 1997). Since all reflections are used to calculate $R_{complete}$, the model can be refined against all data; hence R_{work} is equal to R1. The $R_{complete}$ was calculated with at test set size of 0.2%, running 10 refinement cycles for each run, using the *R_complete* graphical user interface[39].

To compare the quality of the model geometry, we calculated the rmsd of the electron diffraction structure bond lengths from the reference X-ray models (Deffieux *et al.*, 1977; Dai *et al.*, 2010). We assessed the stability of the refined hydrogen atoms by calculating the rmsd of the hydrogen bond lengths (X-H) compared to the idealised hydrogen bond length tables generated by *SHELXL* using the NEUT command (Sheldrick, 2015*a*). Using NEUT will list the inter-nuclei distances as neutron scattering occurs on the nuclei, and since idealised hydrogen bond lengths determined by X-ray diffraction are generally too short, it is therefore appropriate to use the more accurate X-H distances for neutron diffraction as described previously (Gruene, Hahn *et al.*, 2014).

5.2.4 Dynamical scattering corrections

5.2.4.1 Dynamical scattering increases intensity of weaker reflections

We previously observed an overestimation of the weaker reflection intensities in our electron diffraction data (van Genderen *et al.*, 2016; Clabbers *et al.*, 2017). We confirmed that this was also the case for the two organic samples presented here by plotting of the observed structure factor amplitudes $|F_o|$ against the structure factor amplitudes calculated from the model $|F_c|$ (Fig. 3, 4). On average, dynamical scattering increases the intensity of weaker reflections at the expense of the strong reflections, so we can expect the observed structure factor amplitudes $|F_o|$ to be overestimated for the weaker reflections (Dorset, 1995; Weirich *et al.*, 2000; Subramanian *et al.*, 2015; Clabbers & Abrahams, 2018). Assuming that we have a complex-valued dynamical scattering component F_e, which is uncorrelated to the kinematical structure factor $F(\boldsymbol{h})$, we can describe the expected value of $\langle|F_o|\rangle$ using a hyperbolic function defined as:

---

[38] https://github.com/JLuebben/CellOpt
[39] https://github.com/JLuebben/R_complete

$$\langle |F_o| \rangle = \sqrt{|F_c|^2 + |F_e|^2} \tag{23}$$

Using least squares fitting, we can derive the expected dynamical error term $\langle |F_e|^2 \rangle$ over all observations, where the plot $\langle |F_o| \rangle$ *vs.* $|F_c|$, intercepts the *y*-axis at $\langle |F_e| \rangle$ (Fig. 5.3, 5.4).

We determined $\langle |F_e|^2 \rangle$ as a function of resolution by dividing the data in 10 equally sized resolution bins (in terms of number of reflections), and within each bin determined the corresponding, resolution dependent dynamical scattering error $\langle |F_{e,bin}|^2 \rangle$ using least squares fitting. Using the curve fitting tool in *MATLAB* we established the relation between $\langle |F_{e,bin}|^2 \rangle$ and the resolution d (Fig. 5.3C, 5.4C), resulting in the continuous function $\langle |F_e(d)|^2 \rangle$.

In the absence of dynamical scattering (when $\langle |F_e(d)|^2 \rangle = 0$), the measured and calculated structure factors should be linearly correlated. Thus, as a function of resolution d, we can infer the scale $\epsilon_m(d)$ by which an expected observed intensity $\langle |F_o(\boldsymbol{h})| \rangle^2$ is increased due to dynamical scattering, as a function of (*i*) the resolution d, (*ii*) the strength of the dynamic effect $\langle |F_e(d)|^2 \rangle$, and (*iii*) the intensity $|F_c(\boldsymbol{h})|^2$ that would be observed in the absence of errors:

$$\epsilon_m(d) = \frac{|F_c(\boldsymbol{h})|^2 + \langle |F_e(d)|^2 \rangle}{|F_c(\boldsymbol{h})|^2} \tag{24}$$

Thus, $\epsilon_m(d)$ is essentially equivalent to a reciprocal generalised Wiener filter[40].

5.2.4.2 Dynamical scattering corrections for high-resolution data
Refinement in *SHELX* uses scaled, observed intensities, which are assumed to be kinematic, *i.e.* $I(hkl) \propto |F(hkl)|^2$. Instead of refining against the measured intensities, we refined against corrected intensities $|F_{o,corr}(\boldsymbol{h})|^2$, according to:

$$|F_{o,corr}(\boldsymbol{h})|^2 = \frac{|F_o(\boldsymbol{h})|^2}{\epsilon_m(d)} \tag{25}$$

These corrections were applied to each reflection, on both the intensities and their sigma values, using a single line of *Awk* commands that also writes the corrected *HKL* file with the standard formatting required for *SHELX*. The model was refined in *SHELXL* against the corrected data. The positional accuracies of the hydrogen atoms were assessed as before by refinement without any constraints on the hydrogen bonds.

---

[40] Equation (10) in W.K. Pratt (1972) "Generalized Wiener Filtering Computation Techniques" IEEE Trans. On Computers C21, 636-641

5.2.4.3 Dynamical scattering corrections for low-resolution protein diffraction data
In a similar fashion, the likelihood-based corrections for dynamical scattering were applied to low-resolution data. Previously, we observed the same apparent overestimation of the weaker reflections for hen egg-white lysozyme nanocrystals, solved up to a resolution of 2.1Å (5o4x)[41] (Clabbers *et al.*, 2017). As protein data are generally refined against the amplitudes instead of intensities, we corrected structure factors as:

$$\left|F_{o,corr}(\boldsymbol{h})\right| = \frac{|F_o(\boldsymbol{h})|}{\sqrt{\epsilon_m(d)}} \tag{26}$$

The resolution dependency of $\epsilon_m(d)$ was determined from the dynamical scattering errors $\langle\left|F_{e,bin}\right|^2\rangle$ within 10 equally sized resolution bins (in terms of number of reflections). Here a discrete correction was applied for each resolution bin independently in the absence of a continuous function to accurately describe the observed curve of $\langle\left|F_{e,bin}\right|^2\rangle$ as function of resolution d (Fig. 5.5B). In all cases, the models were refined until convergence.

5.3   Results

5.3.1   Structure determination

5.3.1.1. IRELOH
Diffraction data were acquired of three IRELOH nanocrystals over a combined tilt range of 144° using the rotation method (Appendix C2). Merging resulted in 85% completeness up to a resolution of 0.82 Å, which allowed phasing by direct methods (Table 5.1, Appendix C3). The resulting model after phasing reveals the presence of individual hydrogen atoms as indicated by the difference map (Fig. 5.1A).

Automated fixing of the hydrogen atoms can place all but one, indicating the quality of the model (Fig. 5.1B). Although there was clear difference potential, position H5 was not placed automatically based on the model geometry and had to be set manually using AFIX 148, (Fig. 5.1B). The final model after anisotropic refinement shows a well-resolved map (Fig. 5.1C) and high-quality model geometry (Fig. 5.1E), although the standard crystallographic quality indications are rather poor (Table 5.2).

Removing the constraints on the 18 hydrogen atoms allowed refining their positions and bond lengths despite increasing the number of free parameters from 196 to 245, with a total of 156 restraints (Table 5.3). The resulting structural model shows a remarkably high consistency of the hydrogen atoms (Fig. 5.1D).

---

[41] Experimental data are available online at https://doi.org/10.5281/zenodo.1250447

Data were acquired of four EPICZA nanocrystals using the rotation method, with a total tilt range of 213° (Appendix C2). Merging yielded close to complete data up to a resolution of 0.83 Å (Table 5.1, Appendix C3). Although the EPICZA molecule has two-fold symmetry, this was not used or imposed for structure solution or refinement. Direct methods allowed calculating a map revealing difference potential peaks for localising individual hydrogen atoms (Fig. 5.2A).

Most hydrogen atoms could be placed automatically, indicating good data quality (Fig. 5.2B). Hydrogen atoms H3, H5, and H5B were placed with the coordinates of the corresponding difference peaks observed during refinement with respectively AFIX 147, 23 and 147 (Fig. 5.2B). After anisotropic refinement, the final model shows high quality of the map and accurate model geometry (Fig. 5.1C, E), but again with relatively poor quality indicators (Table 5.2).

In contrast to the X-ray model, we did not observe any density indicating the presence of water molecules. Perhaps this was due to evaporation after inserting the sample into the vacuum column. This would also explain the observed shrinkage of the unit cell by 90 $Å^3$ compared to the X-ray model. Given the $P2_12_12_1$ crystal symmetry, this shrinkage corresponds to a cube with sides of 2.83 Å per missing H_2O molecule, which is very close the volume of one H_2O molecule in liquid water, which on average occupies a cube with sides of 3 Å.

All 16 hydrogen atom positions and bond lengths were refined by removing the constraints, increasing the number of free parameters from 259 to 305, with a total of 267 restraints (Table 5.3). Almost all refined hydrogen atom positions are stable, but H7 and H10 are unstable and move away too far to have a bonding interaction with the non-H atom (Fig. 5.1D).

5.3.2 Dynamical scattering corrections

5.3.2.1 IRELOH
We observed considerable dynamical scattering, leading to a clear overestimation of the lower intensities (Fig. 3A, B). The overestimation decreased with increasing resolution. We fitted an exponential curve to the data, thus defining $\langle |F_e(d)|^2 \rangle$ a continuous function of the resolution (Fig. 3C). This allowed likelihood-based corrections of the intensities dependent on both intensity and resolution using Equation 25.

The likelihood-based corrections for dynamical scattering significantly improved the fit of the model to the data, with an R1 of 13.2% and an $R_{complete}$ of 15.1% (Table 5.2). Furthermore, the gap between R1 and $R_{complete}$ decreases after the likelihood-based corrections were applied, indicating reduced bias. Although the model geometry of non-H atoms was unaffected (Table 5.2, Appendix C4), the rms deviations from the idealised hydrogen bond lengths did improve considerably (Table 5.3, Appendix C4). The improvement can also be inferred from the structure factor plots that show an almost linear correlation when least squares fitting the same hyperbolic curve (Fig. 5.3D).

5.3.2.2 EPICZA

Dynamical scattering affected the observed intensities of the structure factors (Fig. 5.4A,B). Again, the dynamical scattering error $\langle|F_e(d)|^2\rangle$ decreased with the resolution, and a linear curve was fitted to the resulting plot (Fig. 5.4C). Using Equation 25 we then made likelihood-based corrections of each reflection as function of intensity and resolution.

The likelihood-based corrections improved the fit between the model and the experimental data, leading to a significant improvement in R1 and R$_{complete}$ (Table 5.2). The model geometry also improved significantly. The rms deviations of bond lengths of the non-H atoms decreased (Table 5.2, Appendix C4), as did the deviations from idealised hydrogen bond lengths (Table 5.3, Appendix C4). After applying the likelihood-based corrections, the F_o versus F_c plot improved and showed linear correlation also for the weaker reflections with a much lower value for $\langle|F_e(d)|^2\rangle$ (Fig. 5.4D).

5.3.2.3 Lysozyme

The effect of dynamical scattering on the intensities was already observed previously from protein data at 2.1Å (5o4x)[6] (Clabbers *et al.*, 2017). Again, the weaker reflections in the F_o versus F_c plot were overestimated (Fig. 5.5A). We plotted the dynamical scattering error $\langle|F_e(d)|^2\rangle$ as function of resolution, showing a non-linear relation (Fig. 5.5B). We assume this non-linearity resulted from the presence of secondary structural elements and solvent contribution (Fig. 5C). We made discrete likelihood-based corrections per resolution bin as described in §5.2.4.3.

The likelihood-based correction resulted in a significant improvement of the model, as indicated by the reduction of R$_{complete}$ from 29.1% to 26.2%. It also reduced model bias, as witnessed by the smaller gap between R1 and R$_{complete}$. Furthermore, the F_o versus F_c plot improves with a lower dynamical error value (Fig. 5.5D). It is unclear why the average B-factor increased upon correcting for dynamical scattering. However, the model geometry after applying the corrections showed a significant improvement, since rms deviations from ideal bond lengths and bond angles dropped by about 10% (Table 5.4).

5.4   Discussion and conclusions

Electron diffraction allows structure solution even when only small crystals are available, and results in increased contrast of hydrogen atoms compared to X-ray diffraction. Here, we show that scattering potential at individual hydrogen atom positions can be visualised after solving the structure by direct methods, even before interactive model improvement (Fig. 5.1, 5.2). The refined coordinates are of comparable quality to the respective X-ray structures, indicating that organic structures can be solved with electron diffraction at sufficient quality to allow for further interpretation, *e.g.* for drug development by modelling. The positions of the individual hydrogen atoms are remarkably stable, allowing unconstrained refinement of the hydrogen atoms.

Dynamical scattering was reduced experimentally by selecting for minimal crystal size, and by collecting rotation data from multiple crystals in random orientations. The signal-to-noise ratio (SNR) was further boosted by measuring at cryogenic temperatures and using a highly sensitive hybrid pixel detector. Profile fitting of weak intensities allowed data to be extracted at or even below the noise level (French & Wilson, 1978; Oatley & French, 1982; Kabsch, 2010*b*).

Acquiring more data to further increase multiplicity will always benefit data accuracy. Although it would allow a more accurate estimation of the kinematic intensity by averaging out orientation-dependent dynamical deviations, it cannot completely eliminate overestimation of weaker reflection intensities. Dynamical scattering effects could in principle also be further reduced by increasing the acceleration voltage, whilst data quality and structure refinement benefit from additional calibrations of the experiment (Gemmi *et al.*, 2015; Yonekura *et al.*, 2015), zero-loss energy filtering to remove inelastically scattered electrons (Yonekura *et al.*, 2002) and modelling of partial charge (Yonekura & Maki-Yonekura, 2016; Yonekura *et al.*, 2018). We consider these additional, independent measures to be important, but beyond the scope of our paper, as they are enhanced, rather than replaced by the methods we discuss here.

Existing methods of structure refinement that compensate for the effects of dynamical scattering rely on knowledge of the atomic crystal structure. Comprehensive modelling of dynamical scattering by either multi-slice or Bloch waves simulations is computationally challenging. Current implementations assume perfect crystallinity and small unit cells with a limited number of atoms; increasing data multiplicity also increases the computational burden (Jansen *et al.*, 1998; Palatinus, Petříček *et al.*, 2015; Palatinus, Corrêa *et al.*, 2015; Palatinus *et al.*, 2017). There is currently no implementation for correcting protein data.

Here, we introduced a likelihood-based approach, akin to the Wiener filter, that applies a straightforward scaling factor $\epsilon_m(d)$ for down-weighting overestimated intensities as function of intensity and resolution. It is computationally undemanding, has no underlying assumptions concerning crystal quality or – thickness, and is sufficiently general to be implemented straightforwardly even for very complex cases. It is sufficiently general to even allow corrections of single particle cryo-EM data. We show it can substantially improve the fit between the model and experimental data, at the expense of only a few extra parameters, reduce bias (as witnessed by a smaller difference between R1 and $R_{complete}$), and has a positive effect on the model geometry (Table 5.2). In the absence of a predictive theory for the dependency of F_e on d, the resolution dependency of F_e needs to be determined heuristically. For plate-like crystals or needle-shaped crystals that are rotated about an axis normal to their longest dimension, F_o *vs.* F_c curves should be checked also as a function of rotation angle, as higher at higher angles the electron beam travels through the crystal for a longer distance. This implies stronger dynamical scattering. In theory, the relationship between F_e and d can be derived for each crystal by means of a full dynamical simulation. However, in that case it would be preferable to use the simulation results. Where dynamical calculations are not possible or practical, we therefore suggest determining F_e as a function of resolution by analysing F_o *vs.* F_c curves for different resolution bins.

Like the other approaches, the implementation of our method requires an initial model to calculate its structure factor amplitudes. It has frequently been observed that the accuracy of structure factor amplitudes is more important for refinement than for phasing, so in practice this restriction may not limit the application of our and other methods. However, it may be possible to infer the required error parameters even in the absence of an initial model, using intensity statistics such as the Wilson plot, but a brute force strategy could also be considered, since the corrections require only a few parameters. In that case, parameters within a reasonable range could be tried, generating corrected data for phasing. In marginal cases where dynamical scattering prevents initial phasing, such an approach could be helpful. The method could be further improved by implementation at the stage of data integration and/or scaling, using the fact that only reflections simultaneously in Bragg condition for that particular frame can be affected.

Electron crystallography can produce accurate atomic models that conform to expected bonding geometries to a remarkably high degree. Yet, based on the current methods of data analysis, these models often do not fit as well to the observed data as can be expected from X-ray diffraction data. As the latter result from the diffraction of crystals that contain many orders of magnitude more molecules, this observation should not be too surprising. Yet, using current methods, X-ray diffraction models fit better to the experimental data, compared to models refined against electron diffraction data. As X-ray data are obtained from crystals that have many millions times more molecules, this observation should not be too surprising. One of those components that compromises the fit between model and data is dynamical electron scattering. Here we demonstrated that experimental approaches aimed at reducing crystal size and computational reduction of the effects of dynamical scattering lead to improved refinement statistics and model geometry. As both approaches are independent, their combined effects are multiplied, leading to structures with improved geometry that are less biased by prior assumptions and fit better to the observed diffraction data.
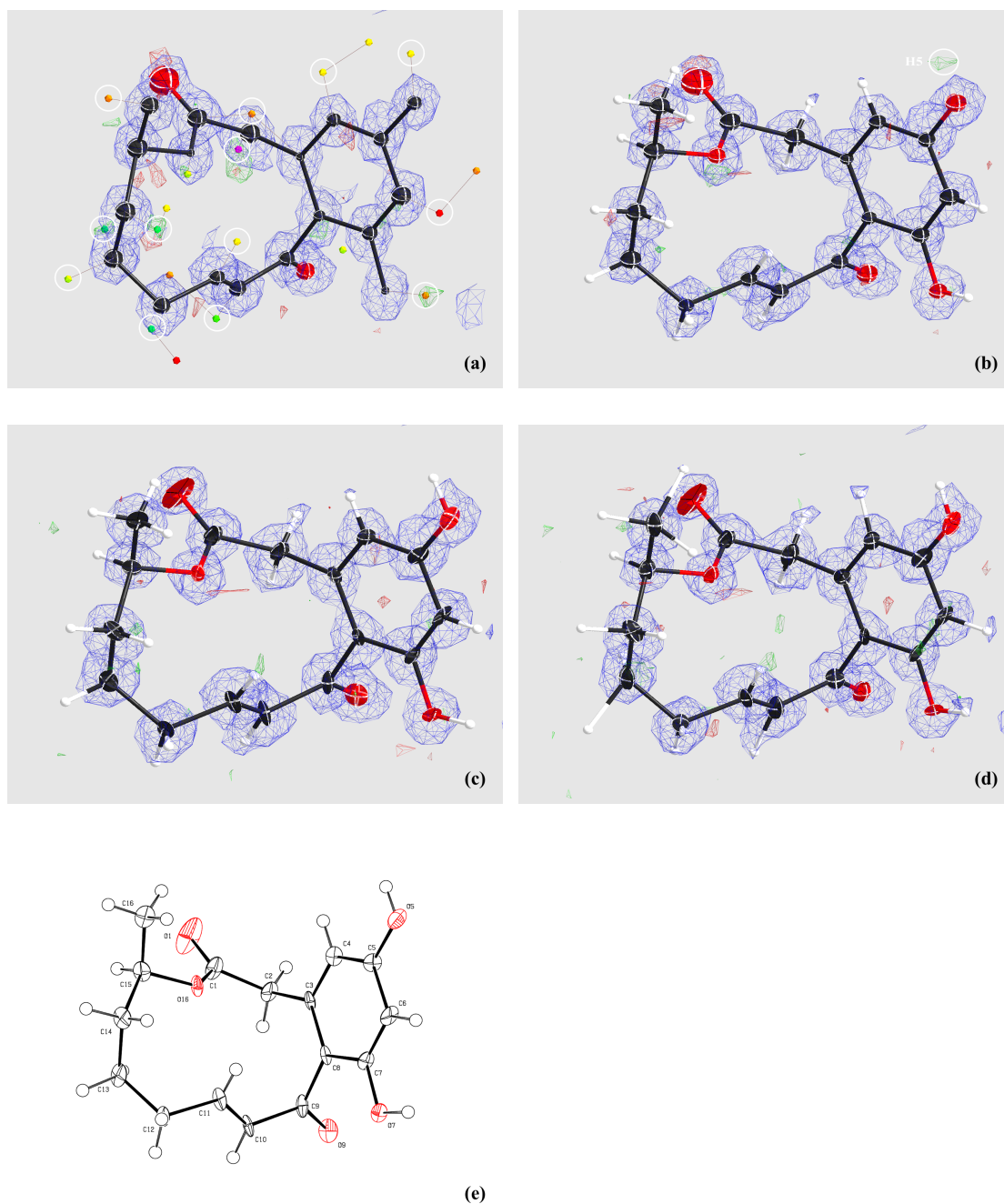
Figure 5.1 Structure determination of IRELOH, (a) the model after phasing using direct methods in *SHELXT*, maxima in the difference map are represented by the small colour coded icosahedra by *SHELXLE* where we encircled difference peaks that indicate the positions of individual hydrogen atoms, (b) automated placement of the hydrogen atoms by *SHELXLE* confirms the quality of the model's geometry, placing all hydrogen atoms correctly except for H5, (c) final model after manual building and anisotropic refinement (RIGU, XNPD 0.001) in *SHELXL*, (d) unconstrained positional refinement of the hydrogen atoms, (e) ORTEP plot with the numbering for the non-H atoms of the final model. Figures a-d were made using *SHELXLE* with standard contour levels of 2.7σ for the $F_o - F_c$ difference map and 1.2σ for the F_o map. The ADP's are colour coded black for carbon, red for oxygen, while the hydrogen atoms are represented in white.
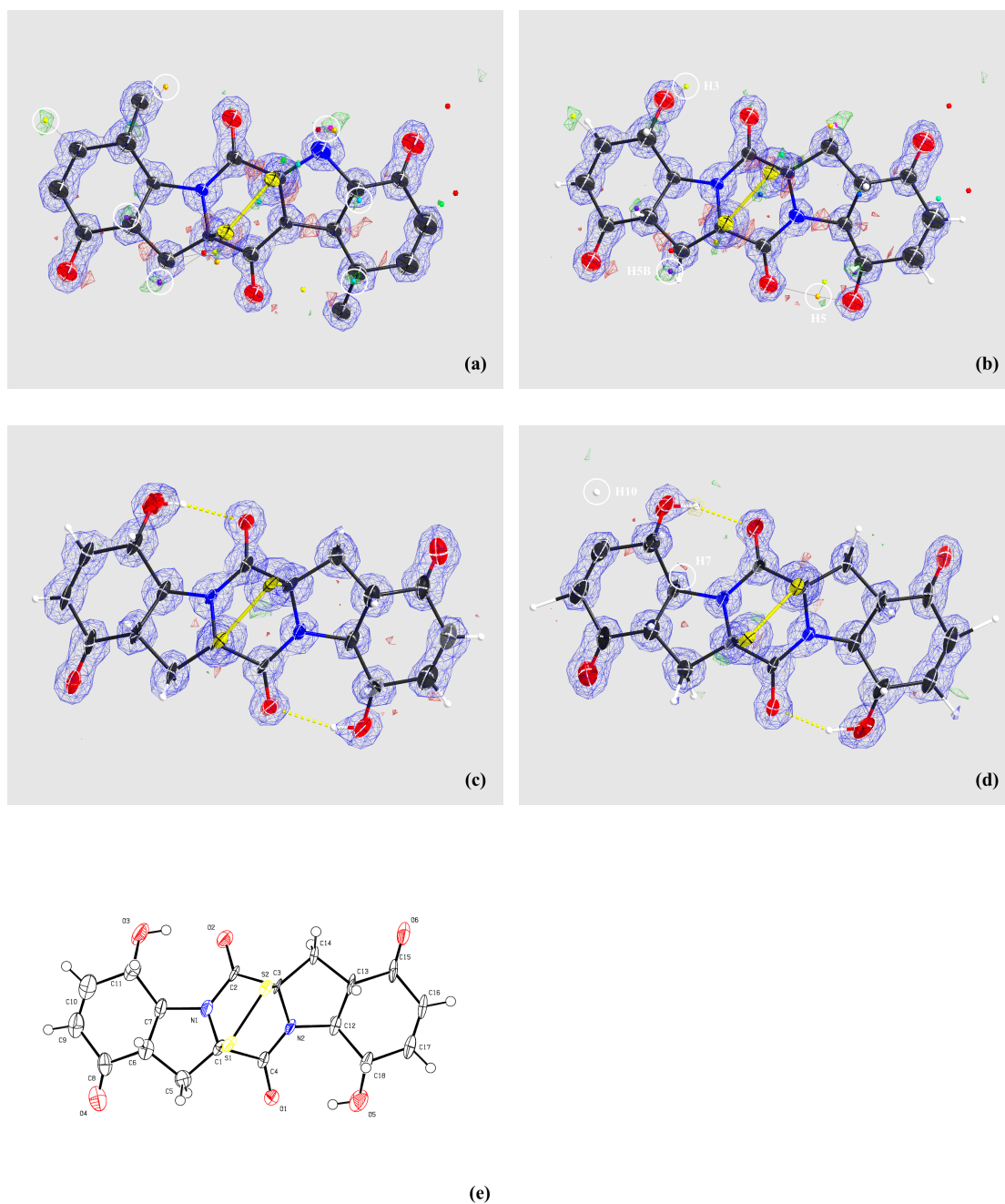
Figure 5.2 Structure determination of EPICZA, (a) the model after phasing using direct methods in *SHELXT*, maxima in the difference map are represented by the small colour coded icosahedra by *SHELXLE* where we encircled difference peaks that indicate the positions of individual hydrogen atoms,, (b) automated placement of the hydrogen atoms by *SHELXLE* showing the quality of the model's geometry, placing all hydrogen atoms apart from H3, H5, and H5B, (c) final model after building and anisotropic refinement (RIGU, XNPD 0.001) in *SHELXL*, (d) unconstrained positional refinement of the hydrogen atoms where the positions of H7 and H10 are unstable and move out of bounds, (e) ORTEP plot with the numbering for the non-H atoms of the final model. Figures a-d were made using *SHELXLE* with standard contour levels of 2.7σ for the F_o-F_c difference map and 1.2σ for the F_o map. The ADP's are colour coded black for carbon, red for oxygen, blue for nitrogen, yellow for sulphur, while the hydrogen atoms are represented in white.
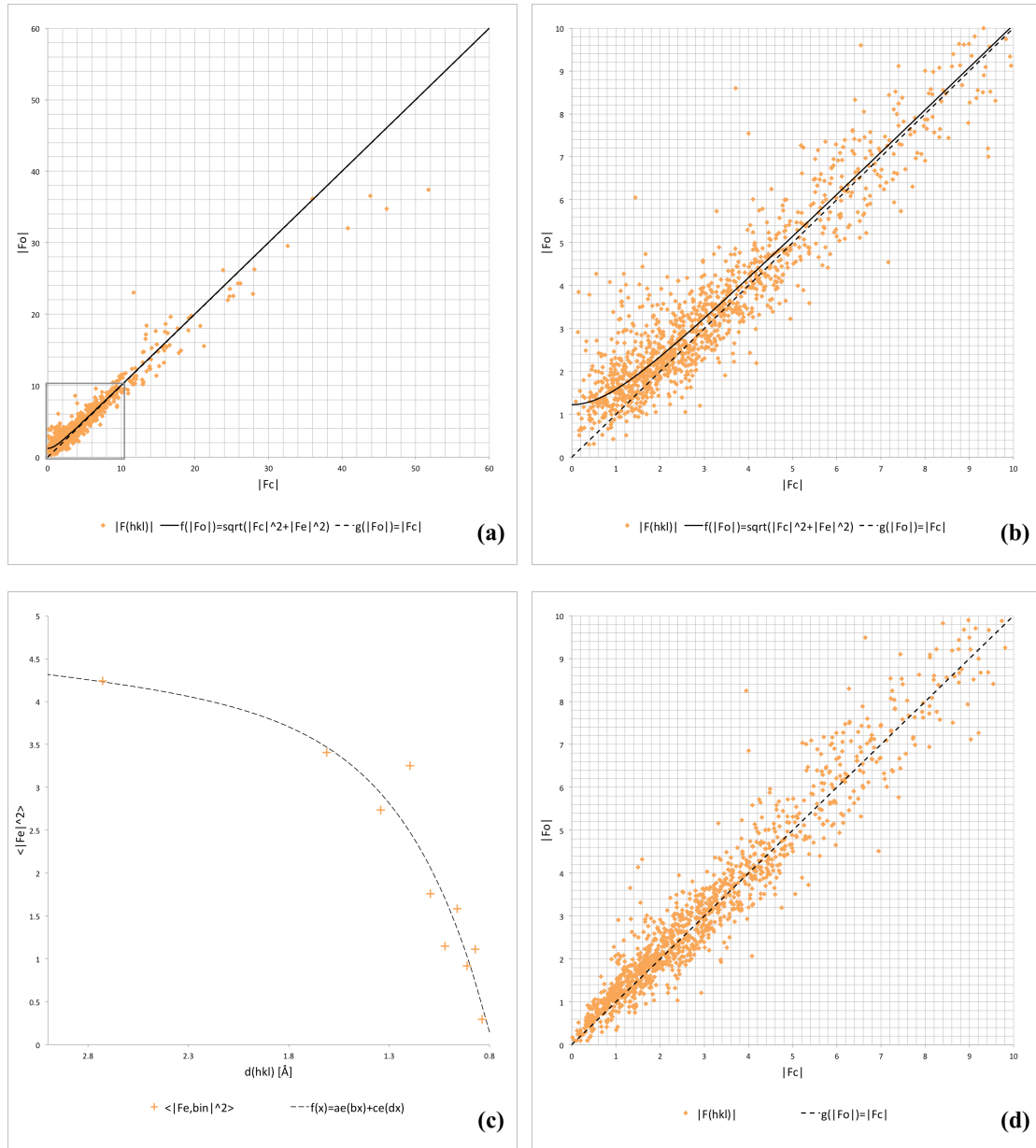
Figure 5.3 Likelihood-based corrections for dynamical scattering, (a) F_o versus F_c plot for IRELOH where $f(|F_o|) = \sqrt{|F_c|^2 + |F_e|^2}$ is a hyperbolic function with an expected dynamical scattering error term of $\langle|F_e|^2\rangle =1.51$ as determined by LS fitting, and where $g(|F_o|) = |F_c|$ shows a perfect linear correlation, (b) zoomed in version on the lower intensity reflections, as marked with a grey box in (a), (c) $\langle\left|F_{e,bin}\right|^2\rangle$ as function of the resolution, where we fitted an exponential curve $f(x) = ae^{bx} + ce^{dx}$ with parameters a=3.64, b=0.058, c=-24.67, and d=-2.38, (d) F_o versus F_c plot after applying the likelihood-based corrections shows an improved correlation between F_o and F_c with an error of $\langle|F_e|^2\rangle$ =0.33.

92

Figure 5.4 Likelihood-based corrections for dynamical scattering, (a) F_o versus F_c plot for EPICZA where $f(|F_o|) = \sqrt{|F_c|^2 + |F_e|^2}$ is a hyperbolic function with an expected dynamical scattering error term of $\langle|F_e|^2\rangle$ =2.20 as determined by LS fitting, and where $g(|F_o|) = |F_c|$ shows a perfect linear correlation, (b) zoomed in version of the same plot on the lower intensity reflections, as marked with a grey box in (a), (c) $\langle|F_{e,bin}|^2\rangle$ as function of the resolution, where we fitted an linear model $f(x) = ax + b$ with parameters a=8.61, and b=-6.98, (d) F_o versus F_c plot after applying the likelihood-based corrections shows an improved correlation between F_o and F_c with an error of $\langle|F_e|^2\rangle$ =0.43.
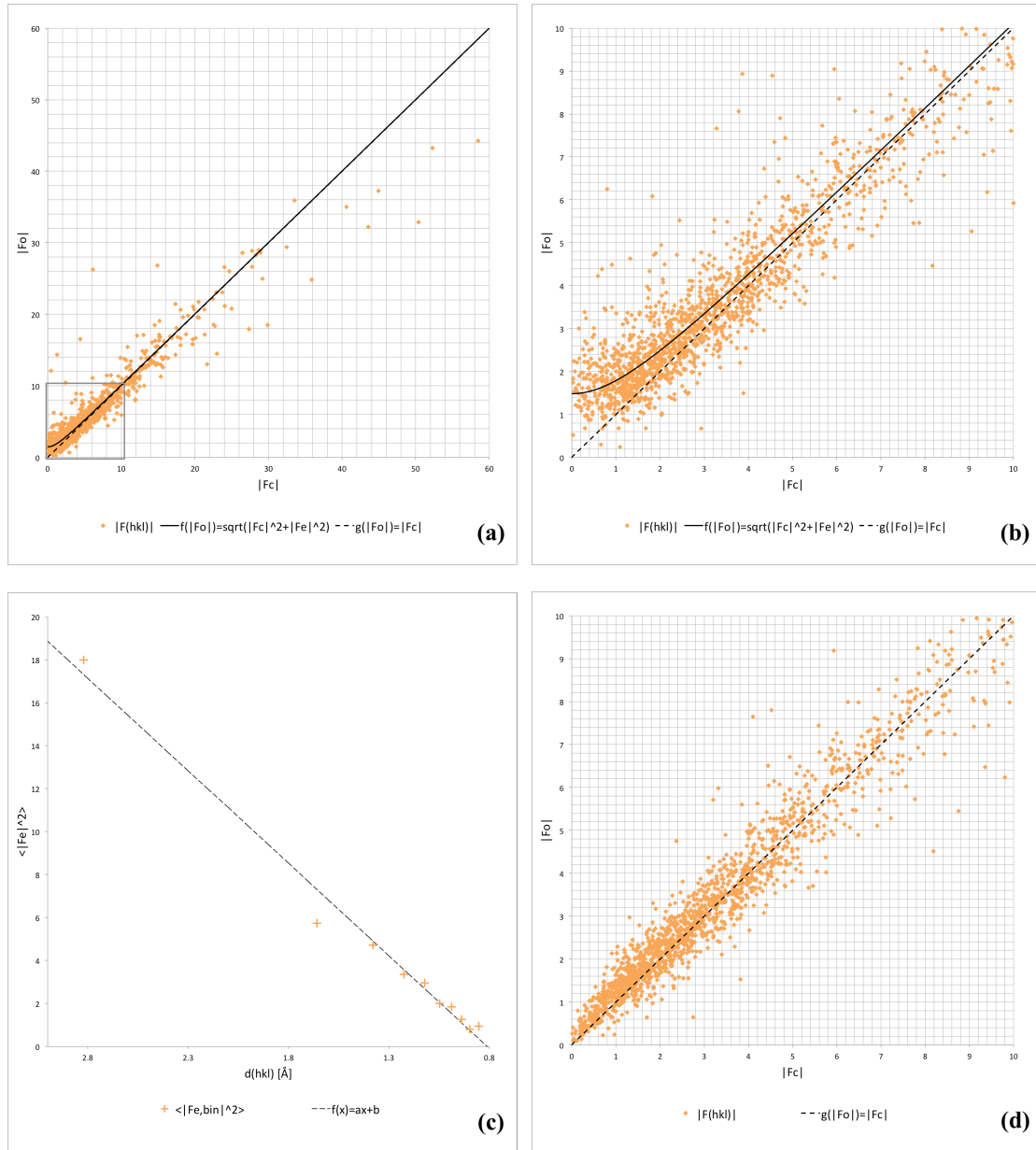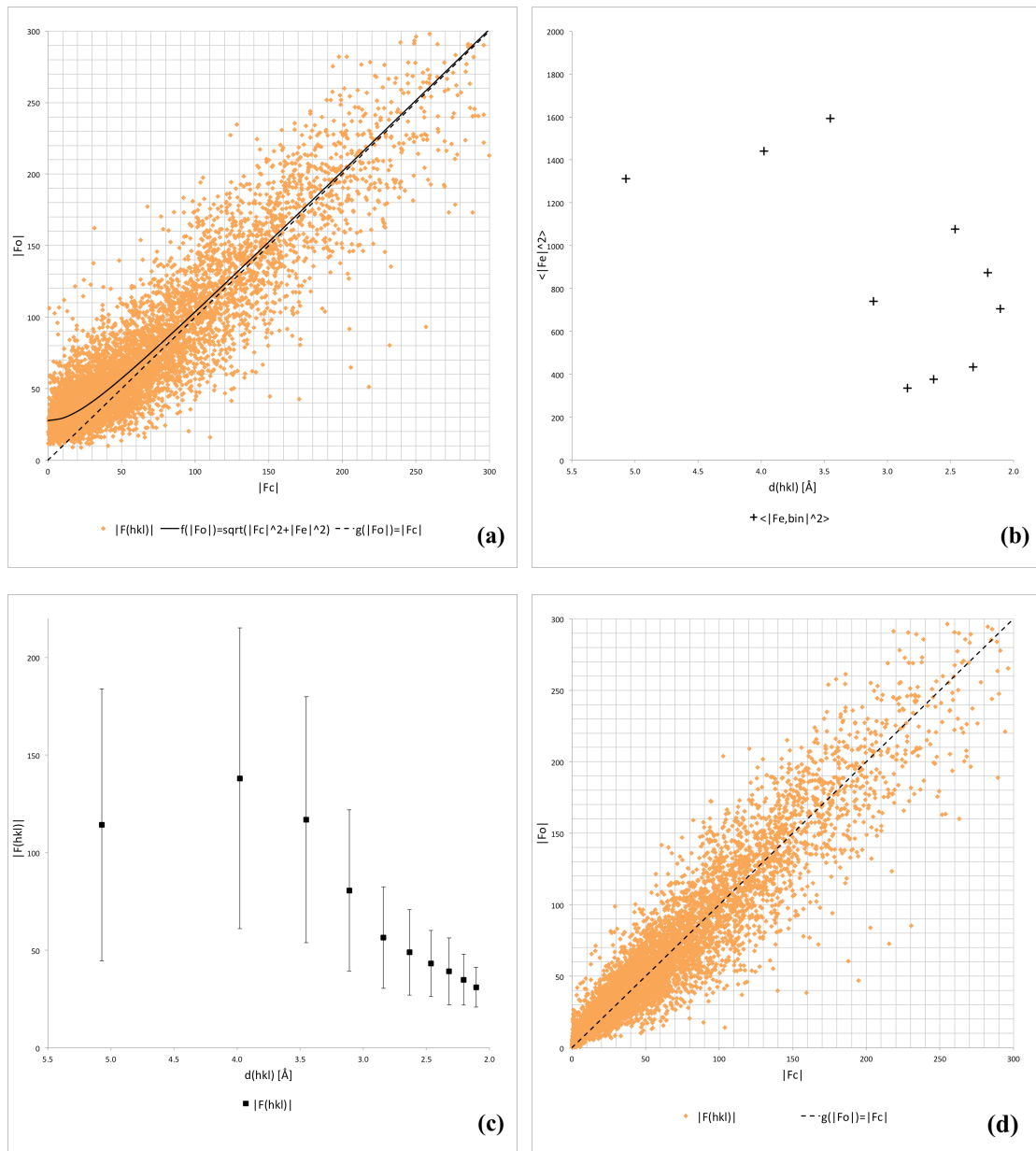
Figure 5.5 Likelihood-based corrections for dynamical scattering for low-resolution protein data, (a) F_o versus F_c plot for hen egg-white lysozyme where $f(|F_o|) = \sqrt{|F_c|^2 + |F_e|^2}$ is a hyperbolic function with an expected dynamical scattering error term of $\langle |F_e|^2 \rangle = 768.13$ as determined by LS fitting, and where $g(|F_o|) = |F_c|$ shows a perfect linear correlation, (b) $\langle |F_{e,bin}|^2 \rangle$ as function of resolution, (c) mean structure factor amplitude $|F(hkl)|$ as function of the resolution, the same resolution bins were used as in (b), errors bar indicate the standard deviation, (d) F_o versus F_c plot after applying discrete likelihood-based corrections, showing an improved correlation between F_o and F_c with an error of $\langle |F_e|^2 \rangle = 53.3$.

Table 5.1 Data merging statistics.

	IRELOH	EPICZA
Literature[42]		
Chemical formula	$C_{16}H_{18}O_5$	$C_{18}H_{16}N_2O_6S_2 \bullet H_2O$
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions		
a, b, c [Å]	8.06, 10.00, 17.73	11.11, 12.61, 13.49
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Data integration		
No. of crystals[43]	3	4
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions		
a, b, c [Å]	8.090(2), 9.940(2), 17.700(4)	11.080(2), 12.580(2), 13.440(3)
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution [Å][44]	17.60-0.82 (0.85-0.82)	12.63-0.83 (0.85-0.83)
I/σI	6.47 (2.50)	5.93 (1.45)
$CC_{1/2}$ [%]	98.2 (46.2)	98.6 (67.7)
R_{merge} [%]	16.6 (50.0)	20.6 (28.6)
R_{meas} [%]	18.5 (61.2)	22.2 (39.8)
Completeness [%]	84.6 (78.9)	89.6 (36.4)
Reflections	6096 (352)	12039 (70)
Unique observations	1332 (135)	1761 (59)

---

[42] Reference values of unit cell dimensions for IRELOH were obtained from Dai *et al.*, 2010, and from Deffieux *et al.,* 1977 for EPICZA.

[43] See Appendix C2 for the data processing statistics of the individual crystal datasets.

[44] Values in parentheses correspond to the highest resolution shell, data were truncated at approximately I/σI ≥ 1.0 and $CC_{1/2}$ ≥ 50% (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013), see also Appendix C3.

Table 5.2 Model building and refinement statistics without refining hydrogen atoms.

	IRELOH	EPICZA
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions[45]		
a, b, c [Å]	8.015(2), 10.015(2), 17.703(4)	10.996 (2), 12.452(2), 13.218(3)
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Parameters	196	259
Restraints[46]	156	267
Reflections	6096	12039
Unique	1332	1761
Refinement		
R1 [%][47]	16.8 (15.0)	17.2 (15.4)
$R_{complete}$ [%][48]	19.7	21.0
wR2 [%]	36.9	39.0
GooF	1.100	1.109
Rmsd bond lengths [Å][49]	0.022(18)	0.027(19)
Dynamical corrections		
R1 [%][47]	13.2 (12.2)	12.7 (11.9)
$R_{complete}$ [%][48]	15.1	14.3
wR2 [%]	29.5	29.1
GooF	0.944	0.879
Rmsd bond lengths [Å][49]	0.022(17)	0.025(13)

---

[45] New unit cell dimensions after lattice refinement, see also Appendix C5.

[46] Enhanced rigid bond restraints (RIGU) were applied for refinement in *SHELXL* (Thorn *et al.*, 2012).

[47] $R1 = \sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||/\sum_{hkl}|F_o(hkl)|$ where the sum is over all reflections, values in parenthesis show R1 for reflections $F_o > 4\sigma(F_o)$.

[48] $R_{complete}$ was calculated over all reflections with a 0.2% test set size as a robust and unbiased validation tool as described (Luebben & Gruene, 2015), since all data are included R_{work} is equivalent to R1.

[49] Rmsd for all non-hydrogen atoms, calculated against reference values from high-resolution X-ray models (Dai *et al.*, 2010; Deffieux *et al.*, 1977), see also Appendix C4.

Table 5.3 Model building and refinement statistics after unconstrained refinement of hydrogen atoms.

	IRELOH	EPICZA
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions[50]		
a, b, c [Å]	8.015(2), 10.015(2), 17.703(4)	10.996 (2), 12.452(2), 13.218(3)
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Parameters	245	305
Restraints[51]	156	267
Reflections	6096	12039
Unique	1332	1761
Refinement		
R1 [%][52]	15.7 (13.9)	16.6 (14.7)
$R_{complete}$ [%][53]	19.9	21.5
*w*R2 [%]	34.6	37.1
GooF	1.031	1.051
Rmsd bond lengths [Å][54]	0.024(18)	0.030(20)
Rmsd H-bond lengths [Å][55]	0.180(72)	0.259(80)
Dynamical corrections		
R1 [%][52]	12.5 (11.5)	12.2 (11.4)
$R_{complete}$ [%][53]	15.2	14.5
*w*R2 [%]	28.1	28.0
GooF	0.907	0.851
Rmsd bond lengths [Å][54]	0.022(13)	0.026(13)
Rmsd H-bond lengths [Å][55]	0.073(52)	0.110(56)

---

[50] New unit cell dimensions after lattice refinement, see also Appendix C5.

[51] Enhanced rigid bond restraints (RIGU) were applied for refinement in *SHELXL* (Thorn *et al.*, 2012).

[52] $R1 = \sum_{hkl}||F_o(hkl)| - |F_c(hkl)||/\sum_{hkl}|F_o(hkl)|$ where the sum is over all reflections, values in parenthesis show R1 for reflections $F_o > 4\sigma(F_o)$.

[53] $R_{complete}$ was calculated over all reflections with a 0.2% test set size as a robust and unbiased validation tool as described (Luebben & Gruene, 2015), since all data are included R_{work} is equivalent to R1.

[54] Rmsd for all non-hydrogen atoms, calculated against reference values from high-resolution X-ray models (Dai *et al.*, 2010; Deffieux *et al.*, 1977), see also Appendix C4.

[55] Rmsd form idealized hydrogen bond lengths after unconstrained refinement of the hydrogen positions (Gruene, Hahn *et al.*, 2014; Sheldrick, 2015*a*), see also Appendix C4.

Table 5.4 Refinement statistics and dynamical scattering corrections of low-resolution protein data from seven lysozyme datasets recorded[56] and solved previously as in §3.3 (Clabbers *et al.*, 2017).

Data integration	
Space group	$P2_12_12$
Unit cell dimensions	
a, b, c [Å]	104.56, 68.05, 32.05
α, β, γ [°]	90.0, 90.0, 90.0
Number of crystals	7
Resolution [Å][57]	57.03-2.11 (2.17-2.11)
R_{merge} [%]	42.1 (57.2)
$CC_{1/2}$ [%]	90.4 (60.3)
I/σI	2.7 (1.0)
Completeness [%]	62.1 (49.8)
Reflections	41191 (1462)
Unique reflections	8560 (545)
Refinement	
Reflections	8503
R1 [%][58]	24.4
$R_{complete}$ [%][59]	29.1
<B> [Å^2]	33.02
Rmsd bond lengths [Å]	0.074
Rmsd bond angles [°]	1.0706
Ramachandran	
Favoured, allowed, outliers [%]	98.4, 1.6, 0.0
Dynamical corrections	
Reflections	8503
R1 [%][57]	24.3
$R_{complete}$ [%][58]	26.2
<B> [Å^2]	41.09
Rmsd bond lengths [Å]	0.066
Rmsd bond angles [°]	1.0072
Ramachandran	
Favoured, allowed, outliers [%]	98.0, 2.0 0.0

---

[56] Experimental data are available online at https://doi.org/10.5281/zenodo.1250447

[57] Values in parentheses correspond to the highest resolution shell, the data were truncated at approximately I/σI > 1.0 and $CC_{1/2} \geq 50\%$ (Diederichs & Karplus, 2013).

[58] $R1 = \sum_{hkl} \left||F_o(hkl)| - |F_c(hkl)|\right| / \sum_{hkl} |F_o(hkl)|$

[59] We present R1 and $R_{complete}$ instead of R_{work} and R_{free}. With less than 10,000 unique reflections $R_{complete}$ is preferred over R_{free} since it is calculated from all reflections (Brunger, 1997; Luebben & Gruene, 2015). Since all structure factors are used in turn this leads to a more robust calculation than R_{free}. With this validation method, the actual refinement uses all reflections; hence R_{work} is equivalent to R1.

CHAPTER 6

Summary and future perspectives

Electron crystallography makes structure elucidation feasible when only small three-dimensional nanometre-sized crystals are available. Until recently, radiation damage inhibited structure determination by electron diffraction from crystals of highly beam-sensitive material such as organic pharmaceuticals and macromolecules. The introduction of novel detectors enabled fast data acquisition to outrun the radiation damage, and offered a substantial improvement in signal-to-noise ratio to accurately measure weak diffraction data at low-dose conditions. Diffraction data in macromolecular crystallography are preferably acquired using the rotation method, allowing (near) complete sampling of reciprocal space and reducing dynamical electron scattering by tilting the non-aligned crystal in a random orientation.

Electrons are charged particles that are scattered very effectively by the electrostatic scattering potential of the crystal. As each inelastic scattering event deposits only a very small amount of energy into the sample, electrons are several orders of magnitude less damaging than X-rays per useful kinematic diffracted quantum. The scattering angles for inelastic scattering are very low, and therefore most inelastic scattering will end up at or close to the incident beam. However, owing to the high scattering probability of electrons, multiple scattering events occur frequently, including combinations of both elastic and inelastic scattering events. As elastic scattering does have a relatively high scattering angle, these multiple scattering events leads to an increase in diffuse background and a broadening of the Bragg peaks. Furthermore, electrons can scatter elastically more than once, also referred to as dynamical scattering. Dynamical scattering affects the reflection intensities as strong reflections become slightly less intense, and weaker neighbouring reflections are becoming stronger. This affects structure determination, as the measured intensities are proportional to the structure factor amplitudes, which holds no longer true in presence of dynamical scattering. By minimising the thickness of the crystals, dynamical scattering can be reduced experimentally, and structure solution is still feasible.

As presented here, a single three-dimensional nanocrystal with a diffracting volume of only 0.14 μm^3 (about 6×10^5 unit cells) provided sufficient information for successfully solving the structure of a rare dimeric polymorph of hen egg-white lysozyme. The limited number of unit cells that contribute to the signal results in a poor signal-to-noise ratio. As protein crystals are highly sensitive to beam-induced radiation damage, the signal gets substantially worse during prolonged exposure by the incident beam. Therefore it is necessary to enhance the signal-to-noise ratio and outrun the radiation damage by cryo-cooling the sample and recording diffraction data using novel hybrid pixel detectors that are fast and highly sensitive. The structure could be solved by molecular replacement using a monomeric poly-Alanine model. Automated model building was then used to successfully place all the side chains in the resulting electrostatic potential map. Although the standard quality indicators are relatively poor, the resulting map is of high quality and the model geometry compares well to those of high-resolution X-ray models.

As electron diffraction shares many similarities with X-ray crystallography, most data processing routines that were originally developed for X-ray diffraction can be adapted for electron crystallography. For example, the *DIALS* and *XDS* software packages can effectively deal with the rather different diffraction geometry. As the electron wavelength is almost 40 times shorter than that of X-rays, the Ewald sphere

100

is nearly flat, and the consequently low diffraction angles require a high sample to detector distance where the flat Ewald sphere implies a strong correlation between lattice parameters and detector distance. Several pathologies encountered in electron diffraction further complicate data integration; these include drift of the incident beam and distortions arising from lens imperfections. Here, recent features of *DIALS* were introduced, enabling refinement using a smoothly varying beam model and applying corrections for distorted diffraction images. These novel features, combined with the existing tools in *DIALS* for problematic geometry refinement, make data integration and refinement feasible for electron crystallography, even in difficult cases.

Besides macromolecular crystallography, electron diffraction is also highly effective in structure determination of inorganic and organic small molecules. Here the structures of two organic pharmaceuticals are presented that were successfully solved from three-dimensional nanocrystals using electron diffraction. Because of the relative short unit cell dimensions and the highly ordered dense packing of the organic pharmaceuticals, the signal gets a substantial Bragg boost compared to that of a protein crystal with similar dimensions. As a result, Bragg peaks can be integrated up to (sub-)atomic resolution, allowing phasing using direct methods and largely unrestrained refinement of the non-hydrogen atoms. Furthermore, electron diffraction has increased contrast for hydrogen atoms when compared to X-ray crystallography. Therefore, individual hydrogen atom positions could already be visualised immediately following phasing without any manual model building. Even more so, the hydrogen atom positions could be refined unconstrained in most cases, indicating the high quality of the data and the high level of detail given by the map.

Despite the high quality of the map and the model geometry, certain pathologies remain in electron crystallography. Although the maps reveal high-resolution detail and allow resolving individual hydrogen positions, standard quality indicators are quite poor. An uncorrelated complex-valued component especially affecting the weaker reflection intensities was observed when comparing the measured structure factor amplitudes to amplitudes that were calculated from the model. Dynamical scattering affects the structure factor amplitudes and results in an overestimation of the weaker reflection intensities. To reduce the effect of dynamical scattering on structure refinement, a likelihood-based method is introduced here that down-weights the overestimated intensities using a single scaling factor that is a function of both intensity and resolution. Applying the likelihood-based corrections improved the fit of the data to the model, as well as the standard quality indicators and the model geometry.

Electron diffraction can thus provide an attractive alternative to X-ray crystallography when only small crystals are available of organics and macromolecules. However, multiple challenges still remain in sample preparation, hardware optimisation, and structure determination. This includes optimising crystallisation of nanometre-sized crystals, improved sample support and grid preparation, a better characterisation and careful calibration of the experimental setup, stable and reliable goniometer tilt movement, and using zero-loss energy filtering which can further boost the signal-to-noise ratio by reducing the diffuse background and sharpening the Bragg peaks. Data quality can further be improved by going to higher acceleration voltages, reducing the probability of (multiple) dynamical scattering events, and thus improving data quality – and/or allowing studying slightly

thicker crystals. Experimental phasing of electron diffraction data from for three-dimensional crystals is currently not possible, and the feasibility of experimental phasing approaches such as extracting spatial phase information from imaging and/or isomorphous replacement using heavy-atom derivatives should be further investigated. Once an initial set of experimental phases is obtained, the phase information can then be extended using existing density modification approaches.

Although data processing, structure solution, model building and refinement can be preformed using standard routines from X-ray crystallography, this is not an optimal treatment of the electron diffraction data. Electrons are charged particles and thus interact differently with the crystal than X-rays, resulting in an electrostatic scattering potential map rather than an electron density map. These differences need to be addressed in structure determination via a better representation of the map by modelling of partial charge and anisotropy, taking into account charge distribution and ionic scattering factors. Accurate modelling of dynamical scattering would further be beneficial in optimising data quality. Dynamical scattering can be modelled by either multi-slice or Bloch wave simulations but are computationally expensive and rely on knowledge of the atomic crystal structure. It is considerably more challenging to minimise the effects of dynamical scattering without prior knowledge of the model. However dynamical scattering might be treated at the stage of data integration based on extensive simulations, interpretation of intensity statistics, and the fact that only reflections that are simultaneously in Bragg condition during one exposure can be affected for that specific diffraction pattern.

Electron crystallography is becoming an increasing relevant research technology for structural biology. With the results described in my thesis, I hope to have made a significant contribution to this development.

APPENDIX A

Supporting information Chapter 3
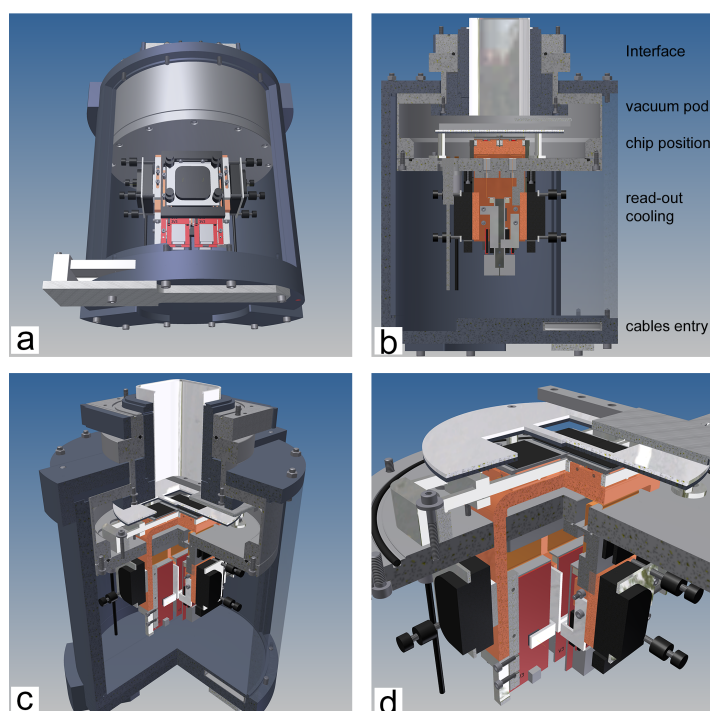
A1   Hybrid pixel detector design

Figure A1: Design drawings of the camera showing (a) bottom to top view of the read-out and cooling of the non-vacuum part, (b) side view of a 180° cross section, (c) side view of a 90° cut-out, (d) 90° cut-out of the bottom vacuum flange showing detector chips, readout, and cooling (van Genderen, 2015). The assembly is holding four 512x512 pixel Timepix quad chips, and was developed to fit on-axis underneath a FEI Titan Krios TEM below a FEI Falcon direct electron detector. The Timepix chip assembly, read-out, and control software were provided by Amsterdam Scientific Instruments (Amsterdam, the Netherlands). The detector chips are covered by a single 300 μm thick Silicon sensor layer, allowing extraction energies of up to 200 kV on a TEM without having any significant damage to the Timepix ASICs (Faruqi & McMullan, 2011). The camera is cooled by a single water loop at four different positions from which the detector chips are being cooled (Peltier controlled) in pairs, to have a stable temperature of 2 ± 0.1 K below ambient temperature. The Relaxd read-out boards (Visser *et al.*, 2011) are directly cooled by the water loop.

A2   Data processing of individual lysozyme datasets using *XDS*

Table A2.1: Data acquisition and integration statistics of individual crystals

	1	2	3	4	5	6	7
Data collection							
Wavelength [Å]	0.02508						
Frame exposure [s]	0.5	0.5	0.2	0.3	0.3	0.3	0.3
φ_{total} [°]	38.15	42.64	20.16	20.11	20.30	20.20	20.15
$\Delta\varphi$ [°/frame]	0.0760	0.1615	0.0344	0.0481	0.0481	0.0481	0.0481
Exposure dose [$e^-.Å^{-2}$]	4.41	10.97	9.82	10.49	10.59	10.55	10.52
Data integration							
Space group	P2$_1$2$_1$2						
Unit cell dimensions							
a, b, c [Å]	104.56(5), 68.05(8), 32.05(3)						
α, β, γ [°]	90.0(0), 90.0(0), 90.0(0)						
Resolution [Å][60]	41.46-2.11 (2.17-2.11)	32.05-2.50 (2.57-2.50)	41.44-3.08 (3.21-3.08)	24.41-2.54 (2.61-2.54)	27.33-2.54 (2.61-2.54)	57.04-3.06 (3.22-3.06)	52.28-3.08 (3.22-3.08)
R_{merge} [%]	26.3 (56.6)	31.7 (107)	19.3 (65.9)	27.5 (64.9)	25.8 (94.2)	21.1 (37.2)	24.1 (87.1)
I/σI	2.6 (1.0)	2.92 (1.10)	2.80 (1.14)	2.34 (1.09)	2.73 (1.02)	2.52 (1.25)	2.44 (1.03)
Completeness [%]	49.5 (49.8)	41.0 (40.5)	28.8 (33.1)	27.5 (28.0)	23.8 (23.5)	21.4 (15.1)	26.0 (25.0)
Reflections	12601 (1462)	9518 (817)	2040 (283)	2141 (361)	3096 (269)	1568 (150)	2092 (270)
Unique reflections	6749 (545)	3445 (236)	1326 (172)	2210 (164)	1920 (626)	1007 (104)	1199 (142)

Table A2.2: Indexing with *XDS* suggests an orthorhombic lattice without prior knowledge about space group or unit cell parameters. The quality of fit makes a large jump before the last line, correctly suggesting 'oP' as Bravais lattice.

Lattice-character	Bravais-lattice	Quality of fit	Unit cell constants (Ångström and degrees)					
			a	b	c	α	β	γ
31	aP	0.0	32.1	67.7	104.5	89.9	89.9	89.9
44	aP	1.9	32.1	67.7	104.5	90.2	90.1	89.9
35	mP	6.7	67.7	32.1	104.5	90.1	90.2	89.9
33	mP	14.7	32.1	67.7	104.5	90.2	90.1	89.9
34	mP	15.6	32.1	104.5	67.7	90.2	89.9	90.1
32	oP	17.5	32.1	67.7	104.5	90.2	90.1	89.9
37	mC	251.0	211.3	32.1	67.7	89.9	90.2	81.4

Table A2.3: *XDS* correctly suggests Laue group *mmm* from the processed data without prior knowledge about space group or unit cell parameters[61], space group number 16 autoselected by *XDS* is marked with a '*'.

	Space group number	Unit cell constants (Ångström and degrees)						Unique[62]	Rmeas[63]	Compared[64]	Lattice-character	Bravais-lattice
		A	b	c	α	β	γ					
	1	32.0	67.6	104.2	89.8	89.9	89.9	876	16.3	200	31	aP
*	16	32.0	67.6	104.2	90.0	90.0	90.0	506	28.5	570	32	oP
	3	67.6	32.0	104.2	90.0	90.2	90.0	589	27.2	487	35	mP
	3	32.0	67.6	104.2	90.0	89.9	90.0	787	21.5	289	33	mP
	3	32.0	104.2	67.6	90.0	90.1	90.0	779	25.1	297	34	mP
	1	32.0	67.6	104.2	90.2	89.9	90.1	876	16.3	200	44	aP

---

[60] Values in parentheses correspond to the highest resolution shell, the data were truncated at I/σ > 1.0 (Diederichs & Karplus, 2013).

[61] *XDS* only makes suggestions about the Laue group, not the space group.

[62] Number of unique reflections.

[63] Redundancy independent R-factor (Diederichs & Karplus, 1997).

[64] Number of reflections used for calculating R_{meas}.

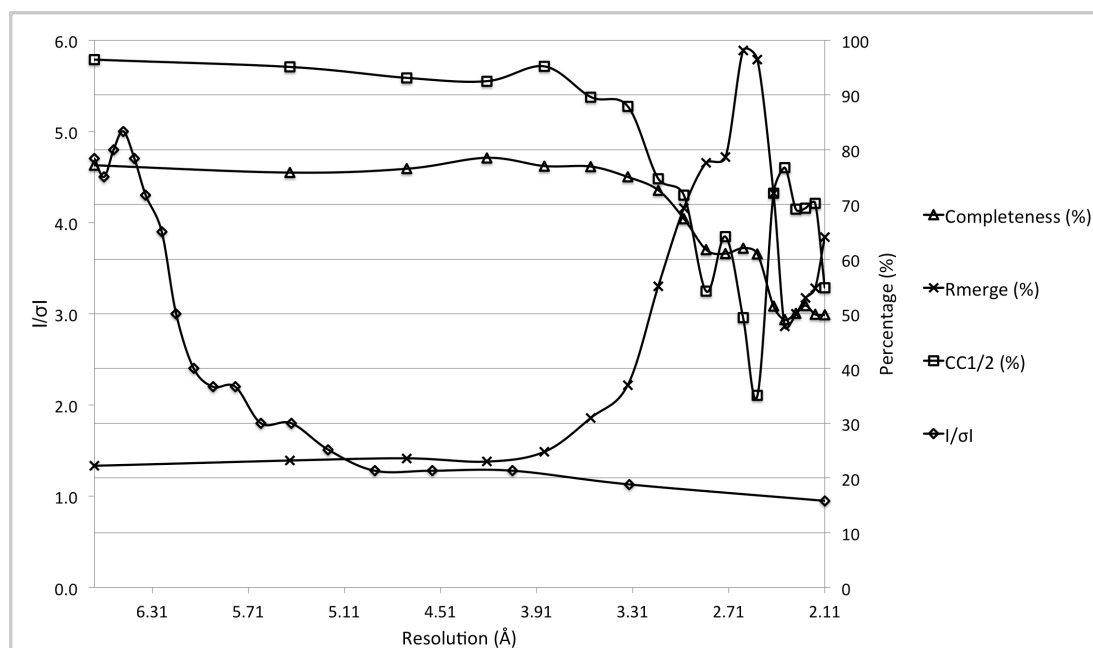## A3   Data merging statics using *XSCALE*

Figure A3.1: Data merging of seven lysozyme nanocrystals (see also Table 3.1), data completeness and data quality indicators are plotted against the resolution. Data completeness is approximately ~75% for the lower resolution bins but drops around 3.0Å and at 2.5Å since not all datasets share the same resolution cut-off (see Appendix A.2). The data were truncated at I/σI ≥ 1.0 (Diederichs & Karplus, 2013).

Table A3.2: Data merging statistics in tabular form presenting data completeness and quality indicators for each resolution bin. The completeness over all data up to 2.1Å increases after merging from ~50% to ~60% compared to the single crystal data. The relatively low completeness can be attributed to radiation damage, preferred crystal orientation and limited goniometer rotation range.

Resolution [Å]	Completeness [%]	R_{merge} [%]	I/σI	$CC_{1/2}$ [%]
9.43	79.4	19.8	5.3	88.6
6.67	77.1	22.2	4.7	96.4
5.45	75.8	23.2	4.5	95.1
4.72	76.5	23.6	4.8	93.1
4.22	78.5	23.0	5.0	92.5
3.86	77.0	24.8	4.7	95.2
3.57	76.9	31.0	4.3	89.6
3.34	75.0	37.0	3.9	87.9
3.15	72.6	55.0	3.0	74.7
2.99	67.4	69.3	2.4	71.7
2.85	61.7	77.6	2.2	54.1
2.73	61.0	78.6	2.2	64.1
2.62	62.0	98.1	1.8	49.3
2.53	60.9	96.4	1.8	35.1
2.43	51.4	72.0	1.5	72.0
2.36	48.9	47.7	1.3	76.7
2.29	50.1	50.0	1.3	69.1
2.23	51.5	52.9	1.3	69.3
2.17	49.9	54.7	1.1	70.2
2.11	49.8	64.0	1.0	54.8
total	61.7	39.8	2.7	91.1

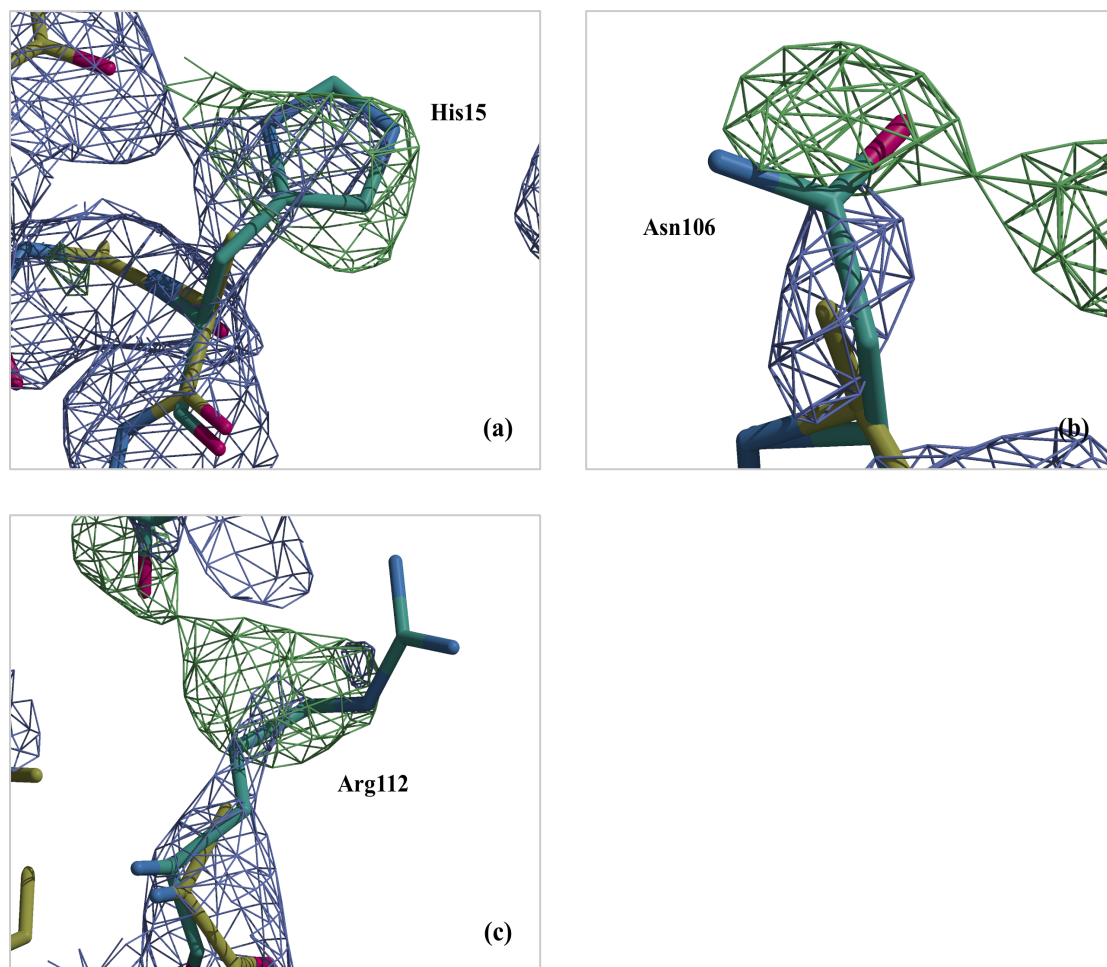## A4 Automated model building using *BUCCANEER* and *REFMAC5*

Figure A4: Electrostatic scattering potential maps of the three side chain residues that were not placed after autobuilding using the merged crystal data, albeit showing clear difference potential in favour of fitting (a) residue His15 in chain A, (b) Asn106 in chain B, and (c) Arg112 in chain B. The yellow carbon represents the model after molecular replacement and autobuilding; the turquoise carbon model represents the fitted side chain residues. All density is shown at a contour level of 1.2 σ.

A5   Model refinement using *REFMAC5*
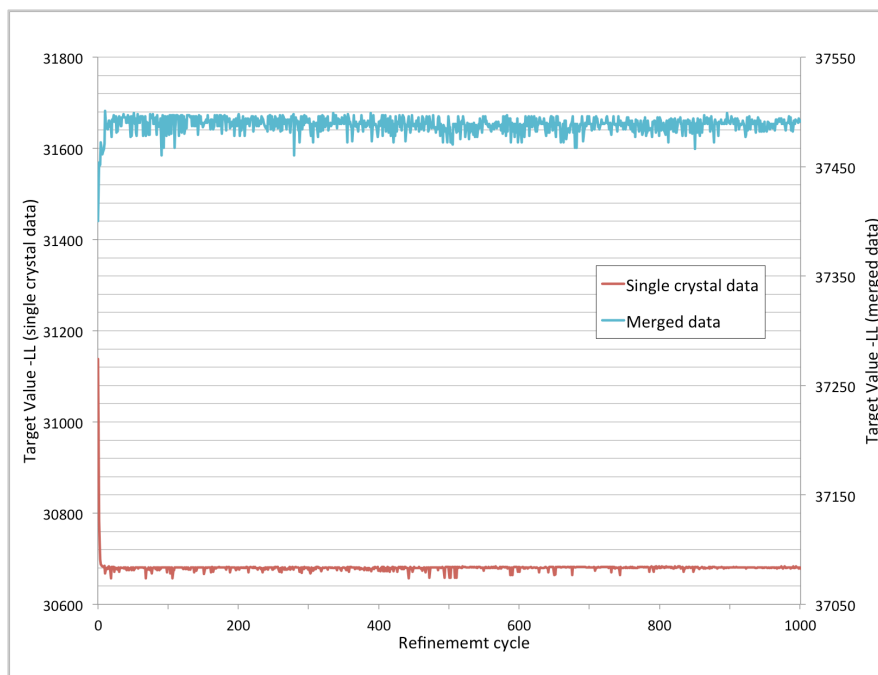
Figure A5: Model refinement with *REFMAC5* for the single and merged diffraction data, indicating convergence well before the 1,000 cycles of refinement used for the final models.

A6 Model geometry comparison between electron and X-ray diffraction

Table A6.1: Model side chain rmsd values calculated for the single and merged lysozyme data compared to an X-ray model of orthorhombic lysozyme. A model of the single crystal data where every side chain was replaced with the most likely rotamer shows significantly higher rmsd values indicative that during autobuilding side chains were placed based on the experimental data.

Reference data		Superimposed data		rmsd [Å][65]
Model	Chain	Model	Chain	
2ybl	A	4r0f	A	0.479
2ybl	A	4r0f	B	0.436
Single crystal	A	4r0f	A	0.739
Single crystal	B	4r0f	B	0.829
Merged data	A	4r0f	A	0.555
Merged data	B	4r0f	B	0.597
Single crystal with most likely rotamer[66]	A	4r0f	A	1.062
Single crystal with most likely rotamer[64]	B	4r0f	B	0.982

[65] Side chain rmsd values were calculated by superimposing a X-ray model of orthorhombic lysozyme (4r0f) (Sharma *et al.*, 2016) and our model and tetragonal lysozyme (2ybl) (De La Mora *et al.*, 2011). Superposition was carried out for each residue using three main chain atoms, and rmsd values were calculated for all atoms using *LSQMAN* (Kleywegt, 1996)

[66] Here a model was created where every rotamer from our refined single crystal model was replaced by the most likely rotamer, ignoring any steric clashes
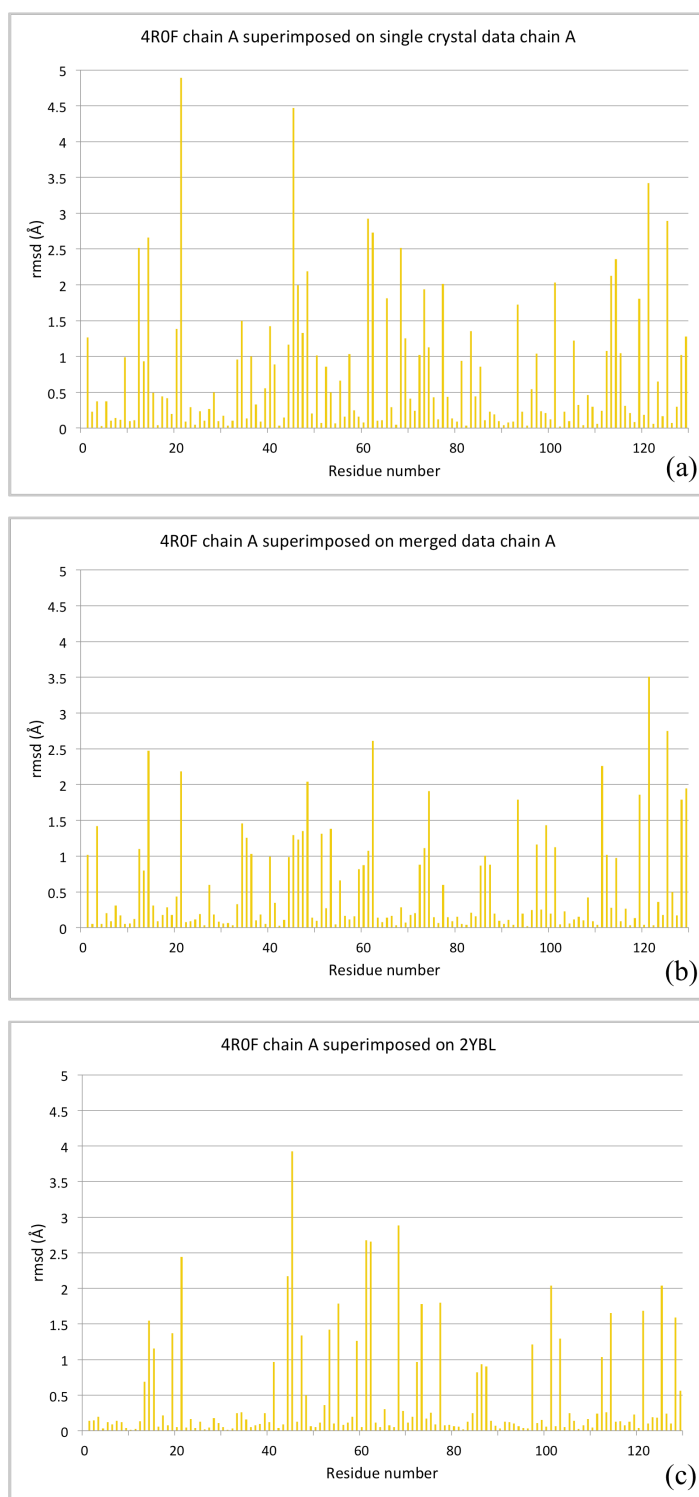
Figure A6.2: Side chain rmsd values plotted for each residue comparing; (a) orthorhombic lysozyme (4r0f) (Sharma *et al.*, 2016) superimposed on chain A of the single crystal data, (b) 4r0f mapped on chain A of the merged data, (c) for comparison 4r0f was superimposed on tetragonal lysozyme (2ybl) (De La Mora *et al.*, 2011), which was used for molecular replacement. Superposition and rmsd were calculated for each residue with *LSQMAN* (Kleywegt, 1996)

A7 F_{obs} vs. F_{calc} graphs for electron diffraction data
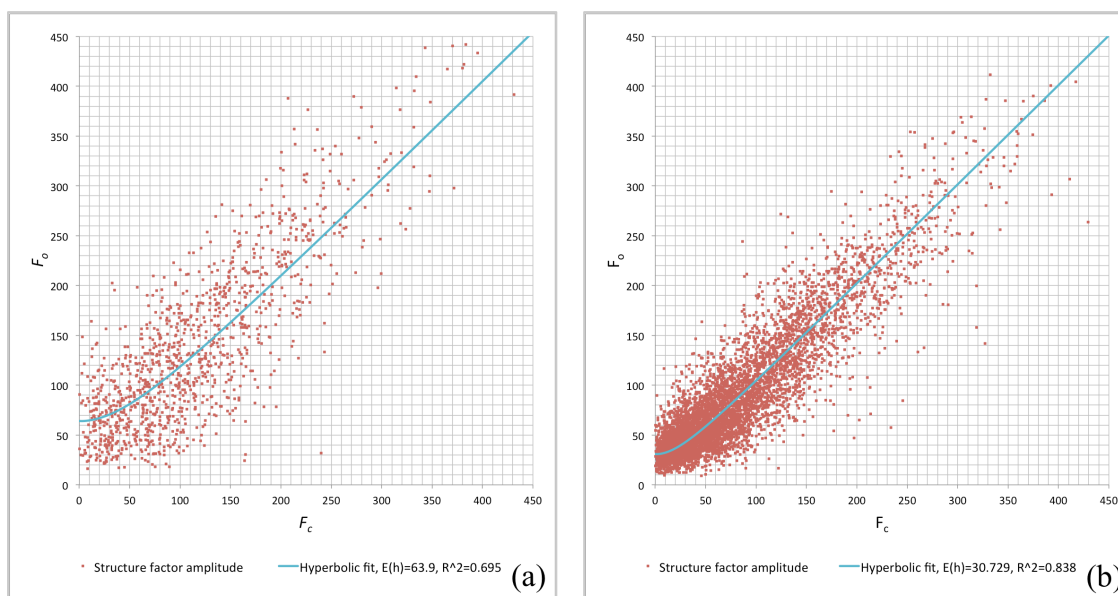
Figure A7: F_o vs. F_c graphs show the extent of dynamical scattering for (a) the low resolution single crystal data with a 4.0Å resolution cut-off, (b) the merged crystal data obtained from merging 7 different crystals (see also Supplement 2). The data were LS fitted with a hyperbolic function described by $\langle|F_o|\rangle = \sqrt{|F_c|^2 + \langle|E(h)|\rangle^2}$.

APPENDIX B

Supporting information Chapter 4

B1 Example parameter files used with *DIALS*

The following options were written to a file that was passed to *dials.import* in order to override the initial model for experimental geometry of the example data set 1:

```
geometry.scan.oscillation=0,0.076
geometry.goniometer.axes=-0.018138,-0.999803,0.008012
geometry.detector.hierarchy{
  fast_axis=1,0,0
  slow_axis=0,-1,0
  origin=-26.3525,30.535,-1890
}
```

Unit-cell restraints to an external target cell were created for data set 6 of the example set by writing the following lines to a file and passing this in at the *dials.index* and *dials.refine* steps:

```
refinement
{
  parameterisation
  {
    crystal
    {
      unit_cell
      {
        restraints
        {
          tie_to_target
          {
            values=32.05,68.05,104.56,90,90,90
            sigmas=0.05,0.05,0.05,0.05,0.05,0.05
          }
        }
      }
    }
  }
}
```

The correct format for these PHIL files may be explored interactively using command-line switches for *DIALS* programs. For example, the command *dials.import* -c -a2 -e2 will show all configuration options for this program along with types and help strings up to an 'expert level' of 2.

Additional parameter files for individual datasets are listed and discussed in the online tutorial available at https://dials.github.io/documentation/tutorials/dials_for_ed.html

114

B2   Simulation for comparison of ED versus MX geometry refinement

To generate simulated spot centroid positions, we started with the real electron diffraction example data set 1, consisting of a continuous rotation scan over 503 images with an angular width of 0.076° per image, for a total scan range of 38.2°. We took the model for the indexed experiments and 'regularised' the geometry of the beam and detector for the purposes of simulation, without changing the crystal model, which had an orthorhombic unit cell with dimensions $a = 31.97$Å, $b = 69.41$ Å, and $c = 104.62$ Å. To regularise the beam and detector models, we forced the beam direction to be exactly aligned to the $-Z$ direction and reoriented the detector model such that the beam intersected the detector in the centre of its square window, and the detector plane was orthogonal to the beam vector. The detector distance remained at the value of 1890 mm, as previously determined and stored in the CBF headers for the images. The real detector consists of 2x2 Timepix quads with large gaps between the active regions. For simplicity we replaced this model with a single panel covering the total extent of the real detector, with no parallax correction, effectively assuming it consists of a perfectly sensitive plane of zero thickness. The updated electron diffraction geometry was written to a new *dxtbx* experiment list and then altered a second time to produce regularised geometry for an X-ray experiment. This involved changing the wavelength from 0.02508Å to 1.0332Å and the detector model such that the total extent and pixel size was equivalent to a Pilatus 6M detector at a distance of 200 mm from the sample. This model was also written to a *dxtbx* experiment list.

The regularised models were used alongside the indexed spot list from the real data set to simulate observed centroid positions for both versions of the experimental geometry. By using the spot list from a real experiment we ensured a realistic distribution of strong spots versus resolution. To make sure that the differences in refinement runs are caused only by the diffraction geometry and not obscured by different sets of input spots, we selected 1571 reflections that could be predicted by both versions of the diffraction geometry.

Simulated centroid positions were calculated for each version of the geometry by predicting their positions then adding random error. The random errors were drawn from a normal distribution with a standard deviation of 0.25 pixels for the X and Y positions and 0.25 images for the Z position. For real data, the centroid position errors in X, Y and Z are neither independent, nor normally distributed. However, the purpose of adding displacements to the centroid positions was merely to ensure that refinement would proceed to convergence with realistic finals rmsds. The centroid positions from spot-finding result from a centre-of-gravity calculation, which also provides estimated errors in these positions that are used to set weights in refinement. These errors have a dependence on the found spot intensity. Rather than simulating new error estimates, we kept the original error estimates from spot-finding on the real data set to give a realistic distribution of weights. The centroid X, Y positions and their errors were rescaled to units of millimetres for use in refinement using the pixel sizes of 55 µm for the electron diffraction detector and 172 µm for the X-ray detector.
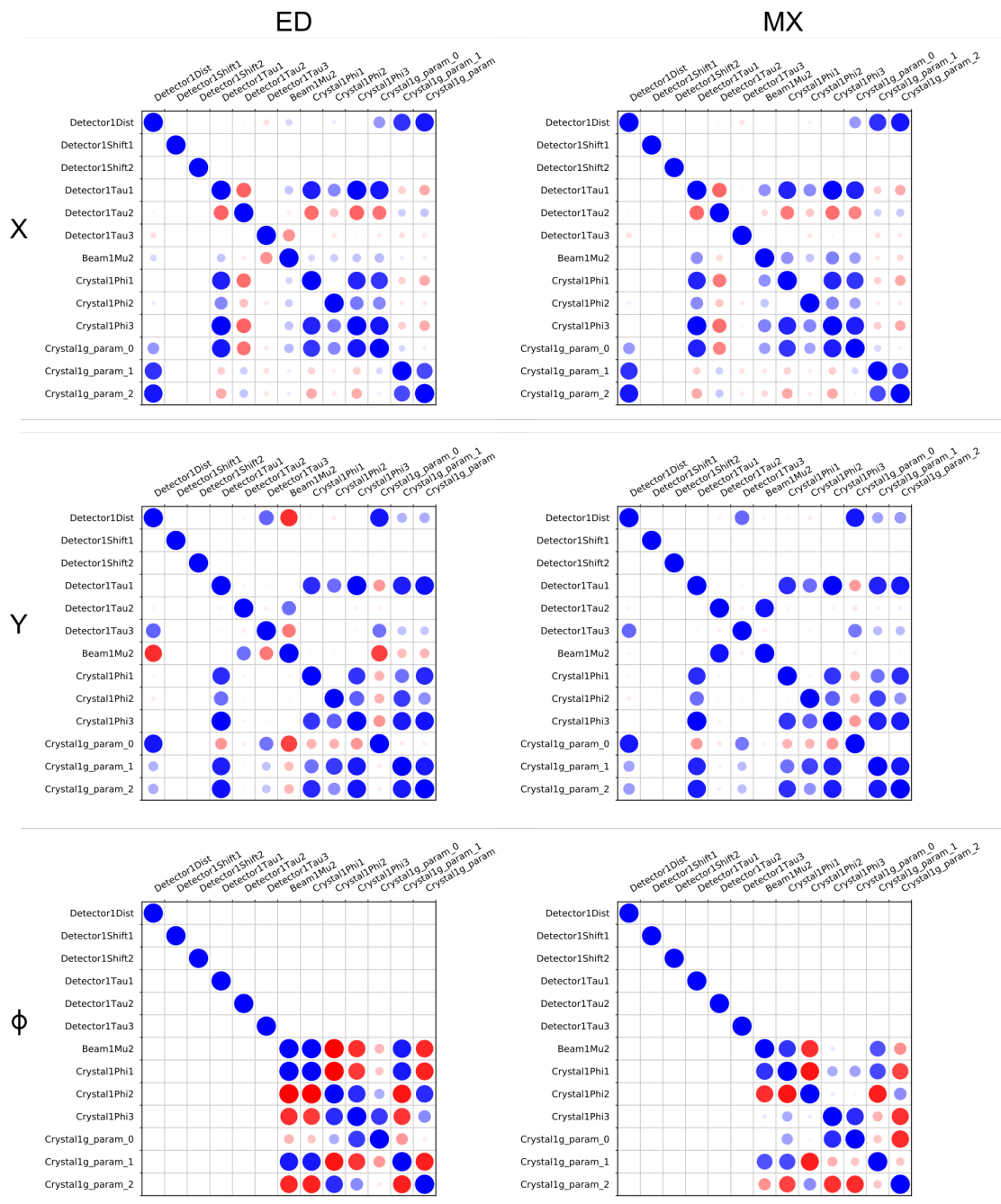
Figure B2: Corrgrams produced for the final step of geometry refinement for simulated data assuming either typical electron diffraction geometry (left column) or X-ray diffraction geometry (right column). The top row shows correlation between effects of different parameters on the positional residuals $(X - X_o)$, the middle row shows those for $(Y - Y_o)$ residuals and the lower row contains the corrgrams for the angular residuals $(\varphi - \varphi_o)$.

B3 Data processing of individual lysozyme datasets using *DIALS*

Table B3: Processing statics for individual datasets as reported by *AIMLESS* (Evans, 2006), values in parenthesis refer to the highest resolution shell. Datasets were processed individually in order to determine suitable resolution cut-offs. These limits were the applied to unscaled data, forming the input to the multiple dataset scaling and merging reported in the main text.

	1	2	3	4	5	6	7
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell							
a [Å]	105.12	104.93	104.25	105.22	103.47	105.00	104.14
b [Å]	68.34	68.51	67.17	69.65	64.73	66.50	68.83
c [Å]	31.98	32.15	31.55	32.35	31.84	31.71	31.73
Resolution [Å]	31.18-2.00 (2.05-2.00)	32.15-2.89 (3.07-2.89)	41.18-2.85 (3.12-2.85)	24.61-2.77 (2.96-2.77)	27.12-2.64 (2.80-2.64)	28.09-3.20 (3.58-3.20)	34.71-3.00 (3.29-3.00)
R_{merge}	0.312 (0.538)	0.218 (0.567)	0.318 (0.538)	0.244 (0.513)	0.248 (0.465)	0.437 (0.613)	0.210 (0.504)
R_{meas}	0.398 (0.667)	0.283 (0.692)	0.437 (0.731)	0.323 (0.659)	0.333 (0.609)	0.583 (0.816)	0.275 (0.665)
$R_{p.i.m.}$	0.244 (0.389)	0.176 (0.381)	0.298 (0.492)	0.210 (0.408)	0.221 (0.387)	0.383 (0.532)	0.175 (0.430)
No. of observations	18907 (1512)	4983 (946)	2034 (571)	2792 (653)	3141 (601)	914 (276)	2012 (513)
Completeness [%]	50.1 (48.4)	42.0 (41.7)	26.4 (28.6)	26.6 (28.7)	29.7 (30.2)	17.3 (18.1)	25.9 (28.1)
Multiplicity	2.4 (2.8)	2.2 (2.7)	1.5 (1.6)	1.7 (2.0)	1.6 (1.9)	1.4 (1.5)	1.6 (1.6)
I/σI	2.0 (1.4)	2.7 (1.5)	2.0 (1.1)	2.2 (1.5)	1.8 (1.4)	2.1 (1.0)	2.2 (1.0)
$CC_{1/2}$ [%]	89.3 (53.1)	91.6 (52.8)	65.5 (51.3)	88.8 (58.8)	85.6 (59.2)	73.1 (55.4)	92.9 (46.4)

APPENDIX C

Supporting information Chapter 5
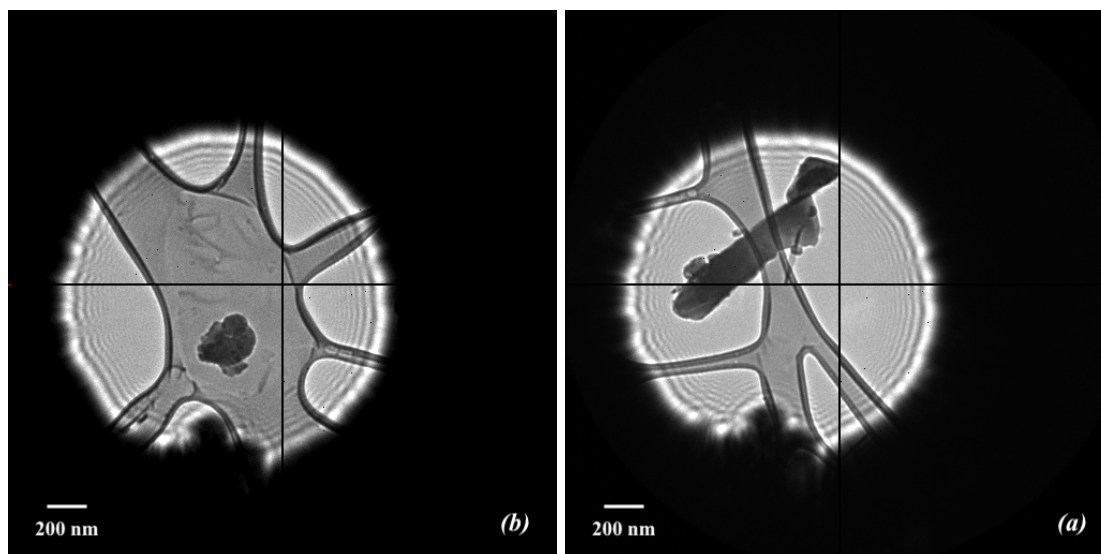
C1   Nanocrystals of organic pharmaceuticals

Figure C1: Micrographs of a 900 by 200 nm IRELOH crystal (a), and a 300 by 200 nm EPICZA crystal (b). Diffraction data were collected with a ~2.0 μm diameter parallel beam that was also used for acquiring the images on the Timepix detector, following directly after diffraction data acquisition.

C2   Data processing statistics of individual datasets using *XDS*

Table C2.1: Data processing statistics of three individual IRELOH crystals

	1	2	3
Data acquisition			
$\Delta\varphi_{frame}$[67] [°]	0.0652	0.0652	0.0652
φ_{total}[68] [°]	56.46	48.90	38.50
Detector distance[69] [mm]	489	489	489
Data integration			
Space group	*P2₁2₁2₁*	*P2₁2₁2₁*	*P2₁2₁2₁*
Unit cell dimensions			
a, b, c [Å]	8.28(5), 9.91(6), 17.84(8)	8.09(3), 9.94(8), 17.70(5)	8.06(4), 10.27(7), 17.59(5)
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution [Å][70]	9.91-1.01 (1.07-1.01)	6.61-0.80 (0.85-0.80)	17.60-0.80 (0.85-0.80)
I/σI	5.03 (1.07)	5.86 (1.48)	4.94 (1.33)
$CC_{1/2}$ [%]	99.6 (77.9)	97.9 (37.9)	98.2 (45.7)
R_{merge} [%]	13.6 (57.1)	11.5 (51.1)	12.0 (51.4)
R_{meas} [%]	16.6 (72.7)	14.3 (67.6)	14.8 (63.8)
Completeness [%]	53.2 (52.8)	60.03 (59.0)	49.8 (35.7)
Reflections	1278 (141)	2734 (269)	2252 (185)
Unique observations	481 (75)	1004 (148)	860 (95)

---

[67] Angular per frame increment during data acquisition, defined as the total tilt range φ_{total} divided by the number of frames (data were acquired with an exposure time of 0.1 s).

[68] The total rotation range over which data were acquired.

[69] The path length between the sample and the detector, the detector distance was calibrated using an aluminum diffraction standard and was not refined during data processing.

[70] Values in parentheses correspond to the highest resolution shell, data were truncated at approximately I/σI ≥ 1.0 and $CC_{1/2}$ ≥ 50% where the correlation is still significant (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

Table C2.2: Data processing statistics of four individual EPICZA crystals

	1	2	3	4
Data acquisition				
$\Delta\varphi_{frame}$[71] [°]	0.0662	0.0664	0.0664	0.0710
φ_{total}[72] [°]	38.33	57.78	57.78	58.93
Detector distance[73] [mm]	489	489	489	489
Data integration				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions				
a, b, c [Å]	11.08(4), 12.58(2), 13.44(1)	11.07(7), 12.63(7), 13.34(9)	11.02(10), 12.78(5), 13.33(1)	11.07(1), 12.13(3), 13.63(2)
α, β, γ [°]	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution [Å][74]	12.58-0.86 (0.91-0.86)	12.63-0.82 (0.87-0.82)	8.35-0.88 (0.94-0.88)	11.07-0.90 (0.95-0.90)
I/σI	6.57 (1.21)	6.02 (1.54)	4.90 (1.07)	5.93 (1.57)
$CC_{1/2}$ [%]	99.2 (70.8)	99.1 (73.2)	99.2 (53.2)	99.0 (41.9)
R_{merge} [%]	9.7 (41.0)	9.5 (22.9)	11.5 (49.3)	14.2 (51.1)
R_{meas} [%]	12.1 (51.2)	12.0 (32.3)	14.0 (63.8)	16.5 (59.3)
Completeness [%]	47.4 (49.3)	68.5 (41.1)	71.0 (68.7)	54.0 (41.5)
Reflections	2198 (226)	3513 (169)	3291 (398)	3068 (315)
Unique observations	835 (135)	1370 (127)	1173 (178)	824 (97)

---

[71] Angular per frame increment during data acquisition, defined as the total tilt range φ_{total} divided by the number of frames (data were acquired with an exposure time of 0.1 s).

[72] The total rotation range over which data were acquired.

[73] The path length between the sample and the detector, the detector distance was calibrated using an aluminum diffraction standard and was not refined during data processing.

[74] Values in parentheses correspond to the highest resolution shell, data were truncated at approximately I/σI ≥ 1.0 and $CC_{1/2}$ ≥ 50% if the correlation is still significant (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

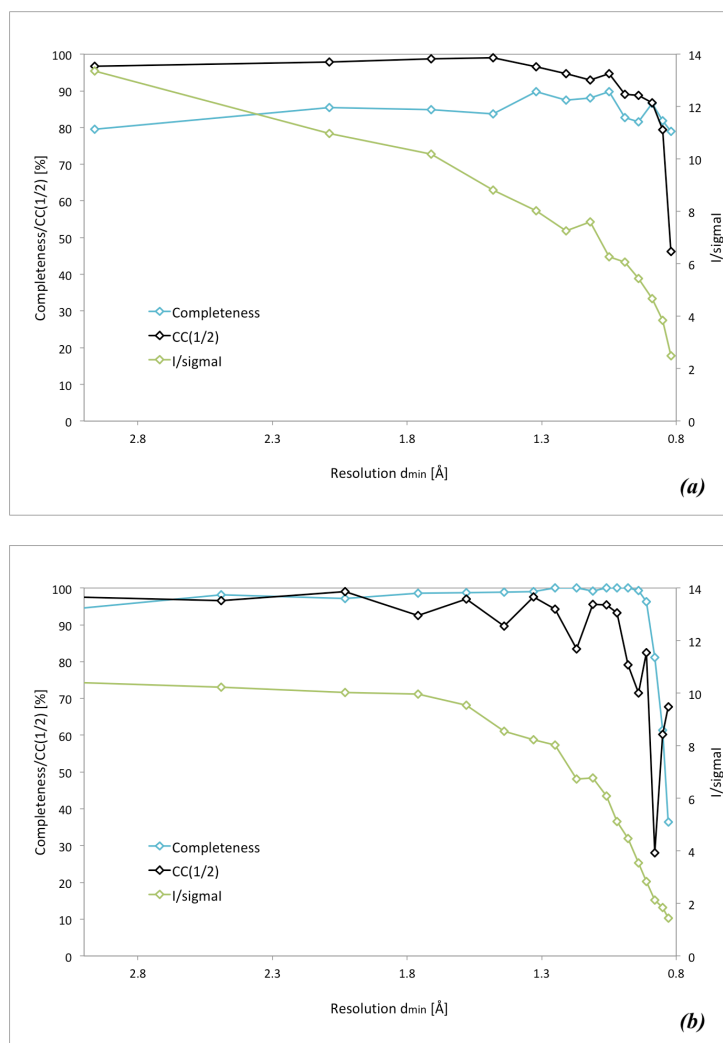## C3   Data merging statistics using *XSCALE*

Figure C3: Merging statistics for IRELOH (a) and EPICZA (b), data completeness and quality indicators $CC_{1/2}$ and $I/\sigma I$ are plotted as function of the resolution d_{min}. Data were integrated, scaled and merged using *XDS* (Kabsch, 2010*b*). Data were truncated at approximately $I/\sigma I \geq 1.0$ and $CC_{1/2} \geq 50\%$ (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

C4 Model geometry comparison

Table C4.1: Bond length comparison for IRELOH between the X-ray model, the electron diffraction model, and the electron diffraction model after applying the dynamical corrections

		X-ray[75]	ED	ED corrected
		d_X [Å]	d_e [Å]	d_e [Å]
C16	C15	1.513(4)	1.47(2)	1.475(15)
C15	C14	1.506(3)	1.523(22)	1.534(13)
C14	C13	1.542(4)	1.54(2)	1.539(15)
C13	C12	1.526(4)	1.538(18)	1.536(13)
C12	C11	1.490(3)	1.493(16)	1.481(10)
C11	C10	1.328(3)	1.348(17)	1.356(10)
C10	C9	1.441(3)	1.476(16)	1.470(11)
C9	O9	1.232(3)	1.201(18)	1.205(12)
C9	C8	1.511(3)	1.503(19)	1.506(13)
C8	C3	1.389(3)	1.445(18)	1.446(13)
C3	C2	1.506(3)	1.505(18)	1.498(12)
C2	C1	1.501(4)	1.517(18)	1.513(12)
C1	O1	1.193(3)	1.210(19)	1.216(13)
C1	O16	1.319(3)	1.330(18)	1.329(12)
C15	O16	1.467(3)	1.461(16)	1.466(11)
C3	C4	1.402(3)	1.385(19)	1.389(14)
C4	C5	1.372(3)	1.350(18)	1.357(12)
C5	O5	1.364(3)	1.382(19)	1.370(13)
C5	C6	1.395(3)	1.41(2)	1.419(14)
C6	C7	1.376(3)	1.36(2)	1.345(15)
C7	C8	1.391(3)	1.393(17)	1.402(11)
C7	O7	1.369(3)	1.360(17)	1.363(12)
			rmsd [Å] [76]	rmsd [Å] [76]
			0.022(18)	0.022(17)

---

[75] Reference bond lengths for equivalent bonds to the electron diffraction model were taken from a previously reported X-ray model (Dai *et al.*, 2010)

[76] $rmsd = \sqrt{\sum_{i=1}^{N}(d_{X,i} - d_{e,i})^2 / N}$

Table C4.2: Hydrogen bond lengths after unconstrained refinement of the hydrogen positions for the electron diffraction model and the electron diffraction model after applying the dynamical corrections.

			ED	ED corrected
		dX-H[77]	d_e [Å]	d_e [Å]
C2	H2A	1.090	1.206(75)	1.085(55)
C2	H2B	1.090	1.066(71)	0.981(51)
C4	H4	1.080	1.091(65)	1.096(49)
O5	H5	0.994	0.906(78)	0.922(56)
C6	H6	1.080	1.112(68)	1.099(47)
O7	H7	1.048	1.025(75)	1.034(54)
C10	H10	1.080	1.258(73)	1.040(52)
C11	H11	1.080	0.929(71)	0.943(53)
C12	H12A	1.090	0.955(75)	0.982(54)
C12	H12B	1.090	0.937(82)	1.088(58)
C13	H13A	1.090	1.132(70)	1.075(46)
C13	H13B	1.090	1.445(68)	1.103(52)
C14	H14A	1.090	1.293(62)	1.178(45)
C14	H14B	1.090	1.186(73)	1.133(53)
C15	H15	1.100	1.282(70)	1.213(52)
C16	H16A	1.060	1.475(75)	1.180(56)
C16	H16B	1.060	1.296(73)	1.148(57)
C16	H16C	1.060	1.218(67)	1.084(51)
			rmsd [Å] [78]	rmsd [Å] [78]
			0.180(72)	0.073(52)

---

[77] Idealized hydrogen bond lengths were generated in *SHELXL* using the NEUT command (Gruene, Hahn *et al.*, 2014; Sheldrick, 2015*a*)

[78] $rmsd = \sqrt{\sum_{i=1}^{N}(d_{X-H,i} - d_{e,i})^2/N}$

Table C4.3: Bond length comparison for EPICZA between the X-ray model, the electron diffraction model, and the electron diffraction model after applying the dynamical corrections.

		X-ray[79]	ED	ED corrected
		d_X [Å]	d_e [Å]	d_e [Å]
S1	S2	2.083(1)	2.096(11)	2.099(7)
S1	C1	1.873(3)	1.864(16)	1.863(10)
S2	C3	1.863(3)	1.876(17)	1.876(11)
C1	N1	1.450(4)	1.409(19)	1.414(12)
N1	C2	1.337(4)	1.39(2)	1.387(13)
C2	C3	1.529(4)	1.529(19)	1.529(12)
C3	N2	1.455(4)	1.422(19)	1.429(12)
N2	C4	1.335(4)	1.33(2)	1.322(12)
C4	C1	1.526(4)	1.554(19)	1.547(11)
N2	C12	1.471(4)	1.468(17)	1.462(11)
C12	C13	1.529(5)	1.55(2)	1.550(14)
C13	C14	1.512(5)	1.47(2)	1.496(12)
C14	C3	1.522(5)	1.54(2)	1.527(13)
C12	C18	1.523(4)	1.540(19)	1.523(12)
C18	C17	1.504(5)	1.48(2)	1.489(13)
C17	C16	1.344(6)	1.334(25)	1.344(15)
C16	C15	1.458(6)	1.47(2)	1.472(13)
C15	C13	1.527(5)	1.52(2)	1.520(13)
C18	O5	1.425(4)	1.41(2)	1.389(14)
C15	O6	1.215(5)	1.22(2)	1.218(14)
C1	C5	1.529(4)	1.53(2)	1.533(13)
C5	C6	1.527(4)	1.48(2)	1.479(13)
C6	C7	1.536(4)	1.53(2)	1.530(14)
C7	N1	1.481(4)	1.504(19)	1.492(12)
C2	O2	1.235(4)	1.22(2)	1.225(13)
C4	O1	1.227(4)	1.26(2)	1.267(13)
C6	C8	1.496(4)	1.52(2)	1.531(14)
C8	C9	1.491(5)	1.53(2)	1.512(15)
C9	C10	1.323(5)	1.28(3)	1.293(17)
C10	C11	1.512(5)	1.53(2)	1.511(15)
C11	C7	1.511(4)	1.54(2)	1.541(14)
C11	O3	1.429(4)	1.39(2)	1.397(13)
C8	O4	1.221(4)	1.17(2)	1.159(14)
			rmsd [Å][80]	rmsd [Å][80]
			0.027(19)	0.025(13)

---

[79] Reference bond lengths for equivalent bonds to the electron diffraction model were taken from a previously reported X-ray model (Deffieux *et al.*, 1977).

[80] $rmsd = \sqrt{\sum_{i=1}^{N}(d_{X,i} - d_{e,i})^2 / N}$

Table C4.4: Hydrogen bond lengths after unconstrained refinement of the hydrogen positions for the electron diffraction model and the electron diffraction model after applying the dynamical corrections.

			ED	ED corrected
		dX-H[81]	d_e [Å]	d_e [Å]
O3	H3	0.980	0.900(56)	0.900(42)
C5	H5A	1.090	1.244(81)	1.116(52)
C5	H5B	1.090	1.232(80)	1.106(48)
O5	H5	0.980	1.021(13)	1.050(64)
C7	H7	1.100	1.626(84)	1.184(54)
C6	H6	1.100	0.950(84)	1.118(52)
C9	H9	1.080	1.434(94)	1.189(60)
C10	H10	1.080	1.680(106)	1.154(66)
C11	H11	1.100	1.193(96)	1.128(57)
C12	H12	1.100	1.103(73)	1.065(46)
C13	H13	1.100	1.303(85)	1.178(51)
C14	H14A	1.090	0.910(87)	1.000(69)
C14	H14B	1.090	1.223(97)	1.154(72)
C16	H16	1.080	1.429(87)	1.170(51)
C17	H17	1.080	1.211(80)	1.119(51)
C18	H18	1.100	1.117(87)	0.744(54)
			rmsd [Å][82]	rmsd [Å][82]
			0.259(80)	0.110(56)

[81] Idealized hydrogen bond lengths were generated in *SHELXL* using the NEUT command (Gruene, Hahn *et al.*, 2014; Sheldrick, 2015*a*)

[82] $rmsd = \sqrt{\sum_{i=1}^{N}(d_{X-H,i} - d_{e,i})^2/N}$

C5   Bond length and bond angle restraints used for lattice refinement

Table C5.1: Restraints on 1,2-distances (DFIX) and 1,3-distances (DANG) used by *SHELXL* and *CellOpt*[83] for lattice refinement of IRELOH, idealised geometrical restraints were generated using the *GRADE* web server.

```
DFIX 1.508 0.016 C16 C15        DANG 2.245 0.022 O1 O16
DFIX 1.518 0.021 C15 C14        DANG 2.406 0.025 C2 O1
DFIX 1.524 0.019 C14 C13        DANG 2.351 0.026 C2 O16
DFIX 1.527 0.024 C13 C12        DANG 2.404 0.022 C1 C15
DFIX 1.497 0.024 C12 C11        DANG 2.405 0.033 C16 O16
DFIX 1.319 0.017 C11 C10        DANG 2.396 0.036 C14 O16
DFIX 1.470 0.015 C10 C9         DANG 2.535 0.033 C16 C14
DFIX 1.238 0.018 C9 O9          DANG 2.524 0.029 C1 C3
DFIX 1.494 0.022 C9 C8          DANG 2.506 0.024 C2 C4
DFIX 1.408 0.013 C8 C3          DANG 2.563 0.024 C2 C8
DFIX 1.512 0.008 C3 C2          DANG 2.419 0.020 C4 C8
DFIX 1.504 0.009 C2 C1          DANG 2.417 0.016 C3 C5
DFIX 1.202 0.015 C1 O1          DANG 2.381 0.033 C4 O5
DFIX 1.341 0.016 C1 O16         DANG 2.411 0.015 C4 C6
DFIX 1.466 0.010 C15 O16        DANG 2.381 0.033 C6 O5
DFIX 1.391 0.010 C3 C4          DANG 2.397 0.014 C5 C7
DFIX 1.387 0.009 C4 C5          DANG 2.372 0.036 C6 O7
DFIX 1.365 0.014 C5 O5          DANG 2.431 0.019 C6 C8
DFIX 1.387 0.009 C5 C6          DANG 2.412 0.027 C8 O7
DFIX 1.385 0.010 C6 C7          DANG 2.499 0.037 C7 C9
DFIX 1.410 0.013 C7 C8          DANG 2.419 0.019 C7 C3
DFIX 1.358 0.015 C7 O7          DANG 2.536 0.033 C3 C9
                                DANG 2.364 0.028 C8 O9
                                DANG 2.576 0.042 C10 C8
                                DANG 2.359 0.027 C10 O9
                                DANG 2.451 0.037 C11 C9
                                DANG 2.511 0.031 C10 C12
                                DANG 2.520 0.048 C11 C13
                                DANG 2.562 0.031 C12 C14
                                DANG 2.567 0.038 C13 C15
```

---

[83] https://github.com/JLuebben/CellOpt

Table C5.2: Restraints on 1,2-distances (DFIX) and 1,3-distances (DANG) used by *SHELXL* and *CellOpt*[84] for lattice refinement of EPICZA, idealised geometrical restraints were generated using the *GRADE* web server.

```
DFIX 1.429 0.011 C11 O3          DANG 2.258 0.020 N1 O2
DFIX 1.499 0.010 C11 C10         DANG 2.381 0.023 C3 O2
DFIX 1.326 0.011 C10 C9          DANG 2.426 0.033 C3 N1
DFIX 1.464 0.013 C9 C8           DANG 2.472 0.036 C2 C1
DFIX 1.222 0.012 C8 O4           DANG 2.486 0.034 C2 C11
DFIX 1.507 0.020 C8 C6           DANG 2.413 0.032 C1 C7
DFIX 1.537 0.012 C6 C5           DANG 2.332 0.026 C5 N1
DFIX 1.530 0.012 C5 C1           DANG 2.761 0.025 N1 S1
DFIX 1.876 0.015 C1 S1           DANG 2.483 0.026 C4 N1
DFIX 2.059 0.026 S1 S2           DANG 2.672 0.031 C4 S1
DFIX 1.876 0.015 C3 S2           DANG 2.851 0.051 C5 S1
DFIX 1.530 0.012 C3 C14          DANG 2.557 0.028 C4 C5
DFIX 1.537 0.012 C14 C13         DANG 2.385 0.030 C1 C6
DFIX 1.546 0.017 C13 C12         DANG 2.364 0.022 C6 N1
DFIX 1.525 0.009 C12 C18         DANG 2.476 0.046 C11 N1
DFIX 1.429 0.011 C18 O5          DANG 2.545 0.045 C11 C6
DFIX 1.499 0.010 C18 C17         DANG 2.556 0.057 C7 C8
DFIX 1.326 0.011 C17 C16         DANG 2.433 0.020 C7 C5
DFIX 1.464 0.013 C16 C15         DANG 2.579 0.071 C5 C8
DFIX 1.507 0.020 C15 C13         DANG 2.390 0.026 C6 O4
DFIX 1.222 0.012 C15 O6          DANG 2.520 0.036 C6 C9
DFIX 1.479 0.011 C12 N2          DANG 2.347 0.024 C9 O4
DFIX 1.456 0.016 C3 N2           DANG 2.432 0.021 C8 C10
DFIX 1.348 0.011 C4 N2           DANG 2.479 0.023 C11 C9
DFIX 1.520 0.011 C4 C1           DANG 2.457 0.026 C7 O3
DFIX 1.216 0.012 C4 O1           DANG 2.467 0.038 C7 C10
DFIX 1.520 0.011 C3 C2           DANG 2.397 0.037 C10 O3
DFIX 1.216 0.012 C2 O2           DANG 2.672 0.031 C2 S2
DFIX 1.348 0.011 C2 N1           DANG 2.483 0.026 C2 N2
DFIX 1.456 0.016 C1 N1           DANG 2.557 0.028 C2 C14
DFIX 1.479 0.011 C7 N1           DANG 2.761 0.025 N2 S2
DFIX 1.525 0.009 C7 C11          DANG 2.855 0.051 C14 S2
DFIX 1.546 0.017 C7 C6           DANG 2.332 0.026 C14 N2
                                 DANG 2.965 0.026 C3 S1
                                 DANG 2.965 0.026 C1 S2
                                 DANG 2.472 0.036 C3 C4
                                 DANG 2.413 0.032 C3 C12
                                 DANG 2.486 0.034 C4 C12
                                 DANG 2.258 0.020 N2 O1
                                 DANG 2.426 0.033 C1 N2
                                 DANG 2.381 0.023 C1 O1
                                 DANG 2.459 0.047 C18 N2
                                 DANG 2.364 0.022 C13 N2
                                 DANG 2.545 0.045 C18 C13
                                 DANG 2.457 0.026 C12 O5
                                 DANG 2.467 0.038 C12 C17
                                 DANG 2.397 0.037 C17 O5
                                 DANG 2.479 0.023 C16 C18
                                 DANG 2.432 0.021 C15 C17
                                 DANG 2.347 0.024 C16 O6
                                 DANG 2.520 0.036 C13 C16
                                 DANG 2.390 0.026 C13 O6
                                 DANG 2.579 0.071 C15 C14
                                 DANG 2.556 0.057 C15 C12
                                 DANG 2.433 0.020 C12 C14
                                 DANG 2.385 0.030 C3 C13
```

[84] https://github.com/JLuebben/CellOpt

C6   Atomic scattering factors for electron diffraction

(a)

(b)
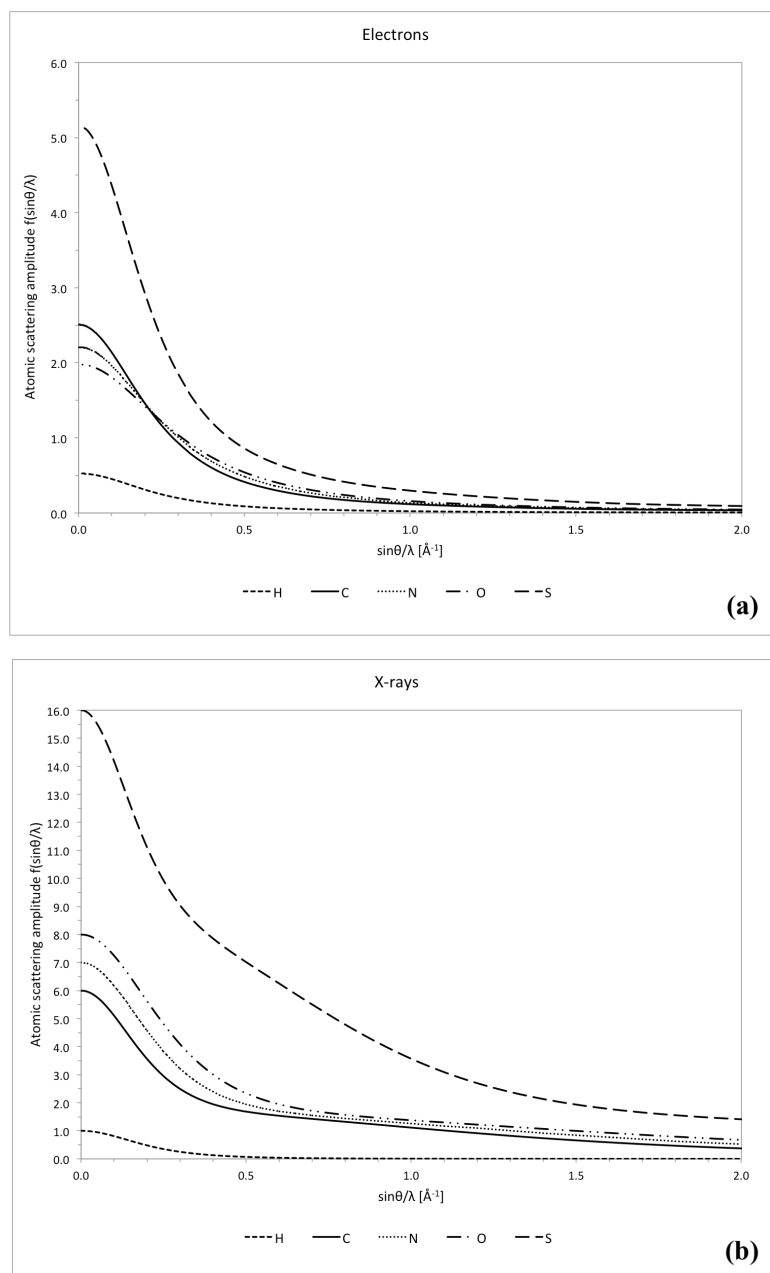
Figure C6.1: Atomic scattering factors for hydrogen, carbon, oxygen, nitrogen and sulphur plotted as function of $sin\theta/\lambda$ for (a) electron diffraction, and (b) X-ray diffraction[85]. Atomic scattering curves for electron diffraction were plotted using the nine Cromer-Mann coefficients with the nine-parameter expansion[86] $f(\sin\theta/\lambda) = \sum_{i=1}^{4} a_i\, e^{(-b_i sin^2\theta/\lambda^2)} + c$. The nine Cromer-Mann coefficients as used by *SHELXL* were obtained by fitting to the eight-parameter fitting described previously by Peng, 1999.

---

[85] Table 6.1.1.1, *International Tables for Crystallography* (2006), Vol. C, Section 6.1, pp. 555-564

[86] Equation 6.1.1.15, *International Tables for Crystallography* (2006), Vol. C, Section 6.1, p. 565
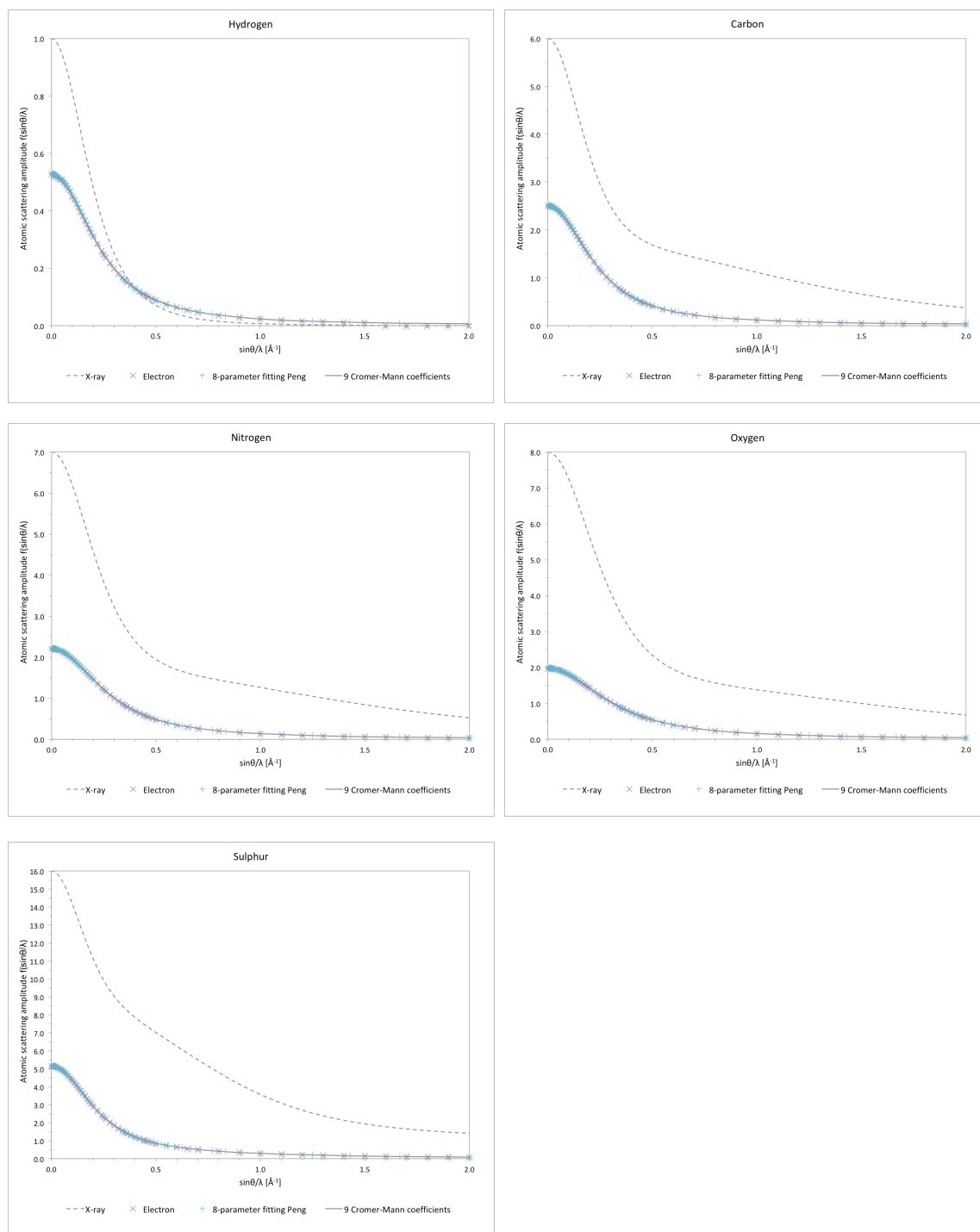
Figure C6.2: Atomic scattering factors plotted as function of $sin\theta/\lambda$. Tabulated values for the mean atomic scattering factors for X-ray[87] and electron[88] diffraction were taken from literature. Values for the 8-parameter fitting for electron scattering factors were taken from Peng, 1999. The atomic electron scattering factors used for modelling and refinement in *SHELXL* are calculated from 9 Cromer-Mann coefficients that were fitted to the 8-parameter fitting from Peng, 1999. Plotting of the curve for the 9 Cromer-Mann coefficients confirms a consistent fit of the scattering factors to the literature values.

---

[87] Table 6.1.1.1, *International Tables for Crystallography* (2006), Vol. C, Section 6.1, pp. 555-564

[88] Table 4.3.1.1, *International Tables for Crystallography* (2006), Vol. C, Section 4.3, pp. 263-271

BIBLIOGRAPHY

Abeyrathne, P. D., Chami, M., Pantelic, R. S., Goldie, K. N. & Stahlberg, H. (2010). *Methods Enzymol.* **481**, 25–43.

Abrahams, J. P. (1993). *Jnt. CCP4/ESF-EACBM Newsl. Prot. Cryst.* **28**, 39–41.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1948–1954.

Adrian, M., Dubochet, J., Lepault, J. & McDowall, A. W. (1984). *Nature.* **308**, 32–36.

van Aert, S., Geuens, P., van Dyck, D., Kisielowski, C. & Jinschek, J. R. (2007). *Ultramicroscopy.* **107**, 551–558.

Arndt, A. W. & Wonacott, A. J. (1977). The rotation method in crystallography. Amsterdam: North-Holland.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S. & Subramaniam, S. (2015). *Science (80-.).* **348**, 1147–1151.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281.

Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N. & Stahlberg, H. (2017). *J. Struct. Biol.* **198**, 124–133.

Blakeley, M. P. (2009). *Crystallogr. Rev.* **15**, 157–218.

Bodenstaff, E. R., Hoedemaeker, F. J., Kuil, M. E., de Vrind, H. P. M. & Abrahams, J. P. (2002). *Acta Cryst. D.* **58**, 1901–1906.

Bricogne, G. (1986*a*). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phases I and II),* Paris: LURE.

Bricogne, G. (1986*b*). *Proceedings of the EEC Cooperative Workshop on Position-Sensitive Detector Software (Phase III),* Paris: LURE.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 2023–2030.

Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *J. Synchrotron Radiat.* **13**, 120–130.

Brunger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.

Calero, G., Cohen, A. E., Luft, J. R., Newman, J. & Snell, E. H. (2014). *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 993–1008.

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B. & Grigorieff, N. (2012). *Structure.* **20**, 1823–1828.

Capitani, G. C., Oleynikov, P., Hovmöller, S. & Mellini, M. (2006). *Ultramicroscopy.* **106**, 66–74.

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Noell, A., Tampé, R., Brandstetter, S., Foerster, A., Mueller, M., Schulze-Briese, C., Bunk, O. & Wang, M. (2016). *Acta Cryst. D.* **72**, 1036–1048.

Chapman, H. N., Fromme, P., Barty, A., White, T. a, Kirian, R. a, Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V, Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B.,

Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K.-U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C.-D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B. & Spence, J. C. H. (2011). *Nature.* **470**, 73–77.

Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. (2015). *Cell.* **161**, 439–449.

Clabbers, M. T. B. & Abrahams, J. P. (2018). *Crystallogr. Rev.* **24**, 176–204.

Clabbers, M. T. B., van Genderen, E., Wan, W., Wiegers, E. L., Gruene, T. & Abrahams, J. P. (2017). *Acta Cryst. D.* **73**, 738–748.

Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P. & Waterman, D. G. (2018). *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 506–518.

Cowley, J. M. (1953*a*). *Acta Crystallogr.* **6**, 516–521.

Cowley, J. M. (1953*b*). *Acta Crystallogr.* **6**, 522–529.

Cowley, J. M. & Moodie, A. F. (1957). *Acta Crystallogr.* **10**, 609–619.

Cowley, J. M. & Moodie, A. F. (1959*a*). *Acta Crystallogr.* **12**, 353–359.

Cowley, J. M. & Moodie, A. F. (1959*b*). *Acta Crystallogr.* **12**, 360–367.

Cowtan, K. D. (2006). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011.

Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A. & Riekel, C. (1998). *Nat. Struct. Biol.* **5**, 634–637.

Dai, J., Krohn, K., Flörke, U., Pescitelli, G., Kerti, G., Papp, T., Kövér, K. E., Bényei, A. C., Draeger, S., Schulz, B. & Kurtán, T. (2010). *European J. Org. Chem.* **2010**, 6928–6937.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. (2014). *Proc. Natl. Acad. Sci.* **111**, 15635–15640.

Dauter, Z. (1999). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 1703–1717.

Dauter, Z. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 389–392.

Dauter, Z. & Wlodawer, A. (2015). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 2217–2226.

Deffieux, G., Gadret, M., Leger, J. M. & Carpy, A. (1977). *Acta Crystallogr. Sect. B.* **33**, 1474–1478.

Diederichs, K. & Karplus, P. A. (1997). *Nat. Struct. Biol.* **4**, 269–275.

Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst. D.* **69**, 1215–1222.

Dorset, D. L. (1976). *J. Appl. Crystallogr.* **9**, 142–144.

Dorset, D. L. (1991). *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5499–5502.

Dorset, D. L. (1992). *Acta Crystallogr. Sect. A.* **48**, 568–574.

Dorset, D. L. (1995). Structural electron crystallography. New York: Plenum Press.

Dorset, D. L. (2007). *Ultramicroscopy.* **107**, 453–461.

Dorset, D. L., Tivol, W. F. & Turner, J. N. (1992). *Acta Crystallogr. Sect. A Found. Crystallogr.* **48**, 562–568.

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W. & Schultz, P. (1988). *Q. Rev. Biophys.* **21**, 129–228.

Egerton, R. F. (2011). Electron Energy-Loss Spectroscopy in the Electron Microscope Springer.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. D. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501.

Evans, G., Axford, D., Waterman, D. & Owen, R. L. (2011). *Crystallogr. Rev.* **17**, 105–142.

Evans, P. R. (2006). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82.

Evans, P. R. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 282–292.

Evans, P. R. & McCoy, A. (2007). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 1–10.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214.

Faruqi, A. R., Henderson, R. & McMullan, G. (2015). *Adv. Imaging Electron Phys.* **190**, 103–141.

Faruqi, A. R. & McMullan, G. (2011). *Q. Rev. Biophys.* **44**, 357–390.

Fischer, H., Polikarpov, I. & Craievich, A. F. (2004). *Protein Sci.* **13**, 2825–2828.

Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V. & Stark, H. (2015). *Nature*. **520**, 567–570.

French, S. & Wilson, K. (1978). *Acta Crystallogr. Sect. A*. **34**, 517–525.

Garman, E. F. & Owen, R. L. (2006). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 32–47.

Gemmi, M., Galanis, A. S., Karavassili, F., Das, P. P., Calamiotou, M., Gantis, A., Kollia, M., Margiolaki, I. & Nicolopoulos, S. (2013). *Microsc. Anal.* **27**, 24–29.

Gemmi, M., Mugnaioli, E., David, J., Tossi, C., Galanis, A. S., Das, P. P., Pop, M., Iordache, C. & Nicolopoulos, S. (2016). *Acta Cryst. A*. **72**, s107.

Gemmi, M. & Oleynikov, P. (2013). *Z. Krist.* **228**, 51–58.

Gemmi, M., La Placa, M. G. I., Galanis, A. S., Rauch, E. F. & Nicolopoulos, S. (2015). *J. Appl. Crystallogr.* **48**, 718–727.

van Genderen, E. (2015). Novel detectors and algorithms for electron nano-crystallography. Biophysical Structural Chemistry (BSC), Leiden Institute of Chemistiry (LIC), Faculty of Science, Leiden University.

van Genderen, E., Clabbers, M. T. B., Das, P. P., Stewart, A. A., Nederlof, I., Barentsen, K. C., Portillo, Q., Pannu, N. S., Nicolopoulos, S., Gruene, T. & Abrahams, J. P. (2016). *Acta Cryst. A*. **72**, 236–242.

van Genderen, E., Li, Y. W., Nederlof, I. & Abrahams, J. P. (2016). *Acta Crystallogr. Sect. D, Struct. Biol.* **72**, 34–39.

Georgieva, D. G., Jansen, J., Sikharulidze, I., Jiang, L., Zandbergen, H. W. & Abrahams, J. P. (2011). *J. Instrum.* **6**, C01033–C01033.

Georgieva, D. G., Kuil, M. E., Oosterkamp, T. H., Zandbergen, H. W. & Abrahams, J. P. (2007). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 564–570.

Gildea, R. J., Waterman, D. G., Parkhurst, J. M., Axford, D., Sutton, G., Stuart, D. I., Sauter, N. K., Evans, G. & Winter, G. (2014). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2652–2666.

Glaeser, R. M. (2016). *Nat Methods*. **13**, 28–32.

Glaeser, R. M. & Downing, K. H. (1993). *Ultramicroscopy*. **52**, 478–486.

Glaeser, R. M., Downing, K. H. & De Rosier, D. J. (2007). Electron crystallography of biological macromolecules Oxford: Oxford University Press.

Glaeser, R. M. & Han, B.-G. (2017). *Biophys. Reports*. **3**, 1–7.

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C. & Walz, T. (2005). *Nature*. **438**, 633–638.

Gorelik, T. E., van De Streek, J., Kilbinger, A. F. M., Brunklaus, G. & Kolb, U. (2012). *Acta Crystallogr. Sect. B Struct. Sci.* **68**, 171–181.

Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. (1996). *J. Mol. Biol.* **259**, 393–421.

Grimes, J. M., Hall, D. R., Ashton, A. W., Evans, G., Owen, R. L., Wagner, A., Mcauley, K. E., von Delft, F., Orville, A. M., Sorensen, T., Walsh, M. A., Ginn, H. M. & Stuart, D. I. (2018). *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 152–166.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. (2002). *J. Appl. Crystallogr.* **35**, 126–136.

Gruene, T., Hahn, H. W., Luebben, A. V., Meilleur, F. & Sheldrick, G. M. (2014). *J. Appl. Crystallogr.* **47**, 462–466.

Gruene, T., Li, T., van Genderen, E., Pinar, A. B. & van Bokhoven, J. A. (2018). *Chem. Eur. J.* **24**, 2384–2388.

Gruene, T., Sheldrick, G. M., Zlatopolskiy, B. D., Kozhushkov, S. I. & de Meijere, A. (2014). *Zeitschrift Für Naturforsch. B*. **69**,.

Guo, P., Shin, J., Greenaway, A. G., Min, J. G., Su, J., Choi, H. J., Liu, L., Cox, P. A., Hong, S. B., Wright, P. A. & Zou, X. (2015). *Nature*. **524**, 74–78.

Hasegawa, K., Hirata, K., Shimizu, T., Shimizu, N., Hikima, T., Baba, S., Kumasaka, T. & Yamamoto, M. (2009). *J. Appl. Crystallogr.* **42**, 1165–1175.

Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., de la Cruz, M. J., Leslie, A. G. W. & Gonen, T. (2015). *Acta Crystallogr. Sect. A Found. Adv.* **71**, 353–360.

Hattne, J., Shi, D., De La Cruz, M. J., Reyes, F. E. & Gonen, T. (2016). *J. Appl. Crystallogr.* **49**, 1029–1034.

Hedman, B., Hodgson, K. O., Helliwell, J. R., Liddington, R. & Papiz, M. Z. (1985). *Proc. Natl. Acad. Sci. U. S. A.* **82**, 7604–7607.

Heijne, E. H. M. (2001). *Nucl. Instruments Methods Phys. Res. Sect. A*. **465**, 1–26.

Helliwell, J. R. (1988). *J. Cryst. Growth*. **90**, 259–272.

Henderson, R. (1995). *Q. Rev. Biophys.* **28**, 171–193.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). *J. Mol. Biol.* **213**, 899–929.

Henderson, R. & Unwin, P. N. T. (1975). *J. Mol. Biol.* **94**, 425–440.

Holton, J. M. & Frankel, K. A. (2010). *Acta Cryst. D*. **66**, 393–408.

Hovmöller, S., Zou, X. & Weirich, T. E. (2002). *Adv. Imaging Electron Phys.* **123**, 257–289.

Hübschle, C. B., Sheldrick, G. M. & Dittrich, B. (2011). *J. Appl. Crystallogr.* **44**, 1281–1284.

Jansen, J., Tang, D., Zandbergen, H. W. & Schenk, H. (1998). *Acta Crystallogr. Sect. A*. **54**, 91–101.

Jiang, L., Georgieva, D. G., Zandbergen, H. W. & Abrahams, J. P. (2009). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 625–632.

Johnson, I., Bergamaschi, A., Buitenhuis, J., Dinapoli, R., Greiffenberg, D., Henrich, B., Ikonen, T., Meier, G., Menzel, A., Mozzanica, A., Radicci, V., Satapathy, D. K., Schmitt, B. & Shi, X. (2012). *J. Synchrotron Radiat.* **19**, 1001–1005.

Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. (2014). *IUCrJ*. **1**, 213–220.

Judge, R. A., Swift, K. & González, C. (2005). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 60–66.

Kabsch, W. (1988*a*). *J. Appl. Crystallogr.* **21**, 67–72.

Kabsch, W. (1988*b*). *J. Appl. Crystallogr.* **21**, 916–924.

Kabsch, W. (2010*a*). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 133–144.

Kabsch, W. (2010*b*). *Acta Cryst. D*. **66**, 125–132.

Karplus, P. A. & Diederichs, K. (2012). *Science (80-.)*. **336**, 1030–1033.

Keegan, R., Waterman, D. G., Hopper, D. J., Coates, L., Taylor, G., Guo, J., Coker, A. R., Erskine, P. T., Wood, S. P. & Cooper, J. B. (2016). *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 933–943.

Khasnis, M. D., Halkidis, K., Bhardwaj, A. & Root, M. J. (2016). *PLoS Pathog.* **12**, 1–29.

Kleywegt, G. J. (1996). *Acta Crystallogr. Sect. D Struct. Biol.* **52**, 842–857.

Kolb, U., Gorelik, T. E., Kübel, C., Otten, M. T. & Hubert, D. (2007). *Ultramicroscopy.* **107**, 507–513.

Kolb, U., Gorelik, T. E., Mugnaioli, E. & Stewart, A. (2010). *Polym. Rev.* **50**, 385–409.

Kolb, U., Gorelik, T. E. & Otten, M. T. (2008). *Ultramicroscopy.* **108**, 763–772.

Kolb, U., Gorelik, T. & Mugnaioli, E. (2009). *MRS Proc.* **1184**, 1184-NaN-5.

Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. (2018). *Acta Cryst. D.* **74**, 215–227.

Kuehlbrandt, W. (2014). *Science (80-.).* **343**, 1443–1444.

De La Cruz, M. J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat. Publ. Gr.* **14**, 399–402.

De La Mora, E., Carmichael, I. & Garman, E. F. (2011). *J. Synchrotron Radiat.* **18**, 346–357.

Leslie, A. G. W. (1999). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D55**, 1696–1702.

Leslie, A. G. W. (2006). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 48–57.

Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, Vol. edited by R.J. Read & J. Sussman, pp. 41–51. Dordrecht: Springer.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A. & Cheng, Y. (2013). *Nat. Methods.* **10**, 584–590.

Llopart, X., Ballabriga, R., Campbell, M., Tlustos, L. & Wong, W. (2007). *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **581**, 485–494.

Llopart, X., Campbell, M., Dinapoli, R., San Segundo, D. & Pernigotti, E. (2002). *IEEE Trans. Nucl. Sci.* **49 I**, 2279–2283.

Luebben, J. & Gruene, T. (2015). *Proc. Natl. Acad. Sci. U. S. A.* 201502136.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

McMullan, G., Cattermole, D. M., Chen, S., Henderson, R., Llopart, X., Summerfield, C., Tlustos, L. & Faruqi, A. R. (2007). *Ultramicroscopy.* **107**, 401–413.

McMullan, G., Chen, S., Henderson, R. & Faruqi, A. R. (2009). *Ultramicroscopy.* **109**, 1126–1143.

McMullan, G., Faruqi, A. R. & Henderson, R. (2016). *Methods Enzymol.* **579**, 1–17.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M. I., Pragani, R., Boxer, M. B., Earl, L. A., Milne, J. L. S. & Subramaniam, S. (2016). *Cell.* **165**, 1698–1707.

Midgley, P. A. & Eggeman, A. S. (2015). *IUCrJ.* **2**, 126–136.

Morris, R. J. & Bricogne, G. (2003). *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 615–617.

Mueller, M., Wang, M. & Schulze-Briese, C. (2012). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 42–56.

Mueller, P. (2009). *Crystallogr. Rev.* **15**, 57–83.

Mugnaioli, E., Gemmi, M., Merlini, M. & Gregorkiewitz, M. (2016). *Acta Crystallogr. Sect. B.* **72**, 893–903.

Mugnaioli, E., Gorelik, T. E. & Kolb, U. (2009). *Ultramicroscopy.* **109**, 758–765.

Mugnaioli, E., Gorelik, T. E., Stewart, A. & Kolb, U. (2011). *Minerals as Advanced*

Materials II, Vol. pp. 41–54. Springer Berlin Heidelberg.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367.

Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. (2014). *Elife.* **3**, e03600.

Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. (2014). *Nat. Methods.* **11**, 927–930.

Nave, C. & Hill, M. A. (2005). *J. Synchotron Rad.* **12**, 299–303.

Nederlof, I., van Genderen, E., Li, Y. W. & Abrahams, J. P. (2013). *Acta Cryst. D.* **69**, 1223–1230.

Nederlof, I., Hosseini, R., Georgieva, D. G., Luo, J., Li, D. & Abrahams, J. P. (2011). *Cryst. Growth Des.* **11**, 1170–1176.

Nederlof, I., Li, Y. W., Van Heel, M. & Abrahams, J. P. (2013). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 852–859.

Oatley, S. & French, S. (1982). *Acta Cryst. A.* **38**, 537–549.

Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4912–4917.

Le Page, Y. (1982). *J. Appl. Crystallogr.* **15**, 255–259.

Palatinus, L., Brázda, P., Boullay, P., Perez, O., Klementová, M., Petit, S., Eigner, V., Zaarour, M. & Mintova, S. (2017). *Science (80-.).* **355**, 166–169.

Palatinus, L., Corrêa, C. A., Steciuk, G., Jacob, D., Roussel, P., Boullay, P., Klementová, M., Gemmi, M., Kopeček, J., Domeneghetti, M. C., Cámara, F. & Petříček, V. (2015). *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **71**, 740–751.

Palatinus, L., Petříček, V. & Corrêa, C. A. (2015). *Acta Crystallogr. Sect. A Found. Adv.* **71**, 235–244.

Parkhurst, J. M., Brewster, A. S., Fuentes-Montero, L., Waterman, D. G., Hattne, J., Ashton, A. W., Echols, N., Evans, G., Sauter, N. K. & Winter, G. (2014). *J. Appl. Crystallogr.* **47**, 1459–1465.

Parkhurst, J. M., Thorn, A., Vollmar, M., Winter, G., Waterman, D. G., Fuentes-Montero, L., Gildea, R. J., Murshudov, G. N. & Evans, G. (2017). *IUCrJ.* **4**, 626–638.

Parkhurst, J. M., Winter, G., Waterman, D. G., Fuentes-Montero, L., Gildea, R. J., Murshudov, G. N. & Evans, G. (2016). *J. Appl. Crystallogr.* **49**, 1912–1921.

Peng, L. M. (1999). *Micron.* **30**, 625–648.

Pflugrath, J. W. (1999). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 1718–1725.

Raunser, S. & Walz, T. (2009). *Annu. Rev. Biophys.* **38**, 89–105.

Read, R. J. & McCoy, A. J. (2016). *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 375–387.

Reeke, G. N. (1984). *J. Appl. Crystallogr.* **17**, 238–243.

Rodriguez, J. A. & Gonen, T. (2016). *Methods Enzymol.* **579**, 369–392.

Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F. E., Shi, D., Sangwan, S., Guenther, E. L., Johnson, L. M., Zhang, M., Jiang, L., Arbing, M. A., Nannenga, B. L., Hattne, J., Whitelegge, J., Brewster, A. S., Messerschmidt, M., Boutet, S., Sauter, N. K., Gonen, T. & Eisenberg, D. S. (2015). *Nature.* **525**, 486–490.

De Rosier, D. J. & Klug, a. (1968). *Nature.* **217**, 130–134.

Sanishvili, R., Yoder, D. W., Babu, S., Rosenbaum, G., Xu, S. & Vogt, S. (2011). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6127–6132.

138

Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. (2006). *J. Appl. Crystallogr.* **39**, 158–168.

Sawaya, M. R., Rodriguez, J., Cascio, D., Collazo, M. J., Shi, D., Reyes, F. E., Hattne, J., Gonen, T. & Eisenberg, D. S. (2016). *Proc. Natl. Acad. Sci.* 201606287.

Schenk, A. D., Castaño-Díez, D., Gipson, B., Arheit, M., Zeng, X. & Stahlberg, H. (2010). *Methods Enzymol.* **482**, 101–129.

Schlichting, I. (2015). *IUCrJ.* **2**, 246–255.

Sharma, P., Verma, N., Singh, P. K., Korpole, S. & Ashish (2016). *Sci. Rep.* **6**, 22475.

Sheldrick, G. M. (1990). *Acta Crystallogr. Sect. A.* **46**, 467–473.

Sheldrick, G. M. (2008). *Acta Crystallogr. Sect. A Found. Crystallogr.* **64**, 112–122.

Sheldrick, G. M. (2015*a*). *Acta Cryst. C.* **71**, 3–8.

Sheldrick, G. M. (2015*b*). *Acta Cryst. A.* **71**, 3–8.

Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *Elife.* **2**, e01345.

Shi, D., Nannenga, B. L., de la Cruz, M. J., Liu, J., Sawtelle, S., Calero, G., Reyes, F. E., Hattne, J. & Gonen, T. (2016). *Nat. Protoc.* **11**, 895–904.

Spence, J. C. H. (2017). *IUCrJ.* **4**, 322–339.

Stahlberg, H., Biyani, N. & Engel, A. (2015). *Arch. Biochem. Biophys.* **581**, 68–77.

Stanton, M. (1993). *Nucl. Inst. Methods Phys. Res. A.* **325**, 550–557.

Stanton, M., Phillips, W. C., Li, Y. & Kalata, K. (1992). *J. Appl. Crystallogr.* **25**, 549–558.

Stevenson, H. P., Lin, G., Barnes, C. O., Sutkeviciute, I., Krzysiak, T., Weiss, S. C., Reynolds, S., Wu, Y., Nagarajan, V., Makhov, A. M., Lawrence, R., Lamm, E., Clark, L., Gardella, T. J., Hogue, B. G., Ogata, C. M., Ahn, J., Gronenborn, A. M., Conway, J. F., Vilardaga, J.-P., Cohen, A. E. & Calero, G. (2016). *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 603–615.

Stevenson, H. P., Makhov, A. M., Calero, M., Edwards, A. L., Zeldin, O. B., Mathews, I. I., Lin, G., Barnes, C. O., Santamaria, H., Ross, T. M., Soltis, S. M., Khosla, C., Nagarajan, V., Conway, J. F., Cohen, A. E. & Calero, G. (2014). *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8470–8475.

Subramanian, G., Basu, S., Liu, H., Zuo, J. M. & Spence, J. C. H. (2015). *Ultramicroscopy.* **148**, 87–93.

Terwilliger, T. C., Stuart, D. & Yokoyama, S. (2009). *Annu. Rev. Biophys.* **38**, 371–383.

Thorn, A., Dittrich, B. & Sheldrick, G. M. (2012). *Acta Cryst. A.* **68**, 448–451.

Thorn, A. & Sheldrick, G. M. (2013). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 2251–2256.

Tinti, G., Frojdh, E., Van Genderen, E., Gruene, T., Schmitt, B., Matthijs De Winter, D. A., Weckhuysen, B. M. & Abrahams, J. P. (2018). *IUCrJ.* **5**, 190–199.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Vainshtein, B. K. (1964). Structure analysis by electron diffraction. Oxford: Pergamon.

Valpuesta, J. M., Carrascosa, J. L. & Henderson, R. (1994). *J. Mol. Biol.* **240**, 281–287.

Vincent, R. & Midgley, P. A. (1994). *Ultramicroscopy.* **53**, 271–282.

Visser, J., Heijden, B. Van Der, Weijers, S. J. A., Vries, R. De & Visschers, J. L. (2011). *Nucl. Inst. Methods Phys. Res. A.* **633**, S22–S25.

Wampler, R. D., Kissick, D. J., Dehen, C. J., Gualtieri, E. J., Grey, J. L., Wang, H.-F., Thompson, D. H., Cheng, J.-X. & Simpson, G. J. (2008). *J. Am. Chem. Soc.* **130**, 14076–14077.

Wan, W., Hovmöller, S. & Zou, X. (2012). *Ultramicroscopy*. **115**, 50–60.

Wan, W., Sun, J., Su, J., Hovmöller, S. & Zou, X. (2013). *J. Appl. Crystallogr.* **46**, 1863–1873.

Wang, Y., Takki, S., Cheung, O., Xu, H., Wan, W., Oehstroem, L. & Ken Inge, A. (2017). *Chem. Commun.* **53**, 7018–7021.

Waterman, D. & Evans, G. (2010). *J. Appl. Crystallogr.* **43**, 1356–1371.

Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K. & Evans, G. (2016). *Acta Crystallogr. Sect. D, Struct. Biol.* **72**, 558–575.

Weirich, T. E., Ramlau, R., Simon, A., Hovmöller, S. & Zou, X. (1996). *Nature*. **382**, 144–146.

Weirich, T. E., Zou, X., Ramlau, R., Simon, A., Cascarano, G. L., Giacovazzo, C. & Hovmöller, S. (2000). *Acta Crystallogr. Sect. A Found. Crystallogr.* **56**, 29–35.

White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. a, Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst. D*. **69**, 1231–1240.

Williams, D. B. & Carter, C. B. (2009). Transmission Electron Microscopy: A Textbook for Materials Science New York: Springer.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A. J., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. A. & Wilson, K. S. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242.

Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 85–97.

Wisedchaisri, G. & Gonen, T. (2011). *Structure*. **19**, 976–987.

Woińska, M., Grabowsky, S., Dominiak, P. M., Woźniak, K. & Jayatilaka, D. (2016). *Sci. Adv.* **2**,.

Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. (2015). *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3368–3373.

Yonekura, K. & Maki-Yonekura, S. (2016). *J. Appl. Cryst.* **49**, 1517–1523.

Yonekura, K., Maki-Yonekura, S. & Namba, K. (2002). *Biophys. J.* **82**, 2784–2797.

Yonekura, K., Matsuoka, R., Yamashita, Y., Yamane, T., Kidera, A. & Maki-yonekura, S. (2018). *IUCrJ*. **5**, 348–353.

Yun, Y., Zou, X., Hovmöller, S. & Wan, W. (2015). *IUCrJ*. **2**, 267–282.

Zhang, D., Oleynikov, P., Hovmöller, S. & Zou, X. (2010). *Zeitschrift Fur Krist*. **225**, 94–102.

Zou, X., Hovmöller, A. & Hovmöller, S. (2004). *Ultramicroscopy*. **98**, 187–193.

Zou, X., Hovmöller, S. & Oleynikov, P. (2011). Electron crystallography. Oxford: Oxford University Press.

ACKNOWLEDGEMENTS

CURRICULUM VITAE

Personal information

Address: Center for Cellular Imaging and NanoAnalytics (C-CINA)
     Biozentrum, University of Basel
     Mattenstrasse 26, 4058 Basel, Switzerland
Office phone: +41 61 387 3232
Mobile phone: +31 65 201 3449
E-mail: max.clabbers@unibas.ch, clabbersm@gmail.com

Birthplace/date: Delft, the Netherlands, 07.02.1987
Nationality: Dutch

Education

2015-2018 PhD, *summa cum laude*
     Biozentrum, University of Basel, Switzerland
     Thesis title**:** *Developing electron crystallography for elucidation of organics and macromolecules.*
     Advisors: Prof. Dr. J.P. Abrahams, Prof. Dr. H. Stahlberg, Dr. T. Gruene

2013-2015 Doctoral student (relocated to University of Basel)
     Leiden Institute of Chemistry, Leiden University, the Netherlands
     Thesis title: *Developing protein nanocrystallography for structural biology.*
     Advisor: Prof. Dr. J.P. Abrahams

2011-2013 MSc, Leiden University, the Netherlands
     Thesis title*: A bioinformatic search for conserved RNA structures.*
     Advisors: Dr. A.P. Gultyaev, Dr. R.C.L. Olsthoorn

2005-2013 BSc, Delft University of Technology, the Netherlands

Publications

- **Clabbers, M.T.B**., T. Gruene, E. van Genderen & J.P. Abrahams. Reducing dynamical electron scattering reveals hydrogen atoms. Manuscript accepted for publication in *Acta Cryst.* A (2018)

- **Clabbers, M.T.B**., T. Gruene, J.M. Parkhurst, J.P. Abrahams & D.G. Waterman. Electron diffraction data processing with *DIALS*. *Acta Cryst.* D74: 506-518 (2018)

- **Clabbers, M.T.B**. & J.P. Abrahams. Electron diffraction and three-dimensional crystallography for structural biology. *Cryst. Rev.* 24: 176-204 (2018)

- Thomas, B., R.K. Dubey, **M.T.B. Clabbers**, K.B.S.S. Gupta, E. van Genderen, W.F. Jager, J.P. Abrahams, E.J.R. Sudholter & H.J.M. de Groot. A molecular level approach to elucidate the supramolecular packing of light-harvesting antenna systems. *Chem. Eur. J.* 24: 1-6 (2018)

- **Clabbers, M.T.B**., E. van Genderen, W. Wan, E. Wiegers, T. Gruene & J.P. Abrahams. Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Cryst.* D73: 738-748 (2017)

- van Genderen, E., **M.T.B. Clabbers**, P. Pratim Das, A. Stewart, I. Nederlof, K.C. Barentsen, J. Portillo, N.S. Pannu, S. Nicolopoulos, T. Gruene & J.P. Abrahams. *Ab initio* structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector. *Acta Cryst.* A72: 236-242 (2016)

- **Clabbers, M.T.B.**, R.C.L. Olsthoorn & A.P. Gultyaev. Tospovirus ambisense genomic RNA segments use almost complete repertoire of stable tetraloops in the intergenic region. *Bioinformatics* 30(13): 1800-1804 (2014)