Chapter 1

Introduction

1.1 Macromolecular X-ray crystallography

The discovery of X-ray diffraction by crystals by Max van Laue [1, 2] marked the origins of modern crystallography. However, it was not until the work of William Lawrence Bragg and William Henry Bragg that X-ray scattering could be translated into atomic positions [3–5]. Since then, X-ray crystallography and the determination of atomic positions in organic and inorganic molecules of has come a long way and shaped the path for many 21st century discoveries. Amongst those ground-breaking discoveries are the earliest structural models of biological molecules including DNA [6], vitamin B12 [7], and the first protein structures [8–11]. These structure elucidations hallmarked the dawn of a new era in biological and biomedical research. At the time of writing, 124,551 structural models were determined by X-ray diffraction studies [12], and thus X-ray crystallography is a key method in biological research.

1.1.1 X-ray scattering

X-rays are high energy photons part of the electromagnetic spectrum with a wavelength 0.1-100Å [13]. X-rays can be described as packets of travelling electromagnetic waves, whose electric field vector interacts with the charged electrons of matter [13]. Such interaction, typically termed scattering, results in the diffraction of the incoming wave, which

X-ray crystallography relies on.

In its simplest form, scattering of X-ray radiation can be explained in the scenario of exposure to a single free electron. The resulting scattering can be classed as elastic (Thomson scattering) or inelastic (Crompton scattering) [13]. The latter — scattering that results in a loss of energy of the emitting photon due to energy transfer onto the electron — does not contribute to discrete scattering, the type of scattering X-ray diffraction relies on. In comparison, Thomson scattering does not result in a loss of energy of the emitting photon. This has significant effects, the incoming photon emits with the same frequency causing the electron to oscillate identically further enhancing the signal.

If we expand the example to include all electrons in an atom and expose the atom to X-ray radiation, our theory needs to be slightly expanded. Given that one or more electrons in an atom are not free but orbit around the atom's nucleus in a stable and defined manner, the distribution of these electrons around the nucleus determines the scattering of the incoming X-ray photons. The distribution of scattered photon waves is thus an overall representation of the probability distributions of each electron in the atom and is refered to as electron density $\rho(\boldsymbol{r})$. In X-ray scattering, it suffices to approximate the shape of the electron density to a sphere. If we now consider the emitting wave $\boldsymbol{s_1}$ of an X-ray photon scattered by any position $\boldsymbol{r}$ in the electron density of an atom, then the phase difference $\Delta\varphi$ to the incoming wave $\boldsymbol{s_0}$ can be described by Eq. (1.1) [13].

$$\Delta\varphi = 2\pi\left(\boldsymbol{s_1} - \boldsymbol{s_0}\right)\boldsymbol{r} = 2\pi \cdot \boldsymbol{Sr} \tag{1.1}$$

If more than one electron in an atom's electron density scatter the incoming X-ray wave, then the emitting partial waves can be described by the atomic scattering function f_s (Eq. (1.2)), which describes the interference of all scattered waves [13]. The total scattering power of an atom is proportional to the number of electrons and element-specific with heavier atoms scattering more strongly. Given the approximation of a centrosymmetric electron density, the atomic scattering function is also symmetric.

$$f_s = \int\limits_{\boldsymbol{r}}^{V(atoms)} \rho\left(\boldsymbol{r}\right) \cdot e^{2\pi i \boldsymbol{S}\boldsymbol{r}} \cdot d\boldsymbol{r} \qquad (1.2)$$

With an enhanced understanding of X-ray scattering of electrons orbiting a single atom, it is important to consider X-ray scattering of adjacent atoms, such as it is typically found in molecules. If the electromagnetic wave of a X-ray photon excites all electrons of adjacent atoms, then the resulting partial waves result in constructive or destructive interference. Maximal interference can be obtained when all partial waves are in phase, and maximal destructive interference when out-of-phase. This leads to varying intensities of the emitting X-ray photon at different points in space. To obtain the overall scattering power F_s of all contributing atoms, Eq. (1.2) needs to be modified to include the sum over all atoms j as described in Eq. (1.3).

$$F_s = \sum_{j=1}^{atoms} f_{s,j}^0 \cdot e^{2\pi i \boldsymbol{S}\boldsymbol{r}_j} \qquad (1.3)$$

If we now translate our hypothetical experiment into a crystal lattice then our understanding described in Eq. (1.3) needs to be expanded from a 1-dimensional distance vector $\boldsymbol{r}$ to the three dimensional lattice translation vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$. The Laue equations (Eq. (1.4)) do exactly that and ultimately determine the positions of the diffraction peaks in 3-dimensional space.

$$\boldsymbol{S} \cdot \boldsymbol{a} = n_1, \qquad \boldsymbol{S} \cdot \boldsymbol{b} = n_2, \qquad \boldsymbol{S} \cdot \boldsymbol{c} = n_3 \qquad (1.4)$$

$$n\lambda = 2d_{hkl}sin\theta \qquad (1.5)$$

Such determination is possible through the findings made by Bragg and Bragg [3], who identified the relationship between the scattering vector $\boldsymbol{S}$ and the planes in the crystal lattice. Today, this relationship is defined by the Bragg equation (Eq. (1.5)) [3], which allows us to interpret X-ray diffraction as reflections on discrete lattice planes, which relates the diffraction angle θ to the lattice spacing d_{hkl} (Fig. 1.1) [13]. For maximum

diffraction n needs to be integer multiples to result in maximum constructive interference of wavelength λ.
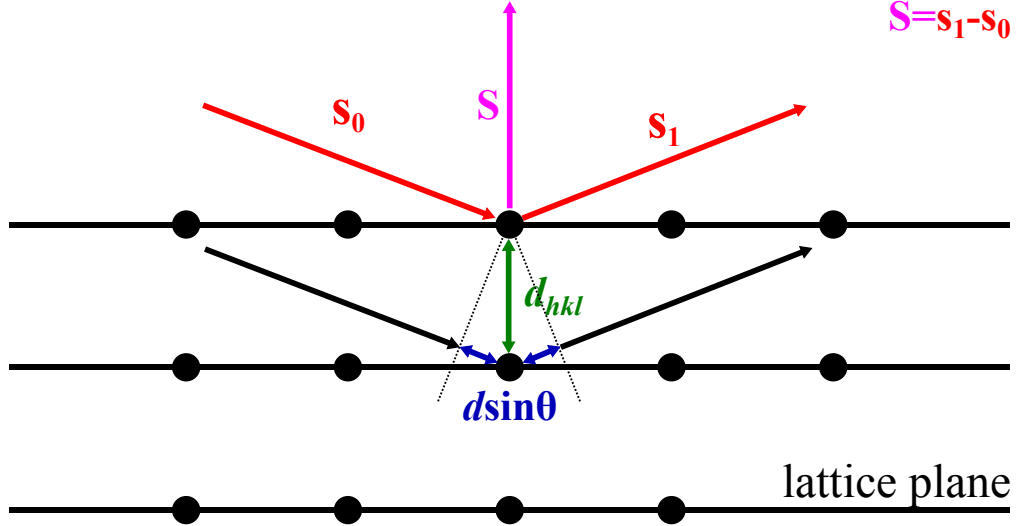

Figure 1.1: Schematic of Bragg scattering.

Lastly, if the hypothetical model is expanded to molecular crystals, then the total scattering from the unit cell is merely a summation of all molecular unit cell scattering contributions in the crystal. Mathematically, this results in Eq. (1.3) being generalised to Eq. (1.6) through the application of the Laue equations (Eq. (1.4)) to express the scattering vector $\boldsymbol{S r}_j$ as Miller indices of the reflection planes $\boldsymbol{h x}_j$.

$$F_h = \sum_{j=1}^{atoms} f_{s,j}^0 \cdot e^{2\pi i \boldsymbol{h x}_j} \tag{1.6}$$

The structure factor equation defines the scattering power from a crystal in a given reciprocal lattice direction $\boldsymbol{h}$. The scattering is enhanced by the number of repeating units of lattice translation vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$, and thus the overall scattering power is proportional to the number of unit cells in the crystal.

It should be noted that Eq. (1.6) is a simplification of the problem at hand. In reality, instrument and experimental corrections need to be applied to the structure factor equation. A correction factor for each experiment-dependent parameter needs to be applied

to the structure factor equation. However, in the scope of this work the details of such correction factors do not need to be discussed.

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=0}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \boldsymbol{F}(hkl) \cdot e^{-2\pi i(hx+ky+lz)} \tag{1.7}$$

Since complex structure factors describe the molecular structure in the reciprocal space domain, the conversion to the real space domain in form of electron density is required. This can be conveniently done through the bijective Fourier transform, which allows to convert complex structure factors to electron density and vice versa without the loss of any information [13]. Thus, electron density can be obtained from the complex structure factors using Eq. (1.7). The normalisation factor $1/V$ provides the correct units for the electron density $\rho(x,y,z)$.

1.1.2 Structure determination

In X-ray crystallographic experiments, X-ray radiation is measured using light detectors. However, the measurement taken is incomplete. Light detectors only capture the intensity of the scattered X-ray photons but crucially lose the phase information. The latter is essential for atomic reconstruction of the molecule in the crystal, and thus needs to be obtained. In Macromolecular Crystallography (MX), experimentalists have a number of alternative techniques to compensate for the lost phase information.

The most popular method to recovering the lost phase information is Molecular Replacement (MR) [14, 15]. In a MR search, a known structure ('search model') similar to the unknown is relocated in the unit cell until the solution with the best fit between calculated and observed diffraction data is obtained [13]. A 6-dimensional search, i.e. a simultaneous rotation and translation search, is possible [16–18], however computationally very expensive and less suitable for challenging cases. In comparison, most modern crystalloraphic applications opt for two distinct sub-searches, the rotation search to orient the search model within the unit cell followed by the translation search to locate it [13]. The benefits over a combined search include search-specific target functions that enable increased sensitivity and additional terms to compensate for imperfect data.

The most successful MR algorithms perform the rotation and translation searches using Patterson methods or Maximum Likelihood functions. Patterson methods — originally developed by Rossmann and Blow [19] — rely on the use of a map of vectors between the scattering atoms, which can be determined for the calculated and observed structure factor amplitudes. Patterson vectors can be sub-classed as intra- and inter-molecular vectors. A distinct separation of the observed vectors is impossible. However, inter-molecular vectors appear further away from the central peak of self-vector (vector from atom to itself) in the Patterson map [13]. The calculated Patterson vectors for the search model allow for a clearer distinction between the intra- and inter-molecular vectors. If the search model is placed in a large unit cell, then inter-molecular vectors must scale with the unit cell dimension [13]. Ultimately, using the intra-molecular Patterson vectors, the search probe can be oriented against the experimentally determined Patterson vectors. Similarly, the inter-molecular vectors can be used to identify the correct translation of the search probe. Patterson methods are very sensitive to small orientation errors of the search probe [13]. Thus, orientations with the highest vector peak overlaps are trialed in the subsequence translation search.

In comparison to the Patterson methods, Maximum Likelihood methods do not rely on inter-atomic vectors in Patterson maps. Instead, Maximum Likelihood methods make use of Bayes' theorem [20] to compare calculated structure factors and observed structure factor amplitudes directly [18]. Bayes' theorem in crystallographic Maximum Likelihood methods is applied to compute the likelihood that an experimental value is observed given the current search model. The maximal likelihood indicates the best search model given the observed experimental data. Since the search model likelihood term is the product of many individual probabilities, which are difficult to represent computationally due to floating point representations, the log of the likelihood is commonly used [13]. The major advantage of Maximum likelihood methods over Patterson methods centres on the more realist target functions, which consider errors and incompleteness of the search model, applies bulk solvent correction and conducts multi-model searches [18]. The latter is of particularl relevance since the Maximum likelihood rotation function can thus consider already placed search model probes in a fixed position whilst trialling additional ones [21], which proves to be a major advantage over Patterson methods. Furthermore, likelihood

target functions consider the structural variance of multiple superposed models in an ensemble search model, which is used to weight structure factors at the various positions to improve the overall likelihood term [18].

Bibliography

[1]   W Friedrich, P Knipping, M Laue, *Ann. Phys.* **1913**, *346*, 971–988.

[2]   M Laue, *Ann. Phys.* **1913**, *346*, 989–1002.

[3]   W. H. Bragg, W. L. Bragg, *Proceedings of the Royal Society A: Mathematical Physical and Engineering Sciences* **July 1913**, *88*, 428–438.

[4]   W. L. Bragg, *Scientia* **1929**, *23*, 153.

[5]   W. L. Bragg, *Nature* **Dec. 1912**, *90*, 410.

[6]   J. D. Watson, F. H. C. Crick, Others, *Nature* **1953**, *171*, 737–738.

[7]   D. C. Hodgkin, J Kamper, M Mackay, J Pickworth, K. N. Trueblood, J. G. White, en, *Nature* **July 1956**, *178*, 64–66.

[8]   T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola, M Vijayan, en, *Nature* **June 1971**, *231*, 506–511.

[9]   C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, V. R. Sarma, en, *Nature* **May 1965**, *206*, 757–761.

[10] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H Muirhead, G Will, A. C. North, en, *Nature* **Feb. 1960**, *185*, 416–422.

[11] J. C. Kendrew, G Bodo, H. M. Dintzis, R. G. Parrish, H Wyckoff, D. C. Phillips, en, *Nature* **Mar. 1958**, *181*, 662–666.

[12] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **Jan. 2000**, *28*, 235–242.

[13] B. Rupp, *Biomolecular crystallography : principles, practice, and application to structural biology*, English, Garland Science, New York, **2010**.

[14] M. G. Rossmann, *Acta Crystallogr. D Biol. Crystallogr.* **2001**, *57*, 1360–1366.

[15] M. G. Rossmann, en, *Acta Crystallogr. A* **Feb. 1990**, *46 (Pt 2)*, 73–82.

[16] C. R. Kissinger, D. K. Gehlhaar, D. B. Fogel, en, *Acta Crystallogr. D Biol. Crystallogr.* **Feb. 1999**, *55*, 484–491.

[17] N. M. Glykos, M Kokkinidis, en, *Acta Crystallogr. D Biol. Crystallogr.* **Feb. 2000**, *56*, 169–174.

[18] R. J. Read, en, *Acta Crystallogr. D Biol. Crystallogr.* **Oct. 2001**, *57*, 1373–1382.

[19] M. G. Rossmann, D. M. Blow, *Acta Crystallogr.* **Jan. 1962**, *15*, 24–31.

[20] M. Bayes, M. Price, *Philosophical Transactions of the Royal Society of London* **Jan. 1763**, *53*, 370–418.

[21] L. C. Storoni, A. J. McCoy, R. J. Read, en, *Acta Crystallogr. D Biol. Crystallogr.* **Mar. 2004**, *60*, 432–438.