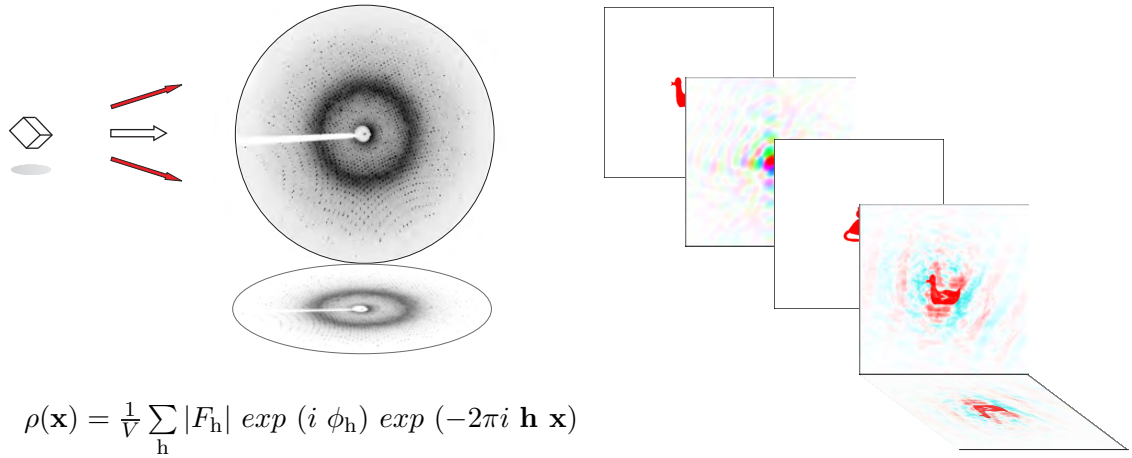
Chapter 1

From experiment to structure



$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathrm{h}} |F_{\mathrm{h}}| \; exp\; (i\; \phi_{\mathrm{h}})\; exp\; (-2\pi i\; \mathbf{h}\; \mathbf{x})$$

Summary

X-ray crystallography is a powerful method to determine the three-dimensional structure of matter. Nowadays it is routinely applied to elucidate the structures of large biological macromolecules such as proteins or nucleic acids.

In the experiment a single crystal is irradiated with X-ray radiation. Position and intensity of the reflection spots are recorded in a diffraction image. Since only the intensity of the scattered waves can be measured, the phase information for each reflection is lost, which is required for the reconstruction of the electron density. This is known as the phase problem in crystallography.

Several methods are applied to solve the phase problem. Molecular replacement and experimental phasing are used in macromolecular crystallography.

1.1 X-ray diffraction

X-ray crystallography is a powerful method to determine three-dimensional structures; it can be used to determine the structure of organic compounds, whether derived from natural sources or chemically synthesised, and organometallic or inorganic compounds (Schenk, 2008).

X-ray crystallography is essential for the structure elucidation of large biological macromolecules such as proteins or nucleic acids. It provides detailed structural models that form the basis of molecular biology. Based on these models biomolecular function such as enzyme catalysis can be understood (Daniel *et al.*, 2003). It is also important for the growing field of structure guided drug design to have a detailed picture of the structure and interaction of interest (Blundell *et al.*, 2002).

For all these different applications the basic experimental setup is very similar (Fig. 1.1). The first step is to obtain a single crystal from the sample material, a task that can be very challenging, especially for biological macromolecules. The crystal is irradiated with monochromated X-ray radiation, typically in the range of 0.5 – 2.5 Å ($\approx$25 000 – 5000 eV). The scattered radiation is recorded at the detector. From the diffraction image as depicted in Fig. 1.1 information can be extracted about the spot positions and their intensity.

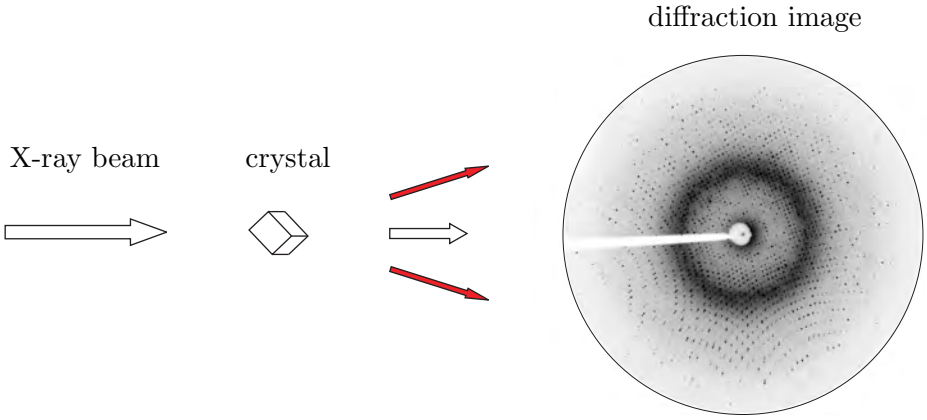


Figure 1.1: Schematic representation of a single-crystal X-ray diffraction experiment. The X-ray beam hits the crystal and interacts with the electrons of the atoms present in the crystal. The primary beam is blocked, only the diffracted radiation (depicted by red arrows) is recorded at the detector. From the diffraction image (right) information can be extracted about the spot positions and their intensity.

The diffraction pattern is caused by interaction of the X-ray beam with the crystal sample. X-ray photons impinge on the electrons of the atoms in the sample crystal. The electrons are induced to oscillate by the electric field vector of the photons. The oscillating electrons re-emit radiation at the same frequency as the incoming photons. This elastic scattering is also known as *Thompson* scattering.

The scattering from a single atom depends on the electron distribution around the nucleus. For practical purposes, the atomic scattering factor f is approximated by a nine-parameter Gaussian summation. The nine parameters, the *Cromer-Mann* coefficients (Cromer and Mann, 1968), are tabulated for each element (Prince, 2004).

The atoms in the crystal form a periodic lattice. They are not fixed to rigid positions, but can vibrate, depending on the chemical environment and the temperature of the sample. This displacement will result in a phase difference in the scattered radiation, attenuating the scattered intensity. The *Debye-Waller* factor B gives the displacement from the mean position. Therefore the isotropic atomic scattering factor has the form:

$$f^B = f \; exp \; (-B \; (\sin\theta/\lambda)^2) \tag{1.1}$$

with f as the scattering of an atom at rest. The scattering from a single atom or even a molecule is extremely weak. In the periodic crystal lattice, the small contribution from each molecule is amplified. The scattered waves from single atoms interfere either constructively or destructively, called diffraction. X-ray diffraction can be simplified to a reflection of the X-rays at lattice planes that pass through the atoms in the crystal. The *Bragg* equation (Eq. 1.2) illustrates the scattering as a reflection at a set of planes, depending on wavelength λ and the scattering angle θ. The *Miller* indices h, k, l denote a set of planes $\mathbf{h}$ with equidistant spacing d_{hkl}.

$$n \; \lambda = 2 \; d_{hkl} \; \sin\theta \tag{1.2}$$

The maximum resolution for a data set corresponds to the distance between a set of of planes d_{hkl} with the smallest value for that data set.

The scattering from a set of planes $\mathbf{h}$ in the crystal corresponds to a discrete spot on the detector, the reflection $\mathbf{h}$ (index vector of reflection h, k, l). The scattered wave that corresponds to that reflection $\mathbf{h}$ may be expressed as the structure factor $\mathbf{F}_\mathrm{h}$, a complex number depending on all atoms j, their atomic scattering factor f_j^B and their position $\mathbf{x}_j$ (in fractional coordinates) in the unit cell:

$$\mathbf{F}_\mathrm{h} = \sum_{j=1}^{atoms} f_j^B \; exp \; (2\pi i \; \mathbf{h} \; \mathbf{x}_j) \tag{1.3}$$

The complex structure factor $\mathbf{F}_\mathrm{h}$ may be separated in the amplitude term $|F_\mathrm{h}|$ and the phase angle term $exp \; (i \; \phi_\mathrm{h})$:

$$\mathbf{F}_\mathrm{h} = |F_\mathrm{h}| \; exp \; (i \; \phi_\mathrm{h}) \tag{1.4}$$

In the complex plane, the amplitude $|F_\mathrm{h}|$ corresponds to the magnitude of the vector $\mathbf{F}_\mathrm{h}$. Its direction is given by the phase angle ϕ_h.

1.2 The phase problem in crystallography

To reconstruct the electron density in the unit cell of the crystal, a Fourier transformation from reciprocal space to real space has to be carried out.

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_\mathrm{h} |F_\mathrm{h}| \; exp \; (i \; \phi_\mathrm{h}) \; exp \; (-2\pi i \; \mathbf{h} \; \mathbf{x}) \tag{1.5}$$

The intensity of the reflections (measured spots on the detector) is proportional to the square of the structure factor amplitude.

$$I_{\mathrm{h}} \propto |F_{\mathrm{h}}|^2 \tag{1.6}$$

Therefore, the first term in Eq. 1.5 can be evaluated. The third term is part of the Fourier transformation to real space. However, the second term, the phase angle term *exp* ($i\ \phi_{\mathrm{h}}$), cannot be determined since the phase information of the reflections is lost in the experiment. This is known as the phase problem in crystallography.

The importance of the phase information for the reconstruction of real space can readily be illustrated by Kevin Cowtan's famous duck and cat thought experiment (Fig. 1.2).

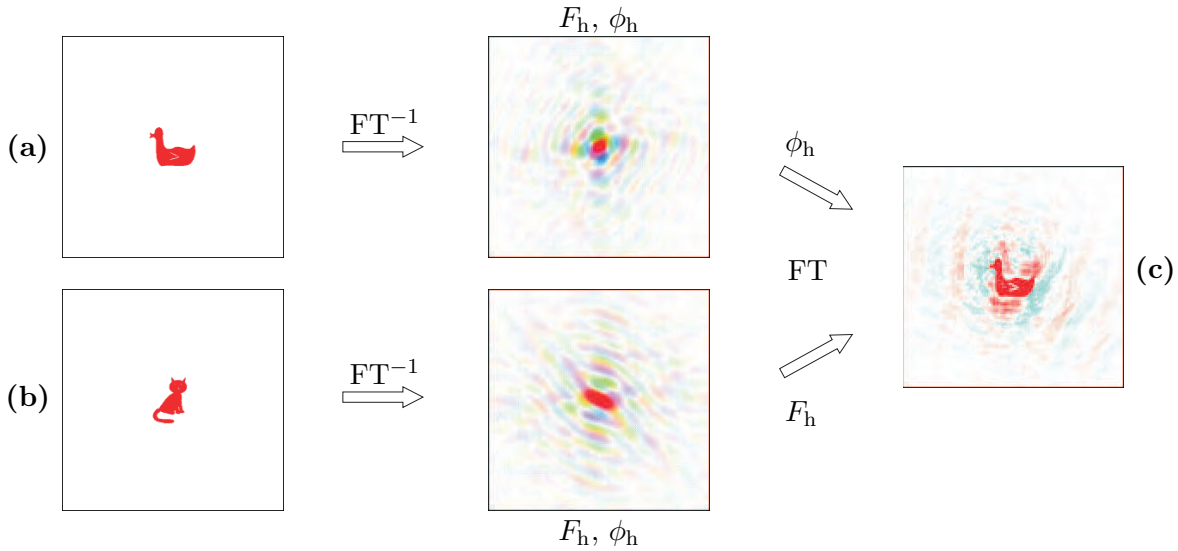


Figure 1.2: The importance of the phase information: **(a)** The image of a duck is transformed *via* inverse Fourier transformation. The diffraction image of a single duck is obtained. The colour (hue) corresponds to the phase information, the colour saturation represents the structure factor amplitude. **(b)** The image of a cat is transformed *via* inverse Fourier transformation. The diffraction image of a single cat is obtained. **(c)** The phase information from the duck (the colour hue) is combined with the structure factor amplitude information from the cat (the colour saturation). A Fourier transformation back to real space is carried out to yield the final image. This resembles clearly the duck image in a), showing the importance of the phase information (Cowtan, 2010).

It is obvious that the phases contain more information than the structure factor amplitudes. Therefore it is really important to determine not only the structure factor amplitude but also to derive the phases for a data set.

Different techniques are used to derive the phase information. *Ab initio* methods, which include **direct methods** and Patterson methods, are used in small-molecule crystallography. They require atomic resolution ($d \leq 1.2$ Å) and work only for small to medium size structures ($\approx$ 1000 light atoms).[1] In macromolecular crystallography the resolution of the data generally does not suffice to carry out *ab initio* phasing. Additionally, the size of the molecules limits the phasing techniques to molecular replacement or experimental phasing.

[1] The presence of heavier elements, e.g. iron, allows structures consisting of up to 2000 atoms to be solved.

If a similar structure is already known and has been deposited with the Protein Data Bank PDB (Bernstein *et al.*, 1977), it may be used as a search model for **molecular replacement**. This method suffers considerably from model bias and requires sufficient structural homology between the search model and target structure.

Methods based on **experimental phasing** are required for *de novo* phasing of proteins and nucleic acids and usually involve the search for heavy atoms. This is discussed in detail in the next chapter.

Once the phase problem is solved, i.e. each reflection has been assigned a phase angle, the electron density is calculated according to Eq. 1.5. The result is the distribution of the electrons in the crystal sample that were diffracting the X-ray radiation in the first place and gave rise to the diffraction pattern. The electron density is then interpreted in subsequent steps. Based on prior information, e.g. the sequence of amino acids or previously known values of chemical bonds, a model is fitted in the electron density and refined against the experimental data (for more details on crystallographic parameters, see Appendix B). After validation, this model can be used to answer chemical and biological questions.