Sticky triangles: New tools for experimental phasing of biological macromolecules

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

vorgelegt von
Diplom-Chemiker
Tobias Beck
aus Hannover

Göttingen 2010

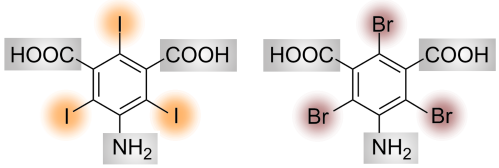
D7
Referent: Prof. George M. Sheldrick, Ph.D.
Korreferent: Prof. Dr. Oliver Einsle
Tag der mündlichen Prüfung: 16. September 2010

Table of contents

Scope of this thesis

X-ray crystallography is the prime method for the elucidation of three-dimensional molecular structures. It enables structure determination of biological macromolecules such as proteins and nucleic acids. These detailed structural models form the basis of molecular biology. The determination of novel macromolecular structures usually requires the incorporation of heavy atoms, which are exploited for experimental phasing. Conventional heavy-atom derivatisation with heavy-metal salts often suffers from non-specific binding, resulting in low occupancy of the heavy-atom sites or derivatisation failing completely. In addition many such soaks require the use of toxic chemicals and therefore stringent safety precautions.

We have developed a new class of compounds that combines heavy atoms for experimental phasing with functional groups for interaction with biological macromolecules. The lead structure is based on a benzene ring that provides a rigid scaffold. The ring is substituted with three functional groups and three heavy atoms, iodine or bromine, respectively.



The three functional groups, e.g. carboxyl or amino groups, but also hydroxyl or methoxy moieties, may interact with protein residues and therefore show enhanced binding properties compared with traditional heavy-atom compounds. The three halogen atoms provide a strong anomalous signal and may be used for experimental phasing *via* the techniques SAD and SIRAS (iodine) or SAD, SIRAS and MAD (bromine). The three heavy atoms form an equilateral triangle, which is easily recognised in the heavy-atom substructure.

Since compounds similar to the ones shown above have been used as contrast agents in medical imaging (Yu and Watson, 1999), a low toxicity is expected for phasing tools based on halogen atoms.

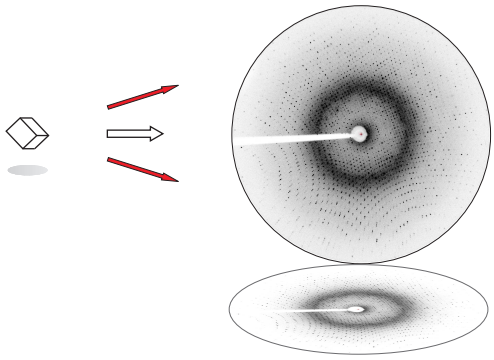
The present work is positioned at the interface of chemistry and structural biology and spans the design, synthesis, characterisation and utilisation of the sticky compounds; the compounds were used for experimental phasing of biological macromolecules, including test proteins and novel structures. This work combines methods from small-molecule crystallography and macromolecular crystallography.

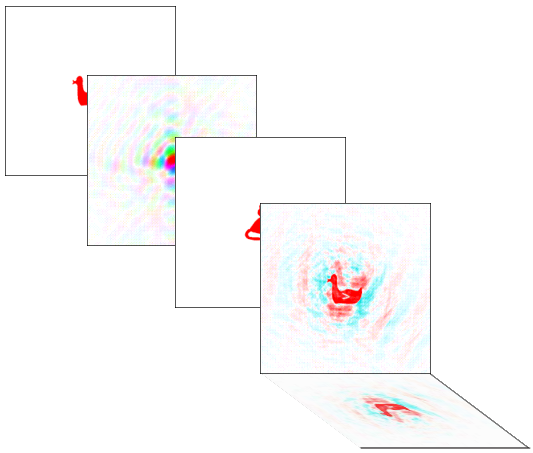
Part I

Crystallographic background

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.

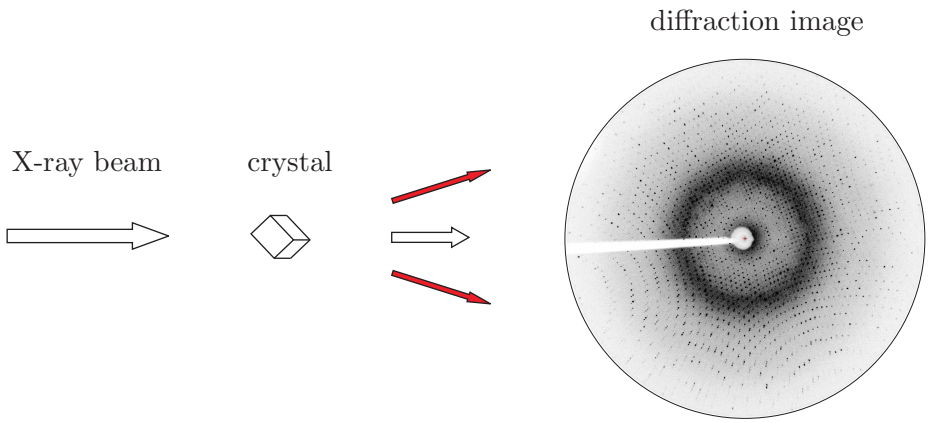


Figure 1.1: Schematic representation of a single-crystal X-ray diffraction experiment. The X-ray beam hits the crystal and interacts with the electrons of the atoms present in the crystal. The primary beam is blocked, only the diffracted radiation (depicted by red arrows) is recorded at the detector. 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).

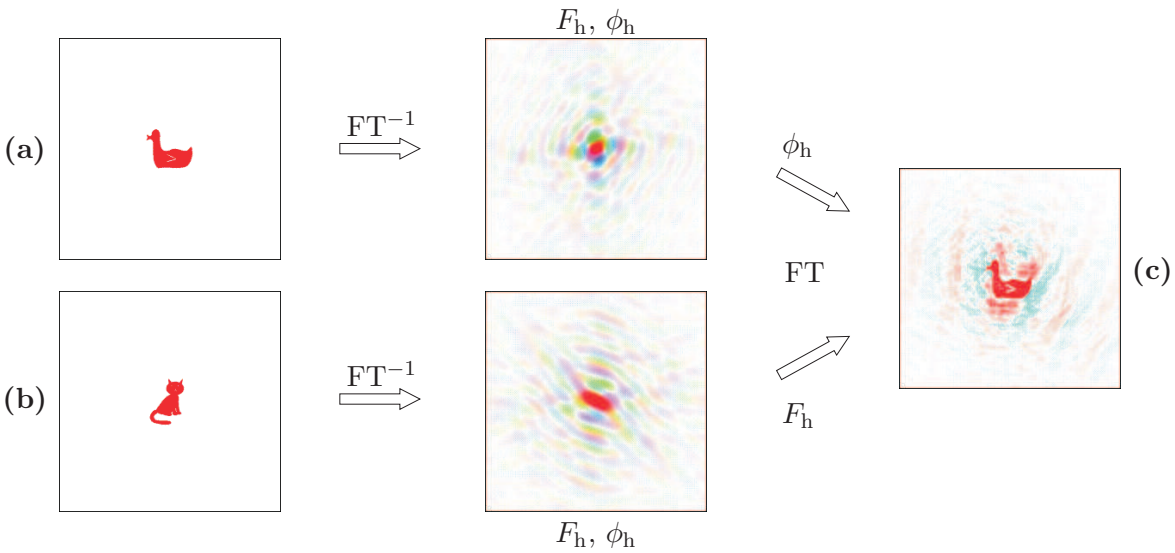


Figure 1.2: The importance of the phase information: **(a)** The image of a duck is transformed *via* inverse Fourier transformation. The diffraction image of a single duck is obtained. The colour (hue) corresponds to the phase information, the colour saturation represents the structure factor amplitude. **(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.

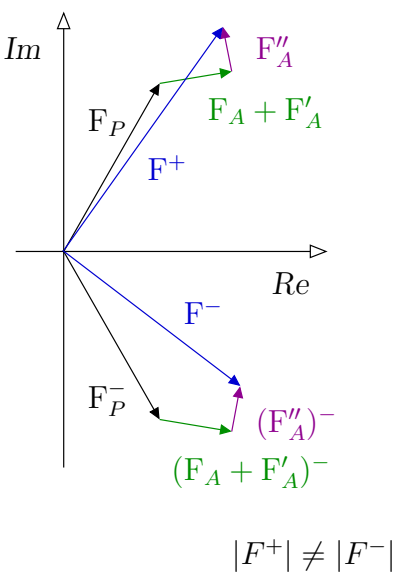
If a similar structure is already known and has been deposited with the Protein Data Bank PDB (Bernstein *et al.*, 1977), it may be used as a search model for **molecular replacement**. This method suffers considerably from model bias and requires sufficient structural homology between the search model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

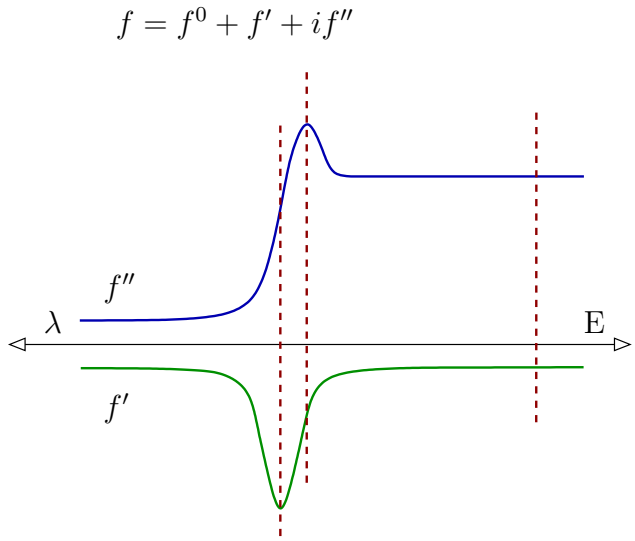
Methods based on **experimental phasing** are required for *de novo* phasing of proteins and nucleic acids and usually involve the search for heavy atoms. This is discussed in detail in the next chapter.

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.

Chapter 2

Experimental phasing





Summary

Experimental phasing is vital for *de novo* phasing of macromolecular structures. Although about two-thirds of newly deposited structures in the Protein Data Bank (PDB) were solved using molecular replacement, experimental phasing suffers less from model bias. It is required for samples that do not have any structurally related PDB entries and therefore it is a prerequisite for uncovering novel folds.

Methods for experimental phasing are based on the difference in the diffraction pattern induced by heavy atoms present in the crystal. These differences may originate from anomalous dispersion. This is due to X-ray absorption which is pronounced for specific elements at certain X-ray energies.

The changes in the diffraction pattern that are due to the heavy atoms are exploited to locate these atoms of interest. This substructure is employed to deduce the phases for the macromolecular structure. An initial electron density map is calculated and further improved, e.g. by density modification techniques.

2.1 X-ray absorption and radiation damage

Only a small fraction of the X-ray radiation is actually scattered by the crystal material, the main portion just passes through the crystal. However if interaction with the atoms in the crystal occurs the radiation may be absorbed (Rupp, 2010).

As already discussed the electric field vector of the X-ray radiation excites the electrons of the atoms in the crystal. When resonance occurs the scattering is not elastic and energy transfer is possible between an incoming photon and an inner-shell electron. The electron is ejected from its original position and the vacancy is filled by an electron from an outer shell of this atom. The transition produces radiation that is emitted in all directions. This X-ray fluorescence will contribute to diffuse background scattering.

The absorption spectrum shows characteristic peaks (Fig. 2.1). These X-ray absorption edges correspond to the discussed electronic transition. They are termed according to the shell where the electron is ejected from, e.g. K-edge or L_1-edge. Radiation with high energy is required to remove the inner-shell electrons of heavy elements.

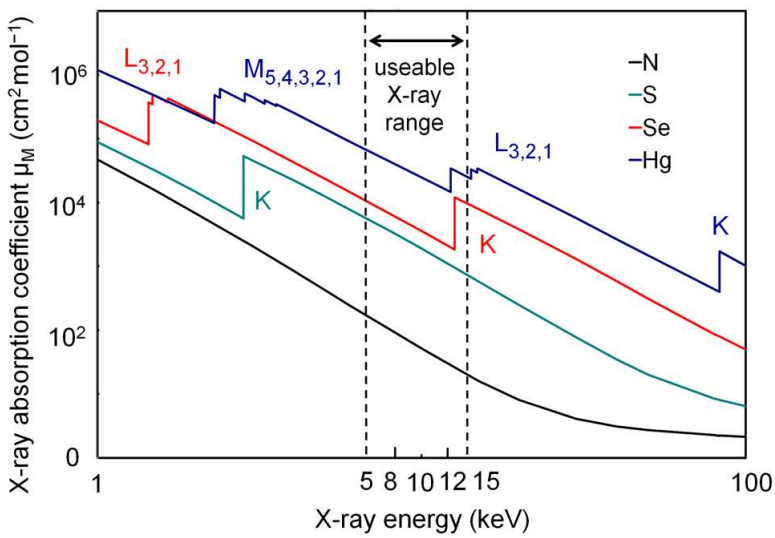


Figure 2.1: Calculated X-ray absorption of selected elements. The absorption edges are labelled according to the electronic transitions. Reproduced courtesy of Bernhard Rupp.

The absorption of a sample is recorded by measuring the X-ray fluorescence (Fig. 2.2). The absorption edge fine structure of an X-ray fluorescence spectrum can yield further information on the heavy atoms. The near-edge fine structure (*XANES*) is caused by electronic transitions in the absorbing atoms (Agarwal, 1991). Therefore, analysis of these features reveals details on the electronic state (e.g. oxidation state).

From the analysis of the extended X-ray absorption fine structure (*EXAFS*) information about the local environment of the absorbing atoms may be deduced (Agarwal, 1991). Weak oscillations in the fluorescence spectrum after the absorption edge are caused by the excited electrons that are scattered at neighbouring atoms. The scattered electrons and outgoing fluorescence photons interfere and cause an oscillation of absorption in dependence of energy (as shown in Fig. 2.2).

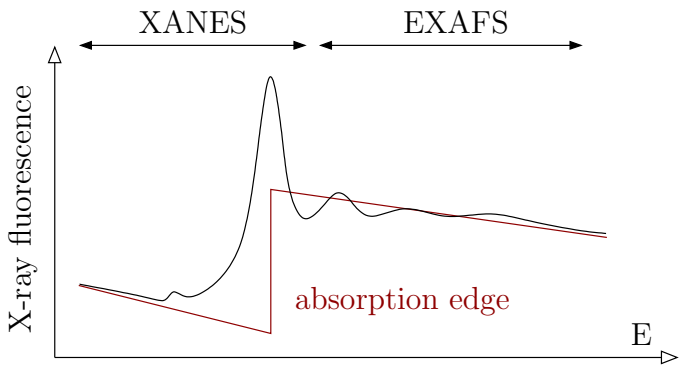


Figure 2.2: Schematic X-ray fluorescence spectrum with the fine structure of the experimental absorption edge shown in black. The theoretically calculated absorption edge is shown in red.

Due to resonance of the X-ray photons with the bound electrons, the atomic scattering factor has to be expressed as:

$$f_\lambda = f^0 + f'_\lambda + i f''_\lambda \tag{2.1}$$

In addition to the normal wavelength-independent scattering f^0, two wavelength-dependent contributions have to be taken into account, the dispersive term f'_λ and the anomalous term f''_λ.

The dispersive term is usually negative and expresses the decrease in scattering. The anomalous scattering contribution is imaginary. Due to the retardation of the absorbed wave the phase of the scattered wave is shifted by 90°. This phase shift applies to all reflections and causes the breakdown of Friedel's law (Fig. 2.3).

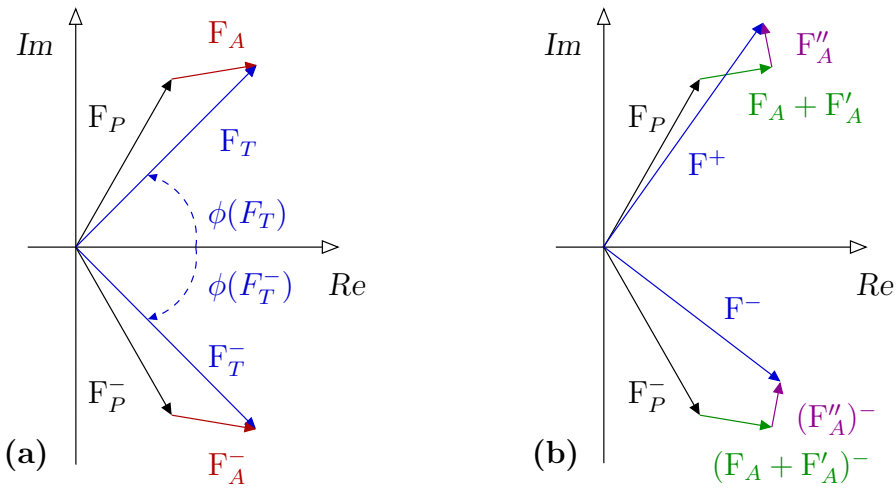


Figure 2.3: Argand diagram illustrating the breakdown of Friedel's law in the presence of anomalous scattering. **(a)** In the absence of anomalous scattering, the contributions of the protein atoms F_P and the heavy atoms F_A add up to yield the total scattering factor F_T. Friedel's law holds, i.e. $|F_T| = |F_T^-|$ and $\phi(F_T) = -\phi(F_T^-)$. **(b)** In the presence of anomalous scattering, the contribution of the heavy atoms has to be expressed as outlined in Eq. 2.1. The dispersive contribution F'_A is usually negative, but in phase with F_A. The anomalous scattering contribution F''_A introduces a phase shift of + 90°. Friedel's law does not hold and $|F^+| \neq |F^-|$.

The difference in intensity between $|F^+|$ and $|F^-|$ is caused by the anomalous scattering

and may be exploited for phasing (anomalous dispersion techniques, see section 2.3).

The absorption of energy in the crystal may cause **radiation damage** to the material. Crystals of biological macromolecules usually diffract only weakly at laboratory sources (rotating anode) due to disorder and high solvent content. Therefore these crystals have to be exposed to a high intensity of radiation (i.e. synchrotron radiation) to obtain suitable diffraction images. However X-ray radiation can ionise biological samples, bringing about damage to the crystal. The development of suitable cryocooling techniques has decreased the problem for a period of time, but with the advent of more brilliant beamlines (at second or third generation synchrotrons) radiation damage has come into focus again (Garman, 2010).

Absorption of X-ray energy in the crystal causes specific and global damage. Global radiation damage indicators are the increase of unit-cell volume, decay in diffracted intensity, increase in merging factor of the data and increase in the atomic B value (Owen *et al.*, 2006). Specific radiation damage inflicts specific structural damage to certain residues in the protein, e.g. cleavage of disulfide bonds or decarboxylation of glutamate and aspartate residues (Garman and Owen, 2006).

The radiation that is absorbed in the sample is mainly converted to free electrons (photoelectric effect). These photoelectrons may further ionise the sample, producing secondary electrons and radicals. At 100 K, the typical data collection temperature for macromolecular crystallography experiments, radical species are immobilised, including ·OH radicals from the solvent. However, the generated electrons are free to diffuse through the sample and go to the most electron-affinic sites, e.g. disulfide bonds or metal ions (Garman, 2010). The reduction of metal ions at active sites can lead to wrong assignment of oxidation states (Yano *et al.*, 2005; Corbett *et al.*, 2007). The radiolysis of anomalous scatterers, e.g. the bromine atom in brominated nucleotides, can prevent structure solution (Ennifar *et al.*, 2002).

Although radiation damage can seriously obstruct data collection, it may also be exploited for generating phase information. The radiation induced phasing (RIP) approach utilises the damage to specific sites for phasing (Ravelli *et al.*, 2003).

2.2 Isomorphous replacement

Isomorphous replacement is the traditional method to obtain experimental phase information for macromolecules (Kendrew *et al.*, 1958; Blake *et al.*, 1962). In the early days of protein X-ray crystallography, anomalous differences could not be measured accurately due to hardware limitations. Isomorphous replacement experiments exploit the differences in the diffraction from a native crystal and a derivatised crystal (see chapter 3). After a data set from a native crystal and a data set from a derivatised crystal have been collected, the difference in the intensities is derived. The differences are ideally only due to the presence of heavy atoms in the derivatised sample (Fig. 2.4). These differences may be used to locate the position of the heavy atoms in the unit cell (single isomorphous replacement, SIR).

The derivatisation with heavy atoms often induces changes to the arrangement of the macromolecules in the crystal and may therefore cause changes to the dimensions of the unit cell.

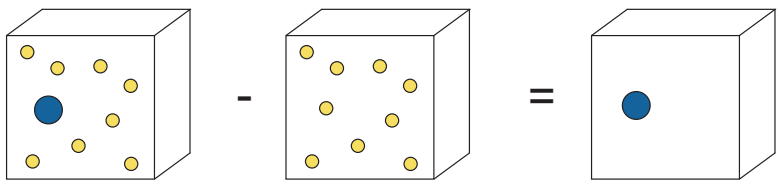


Figure 2.4: Isomorphous replacement. One unit cell for a protein crystal is shown; the blue sphere represents one heavy atom, yellow spheres represent protein atoms and solvent molecules. In theory a subtraction in real space of the native unit cell (*middle*) from the derivatised crystal (*left*) yields a unit cell that only contains the heavy atom (*right*). Similarly, the subtraction of the diffraction data in reciprocal space yields a data set that is ideally only caused by the heavy atom. These isomorphous differences may be used to locate the heavy atom in the unit cell.

Isomorphism is a prerequisite for the correct estimation of the isomorphous differences and the lack thereof is the main problem associated with isomorphous replacement, especially when several derivatives are used (multiple isomorphous replacement, MIR). Isomorphous replacement may be combined with methods using anomalous scattering (SIRAS, see below).

2.3 Anomalous dispersion

Methods for experimental phasing based on the anomalous scattering of certain atoms have largely replaced traditional methods such as multiple isomorphous replacement (Terwilliger and Berendzen, 1997; Hendrickson, 1999).[1] The difference in intensity of the *Bijvoet* pair[2] $|F^+|$ and $|F^-|$ is due to the presence of anomalously scattering atoms in the sample (Fig. 2.3). These differences are employed to find the **substructure**, i.e. the positions of the anomalously scattering atoms. From the positions of these heavy atoms,[3] also referred to as **marker atoms**, the reference phases ϕ_A are calculated. These can be used to derive starting phases ϕ_T for the macromolecular structure:

$$\phi_T = \phi_A + \alpha \tag{2.2}$$

The phase shift α is estimated from the experimental data as described below.

[1]Until the year 2000, more structures were solved using MIR or SIR (single isomorphous replacement) than with MAD/SAD techniques (a total of 394 structures using MIR/SIR vs. 109 using MAD/SAD were deposited from 1995 to 2000). However since then more structures have been solved with SAD or MAD methods (a total of 5760 using MAD/SAD vs. 1101 using MIR/SIR have been deposited since 2000), with the fraction of structures being solved with MAD/SAD still increasing. PDB statistics were kindly provided by Clemens Vonrhein. PDB entries with REMARK 200 set to NULL or OTHER were ignored. Date of PDB analysis: June 10, 2009.

[2]Reflection **h** and reflection -**h** form a *Friedel* pair. A *Bijvoet* pair is formed between a reflection **h** and any symmetry equivalent of reflection -**h**.

[3]The term heavy atom in macromolecular crystallography usually refers to any element that can be used for substructure solution, e.g. sulfur. The term marker atom is more appropriate; however to be consistent with the literature, atoms used for phasing are referred to as heavy atoms throughout this text.

Multi-wavelength anomalous dispersion (MAD)

In a multi-wavelength anomalous dispersion (MAD) experiment, several data sets around the absorption edge of the particular element are collected. In addition to taking advantage of the anomalous differences that are present within one single data set (*Bijvoet* pairs), the difference between data sets collected at different wavelengths, the dispersive signal f'_λ, is also exploited (e.g. difference between data set 'peak' and 'inflection point' in Fig. 2.5). This yields additional information for the location of the heavy atoms.

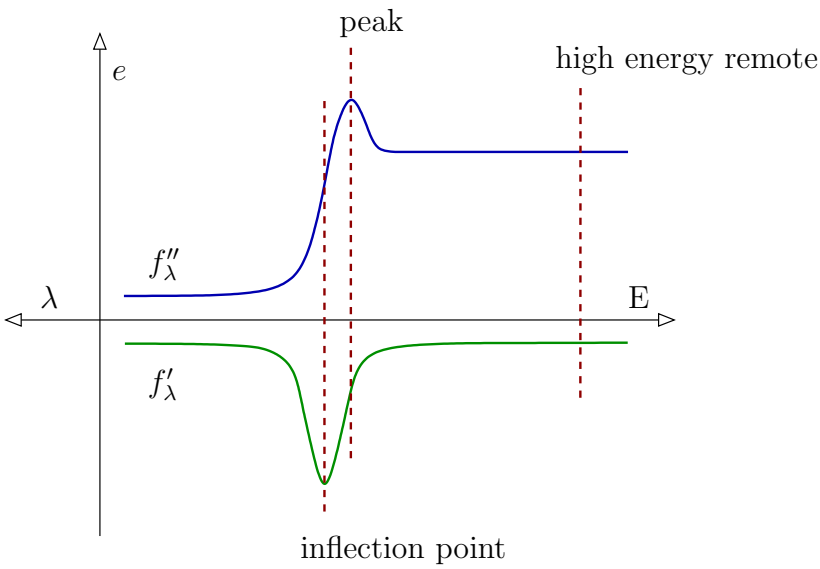


Figure 2.5: Schematic representation of the anomalous contribution f'_λ and f''_λ. The wavelength dependence is exploited in a MAD experiment. Data sets are collected at different wavelengths, marked as 'peak', 'inflection point' and 'high-energy remote', each having different f'_λ and f''_λ.

Although the anomalous scattering factors f'_λ and f''_λ are available from theoretical data, it is always advised to carry out an X-ray fluorescence scan in practice. Due to the chemical environment of the absorbing atoms (or beamline related issues) the wavelength of the absorption edge might be slightly shifted. The anomalous scattering contribution f''_λ is proportional to the atomic absorption coefficient at that wavelength and can therefore be directly obtained from the fluorescence scan. From f''_λ the real part f'_λ can be derived using the *Kramers-Kronig* transform (de L. Kronig, 1926).

Karle (1980) and Hendrickson *et al.* (1985) showed that for a MAD experiment the measured intensities for the *Bijvoet* pair $|F^+|$ and $|F^-|$ are represented by

$$|F^+|^2 = |F_T|^2 + a\ |F_A|^2 + b\ |F_T||F_A| \cos\alpha + c\ |F_T||F_A| \sin\alpha \tag{2.3}$$

$$|F^-|^2 = |F_T|^2 + a\ |F_A|^2 + b\ |F_T||F_A| \cos\alpha - c\ |F_T||F_A| \sin\alpha \tag{2.4}$$

with $a = (f''^2 + f'^2)/f_0^2$, $b = 2f'/f_0$ and $c = 2f''/f_0$. This is based on the assumption that only one type of heavy atom is present. a, b and c are different for each wavelength and depend on the scattering angle (since f_0 does).

Given that $|F^+|$ and $|F^-|$ were measured at least at two different wavelengths, the phase shift α and the amplitudes for the heavy-atom contribution $|F_A|$ and the total scattering $|F_T|$ can be derived. The amplitudes $|F_A|$ are used to locate the heavy atoms (see section 2.4).

From Fig. 2.5 it is clear that data should be collected at two wavelengths to maximise the dispersive differences (the difference in f'). Additionally for each data set the anomalous signal should be strong: a large f'' value leads to a large difference for the *Bijvoet* pair $|F^+|$ and $|F^-|$.

Different strategies exist about the collection order of the data sets. The peak data set (see Fig. 2.5) is usually collected first, since at this wavelength the anomalous signal is maximal; if MAD phasing fails (due to radiation damage), this data set may still be used for SAD phasing (see below). The next data sets are the high-energy remote and the inflection point data set. Usually the inflection point is collected last to ensure that possible radiation damage to the heavy-atom sites has the same sign as the dispersive differences.

Single-wavelength anomalous dispersion (SAD)

For a single-wavelength anomalous dispersion (SAD) experiment, a data set is collected at a wavelength that gives rise to a strong anomalous signal for the heavy atoms present in the crystal. At the synchrotron beamline the wavelength may be adjusted to maximise the signal, e.g. close to an absorption edge. However, SAD phasing may also be carried out in-house, with elements that have a suitable signal at CuKα, e.g. iodine or sulfur.

For a single wavelength experiment, there are only two equations for each *Bijvoet* pair $|F^+|$ and $|F^-|$. Subtracting Eq. 2.3 from Eq. 2.4 results in

$$|F^+|^2 - |F^-|^2 = 2\ c\ |F_T||F_A| \sin\alpha \tag{2.5}$$

with $c = 2f''/f_0$. With $|F_A| \ll |F_T|$ the total scattering can be approximated as the mean value of $|F^+|$ and $|F^-|$:

$$|F_T| = \frac{1}{2}\left(|F^+| + |F^-|\right) \tag{2.6}$$

With Eqs. 2.5 and 2.6 and help of the expression $a^2 - b^2 = (a+b)(a-b)$ the *Bijvoet* difference may be expressed as:

$$|F^+| - |F^-| = c\ |F_A| \sin\alpha \tag{2.7}$$

For large differences, the phase shift α may be deduced. In contrast to a MAD experiment where values between 0 and 360° for α may be derived, the SAD approximation yields only values of 90° or 270° (Fig. 2.6). For small differences, a two-fold ambiguity is present that can be resolved by means of density modification (see section 2.5).

The findings from Fig. 2.6 are in agreement with Eq. 2.7: at $\alpha = 90°$ or 270° the sine value is ± 1. Therefore, for large differences:

$$\big||F^+| - |F^-|\big| \approx c\ |F_A| \tag{2.8}$$

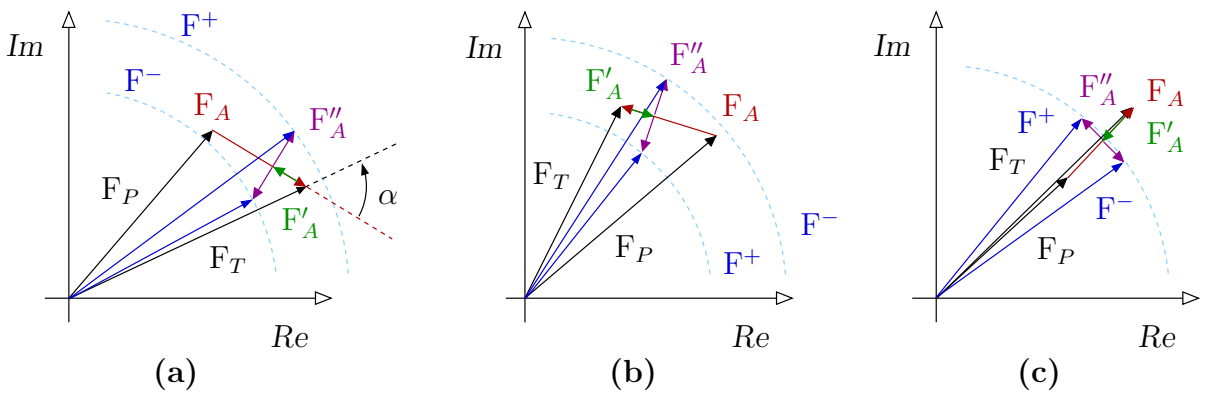


Figure 2.6: Estimating the phase shift α for SAD. For convenience, the complex conjugate of F^- is shown. The normal scattering F_A, the dispersive (negative) F'_A and the anomalous contribution F'_A are depicted, resulting in the *Bijvoet* pair $|F^+|$ and $|F^-|$. **(a)** If $|F^+| \gg |F^-|$, adding a phase shift of $\alpha = 90°$ to the phase of F_A is a good approximation for the phase of F_T. **(b)** If $|F^+| \ll |F_-|$, a phase shift of $\alpha = 270°$ is a good approximation. **(c)** If $|F^+|$ and $|F^-|$ are approximately equal, the reflection has either $\phi_T = \phi_A$ or $\phi_T = \phi_A + 180°$, a two-fold ambiguity (Sheldrick, 2002)

Single isomorphous replacement plus anomalous scattering (SIRAS)

For a single isomorphous replacement (SIR) experiment, the derivative structure factor $|F_D|$ is expressed, similarly to a SAD experiment, as:

$$|F_D|^2 = |F_T|^2 + a\ |F_A|^2 + b\ |F_T||F_A| \cos\alpha \tag{2.9}$$

with $a = (f''^2 + f'^2)/f_0^2$ and $b = 2f'/f_0$, after averaging Friedel opposites and assuming perfect isomorphism. The difference between the derivative structure factor $|F_D|$ and the native structure factor $|F_T|$ is usually small, therefore:

$$|F_D|^2 - |F_T|^2 = (|F_D| + |F_T|)(|F_D| - |F_T|) \approx 2|F_T|(|F_D| - |F_T|) \tag{2.10}$$

Only reflections with large isomorphous differences and therefore $\alpha = 0°$ or $180°$ contribute useful information. With the additional assumption that $|F_T| \gg |F_A|$, the isomorphous difference is expressed as:

$$|F_D| - |F_T| \approx \frac{1}{2}\ b\ |F_A| \cos\alpha \tag{2.11}$$

The assumptions used here (especially perfect isomorphism) are more severe than for SAD. However, the combination of a SAD experiment with a native data set may improve the extent of phase information (single isomorphous replacement plus anomalous scattering; SIRAS). After normalisation the SAD information $|F_A| \sin\alpha$ and the SIR information $|F_A| \cos\alpha$ are combined:

$$|F_A| = \left[(|F_A| \sin\alpha)^2 + (|F_A| \cos\alpha)^2\right]^{\frac{1}{2}} \tag{2.12}$$

With the signs of sin α and cos α taken into account the phase angle α can be calculated:

$$\frac{|F_A| \sin \alpha}{|F_A| \cos \alpha} = tan\ \alpha \tag{2.13}$$

in a full range of 0° to 360°, similarly to a MAD experiment. Since the phase information from the two sources are orthogonal to each other, this technique is very useful if weak phase information from a SAD experiment is available in conjunction with an isomorphous native data set of higher resolution (see below).

Anisotropy of anomalous scattering (AAS)

The anomalous scattering from isolated atoms is a scalar quantity. However the local chemical environment, i.e. the bonds to other atoms, introduces a directional dependence of the dispersive and anomalous contributions on the direction of linear polarisation of the X-ray beam. This is only important for synchrotron radiation where the X-ray beam is linearly polarised in the plane defined by the orbit of the electron beam, that is the horizontal plane (Schiltz and Bricogne, 2010). The anisotropy of the anomalous scattering (AAS) is significant close to the absorption edge, resulting in different fluorescence spectra that depend on the orientation of the molecular principal directions. The anisotropy of anomalous scattering can be exploited to enhance the phase information present in collected data (Schiltz and Bricogne, 2008, 2010).

2.4 Substructure determination

Although the resolution of data obtained from macromolecular crystals usually does not extend to atomic resolution (i.e. < 1.2 Å), *ab initio* methods used for the solutions of small-molecule structures may still be employed for substructure solution in macromolecules. Difference structure factor data (anomalous data) to 2.5-3.0 Å is usually sufficient since the distance between the atoms in a substructure is usually greater than 3.0 Å. The methods outlined above (except AAS) are incorporated in the *SHELXC/D/E* suite (Sheldrick, 2002, 2008a, 2010).

SHELXC (or *XPREP*; Bruker, 2007) takes the unmerged reflection file and estimates the amplitudes $|F_A|$ of the heavy atom substructure. For MAD data separate estimates for the dispersive and anomalous contribution are combined, similarly to the SIRAS formalism (Eq. 2.13). The amplitudes $|F_A|$ are used for substructure solution.

In the program *SHELXD* (Usón and Sheldrick, 1999; Schneider and Sheldrick, 2002; Sheldrick, 2008a), originally written to solve large and difficult small-molecule structures, the structure factor amplitudes are normalised to disregard the atomic thermal motion and the electron density distribution:

$$E_{\mathrm{h}}^2 = \frac{F_{\mathrm{h}}^2/\epsilon}{< F^2/\epsilon >} \tag{2.14}$$

with ϵ as statistical factor for proper treatment of special reflections. $< F^2/\epsilon >$ is the mean value per resolution shell. These normalised structure factors correspond to structure

factors calculated for a point-atom structure. In the case of SAD and SIR, the normalisation deletes the constants in Eq. 2.7, 2.8 and 2.11.

For substructure determination only the largest ∼15% of the observed difference structure factors ΔF (SAD or SIR) or F_A (MAD or SIRAS) are used. These correspond to the strongest normalised difference structure factors E.

Starting with atoms that are consistent with the Patterson function improves the substructure solution substantially. A two-atom fragment is chosen, based on one of the highest peaks in the Patterson function. A translational search is carried out by using a special Patterson minimum function (PMF). The two-atom fragment is translated randomly in the unit cell. For each position the Patterson function values for the vectors between the two atoms and all their symmetry equivalents are stored. The PMF is calculated as the mean value of the weakest 30% of this list. The position of the two-atom fragment with the highest PMF is kept.

These two starting atoms are expanded by using a full-symmetry Patterson superposition minimum function (PSMF). All symmetry equivalents are calculated. A dummy atom is placed on a pixel and all vectors between it and the atoms are evaluated. Each pixel in the PSMF map is assigned a value equal to the PMF value for these vectors. This map is used for peak search to expand the substructure (Schneider and Sheldrick, 2002).

Based on these atoms a first set of phases is calculated. The tangent formula (Karle and Hauptman, 1956; Karle, 1968) is used to refine these phases in reciprocal space:

$$tan(\phi_{\mathrm{h}}) = \frac{\sum_{\mathrm{h}'} |E_{\mathrm{h}'} E_{\mathrm{h-h}'}| \sin(\phi_{\mathrm{h}'} + \phi_{\mathrm{h-h}'})}{\sum_{\mathrm{h}'} |E_{\mathrm{h}'} E_{\mathrm{h-h}'}| \cos(\phi_{\mathrm{h}'} + \phi_{\mathrm{h-h}'})} \tag{2.15}$$

The refined phases are used in combination with the normalised structure factors to calculate an improved map. A new round of peak picking is conducted to select a new set of atoms. The iteration between real space and reciprocal space is known as **dual space recycling** (Fig. 2.7; Sheldrick *et al.*, 2001).

Instead of starting with a two-atom fragment based on the Patterson function, a vector with fixed length (e.g. disulfide bond) or even a small molecular fragment may be used for seeding (Schneider and Sheldrick, 2002). The use of random omit maps instead of probabilistic Patterson sampling is usually a good alternative for *ab initio* structure solution at atomic resolution.

The dual space recycling is carried out only with the strongest normalised structure factors. The correlation coefficient (CC) between the calculated structure factors E_c and the observed ones E_o is a measure for the quality of the solution (Fujinaga and Read, 1987):

$$CC = \frac{100[\sum(wE_oE_c)\sum w - \sum(wE_o)\sum(wE_c)]}{\{[\sum(wE_o^2)\sum w - (\sum wE_o)^2][\sum(wE_c^2)\sum w - (\sum wE_c)^2]\}^{\frac{1}{2}}} \tag{2.16}$$

The weights w are employed to weight down less reliable difference structure factors. The correlation coefficient calculated with all data (CC) and the correlation coefficient calculated with only the E values not used for substructure determination [CC_{weak}; similar to the R_{free} factor used in refinement (Brünger, 1992)] are good indicators whether the sub-

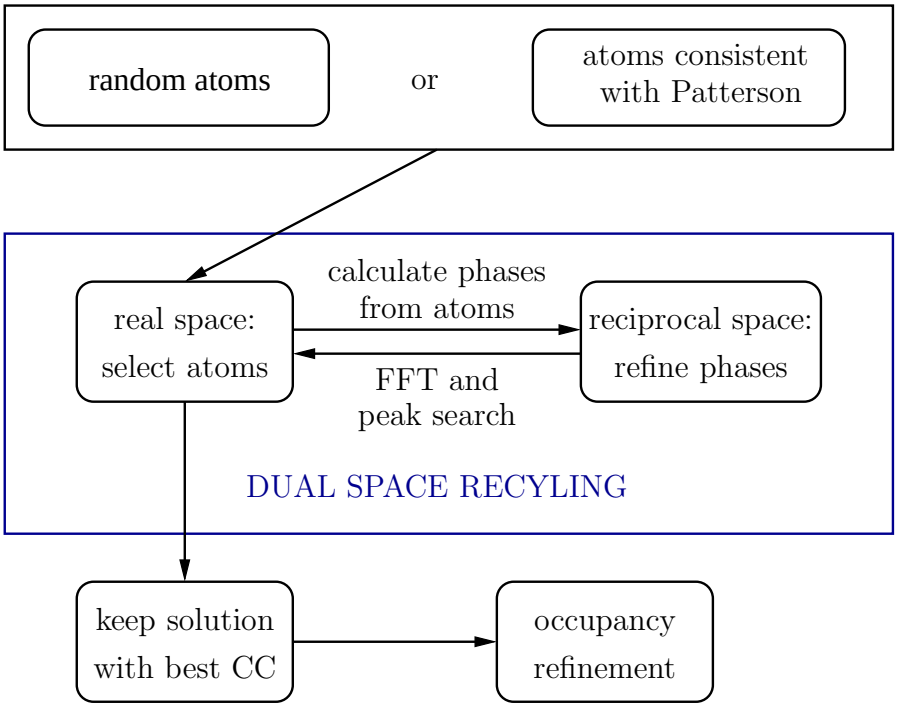


Figure 2.7: Dual space recycling in *SHELXD*. Starting from a set of atoms, the phases are refined in reciprocal space. A new map is calculated that is used for peak search. The peak picking imposes a strong atomicity constraint and is an extreme form of density modification. The new set of atoms is used to determine a new set of phases.

structure was solved. For a MAD experiment CC values greater than 35% are expected. A SAD experiment usually gives lower values ($CC \geq 25\%$, $CC_{weak} \geq 15\%$); however there have also been cases when structures were solved that showed considerably lower values.

A critical parameter for substructure solution is the resolution cutoff for the anomalous data. Direct methods are based on normalised structure factors, which emphasise the high resolution data. Therefore, using data to a higher resolution than there is considerable anomalous (or dispersive) signal will deteriorate structure solution. To assess the quality of the data, the correlation coefficient for the anomalous signal CC_{anom} is a good indicator (see Appendix B: Crystallographic parameters). It compares the signed anomalous differences for different data sets. In general, data should be truncated at the resolution where the CC_{anom} falls below 30%. For a SAD experiment the mean *Bijvoet* difference for each resolution shell is divided by its estimated standard deviation. Additionally, instead of the estimated standard deviations the variance of the differences is also used. This does not depend on (error-prone) estimation of standard deviations. The correlation coefficient between the signed anomalous differences of two randomly chosen subsets is another option for assessing the data quality, but is less reliable than for a MAD experiment.

In the final dual-space cycle the occupancy of the atoms is refined in *SHELXD* (Fig. 2.7). This provides additional information since a sharp drop is expected after the last site for covalently bound anomalously scattering atoms. For crystals that have been derivatised *via* soaking with a heavy-atom solution, a continuous fall-off is usually observed, due to many sites with decreasing occupancy. The result from the substructure determination step is a file that contains the heavy atoms with their coordinates and occupancies.

2.5 Density modification

The positions of the heavy-atom sites can be used to calculate the phases ϕ_A. The starting phases ϕ_T for the macromolecular structure are calculated by adding the phase shift α to these reference phases (Eq. 2.2). Since both enantiomorphs of the substructure are in equally good agreement with the structure factor amplitudes $|F_A|$, both enantiomorphs (and therefore both enantiomorphs of enantiomorphic space groups) have to be tried for phase extension (except when the substructure is centrosymmetric).

Several features of the program *SHELXE*, which improve the experimental electron density map, are briefly discussed:

- resolving the two-fold ambiguity (SAD and SIR only)
- density modification with the sphere of influence algorithm
- extending the data to higher resolution than actually measured (free lunch algorithm)
- main chain autotracing

For SAD and SIR, a **two-fold ambiguity** for those reflections is present where the difference ΔF between $|F^+|$ and $|F^-|$ or $|F_D|$ and $|F_T|$, respectively, is small, i.e. the phase shift α cannot be estimated. To resolve this ambiguity, the program *SHELXE* calculates a first electron density map with centroid phases ϕ_T, $|F_T|$ and suitable weights (figure of merit, based on the magnitude of ΔF). Only a small fraction of the pixels with highest intensity (usually 7%) is retained, the rest is set to zero. After inverse Fourier transformation phases for the reflections with small ΔF and for centric reflections are derived (Sheldrick, 2002). This is similar to the low density elimination (Shiono and Woolfson, 1992; Giacovazzo and Siliqi, 1997).

The phases ϕ_T are further improved by means of density modification. The electron density of the solvent region is usually flat and featureless whereas the density in the protein region has a large variance. This has been exploited in several programs, e.g. *DM* (Cowtan, 1994) or *RESOLVE* (Terwilliger, 2000). The program *SHELXE* uses the **sphere of influence** algorithm to introduce chemical information in the electron density (Sheldrick, 2002). The variance of the electron density on a spherical surface (approximated by the vectors from the centre to each atom and from the centre to all faces in fullerene) is calculated for each pixel. Pixels with low variance are flipped (i.e. solvent region) and for those with high variance the density is sharpened (i.e. protein region). The *fuzzy solvent boundary* is employed for pixels in between to allow for a weighted mean of the two treatments. The radius of the sphere of influence is 2.42 Å, a typical 1,3 distance in proteins and DNA/RNA.

Figures of merit for the density modification are *contrast* and *connectivity* (Sheldrick, 2002). Additionally, after leaving out ∼10% of the reflections for one density modification step, the pseudo-free CC is calculated between the calculated normalised structure factors and the observed ones for these reflections. Values $> 70\%$ usually indicate an interpretable map. These indicators may be used to distinguish between the correct and wrong substructure enantiomorph.

The phase information from the centroid phases and the low density elimination step (only applicable for SAD and SIR) are combined using σ_A-weights (Read, 1986). For SAD,

if the heavy atoms are present in the native data, their direct contribution to the phases may also be added using σ_A-weights (similar to the Patterson method for small molecules). Since this information does not come from the differences ΔF, it is orthogonal to the centroid phases (and may help in resolving the two-fold ambiguity).

SAD phasing usually works well if the solvent content is high and native data are available to a high resolution (< 2.0 Å) since density modification will improve and extend the rather crude starting phases considerably. However, if the resolution is worse, MAD phasing is needed since it provides more phase information than SAD. In some cases at low resolution, density modification might not be required or even fail to improve the electron density map.

The **free lunch algorithm** (FLA) as incorporated in *SHELXE* extends data to a higher resolution than actually measured, based on the Fourier transformation of the density-modified map (Usón *et al.*, 2007). This seems to work well since *'zero'* is a bad estimate of the amplitudes of the reflections not measured and therefore corrects for Fourier truncation errors. Additionally the phases derived from density modification for these reflections play a more important role than the amplitudes (see Fig. 1.2).

In a recent version of *SHELXE*, an **autotracing** algorithm for the protein main chain was introduced that is iterated with the density modification procedure to "get a toehold in a noisy map" (Sheldrick, 2010). Although designed for data where the usual density modification procedures only produce an uninterpretable map, it is also useful for straightforward structures to improve the experimental electron density map, before it is passed to programs like *ARP/wARP* (Langer *et al.*, 2008), *RESOLVE* (Terwilliger, 2000) or *Buccaneer* (Cowtan, 2006) for tracing and side-chain fitting. It may be used as a fast validation tool for data quality and completeness during a synchrotron session: if a main chain trace can be found, the crystal may be removed to save beam time.

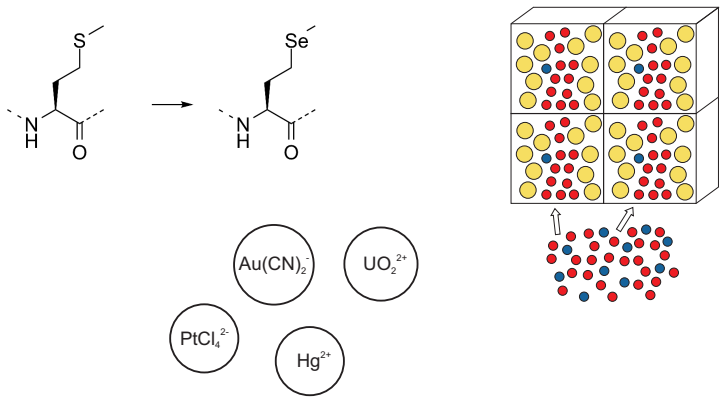
Molecular replacement usually gives a solution that will depend very much on the search model. Anomalous data (maybe not sufficient to solve the structure *de novo*) can be used to eliminate this *model bias*. The combination of phase information from a partial molecular replacement solution and weak experimental phases (MRSAD) has led to a number of successes (Tereshko *et al.*, 2008; Schuermann and Tanner, 2003; Roversi *et al.*, 2010) and is also included in the *Auto-Rickshaw* server (Panjikar *et al.*, 2009). In *SHELXE* a partial protein structure may be read in to carry out density modification and locate the heavy atoms based on the provided anomalous data. In a second round, these heavy atoms can be the starting point for new reference phases (and therefore eliminating the model bias) and a new run of autotracing. Another application for the autotracing algorithm in *SHELXE* is the program *ARCIMBOLDO* (Rodriguez *et al.*, 2009).

For difficult cases it is sometimes advisable to carry out a refinement of the heavy-atom positions to improve the reference phases ϕ_A. This may either be done with the program *SHARP* (Vonrhein *et al.*, 2007) or by recycling the heavy-atom positions in *SHELXE*: after one round of density modification (e.g. combined with autotracing) the improved heavy-atom positions are read in again in the next round. This sometimes significantly improves the final electron density map (and the main chain trace), especially if the FLA has been used since it can improve the precision of the heavy-atom positions.

The result from the steps outlined above is an experimental electron density map, which is transferred to programs that will trace the protein according to the provided sequence.

Chapter 3

Heavy-atom derivatisation



Summary

Heavy atoms are required for experimental phasing. Intrinsic scatterers like sulfur, phosphorus or endogenous metals may be used, but usually their anomalous signal does not suffice for phasing. Therefore external scatterers need to be introduced. The main route to experimental phasing is without doubt the use of a selenomethionine mutant of the protein under investigation.

A less demanding approach is to soak native crystals in a solution containing anomalously scattering ions or compounds. Traditional heavy-atom compounds include mercury, platinum, uranium or gold salts. Some problems associated with conventional soaks include toxicity and non-specific binding, resulting in low occupancy of the heavy-atom sites or derivatisation failing completely. The quick soak approach utilises a high concentration of heavy-atom compound in conjunction with a short soaking time, minimising the risk of damaging the crystal.

Another approach is the co-crystallisation of the macromolecules with anomalously scattering atoms. This usually requires the screening for new crystallisation conditions since the new compound will influence the solubility and crystallisation of biological macromolecules.

3.1 Introduction

With the latest advances in synchrotron hardware, e.g. improved detectors, phasing with the weak anomalous signal from intrinsic scatterers, e.g. sulfur (Debreczeni *et al.*, 2003a; Dauter *et al.*, 1999), phosphorus (Dauter and Adamiak, 2001) or metal ions (Coelho *et al.*, 2001; Lukat *et al.*, 2008), has become a possible option, though high quality data are required. This is more readily obtained when high symmetry enables a high data multiplicity to be achieved, but low-symmetry examples have also proved successful (Lakomek *et al.*, 2009).

However in case of sub-optimal data quality or absence of suitable intrinsic scatterers, external anomalously scattering atoms have to be introduced into the protein crystal. The main route for solving novel protein structures is without doubt the use of a selenomethionine derivative (Hendrickson *et al.*, 1991). Similarly, the chemical incorporation of brominated nucleobases has become an important technique for nucleic acid structure determination (Correll *et al.*, 1997). The substitution of endogenous metal ions by heavier metals with similar valency is also a possible option (Holland *et al.*, 1995). Anomalously scattering atoms can also be introduced by chemical modifications of the protein or nucleic acid (Xie *et al.*, 2004; Miyatake *et al.*, 2006; Oliéric *et al.*, 2009).

A less demanding approach for anomalous scattering experiments is to soak native crystals in a solution containing anomalously scattering ions or compounds. Another possible option is the co-crystallisation of the macromolecule with anomalously scattering atoms. These two techniques are discussed in the next sections.

3.2 Soaking

Protein crystals have a high solvent content (usually in the range of 30 to 70 %). Due to their mostly globular shape, the protein molecules are loosely packed in the crystal lattice, with solvent occupying the space and solvent channels around them. This aqueous environment[1] can be exploited when derivatising a native crystal. The protein crystal is transferred to a solution containing heavy atoms. These will diffuse into the crystal through the solvent channels and replace atoms (solvent molecules or light ions) at the surface of the protein molecule. If the heavy-atom compounds show some lipophilicity, they may sometimes diffuse into the hydrophobic core area of the protein and bind to residues in the protein core.

Traditionally, derivatisation was carried out with heavy atoms such as mercury, platinum, uranium or gold for MIR experiments. The heavy atoms form complexes with the functional groups of amino acids as ligands, e.g. mercury is known for forming stable covalent bonds with sulfhydryl groups in free cysteines. For a list of popular heavy-atom salts see Carvin *et al.* (2001).

In general soaking is carried out with a solution containing the heavy-atom salt but also the precipitant and buffer of the crystallisation condition. For conventional heavy-atom soaks, soaking times last from 10 min up to several days. The concentration for soaks

[1] It resembles *in vivo* conditions and is the reason why some enzymes still show activity in the crystalline state (Pearson and Owen, 2009)

is usually in the range of 0.1 – 100 mM, sometimes limited by the solubility of the heavy-atom compound. Heavy-atom soaks traditionally have a low success rate, but systematic heavy-atom screening with conventional heavy-metal ions has been performed using gel-electrophoresis (Boggon and Shapiro, 2000), mass spectrometry (Agniswamy *et al.*, 2008; Joyce *et al.*, 2010) or a database approach (Sugahara *et al.*, 2005). The incorporation of heavy atoms may also be monitored using microPIXE (Garman, 1999).

The quick soak approach utilises a high concentration of heavy halides [bromine or iodine salts (Dauter *et al.*, 2000, 2001) or triiodide (Evans and Bricogne, 2003)] or heavy alkali metals [cesium or rubidium salts (Nagem *et al.*, 2003)]. Soaking times are usually less than a minute, reducing the risk of damage to the crystal. Any degradation of the crystal is halted by quickly flash-cooling the crystal in liquid nitrogen. However, the ions usually bind non-specifically to the surface of the protein, resulting in many sites with low average occupancy. Nevertheless, due to the high anomalous signal, these partial sites may still be used for phasing. The quick soak approach has also successfully been applied to conventional heavy-atom soaks (Sun *et al.*, 2002; Sun and Radaev, 2002).

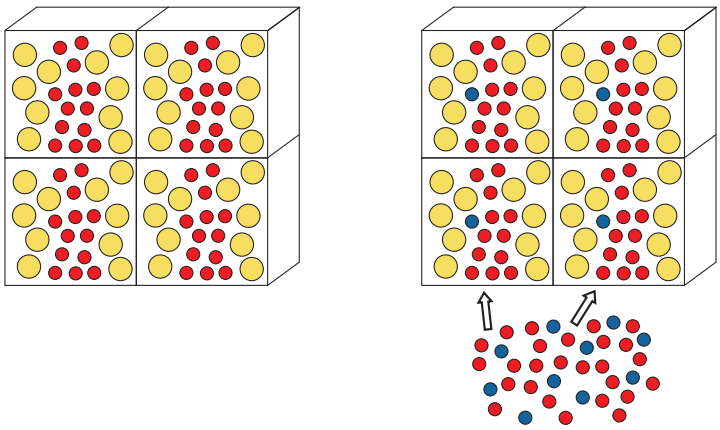


Figure 3.1: Schematic representation of heavy-atom derivatisation *via* soaking. Four unit cells of a macromolecular crystal are shown, with the solvent channels clearly visible. The native crystal (*left*) is transferred to a solution (*right*) containing the heavy-atom compound, but also buffer and precipitant at the same concentrations as in the crystallisation condition. The heavy-atom ions (blue spheres) diffuse into the solvent channels, replace solvent molecules and bind to the macromolecule.

Other heavy-atom compounds used for experimental phasing are lanthanoid complexes (Girard *et al.*, 2003). The strong anomalous signal of lanthanoid may also be exploited by incorporating lanthanoid binding tags in proteins (Silvaggi *et al.*, 2007). For phasing of larger macromolecular entities, several sites with heavy atoms are required. A successful approach is the utilisation of multi-metal cluster compounds, e.g. the $Ta_6Br_{12}^{2-}$ cluster (Schneider and Lindqvist, 1994; Banumathi *et al.*, 2003), tungsten or gold clusters (Thygesen *et al.*, 1996). Noble gases may also be used for heavy-atom derivatisation, e.g. xenon (Djinovic-Carugo *et al.*, 1998).

In contrast to multi-isomorphous replacement experiments, only one derivative is required for SAD and MAD phasing. Non-isomorphism is not an issue since the data are usually collected from one crystal only, i.e. the derivatised one. However the derivatisation may still cause a disruption of the lattice, leading to a decay in diffraction intensity. To minimise these deleterious effects, soaking is carried out in a stabilising solution, containing the heavy-atom compound, but also buffer and precipitant at the same concentrations as in the crystallisation condition. A gradient soak, i.e. stepwise increasing the heavy-atom

content of the soaking solution, may sometimes be a valuable option for fragile crystals. Heavy-atom ions that are present in the solvent channels but not bound to the protein may contribute to diffuse background scattering (and radiation damage). Therefore, back soaking in the cryo solution before flash-cooling is recommended. However if binding is not very strong (e.g. when using halide soaks), a back soak may not be possible since it will remove most anomalously scattering atoms from the protein crystal. In these cases, the cryo solution may contain a small amount of heavy-atom compound. For further practical advice see Garman and Murray (2003). When assessing the anomalous signal of certain elements Ethan Merrit's website provides a good resource for the strength of the anomalous signal vs. wavelength (Merrit, 2010).

3.3 Co-crystallisation

Heavy-atom compounds may also be incorporated by means of co-crystallisation. The heavy-atom compound is present in the crystallisation drop and is incorporated in the protein crystal during crystal growth. The presence of heavy-atom salts usually changes the crystallisation condition. Therefore, a screening has to be carried out to establish new crystallisation conditions. However the effect of additives on the crystallisation properties of a protein may be exploited. Additive screens affect the solubility and crystallisation of proteins. Some molecules may act as linkers between protein molecules (McPherson and Cudney, 2006). If heavy-atom compounds or small molecules, e.g. inhibitors, are successfully incorporated *via* co-crystallisation, the occupancy in the derivatised crystal is usually high, since the molecules or ions are built in during crystallisation.

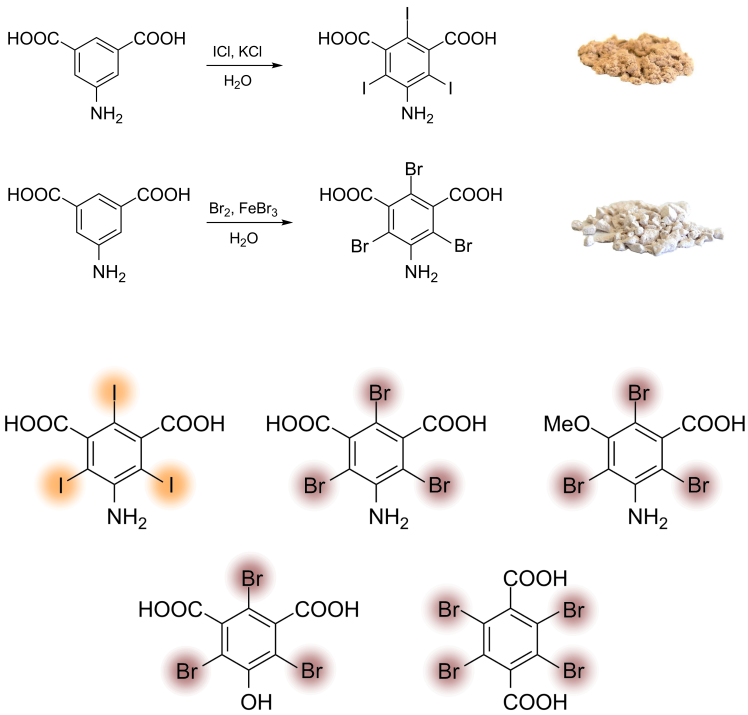
Since co-crystallisation usually requires the screening for new conditions and is therefore more demanding than soaking, it is not applied as often as soaking for derivatisation. Examples include lanthanoid compounds (Girard *et al.*, 2003; Pompidor *et al.*, 2010) or ammonium iodide (Yogavel *et al.*, 2007).

Part II

Materials and methods

Chapter 4

Synthesis of sticky triangles



Summary

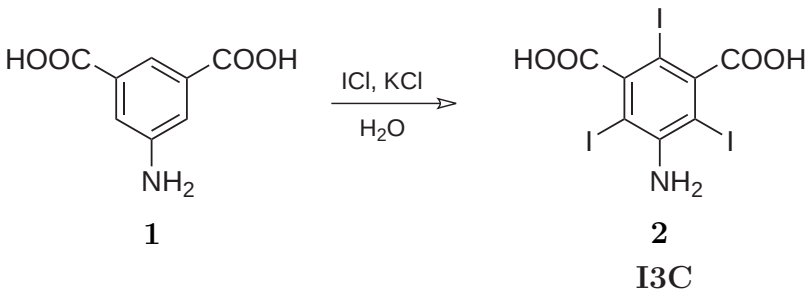
Several compounds containing three halogen atoms and three functional groups were synthesised and characterised. Tri-substituted benzene derivatives were either iodinated or brominated. Apart from the carboxyl and amino group, other functional groups were tested, e.g. replacing the amino group with a guanidinium functionality, similar to the side chain in the amino acid arginine. However bromination in the presence of the deactivating guanidium group is only successful if other activating substituents are present. Five compounds were obtained that can be used for experimental phasing.

4.1 Introduction

Based on the lead structure with three heavy atoms for experimental phasing and three functional groups for protein interaction, several compounds were synthesised to be used for experimental phasing of macromolecules. For further details on the synthesis and characterisation see Appendix E: Experimental data for small molecules.

4.2 Iodinated compounds

Iodination of 5-aminoisophthalic acid **1** afforded the compound 5-amino-2,4,6-triiodoisophthalic acid **2** (hereafter referred to as I3C).

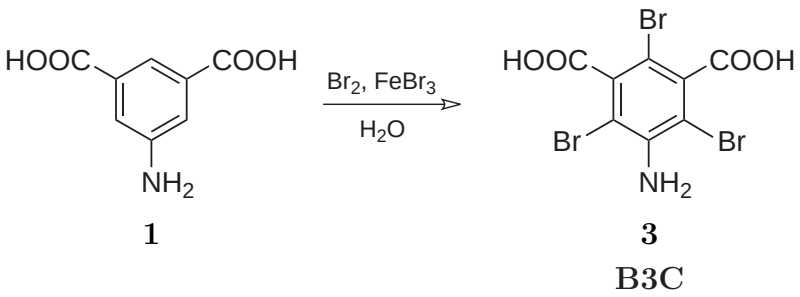


Scheme 4.1: Synthesis of 5-amino-2,4,6-triiodoisophthalic acid, **I3C**.

Iodomonochloride was employed as the iodination reagent, dissolved in potassium chloride solution (Parkesh *et al.*, 2006; Larsen *et al.*, 1956). I3C may also be obtained from different chemical suppliers (CAS number 35453-19-1). I3C and its use for experimental phasing are discussed in detail in chapter 7.

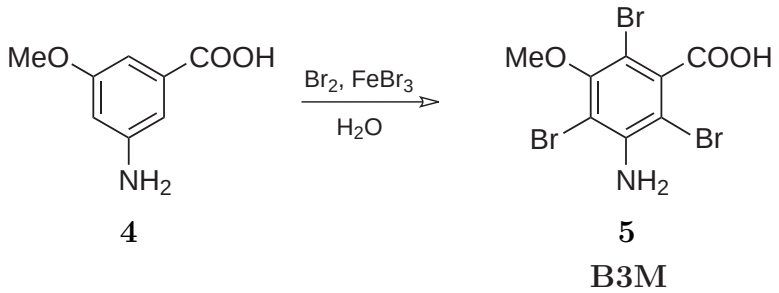
4.3 Brominated compounds

Bromination of 5-aminoisophthalic acid **1** with elemental bromine and iron(III)bromide as catalyst afforded the compound 5-amino-2,4,6-tribromoisophthalic acid **3** (hereafter referred to as B3C).



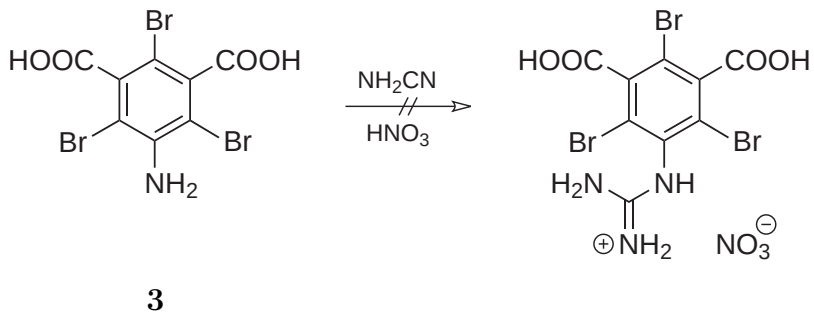
Scheme 4.2: Synthesis of 5-amino-2,4,6-tribromoisophthalic acid, **B3C**.

Different functional groups were introduced to enhance the binding capabilities of the small molecules. One carboxyl group was substituted with a methoxy group to introduce a functional group suitable for interaction with hydrophobic residues. Bromination yielded the compound 3-amino-2,4,6-tribromo-5-methoxybenzoic acid **5** (hereafter referred to as B3M).



Scheme 4.3: Synthesis of 3-amino-2,4,6-tribromo-5-methoxybenzoic acid, **B3M**.

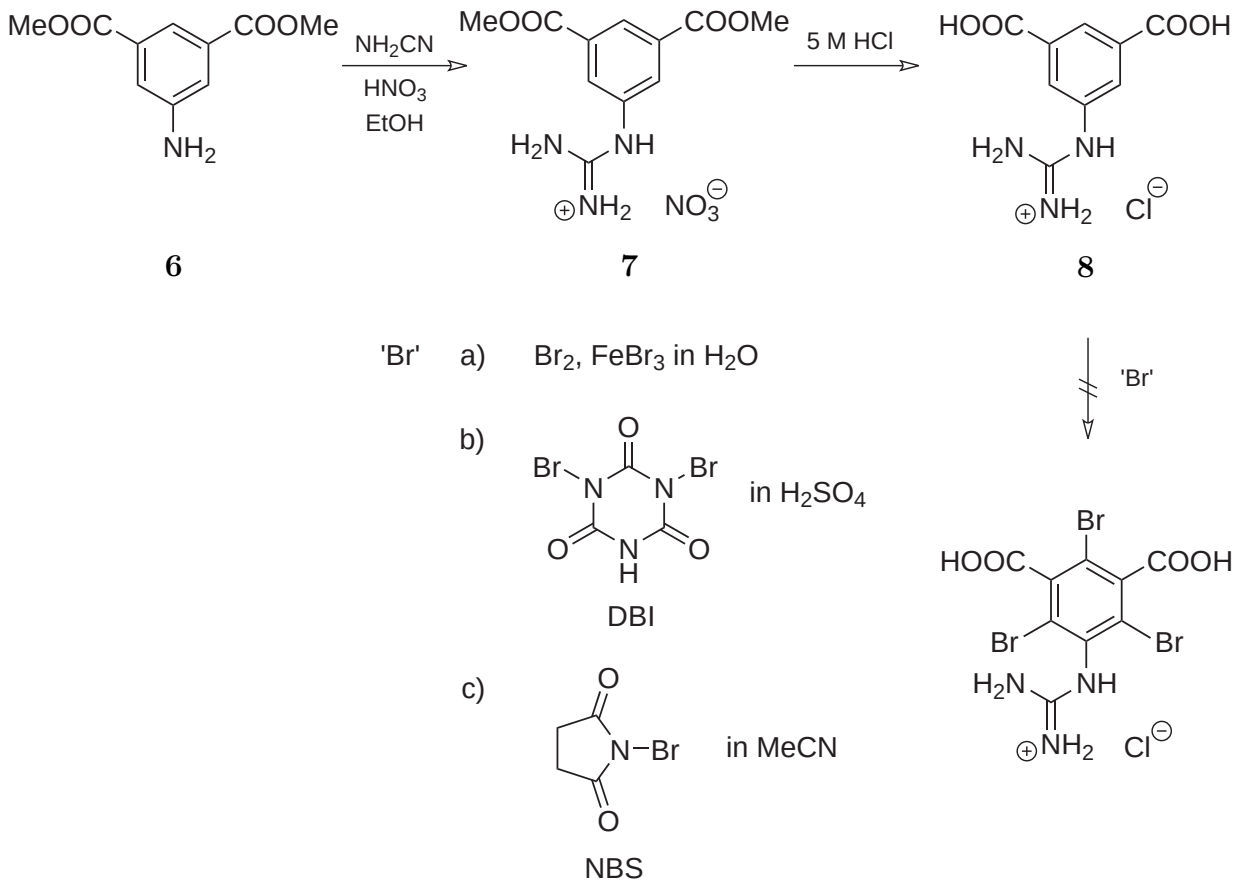
To improve the binding of the amino group, different strategies were tested to obtain the guanidinium derivative. A direct guanidation of B3C was not successful since steric hindrance of the neighbouring bromine prevents the reaction of cyanamide with the amino functionality.



Scheme 4.4: Direct synthesis of the guanidinium derivative of B3C was not successful.

Additionally, the electron-withdrawing power of the three bromine atoms and two carboxyl groups decreases the reaction rate. Guanidation with 3,5-dimethyl-1-pyrazolylformamidinium nitrate did not give any product either. The carboxyl groups may undergo side reactions, therefore they should be protected (see below).

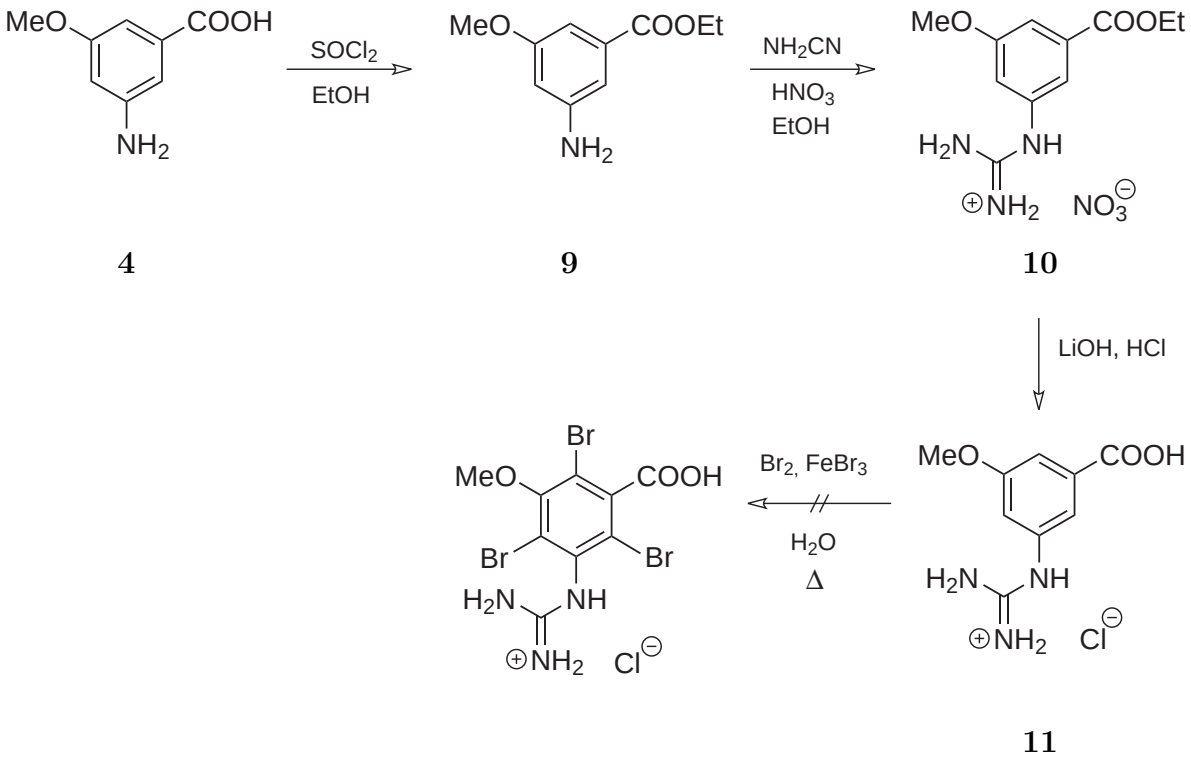
Therefore the guanidation reaction was carried out as the first step (Scheme 4.5) with the methyl ester **6**, according to literature procedures (Tavares *et al.*, 2004), followed by deprotection of the carboxyl groups to afford **8**. Different bromination reagents were tested for the bromination of the guanidinium derivative (Fig. 4.5). However bromination of **8** did not afford any product due to the deactivating effect of two carboxyl and one guanidinium group on the benzene ring.



Scheme 4.5: Reaction of **8** with several bromination agents did not yield any product. Bromination was tried with **a)** elemental bromine in aqueous solution with iron(III)bromide as catalyst, **b)** dibromoisocyanuric acid (DBI) in concentrated sulfuric acid, **c)** *N*-bromosuccinimide (NBS) in acetonitrile.

In dibromoisocyanuric acid (DBI) and *N*-bromosuccinimide (NBS) the bromine-nitrogen bond is polarised and the bromine atoms carry a partial positive charge. These reagents are well suited for aromatic electrophilic substitution reactions. Especially the reaction conditions with DBI in concentrated sulfuric acid are suitable for perbromination of deactivated substrates (Gottardi, 1969). However the substrate **8** is not reactive enough for three-fold bromination, even at these harsh conditions; no product formation was detected. High water solubility of the starting material **8** and possibly the product render work-up very difficult. Additionally, the guanidinium group is likely to be cleaved off at these harsh conditions.

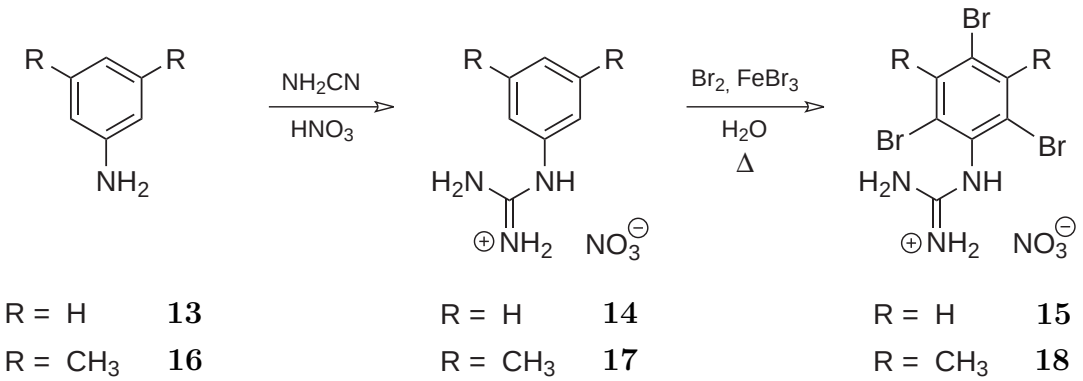
Since the highly deactivated compound **8** could not be brominated, a more reactive precursor was chosen (Scheme 4.6). First, the carboxyl group in **4** was protected with thionylchloride in ethanol as ethyl ester **9**. Reaction with cyanamide followed by deprotection of **10** yielded the guanidinium chloride salt **11**. Surprisingly, **11** is not well soluble at neutral pH: heavy precipitate forms when the basic lithium carboxylate solution is brought to neutral pH by adding hydrochloric acid.



Scheme 4.6: Reaction of **11** with elemental bromine and iron(III)bromide as catalyst did not yield tribrominated product, even in reaction conditions under reflux.

The reactivity of **11** is also not high enough for perbromination; complete bromination was not achieved. Additionally, the product from the bromination reaction is not soluble in aqueous solution at neutral pH, making it a bad candidate for heavy-atom derivatisation of macromolecules. Bromination with NBS or DBI was not considered since it is prone to either side reactions or cleavage of the guanidinium group.

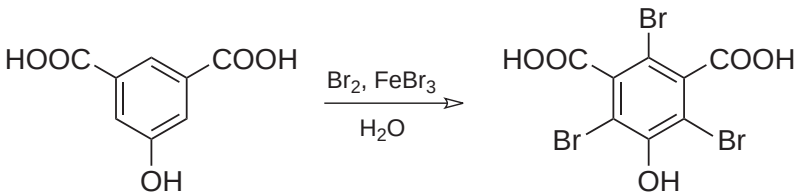
Since the reactivity of the benzene ring determines whether the tribrominated species can be obtained, substrates with a higher reactivity were chosen, i.e. compounds carrying no deactivating groups or even having activating substituents.



Scheme 4.7: Synthesis of the tribrominated guanidinium derivatives **15** and **18**.

Reaction of aniline **13** and 3,5-dimethylaniline **16** with cyanamide preceded smoothly and tribrominated compounds could be obtained (reflux necessary). Compounds **15** and **18** are however only soluble at acidic pH values in aqueous solution, rendering these compounds not suitable for heavy-atom derivatisation of macromolecules in aqueous media.

To investigate the effect of one functional group on interaction with proteins the amino group was exchanged with a hydroxyl group. The amino group in I3C and B3C is not protonated at pH values used for crystallisation and derivatisation of biological macromolecules. It carries two hydrogen atoms and one electron lone pair for hydrogen bonding. In contrast the hydroxyl group carries two electron lone pairs at the oxygen atom that are suitable as acceptors for hydrogen bonds, but only one hydrogen atom. The bromination of **19** afforded 2,4,6-tribromo-5-hydroxyisophthalic acid **20** (hereafter referred to as B3O).

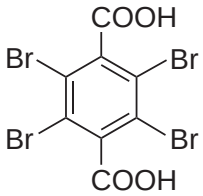


19 **20**

B3O

Scheme 4.8: Synthesis of 2,4,6-tribromo-5-hydroxyisophthalic acid, **B3O**.

Another brominated compound that can be used for experimental phasing is tetrabromoterephthalic acid **21** (hereafter referred to as B4C).



21

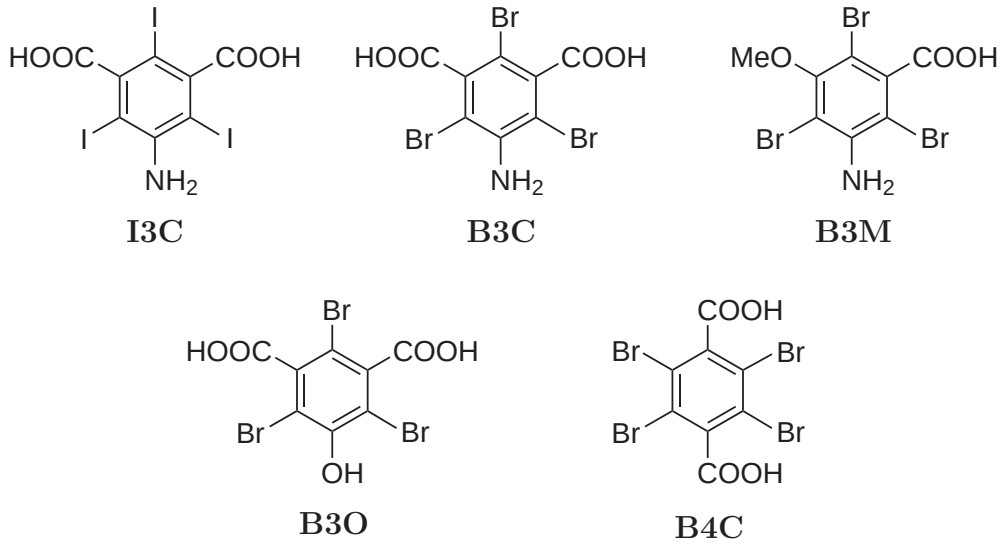
B4C

Scheme 4.9: Tetrabromoterephthalic acid, **B4C**.

B4C is available commercially and may be obtained from different chemical suppliers (CAS number 5411-70-1).

4.4 Conclusion

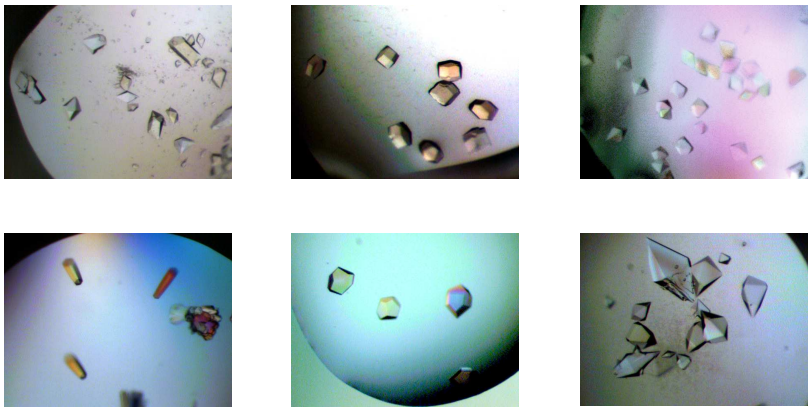
Five compounds were obtained that are suitable for derivatisation of biological macromolecules. I3C, B3C, B3M, B3O and B4C and their use for experimental phasing are discussed in detail in chapters 7 to 10.

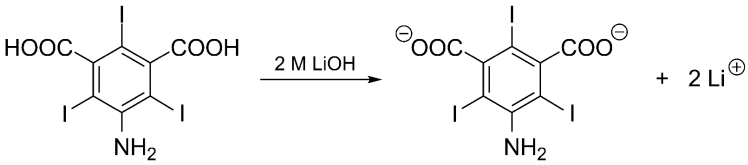


Scheme 4.10: **I3C**, **B3C**, **B3M**, **B3O** and **B4C**, five compounds used for protein derivatisation and experimental phasing.

Chapter 5

Crystallisation and heavy-atom derivatisation





Summary

Test proteins were crystallised and derivatised with the phasing tools I3C, B3C, B3M, B3O and B4C *via* soaking or co-crystallisation. To obtain a stock solution at high concentration for heavy-atom derivatisation the phasing tools were dissolved by adding double equimolar amounts of base.

5.1 Crystallisation of small molecules

The compounds I3C, B3C, B3M, B3O and B4C were dissolved in appropriate solvents to obtain a nearly saturated solution. All solutions were filtered prior to setting up the crystallisation trials. For details see Appendix E: Experimental data for small molecules.

5.2 General remarks on protein crystallisation and heavy-atom derivatisation

Protein samples were obtained from *Sigma Aldrich* (unless otherwise stated) as lyophilised powder and were used as received. Protein crystals were obtained by variations of the vapour diffusion method, either as sitting or hanging drops. Crystals for thaumatin (Ko *et al.*, 1994), porcine pancreatic elastase (Weiss *et al.*, 2002), ribonuclease A (Mueller-Dieckmann *et al.*, 2007) and thermolysin (Mueller-Dieckmann *et al.*, 2007) were obtained in crystallisation conditions similar to the published procedures.

For heavy-atom derivatisation a stock solution with a high concentration of the phasing tool is desirable, especially for soaking experiments. To obtain a high concentration in the stock solution, the compounds cannot be dissolved directly in water since they have a poor solubility. The two carboxyl groups are deprotonated with double equimolar amounts of aqueous base to obtain a salt solution (Fig. 5.1). For B3M, having only one carboxyl group, equimolar amounts of base suffice.

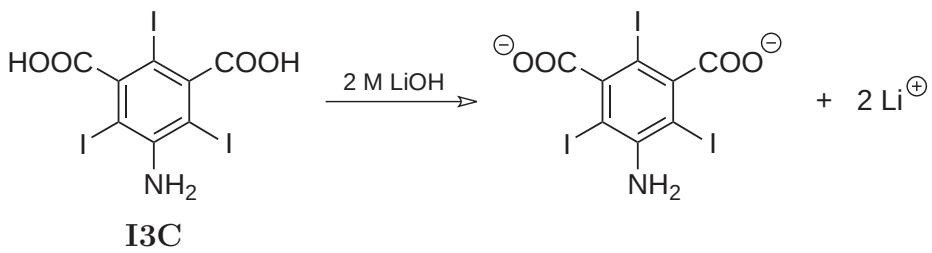


Figure 5.1: Deprotonation of I3C with lithium hydroxide solution yields the highly soluble lithium I3C salt.

In case of I3C, only lithium hydroxide solution (LiOH) as a base produces an I3C salt with high solubility. If sodium or potassium hydroxide solution or bases such as triethylamine or ammonia are used the resulting salt has limited solubility. Lithium cations as counterions have a high hydration energy with the result that the solvation energy exceeds the lattice energy of the salt formed with the doubly charged anion of I3C, so the salt is highly soluble. For co-crystallisation experiments lower concentrations of I3C are sufficient; therefore sodium hydroxide was also used as a base (see lysozyme).

Table 5.1: Preparation of the stock solutions.

	I3C	B3C	B3M	B3O	B4C
molecular weight [g mol^{-1}]	576.84	417.85	403.85	418.82	481.72
solid material [mg]	120	84	81	84	48
2 M LiOH [µL]	200	200	100	200	100
water [µL]	-	-	100	-	100
concentration [M]	1.0	1.0	1.0	1.0	0.5

In case of B3C, sodium hydroxide solution gives satisfactory results, possibly because the smaller dianion B3C has a higher solvation energy. Nevertheless, unless otherwise indicated, lithium hydroxide solution was used for the bromine compounds as well. The pK_A values of the two carboxyl groups, e.g. in B3C, could not be determined experimentally by potentiometric titration, probably due to the fact that the molecule forms a zwitterion in solution; the pK_A values of the non-brominated compound have not been reported in the literature.

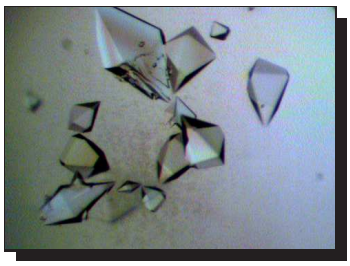
The lithium B4C salt has a lower solubility since only two hydrophilic groups are present. Therefore only 0.5 M B4C solution can be obtained.

The stock solutions were prepared by adding lithium hydroxide (or sodium hydroxide) solution (and water) to the solid material, according to Tab. 5.1. The pH of the stock solution was usually in the range of 7 to 8 (and was adjusted with small amounts of 1 M hydrochloric acid or lithium hydroxide solution if necessary). The solutions were filtered with a 0.2 µm syringe filter and stored under exclusion of light. Solutions older than two weeks were discarded.

5.3 Soaking

The quick soak procedure was carried out to incorporate the phasing tools, with soaking times ranging from 10 seconds to several minutes.

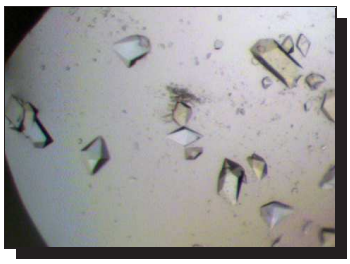
Thaumatin with I3C



Thaumatin (207 residues, 22.2 kDa) was crystallised by mixing 2 µL of 40 mg mL^{-1} protein solution with an equal volume of precipitant solution, which contained 0.05 M ADA (pH = 6.8) and 0.8 M potassium sodium tartrate. Crystals were grown at 293 K using the sitting-drop vapour diffusion method. Crystals appeared after a few days. A quick soak (2 min) was performed with a solution containing 0.5 M lithium I3C, 0.05 M ADA (pH = 6.8) and 0.8 M potassium sodium tartrate. The crystal was back-soaked in a solution containing 30% (v/v) glycerol,

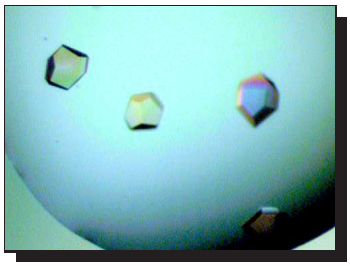
0.05 M ADA (pH = 6.8) and 0.8 M potassium sodium tartrate. The crystal was flash-cooled in liquid nitrogen.

Porcine pancreatic elastase (PPE) with I3C



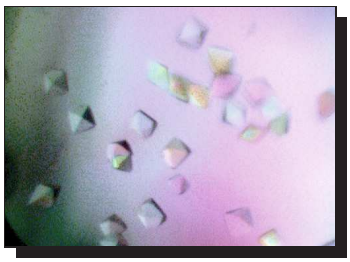
Porcine pancreatic elastase (240 residues, 25.9 kDa, Boehringer Mannheim) was crystallised at 277 K by mixing 2 μL of 40 mg mL^{-1} protein solution with an equal volume of precipitant solution containing 0.6 M sodium sulfate and 0.1 M HEPES (pH = 8.0). Crystals were grown using the hanging-drop vapour diffusion method. Crystals appeared after several weeks. A gradient soak was performed to incorporate I3C into the crystal: starting with a solution containing 0.15 M lithium I3C solution, 0.1 M HEPES (pH = 8.0) and 0.6 M sodium sulfate (2 min) and continuing with a solution with similar composition, but containing 0.25 M lithium I3C solution (5 min). The crystal was transferred to a solution containing 30% (v/v) glycerol, 0.25 M lithium I3C solution, 0.1 M HEPES (pH = 8.0) and 0.6 M sodium sulfate (10 s). The crystal was flash-cooled in liquid nitrogen.

Ribonuclease A with I3C



Ribonuclease A (124 residues, 13.7 kDa) was crystallised at 293 K by mixing 2.0 μL of 10 mg mL^{-1} protein solution with an equal volume of precipitant solution, which contained 1.4 M ammonium sulfate, 2 M sodium chloride and 0.1 M sodium acetate (pH = 5.4). Crystals were grown using the hanging-drop vapour diffusion method. Crystals appeared within a few days. A quick soak (30 s) was performed with a solution containing 0.5 M lithium I3C, obtained by adding 2 μL of 1 M lithium I3C to 2 μL reservoir solution. The crystal was back-soaked in a solution containing 30% (v/v) glycerol, 1.4 M ammonium sulfate, 2 M sodium chloride and 0.1 M sodium acetate (pH = 5.4). The crystal was flash-cooled in liquid nitrogen.

Proteinase K with I3C and B3C



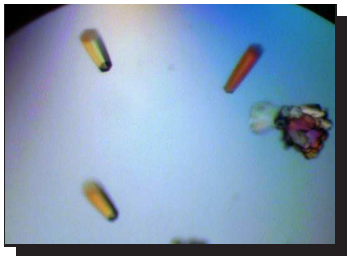
Proteinase K (279 residues, 28.9 kDa) was crystallised at 293 K by mixing 2.5 μL of 20 mg mL^{-1} protein solution with an equal volume of precipitant, containing 0.1 M Tris (pH = 7.2) and 1.28 M ammonium sulfate. Crystals were grown using the sitting-drop vapour diffusion method. Crystals appeared within one week. Protein crystals were soaked for about 10 s in 0.5 M B3C or I3C solution, respectively, which also contained 0.1 M Tris (pH = 7.2) and 1.28 M ammonium

sulfate. The crystals were back-soaked for 5 s in the cryo solution of 30% (v/v) glycerol containing 0.1 M Tris (pH = 7.2) and 1.28 M ammonium sulfate, but no heavy-atom compound. The crystals were flash-cooled in liquid nitrogen.

Thaumatin with B3C, B3M, B3O and B4C

Thaumatin was crystallised as for the I3C derivative. A quick soak (1 min) was performed with a solution containing 0.5 M lithium B3C, obtained by adding 2 μL of 1 M lithium B3C to 2 μL reservoir solution. The same was done for B3M and B3O. The soak with B4C was carried out with a final concentration of 0.25 M lithium B4C. The crystals were back-soaked in a solution containing 30% (v/v) glycerol, 0.5 M ADA (pH = 6.8) and 0.8 M potassium sodium tartrate. The crystals were flash-cooled in liquid nitrogen.

Thermolysin with B4C

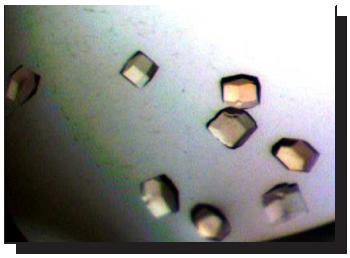


Thermolysin (316 residues, 34.4 kDa) was dissolved to a final concentration of 100 mg mL^{-1} in a solution containing 50 mM MES (pH = 6.4) and 45% (v/v) dimethylsulfoxide (DMSO). Undissolved protein was removed by centrifugation for 10 min at 14 000*g* at 277 K. Crystals were grown using the hanging-drop vapour diffusion method. 1.0 μL of protein solution were mixed with an equal volume of precipitant, containing 50 mM MES (pH = 6.4), 1 M NaCl and 45% (v/v) DMSO. The drops were equilibrated over wells of 35% (v/v) saturated ammonium sulfate solution. Crystals appeared within a few days. Protein crystals were soaked for about 25 s in 0.25 M B4C solution, obtained by adding 2 μL of 0.5 M lithium B4C to 2 μL reservoir solution. The crystals were back-soaked for 5 s in the cryo solution, containing 25 mM MES (pH = 6.4), 250 mM NaCl, 20% (v/v) DMSO and 5% (v/v) glycerol. The crystals were then flash-cooled in liquid nitrogen.

5.4 Co-crystallisation

I3C and B3C were also incorporated into biological macromolecules *via* co-crystallisation. The compounds were added either to the protein solution before crystallisation set-up or added to the reservoir solution. It is generally advised to use at least a ten-fold excess of I3C or B3C for incorporation via co-crystallisation.

Lysozyme with I3C



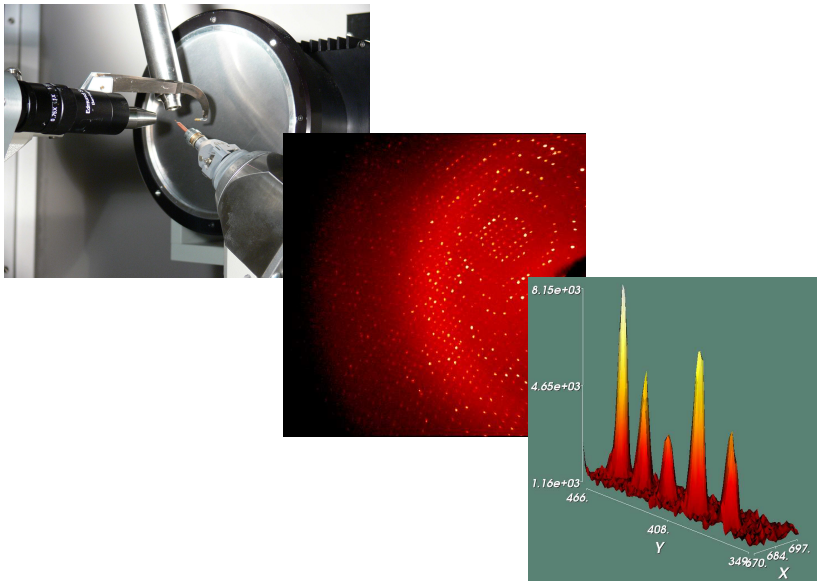
Hen egg white lysozyme (129 residues, 14.3 kDa) was crystallised at 293 K by mixing 5 μL of 20 mg mL^{-1} protein solution with an equal volume of precipitant, containing 0.1 M HEPES (pH = 7.0), 15%(w/v) PEG 3350 and 8 mM sodium I3C solution (about three-fold excess of I3C with respect to the protein concentration). Crystals containing I3C appeared within one week. The crystal was soaked for 5 s in cryo solution containing 15% (v/v) PEG-400, 0.1 M HEPES (pH = 7.0) and 15%(w/v) PEG 3350. The crystal was flash-cooled in liquid nitrogen.

Trypsin with B3C

Bovine trypsin (223 residues, 23.3 kDa, Sigma Aldrich) was dissolved to a final concentration of 60 mg mL^{-1} in a solution containing 10 mg mL^{-1} benzamidine (inhibitor to prevent auto-cleavage), 3 mM $CaCl_2$, 0.02% NaN_3 and 30 mM sodium B3C (about ten-fold excess of B3C with respect to the protein concentration). Crystals were grown using the sitting-drop vapour diffusion method. The Wizard Screen I (*Emerald BioSystems*) was set up with a *Tecan* dispenser (100 μL reservoir volume). 0.1 μL of the protein solution was mixed with an equal volume of the reservoir solution using a *Moscito* robot. The plate was stored at 293 K and crystals appeared within one week in several conditions. Data were collected from a crystal taken from a drop with the reservoir solution containing 30% (w/v) PEG-3000 and 0.1 M CHES (pH = 9.5). The crystal was flash-cooled in liquid nitrogen without any further cryoprotection.

Chapter 6

Data collection, processing and refinement



Summary

Data for the small-molecule compounds were collected in-house on a Bruker SMART 6000 at CuKα. The structures were refined with *SHELXL-97*. For details on the small-molecule data and processing, please refer to Appendix D: Crystallographic data for small molecules. Restraints were generated based on these structures for the refinement of the protein derivative data sets.

Protein data were collected in-house on a Bruker SMART 6000 or a Marresearch MAR345 at CuKα and at the synchrotron (SLS, BESSY II). The heavy-atom substructures were solved with *SHELXD* and density modification was done in *SHELXE*. Initial models were obtained from *ARP/wARP* and refined with *SHELXL-97* or *REFMAC*.

6.1 Small-molecule crystallography

For details on the small-molecule data and processing, please refer to Appendix D: Crystallographic data for small molecules.

6.2 In-house data collection

Data for the elastase I3C, trypsin B3C and ribonuclease I3C data sets were recorded at 100 K in-house on a Bruker three-circle diffractometer equipped with a SMART 6000 CCD detector, a MacScience rotating anode and Incoatec Helios optics ($\lambda = 1.54178$ Å).

Data for the lysozyme I3C and thaumatin I3C data sets were collected in-house at 100 K on a Marresearch MAR345 image-plate detector using CuKα radiation ($\lambda = 1.54178$ Å) from a MacScience rotating anode X-ray generator equipped with Osmic optics.

6.3 Data collection at the synchrotron

Swiss Light Source (SLS)

The proteinase K I3C and B3C data sets were collected at 106 K at beamline X10SA (PXII) at SLS, Villigen, Switzerland. For the B3C data, a fluorescence scan was performed to locate the bromine K edge. Interestingly, the spectrum showed two peaks at the bromine edge (Fig. 8.6, p. 68). Therefore, two peak datasets were collected (13480 and 13473 eV), followed by a high-energy remote (13580 eV) and an inflection point data set (13471 eV). For details, see Tab. 8.3 (p. 68). For the I3C data, only one dataset was collected, at the same wavelength as in-house CuKα (1.5418 Å, Tab. 8.3, p. 68). Data collection for all datasets was carried out with 0.5 s exposure per frame and 1° frame width.

Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY II)

The thaumatin B3C, B3M, B3O and B4C and the thermolysin B4C data sets were collected at beamline 14.2 at BESSY II, Berlin. For the thaumatin B3C and B4C data, a fluorescence scan was performed to locate the bromine K edge (see Fig. 9.7, p. 81 and Fig. 10.5, p. 91). For the B3M and B3O crystal, data sets were collected at the high-energy remote (13580 eV). For the thaumatin B3C crystal, data sets were collected at peak (13478 eV), inflection point (13470 eV) and high-energy remote (13578 eV). For details, see Tab. 9.3 (p. 80). For the thaumatin B4C crystal, data sets were collected at peak (13481 eV) and inflection point (13469 eV). For details, see Tab. 10.2 (p. 90). For the thermolysin B4C crystal, a data set was collected at high-energy remote (13478 eV). For details, see Tab. 10.2 (p. 90). Data collection for all data sets was carried out with exposure times ranging from 3.2 to 5.2 s per frame and 1° frame width.

6.4 Integration and scaling

Data collected on the SMART 6000 were integrated with *SAINT* (Bruker, 2009). Protein data collected on the MAR345 or at the synchrotron site were integrated with *XDS* (Kabsch, 1993). Absorption correction (semi-empirical from equivalents) and scaling were carried out with *SADABS* (Sheldrick, 2009).

6.5 Experimental phasing

Unless otherwise stated, protein data sets were processed as follows. Data were prepared with *XPREP* (Bruker, 2007), i.e. extracting the anomalous data and selecting a proper resolution cut-off for the anomalous data. Substructure solution was carried out with *SHELXD* (Usón and Sheldrick, 1999; Schneider and Sheldrick, 2002; Sheldrick, 2008a). Density modification was done with *SHELXE* (Sheldrick, 2002, 2008a), in some cases in conjunction with the autotracing feature (Sheldrick, 2010).

6.6 Model building, refinement and validation

The experimental electron density was transferred to *ARP/wARP* (Langer *et al.*, 2008; Murshudov *et al.*, 1997) for main chain tracing, sequence docking, side chain fitting and addition of water molecules. The first model from *ARP/wARP* was transferred to *SHELX* and refined with *SHELXL-97* (Sheldrick, 2008a). Water molecules were updated with *SHELXPRO* (Sheldrick, 2008a) and validated with *COOT* (Emsley *et al.*, 2010). Refined structures were manually inspected and improved in *COOT*, e.g. adding alternate conformations for side chains. Ions binding to the surface of the protein were differentiated from water molecules according to the anomalous density map (where applicable) calculated with *SHELXE* (Sheldrick, 2002), their coordination environment and thermal displacement parameters.

Refinement in *SHELXL-97* was carried out by conjugate-gradient least-squares methods on F^2, with 5% of the data set aside (randomly chosen) for calculation of the R_{free} factor (Brünger, 1992). Restraints for the bond distances in the proteins were used as provided by *SHELXPRO*, which are based on Engh and Huber (1991). The restraints for the phasing tools were derived from the small-molecule structures as described in this work. The bond lengths of the carboxyl group were adjusted with *HIMP* in *XP* (Bruker, 2008) to fit the deprotonated form. Restraints for other ligands (buffer compounds, precipitants, etc.) were obtained from the *PRODRG* server (van Aalten *et al.*, 1996).

The ribonuclease A I3C and thermolysin B4C data were first refined with *SHELXL-97* and then transferred to *REFMAC* (Murshudov *et al.*, 1997) for further refinement. Conversion of the reflection files to different formats was carried out within *CPP4i* (Collaborative Computational Project Number 4, 1994).

Validation of the refined structures was carried out in *COOT*, e.g. validation of water molecules or temperature factor variance, and with the Molprobity server (Chen *et al.*, 2010).

Figures were prepared with PyMOL (DeLano, 2010) and *GNUPLOT* (Gnuplot, 2010). The anomalous electron density (SAD density or MAD density) was calculated in *SHELXE* using the final refined model and the anomalous data. Anomalous data statistics were prepared with *HKL2MAP* (Pape and Schneider, 2004), which uses *SHELXC* (Sheldrick, 2008a) for data preparation.

6.7 Radiation damage

To investigate radiation damage on B3C, a series of experiments was carried out at 106 K at beamline X10SA (PXII) at the SLS, Villigen, Switzerland. The same crystal that had been used for the proteinase K B3C MAD experiment was exposed to full photon flux (1.82×10^{12} photons s^{-1} at 13473 eV), followed by data collection at the same energy (corresponding to wavelength at peak2, Table 9.3). Full beam exposure was maintained for 2 s and was increased to 30 s and 60 s for the last two burn exposures, respectively. Six datasets were collected in total from the same crystal following this procedure.

The program *RADDOSE* (Murray *et al.*, 2004) was used for dose calculations. Heavy-atom content of the unit cell was specified, including sulfur atoms from the protein and chloride and sulfur atoms from the buffer. The absorbed dose for the crystal is given in MGy (Gray = J kg^{-1}).

Part III

Results and discussion

Part III - Results and discussion

Outline

The following chapters discuss experimental phasing of proteins with the halogen compounds introduced in Part II - Materials and methods (hereafter referred to as phasing tools).

The **magic triangle I3C** contains three iodine atoms for SAD or SIRAS phasing and is discussed in chapter 7. The analogous bromine compound, the **MAD triangle B3C**, was employed for MAD and SAD phasing (chapter 8). Two other bromine compounds, **B3M** and **B3O**, carrying different functional groups, were also utilised for experimental phasing and are discussed in chapter 9. The **MAD tetragon B4C** contains four bromine atoms and was utilised for SAD and MAD phasing (chapter 10).

The molecular structure of each phasing tool was established by X-ray crystallography. The molecular structures were used to generate bond length and angle restraints for the refinement of the derivatised protein structures.

X-ray absorption may lead to **radiation damage** in the crystal, especially damage to the heavy-atom sites. The effect of radiation damage on the bromine-carbon bonds was investigated for B3C (chapter 8).

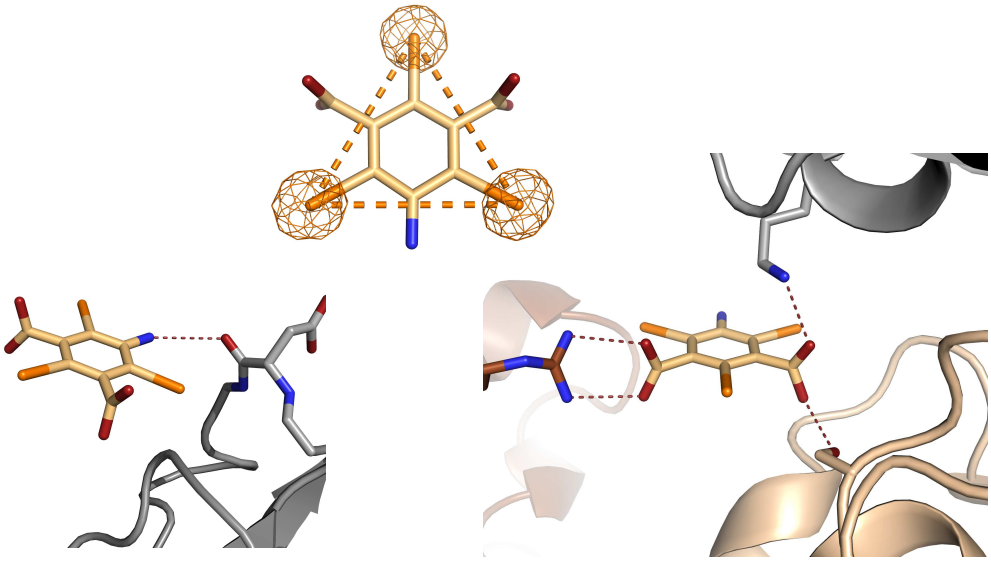
The structural information about the **triangular arrangement** of the heavy atoms provides an independent test for substructure validation. A computer script was developed that **searches for triangles** within the heavy-atom coordinate file to facilitate substructure validation. Three examples for difficult structures are presented where the triangular arrangement of the heavy-atoms was crucial (chapter 11).

The phasing tools bind to proteins through **hydrogen bonds** and **halogen bonds**. The latter can be rationalised by a positive electrostatic potential at the halogen atom.

In this study the phasing tools were used to derivatise **test proteins**, which were readily available. Further information on the structural details and biological function of these proteins may be found in the current literature: hen egg-white lysozyme (Blake *et al.*, 1965; Vocadlo *et al.*, 2001), thaumatin (de Vos *et al.*, 1985; Ko *et al.*, 1994), porcine pancreatic elastase (Sawyer *et al.*, 1978; Fodor *et al.*, 2006), proteinase K (Pal *et al.*, 1994; Saxena *et al.*, 1996), bovine trypsin (Recacha *et al.*, 1999; Polgar, 2005), ribonuclease A (Borkakoti *et al.*, 1982; Schultz *et al.*, 1998) and thermolysin (Tajima *et al.*, 1976; Holmes and Matthews, 1982). A brief discussion of **novel protein structures** that were solved with the magic triangle I3C can be found in Appendix A.

Chapter 7

I3C - The magic triangle



Summary

The magic triangle I3C has three functional groups for interaction with biological macromolecules and three iodine atoms for experimental phasing. The molecular structure of I3C was determined by X-ray crystallography.

I3C was incorporated into three test proteins *via* soaking or co-crystallisation. Data were collected in-house and SAD phasing was carried out. I3C binds to the proteins mainly *via* hydrogen bonds to side chains and the main chain.

7.1 Introduction

The compound 5-amino-2,4,6-triiodoisophthalic acid (I3C, Fig. 7.1a) is the first representative of a new class of compounds for heavy-atom derivatisation of biological macromolecules. It combines functional groups for interaction with proteins and nucleic acids with heavy atoms for experimental phasing. Three iodine atoms per molecule provide a strong anomalous signal for phasing, even at in-house X-ray sources: the anomalous signal f'' of iodine is 6.85 e at the CuKα wavelength.

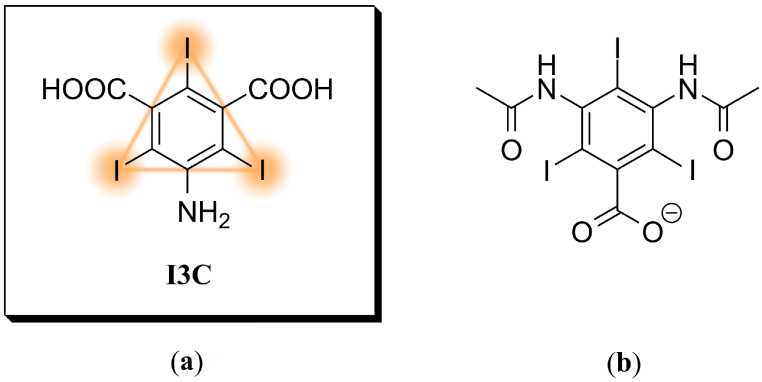


Figure 7.1: Iodinated compounds. **(a)** 5-Amino-2,4,6-triiodoisophthalic (I3C) is the first representative of a new class of compounds suitable for experimental phasing. **(b)** The X-ray contrast agent diatrizoate. The 1,3,5-triiodobenzene core is the basis of many iodine-containing contrast agents.

In comparison to other heavy atom reagents I3C has a low toxicity. Its derivatives have been employed as X-ray contrast reagents in medical diagnosis for almost 100 years. The ionic monomer diatrizoate (Fig. 7.1b) was one of the first compounds used (see Tonnessen *et al.*, 1996, for the X-ray structure), although nowadays nonionic compounds are preferred due to their reduced osmolality (Almen, 1969; Dawson *et al.*, 1999). The 1,3,5-triiodobenzene core has been the basis of many iodine-containing contrast agents (Yu and Watson, 1999).

The three iodine atoms are arranged in an equilateral triangle in I3C (Fig. 7.1a). This geometrical arrangement is easily identified in the heavy-atom substructure. The molecular structure of I3C was determined by single-crystal X-ray analysis (Beck and Sheldrick, 2008) and used for the generation of bond length and bond angle restraints. These are required for the correct refinement of small-molecule compounds in macromolecular structures.

7.2 I3C crystal structure

In the crystal environment I3C has no internal crystallographic symmetry and crystallises in the space group *Pbca* with one molecule in the asymmetric unit. The bond lengths and angles fall within normal ranges. A displacement ellipsoid plot is shown in Fig. 7.2. I3C crystallises as a monohydrate due to water impurities in the crystallisation solution.

The three iodine atoms in I3C form an equilateral triangle with a side of 6.0 Å [cf. I1−I2 6.0425(6), I2−I3 6.0388(7), I1−I3 6.0267(7) Å]. In I3C the carboxyl groups are per-

pendicular to the ring and clear indications for protonated carboxyl groups can be found. In both carboxyl groups, one C−O bond is significantly shorter than the other one, indicating one double bond and one single bond [cf. C7−O9 1.214(6) *vs.* C7−O8 1.308(6) Å and C10−O12 1.223(6) *vs.* C10−O11 1.299(6) Å]. The unsubstituted 5-aminoisophthalic acid crystallises as a zwitterion, with one carboxylate and an -NH_3^+ group (Dobson and Gerkin, 1998). The carboxyl groups are in plane with the aromatic ring. The negative charge of the carboxylate group can therefore be delocalised across the π-system, leading to an increase in acid strength compared to I3C.

In the crystal structure I3C forms hydrogen bonds with all potential donors and the lattice water molecule (Fig. 7.3, Tab. 7.1); however, the interaction between N13 and O12 is slightly weaker. Hydrogen bonds to a lattice water molecule are also present in the unsubstituted 5-aminoisophthalic acid.

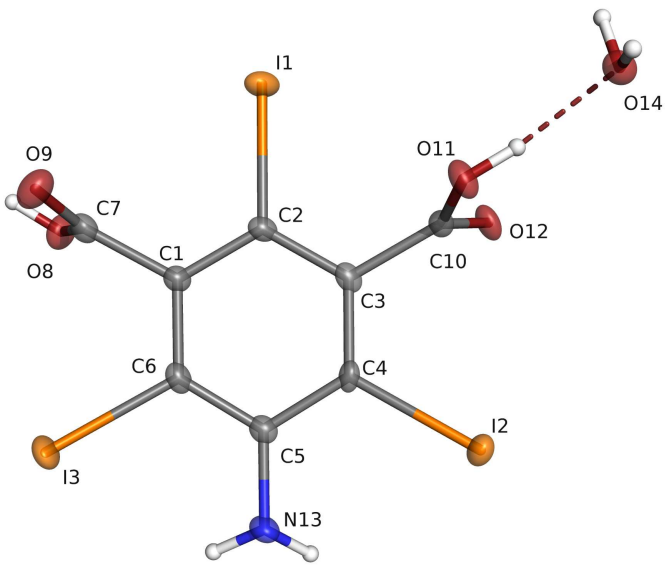


Figure 7.2: The asymmetric unit of I3C. Displacement ellipsoids are drawn at the 50% probability level and hydrogen atoms are shown as small spheres with arbitrary radii. The hydrogen bond to the lattice water molecule is shown as a dashed line.

Table 7.1: Hydrogen-bond geometry (Å, °) for I3C.

D−H$\cdots A$	D−H	H$\cdots A$	$D\cdots A$	D−H$\cdots A$
O8−H8$\cdots$O12^{i}	0.80(5)	1.90(5)	2.662(5)	161(7)
O11−H11$\cdots$O14	0.79(5)	1.75(5)	2.540(5)	173(7)
O14−H14A$\cdots$O9^{ii}	0.81(4)	1.95(4)	2.751(5)	170(6)
O14−H14B$\cdots$N13^{iii}	0.81(4)	2.05(4)	2.841(5)	166(6)
N13−H13A$\cdots$O14^{iv}	0.88(4)	2.30(5)	3.067(6)	147(5)
N13−H13B$\cdots$O12^{iv}	0.88(4)	2.68(5)	3.478(5)	152(5)

Symmetry codes: (i) $x+\frac{1}{2}, -y+\frac{3}{2}, -z; (ii) -x+\frac{5}{2}, y-\frac{1}{2}, z;$ $(iii) -x+2, y-\frac{1}{2}, -z+\frac{1}{2}; (iv) -x+\frac{3}{2}, y+\frac{1}{2}, z.$

Formation of cyclic dimers between two carboxyl groups is frequently observed for carboxylic acids in the solid state (Leiserowitz, 1976), but not found in I3C. Due to the large van der Waals radius for iodine (1.98 Å; Bondi, 1964), cyclic dimeric hydrogen bonds

of two carboxyl groups are not favoured for I3C: The iodine atoms of two molecules would come in close proximity and repel each other.

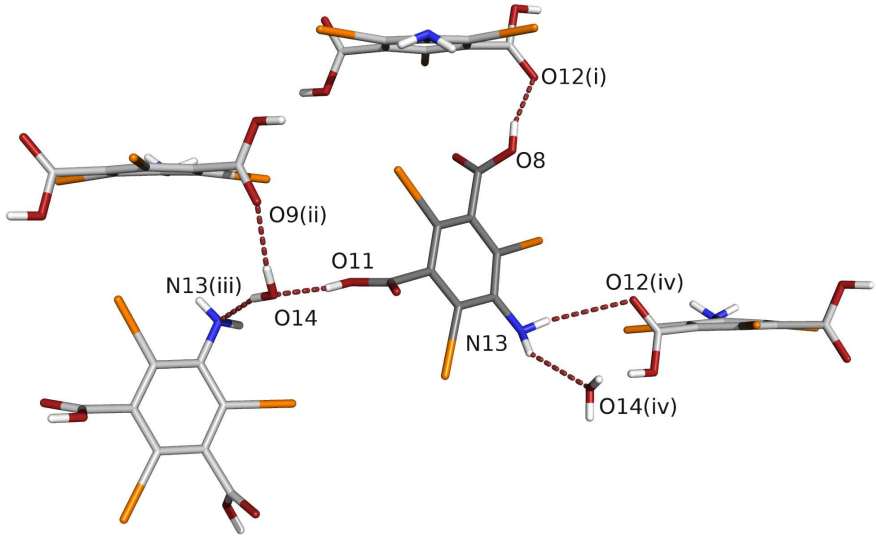


Figure 7.3: Hydrogen bonding of I3C. The hydrogen bonds for hydrogen donor atoms of the central I3C are depicted as dashed lines. Symmetry-equivalent molecules are shown in lighter shade. For the symmetry operators, please refer to Tab. 7.1.

In the crystal lattice, the molecules are positioned perpendicular to each other, showing no π-π interactions of the benzene rings. The packing is different from that observed in the unsubstituted acid where the substituents are missing (Dobson and Gerkin, 1998) and the molecules are packed in layers showing π-π interaction.

The molecular arrangement of I3C in the crystal lattice shows that hydrogen bonds are the main type of interaction. This becomes important when I3C is used as a ligand for heavy-atom derivatisation of proteins.

7.3 Experimental phasing with I3C

In this study the incorporation of I3C into three protein samples was investigated. Stock solutions of I3C were obtained by dissolving the solid material in water, adding double equimolar amounts of aqueous base to fully deprotonate the carboxyl groups. I3C was incorporated into lysozyme crystals *via* co-crystallisation. A quick soak was carried out with thaumatin and elastase crystals. SAD phasing was carried out with data collected in-house at CuKα ($\lambda = 1.54178$ Å).

Substructure solution and data analysis

The substructure solution with *SHELXD* using Patterson seeding and dual space direct methods followed by density modification with *SHELXE* resulted in high-quality starting phases (Tab. 7.2).

The SAD data sets show a strong anomalous signal. This is also indicated by the ratio of R_{anom} to $R_{\text{p.i.m.}}$, a quality measure for the anomalous signal (Weiss, 2001; Tab 7.2). Investigations of the data show that only a small multiplicity of data is required to find the

heavy-atom positions (Tab. 7.2), although of course the ratio of R_{anom} to $R_{\text{p.i.m.}}$ increases with higher multiplicity.

Table 7.2: Data collection details for lysozyme, thaumatin and elastase. For lysozyme and thaumatin a comparison of the different data sets shows that a much lower multiplicity is sufficient to solve the heavy-atom substructure. With increasing map correlation coefficient and decreasing mean phase error, more residues can be traced by automated procedures. A similar comparison for elastase was not carried out because of different detector and goniometer settings (three-circle goniometer) between the data sets. Values in parentheses denote the highest resolution shell.

a) Lysozyme

Unit-cell parameters (Å, °)	$a = b = 76.83, c = 38.87,\ \alpha = \beta = \gamma = 90$				
Space group	$P4_32_12$				
Resolution (Å)	19.4−1.55 (1.65−1.55)				
Degrees collected (°)	30	45	60	90	360
R_{merge}	0.060 (0.14)	0.063 (0.15)	0.066 (0.16)	0.068 (0.18)	0.069 (0.15)
Completeness	0.86 (0.83)	0.96 (0.93)	0.98 (0.98)	0.99 (0.98)	0.99 (0.99)
Multiplicity	2.14 (1.99)	3.23 (2.96)	4.32 (3.97)	6.50 (5.94)	26.1 (24.0)
$< I/\sigma(I) >$	10.3 (5.06)	12.0 (5.62)	13.8 (6.43)	17.1 (7.80)	38.2 (21.2)
R_{anom}	0.117 (0.238)	0.115 (0.236)	0.114 (0.229)	0.109 (0.205)	0.0974 (0.122)
$R_{\text{anom}}/R_{\text{p.i.m.}}$	3.68 (2.17)	3.61 (2.20)	3.71 (2.12)	4.31 (2.19)	7.12 (3.00)
No. sites found	10 [83%]	12 [100%]	12 [100%]	12 [100%]	12 [100%]
Mean phase error (°)	38.6	32.6	32.1	30.4	31.1
Mean map CC	0.731	0.801	0.817	0.833	0.852
No. residues built	53 [41%]	83 [64%]	105 [81%]	111[86%]	120 [93%]

b) Thaumatin

Unit-cell parameters (Å, °)	$a = b = 57.71, c = 149.58,\ \alpha = \beta = \gamma = 90$				
Space group	$P4_12_12$				
Resolution (Å)	19.7−1.73 (1.83−1.73)				
Degrees collected (°)	30	45	60	90	360
R_{merge}	0.038 (0.095)	0.041 (0.099)	0.044 (0.11)	0.046 (0.11)	0.056 (0.12)
Completeness	0.88 (0.86)	0.97 (0.95)	0.99 (0.97)	0.99 (0.97)	0.99 (0.98)
Multiplicity	2.19 (1.99)	3.29 (2.97)	4.40 (4.00)	6.65 (6.12)	26.8 (24.2)
$< I/\sigma(I) >$	16.3 (8.36)	18.7 (9.75)	20.8 (10.9)	26.0 (13.9)	47.5 (25.5)
R_{anom}	0.0774 (0.157)	0.0750 (0.143)	0.0712 (0.135)	0.0668 (0.115)	0.0620 (0.0886)
$R_{\text{anom}}/R_{\text{p.i.m.}}$	3.41 (2.04)	3.29 (1.85)	3.07 (1.88)	3.36 (1.90)	4.96 (2.69)
No. sites found	15 [100%]	15 [100%]	15 [100%]	15 [100%]	15 [100%]
Mean phase error (°)	50.2	32.5	26.7	26.0	28.4
Mean map CC	0.608	0.811	0.886	0.891	0.860
No. residues built	0 [0%]	199 [96%]	198 [96%]	199[96%]	196 [95%]

c) Elastase

Unit-cell parameters (Å, °)	$a = 50.10, b = 57.97, c = 74.40,\ \alpha = \beta = \gamma = 90$
Space group	$P2_12_12_1$
Resolution (Å)	37.2−1.60 (1.70−1.60)
Degrees collected (°)	360
R_{merge}	0.038 (0.10)
Completeness	0.99 (0.92)
Multiplicity	11.2 (2.76)
$< I/\sigma(I) >$	40.6 (8.38)
R_{anom}	0.0310 (0.109)
$R_{\text{anom}}/R_{\text{p.i.m.}}$	2.61 (1.17)
No. sites found	12 [100%]
Mean phase error (°)	35.3
Mean map CC	0.806
No. residues built	227 [95%]

For crystallographic parameters, see Appendix B: Crystallographic parameters.

The anomalous signal is strong up to high resolution for the thaumatin and lysozyme

data sets (Fig. 7.4). Therefore, no truncation of the data was necessary for substructure solution. However, the anomalous data for elastase was truncated at 2.2 Å where the anomalous signal drops below $d''/\text{sig} = 1.0$.

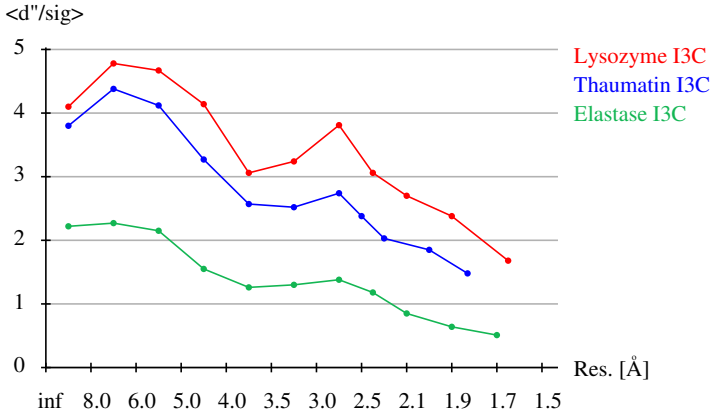


Figure 7.4: Anomalous signal for the three SAD data sets; pure noise would correspond to $d''/\text{sig} \simeq 0.798$. Especially the lysozyme and thaumatin data sets show a strong anomalous signal; the anomalous data for elastase were truncated at 2.2 Å.

Substructure solution with *SHELXD* gave large figures of merit: lysozyme I3C with CC: 47.09, CC_{weak}: 26.81; thaumatin I3C with CC: 43.56, CC_{weak}: 25.59; elastase I3C with CC: 33.86, CC_{weak}: 19.06. Since the heavy atoms form an equilateral triangle (Fig. 7.5), a successful substructure solution in *SHELXD* is readily identified when inspecting the heavy-atom positions, which facilitates structure solution.

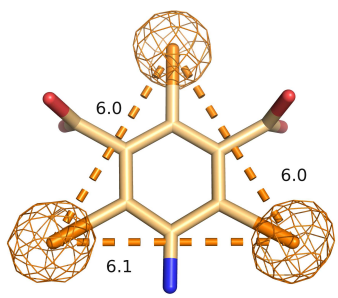


Figure 7.5: One molecule of I3C in lysozyme, anomalous electron density at 4σ (orange), distances in Å. The equilateral triangle formed by the iodine atoms is clearly visible.

I3C binding sites

Refinement was carried out with the *SHELX* suite (Tab. C.1) with the restraints for I3C generated from the small-molecule crystal structure.

The molecular structures of the proteins are similar to those previously reported (e.g. lysozyme: Weiss *et al.*, 2000; thaumatin: Ko *et al.*, 1994; elastase: Würtele *et al.*, 2000), since I3C mostly replaces solvent water molecules in the crystal lattice and binds to the surface of the proteins (Fig. 7.6).

Inspection of the I3C sites reveals several modes of interaction. The amino group of I3C forms hydrogen bonds to a main chain carbonyl oxygen (e.g. Fig. 7.7a) or to a hydrogen bond acceptor in a side chain (e.g. the oxygen atom of asparagine or glutamine). The carboxylate groups interact with side chains of the protein *via* hydrogen bonds with donor

groups in serine, lysine, tyrosine or threonine residues. The prominent interaction of the carboxylate group is the interaction with arginine (Fig. 7.7b), forming an ion pair.

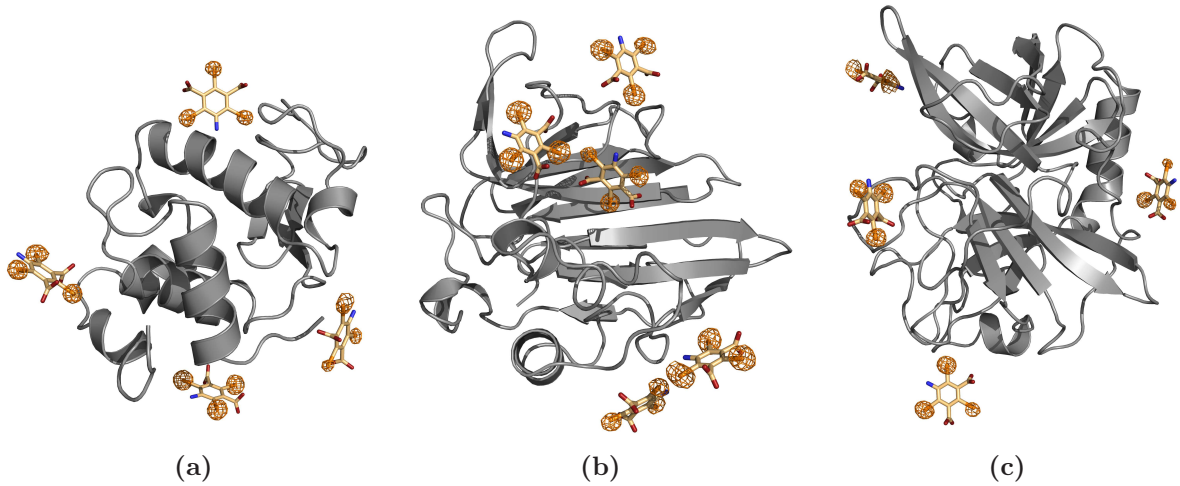


Figure 7.6: I3C in protein crystals: anomalous electron density shown around I3C at 4σ (orange). I3C molecules are found at the surface of the protein molecules. **(a)** Lysozyme with four molecules of I3C in the asymmetric unit. **(b)** Thaumatin with five molecules of I3C in the asymmetric unit. **(c)** Elastase with four molecules of I3C in the asymmetric unit.

Additional hydrogen bonds are also formed to solvent water molecules. These strong interactions lead to relatively high occupancies of the sites. The occupancies of I3C were refined with *SHELXL* to values of 0.60, 0.32, 0.32, 0.25 for lysozyme, 0.49, 0.44, 0.27, 0.26, 0.19 for thaumatin and 0.45, 0.23, 0.21, 0.16 for elastase.

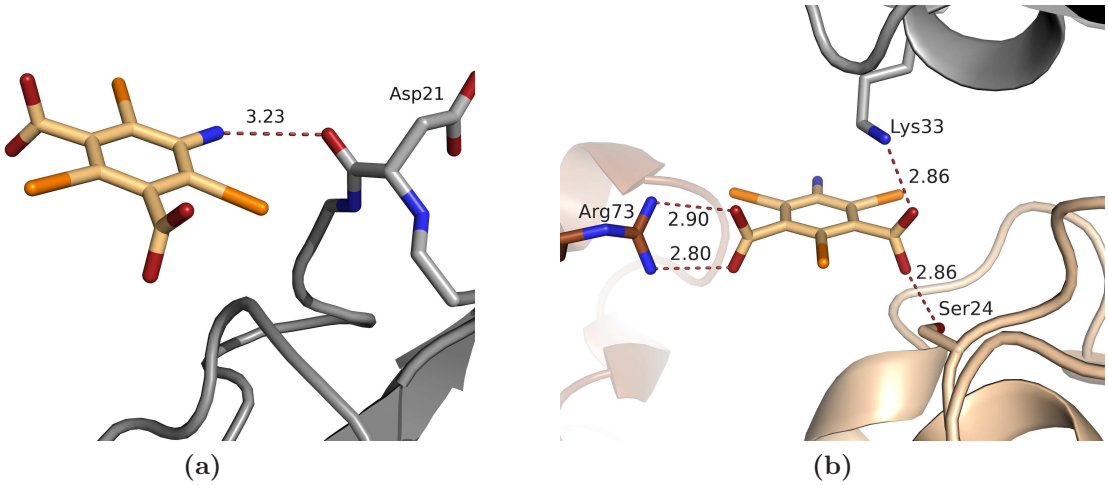


Figure 7.7: I3C interaction with proteins. **(a)** Hydrogen bonding of the amino group of I3C with the main chain carbonyl oxygen of Asp21 in thaumatin (other interactions not shown). **(b)** Interaction of I3C with three lysozyme molecules. Hydrogen bonds to Arg73, Lys33 and Ser24 are shown as dashed lines.

Interactions of the benzene ring or the iodine atoms with hydrophobic residues (e.g. phenylalanine, tryptophane) are not observed. This is probably due to the fact that the iodine atoms have a large van der Waals radius. Additionally, the carboxylate groups are

perpendicular to the benzene ring. π-π interactions are not favoured (cf. small-molecule crystal structure, p. 56) due to high steric demand of the carboxylate groups and the iodine atoms. Non-covalent contacts observed for the iodine atoms will be discussed in detail in chapter 12.

7.4 Conclusion

I3C represents a new class of compounds; it may be used for heavy-atom derivatisation for SAD or SIRAS phasing, combining an easily recognisable arrangement of three anomalous scatterers with functional groups for hydrogen bonding to a protein molecule. Low toxicity and ready commercial availability are further advantages (Beck *et al.*, 2008).

Three functional groups (two carboxylates and one amino group) of I3C interact through hydrogen bonds with the main chain as well as with the side chains of proteins. This results in a relatively high occupancy of the bound ligands. The three iodine atoms form an equilateral triangle (I$\cdots$I 6.0 Å), which is easy to recognise in the heavy-atom substructure when this compound is used as a heavy-atom derivative for macromolecular phasing. Up to date I3C has been used to solve four novel structures (see Appendix A).

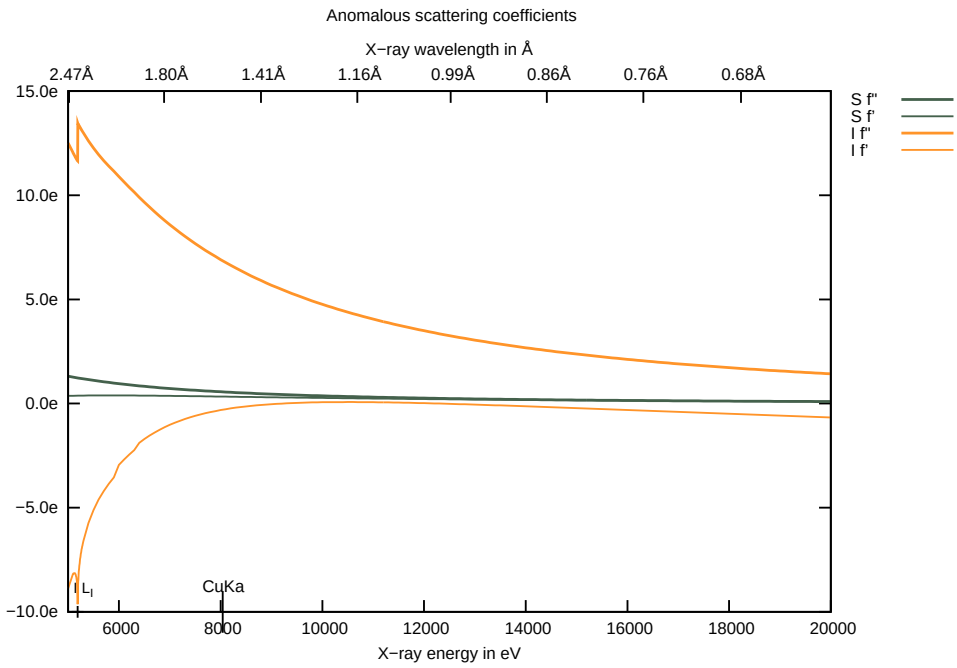


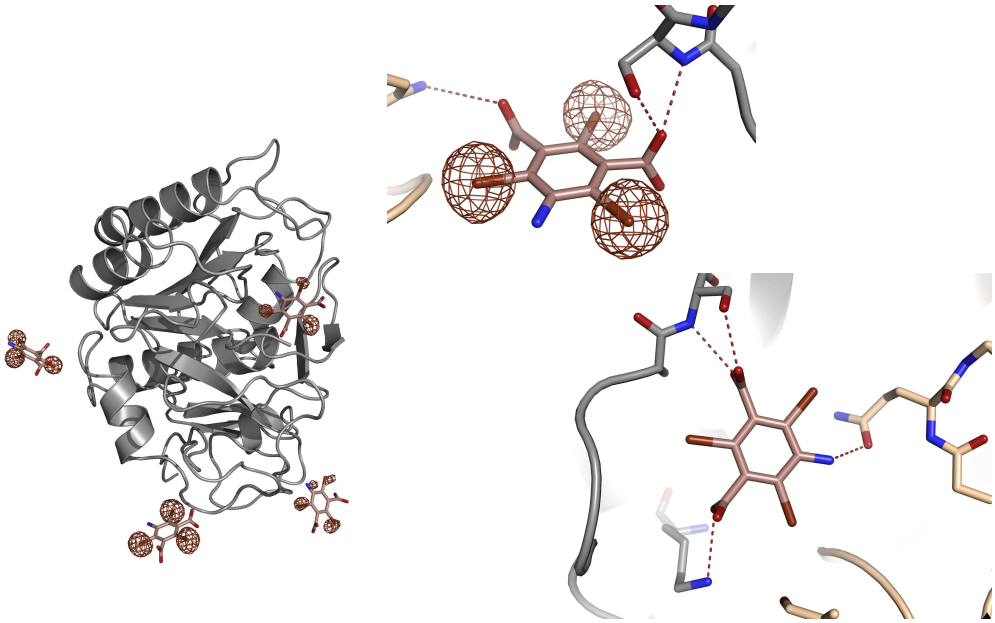
Figure 7.8: Calculated anomalous scattering coefficients for iodine and sulfur (Merrit, 2010). The anomalous signal of sulfur, an intrinsic anomalous scatterer found in cysteine or methionine residues, may also be used for phasing; however the signal is very weak and requires high-quality data.

The strong anomalous signal of the iodine atoms (Fig. 7.8) renders I3C a powerful phasing tool, both for in-house and synchrotron data. However the iodine K edge ($\lambda = 0.374$ Å) and L_3 edge ($\lambda = 2.72$ Å) are not commonly accessible on synchrotron beamlines. Therefore, I3C can only be used for SAD or SIRAS phasing. Similar compounds could

be used to introduce elements such as bromine or selenium that would be more suitable than iodine for MAD experiments, or exploit hydrophobic interactions as well as hydrogen bonds for binding to protein molecules.

Chapter 8

B3C - The MAD triangle



Summary

The MAD triangle B3C carries three bromine atoms for multi-wavelength anomalous dispersion (MAD) experiments. B3C has, analogously to I3C, three functional groups for interaction with proteins. The molecular structure of B3C was determined by X-ray crystallography.

The test protein proteinase K was derivatised with B3C *via* soaking and a MAD experiment was carried out. The effect of radiation damage to B3C was investigated to see if radiolysis of the bromine-carbon bond can obstruct structure solution. B3C was also incorporated by co-crystallisation into bovine trypsin. A SAD experiment was carried out with data collected in-house.

8.1 Introduction

After the success with the iodine compound I3C there was interest in expanding the introduced concept (heavy atoms for phasing plus functional groups for interaction) to elements that are suitable for multi-wavelength anomalous dispersion (MAD) experiments. As discussed in chapter 2 these multi-wavelength experiments are carried out around the absorption edge of a particular anomalous scatterer and yield more phase information. The anomalous signal of iodine is very strong at longer wavelengths, but the absorption edges of iodine are not accessible at a typical macromolecular beamline ($\lambda < 2.0\,\text{Å}$).

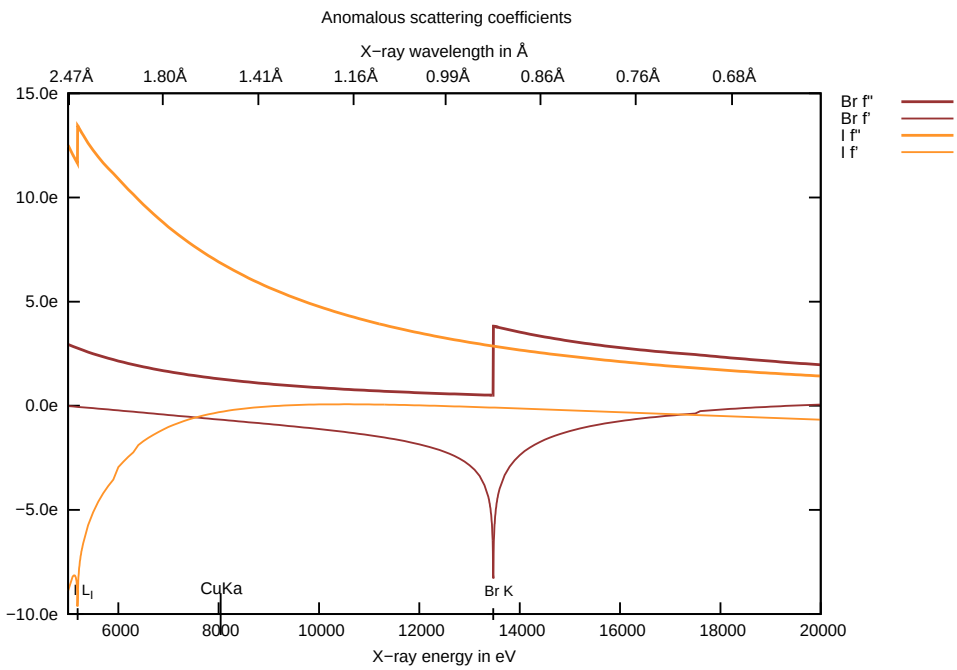
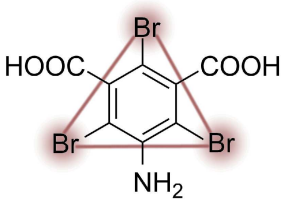


Figure 8.1: Calculated anomalous scattering coefficients for iodine and bromine (Merrit, 2010).

However the bromine K edge ($\lambda = 0.920\,\text{Å}$) falls within the normal energy range of a macromolecular crystallography beamline (Fig. 8.1). The analogous bromine compound 5-amino-2,4,6-tribromoisophthalic acid (B3C, Fig. 8.2) contains three functional groups for interaction with proteins or nucleic acids but also three bromine atoms. These render B3C a suitable phasing tool for MAD experiments.



B3C

Figure 8.2: B3C with its three bromine atoms is suitable for MAD experiments.

The three bromine atoms in B3C form an equilateral triangle (Fig. 8.2), which is easily recognised in the heavy-atom substructure. The molecular structure of B3C was determined by single-crystal X-ray analysis (Beck *et al.*, 2009b) and was used for the generation of bond length and bond angle restraints.

8.2 B3C crystal structure

In the crystal environment B3C has no internal crystallographic symmetry and crystallises in the space group $P2_1/c$ with four molecules in the asymmetric unit. The bond lengths and angles fall within normal ranges. A thermal ellipsoid plot is shown in Fig. 8.3.

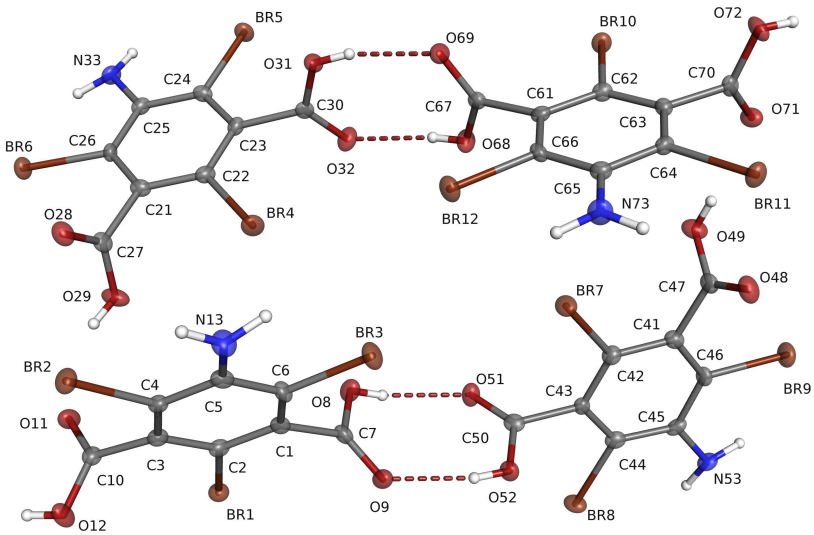


Figure 8.3: The asymmetric unit of B3C. Displacement ellipsoids are drawn at the 50% probability level and hydrogen atoms are shown as small spheres with arbitrary radii. Hydrogen bonds within the asymmetric unit are shown as dashed lines.

The distances between bromine atoms in the four B3C molecules vary from 5.652(1) to 5.695(1) Å. For each B3C molecule, the bromine atoms form an equilateral triangle.

Table 8.1: Selected bond lengths (Å) for B3C.

C7−O9	1.212(5)	C47−O48	1.207(5)
C7−O8	1.312(5)	C47−O49	1.321(5)
C10−O11	1.219(5)	C50−O51	1.212(5)
C10−O12	1.303(5)	C50−O52	1.312(5)
C27−O28	1.213(5)	C67−O69	1.236(5)
C27−O29	1.309(5)	C67−O68	1.289(5)
C30−O32	1.240(5)	C70−O71	1.215(5)
C30−O31	1.288(5)	C70−O72	1.300(5)

As well as for I3C, there are clear indications for protonated carboxyl groups in B3C (Tab. 8.1). In all carboxyl groups, one C$\cdots$O bond is significantly shorter than the other

one, indicating one double bond and one single bond. Interestingly, the bond lengths for C30···O31/O32 and C67···O68/O69 do not show such large differences, but still indicate a protonation of O31 and O68, respectively (Fig. 8.4).

In the crystal, molecules are linked to chains *via* COO−H···O bonds (Fig. 8.4). The first carboxyl group of each molecule forms cyclic dimers with a neighbouring carboxyl group (e.g. O8−H8···O51 and O52−H52···O9, see Tab. 8.2 for hydrogen-bond geometry), as observed frequently for carboxylic acids in the solid state (Leiserowitz, 1976). The second carboxyl group is involved in hydrogen bonding as well to form a 1-dimensional chain (e.g. O12−H12···O48^{i}, see Tab. 8.2 for symmetry codes). Instead of a dimeric interaction, one hydrogen bond is formed to an adjacent chain (e.g. O29−H29···O11, Tab. 8.2). Therefore a two-dimensional hydrogen bond network is found in the crystal lattice of B3C. In trimesic acid with its three carboxyl groups, a two-dimensional network is found, not limited to chains but extending within a plane (Duchamp and Marsh, 1969).

Table 8.2: Hydrogen-bond geometry (Å, °) for B3C.

D−H···A	D−H	H···A	D···A	D−H···A
O8−H8···O51	0.75(3)	1.97(3)	2.713(4)	167(6)
O12−H12···O48^{i}	0.76(3)	1.96(3)	2.719(4)	177(6)
O52−H52···O9	0.76(3)	1.92(3)	2.667(4)	170(6)
O49−H49···O71	0.75(3)	1.93(3)	2.664(4)	165(7)
O31−H31···O69	0.76(3)	1.91(3)	2.649(4)	165(6)
O68−H68···O32	0.75(3)	1.91(3)	2.659(4)	171(6)
O72−H72···O28^{ii}	0.76(3)	1.92(3)	2.673(4)	171(6)
O29−H29···O11	0.76(3)	1.99(4)	2.662(4)	147(6)
N33−H33A···O68^{iii}	0.88(2)	2.39(3)	3.237(5)	164(5)

Symmetry codes: (i) $x+1, y, z$; (ii) $x-1, y, z$; (iii) $-x+1, y-\frac{1}{2}, -z+\frac{3}{2}$.

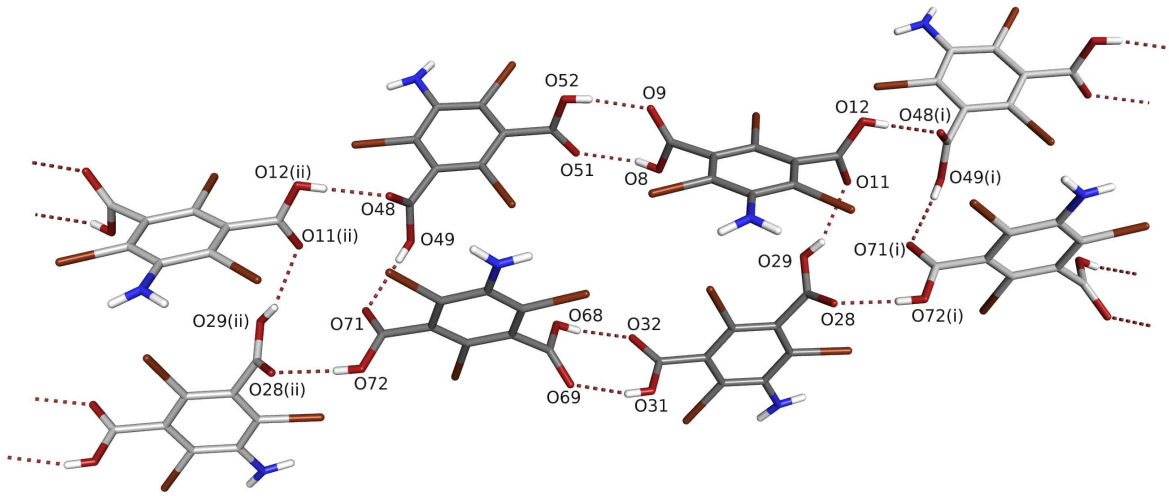


Figure 8.4: Hydrogen bonding of B3C. Molecules form two-dimensional chains. The chains are connected pairwise to each other. Symmetry equivalent carbon atoms are depicted in lighter shade. For symmetry operators, please refer to Tab. 8.2.

A bundle of chains shows weak hydrogen bonding *via* one amino group (N33) to adjacent chain bundles (Fig. 8.5, Tab. 8.2). In addition, stacking of benzene rings (C21 to C26) with symmetry equivalents of C41 to C46 is observed as a contact between the chain bundles. No direct face-to-face contact is present, but rather an offset-geometry, favoured by decreased π-π repulsion and increased σ-π attraction (Hunter and Sanders, 1990). Close carbon-carbon contacts are observed between C25 and C42^{iii} [3.690(5)Å], C26 and C42^{iii} [3.889(5) Å], C25 and C43^{iii} [3.803(5) Å], C26 and C43^{iii} [3.843(5) Å]. In addition, a close carbon-bromine distance is observed [Br6$\cdots$C46^{iii} 3.620(4) Å]. The angle between the two benzene ring planes is 7.9(2)°.

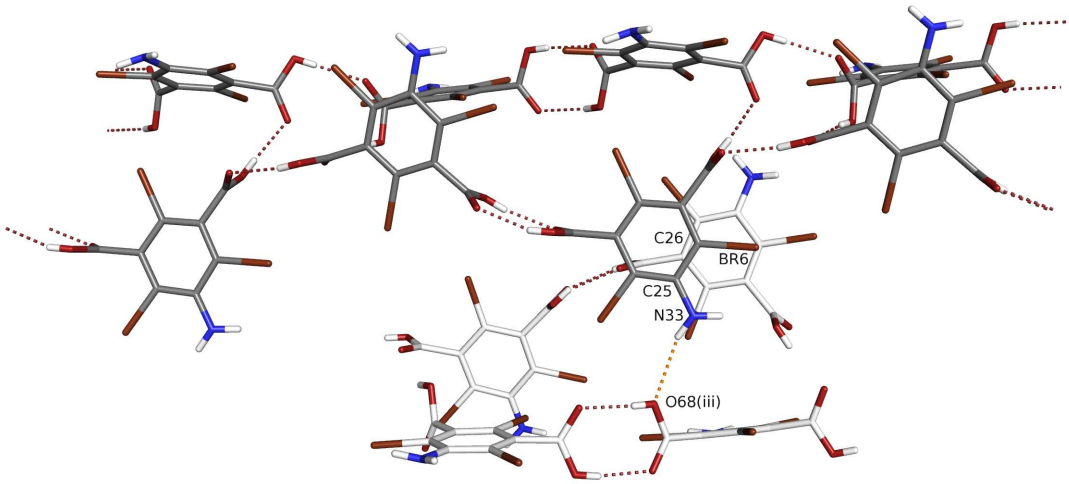


Figure 8.5: Molecules form chain bundles (see Fig. 8.4) that are connected via offset π-π stacking and weak hydrogen bonding. Atoms from the first chain bundle that are involved in the π-π interaction are labelled. Molecules of a second chain pair are depicted in lighter shade. The weak hydrogen bond between N33 and O68^{iii} is shown as a yellow dashed line.

The molecular arrangement in the crystal lattice differs from the packing found in crystals of the iodine derivative I3C. In B3C, no lattice water molecule is present. Decreased halogen-carbon bond lengths and van der Waals radii (1.85 Å for Br *vs.* 1.98 Å for I; Bondi (1964)) explain why the cyclic dimeric hydrogen bonds of two neighbouring carboxyl groups are found for B3C but not for I3C.

The crystallographic analysis makes clear that as for I3C hydrogen bonds are a crucial interaction of the B3C molecules in the crystal lattice.

8.3 MAD phasing with B3C

B3C was incorporated into crystals of proteinase K *via* soaking and a four-wavelength MAD experiment was carried out. Radiation damage was investigated since it is known that the radiolysis of the anomalous scatterers, e.g. the bromine atom in brominated nucleotides,

can prevent structure solution (Ennifar *et al.*, 2002). The phasing ability of B3C was compared with I3C for proteinase K.

Substructure solution and data analysis

For B3C four data sets (two peak, one high-energy remote and one inflection point) were collected for MAD phasing (see Fig. 8.7 for anomalous data statistics). Data sets were prepared with *XPREP*. Substructure solution was carried out with *SHELXD*. Inspection of the heavy-atom sites revealed the presence of equilateral triangles (with side lengths of about 5.6 Å).

Table 8.3: Data collection details for proteinase K with B3C and I3C. Values in parentheses denote the highest resolution shell.

	B3C				I3C
	Peak1	Peak2	High-energy remote	Inflection	
Unit-cell parameters (Å)	$a = b = 67.84$, $c = 101.77$				$a = b = 67.78$, $c = 101.84$
Space group	$P4_32_12$				$P4_32_12$
Wavelength (Å)	0.9197	0.9202	0.9129	0.9204	1.5418
Photon energy (eV)	13480	13473	13580	13471	8042
Resolution (Å)	48.0−1.50 (1.60−1.50)	48.0−1.50 (1.60−1.50)	48.0−1.50 (1.60−1.50)	48.0−1.50 (1.60−1.50)	47.9−1.76 (1.86−1.76)
Degrees collected (°)	180	180	100	100	180
R_{merge}	0.072 (0.23)	0.068 (0.28)	0.065 (0.23)	0.062 (0.22)	0.048 (0.13)
Completeness (%)	99.8 (99.6)	99.8 (99.2)	99.8 (99.4)	99.7 (99.2)	93.8 (62.0)
Multiplicity	14.1 (12.9)	14.1 (12.9)	7.9 (7.4)	7.8 (7.2)	10.14 (1.73)
$< I/\sigma(I) >$	27.3 (10.0)	24.9 (8.3)	22.0 (7.6)	23.0 (7.8)	33.8 (8.3)

For crystallographic parameters, see Appendix B: Crystallographic parameters.

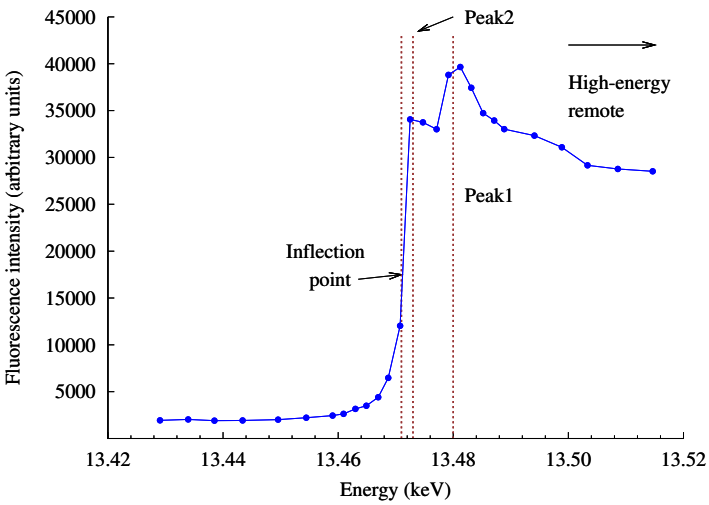


Figure 8.6: Fluorescence scan of proteinase K with B3C incorporated. The data collection energies of peak1, peak2 and inflection point (see Tab. 8.3) are marked by vertical lines. The two peaks close to the bromine K edge are clearly visible.

The two peaks in the fluorescence spectrum (Fig. 8.6) can be rationalised by the anisotropy of the anomalous signal, although fluorescence scans at different orientations of the crystal (not carried out) would be required to confirm this. Similar effects have been observed for proteins containing selenomethionine and brominated nucleotides (Schiltz and Bricogne, 2008).

For I3C, SAD phasing was carried out using the data collected at 1.5418 Å (see Fig. 8.7b for anomalous data statistics). Substructure solution with *SHELXD* resulted in heavy-atom sites that form equilateral triangles (with a side length of about 6 Å).

Refinement was carried out with the *SHELX* suite (Tab. C.2) with the restraints for B3C and I3C generated from the small-molecule crystal structures. Free variables were introduced to refine the occupancy of each B3C or I3C site and later also the occupancy of each single halogen atom separately in B3C or I3C. For the occupancy refinement, thermal displacement parameters were kept fixed for the halogen atoms (at B = 15.8 Å^2 for both bromine and iodine).

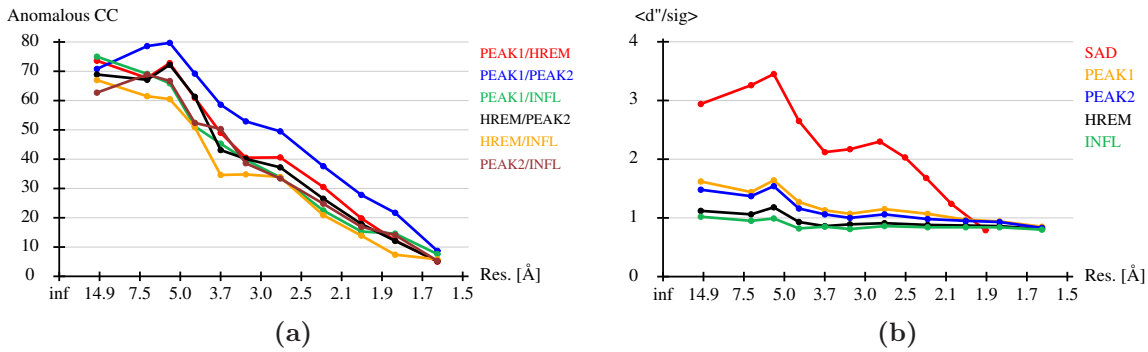


Figure 8.7: Anomalous data statistics. **(a)** Anomalous correlation coefficient (CC) between the MAD data sets. HREM = high-energy remote, INFL = inflection point (data collection wavelengths can be found in Tab. 8.3). The correlation coefficients do not depend on the (estimated) sigma values. Data were truncated at 2.5 Å for heavy-atom substructure solution (where CC falls below 30% for all data sets). **(b)** Anomalous signal for the four bromine MAD data sets and the iodine SAD data set (red); pure noise would correspond to $d''/\text{sig} \simeq 0.798$. Here, the cut-off for the MAD data sets cannot be determined easily. The iodine SAD data set shows a strong anomalous signal (data collected at 1.5418 Å). Data were truncated at 2.0 Å for heavy atom substructure solution for the iodine SAD data set (where d''/sig falls below 1.2).

Table 8.4: Comparison of the phasing statistics for the B3C and I3C data sets. The mean phase error and mean map correlation coefficient compare the first experimental map obtained after density modification with *SHELXE* with the final refined map (*SHELXL*). MAD phasing with B3C results in better starting phases: the mean phase error is lower, the mean map correlation coefficient is higher and more residues can be traced compared to the SAD data set.

	B3C (MAD)	I3C (SAD)
SHELXD CC/CC_{weak}	46.6 / 37.6	52.0 / 32.5
mean phase error (°)	31.2	36.9
mean map CC	0.861	0.786
residues traced (*ARP/wARP*)	272/279 [97.5 %]	267/279 [95.7 %]

The final model from the MAD data set refinement (peak2) was used for further refinement of the radiation damage data sets. A comparison of the results from the MAD and the SAD phasing experiments can be found in Tab. 8.4.

B3C and I3C binding sites in proteinase K

For B3C in proteinase K four binding sites (Fig. 8.8) and for I3C three binding sites are observed. Two sites coincide for both derivatives; one of these is the main site with the highest occupancy, shown in Fig. 8.8b.

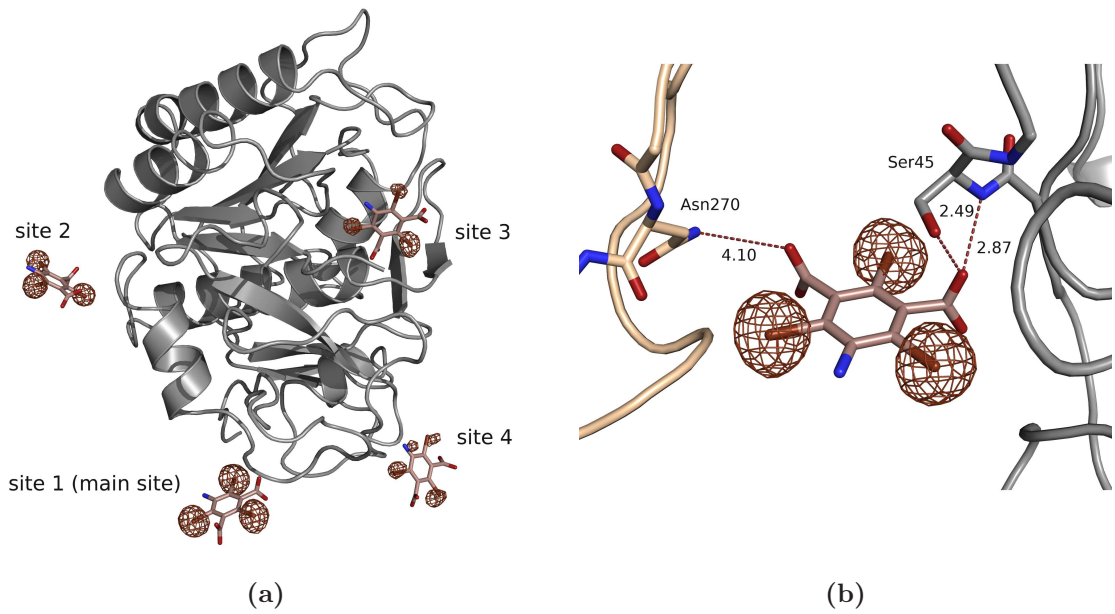


Figure 8.8: B3C in proteinase K. **(a)** The asymmetric unit with four molecules of B3C. Substructure density calculated with *SHELXE* (using F_A and α derived from the dispersive and anomalous differences) is contoured at 4σ. Clear density can be seen for the bromine atoms of the four B3C molecules. **(b)** B3C (site 1) in proteinase K at the interface of two proteinase K molecules. Substructure density shown as for a). Hydrogen bonds are depicted as dashed lines and distances are given in Å. The two carboxylate groups interact with Asn270 and Ser45.

The common occupancies for all three halogen atoms per site were refined with *SHELXL* to 0.42, 0.13, 0.10 and 0.09 for B3C (refinement against peak2 data set;[1] Tab. C.2) and 0.80, 0.19 and 0.14 for I3C (refinement details in Tab. C.2). Interestingly, the occupancy of the main site differs significantly although similar soaking conditions were used. The difference might be due to different soaking times or crystal properties or can be attributed to the different chemical properties of the two compounds (containing either iodine or bromine atoms, see chapter 12).

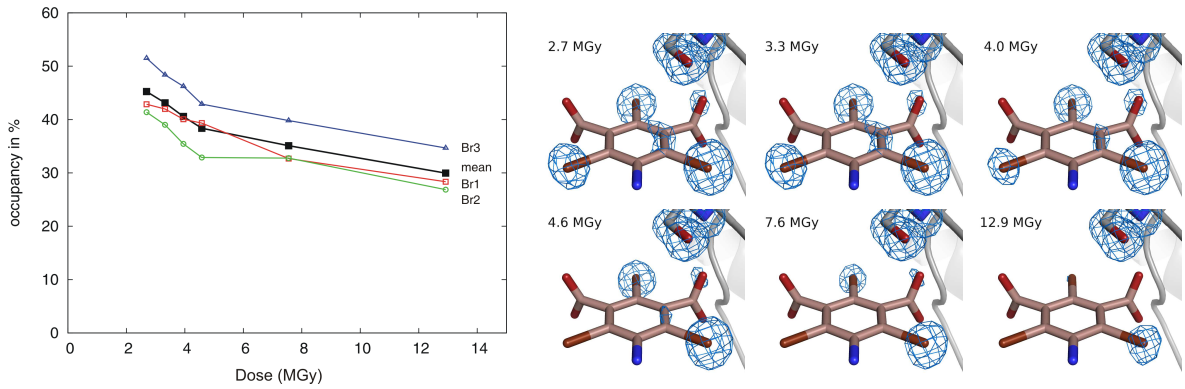
Interactions of the small molecules in proteinase K are very similar to those previously reported for lysozyme, thaumatin and elastase (Beck *et al.*, 2008). The three functional

[1]The atomic scattering factor of bromine was adjusted to the given wavelength (*SHELXL* DISP $f'f''$ command with $f' = -8.7\,e$ and $f'' = 3.47\,e$, experimental values).

groups of the phasing tools form hydrogen bonds to side chains or the main chain of the protein and to water molecules. Interactions for the main site of B3C (site 1) are shown in Fig. 8.8b. One carboxylate group interacts with a serine residue and the amide hydrogen of the protein backbone. The other carboxylate group interacts with the amino group of an asparagine through hydrogen bonding. Interactions of the amino group with the protein and other interactions of the carboxylate groups are also observed.

Radiation damage

The effect of irradiation is depicted in Fig. 8.9. Due to changes in the experimental setup after the MAD experiment (detector distance was changed from 160 mm to 200 mm; crystal re-centring after de-icing), the occupancies from the refinement against peak2 (see above) deviate from the occupancies obtained from the radiation damage experiments and are therefore not depicted in Fig. 8.9.



(a) (b)

Figure 8.9: Radiation damage for B3C (site 1) in proteinase K. The absorbed dose for the crystal is given in MGy (Gray = J kg^{-1}). **(a)** Refined occupancies (*SHELXL*) for the three bromine atoms and the mean value (black) of site 1. The first data point is the occupancy after MAD data collection and a single burn, and the subsequent five points show the change in occupancy with increasing dose. A drop of the refined occupancies of about 15% can be observed after a dose of 12.9 MGy. **(b)** Electron density at site 1 contoured at 1σ. The serine residue can be used as a reference since its density is not affected by irradiation. The density at the bromine atoms clearly decreases with dose, but not all three bromine atoms are affected equally.

A loss of about 15% in occupancy of the bromine sites can be observed (Fig. 8.9a) after the MAD experiment and six consecutive radiation damage experiments (burn-collect). The program *RADDOSE* was used for dose calculations. Interestingly, not all bromine-carbon bonds suffer to the same extent from irradiation (Fig. 8.9b). Although the bromine-carbon bonds were cleaved, the MAD experiment was still successful. Further MAD experiments with B3C will show if radiation damage can actually obstruct structure solution with B3C.

Preliminary results from radiation damage experiments with I3C (results not shown here) indicate that I3C is at least as susceptible as B3C to radiation, probably even more. Similar results have been reported for halogenated nucleotides (Oliéric *et al.*, 2009).

8.4 SAD phasing with B3C

In addition to derivatisation of native protein crystals *via* soaking B3C was incorporated by means of co-crystallisation. Although MAD experiments yield more phase information, it was tested if SAD phasing with B3C in-house data could be carried out.

Substructure solution and data analysis

The anomalous signal of Br is relatively weak at CuKα ($f'' = 1.25\,e$); Fig. 8.10 shows the anomalous signal for this data set. However the ratio of R_{anom} to $R_{\text{p.i.m.}}$ indicates that structure solution should be possible using in-house data (Tab. 8.5). The presence of one calcium ion that also contributes to the anomalous signal at this wavelength (f'' of Ca about the same as for Br) is probably helpful for structure solution.

Table 8.5: Data collection details for bovine trypsin with B3C. Values in parentheses denote the highest resolution shell.

Unit-cell parameters (Å, °)	$a = 53.66, b = 56.88, c = 66.81$; $\alpha = \beta = \gamma = 90$
Space group	$P2_12_12_1$
Wavelength (Å)	1.54178
Resolution (Å)	41.8−1.55 (1.65−1.55)
R_{merge}	0.0658 (0.288)
Completeness	99.9 (99.7)
Multiplicity	7.69 (3.61)
$< I/\sigma(I) >$	18.9 (3.87)
$R_{\text{anom}}/R_{\text{p.i.m.}}$	2.81 (2.20)

For crystallographic parameters, see Appendix B: Crystallographic parameters.

The heavy-atom substructure was successfully solved with *SHELXD* using the single-wavelength anomalous dispersion (SAD) method (CC: 18.31, CC_{weak}: 9.97). The first four peaks found with *SHELXD* are the calcium atom and three atoms forming an equilateral triangle with a side of about 5.6 Å, i.e. the bromine atoms in B3C.

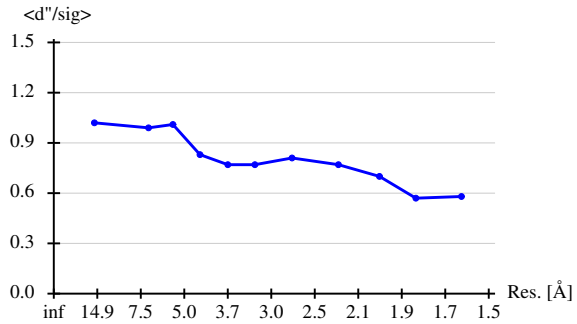


Figure 8.10: Anomalous data statistics for B3C in bovine trypsin. Three bromine atoms (B3C) and one calcium atom contribute to the anomalous signal. Data were truncated at 1.9 Å for heavy-atom substructure solution. Although the anomalous signal is rather weak, heavy-atom substructure solution and phase extension could successfully be carried out.

SHELXE was used for phase extension and density modification. The new beta-test version, which includes a protein main-chain tracing algorithm (Sheldrick, 2010), was used. 218 out of 223 residues were traced, belonging to five different fragments. The experimental

electron density map obtained from *SHELXE* shows a high correlation with the final refined map (*CC*: 0.892) and a low mean phase error (26.8°).

Although experimental phasing with B3C using data collected at a single wavelength was successful in this example, it is recommended to use B3C for MAD experiments to take advantage of the extra phase information available from different wavelengths.

B3C binding site in trypsin

The initial backbone model from *SHELXE* was first manually completed with *COOT* (Emsley *et al.*, 2010). The side-chains were docked within *COOT* using the 'dock sequence' extension. Refinement was carried out with *SHELXL* (Tab. C.3).

There is a single site for B3C in the crystal structure; its occupancy was refined to 0.9. B3C interacts with two molecules of trypsin and shows similar hydrogen bond interactions to those observed in other protein structures (Fig. 8.11).

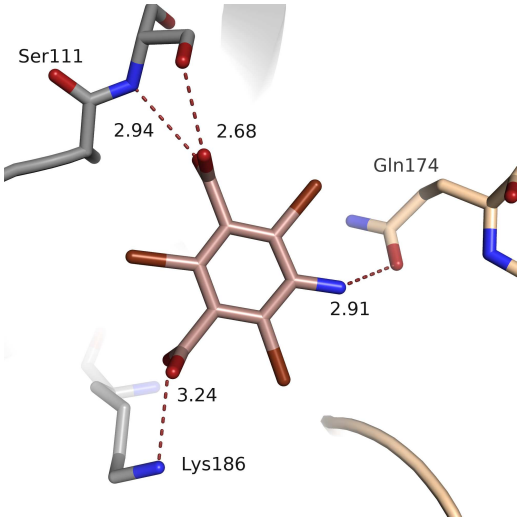


Figure 8.11: The B3C binding site in bovine trypsin at the interface of two protein molecules. Hydrogen bonds with Lys186, Ser111 (including the interaction with the amide proton) and Gln174 are shown as dashed lines, distances are given in Å.

8.5 Conclusion

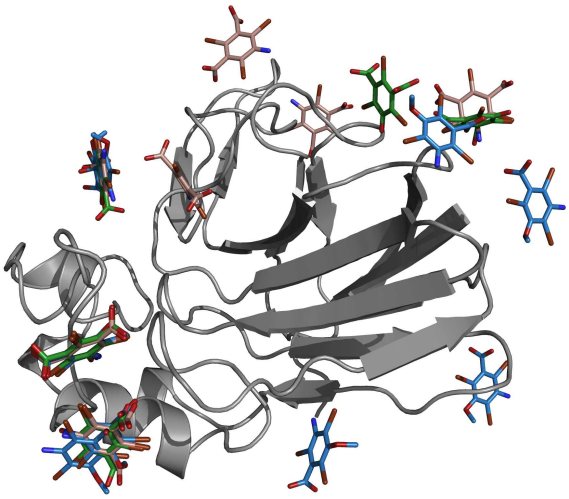
B3C and I3C represent a novel class of compounds that show interaction with protein molecules. These sticky phasing tools may be utilised for experimental phasing. I3C is the compound of choice for in-house data collection since the iodine atoms give rise to a strong anomalous signal at CuKα radiation (SAD or SIRAS phasing). If diffraction is too weak for in-house phasing, B3C is suitable for MAD data collection at a synchrotron beamline, taking advantage of the additional phase information from data collection at different wavelengths. The fixed geometrical arrangement of the heavy atoms facilitates structure solution since the triangles are readily identified in the heavy-atom substructure. The effect of radiation damage on B3C and its phasing capabilities has been investigated. Although the bromine-carbon bond in B3C suffers considerably from radiolysis, a MAD experiment could still be carried out successfully (Beck *et al.*, 2010). In order to take advantage of the radiolysis of the anomalous scatterers, it is recommended to collect the

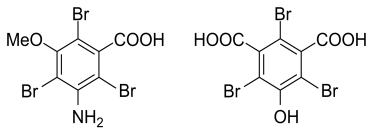
inflection data set at the end of the MAD experiment. It was found here that B3C binds to the surface of the protein at the periphery. Interestingly, although several modes of hydrogen bonding interactions are observed for B3C and I3C, no aromatic interactions are found in the crystals investigated so far. As for I3C the bulky halogen atoms and the carboxylate groups arranged perpendicular to the benzene ring may hinder π-π interactions with aromatic side chains.

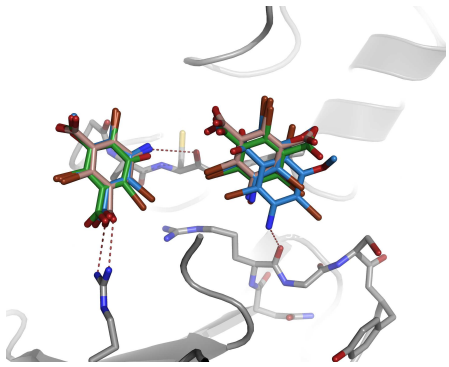
It was shown that co-crystallisation is another practical method, in addition to soaking of native protein crystals, for the incorporation of B3C molecules for phasing. Co-crystallisation of B3C with bovine trypsin results in a single site with high occupancy (Beck *et al.*, 2009a). In-house SAD phasing was carried out based on the anomalous signal of one Ca and three Br atoms. However, the recommended usage for B3C is to carry out a multi-wavelength anomalous dispersion (MAD) experiment. It was shown that MAD phasing with B3C yields improved experimental phases. Since some crystals do not tolerate the soaking with high concentrations of I3C or B3C due to disturbance of the crystal lattice, co-crystallisation with these compounds might prove to be a valuable option for incorporating the phasing tools, especially if no suitable native crystals could be obtained in the first place. Samples of B3C can be provided to interested researchers.

Chapter 9

More sticky triangles







Summary

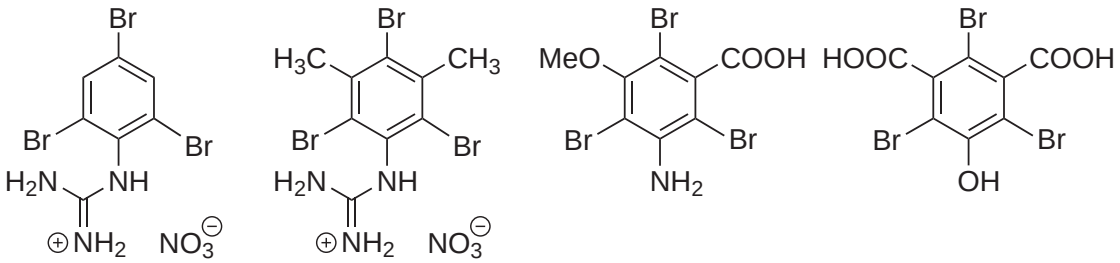
The two bromine compounds B3M and B3O contain other functional groups: a methoxy group and a hydroxyl group. The molecular structures of B3M and B3O were determined by X-ray crystallography.

B3M and B3O were utilised for heavy-atom derivatisation of thaumatin. Experimental phasing was carried out with data collected at the synchrotron. The binding sites of B3M and B3O were compared with the binding sites of B3C. Although they carry different functional groups the binding is very similar to B3C.

9.1 Introduction

I3C and B3C show several modes of interaction with protein residues, but the binding might still be improved. Especially the incorporation of functional groups for interaction with hydrophobic residues could be beneficial. Therefore, a series of reactions was carried out to produce sticky triangle compounds with different functional groups. The focus was put on brominated compounds since these can be used for MAD experiments, which yield more phase information.

As discussed in chapter 4, several strategies were tested to substitute the amino group with a guanidinium moiety. This group carries a permanent positive charge at neutral pH, enabling ionic interactions (as for the carboxylate groups). Additionally the increased size compared to the single atom of the amino group (e.g. compare **15** and B3M in Fig. 9.1) may facilitate interactions since the functional group is better accessible (no steric hindrance due to the adjacent bromine atoms). However due to the deactivating effect of the guanidinium group on the benzene ring only the two guanidinium derivatives **15** and **18** (Fig. 9.1) could be synthesised. The compounds are not soluble at neutral pH in aqueous media. However derivatisation with these compounds in solutions that contain organic solvents is possible and especially suitable for macromolecules that contain negatively charged residues (e.g. nucleic acids).[1]



15 **18** **B3M** **B3O**

Figure 9.1: Sticky triangles that were synthesised for heavy-atom derivatisation. B3M and B3O were utilised for heavy-atom derivatisation. **15** and **18** show only poor solubility at neutral pH.

Two bromine derivatives were synthesised that can be utilised for heavy-atom derivatisation. B3M contains a methoxy group for possible interactions with hydrophobic residues. B3O contains a hydroxyl group instead of the amino group in B3C. The amino group in B3C is not protonated at pH values used for crystallisation and derivatisation of biological macromolecules. It carries two hydrogen atoms and one electron lone pair for hydrogen bonding. In contrast the hydroxyl group carries two electron lone pairs at the oxygen atom that are suitable as acceptors for hydrogen bonds, but only one hydrogen atom.

As for the halogen atoms in I3C and B3C the three bromine atoms in B3M and B3O form an equilateral triangle. The molecular structures of B3M and B3O were determined

[1] At the time of writing **15** and **18** are tested for derivatisation of a DNA/intercalator complex. The crystallisation conditions include DMSO, allowing the bromine compounds to be dissolved for co-crystallisation experiments.

by single-crystal X-ray analysis and were used for the generation of bond length and bond angle restraints, required for the refinement of derivative data sets.

9.2 Crystal structures

B3M crystal structure

B3M has no internal crystallographic symmetry and crystallises as a monohydrate in the space group $P6_5$ with one molecule in the asymmetric unit. The bond lengths and angles fall within normal ranges. A thermal ellipsoid plot is shown in Fig. 9.2.

The three bromine atoms in B3M form an equilateral triangle [cf. Br1−Br2 5.6837(13), Br2−Br3 5.6521(13), Br1−Br3 5.6764(13) Å]. As for I3C and B3C, there are clear indications for a protonated carboxyl group. One C−O bond is significantly shorter than the other one, indicating one double bond and one single bond [cf. C7−O9 1.215(4) *vs.* C7−O8 1.305(4) Å]. The carboxyl groups are arranged perpendicular to the ring.

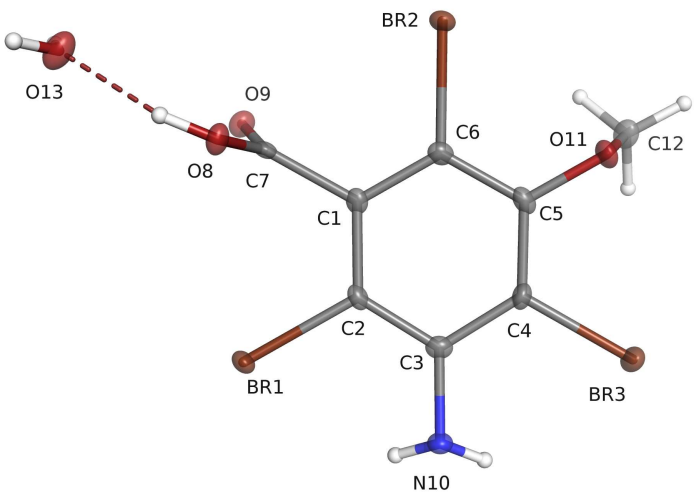


Figure 9.2: The asymmetric unit of B3M. Displacement ellipsoids are drawn at the 50% probability level and hydrogen atoms are shown as small spheres with arbitrary radii. The hydrogen bond to the lattice water molecule is shown as a dashed line.

Hydrogen bonding involves the lattice water molecule as well as the carboxyl group and the amino group (Fig. 9.3, Tab. 9.1). Only one carboxyl group is present per molecule and no dimeric hydrogen bonding of carboxyl groups as for B3C is observed here.

Table 9.1: Hydrogen-bond geometry (Å, °) for B3M.

D−H$\cdots A$	D−H	H$\cdots A$	$D\cdots A$	D−H$\cdots A$
O8−H8$\cdots$O13	0.85(4)	1.73(4)	2.553(3)	163(4)
N10−H10A$\cdots$O8^{i}	0.79(3)	2.37(4)	3.113(4)	158(4)
O13−H13B$\cdots$O9^{ii}	0.76(3)	2.21(3)	2.909(3)	154(4)
O13−H13A$\cdots$N10^{iii}	0.75(3)	2.12(3)	2.849(4)	165(5)

Symmetry codes: (i) $x-y, x, z-\frac{1}{6}$; (ii) $y+1, -x+y+1, z+\frac{1}{6}$; (iii) $x+1, y, z$.

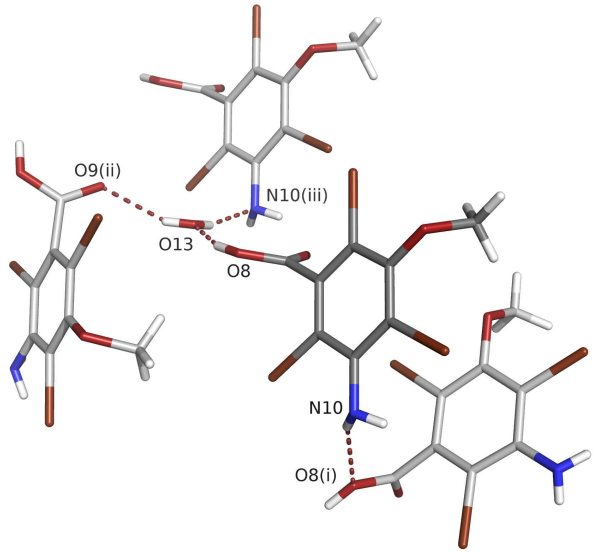


Figure 9.3: Hydrogen bonding of B3M. The hydrogen bonds for hydrogen donor atoms of the central B3M are depicted as dashed lines. Symmetry-equivalent molecules are shown in lighter shade. For symmetry operators, please refer to Tab. 9.1.

The molecules are arranged in layers in the crystal lattice. However, only hydrogen bonding connects these layers and no π-π interactions are observed (probably due to the steric hindrance of the carboxyl groups and bromine atoms). The small-molecule structure of the unsubstituted 3-amino-5-methoxybenzoic acid has not been determined yet.

B3O crystal structure

In the crystal environment B3O has no internal crystallographic symmetry and crystallises as a monohydrate in the space group $P2_1/c$ with one molecule in the asymmetric unit. The bond lengths and angles fall within normal ranges. A thermal ellipsoid plot is shown in Fig. 9.4.

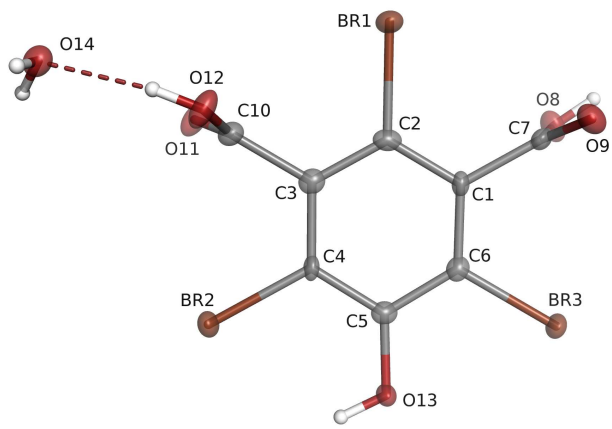


Figure 9.4: The asymmetric unit of B3O. Displacement ellipsoids are drawn at the 50% probability level and hydrogen atoms are shown as small spheres with arbitrary radii. The hydrogen bond to the lattice water molecule is shown as a dashed line.

As well as for B3C and B3M, the three bromine atoms in B3M form an equilateral triangle [cf. Br1−Br2 5.6617(13), Br2−Br3 5.6725(21), Br1−Br3 5.6529(12) Å]. The two

carboxyl groups in B3O are protonated: one C−O bond is significantly shorter than the other one, indicating one double bond and one single bond [cf. C7−O9 1.225(5) *vs.* C7−O8 1.305(5) Å and C10−O11 1.229(5) *vs.* C10−O12 1.292(5) Å]. The carboxyl groups are arranged perpendicular to the ring.

In the crystal structure B3O forms hydrogen bonds with all potential donors and the lattice water molecule (Fig.9.5, Tab. 9.2); however the angle in the hydrogen bond from O13 to O14^{ii} deviates from ideal values.

Table 9.2: Hydrogen-bond geometry (Å, °) for B3O.

D−H$\cdots A$	D−H	H$\cdots A$	$D\cdots A$	D−H$\cdots A$
O8−H8$\cdots$O13^{i}	0.83(2)	1.98(3)	2.765(4)	159(6)
O12−H12$\cdots$O14	0.83(2)	1.83(2)	2.652(4)	172(6)
O13−H13$\cdots$O14^{ii}	0.82(2)	2.02(3)	2.722(4)	144(5)
O14−H14A$\cdots$O11^{iii}	0.75(4)	1.93(4)	2.670(4)	173(6)
O14−H14B$\cdots$O9^{iv}	0.75(4)	2.08(4)	2.812(4)	165(6)

Symmetry codes: (i) $-x+2, y-\frac{1}{2}, -z+\frac{3}{2}$; (ii) $-x+1, y+\frac{1}{2}, -z+\frac{1}{2}$; (iii) $x, -y+\frac{3}{2}, z-\frac{1}{2}$; (iv) $x-1, -y+\frac{3}{2}, z-\frac{1}{2}$.

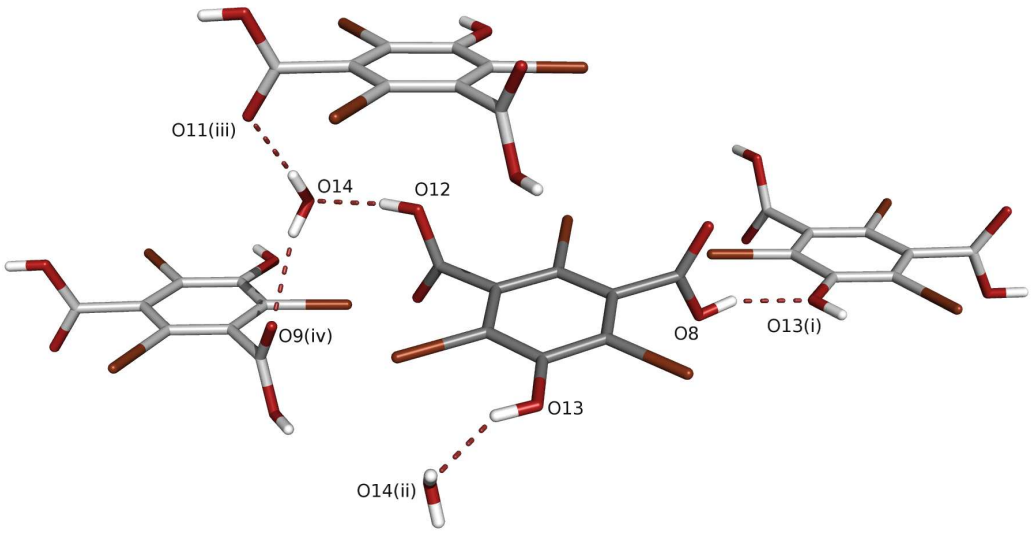


Figure 9.5: Hydrogen bonding of B3O. The hydrogen bonds for hydrogen donor atoms of the central B3O are depicted as dashed lines. Symmetry-equivalent molecules are shown in lighter shade. For symmetry operators, please refer to Tab. 9.2.

The non-brominated compound 5-hydroxyisophthalic acid shows a different molecular arrangement in the crystal lattice. As for the unsubstituted 5-aminoisophthalic acid, the carboxyl groups are in the same plane as the benzene ring and π-π stacking is found (Ermer and Neudörfl, 2001). For B3O, there are no π-π interactions present (probably due to the steric hindrance of the carboxyl groups and bromine atoms), although the molecules are packed in layers.

9.3 Experimental phasing

The primary objective was to investigate the effect of different functional groups on binding of the bromine compounds. B3M, B3O and B3C were incorporated in thaumatin *via* soaking native crystals as described in chapter 5.

B3M and B3O: substructure solution and data analysis

For the thaumatin crystals derivatised with B3M and B3O data collection was carried out at $\lambda = 0.9130$ Å, which corresponds to the high-energy remote in a bromine MAD experiment.[2] The anomalous signal of bromine is strong at this wavelength ($f'' = 3.7\,e$; Merrit, 2010).

Table 9.3: Data collection details for thaumatin with B3M, B3O and B3C. Values in parentheses denote the highest resolution shell.

	B3M	B3O	B3C		
	High-energy remote	High-energy remote	Peak	Inflection	High-energy remote
Unit-cell parameters (Å)	$a = b = 58.08$, $c = 150.04$	$a = b = 57.95$, $c = 150.20$	$a = b = 57.95$, $c = 150.36$		
Space group	$P4_12_12$				
Wavelength (Å)	0.9130	0.9130	0.9199	0.9205	0.9131
Photon energy (eV)	13580	13580	13478	13470	13578
Resolution (Å)	24.6−1.60 (1.69−1.60)	24.5−1.60 (1.70−1.60)	24.5−1.60 (1.69−1.60)	24.5−1.60 (1.69−1.60)	24.6−1.60 (1.69−1.60)
Degrees collected (°)	120	120	90	90	90
R_{merge}	0.044 (0.16)	0.050 (0.20)	0.052 (0.22)	0.040 (0.18)	0.045 (0.26)
Completeness (%)	99.3 (95.6)	99.5 (97.3)	98.3 (91.6)	98.3 (91.4)	98.5 (92.4)
Multiplicity	8.12 (3.98)	8.14 (4.12)	6.00 (2.83)	5.99 (2.79)	6.08 (2.97)
$< I/\sigma(I) >$	29.6 (6.68)	25.4 (5.64)	15.7 (4.10)	25.7 (4.95)	22.67 (3.72)

For crystallographic parameters, see Appendix B: Crystallographic parameters.

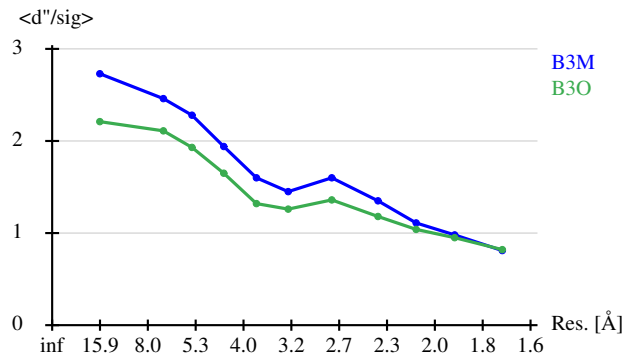


Figure 9.6: Anomalous data statistics for B3M and B3O. Both data sets contain a strong anomalous signal. The data were truncated at 1.9 Å and 2.0 Å for the B3M and B3O substructure solution, respectively.

[2]This wavelength was chosen to ensure a strong anomalous signal without performing a fluorescence scan.

Fig. 9.6 shows the anomalous data statistics for the B3M and B3O data. SAD phasing was carried out with these data sets. Substructure solution was carried out with *SHELXD* (B3M CC: 31.16, CC_{weak}: 16.85; B3O CC: 34.20, CC_{weak}: 18.93). Equilateral triangles with side lengths of about 5.7 Å were observed in each substructure.

B3C: substructure solution and data analysis

A three-wavelength MAD experiment was carried out for the thaumatin B3C derivative (Tab. 9.3). The fluorescence scan is depicted in Fig. 9.7; as for the proteinase K B3C data two peaks are observed in the spectrum.

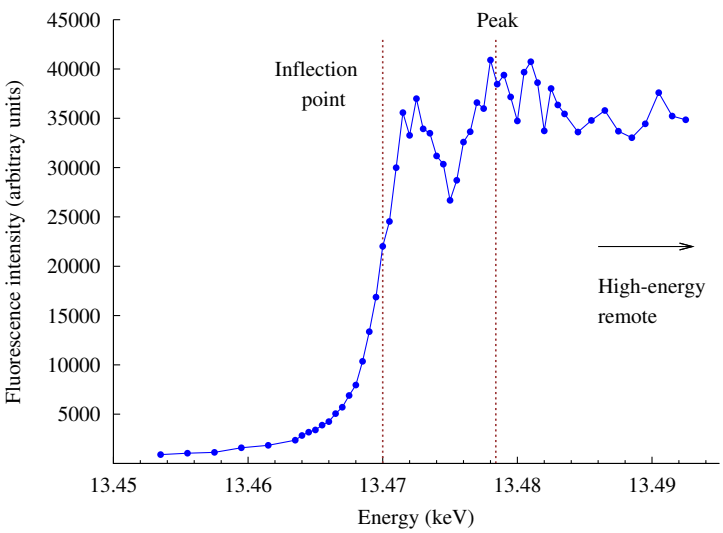


Figure 9.7: Fluorescence scan of thaumatin with B3C incorporated. The data collection energies of peak and inflection point (see Tab. 9.3) are marked by vertical lines. The two peaks close to the bromine K edge are clearly visible.

Data collection for the peak data set was carried out at a higher energy than usual to enhance the difference in dispersive signal between the peak and the inflection point data set. Although it is recommended to collect the high-energy remote data set between peak and inflection point it was collected at the end of the data collection (due to beamline related issues). The anomalous signal of B3C is compared with the B3M and B3O data in Fig. 9.8.

The graph in Fig. 9.8b shows that there is no significant difference in anomalous signal between the B3O high-energy remote data set and the B3C peak data. However compared with the B3C peak the B3C high-energy remote data set (collected at the end) shows a lower signal. While this is theoretically expected (lower f'' at the high-energy remote wavelength) it can also be explained by radiation damage causing radiolysis of the bromine-carbon bonds.

Data sets from the MAD experiment were prepared with *XPREP*. B3C substructure solution was carried out with *SHELXD* (CC: 33.39, CC_{weak}: 24.79; 3 hits with $CC > 30$ out of 1000 trials). Interestingly substructure solution with anomalous data derived only from the peak and the inflection point data sets gives better figures of merit compared to the three-wavelength experiment (CC: 38.35, CC_{weak}: 27.73; 191 hits with $CC > 36$ out of 1000 trials). Additionally the initial experimental map obtained with *SHELXE* without any density modification shows a higher map correlation coefficient for the two-wavelength

experiment (0.542 *vs.* 0.447) and a lower mean phase error (50.3° *vs.* 56.3°) compared to the three-wavelength experiment. This stresses the point that the inflection point data set should always be collected at the end to ensure that the decrease in scattering due to radiation damage carries the same sign as the dispersive signal of the heavy atoms.

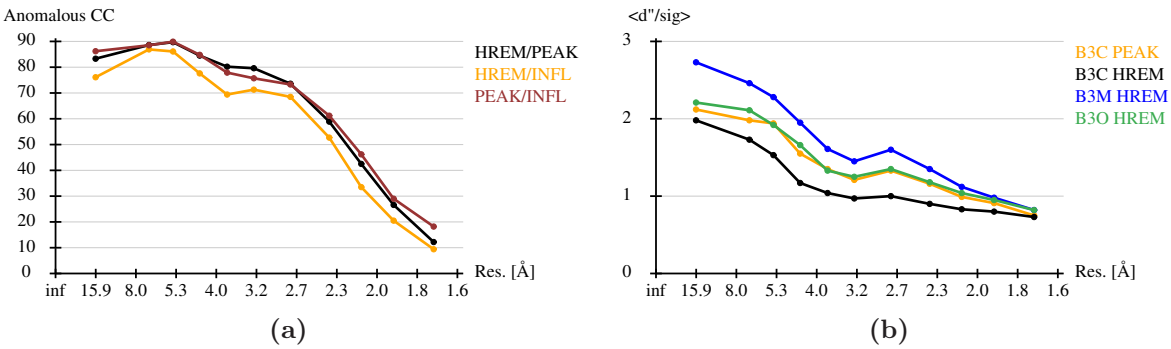


Figure 9.8: Anomalous data statistics for B3C and comparison with B3M and B3O. **(a)** Anomalous correlation coefficient (*CC*) between the B3C MAD data sets. HREM = high-energy remote, INFL = inflection point (data collection wavelengths can be found in Tab. 9.3). The correlation coefficients do not depend on the (estimated) sigma values. Data were truncated at 2.0 Å for heavy-atom substructure solution (where *CC* falls below 30% for all data sets). **(b)** Anomalous signal for the B3C peak and B3C high-energy remote data compared with the high-energy remote data for B3M and B3O (see Fig. 9.6); pure noise would correspond to $d''/\text{sig} \simeq 0.798$. The B3C peak data set contains as much anomalous signal as the B3O high-energy remote data set; however the B3C high-energy remote contains less, probably due to radiation damage (it was the third data set to be collected).

9.4 Binding sites for B3M, B3O and B3C

Refinement was carried out with the *SHELX* suite (Tab. C.4) with restraints for B3M, B3O and B3C generated from the small-molecule crystal structures.

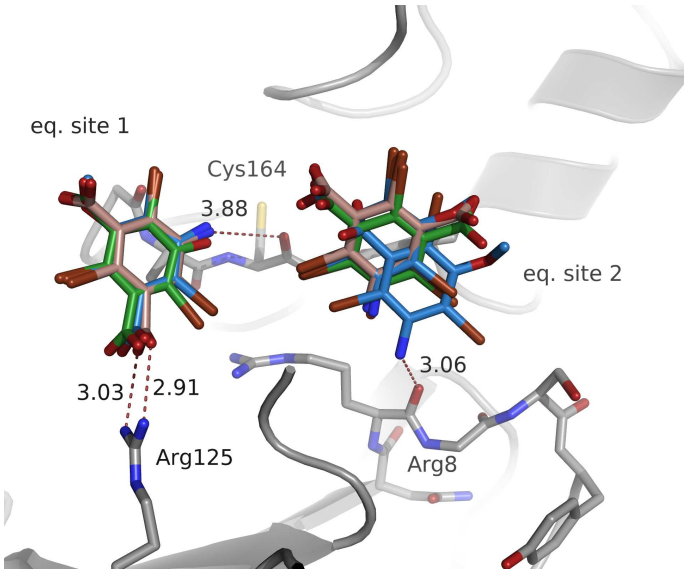


Figure 9.9: Hydrogen bonding for the two equivalent sites 1 and 2, depicted as dashed lines for B3C and B3M. Distances are given in Å. For site 1, the mean occupancy of the three bromine atoms are 0.37 for B3M, 0.41 for B3O and 0.35 for B3C. For site 2, the occupancies are 0.35 for B3M (sum of the two disordered sites), 0.33 for B3O and 0.30 for B3C.

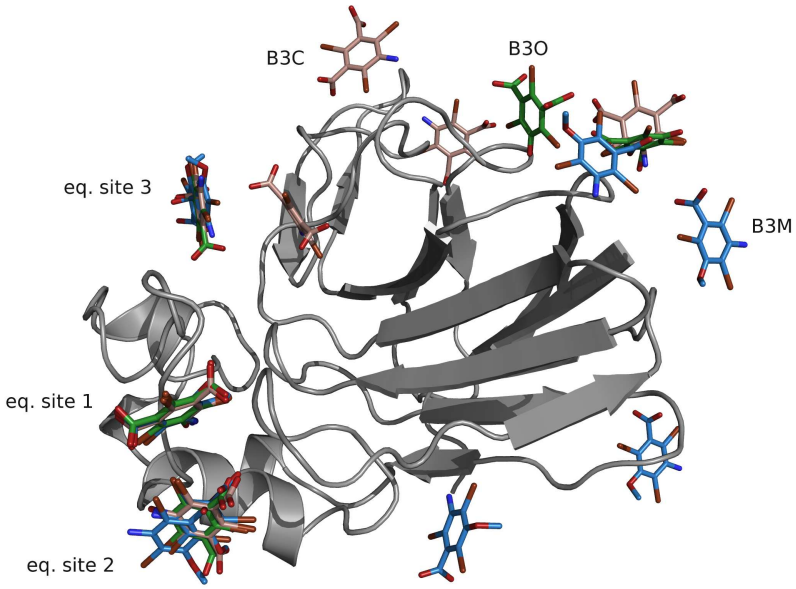
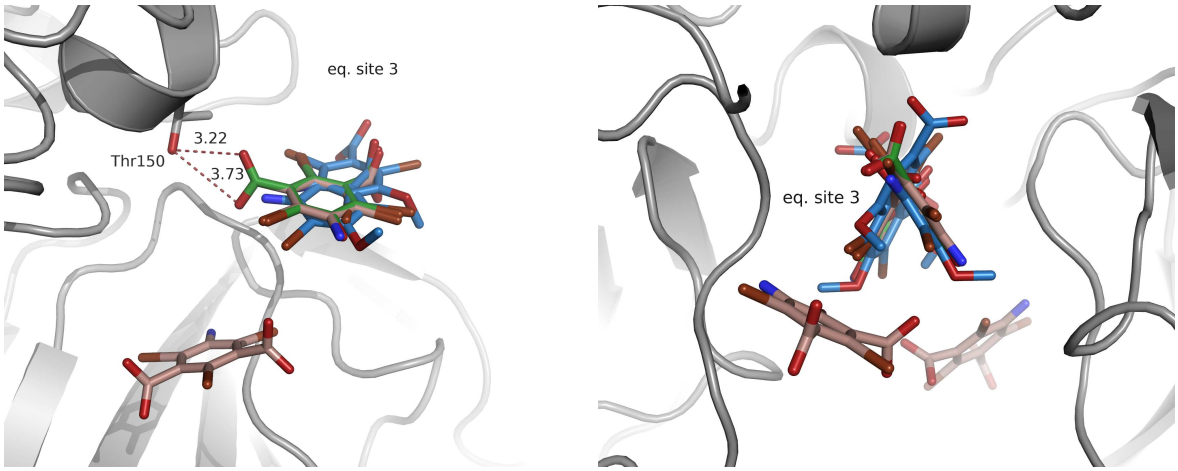


Figure 9.10: Superposition of the three thaumatin structures. B3M molecules are displayed in blue, B3O molecules in green and B3C molecules in pale red. The three sites that are equivalent for the three bromine compounds are labelled.

For refinement those data were used that were collected first from each crystal, i.e. the high-energy remote for B3M and B3O and the peak data set for B3C, to reduce any effect of radiation damage on the occupancy of the bromine sites.



(a) **(b)**

Figure 9.11: Equivalent site 3. **(a)** Hydrogen bonding for B3O is shown as a dashed line. Distances are given in Å. Other interactions are not shown. The B3M molecule is disordered about two sites. There is a binding site for another B3C molecule in close proximity. Occupancies were refined to 0.42 for B3M (sum of the two disordered sites), 0.38 for B3O and 0.37 for B3C (and 0.21 for the other B3C site). **(b)** Different view on the binding site with symmetry equivalent molecules shown. The molecules occupy a special position in the crystal lattice. A twofold axis runs vertical through the molecules and relates the protein and the other B3C molecules to each other.

Free variables were introduced to refine the occupancy of each B3M, B3O or B3C site and later also the occupancy of each single bromine atom separately. For the occupancy refinement, thermal displacement parameters were kept fixed for the bromine atoms (at $B = 19.7\,\text{Å}^2$).[3]

Thaumatin shows seven binding sites for B3M, five binding sites for B3O and seven binding sites for B3C. The mean occupancies for the three bromine atoms of each site are:

B3M	0.42[1]	0.37	0.35[2]	0.29	0.27	0.26	0.22
B3O	0.41	0.38	0.33	0.31	0.25		
B3C	0.37	0.35	0.30	0.22	0.21	0.18	0.12

[1] two conformations with 0.21 each.
[2] two conformations with 0.22 and 0.13.

The binding sites are depicted in Fig. 9.10. There are three sites that are equivalent for all three derivatives. These three sites show high occupancies (Figs. 9.9 and 9.11). The binding sites for B3C in thaumatin are very similar to those found for I3C.

The unique sites show similar hydrogen bonding as found for I3C or the bromine compounds in general, but have lower occupancies (B3M: 0.29, 0.27, 0.26 and 0.22; B3O: 0.31 and 0.25; B3C: 0.22, 0.21, 0.18 and 0.12). There are no clear indications of any hydrophobic contact of the methoxy group in B3M with hydrophobic residues; however the methyl group of the methoxy moiety is free to rotate and therefore difficult to localise in the electron density map.

9.5 Conclusion

In summary, the three bromine compounds B3M, B3O and B3C are very similar in their binding to thaumatin. The occupancies for the compounds do not show large differences. The three major sites are equivalent for all bromine compounds. Trials with other biomolecules have to be carried out to confirm these findings. Although the differences are not significant for thaumatin, soaking with a cocktail of all three bromine compounds (and I3C) may be a valuable option since it will be difficult to predict which compound shows the highest affinity to a protein prior to structure determination.

The similarity of the interaction sites may be accounted for by the similarity of the functional groups. Each bromine compound carries at least one carboxyl group, for which many types of interaction could be observed for each compound. Additionally the exchange of the amino group (B3C) with a hydroxyl group (B3O) did not show any significant differences in binding for thaumatin. B3M shows disorder for two sites and no clear indications of interactions of the methoxy group.

However the similarity in binding may also be attributed to the similar arrangement of the functional groups in the bromine compounds (Fig. 9.12, 1,3,5-substitution of the

[3] The atomic scattering factor of bromine was adjusted to the given wavelength with the *SHELXL* DISP $f'f''$ command. B3M and B3O: $f' = -4.2\,e$, $f'' = 3.75\,e$, theoretical values (Merrit, 2010); B3C: $f' = -6.5\,e$, $f'' = 4.0\,e$, experimental values from fluorescence scan.

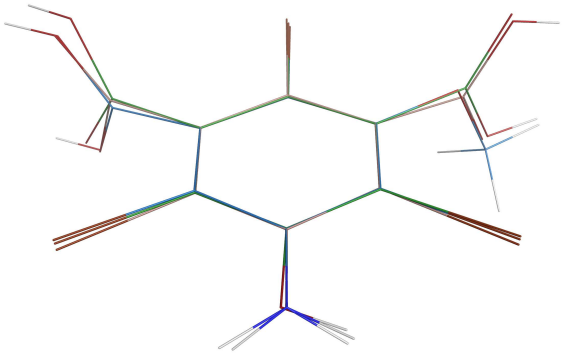
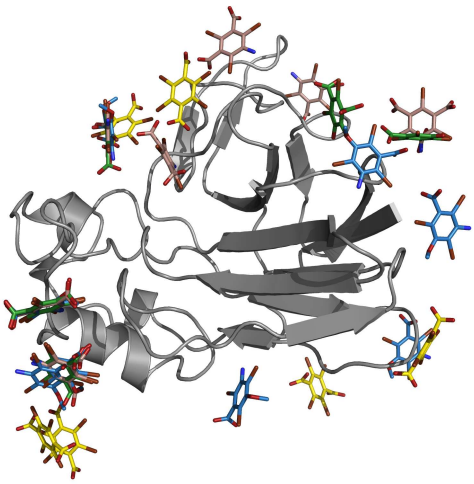


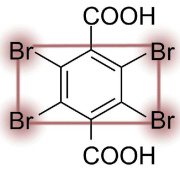
Figure 9.12: Superposition of the molecular structures as determined by X-ray crystallography for B3M, B3O and B3C. Colour code as for the previous figures: B3M in blue, B3O in green and B3C in pale red.

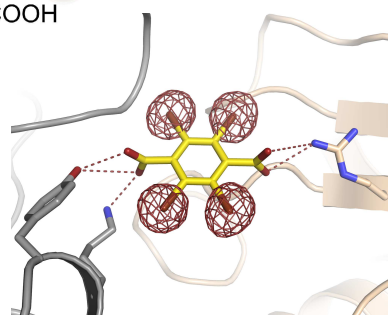
benzene ring). Therefore other substitution patterns have to be tested to investigate if they show different interaction modes.

Chapter 10

B4C - The MAD tetragon







Summary

The MAD tetragon B4C has four bromine atoms for experimental phasing and two carboxyl groups for protein interaction. The molecular structure of B4C was determined by X-ray crystallography.

B4C was incorporated in thaumatin and thermolysin crystals. MAD and SAD phasing was carried out with data collected at the synchrotron. B4C shows different modes of interaction compared with the sticky triangles since the two carboxyl groups in B4C are arranged in *para* position.

10.1 Introduction

The phasing tools that were already discussed are well suited for experimental phasing. However their interaction pattern is very similar due to the arrangement of functional groups (1,3,5-substituted benzene rings). A molecule with substituents at different positions, e.g. 1,4-substitution (*para* arrangement) may show different interaction modes with biological macromolecules. B4C (Fig. 10.1) carries four bromine atoms for experimental phasing and two carboxyl groups in 1,4 position.

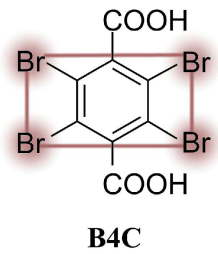


Figure 10.1: B4C with its four bromine atoms is well suitable for MAD experiments. In contrast to the other triangle compounds, the two carboxyl groups are in 1,4 position.

The molecular structure of B4C was determined by single-crystal X-ray analysis and used for the generation of bond length and bond angle restraints.

10.2 B4C crystal structure

In the crystal environment B4C has no internal crystallographic symmetry and crystallises as a dihydrate in the space group $P2_1/n$ with one molecule in the asymmetric unit. The bond lengths and angles fall within normal ranges. A thermal ellipsoid plot is shown in Fig. 10.2.

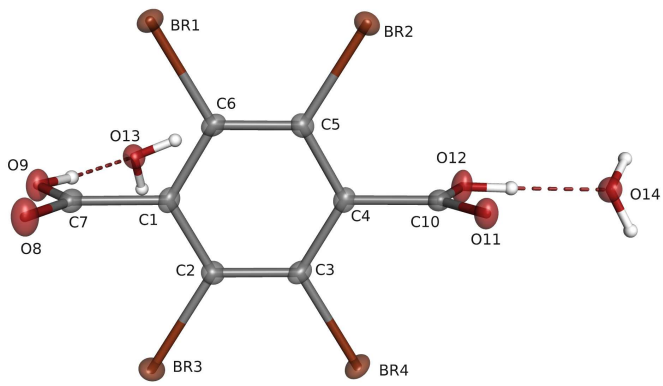


Figure 10.2: The asymmetric unit of B4C. Displacement ellipsoids are drawn at the 50% probability level and hydrogen atoms are shown as small spheres with arbitrary radii. The hydrogen bonds to the lattice water molecules are shown as dashed lines.

The four bromine atoms in B4C form a rectangle with side lengths of about 3.4 and 5.6 Å [cf. Br1−Br2 3.3531(8), Br3−Br4 3.3560(8), Br1−Br3 5.6312(10), Br2−Br4 5.6291(10) Å].

In B4C the carboxyl groups are perpendicular to the ring and clear indications for protonated carboxyl groups can be found. In both carboxyl groups, one C−O bond is

significantly shorter than the other one, indicating one double bond and one single bond [cf. C7−O8 1.216(3) *vs.* C7−O9 1.295(3) Å and C10−O11 1.223(3) *vs.* C10−O12 1.284(3) Å].

In the crystal structure B4C forms hydrogen bonds with all potential donors and the lattice water molecules (Fig.10.3, Tab. 10.1). However, compared with the other molecules discussed, B4C carries only two groups for hydrogen bonding. The hydrogen-bond network is mediated by two lattice water molecules.

Table 10.1: Hydrogen-bond geometry (Å, °) for B4C.

D−H$\cdots A$	D−H	H$\cdots A$	$D\cdots A$	D−H$\cdots A$
O9−H9$\cdots$O13	0.81(3)	1.78(3)	2.591(2)	175(3)
O12−H12$\cdots$O14	0.83(3)	1.68(3)	2.504(2)	178(4)
O13−H13A$\cdots$O11^{i}	0.80(2)	2.00(2)	2.791(3)	172(3)
O13−H13B$\cdots$O8^{ii}	0.80(2)	1.95(2)	2.754(3)	178(3)
O14−H14A$\cdots$O11^{ii}	0.80(2)	2.00(2)	2.790(2)	170(3)
O14−H14B$\cdots$O13^{iii}	0.81(2)	1.94(2)	2.743(3)	172(3)

Symmetry codes: (i) $-x+\frac{1}{2}, y+\frac{1}{2}, -z+\frac{1}{2}$; (ii) $x-1, y, z$; (iii) $x-\frac{1}{2}, -y+\frac{3}{2}, z+\frac{1}{2}$.

There are no π-π interactions in the crystal lattice. These are also not present in the unsubstituted terephthalic acid (Domenicano *et al.*, 1990). There, the carboxyl groups are again in plane with the benzene ring. Additionally, dimeric hydrogen bonding for two carboxyl groups are observed, which is not present in B4C.

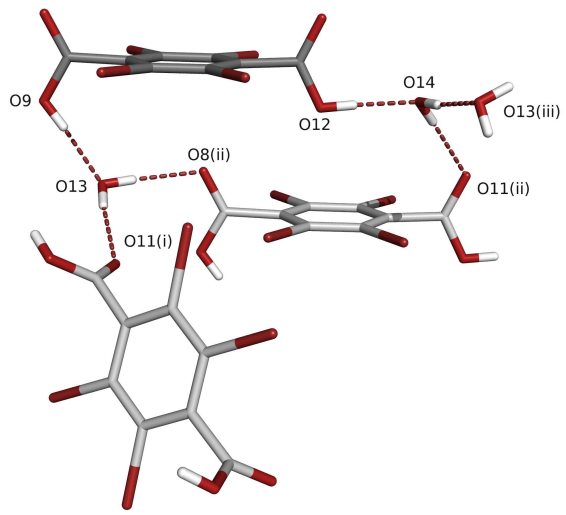


Figure 10.3: Hydrogen bonding of B4C. The hydrogen bonds for hydrogen donor atoms of the central B4C are depicted as dashed lines. Symmetry-equivalent molecules are shown in lighter shade. For symmetry operators, please refer to Tab. 10.1.

10.3 Experimental phasing

B4C was incorporated into thaumatin crystals to investigate if the different arrangement of functional groups in B4C leads to different binding sites in the protein. B4C was incorporated into thermolysin crystals where derivatisation with the sticky triangles was previously

not successful. Additionally the four bromine atoms per molecule in B4C should provide more phasing power than the three bromine atoms in the triangle compounds.

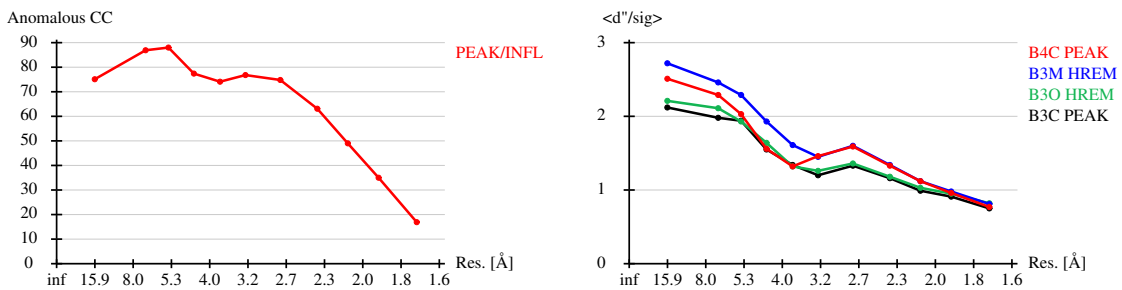
Thaumatin B4C: substructure solution and data analysis

For the thaumatin B4C derivative two data sets (peak and inflection point) were collected for MAD phasing (see Fig. 10.4 for anomalous data statistics).

Table 10.2: Data collection details for B4C in thaumatin and thermolysin. Values in parentheses denote the highest resolution shell.

	Thaumatin		Thermolysin
	Peak	Inflection	High-energy remote
Unit-cell parameters (Å, °)	$a = b = 58.03$, $c = 150.68$		$a = b = 92.75$, $c = 129.28; \gamma = 120$
Space group	$P4_12_12$		$P6_122$
Wavelength (Å)	0.9197	0.9205	0.9131
Photon energy (eV)	13481	13469	13578
Resolution (Å)	24.5−1.60 (1.69−1.60)	24.6−1.60 (1.69−1.60)	24.8−1.80 (1.89−1.80)
Degrees collected (°)	120	120	120
R_{merge}	0.047 (0.17)	0.035 (0.13)	0.092 (0.37)
Completeness (%)	97.7 (87.3)	97.7 (87.0)	96.4 (93.5)
Multiplicity	8.02 (3.77)	8.02 (3.74)	14.1 (13.8)
$< I/\sigma(I) >$	26.8 (6.39)	36.80 (8.88)	22.7 (6.52)

For crystallographic parameters, see Appendix B: Crystallographic parameters.



(a) **(b)**

Figure 10.4: Anomalous data statistics for B4C in thaumatin and comparison with thaumatin B3M, B3O and B3C data. **(a)** Anomalous correlation coefficient (CC) between the B4C MAD data sets. INFL = inflection point (data collection wavelengths can be found in Tab. 10.2). The correlation coefficient does not depend on the (estimated) sigma values. Data were truncated at 2.0 Å for heavy-atom substructure solution. **(b)** Anomalous signal for the B4C peak compared with the high-energy remote data from B3M and B3O and peak data from B3C.

The fluorescence scan shows two peaks close to the absorption edge as for B3C in thaumatin and proteinase K. The peak data were collected not directly at the absorption

edge, but rather at a higher energy (marked as Peak in Fig. 10.5) to maximise the difference in dispersive signal.

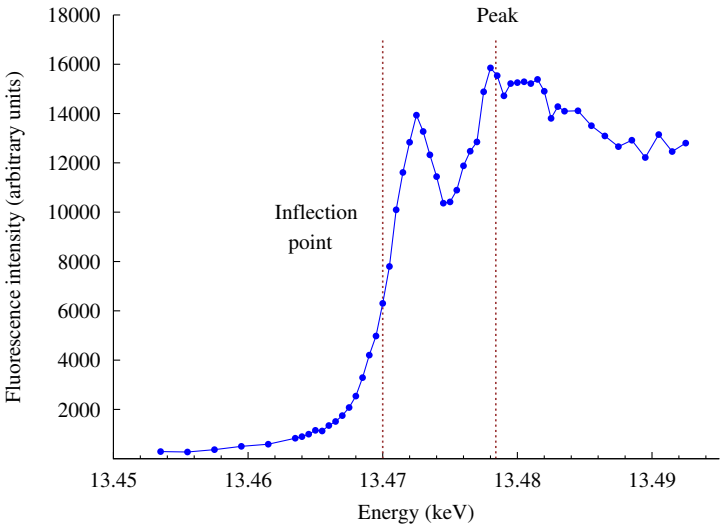


Figure 10.5: Fluorescence scan of thaumatin with B4C incorporated. The data collection energies of peak and inflection point (see Tab. 10.2) are marked by vertical lines. The two peaks close to the bromine K edge are clearly visible.

Data sets were prepared with *XPREP*. Substructure solution with *SHELXD* gave high figures of merit (CC: 46.17, CC_{weak}: 34.69; 596 hits with $CC > 44$ out of 1000 trials) for the MAD data and resulted in a substructure with rectangles with side lengths of about 3.3 and 5.7 Å. SAD phasing with the peak data set alone was also successful and gave high figures of merit (CC: 42.80, CC_{weak}: 23.39; 230 hits with $CC > 41$ out of 1000 trials).

Table 10.3: Comparison of experimental electron density maps obtained for the thaumatin bromine derivatives. The weighted mean phase error (wMPE) and map correlation coefficient (map CC) is given for each data set. Experimental maps were obtained from *SHELXE* with initial phases (only resolving the two-fold ambiguity for SAD, no heavy-atom contribution), after 20 cycles of density modification and after five cycles of autotracing (iterated with density modification).

	initial phases		density modification		density modification and autotracing	
	wMPE (°)	map CC	wMPE (°)	map CC	wMPE (°)	map CC
B3C						
SAD (peak)	54.3	0.510	24.4	0.899	23.8	0.904
MAD (peak, infl.)	48.4	0.579	22.5	0.916	22.4	0.916
MAD (peak, infl., high-energy rem.)	55.6	0.462	24.0	0.904	23.7	0.905
B3M						
SAD (high-energy rem.)	54.9	0.513	25.4	0.885	25.2	0.892
B3O						
SAD (high-energy rem.)	55.3	0.489	23.6	0.899	23.8	0.904
B4C						
SAD (peak)	48.3	0.593	22.5	0.910	23.1	0.909
MAD (peak, infl.)	44.6	0.620	21.2	0.920	21.2	0.920

Tab. 10.3 shows that the MAD experiments yield higher quality electron density maps

(lower mean phase error, higher map correlation coefficient) than the SAD experiments, except for the three-wavelength MAD from the B3C derivative; here radiation damage at the bromine sites may reduce the quality of the phase information. Although the anomalous signal is very similar for bromine triangles and the tetragon data sets (Fig. 10.4b), especially the experimental electron density maps with initial phases (no density modification) obtained from the B4C derivative are of a slightly higher quality (Tab. 10.3).

Thermolysin B4C: substructure solution and data analysis

Data for the thermolysin derivative was collected at $\lambda = 0.9131$ Å (Tab. 10.2), which corresponds to the high-energy remote in a bromine MAD experiment. The anomalous signal of bromine is strong at this wavelength ($f'' = 3.7\,e$; Merrit, 2010).

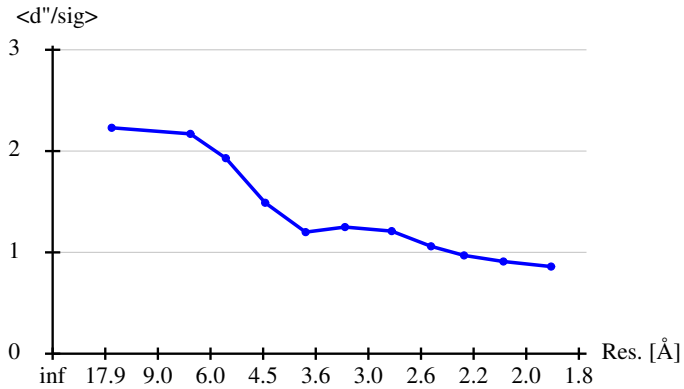


Figure 10.6: Anomalous data statistics for B4C in thermolysin. Data were truncated at 2.40 Å where the signal drops below 1.0.

The anomalous signal is depicted in Fig. 10.6. SAD phasing was carried out with *SHELXD* (CC: 38.96, CC_{weak}: 22.08) and resulted in a substructure with rectangles (side length of about 3.3 and 5.7Å).

Although SAD phasing was successful here at this wavelength close to the absorption edge, it is always recommended to take advantage of the extra phase information that is available from a MAD experiment.

10.4 B4C binding sites

Thaumatin

Refinement against peak data was carried out with the *SHELX* suite (Tab. C.5) with restraints for B4C generated from the small-molecule crystal structure. There are six B4C sites in thaumatin (Fig. 10.7). Free variables were introduced to refine the occupancy of each B4C site and later also the occupancy of each single bromine atom separately. For the occupancy refinement, thermal displacement parameters were kept fixed for the bromine atoms (at $B = 19.7\,\text{Å}^2$). The occupancies (mean value of the four bromine atoms for each molecule) are 0.60 (two orientations with 0.44 and 0.16 each), 0.36, 0.14, 0.14 and 0.11.[1]

[1]The atomic scattering factor of bromine was adjusted to the given wavelength (*SHELXL* DISP $f'f''$ command with $f' = -6.0\,e$ and $f'' = 4.0\,e$, experimental values).

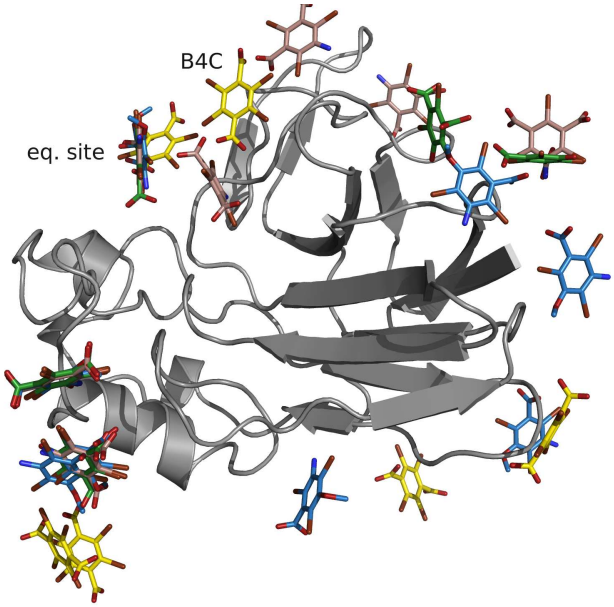
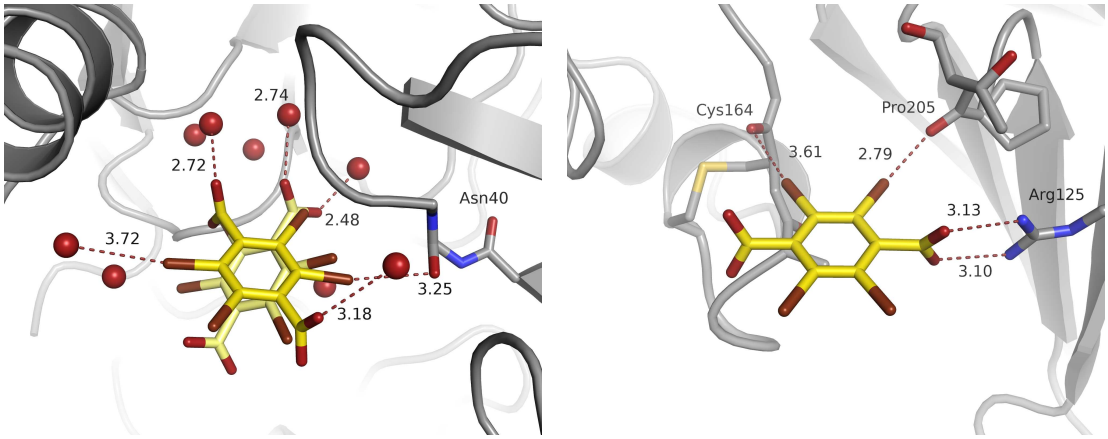


Figure 10.7: B4C in thaumatin; superposition with the three thaumatin triangle derivative structures. B4C molecules are displayed in yellow, B3M molecules in blue, B3O molecules in green and B3C molecules in pale red. In contrast to B3M, B3O and B3C there is only one equivalent site; the other B4C molecules bind to different sites in the protein.

B4C mainly interacts through hydrogen bonds with the protein and water molecules in the crystal lattice. All binding sites except for one differ from those observed for the other bromine derivatives (Fig. 10.7).



(a) **(b)**

Figure 10.8: Two binding sites of B4C in thaumatin. Distances in Å. **(a)** The B4C main site shows two conformations of B4C (yellow: occupancy of 0.44; pale yellow: occupancy of 0.14). Interactions of B4C are mainly formed with lattice water molecules. There are also close contacts of bromine atoms with oxygen atoms (carbonyl group of the main chain, lattice water molecules). **(b)** Interactions for site 6 (occupancy 0.11). In addition to the prominent interaction of the carboxylate group with arginine (Arg125) as found for the triangles there are close contacts of bromine atoms with oxygen atoms (from carbonyl groups of the main chain).

Thermolysin

Refinement was carried out first with *SHELXL* (Tab. C.5) with restraints for B4C generated from the small-molecule crystal structure. Since the data have a moderate resolution (1.8 Å) the initial model was transferred to *REFMAC* and TLS refinement was carried out (Winn *et al.*, 2001).

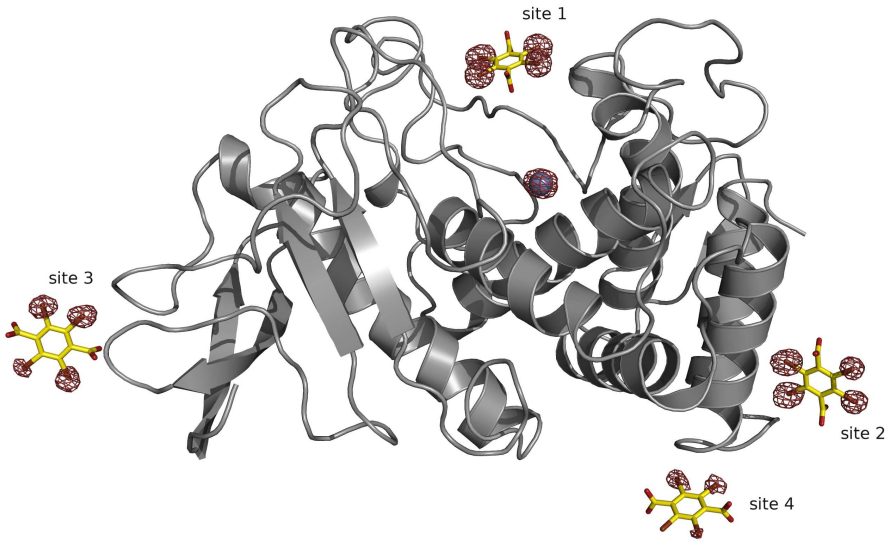


Figure 10.9: B4C in thermolysin. There are four B4C binding sites; anomalous electron density displayed at 4σ (red) around the bromine atoms. The zinc ion at the active site is depicted as a blue sphere.

There are four B4C sites in thermolysin (Fig. 10.9). The occupancies of the B4C molecules were estimated from the initial *SHELXL* refinement to values of 0.5, 0.4, 0.3 and 0.2 and kept at these values for the *REFMAC* refinement.[2]

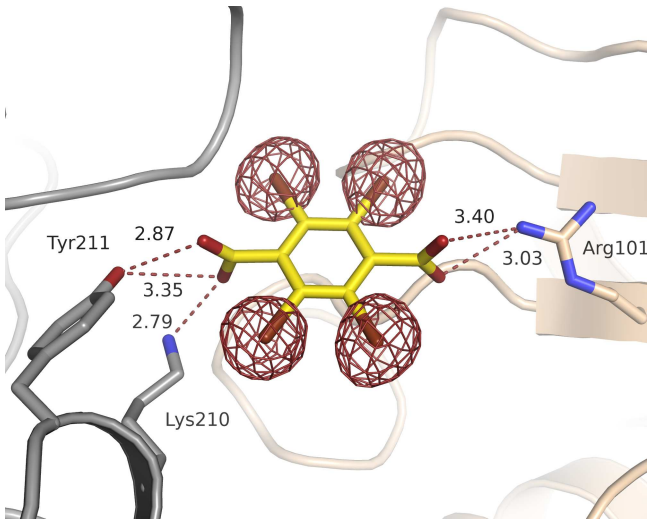


Figure 10.10: Main B4C binding site in thermolysin (occupancy 0.5). Anomalous electron density is shown at 4σ (red), distances in Å. Hydrogen bonds are formed from B4C to two protein molecules (depicted in grey and sand colour).

[2]The atomic scattering factor of bromine was adjusted to the given wavelength [*REFMAC anomalous formfactor* $f'f''$ command with $f' = -4.2\,e$ and $f'' = 3.75\,e$, theoretical values (Merrit, 2010)].

Interestingly incorporation of sticky triangle compounds (B3M, B3O or B3C *via* co-crystallisation or soaking with B3M, B3O, B3C or I3C solutions) was not successful for thermolysin.

10.5 Conclusion

Derivatisation of thermolysin is only possible with B4C and not with any triangle compound. Additionally, B4C binds at different sites in thaumatin. These different binding capabilities can be rationalised by the arrangement of the functional groups in B4C compared to the triangle compounds. The carboxyl groups in B4C have a linear arrangement (1,4 substitution) and span a larger distance than the two carboxyl groups in B3C (Fig. 10.11).

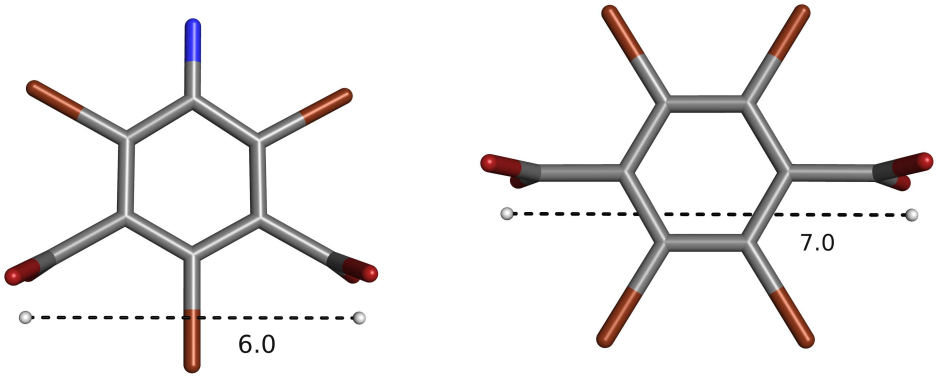
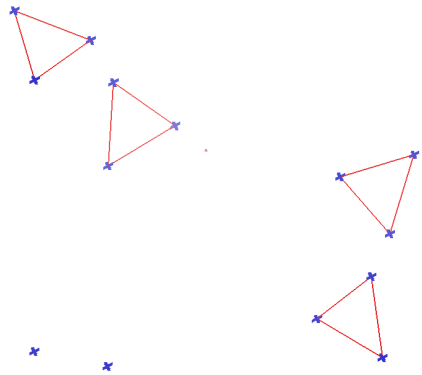


Figure 10.11: Comparison of B3C (*left*) and B4C (*right*). Distances are given in Å between the centroid of each carboxyl group in B3C and B4C.

B4C is very suitable for MAD experiments since it introduces four bromine atoms per molecule in the protein crystal (in contrast to three in B3C) and therefore exhibits a strong anomalous (and dispersive) signal for MAD experiments. The occupancies of the B4C sites are very similar to those observed for the triangle compounds. However the binding as observed for thaumatin and thermolysin differs from the triangle compounds due to the different arrangement of the carboxyl groups and increased distance between the carboxyl groups. B4C may be used as an orthogonal phasing tool when binding of I3C or B3C is not successful. Further advantages are that B4C is readily available from different chemical suppliers and that (as for the sticky triangle compounds) the usage is straightforward and does not involve any toxic chemicals.

Chapter 11

Exploiting the triangles for phasing



```
triangle search in a heavy-atom coordinate file from SHELX

Version 2010/1 (c) Tobias Beck

Using default minimum and maximum distances for B3C search

Reading coordinate file: phase.res

These are the unit cell constants:

 58.082 58.082 150.041 90.0 90.0 90.0

Number of atoms found in the coordinate file: 17

These are the distances within the specified range of 5.2 and 6.2 Angstroms:

['BR01', 'BR02', '5.608']
['BR01', 'BR04', '5.757']
['BR02', 'BR04', '5.597']
['BR05', 'BR12', '5.475']
['BR05', 'BR14', '5.810']
['BR06', 'BR08', '5.688']
```

Summary

The triangular arrangement of the halogen atoms in I3C and B3C provides an independent validation tool for substructure solution. The script *TRIFIND* was developed to facilitate this validation process.

In difficult cases the presence of a triangle in the substructure is an indication of a correct substructure. The information about the geometrical arrangement of the heavy atoms could also be incorporated in the substructure solution process.

11.1 Searching for triangles: *TRIFIND*

When the triangle compounds are used for experimental phasing the substructure contains atom coordinates that form equilateral triangles. These triangles have side lengths of about 6.0 Å for I3C and about 5.7 Å for the bromine compounds. The substructure coordinate file (usually *.res* format from *SHELXD*) may be read into programs like *COOT* or *XP* that can display these files. The heavy-atom sites can be inspected manually and triangles should be visible.

To speed up the substructure validation a simple *PYTHON* script was developed that reads in the substructure atom coordinates and searches for triangles. The *TRIFIND* script involves three steps:

- parse the coordinate file and calculate distances
- find triangles (within a specified side length range)
- open *COOT* and display the triangles (optional)

The script may also read the revised heavy-atom coordinate file from *SHELXE* (with file ending *.hat*; the file format is equivalent to *.res*).

The script is invoked from the command line. First the substructure coordinate file is read and the atom coordinates are extracted along with the unit cell constants and the symmetry operators (always present in *SHELX .res* or *.hat* files). Distances are calculated between all atoms, but are only stored and printed to the terminal if the distance is within a specified range. Triangles are present if two atoms that are within the specified range share entries with an equivalent neighbour. A list of triangles (and two-atom fragments) is printed to the terminal. A typical output is shown in Fig. 11.1.

The user only has to enter the coordinate file stem (file name without *.res* or *.hat*) and the range for the side lengths of the triangles (minimum and maximum distance in Å). Instead of giving two numbers the keywords `i3c` or `b3c` can be provided for a default range suitable for the corresponding phasing tools.

The distance between two atoms at (x_1, y_1, z_1) and (x_2, y_2, z_2) is calculated by:

$$\begin{aligned} d^2 = (\Delta x\ a)^2 + (\Delta y\ b)^2 + (\Delta z\ c)^2 &+ 2\ bc\ \Delta y \Delta z\ \cos(\alpha) \\ +2\ ac\ \Delta x \Delta z\ \cos(\beta) + 2\ ab\ \Delta x \Delta y\ &\cos(\gamma) \end{aligned} \tag{11.1}$$

a, b, c, α, β and γ are the unit cell constants, which are read from the coordinate file. The atomic coordinates are given as fractional coordinates in *SHELX* format, therefore no conversion is necessary. The script requires a version of *PYTHON* to be installed.

```
   _______    _    ___ _           _
  (_______)  (_)  / __|_)         | |
     _  ____ _ _| |__ _ ____   __| |
    | |/ ___) (_   __) |  _ \ / _  |
    | | |   | | | |  | | | | ( (_| |
    |_|_|   |_| |_|  |_|_| |_|\____|

    triangle search in a heavy-atom coordinate file from SHELX

    Version 2010/1 (c) Tobias Beck

 Using default minimum and maximum distances for B3C search

 Reading coordinate file: phase.res

 These are the unit cell constants:

  58.082 58.082 150.041 90.0 90.0 90.0

 Number of atoms found in the coordinate file: 17

 These are the distances within the specified range of 5.2 and 6.2 Angstroms:

['BR01', 'BR02', '5.608']
['BR01', 'BR04', '5.757']
['BR02', 'BR04', '5.597']
['BR05', 'BR12', '5.475']
['BR05', 'BR14', '5.810']
['BR06', 'BR08', '5.688']
['BR06', 'BR11', '5.858']
['BR07', 'BR10', '5.502']
['BR08', 'BR11', '5.675']
['BR09', 'BR15', '5.921']
['BR09', 'BR16', '5.885']
['BR12', 'BR14', '5.502']
['BR15', 'BR16', '5.767']

 There are 4 triangles. These are the atoms that are part of each triangle:

['BR01', 'BR02', 'BR04']
['BR05', 'BR12', 'BR14']
['BR06', 'BR08', 'BR11']
['BR09', 'BR15', 'BR16']

 And these are the atoms that form atom pairs:

['BR07', 'BR10']

 Open COOT and display the triangle(s) found? y/n [y] y

 Display also the two-atom pairs in COOT? y/n [n]
```

Figure 11.1: Terminal output of *TRIFIND* for the substructure coordinate file of thaumatin B3M. The script was started with the command `trifind phase b3c` from the terminal.

The output from *TRIFIND* may be displayed in *COOT* (Fig. 11.2) to inspect the triangles found. In addition, the two-atom fragments can also be highlighted.

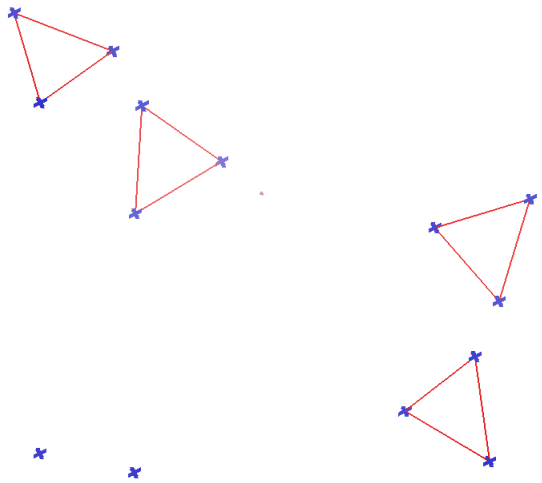


Figure 11.2: Screenshot of *COOT* displaying the four triangles found by *TRIFIND* in the substructure file from thaumatin B3M. The two-atom fragments can also be highlighted (Fig. 11.1).

11.2 Difficult structures with I3C

Three examples are discussed where the triangular nature of the phasing tool I3C was crucial for structure solution.

Ribonuclease A - Look for triangles in the substructure

A quick soak was carried out to incorporate I3C into ribonuclease A. Data were collected in-house at CuKα (Tab. 11.1).

Table 11.1: Data collection details for ribonuclease A with I3C. Values in parentheses denote the highest resolution shell.

Unit-cell parameters (Å, °)	$a = b = 63.88, c = 63.53$; $\alpha = \beta = 90, \gamma = 120$
Space group	$P3_221$
Wavelength (Å)	1.54178
Resolution (Å)	28.5−1.80 (1.90−1.80)
R_{merge}	0.086 (0.36)
Completeness	93.4 (86.2)
Multiplicity	5.31 (3.14)
$< I/\sigma(I) >$	12.03 (2.63)
$R_{\text{anom}}/R_{\text{p.i.m.}}$	2.64 (2.12)

For crystallographic parameters, see Appendix B: Crystallographic parameters.

The anomalous signal is weak for this data set (Fig. 11.3) and the ratio R_{anom} to $R_{\text{p.i.m.}}$ is quite low (compared with other I3C data sets like lysozyme I3C and thaumatin I3C). The anomalous data were truncated to 2.6 Å. Substructure solution with *SHELXD* resulted in a

substructure with low figures of merit (CC: 23.59, CC_{weak}: 7.76). One triangle is present, which consists of peaks I1, I2 and I5. The presence of a triangle is a reliable criterion that substructure solution was successful. However standard procedures for phase extension with *SHELXE*, i.e. using all heavy atoms from the substructure coordinate file (switch *-h*), did not result in an interpretable map. The autotracing in *SHELXE* did also not improve the results. However phase extension with the first two peaks only was successful (switch *-h2*, i.e. selects only the first two peaks from the heavy-atom file). Hand-editing the file to include only the triangle (peaks I1, I2 and I5) and phase extension with the modified file was also successful. These trials resulted in a decent experimental map (the autotracing algorithm in *SHELXE* placed 110/124 residues). Refinement was carried out with *REFMAC* with TLS parameters (Tab. C.6). One I3C binding site was found and its occupancy was set to 0.5.

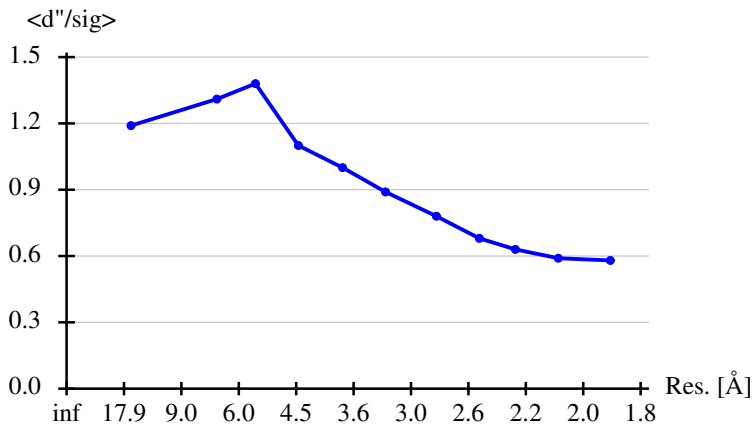


Figure 11.3: Anomalous data statistics for I3C in ribonuclease A. The anomalous signal is weak for this data set. Data were truncated at 2.6 Å where the signal drops clearly below the noise level of ≈ 0.8.

Analysis of the peak positions from the substructure shows that the peaks I1, I2 and I5 are indeed the three iodine atoms from I3C. The other peaks are not positions of other anomalous scatterers (e.g. sulfur atoms in cysteine/methionine residues or chloride ions in the solvent region) as could be expected for data collected at CuKα, but are false peaks. Recycling of the improved heavy-atom file from *SHELXE* (*.hat* was copied to *.res*) and another round of density modification yielded an improved experimental map with even more residues being placed (122 out of 124). From the seven peaks of the final heavy-atom substructure file three belong to the I3C triangle and only one peak to a sulfur atom from a disulfide bond and one to a chloride ion (out of eight sulfur atoms in cysteine residues, four in methionine residues and four chloride ions). The remaining peaks are again false positives.

In many cases phase extension with all atoms present in the heavy-atom substructure file results in a better experimental map than phasing with the strong peaks only. The phase information even from weak anomalous scatterers (low anomalous scattering at this wavelength or low occupancy) usually improves the results. However if the peaks from the substructure solution do not correspond to any heavy atom in the final structure (false positives) it is important to identify the correct peaks and use only those for phasing. The triangular arrangement of iodine atoms in I3C provides a convenient way to identify the correct atoms for phasing, but it is also an important general validation tool for substructure solution.

SAD phasing of a novel structure - Recycle the triangles

Data were kindly provided by Christoph Parthier, Ulrike Bräuer and Milton Stubbs from the Martin-Luther-Universität in Halle-Wittenberg, Germany.[1] They had collected in-house data (CuKα, 1.54178 Å) from a protein crystal derivatised with I3C.

Analysis of the substructure solution (SAD phasing) revealed that two triangles and two two-atom fragments in the range of 5.5 to 7.0 Å were present. The figures of merit (CC: 28.27, CC_{weak}: 16.59) indicated that the heavy atom-coordinates were determined correctly. However phase extension and density modification resulted only in a mediocre map.

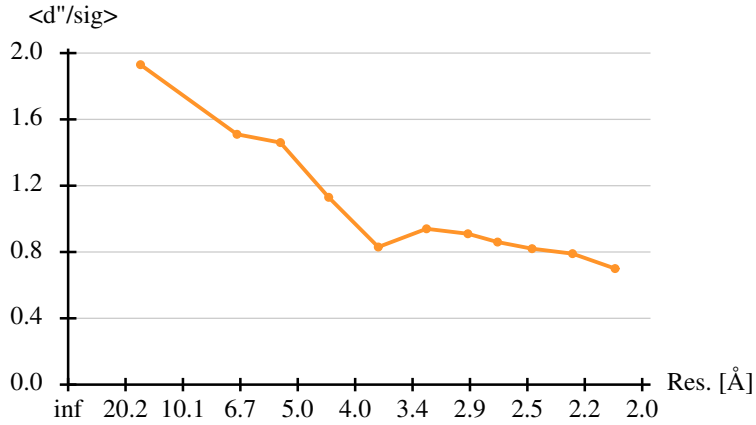


Figure 11.4: Anomalous data statistics for the I3C data from Halle; resolution: 2.10 Å. The data show a strong anomalous signal at low resolution. Data were truncated at 3.0 Å for substructure solution. Data courtesy of Christoph Parthier, Ulrike Bräuer and Milton Stubbs from the Martin-Luther-Universität in Halle-Wittenberg, Germany.

Phase extension and density modification in conjunction with the autotracing algorithm (40 cycles) in *SHELXE* and exploiting the twofold NCS resulted in an improved electron density map, but still only ~65% of the residues (main chain) could be traced. Recycling of the heavy-atom positions and another round of autotracing iterated with density modification with *SHELXE* gave a good experimental electron density map (~80% of all main chain residues were placed).

Phase extension carried out by the Stubbs group using our final model from *SHELXE* and native data (higher resolution), which they had collected at the synchrotron, improved the experimental map and allowed >95% of the residues to be traced. This novel protein structure solved with I3C is to be published. The binding sites of I3C are briefly discussed in Appendix A.

SIRAS phasing of a novel structure - One triangle is there

Data were kindly provided Jeanine Amacher and Dean Madden from Dartmouth Medical School, Hanover, USA.[1] They had collected data at the synchrotron from an I3C derivatised crystal at $\lambda = 1.75$ Å extending to 1.96 Å resolution and native data at $\lambda = 0.9181$ Å extending to 1.40 Å resolution.

Data were prepared for SIRAS phasing with *XPREP* and substructure solution was carried out with *SHELXD*. Analysis of the substructure coordinates reveals one triangle

[1] Data and structure are not published yet; therefore only few data statistics are given.

(peaks I1, I2 and I5). However the figures of merit (CC: 19.35, CC_{weak}: 11.98) were quite low.

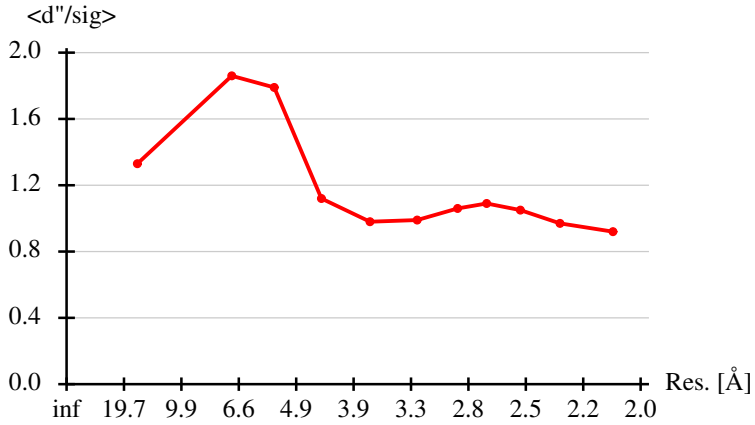


Figure 11.5: Anomalous data statistics for the I3C data from Hanover, USA; resolution: 1.96 Å. The data show a weak anomalous signal. Data were combined for a SIRAS experiment with the native data. Data courtesy of Jeanine Amacher and Dean Madden from Dartmouth Medical School, Hanover, USA.

Phase extension and autotracing combined with density modification in *SHELXE* was successfully carried out and >80% of the main-chain residues could be traced.

Structure solution with *HKL2MAP* (with *SHELXC* and *SHELXD*) gave no suitable heavy-atom coordinate file. Different resolution cut-offs for the difference data were also tested. For these substructure files either no distances were observed within the I3C range or if so they deviated clearly from 6.0 Å. The scaling of the dispersive differences is done differently in *SHELXC* and *XPREP* (G.M. Sheldrick, *private communication*), with the latter usually giving better results. Therefore in borderline cases as seen here *XPREP* should be used to prepare the difference data for *SHELXD*. This novel protein structure solved with I3C is to be published. The binding site of I3C is briefly discussed in Appendix A.

The strategy employed here reflects the recommendations for the usage of I3C. Anomalous data for I3C derivatised crystals should be collected at longer wavelengths, i.e. >1.50 Å, to exploit the strong anomalous signal of iodine at these wavelengths (CuKα gives good results). Data collection at long wavelengths suffers to an increased extent from radiation absorption issues such as radiation damage to the iodine sites or air absorption. Therefore a compromise has to be found and a wavelength around 1.7 −1.8 Å seems to be a good choice. In addition to the anomalous data set a data collection at shorter wavelength is recommended with a native crystal to extend the resolution. Provided that the two crystals are isomorphous SIRAS phasing can be carried out. For weakly diffracting crystals MAD phasing is necessary since density modification is not able to improve the phases as much as it can at medium or high resolution. For these cases phasing with the three bromine atoms in B3C or the four bromine atoms in B4C is recommended.

11.3 Tests with *SHELXD*

The triangular arrangement of the heavy atoms in the sticky triangle compounds provides a good additional quality measure for substructure solution. The presence of triangles (or at least two-atom pairs) within the correct distance range is a reliable criterion that the substructure is correct. In difficult cases the figures of merit (e.g. CC and CC_{weak}

in *SHELXD*) may not clearly indicate a true solution, i.e. there is only a small difference between CC/CC_{weak} for a correct substructure solution and wrong solution. Although it is very helpful to have an independent test for identifying a correct solution, it may be valuable to include the structural information about the equilateral triangles in the substructure solution process.

Initial tests were carried out with *SHELXD* to investigate if substructure solution can be improved. Different options already present in *SHELXD* (GROP, PATS) and a test version with a new algorithm (MAGI) were tested. The GROP option reads in a molecule fragment (in PDB format) and *SHELXD* uses this fragment as a seed for a six-dimensional small rigid group search; in practice, this search is rather slow. The PATS option allows to specify the distance of the vector used for Patterson seeding. The MAGI command assigns a higher figure of merit to those solutions where a certain distance, e.g. 6 Å, is present. This was chosen to filter out solutions that otherwise would be lost since their CC values are worse than those from other (false) solutions. These options were tested for straight-forward structures (e.g. lysozyme I3C) but also for difficult data (e.g. ribonuclease A). However these approaches did not give better results than the default settings in *SHELXD*, i.e. no lower r.m.s.d. values for the substructure solution in comparison with the heavy-atom coordinates from the final refined structures, provided that a sufficient number of trials were carried out.

The search for triangles could be carried out in real space, i.e. in the peak-search routine that is used in each dual-space refinement cycle in *SHELXD*; this is similar to the resolution of super-sulfurs (DSUL option) where one peak at low resolution is resolved into a disulfide bridge (Debreczeni *et al.*, 2003b). For triangles the peak-picking procedure could be modified such that those peaks that have another peak within a specified distance are retained whereas isolated peaks are discarded. Since especially at longer wavelengths other anomalous scatterers such as sulfur could be present in addition to the iodines from I3C it would be reasonable to retain also the highest isolated peaks. The fraction of peaks belonging to a triangle to isolated peaks could be refined within the program or defined by the user. It is important to note that the triangle information would not be exploited completely at this point. Due to radiation damage or high thermal displacement only two atoms of a triangle might contribute to anomalous scattering; in those cases a triangle search would fail but the distance criterion in the peak-picking step would still identify both peaks. The peak-picking should be treated as a restraint rather than a constraint, i.e. no specific number of peaks has to fulfil the distance criterion because this depends on the structure and data quality.

Another possibility is to write out a new coordinate file every time a certain distance, e.g. 6.0 Å, is present in the heavy-atom coordinates. This would facilitate an automated approach where each separate heavy-atom substructure could be expanded with *SHELXE*. This procedure is analogous to the approach in *ARCIMBOLDO* (Rodriguez *et al.*, 2009).

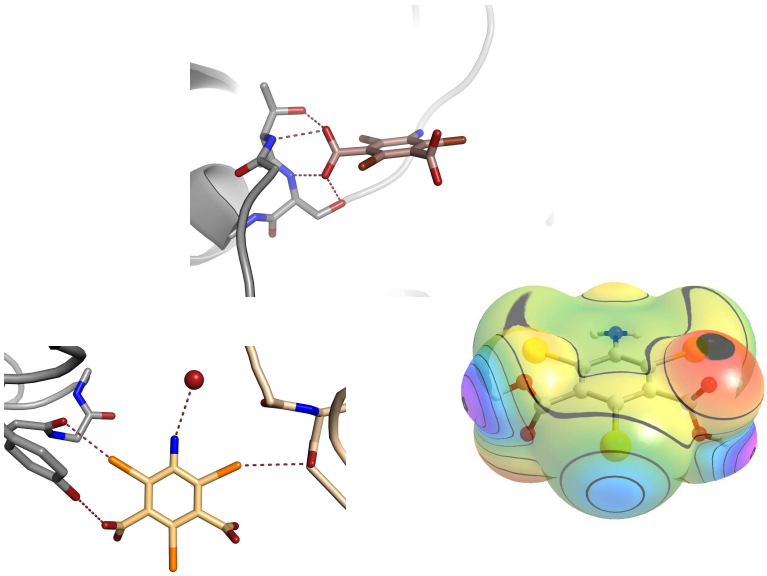
11.4 Conclusion

The triangular arrangement of the heavy atoms in the phasing tools provides an independent test for substructure validation: if triangles are present in the substructure coordinate

file the substructure is probably correct. The validation can readily be carried out with the *TRIFIND* script. For difficult structures this validation proved to be crucial to ensure that the substructure solution was correct and if necessary further adjustments were required at the later steps of the structure solution process. Tests with *SHELXD* were carried out to investigate if the distance information can be exploited in the substructure solution process, but there are still suggestions for modified search procedures that could be tested.

Chapter 12

What makes the triangles sticky?



Summary

The types of potential interactions between the phasing tools and other molecules are discussed. The compounds form hydrogen bonds to protein residues and water molecules. The halogen atoms show short non-covalent contacts with oxygen atoms.

The number and type of contacts are analysed quantitatively. Most interactions involve carboxylate groups. Since the phasing tools mainly replace lattice water molecules upon derivatisation, the changes in the final refined structures are small compared with native structures.

The phasing tools show several crystal contacts in the protein structures where they bridge different protein molecules. It is concluded that the inclusion of phasing tools in crystallisation screens might be beneficial for crystal growth.

12.1 Interaction modes of the phasing tools

Hydrogen bonding

Intermolecular contacts such as hydrogen bonds are generally weaker than intramolecular, covalent bonds. These weak interactions determine molecular conformation or molecular aggregation (Jeffrey and Saenger, 1991; Steiner, 2002) and are by far the most important specific interaction in biological systems (Bissantz *et al.*, 2010).

Hydrogen bonds generally consist of a group that donates a hydrogen atom (hydrogen donor) and an atom that accepts this hydrogen atom (hydrogen acceptor, Fig. 12.1).

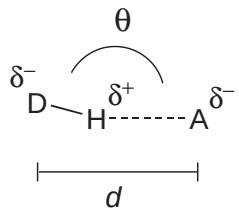
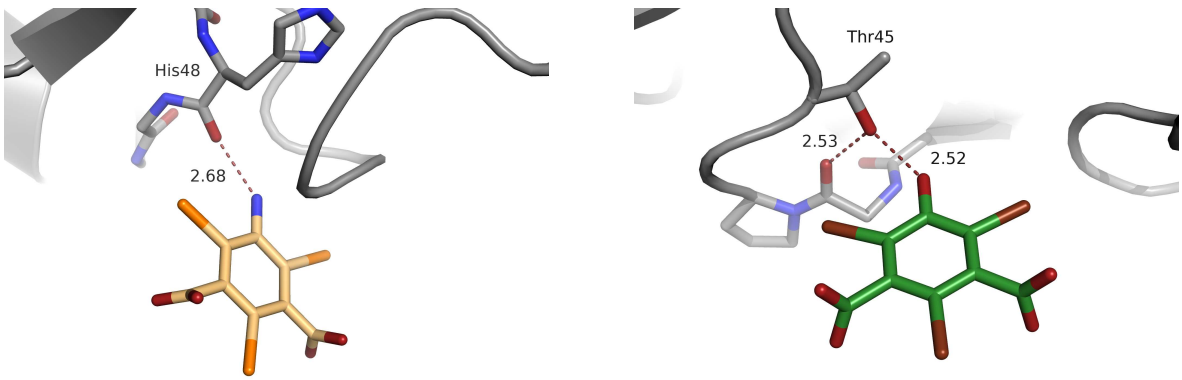


Figure 12.1: Schematic representation of a hydrogen bond; $D =$ donor, $A =$ acceptor. For strong interactions the distance d is in the range of 2.3 and 3.6 Å. The angle θ is usually $> 150°$.

The hydrogen bond is a complex interplay of several energetic contributions. It can conceptually be subdivided into electrostatic, polarisation, charge transfer, dispersion and exchange repulsion contributions. These depend differently on distance and interaction angle; therefore the portion of each contribution is different for each hydrogen bond (Steiner, 2002). The electrostatic contribution is directional and has a long range effect; it falls off with r^{-2} for dipole-monopole interactions. The hydrogen bond may also be described as a three-centre (donor, hydrogen and acceptor atom) four-electron bond.



(a) **(b)**

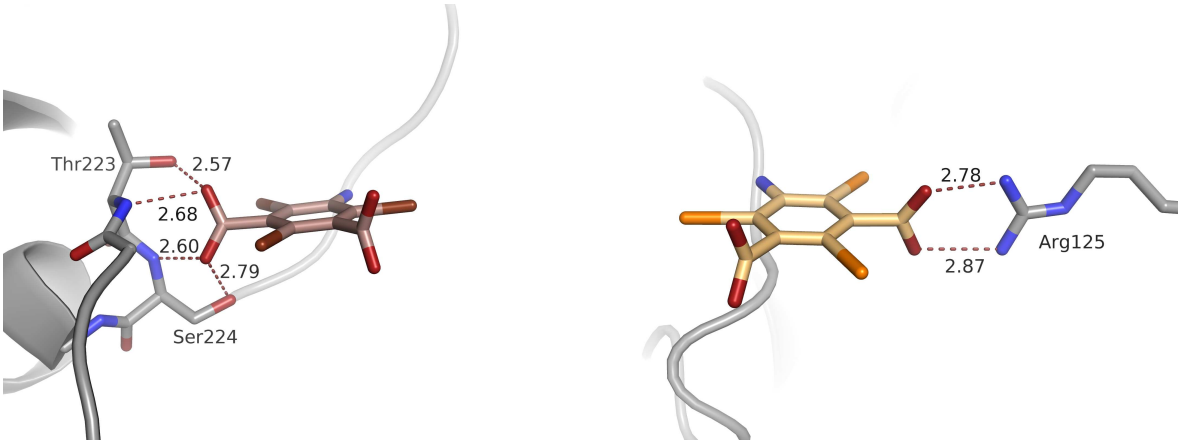
Figure 12.2: Hydrogen bonding between neutral partners. Hydrogen bonds are depicted as dashed lines and distances are given in Å. **(a)** The binding site of I3C in ribonuclease A. The amino group forms a hydrogen bond to the carbonyl oxygen of His48. **(b)** One molecule of B3O in thaumatin forms a hydrogen bond to Thr45. The hydroxyl group of B3O probably acts as a hydrogen donor since Thr45 also forms a hydrogen bond to a carbonyl oxygen (hydrogen acceptor) of the main chain.

The bonding strength increases either if the donor-hydrogen bond is strongly polarised towards the donor or if the acceptor is negatively polarised (e.g. both D and A are elec-

tronegative atoms). However weak hydrogen bonds can be found where the donor-hydrogen bond is only weakly polarised, e.g. in C–H. A hydrogen acceptor can be a delocalised π-system, e.g. phenylalanine residues.

Two examples for hydrogen bonds between neutral partners[1] that involve the phasing tools are shown in Fig. 12.2. The amino (I3C, B3C, B3M) and hydroxyl groups (B3O) may act as a hydrogen donor or acceptor group, but in some cases the assignment may not be clear (e.g. Fig. 12.2b). Other examples can be found in Fig. 7.7a (p. 59), Fig. 8.11 (p. 73) and Fig. 9.9 (p. 82).

The strength of a hydrogen bond increases further if one or both of the partners are charged ("salt bridges" in proteins; Jeffrey and Saenger, 1991). However the desolvation energy for charged residues is larger than for neutral residues and has to be taken into account, also for protein-ligand interactions. These hydrogen bonds usually show shorter distances d between the donor and acceptor group.



(a) **(b)**

Figure 12.3: Hydrogen bonding that involves charged residues. Hydrogen bonds are depicted as dashed lines and distances are given in Å. **(a)** One molecule of B3C in proteinase K. The carboxylate group acts as an hydrogen acceptor and forms hydrogen bonds with the amide protons of the main chain and the hydroxyl groups of Thr223 and Ser224. **(b)** One molecule of I3C in thaumatin. The carboxylate group binds to Arg125 and forms an ion pair (salt bridge).

Lysine or arginine residues are commonly protonated at neutral pH values. Examples for hydrogen bonds with charged partners[2] are shown in Fig. 12.3. The carboxylate group also interacts with neutral groups, such as the main chain amide protons or hydroxyl groups found in serine, threonine or tyrosin (e.g. Fig. 12.3a). Other examples for hydrogen bonds that involve charged functional groups can be found in Fig. 7.7b (p. 59), Fig. 8.8b (p. 70), Fig. 8.11 (p. 73), Fig. 9.9 (p. 82), Fig. 10.8 (p. 93) and Fig. 10.10 (p. 94).

[1]The pK_A value for the amino group in the phasing tools could not be determined experimentally. A comparison with similar compounds results in an estimated value of $pK_A < 4$ for the $-NH_2$ group. Therefore at neutral pH the amino group is not protonated.

[2]The pK_A value for the carboxyl groups could not be determined experimentally. A comparison with similar compounds results in an estimated value of $pK_A < 4$ for both –COOH groups. Therefore the carboxyl groups are almost fully deprotonated at neutral pH.

Many interactions of the functional groups of the phasing tools involve hydrogen bonds with water molecules. These water molecules usually constitute the first solvation shell around the protein. They are usually well ordered since they form hydrogen bonds to residues at the surface of the protein.

Halogen bonding

The halogen atoms of the phasing tools show several short contacts with electronegative moieties of the protein. This can be rationalised by an attractive interaction that is based on the aspherical electrostatic potential of the halogen atoms. Theoretical studies reveal that halogen atoms show a positive electrostatic potential opposite of the halogen σ bond (Politzer *et al.*, 2007, 2010). This σ hole (Fig. 12.4) may interact with electronegative atoms.

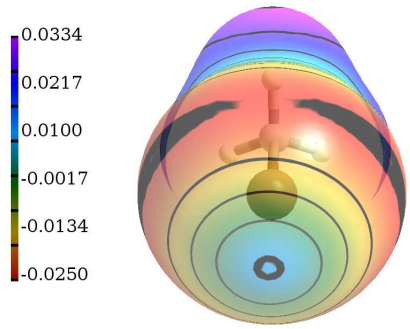


Figure 12.4: Computed electrostatic potential, MP2/DGDZVP, for CH_3Br in a.u. mapped on the surface of molecular electron density at 0.001 a.u.. The bromine atom shows a region of positive electrostatic potential surrounded by a ring of negative potential. Figure prepared with *MOLISO* (Hübschle and Luger, 2006) within *MOLECOOLQT* (Hübschle, 2010).

Halogen bonding involves a halogen atom that is attached to a donor group, usually a carbon atom, and an acceptor atom that is electronegative (Fig. 12.5). The interaction is related to the maximum value of the local positive surface potential $V_{S,max}$ at the halogen atom. $V_{S,max}$ depends on *(a)* the electronegativity of the halogen atoms, *(b)* its polarisability and *(c)* any electron-withdrawing groups (Politzer *et al.*, 2010). Therefore $V_{S,max}$ increases in the order F < Cl < Br < I. Electronegative substituents at the halogen atom or in the remainder of the molecule can further increase $V_{S,max}$.

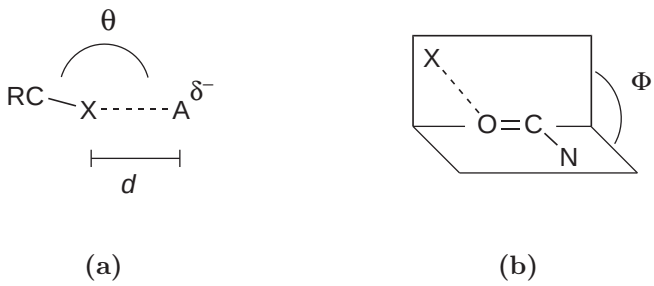


Figure 12.5: Schematic representation of halogen bonding. **(a)** Halogen bonding geometry. The angle θ is usually $>150°$. The distance d depends on the halogen atom X, but is usually smaller than the sum of the van der Waals radii of X and acceptor A. **(b)** The dihedral angle Φ is found to be close to 90° for interactions with carbonyl oxygen atoms (Auffinger *et al.*, 2004). The halogen atom may interact with the π system of the peptide bond.

For halogen bonding the distance d in Fig. 12.5 is found to be smaller than the sum of the van der Waals radii, i.e. $d < r_{vdW}(\mathrm{Br}\cdots\mathrm{O}) = 3.37\,\text{Å}$ or $d < r_{vdW}(\mathrm{I}\cdots\mathrm{O}) = 3.50\,\text{Å}$ with $r_{Br} = 1.85\,\text{Å}$, $r_I = 1.98\,\text{Å}$ and $r_O = 1.52\,\text{Å}$ (Bondi, 1964).[3] The angle θ was postulated to be close to 180°, i.e. a linear arrangement of the carbon-halogen-acceptor atoms is present (Politzer *et al.*, 2007). However investigations of halogen bonding in biological macromolecules show that the angle has a bimodal distribution with peaks at 160-170° and 145-150°. The second maximum is rationalised by polarisation effects induced by the acceptor atom (e.g. carbonyl oxygen) that deform the electron density at the halogen atom (Auffinger *et al.*, 2004). For the interaction of a halogen atom with the carbonyl oxygen involved in a peptide bond the dihedral angle Φ is found to be close to 90°. Here the positive electrostatic potential of the halogen atom interacts with the π system of the peptide bond. Therefore halogen bonding does not necessarily compete with hydrogen bonding since hydrogen bonds are formed to the oxygen electron lone pairs, which are in plane with the peptide bond. Halogen bonding may be seen as an orthogonal molecular interaction to hydrogen bonding (Voth *et al.*, 2009). Both can be considered as interactions of a positive electrostatic potential, the σ-hole, with an electronegative partner. Since the positive electrostatic potential at the hydrogen atom is more smeared out (due to the s character of the potential) hydrogen bonding is considered less direction-dependent than halogen bonding.

To study halogen bonding in more detail the electrostatic potential of the phasing tools was evaluated. Since the protonation state can not clearly be determined from the macromolecular crystal structure the neutral molecule and the two anions for each phasing tool were studied. Molecular models of I3C, B3C, B4C and their anions (one or both carboxyl groups deprotonated) as well as CH_3Br were subjected to geometry optimisation and frequency calculations with MP2/DGDZVP (Clark *et al.*, 2007) in *GAUSSIAN 09* (Frisch *et al.*, 2009).[4] The electrostatic potential was mapped on the isosurface of molecular electron density at 0.001 electrons/bohr^3 (= a.u.) with *MOLISO* (Hübschle and Luger, 2006) within *MOLECOOLQT* (Hübschle, 2010).

Fig. 12.6 shows the electrostatic potential for each phasing tool in three protonation states. The same colour scale was used only for each protonation state, e.g. the electrostatic potential of neutral I3C, neutral B3C and neutral B4C is plotted with the same colour scale. Therefore a quantitative comparison can only be carried out for equivalent protonation states. For neutral I3C the iodine atoms show positive maxima (Fig. 12.6a) at the centre of the iodine atoms, opposite of the iodine-carbon bond (σ-hole). For I3C mono- and dianions the potential is negative at the iodine atoms, but shows a central region of less negative potential (Fig. 12.6b and c). For neutral B3C the bromine atoms show a positive potential (Fig. 12.6d) although the effect is less pronounced compared with neutral I3C since the polarisability of iodine is larger than for bromine. The B3C mono- and dianions also show a negative potential at the bromine atoms and the characteristical features of the σ-hole with a surrounding ring of more negative potential (Fig. 12.6e and f) as for the I3C anions.

[3]The occurrence of non-covalent contacts in crystal structures with distances smaller than the sum of the van der Waals radii was referred to as "polar flattening" (Nyburg, 1979) in earlier literature and led to a revision of the van der Waals radii for atoms bonded to carbon (Nyburg and Faerman, 1985).

[4]The calculations were kindly performed by Kathrin Meindl, Göttingen.

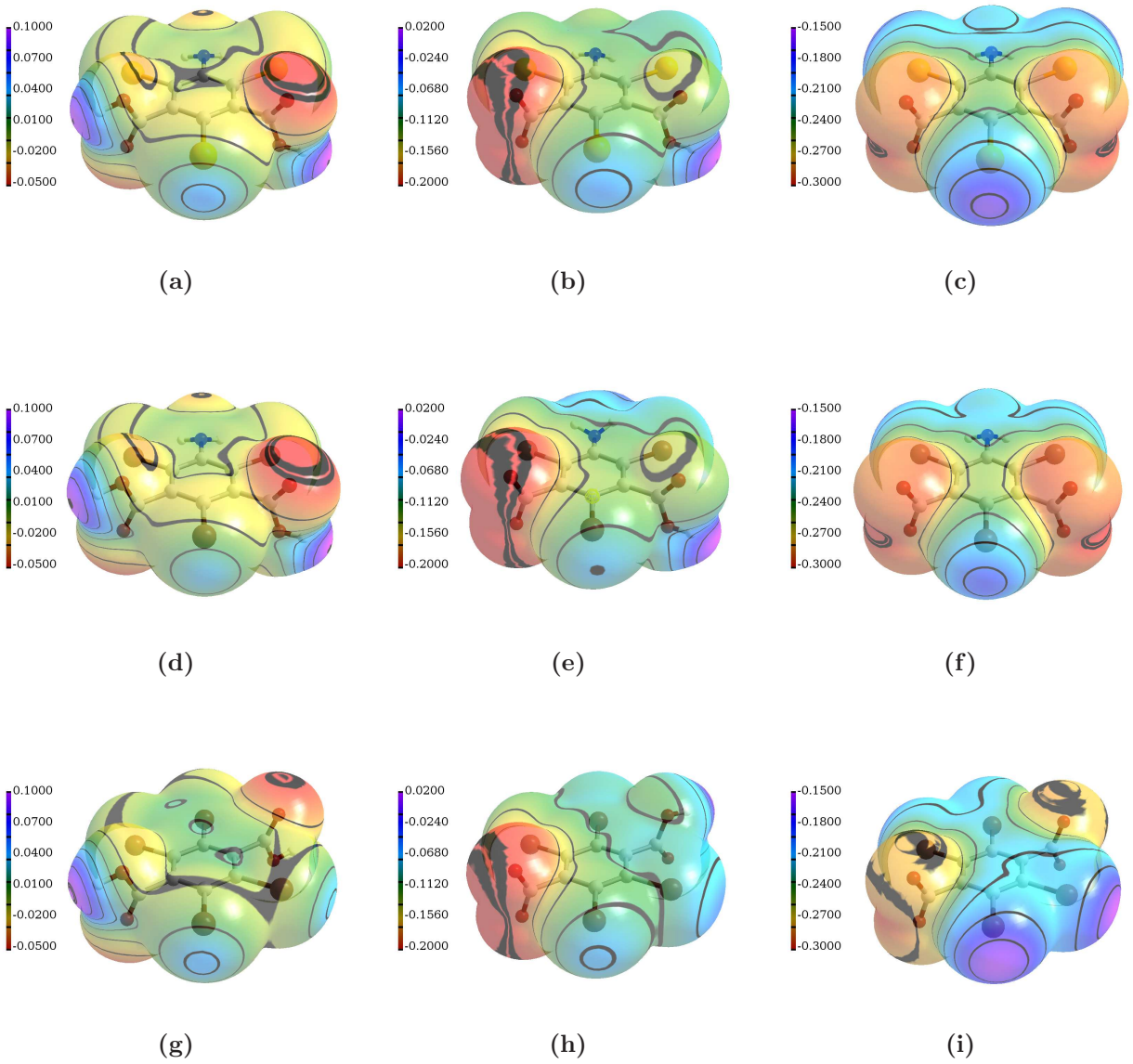


Figure 12.6: Computed electrostatic potential, MP2/DGDZVP, for I3C, B3C and B4C in a.u. mapped on the surface of molecular electron density at 0.001 a.u.. The same colour scale was used for each protonation state, e.g. the electrostatic potential of neutral I3C, neutral B3C and neutral B4C is plotted with the same colour scale. Therefore a quantitative comparison can only be carried out for equivalent protonation states. The halogen atoms show a positive or more positive potential in the central region. **(a)** I3C with one iodine atom and the carboxyl groups pointing towards the reader. **(b)** I3C monoanion, same orientation. **(c)** I3C dianion, same orientation. **(d)** B3C with one bromine atom and the carboxyl groups pointing towards the reader. **(e)** B3C monoanion, same orientation. **(f)** B3C dianion, same orientation. **(g)** B4C with two bromine atoms and one carboxyl group pointing towards the reader. **(h)** B4C monoanion, same orientation. **(i)** B4C dianion, same orientation. Figures were prepared with *MOLISO* (Hübschle and Luger, 2006) within *MOLECOOLQT* (Hübschle, 2010).

In B4C the bromine atoms show a positive potential (Fig. 12.6g) where in the B4C monoanion a small negative potential at the bromine atoms with a ring of larger negative potential around it (Fig. 12.6h) can be observed. For the B4C dianion the potential is negative at the bromine atom centres but is the least negative compared with the whole molecule (Fig. 12.6i).

Obviously the electrostatic potential is very sensitive to the protonation state. Iodine as the more polarisable element shows a larger positive (less negative) potential at the σ-hole compared with bromine; therefore the halogen bonding for iodine compounds is expected to be stronger. The carboxyl groups of the phasing tools are most probably present in the deprotonated form (i.e. as carboxylate groups) at the pH range usually used for co-crystallisation or soaking experiments. However, the σ-hole is still found in the anions of the phasing tools where the halogen atoms have only little negative potential at the centre of each atom. The polarising effect of the oxygen acceptor atom can induce a positive potential on the halogen atom as shown in previous studies (Riley and Hobza, 2008). Additionally cooperative effects due to more than one halogen (or hydrogen) bond may locally enhance the σ-hole and therefore the halogen bonding.

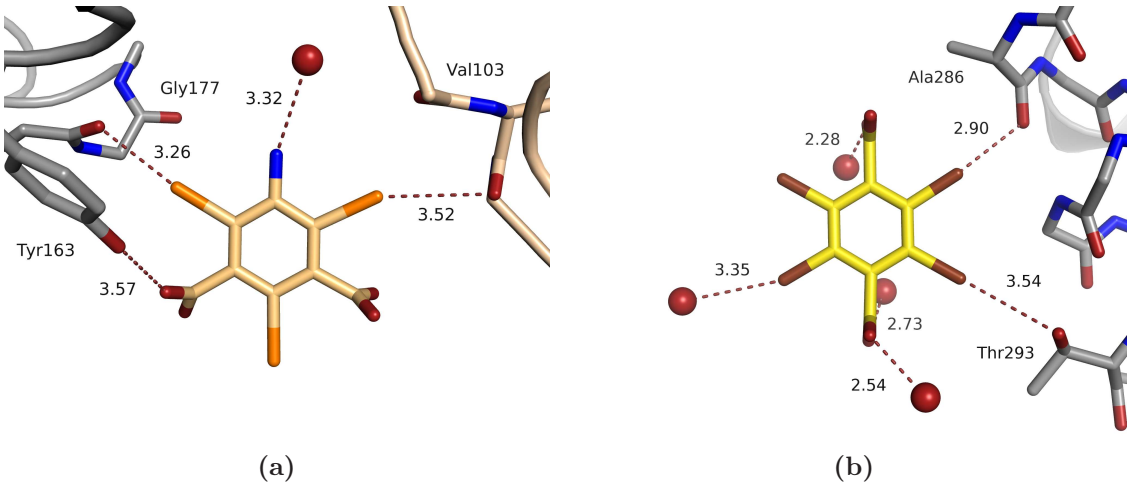


Figure 12.7: Halogen bonding, distances given in Å. **(a)** One molecule of I3C in elastase. The iodine atoms form halogen bonds with the carbonyl oxygen atoms of Gly177 and Val103 (symmetry equivalent, depicted in sand colour). The halogen bond to the carbonyl oxygen in Val103 is clearly perpendicular to the plane of the peptide bond, i.e. the iodine atom interacts with the π system of the carbonyl group, not the oxygen lone pair electrons. **(b)** One molecule of B4C in thermolysin. The bromine atoms form halogen bonds with the carbonyl oxygen of Ala286 and (a weaker bond) with the hydroxyl group of Thr293. There is also a halogen bond between a bromine atom and a water molecule.

Examples for halogen bonding are shown in Fig. 12.7. Halogen bonding is also observed for I3C in Mh-p37 (Fig. A.2, p. 125) where the I3C molecule is also found in the neutral form due to the low pH value of the crystallisation conditions. According the Fig. 12.6a the iodine atoms have a clearly positive electrostatic potential for the neutral state of I3C. Other examples can be found in Fig. 10.8 (p. 93), Fig. A.3 (p. 125) and Fig. A.5 (p. 127).

The I3C molecule shows several iodine-water contacts at the main site of proteinase K with the more polarisable iodine atoms whereas fewer (and probably weaker) bromine-water contacts are present in the proteinase K B3C structure. This could be one reason

why the occupancy for B3C is lower than for I3C at the main site in proteinase K.

In the small-molecule crystal structures there are a few short halogen-oxygen contacts and hydrogen bonding dominates.

Comparison of interaction

The interactions for the phasing tools I3C, B3C and B4C were compared. The statistics are based on the test protein structures discussed in this work and the Mh-p37 I3C structure (Sippel *et al.*, 2008; Appendix A). B3M and B3O were not included in the statistics since only one derivative data set is available for either phasing tool. Tab. 12.1 and 12.2 show that the main interaction takes place between the phasing tools and water molecules. Considering all phasing tools the carboxylate groups are responsible for more than two thirds of the interactions.

Table 12.1: Summary of interaction types for I3C. Hydrogen bonds were included in the list if the distance d is between 2.30 and 3.60 Å and halogen bonds if the distance between the halogen atom and the acceptor is smaller than the sum of their van der Waals radii.

	moiety	interaction type[a]	examples[a]	# of contacts	perc.[b]
I3C	$-COO^-$	O–H···**OOC**	HO–H···**OOC**	38	47%
			Ser–O–H···**OOC**	4	
			Thr–O–H···**OOC**	1	
			Tyr–O–H···**OOC**	1	
		N–H···**OOC**	N_{mc}–H···**OOC**	6	24%
			Arg–NH–H···**OOC**	5	
			Lys–NH_2–H···**OOC**	5	
			Asn–NH–H ···**OOC**	4	
			Arg–N_ϵ–H ···**OOC**	1	
			Gln–NH–H ···**OOC**	1	
	$-NH_2$	**N–H**···A	**N–H**···O=C_{mc}	6	13%
			N–H···OH_2	5	
			N–H···O–Gln	1	
	–I	**C–I**···O	**C–I**···O=C_{mc}	7	16%
			C–I···OH_2	6	
			C–I···OH–Tyr	1	
			C–I···O–Glu	1	

[a] Charges are omitted for clarity reasons. At neutral pH carboxyl groups are almost fully deprotonated whereas the amino group is not protonated. mc = main chain.

[b] Percentage of this interaction type compared with all interactions of the phasing tool. There is a total number of 93 interactions for I3C.

Table 12.2: Summary of interaction types for B3C and B4C. Hydrogen bonds were included in the list if the distance d is between 2.30 and 3.60 Å and halogen bonds if the distance between the halogen atom and the acceptor is smaller than the sum of their van der Waals radii.

	moiety	interaction type[a]	examples[a]	# of contacts	perc.[b]
B3C	–COO$^-$	O–H···**OOC**	HO–H···**OOC**	28	51%
			Ser–O–H···**OOC**	2	
			Thr–O–H···**OOC**	1	
		N–H···**OOC**	N_{mc}–H···**OOC**	5	16%
			Arg–NH–H···**OOC**	2	
			Lys–NH_2–H···**OOC**	2	
			Arg–N_ϵ–H ···**OOC**	1	
			N–H···OH_2	7	23%
	–NH_2	**N–H**···A	**N–H**···O=C_{mc}	6	
			N–H···O–Gln	1	
			C–Br···OH_2	3	10%
	–Br	**C–Br**···O	**C–Br**···O=C_{mc}	2	
			C–Br···OH–Ser	1	
B4C	–COO$^-$	O–H···**OOC**	HO–H···**OOC**	28	67%
			Tyr–O–H···**OOC**	3	
			Ser–O–H···**OOC**	1	
			Arg–NH–H···**OOC**	5	12%
			Lys–NH_2–H···**OOC**	1	
	–Br	**C–Br**···O	**C–Br**···O=C_{mc}	5	21%
			C–Br···OH_2	4	
			C–Br···OH–Thr	1	

[a] Charges are omitted for clarity reasons. At neutral pH carboxyl groups are almost fully deprotonated whereas the amino group is not protonated. mc = main chain.
[b] Percentage of this interaction type compared with all interactions of the phasing tool. There is a total number of 61 interactions for B3C and 48 interactions for B4C.

Halogen bonds play an important role especially for I3C and B4C. For I3C the halogen bonds involve iodine atoms and are therefore more stabilising than halogen bonds with bromine atoms. In B4C there are four instead of three halogen atoms; therefore more halogen-acceptor contacts are possible. Halogen bonding with the water molecules in the crystal structure may also be interpreted as hydrogen bonding between water and halogen since the position of the hydrogen atoms cannot be determined from the data; however there is a recent report on halogen-water bonding that found interactions between a halogen atom, a water molecule and a hydrogen bond acceptor, termed halogen-water-hydrogen bridge (Zhou *et al.*, 2010).

On average the phasing tools form more than five contacts per molecule to protein residues or water molecules. The number of contacts per phasing tool can be correlated with the occupancy of that molecule; however the occupancy also depends on the strength of the non-covalent bonds.

In addition to enthalpic effects from hydrogen and halogen bonding the hydrophobic effect, which is also important for protein folding and crystallisation, makes an additional contribution to ligand binding. Water molecules around the ligand are removed and replaced with hydrogen or halogen bonds to the protein residues; this leads to an increase in entropy. Hydrophobic interactions have so far only been observed in one structure (Fig. A.4, p. 126). The bulky halogen atoms and the carboxylate groups arranged perpendicular to the benzene ring probably hinder π-π interactions with aromatic side chains (see small-molecule structures).

There is no specific motif or residue required for the binding of the phasing tools. Halogen bonds are formed either to water molecules or to the carbonyl oxygen atom of the main chain. Apart from hydrogen bonds to water molecules interactions of the carboxylate group with arginine residues are found quite often. Hydrogen bonds with the main chain amide proton are frequently observed. The amino groups often forms hydrogen bonds to the carbonyl oxygen of the main chain. There are some examples where the hydroxyl group of a serine or threonine residue and the amide proton of that same residue form hydrogen bonds to a carboxylate group of a phasing tool (e.g. Fig. 8.8 or Fig. 12.3).

The phasing tools mainly replace water molecules in the crystal lattice when the protein crystals are derivatised. Therefore changes induced in the final refined structures are only minor (reorientation of side chains, replacement of water molecules).[5]

12.2 Triangles as additives in crystallisation screens

Co-crystallisation with the phasing tools usually results in high occupancies (e.g. lysozyme I3C and trypsin B3C) and may be a valuable option if weak binding or crystal degradation do not permit derivatisation *via* soaking. Whether incorporated *via* soaking or co-crystallisation the phasing tools show a number of crystal contacts in the protein structures: they act as bridges between different protein molecules. Examples include the B3C binding site in trypsin (Fig. 8.11, p. 73), some I3C binding sites in lysozyme, thaumatin, elastase and proteinase K (e.g. Fig. 7.7b, p. 59), one I3C binding site in Mh-p37 (Fig. A.2b, p. 125), the main site of B3C in proteinase K (Fig. 8.8b, p. 70) and the main B4C binding site in thermolysin (Fig. 10.10, p. 94).

It is possible that the hydrogen donor and acceptor groups of the phasing tools as well as halogen bonding could promote crystal growth. It has been shown that similar small molecules without halogen atoms promote crystal growth (McPherson and Cudney, 2006) and a crystallisation screen was introduced that includes such molecules (Silver Bullets Screen; Hampton Research, 2008). Therefore an introduction of the phasing tools into standard crystallisation screens is desirable: it could promote crystal growth and introduce heavy atoms for phasing at the same time. Initial tests were carried out with I3C, B3C,

[5]For the structures discussed in this work the r.m.s.d. between native and derivatised structure is in general < 0.3 Å for C_α atoms and usually < 0.7 Å for all atoms.

B3M and B3O at the Max-lab crystallisation facility in Lund, Sweden, so far without clear indications of a positive effect. Additional tests are required to further investigate the effects of such small molecules on crystallisation.

12.3 Conclusion

The phasing tools show several types of hydrogen and halogen bonding. No specific residue or sequence of amino acids are required for binding. The compounds show disorder only in a few cases, which is often observed for highly symmetrical heavy-atom cluster compounds, e.g. the tantalum bromide cluster. Halogen and hydrogen bonding can be considered to be orthogonal, i.e. halogen bonding does not disrupt an existing hydrogen bond network. I3C and B4C seem to be the best candidates for halogen bonding since the iodine atoms in I3C show the largest positive electrostatic potential compared with the bromine compounds; B4C has four bromine atoms that can form halogen bonds to protein residues.

Small-molecule compounds similar to the phasing tools were shown to promote crystal growth. Further tests have to reveal if the inclusion of compounds containing functional groups for linking protein molecules and heavy atoms suitable for phasing is beneficial.

Chapter 13

Summary and perspective

A new class of compounds was developed that is suitable for the derivatisation of biological macromolecules. Sticky triangle and sticky rectangle compounds were used to introduce iodine or bromine atoms. The iodine atoms exhibit a strong anomalous signal at longer wavelengths (data collection in-house and at the synchrotron). The bromine atoms are suited for multi-wavelength anomalous dispersion (MAD) experiments at the synchrotron, but were also employed for single-wavelength anomalous dispersion (SAD) experiments. Compared with traditional heavy-atom compounds the phasing tools presented here have low toxicity.

The sticky compounds interact *via* hydrogen and halogen bonding with proteins. There is no specific residue or sequence required for binding. The compounds can be introduced either by soaking or by co-crystallisation. In particular the magic triangle I3C and the MAD tetragon B4C have several advantages for heavy-atom derivatisation. The strong anomalous signal of the three iodine atoms in I3C can be exploited in SAD and SIRAS experiments. For strongly diffracting crystals it is recommended to derivatise with I3C and carry out data collection in-house. The four bromine atoms in B4C are suitable for MAD (or SAD) experiments. If due to the crystal quality synchrotron radiation is required for data collection, multi-wavelength experiments with B4C are recommended for weakly diffracting crystals. First indications are that the binding of B4C is orthogonal to I3C. These two compounds are inexpensive and readily commercially available. Therefore they are easily accessible for many structural biology laboratories, which would not have the facilities to carry out chemical synthesis.

The triangular arrangement of the heavy atoms in the sticky triangle compounds provides an independent criterion for substructure validation, especially in borderline cases where it is difficult to discriminate between a correct and a wrong substructure solution. As a proof of principle several test proteins were derivatised with the sticky compounds and experimental phasing was carried out. Up to date four novel structures have been solved with the magic triangle I3C. There are several phasing kits available (see Appendix F).

Further investigations will be carried out to study the halogen bonding of the sticky compounds. Theoretical studies will be conducted with emphasis on the interaction of the acceptor with the halogen atom. It was found in previous studies that the electronegative acceptor can influence the electrostatic potential at the halogen atom; cooperative effects could enhance the strength of the halogen bonding.

The bromine compounds discussed in the present work can serve as candidates for further investigations of the anisotropy of anomalous scattering. The molecular arrangement of the anomalous scatterers provides additional information for the refinement of the parameters that describe the anisotropy of the anomalous scattering. Investigations are planned in cooperation with Dr. Marc Schiltz and Dr. Gérard Bricogne at the ESRF, Grenoble, France. These investigations might also elucidate the origin of the double peak frequently observed in the fluorescence spectra of protein crystals that were derivatised with the bromine compounds.

Since radiation damage can compromise experimental phasing with halogenated compounds it will be important to further study the effects. The usage of radical scavengers, although already discussed and tested in previous studies (e.g. Murray and Garman, 2002), should be reconsidered for the bromine compounds of this study. On the other hand radiation damage could also be exploited to generate phase information. Different data collection strategies at the synchrotron should be carried out to shed light on these issues. Additionally the iodine compound could be exploited for a pseudo-MAD experiment with data collection carried out at several long wavelengths at the synchrotron.

The phasing tools show a number of crystal contacts: they form a link between different protein molecules. A crystallisation screen could be designed that includes the phasing tools *a priori*; the functional groups could be exploited as buffer components. The compounds could serve as linkers between protein molecules and possibly promote crystal growth while introducing heavy atoms for phasing at the same time. New linker compounds that span larger distances (e.g. polycyclic aromatic hydrocarbons) could also be tested. However halogen-carbon bonds are sensitive to light and care must be taken in handling these solutions.

The concept of combining heavy atoms for phasing with functional groups for interaction could be extended to multi-metal compounds such as clusters. For larger structures more anomalous scatterers are required. In particular metal cluster based on lanthanoid elements could provide a strong anomalous signal. Halogenated compounds with hydrophobic functionalities could also be designed. These could be beneficial for the derivatisation of water-insoluble biomolecules, which sometimes incorporate detergent molecules in the crystal lattice.

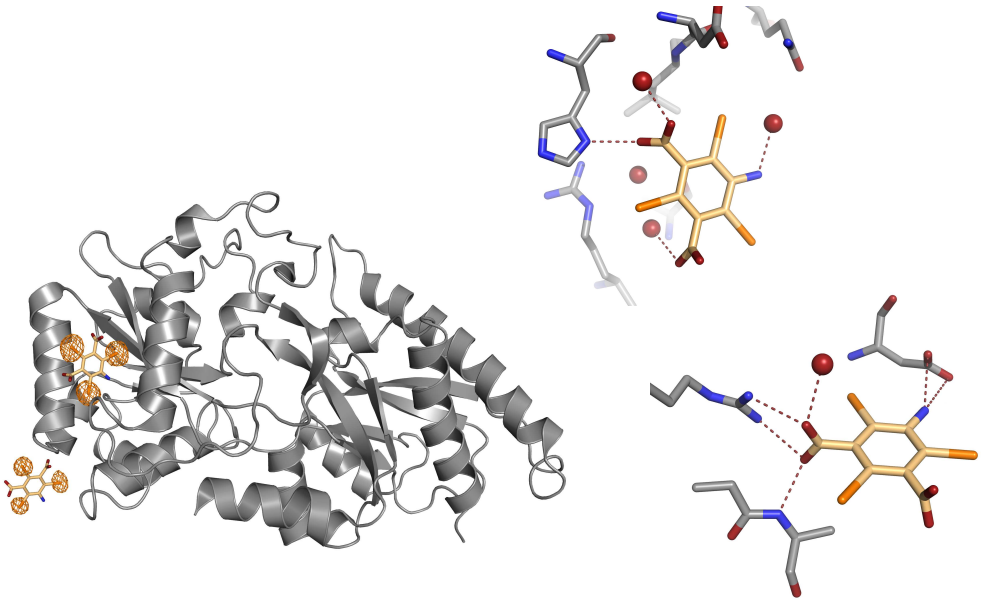
The sticky compounds of the current work are readily available and easy to use. Therefore it is expected that they will find broad application in experimental phasing.

Part IV

Appendix

Appendix A

Novel protein structures solved with the magic triangle I3C



Summary

At the time of writing four novel protein structures have been solved with the magic triangle I3C. All data presented in this chapter were kindly provided by the I3C users. I3C shows several types of interaction in these proteins, which are briefly discussed.

Mh-p37

The first novel structure solved with I3C was the protein Mh-p37 (38 kDa) from *M. hyorhinis*, carried out by Katherine Sippel, Robert McKenna and co-workers from the University of Florida, Gainesville, USA (Sippel *et al.*, 2008). Protein crystals were grown by mixing 3 µL protein solution at a concentration of 10 mg mL^{-1} and 3 µL precipitant solution (100 mM citric acid at pH = 3.0, 40% PEG 4000 and 100 mM ammonium bromide). Crystals were derivatised with I3C *via* a quick soak: 0.5 µL of 0.5 M I3C solution were added to the crystallisation drop. Data were collected in-house (CuKα, 1.54178 Å) from a derivatised crystal and a native crystal (resolution 1.9 Å). Data were prepared for SIRAS phasing with *SHELXC* (Sheldrick, 2008a) and structure solution was carried out with SHELXD (Schneider and Sheldrick, 2002).

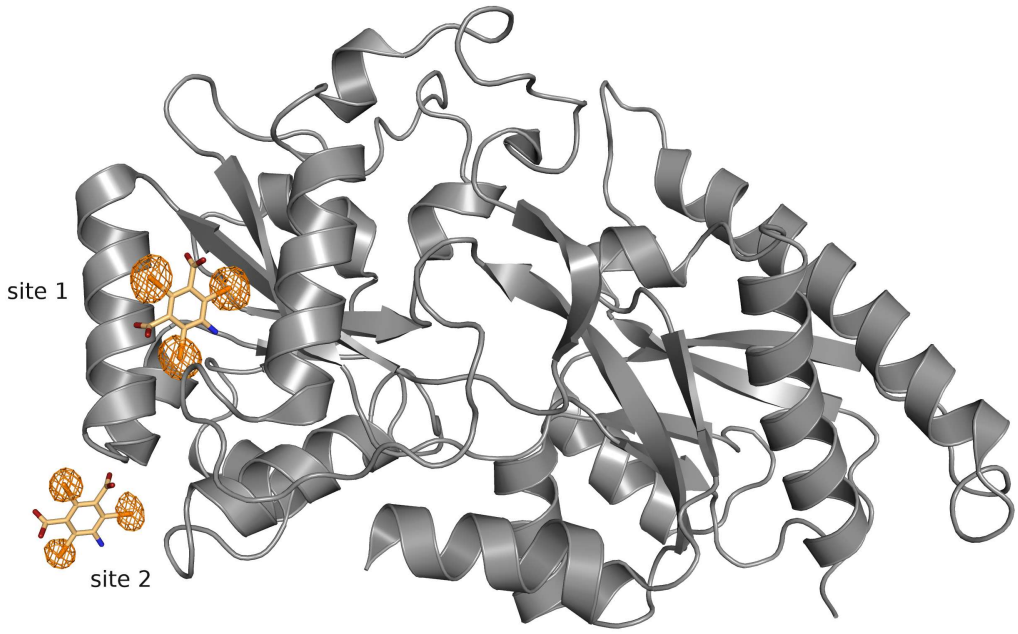
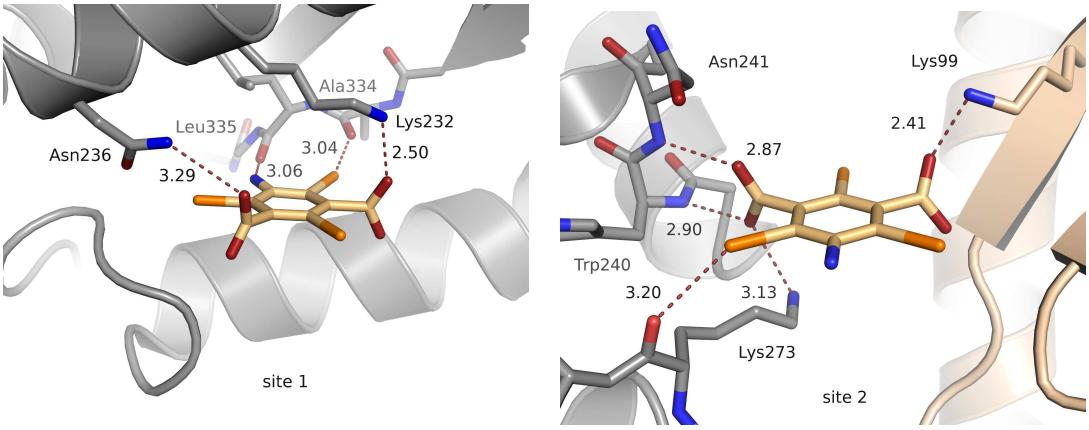


Figure A.1: The protein Mh-p37 with the two binding sites of I3C. Anomalous electron density is shown at 4σ (orange).

There are two binding sites for I3C (Fig. A.1) and their occupancies were set to 1.0 and 0.5, respectively. I3C at site 2 makes hydrogen bonds to two separate protein molecules (Fig. A.2b).

The derivatisation with I3C was successful, even though the crystals were grown at acidic conditions (pH = 3). Incorporation of other heavy atoms failed due to these conditions. Low sequence identity to previously solved structures prevented structure solution by molecular replacement. A selenomethionine mutant was not prepared since the sequence contained only two methionine residues (terminal residue not included) and "thus the structure remained unsolved for 6 y despite much effort" (Sippel *et al.*, 2008).

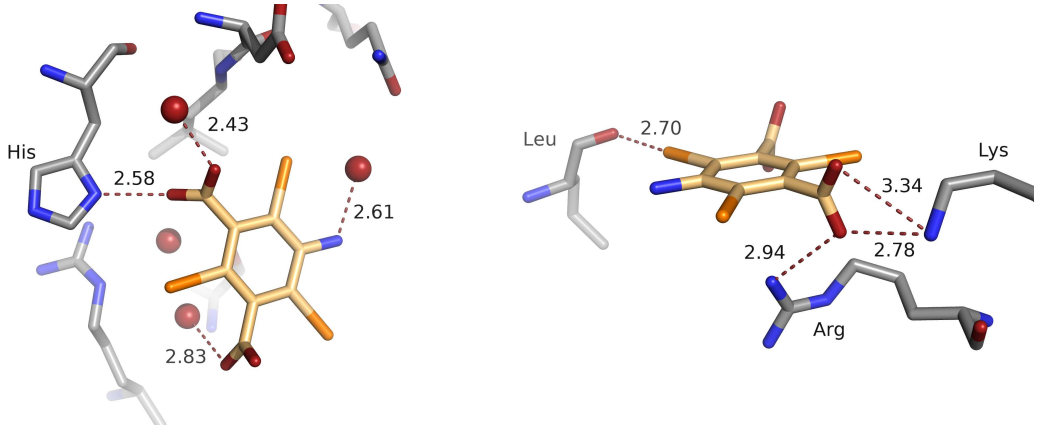


(a) (b)

Figure A.2: The two binding sites of I3C in Mh-p37. Hydrogen bonds are shown as dashed lines and distances are given in Å. **(a)** The main site with I3C hydrogen bonds to Asn236 and Lys232 (carboxylate groups) and the carbonyl oxygen atom of Leu335 (amino group). Halogen bonding is observed between iodine and the carbonyl oxygen of Ala334. **(b)** I3C at the interface of two protein molecules (symmetry equivalent in sand colour). Hydrogen bonds are formed to the amide nitrogen atom of Trp240 and Asn241. The carboxylate groups also interact with Lys99 (symmetry equivalent) and Lys273. Halogen bonding is observed between iodine and the carbonyl oxygen of Lys273.

CTP

Scott Lovell from the University of Kansas, Lawrence, USA, kindly provided data from an unpublished novel structure that was solved with I3C. Data were collected in-house (CuKα, 1.54178 Å) from an I3C-derivatised crystal (5 min soak in 50 mM I3C solution) and a native crystal (resolution 2.10 Å). SIRAS phasing was carried out to solve the structure.



(a) (b)

Figure A.3: Two of the three bindings sites of I3C in CTP. Hydrogen bonds are shown as dashed lines and distances are given in Å. **(a)** The main site with I3C contacts to His and water molecules. **(b)** I3C hydrogen bonds to Lys and Arg. A short iodine-oxygen contact is observed with Leu. Data courtesy of Scott Lovell.

In the final refined structure there are three binding sites for I3C (Fig. A.3). The

occupancies were set to 1.0, 0.6 and 0.5 for the refinement (based on peak height at the iodine positions in the anomalous electron density map).

PFH

Data and the final refined structure were kindly provided by Christoph Parthier, Ulrike Bräuer and Milton Stubbs from the Martin-Luther-Universität in Halle-Wittenberg, Germany. The structure solution is discussed in chapter 11. The structure is to be published. The protein crystallises as a dimer with twofold NCS-symmetry. There are five I3C binding sites (Fig. A.4) with two equivalent sites per protein chain.

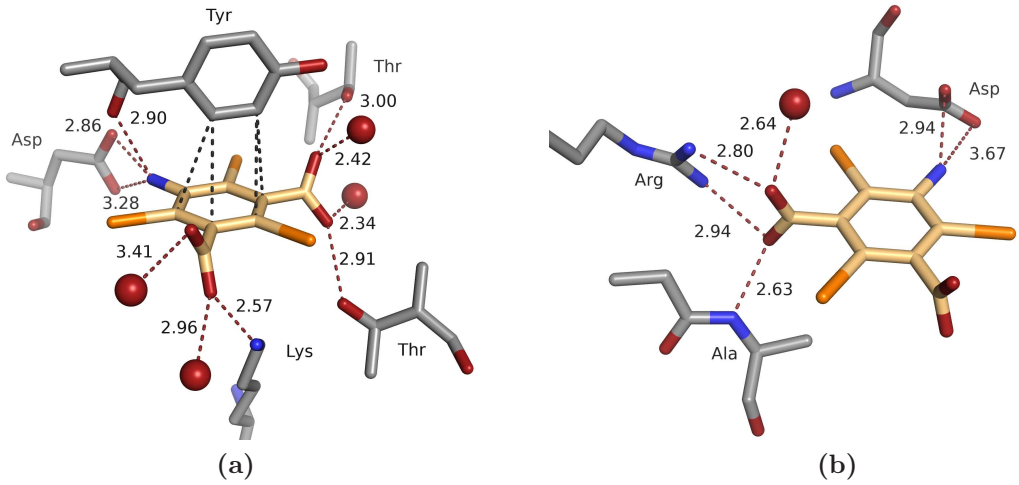


Figure A.4: Two of the five bindings sites of I3C in PFH. Hydrogen bonds are shown as red dashed lines and distances are given in Å. **(a)** I3C at site 1 interacts with the protein through hydrogen bonds: the amino group makes contacts to Asp and the carbonyl oxygen of Tyr; the two I3C carboxylate groups interact with lattice water molecules and Lys and Thr residues. There is a hydrophobic contact of the I3C benzene ring with Tyr side chain (edge-to-face, black dashed lines, contacts in the range of 3.57 to 3.61 Å.). Site 2 shows a similar coordination environment **(b)** I3C at site 3 shows hydrogen bonds to Arg, the amide nitrogen of Ala and a lattice water molecule (I3C carboxylate group) and to Asp (I3C amino group). Site 4 shows a similar coordination environment. Data courtesy of Christoph Parthier, Ulrike Bräuer and Milton Stubbs.

PDZP

Data and the final refined structure were kindly provided Jeanine Amacher and Dean Madden from Dartmouth Medical School, Hanover, USA. The structure solution is discussed in chapter 11. The structure is to be published. There is one binding site for I3C in PDZP, which is shown in Fig. A.5.

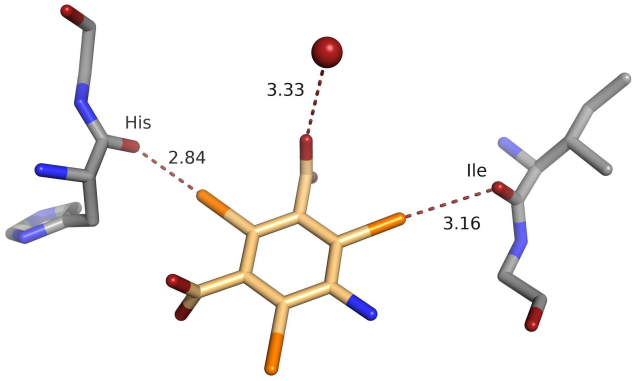


Figure A.5: The I3C binding site in PDZP. The carboxylate group interacts with a lattice water molecule and two iodine atoms are in contact with the carbonyl oxygen atoms of His and Ile. Data courtesy of Jeanine Amacher and Dean Madden.

Appendix B

Crystallographic parameters

Data statistics

Resolution: The crystallographic resolution of a data set is based on the maximum Bragg diffraction angle 2θ (Eq. 1.2, p. 7) for a given wavelength. The crystallographic resolution is related to the optical resolution, which is the minimum distance between two objects that allows their images in the electron density map to be distinguished from each other. In protein crystallography high resolution data are usually truncated when the mean signal-to-noise ratio $< I/\sigma(I) >$ for this resolution shell is smaller than 2.0.

Completeness: The completeness is defined as the fraction of independent reflections that have been measured and the theoretical possible reflections for a given space group and resolution.

Multiplicity: The multiplicity of observations (or redundancy) is defined as the average of how many times a reflection (including symmetry equivalents) has independently been measured during data collection.

Mean signal-to-noise ratio $< I/\sigma(I) >$: The mean signal-to-noise ratio indicates for a particular resolution shell the average of the ratio of intensities to their standard deviations. The mean signal-to-noise ratio in the outer resolution shells may be used to determine the resolution of a data set.

Merging R-factor R_{merge} or R_{int}: The merging residual factor describes the deviation of an individual intensity measurement $I_i(hkl)$ from the mean value $< I(hkl) >$ of all its symmetry-equivalent reflections.

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - < I(hkl) >|}{\sum_{hkl} \sum_i I_i(hkl)} \tag{B.1}$$

This includes all reflections for which more than one symmetry equivalent has been measured.

Redundancy-independent merging R-factor $R_{\text{r.i.m.}}$ or R_{meas}: The redundancy-independent merging residual factor is defined as:

$$R_{\text{r.i.m.}} = \frac{\sum_{hkl}\left(\frac{N}{N-1}\right)^{1/2}\sum_i |I_i(hkl) - < I(hkl) >|}{\sum_{hkl}\sum_i I_i(hkl)} = R_{\text{meas}} \tag{B.2}$$

$R_{\text{r.i.m.}}$ (R_{meas}) indicates the precision of an individual intensity measurement I_i independent of the multiplicity of that measurement (Weiss, 2001; Diederichs and Karplus, 1997). Since the R_{merge} factor is not independent of multiplicity – it will rise with increased multiplicity although the averaged intensity values should be better determined – it has been suggested to replace the R_{merge} factor with $R_{\text{r.i.m.}}$ or R_{meas} (Weiss, 2001; Diederichs and Karplus, 1997).

Precision-indicating merging R-factor $R_{\text{p.i.m.}}$: The precision-indicating merging residual factor describes the precision of the averaged intensity measurement $< I(hkl) >$ (Weiss, 2001) and is defined as:

$$R_{\text{p.i.m.}} = \frac{\sum_{hkl}\left(\frac{1}{N-1}\right)^{1/2}\sum_i |I_i(hkl) - < I(hkl) >|}{\sum_{hkl}\sum_i I_i(hkl)} \tag{B.3}$$

$R_{\text{p.i.m.}}$ decreases with increasing multiplicity and seems to be a useful measure for the estimation of data quality for anomalous data sets.

Anomalous R-factor R_{anom}: The anomalous residual factor quantifies the relative differences of Friedel-related reflections $I(hkl)$ and $I(\overline{hkl})$:

$$R_{\text{anom}} = \frac{\sum_{hkl} 2\left|I(hkl) - I(\overline{hkl})\right|}{\sum_{hkl}[I(hkl) + I(\overline{hkl})]} \tag{B.4}$$

A strong anomalous signal is usually identified by a large R_{anom}; however the ratio R_{anom} to $R_{\text{p.i.m.}}$ has been proposed as a better indicator if the substructure can be solved (Weiss, 2001; Mueller-Dieckmann *et al.*, 2005), although in practice values greater than the proposed ratio of 1.5 are necessary for substructure solution.

Anomalous correlation coefficient CC_{anom}: The anomalous correlation coefficient is the correlation coefficient between the signed anomalous differences from two data sets (e.g. from a multi-wavelength anomalous dispersion experiment). Values greater than 0.30 are an indication of an anomalous signal suitable for experimental phasing.

Mean anomalous signal-to-noise d''/sig: For single-wavelength experiments the anomalous signal is quantified by the ratio of the observed anomalous difference (d'') and the corresponding estimated standard deviation (sig or σ). There are different scales used in *XPREP* and *SHELXC*. For *SHELXC* the noise level (i.e. no anomalous signal) corresponds to $d''/\text{sig} \simeq 0.798$. Throughout this text, the statistics provided by *SHELXC* are used. Additionally, instead of the estimated standard deviations the variance of the differences can also be used. This does not depend on (often

error-prone) estimation of standard deviations. The anomalous signal may also be assessed by calculating the self-anomalous CC, the anomalous correlation coefficient between two randomly chosen subsets of reflections.

Experimental phasing statistics

SHELXD CC **and** CC_{weak}**:** The correlation coefficient (CC) between the calculated normalised structure factors E_c and the observed ones E_o is a measure for the quality of the substructure solution (Schneider and Sheldrick, 2002):

$$CC = \frac{100[\sum(wE_oE_c)\sum w - \sum(wE_o)\sum(wE_c)]}{\{[\sum(wE_o^2)\sum w - (\sum wE_o)^2][\sum(wE_c^2)\sum w - (\sum wE_c)^2]\}^{\frac{1}{2}}} \tag{B.5}$$

The weights $w = [1 + g\sigma^2(E)]^{-1}$ with $g = 0.1$ (default) are employed to weight down less reliable difference structure factors. The correlation coefficient calculated with all data (CC) and the correlation coefficient calculated with only the E values not used for substructure determination [CC_{weak}; similar to the R_{free} factor used in refinement (Brünger, 1992)] are good indicators whether the substructure was solved. For a MAD experiment CC values greater than 35% are expected. A SAD experiment usually gives lower values ($CC \geq 25\%$, $CC_{weak} \geq 15\%$); however there have also been cases when structures were solved that showed considerably lower values.

SHELXE **pseudo-free** CC**:** During density modification in *SHELXE* (Sheldrick, 2002) ~10% of the reflections are left out for one density modification step to determine the pseudo-free CC. It is calculated as the correlation coefficient between the calculated normalised structure factors and the observed ones for these reflections. Values > 70% usually indicate an interpretable map. The pseudo-free CC (along with *contrast* and *connectivity*) may also be used to distinguish between the correct and wrong substructure enantiomorph (Sheldrick, 2002).

Experimental electron density map analysis

weighted mean phase error wMPE: When comparing an experimental electron density map with the map obtained from the final refined model, the weighted mean phase error is used to estimate the quality of the initial experimental map. Ideally the phases from the experimental map should be close to the phases obtained from the final model (assumed as the true phases). wMPE can be calculated with *SHELXPRO* (Sheldrick, 2008a): the *.fcf* file from the *SHELXL-97* refinement is compared with the *.phs* file from *SHELXE*.

map correlation coefficient, map CC**:** The map CC is another measure for the quality of an initial experimental electron density map. It is also calculated in *SHELXPRO*, based on the wMPE (Lunin and Woolfson, 1993). The *.fcf* file from the final refinement is compared with the experimental map as *.phs* file.

Refinement statistics

During refinement the initial model (obtained from automated tracing programs such as *ARP/wARP* or *SHELXE*) is improved to give a better fit to the measured data. In least-squares refinement the parameters of the model (atom positions, thermal displacement parameters) are varied to minimise M:

$$M = \sum_{hkl} w \left(F_o^2(hkl) - kF_c^2(hkl)\right)^2 \tag{B.6}$$

with k as the relative scaling factor, $F_o^2(hkl)$ proportional to the measured intensities and $F_c(hkl)$ as the structure factor amplitude calculated from the model for the reflection hkl. The weight w is based on the standard deviation of that reflection. Structure refinement is an iterative procedure and user intervention is necessary to manually improve the model (e.g. include new atoms or delete those wrongly assigned). The least-squares method (as implemented e.g. in *SHELXL-97*; Sheldrick, 2008) is suitable for small-molecule structures, but also for macromolecular structures where data to at least medium resolution ($\leq 1.6\,\text{Å}$) are available.

At lower resolution the maximum likelihood method is applied for refinement. The quality of a model is determined by its consistency with the observed data. The likelihood function L is the joint probability of making a set of observations assuming a particular model and has to be maximised (Pannu and Read, 1996):

$$L = \prod_{hkl} P\left(|F_o|; |F_c|\right) \tag{B.7}$$

Especially for macromolecular structure refinement the ratio of data to parameters is low. Restraints are used to introduce prior knowledge. Restraints may be derived from previously determined small-molecule structures, e.g. the *Engh* and *Huber* parameters (Engh and Huber, 1991) for proteins or the restraints for the phasing tools as discussed in this work. Restraints have a standard deviation and can therefore be treated as additional data. On the other hand constraints reduce the number of parameters, e.g. in the refinement of hydrogen atoms as riding atoms.

Crystallographic R-factor R_{cryst} (or R_{work}): The quality of the model is assessed by comparison of the observed structure factor amplitudes and the structure factor amplitudes calculated from the model.

$$R_{\text{cryst}} = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|} \tag{B.8}$$

Free R-factor $\mathbf{R}_{\text{free}}$: The free residual factor was introduced to prevent over-fitting (Brünger, 1992). It is defined as R_{cryst}, but calculated from a set of reflections (500-1000 reflections are suggested, usually 5% of all reflections) that was set aside at random before refinement was commenced. Consequently the R_{cryst} is calculated with the remaining 95% of the reflections.

Crystallographic R-factors $R1$ and $wR2$: For small-molecule structures the crystallographic residual factor is termed $R1$:

$$R1 = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|} \tag{B.9}$$

The weighted residual factor $wR2$ based on F^2 is closely related to the least-square refinement:

$$wR2 = \left[\frac{\sum_{hkl} w \left\{ F_o^2(hkl) - kF_c^2(hkl) \right\}^2}{\sum_{hkl} wF_o^2(hkl)} \right]^{1/2} \tag{B.10}$$

Although refinement is carried out against F^2 and therefore $wR2$ should be used as a quality indicator the $R1$ value is still quoted.

Thermal displacement parameters, B and U values: During refinement one thermal displacement parameter per atom is refined (for anisotropic refinement: six parameters). The Debye-Waller factor B (frequently used in protein crystallography) as introduced in Eq. 1.3 (p. 7) is related to the thermal displacement parameter U (used in small-molecule crystallography) by:

$$B = 8\pi^2 U \tag{B.11}$$

R.m.s. deviations: The root-mean square deviation of a geometric parameter (e.g. bond length, bond angle or chiral volume) is a measure of how much the parameter deviates from its ideal value. Generally higher deviations are tolerated for high-resolution data sets.

Ramachandran plot: The plot of the protein backbone torsion angles ϕ and ψ shows preferred regions where specific ϕ-ψ combinations are often observed. (Ramachandran *et al.*, 1963). For model validation the torsion-angle combinations observed in the protein are compared with data derived from other structures. The regions in the plot are classified as 'favoured', 'allowed' and 'disallowed'. Usually 98% of all residues are expected to lie within the favoured core regions and only 0.2% outside the allowed regions (Lovell *et al.*, 2003).

Appendix C

Crystallographic data for proteins

The following tables contain the refinement details for the protein derivative data sets. Refinement was carried out with *SHELXL-97* (Sheldrick, 2008a), except when otherwise indicated as *REFMAC* (Murshudov *et al.*, 1997). For details on data collection and structure solution please refer to the appropriate chapters in Part III: Results and discussion. For crystallographic parameters, see Appendix B: Crystallographic parameters.

I3C in lysozyme, thaumatin and elastase

Table C.1: Refinement details for lysozyme, thaumatin and elastase.

	Lysozyme I3C	Thaumatin I3C	Elastase I3C
PDB Code	3e3d	3e3s	3e3t
Resolution (Å)	19.4−1.55	19.7−1.73	37.2−1.60
No. of reflections	17445	27271	28897
$R_{\mathrm{cryst}}/R_{\mathrm{free}}$ (%)	17.1/22.0	15.8/21.0	15.7/21.3
No. of protein atoms	993	1531	1860
No. of ligand/ion atoms	79	94	75
No. of waters	147	262	300
B factors (Å^2)			
Protein	9.58	12.80	14.22
Ligands	11.00	21.90	44.68
Waters	21.23	25.50	30.10
R.m.s. deviations			
Bond length (Å)	0.007	0.008	0.008
Angle distance (Å)	0.023	0.024	0.024
Ramachandran plot			
Favoured (%)	99.2	98.5	97.9
Disallowed (%)	0	0.5	0

I3C and B3C in proteinase K

Table C.2: Refinement details for proteinase K.

	Proteinase K B3C	Proteinase K I3C
PDB code	3gt3	3gt4
Data set used for refinement	peak2	
Resolution (Å)	48.0−1.50	47.9−1.76
No. of reflections	38826	22770
$R_{\text{cryst}}/R_{\text{free}}$ (%)	14.3/18.4	14.0/19.1
No. of protein atoms	2045	2001
No. of ligand/ions atoms	74	53
No. of water atoms	417	343
B factors (Å^2)		
Protein	11.04	7.91
Ligands	15.77	15.80
Waters	28.32	22.70
R.m.s. deviations		
Bond length (Å)	0.009	0.007
Angle distance (Å)	0.024	0.023
Ramachandran plot		
Favoured (%)	97.8	97.8
Disallowed (%)	0	0

B3C in trypsin

Table C.3: Refinement details for bovine trypsin.

	Trypsin B3C
PDB code	3iti
Resolution (Å)	41.8−1.55
No. of reflections	30344
$R_{\text{cryst}}/R_{\text{free}}$ (%)	16.7/21.3
No. of protein atoms	1615
No. of ligand/ions atoms	27
No. of water atoms	243
B factors (Å^2)	
Protein	13.51
Ligands	14.30
Waters	28.30
R.m.s. deviations	
Bond length (Å)	0.009
Angle distance (Å)	0.024
Ramachandran plot	
Favoured (%)	98.64
Disallowed (%)	0

B3M, B3O and B3C in thaumatin

Table C.4: Refinement details for thaumatin data sets.

	Thaumatin B3M	Thaumatin B3O	Thaumatin B3C
Data set used for refinement	high-energy remote	high-energy remote	peak
Resolution (Å)	24.6−1.60	24.5−1.60	24.5−1.60
No. of reflections	34601	34552	34193
$R_{\mathrm{cryst}}/R_{\mathrm{free}}$ (%)	16.8/20.7	17.02/20.7	17.6/22.2
No. of protein atoms	1516	1503	1493
No. of ligand/ion atoms	145	90	122
No. of waters	265	285	267
B factors (Å^2)			
Protein	15.7	14.8	16.5
Ligands	19.4	19.0	19.4
Waters	29.9	30.6	31.9
R.m.s. deviations			
Bond length (Å)	0.008	0.009	0.008
Angle distance (Å)	0.024	0.025	0.024
Ramachandran plot			
Favoured (%)	99.02	99.02	98.04
Disallowed (%)	0	0	0

B4C in thaumatin and thermolysin

Table C.5: Refinement details for thaumatin and thermolysin.

	Thaumatin B4C	Thermolysin B4C (*REFMAC*)
Data set used for refinement	peak	high-energy remote
Resolution (Å)	24.5−1.60	24.8−1.80
No. of reflections	34179	28511
$R_{\mathrm{cryst}}/R_{\mathrm{free}}$ (%)	16.3/19.5	14.4/17.9
No. of protein atoms	1509	2430
No. of ligand/ions atoms	106	95
No. of water atoms	294	445
B factors (Å^2)		
Protein	14.67	14.09
Ligands	19.18	20.81
Waters	29.17	25.35
R.m.s. deviations		
Bond length (Å)	0.008	0.017
Angle distance/Angle	0.024 Å	1.53°
Ramachandran plot		
Favoured (%)	98.5	96.5
Disallowed (%)	0	0

I3C in ribonuclease A

Table C.6: Refinement details for ribonuclease A.

	Ribonuclease A (*REFMAC*)
Resolution (Å)	28.5−1.80
No. of reflections	13418
$R_{\mathrm{cryst}}/R_{\mathrm{free}}$ (%)	19.2/22.6
No. of protein atoms	954
No. of ligand/ions atoms	25
No. of water atoms	141
B factors (Å^2)	
Protein	16.86
Ligands	31.22
Waters	20.27
R.m.s. deviations	
Bond length (Å)	0.017
Angle (°)	1.53
Ramachandran plot	
Favoured (%)	95.08
Disallowed (%)	0

Appendix D

Crystallographic data for small molecules

General

Crystals were transferred to paratone-N oil (Hampton Research) and flash-cooled in liquid nitrogen. Data were recorded at 100 K on a Bruker three-circle diffractometer equipped with a SMART 6000 CCD detector, a MacScience rotating anode and Incoatec Helios optics (CuKα, $\lambda = 1.54178$ Å).

Data collected were integrated with *SAINT* (Bruker, 2009). Absorption correction (semi-empirical from equivalents) and scaling were carried out with *SADABS* (Sheldrick, 2009), except for the B3C small-molecule structure where *TWINABS* was used (Sheldrick, 2008b).

The structures were solved using direct methods within *SHELXS* (Sheldrick, 2008a) or *SHELXD* (Schneider and Sheldrick, 2002; Sheldrick, 2008a).

Structures were refined against all data by full-matrix least-squares methods on F^2 with *SHELXL-97* (Sheldrick, 2008a). All non-hydrogen atoms were refined with anisotropic displacement parameters. The hydrogen atoms attached to nitrogen or oxygen atoms were refined isotropically with distance restraints (*SADI* instructions in *SHELXL-97* for I3C, *SADI* for B3C, *SADI* and *DFIX* for B3M, *DFIX* for B3O, *SADI* for B4C) and their U_{iso} values constrained to $1.5 \times U_{eq}$ of their pivot atom. The hydrogen atoms of the methoxy group in B3M were refined isotropically on calculated positions using a riding model with their U_{iso} values constrained to $1.5 \times U_{eq}$ of their pivot atom. The benzene ring and all substituents were restrained to planarity (*FLAT* and *CHIV*). The carboxyl groups were restrained to planarity (*CHIV*). For all structures the extinction coefficient was refined, but only kept for further refinement if the absolute value was not within three times of its estimated standard deviation. Structure validation was carried out with the CIFcheck server (Spek, 2009).

5-Amino-2,4,6-triiodoisophthalic acid (I3C)

Table D.1: Crystal data and refinement details for **2**.

5-Amino-2,4,6-triiodoisophthalic acid (I3C)	
empirical formula	$C_8H_4I_3NO_4{\cdot}H_2O$
formula weight	576.84
T (K)	100(2)
λ (Å)	1.54178
crystal system	orthorhombic
space group	*Pbca* (No. 61)
a (Å)	9.214(1)
b (Å)	15.735(2)
c (Å)	18.816(2)
α (°)	90
β (°)	90
γ (°)	90
V (Å^3)	2728.0(5)
Z	8
$\rho_{calcd.}$ (g cm^{-3})	2.809
μ (mm^{-1})	54.111
$F(000)$	2080
crystal size (mm)	$0.08 \times 0.05 \times 0.03$
θ range (°)	$4.70 - 74.28$
hkl range	± 10, $-17 \rightarrow 19$, ± 23
reflections collected	49139
independent reflections (R_{int})	2716 (0.0433)
completeness to theta = 70.00° (%)	99.6
absorption correction	semi-empirical from equivalents
$T_{\text{max}} / T_{\text{min}}$	0.3448 / 0.1064
refinement method	full-matrix least-squares on F^2
data	2716
refined parameters	173
restraints	14
goodness-of-fit	1.032
$R1, wR2$ [$I > 2\sigma(I)$]	0.0256, 0.0598
$R1, wR2$ [all data]	0.0276, 0.0610
extinction coefficient	0.000058(8)
residual electron density (e Å^{-3})	0.715 / −1.710

5-Amino-2,4,6-tribromoisophthalic acid (B3C)

The B3C crystal was non-merohedrally twinned. The first and second domain are related by a two-fold axis about the real axis 1 0 0. Integration using both orientation matrices simultaneously was carried out with *SAINT*. The fractional contribution of the second domain was refined to 0.3468(6).

Table D.2: Crystal data and refinement details for **3**.

5-Amino-2,4,6-tribromoisophthalic acid (B3C)	
empirical formula	$C_8H_4Br_3NO_4$
formula weight	417.85
T (K)	100(2)
λ (Å)	1.54178
crystal system	monoclinic
space group	$P2_1/c$ (No. 14)
a (Å)	16.728(2)
b (Å)	11.449(2)
c (Å)	23.649(3)
α (°)	90
β (°)	99.69(3)
γ (°)	90
V (Å^3)	4464.6(11)
Z	16
$\rho_{calcd.}$ (g cm^{-3})	2.487
μ (mm^{-1})	13.443
$F(000)$	3136
crystal size (mm)	$0.10 \times 0.10 \times 0.05$
θ range (°)	$2.68 - 73.98$
hkl range	± 20, ± 13, $-28 \rightarrow 27$
reflections collected	97602
independent reflections (R_{int})	8717 (0.0484)
completeness to theta = 70.00° (%)	98.3
absorption correction	semi-empirical from equivalents
$T_{\max} / T_{\min}$	0.5529 / 0.3466
refinement method	full-matrix least-squares on F^2
data	8717
refined parameters	626
restraints	106
goodness-of-fit	1.132
$R1, wR2$ [$I > 2\sigma(I)$]	0.0258, 0.0631
$R1, wR2$ [all data]	0.0276, 0.0641
residual electron density (e Å^{-3})	0.640 / −0.546

3-Amino-2,4,6-tribromo-5-methoxybenzoic acid (B3M)

B3M crystallises in the chiral space group $P6_5$. The absolute structure parameter (Flack and Bernardinelli, 2000) was refined to 0.06(2).

Table D.3: Crystal data and refinement details for **5**.

3-Amino-2,4,6-tribromo-5-methoxybenzoic acid (B3M)	
empirical formula	$C_8H_6Br_3NO_3{\cdot}H_2O$
formula weight	421.88
T (K)	100(2)
λ (Å)	1.54178
crystal system	hexagonal
space group	$P6_5$ (No. 170)
a (Å)	9.002(2)
b (Å)	9.002(2)
c (Å)	25.333(3)
α (°)	90
β (°)	90
γ (°)	120
V (Å^3)	1777.8(6)
Z	6
$\rho_{calcd.}$ (g cm^{-3})	2.364
μ (mm^{-1})	12.660
$F(000)$	1200
crystal size (mm)	$0.20 \times 0.10 \times 0.10$
θ range (°)	$5.67 - 72.13$
hkl range	$-11 \rightarrow 8,\ -6 \rightarrow 11,\ -31 \rightarrow 27$
reflections collected	17069
independent reflections (R_{int})	2260 (0.0234)
completeness to theta = 67.50° (%)	100.0
absorption correction	semi-empirical from equivalents
$T_{\max} / T_{\min}$	0.3641 / 0.1862
refinement method	full-matrix least-squares on F^2
data	2260
refined parameters	166
restraints	13
goodness-of-fit	1.182
$R1, wR2$ [$I > 2\sigma(I)$]	0.0156, 0.0402
$R1, wR2$ [all data]	0.0156, 0.0402
absolute structure parameter	0.06(2)
extinction coefficient	0.00014(2)
residual electron density (e Å^{-3})	1.161 / −0.352

2,4,6-Tribromo-5-hydroxyisophthalic acid (B3O)

Table D.4: Crystal data and refinement details for **20**.

2,4,6-Tribromo-5-hydroxyisophthalic acid (B3O)	
empirical formula	$C_8H_3Br_3O_5{\cdot}H_2O$
formula weight	436.85
T (K)	100(2)
λ (Å)	1.54178
crystal system	monoclinic
space group	$P2_1/c$ (No. 14)
a (Å)	8.745(2)
b (Å)	13.105(3)
c (Å)	10.652(2)
α (°)	90
β (°)	103.07(3)
γ (°)	90
V (Å^3)	1189.1(4)
Z	4
$\rho_{calcd.}$ (g cm^{-3})	2.440
μ (mm^{-1})	12.764
$F(000)$	824
crystal size (mm)	$0.05 \times 0.05 \times 0.05$
θ range (°)	$5.19 - 72.50$
hkl range	± 10, ± 16, ± 13
reflections collected	21124
independent reflections (R_{int})	2282 (0.0467)
completeness to theta = 67.50° (%)	98.1
absorption correction	semi-empirical from equivalents
$T_{\max} / T_{\min}$	0.5678 / 0.5678
refinement method	full-matrix least-squares on F^2
data	2282
refined parameters	169
restraints	15
goodness-of-fit	1.118
$R1, wR2$ [$I > 2\sigma(I)$]	0.0291, 0.0718
$R1, wR2$ [all data]	0.0310, 0.0763
residual electron density (e Å^{-3})	1.231 / −0.470

Tetrabromoterephthalic acid (B4C)

Table D.5: Crystal data and refinement details for **21**.

Tetrabromoterephthalic acid (B4C)	
empirical formula	$C_8H_2Br_4O_4{\cdot}2\,H_2O$
formula weight	517.77
T (K)	100(2)
λ (Å)	1.54178
crystal system	monoclinic
space group	$P2_1/n$ (No. 14)
a (Å)	5.8050(10)
b (Å)	14.605(3)
c (Å)	16.433(3)
α (°)	90
β (°)	97.17(3)
γ (°)	90
V (Å^3)	1382.3(4)
Z	4
$\rho_{calcd.}$ (g cm^{-3})	2.488
μ (mm^{-1})	14.395
$F(000)$	968
crystal size (mm)	$0.08 \times 0.05 \times 0.05$
θ range (°)	$4.06 - 72.95$
hkl range	$\pm 7,\ -14 \rightarrow 17,\ -18 \rightarrow 19$
reflections collected	16660
independent reflections (R_{int})	2671 (0.0327)
completeness to theta = 67.50° (%)	98.5
absorption correction	semi-empirical from equivalents
T_{max} / T_{min}	0.5330 / 0.3922
refinement method	full-matrix least-squares on F^2
data	2671
refined parameters	181
restraints	18
goodness-of-fit	1.094
$R1, wR2$ [$I > 2\sigma(I)$]	0.0190, 0.0447
$R1, wR2$ [all data]	0.0196, 0.0452
residual electron density (e Å^{-3})	0.378 / −0.502

Appendix E

Experimental data for small molecules

General remarks

All reagents obtained from commercial sources were used without further purification. Thin-layer chromatography (TLC) was performed on precoated silica gel plates with UV indicator (Fluka, Sigma-Aldrich Chemie GmbH). ^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker Avance 300 MHz or 500 MHz spectrometer. ^{1}H and ^{13}C chemical shifts are reported in ppm relative to residual solvent signals of DMSO-d_6 (2.49 ppm and 39.5 ppm). Signals were assigned as: s (singlet), s_{br} (broad singlet) or m (multiplet). Microanalyses were performed by the Analytical Laboratory of the Institute for Inorganic Chemistry at the University of Göttingen. EI mass spectra were measured on a Finnigan MAT 95, ESI mass and ESI HRMS spectra on a Bruker mircOTOF instrument.

5-Amino-2,4,6-triiodoisophthalic acid I3C (2)

Iodomonochloride (10.5 g, 60 mmol) was added to a solution of KCl (7.2 g, 97 mmol) in 100 mL demineralised water. The resulting orange solution was stirred for 5 min and filtered. 5-Aminoisophthalic acid **1** (3.32 g, 18 mmol) was suspended in 210 mL water. The mixture was heated to 60 °C. The iodomonochloride solution was slowly added to the suspension. After stirring overnight, the solid material was filtered off and dissolved in 25 mL 1 M KOH. After adding charcoal to the solution the mixture was heated to 60 °C. The charcoal was filtered off and the solution was allowed to cool to room temperature. Concentrated hydrochloric acid was added and the precipitate formed was filtered off. After drying *in vacuo* I3C (**2**) was obtained as brown powder (2.43 g, 4.4 mmol, 24 %, not optimised). ^{1}H NMR (500 MHz, DMSO): $\delta = 5.53$ (s_{br}, NH_2). ^{13}C NMR (125 MHz, DMSO): $\delta = 70.4$ (Ar-C), 78.1 (Ar-C), 147.8 (Ar-C), 148.5 (Ar-C), 169.9 (COOH). MS (ESI-): m/z (%) = 557.6 (18) $[M\text{-}H]^-$, 513.7 (13) $[M\text{-}COOH]^-$, 1116.3 (100) $[2M\text{-}H]^-$. Elemental analysis: Calcd. (%) for $C_8H_4I_3NO_4{\cdot}H_2O$: C 16.66, H 1.05, N 2.43. Found: C 16.53, H 1.02, N 2.43. I3C was recrystallised from a methanol-acetonitrile solution with slowly evaporating the solvents to obtain crystals suitable for X-ray single crystal diffraction.

5-Amino-2,4,6-tribromoisophthalic acid B3C (3)

5-Aminoisophthalic acid **1** (5.43 g, 30 mmol) was added to 30 mL demineralised water. A small amount of iron(III)bromide was added. With stirring bromine (5.4 mL, 17.0 g, 106 mmol) was added slowly with a syringe within 1 h to the suspension. After stirring for 5 days the precipitate was filtered off and washed with small amounts of cold water. After drying *in vacuo* B3C (**3**) was obtained as pale pink powder (10.6 g, 25 mmol, 84 %). ^{1}H NMR (300 MHz, DMSO): $\delta = 5.91$ (s_{br}, NH_2). ^{13}C NMR (75 MHz, DMSO): $\delta = 98.1$ (Ar-C), 103.3 (Ar-C), 139.1 (Ar-C), 143.4 (Ar-C), 166.6 (COOH). MS (ESI-): m/z (%) = 415.8 (100) [M-H]$^-$, 371.8 (72) [M-COOH]$^-$. Elemental analysis: Calcd. (%) for $C_8H_4Br_3NO_4$: C 23.00, H 0.96, N 3.35. Found: C 22.70, H 0.92, N 3.37. B3C was recrystallised from a methanol-acetonitrile solution by slowly evaporating the solvents to obtain crystals suitable for single crystal X-ray diffraction.

3-Amino-2,4,6-tribromo-5-methoxybenzoic acid B3M (5)

3-Amino-5-methoxybenzoic acid **4** (1.67 g, 10 mmol) was added to 10 mL demineralised water. A small amount of iron(III)bromide was added. With stirring bromine (2.1 mL, 6.6 g, 42 mmol) was added slowly with a syringe within 1 h to the suspension. After stirring for 4 days the precipitate was filtered off and washed with small amounts of cold water. After drying *in vacuo* B3M (**5**) was obtained as pale orange powder (3.76 g, 9.3 mmol, 93 %). ^{1}H NMR (300 MHz, DMSO): $\delta = 3.75$ (s, MeO), 5.75 (s_{br}, NH_2). ^{13}C NMR (75 MHz, DMSO): $\delta = 60.2$ (MeO), 98.0 (Ar-C), 98.6 (Ar-C), 103.2 (Ar-C), 139.0 (Ar-C), 144.3 (Ar-C), 153.6 (Ar-C), 166.7 (COOH). MS (ESI-): m/z (%) = 401.8 (100) [M-H]$^-$, 359.8 (63) [M-COOH]$^-$. Elemental analysis: Calcd. (%) for $C_8H_6Br_3NO_3$: C 23.79, H 1.50, N 3.47. Found: C 23.46, H 1.48, N 3.47. B3M was recrystallised from a methanol solution by slowly evaporating the solvent to obtain crystals suitable for single crystal X-ray diffraction.

2,4,6-Tribromo-5-hydroxyisophthalic acid B3O (20)

5-Hydroxyisophthalic acid **19** (5.46 g, 30 mmol) was added to 110 mL demineralised water. A small amount of iron(III)bromide was added. With stirring bromine (5.4 mL, 17.0 g, 106 mmol) was added slowly with a syringe within 1 h to the suspension. After stirring for 7 days, the suspension was heated to 80 °C for 1 h. Since the reaction was not complete (monitored by TLC), additional bromine was added (1.0 mL) and the suspension was stirred for 2 days at room temperature. The precipitate was filtered off and washed with small amounts of cold water. After drying *in vacuo* B3O (**20**) was obtained as pale yellow powder (9.3 g, 22 mmol, 74 %). ^{1}H NMR (500 MHz, DMSO): $\delta = 10.84$ (s_{br}, OH). ^{13}C NMR (125 MHz, DMSO): $\delta = 103.6$ (Ar-C), 109.2 (Ar-C), 139.7 (Ar-C), 151.4 (Ar-C), 166.3 (COOH). MS (ESI-): m/z (%) = 416.7 (50) [M-H]$^-$, 372.8 (41) [M-COOH]$^-$, 248.8 (77) [M-Br-2COOH]$^-$, 207.9 (100) [M-2H]$^{2-}$. HRMS (ESI-): Calcd. (m/z) for $C_8H_2Br_3O_5$ [M-H]$^-$: 414.7458. Found 414.7455. B3O was recrystallised from an ethanol solution with a few drops of toluene added by slowly evaporating the solvents to obtain crystals suitable for single crystal X-ray diffraction.

N-(2,4,6-Tribromophenyl)guanidinium nitrate (15)

N-Phenylguanidinium nitrate (14)

Aniline **13** (1.9 mL, 1.86 g, 20 mmol) was dissolved in 15 mL ethanol. Nitric acid (2.0 mL) was added slowly. Cyanamide solution (50 % in water, 2.5 mL, 30 mmol) was added. The solution was stirred under reflux for 20 h. Complete conversion was confirmed by TLC and ninhydrin staining. After cooling to room temperature, diethylether (50 mL) was added and the reaction mixture was stored overnight at 4 °C. The resulting precipitate was filtered off, washed with diethylether and dried *in vacuo*. **14** was obtained as beige solid (2.54 g, 13 mmol, 64 %). ^{1}H NMR (500 MHz, DMSO): $\delta = 7.22 - 7.45$ (m, 9 H, 5 Ar-H, 4 NH), 9.61 (s_{br}, 1 H, NH-Ph). ^{13}C NMR (125 MHz, DMSO): $\delta = 124.5$ (Ar-C), 126.5 (Ar-C), 129.7 (Ar-H), 135.3 (Ar-H), 155.7 (Ph-NH-C-NH(NH_2)). MS (EI+): m/z (%) = 135 (85) $[M-H]^+$, 93 (100) $[M-C-NH(NH_2)]^+$.

N-(2,4,6-Tribromophenyl)guanidinium nitrate (15)

14 (0.99 g, 5 mmol) was added to 15 mL demineralised water. A small amount of iron(III)-bromide was added. With stirring bromine (0.8 mL, 2.5 g, 16 mmol) was added slowly with a syringe. After stirring for 3 days at room temperature, the reaction mixture was heated to 80 °C and stirred for 3 days. Additional bromine (0.5 mL) was added and the mixture was stirred at elevated temperature for 3 days. After cooling to room temperature, the solid material was filtered off. The fine yellow solid was separated from a small ball of red solid. After drying *in vacuo* **15** was obtained (0.25 g, 6 mmol, 12 %, not optimised). ^{1}H NMR (500 MHz, DMSO): $\delta = 7.43$ (s_{br}, 4 H, NH), 8.09 (m, 2 H, Ar-H), 9.60 (s_{br}, 1 H, NH-Ph). ^{13}C NMR (125 MHz, DMSO): $\delta = 123.1$ (Ar-C), 126.1 (Ar-C), 132.3 (Ar-C), 135.0 (Ar-C), 155.4 (Ph-NH-C-NH(NH_2)). MS (ESI+): m/z (%) = 371.8 (100) $[M]^+$, 293.9 (9) $[M-Br]^+$. HRMS (ESI+): Calcd. (m/z) for $C_7H_6Br_3$ $[M]^+$: 369.8185. Found 369.8185.

N-(2,4,6-Tribromo-3-5-dimethylphenyl)guanidinium nitrate (18)

N-(3-5-Dimethylphenyl)guanidinium nitrate (17)

3,5-Dimethylaniline **16** (3.75 mL, 3.64 g, 30 mmol) was dissolved in 20 mL ethanol. Nitric acid (3.0 mL) was added slowly. Cyanamide solution (50 % in water, 3.75 mL, 45 mmol) was added. The solution was stirred under reflux for 20 h. Complete conversion was confirmed by TLC and ninhydrin staining. After cooling to room temperature, diethylether (40 mL) was added and the reaction mixture was stored overnight at 4 °C. The resulting precipitate was filtered off, washed with diethylether and dried *in vacuo*. **17** was obtained as beige solid (3.70 g, 16 mmol, 55 %). ^{1}H NMR (300 MHz, DMSO): $\delta = 2.27$ (s, 6 H, CH_3), 6.83 (m, 2 H, Ar-H), 6.93 (m, 1 H, Ar-H), 7.23 (s, 4 H, NH), 9.47 (s, 1 H, NH-Ph). ^{13}C NMR (75 MHz, DMSO): $\delta = 20.8$ (CH_3), 122.1 (Ar-C), 128.0 (Ar-C), 134.9 (Ar-C), 139.0 (Ar-C), 155.7 (Ph-NH-C-NH(NH_2)). MS (EI+): m/z (%) = 163 (100) $[M-H]^+$, 121 (100) $[M-C-NH(NH_2)]^+$.

N-(2,4,6-Tribromo-3-5-dimethylphenyl)guanidinium nitrate (18)

17 (0.57 g, 2.5 mmol) was added to 20 mL demineralised water. A small amount of iron(III)-bromide was added. With stirring bromine (0.5 mL, 1.6 g, 10 mmol) was added slowly with a syringe. After stirring for 4 days at room temperature the reaction mixture was heated to 80 °C and stirred for 2 days at elevated temperature. After cooling to room temperature, the solid material was filtered off. After drying *in vacuo* **15** was obtained as yellow solid (0.90 g, 2 mmol, 80 %). ^{1}H NMR (500 MHz, DMSO): $\delta = 2.62$ (s, 6 H, CH_3), 7.38 (s_{br}, 4 H, NH), 9.72 (s, 1 H, NH-Ph). ^{13}C NMR (125 MHz, DMSO): $\delta = 25.8$ (CH_3), 125.2 (Ar-C), 127.8 (Ar-C), 132.0 (Ar-C), 138.0 (Ar-C), 155.4 (Ph-NH-C-NH(NH_2)). MS (ESI+): m/z (%) = 399.9 (100) $[M]^+$, 322.0 (4) $[M\text{-}Br]^+$. HRMS (ESI+): Calcd. (m/z) for $C_9H_6Br_3$ $[M]^+$: 397.8498. Found 397.8496.

Tetrabromoterephthalic acid B4C (21)

B4C was obtained from Sigma-Aldrich and recrystallised from a water-acetone solution by slowly evaporating the solvents to obtain crystals suitable for single crystal X-ray diffraction.

Appendix F

Sticky triangles in the news

This is a list of publications and websites about the sticky triangles.[1]

- available phasing kits: *JBS Magic Triangle* (Jena Bioscience)
  I3C phasing kit (Hampton Research)
  I3C phasing kit (Molecular Dimensions)
- highlighted article in the IUCr newsletter: Beck, T., Krasauskas, A., Gruene, T., and Sheldrick, G. M. (2008). A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr. Section D*, 64:1179–1182.
- journal front cover: Beck, T., da Cunha, C. E. and Sheldrick, G. M. (2009). How to get the magic triangle and the MAD triangle into your protein crystal. *Acta Crystallogr. Section F*, 65:1068–1070.
- report about I3C and B3C in the newsletter of the German Society for Crystallography (DGK):
  `http://www.uni-leipzig.de/~straeter/newsletter/AK1_newsletter_06_2009.html`
- Mh-p37, the first novel protein structure solved with I3C (Sippel *et al.*, 2008), featured as Rigaku Structure of the Month:
  `http://www.rigaku.com/protein/sotm-006.html`
- Michael Sawaya's instructions on how to prepare an I3C derivative:
  `http://www.doe-mbi.ucla.edu/~sawaya/tutorials/Phasing/iodide_recipes.html`

[1] As of October 2010. Please refer to `http://shelx.uni-ac.gwdg.de/tbeck` for further information.

Bibliography

Agarwal, B. K., editor (1991). *X-ray spectroscopy: an introduction*. Berlin, Germany: Springer.

Agniswamy, J., Joyce, M. G., Hammer, C. H., and Sun, P. D. (2008). Towards a rational approach for heavy-atom derivative screening in protein crystallography. *Acta Crystallogr. Section D*, 64:354–367.

Almen, T. (1969). Contrast agent design. Some aspects on the synthesis of water soluble contrast agents of low osmolality. *J. Theor. Biol.*, 24:216–226.

Auffinger, P., Hays, F. A., Westhof, E., and Ho, P. S. (2004). Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.*, 101:16789–16794.

Banumathi, S., Dauter, M., and Dauter, Z. (2003). Phasing at high resolution using Ta_6Br_{12} cluster. *Acta Crystallogr. Section D*, 59:492–498.

Beck, T., da Cunha, C. E., and Sheldrick, G. M. (2009a). How to get the magic triangle and the MAD triangle into your protein crystal. *Acta Crystallogr. Section F*, 65:1068–1070.

Beck, T., Gruene, T., and Sheldrick, G. M. (2010). The magic triangle goes MAD: experimental phasing with a bromine derivative. *Acta Crystallogr. Section D*, 66:374–380.

Beck, T., Herbst-Irmer, R., and Sheldrick, G. M. (2009b). 5-Amino-2,4,6-tribromoisophthalic acid: the MAD triangle for experimental phasing. *Acta Crystallogr. Section C*, 65:o237–o239.

Beck, T., Krasauskas, A., Gruene, T., and Sheldrick, G. M. (2008). A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr. Section D*, 64:1179–1182.

Beck, T. and Sheldrick, G. M. (2008). 5-Amino-2,4,6-triiodoisophthalic acid monohydrate. *Acta Crystallogr. Section E*, 64:o1286.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Jr., E. E. M., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–542.

Bissantz, C., Kuhn, B., and Stahl, M. (2010). A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.*, 53:5061–5084.

Blake, C. C., Fenn, R. H., North, A. C., Phillips, D. C., and Poljak, R. J. (1962). Structure of lysozyme. A Fourier map of the electron density at 6 angstrom resolution obtained by x-ray diffraction. *Nature*, 196:1173–1176.

Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C., and Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*, 206:757–761.

Blundell, T. L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.*, 1:45–54.

Boggon, T. J. and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure*, 8:R143–149.

Bondi, A. (1964). van der Waals Volumes and Radii. *J. Phys. Chem.*, 68:441–451.

Borkakoti, N., Moss, D. S., and Palmer, R. A. (1982). Ribonuclease-A: least-squares refinement of the structure at 1.45 Å resolution. *Acta Crystallogr. Section B*, 38:2210–2217.

Bruker (2007). XPREP. *Bruker AXS Inc., Madison, Wisconsin, USA.*

Bruker (2008). XP. *Bruker AXS Inc., Madison, Wisconsin, USA.*

Bruker (2009). SAINT. *Bruker AXS Inc., Madison, Wisconsin, USA.*

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472–475.

Carvin, D., Islam, S. A., Sternberg, M. J. E., and Blundell, T. L. (2001). *International Tables for Crystallography, Volume F: Crystallography of biological macromolecules*, chapter 12.1: The preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering, pages 247–255. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 66:12–21.

Clark, T., Hennemann, M., Murray, J. S., and Politzer, P. (2007). Halogen bonding: the sigma-hole. Proceedings of "Modeling interactions in biomolecules II", Prague, September 5th-9th, 2005. *J Mol Model*, 13:291–296.

Coelho, A. V., Macedo, S., Matias, P. M., Thompson, A. W., LeGall, J., and Carrondo, M. A. (2001). Structure determination of bacterioferritin from *Desulfovibrio desulfuricans* by the MAD method at the Fe *K*-edge. *Acta Crystallogr. Section D*, 57:326–329.

Collaborative Computational Project Number 4 (1994). The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. Section D*, 50:760–763.

Corbett, M. C., Latimer, M. J., Poulos, T. L., Sevrioukova, I. F., Hodgson, K. O., and Hedman, B. (2007). Photoreduction of the active site of the metalloprotein putidaredoxin by synchrotron radiation. *Acta Crystallogr. Section D*, 63:951–960.

Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997). Use of chemically modified nucleotides to determine a 62-nucleotide RNA crystal structure: a survey of phosphorothioates, Br, Pt and Hg. *J. Biomol. Struct. Dyn.*, 15:165–172.

Cowtan, K. (1994). 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, 31:34–38.

Cowtan, K. (2006). The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Section D*, 62:1002–1011.

Cowtan, K. (2010). `http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html`.

Cromer, D. T. and Mann, J. B. (1968). X-ray scattering factors computed from numerical Hartree–Fock wave functions. *Acta Crystallogr. Section A*, 24:321–324.

Daniel, R. M., Dunn, R. V., Finney, J. L., and Smith, J. C. (2003). The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, 32:69–92.

Dauter, Z. and Adamiak, D. A. (2001). Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallogr. Section D*, 57:990–995.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., and Sheldrick, G. M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol.*, 289:83–92.

Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr. Section D*, 56:232–237.

Dauter, Z., Li, M., and Wlodawer, A. (2001). Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics. *Acta Crystallogr. D Biol. Crystallogr.*, 57:239–249.

Dawson, P., Cosgrove, D. O., and Grainger, R. G. (1999). *Textbook of Contrast Media.* Oxford, UK: Isis Medical Media Ltd.

de L. Kronig, R. (1926). On the theory of dispersion of X-rays. *J. Opt. Soc. Am.*, 12:547–557.

de Vos, A. M., Hatada, M., van der Wel, H., Krabbendam, H., Peerdeman, A. F., and Kim, S. H. (1985). Three-dimensional structure of thaumatin I, an intensely sweet protein. *Proc. Natl. Acad. Sci. U.S.A.*, 82:1406–1409.

Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H., and Sheldrick, G. M. (2003a). In-house measurement of the sulfur anomalous signal and its use for phasing. *Acta Crystallogr. Section D*, 59:688–696.

Debreczeni, J. É., Girmann, B., Zeeck, A., Krätzner, R., and Sheldrick, G. M. (2003b). Structure of viscotoxin A3: disulfide location from weak SAD data. *Acta Crystallogr. Section D*, 59:2125–2132.

DeLano, W. (2010). The PyMOL molecular graphics system. *Version 1.2r3pre, Schrödinger, LLC.*

Diederichs, K. and Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.*, 4:269–275.

Djinovic-Carugo, K., Everitt, P., and Tucker, P. A. (1998). A cell for producing xenon-derivative crystals for cryocrystallographic analysis. *J. Appl. Cryst.*, 31:812–814.

Dobson, A. J. and Gerkin, R. E. (1998). 5-Aminoisophthalic Acid Hemihydrate. *Acta Crystallogr. Section C*, 54:1503–1505.

Domenicano, A., Schultz, G., Hargittai, I., Colapietro, M., Portalone, G., George, P., and Bock, C. W. (1990). Molecular structure of nitrobenzene in the planar and orthogonal conformations. *Struct. Chem.*, 1:107–122.

Duchamp, D. J. and Marsh, R. E. (1969). The crystal structure of trimesic acid (benzene-1,3,5-tricarboxylic acid). *Acta Crystallogr. Section B*, 25:5–19.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of *Coot*. *Acta Crystallogr. Section D*, 66:486–501.

Engh, R. A. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. Section A*, 47:392–400.

Ennifar, E., Carpentier, P., Ferrer, J.-L., Walter, P., and Dumas, P. (2002). X-ray-induced debromination of nucleic acids at the BrK absorption edge and implications for MAD phasing. *Acta Crystallogr. Section D*, 58:1262–1268.

Ermer, O. and Neudörfl, J. (2001). Comparative Supramolecular Chemistry of Coronene, Hexahelicene, and [18]Crown-6: Hydrated and Solvated Molecular Complexes of [18]Crown-6 with 5-Hydroxyisophthalic Acid and Related Di- and Tricarboxylic Acids. *Chem. Eur. J.*, 7:4961–4980.

Evans, G. and Bricogne, G. (2003). Triiodide derivatization in protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 59:1923–1929.

Flack, H. D. and Bernardinelli, G. (2000). Reporting and evaluating absolute-structure and absolute-configuration determinations. *J. Appl. Cryst.*, 33:1143–1148.

Fodor, K., Harmat, V., Neutze, R., Szilagyi, L., Graf, L., and Katona, G. (2006). Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry*, 45:2114–2121.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., J., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009). Gaussian 09, revision a.02. *Gaussian, Inc., Wallingford CT*.

Fujinaga, M. and Read, R. J. (1987). Experiences with a new translation-function program. *J. Appl. Cryst.*, 20:517–521.

Garman, E. (1999). Leaving no element of doubt: analysis of proteins using microPIXE. *Structure*, 7:R291–299.

Garman, E. and Murray, J. W. (2003). Heavy-atom derivatization. *Acta Crystallogr. D Biol. Crystallogr.*, 59:1903–1913.

Garman, E. F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. Section D*, 66:339–351.

Garman, E. F. and Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr. Section D*, 62:32–47.

Giacovazzo, C. and Siliqi, D. (1997). Improving Direct-Methods Phases by Heavy-Atom Information and Solvent Flattening. *Acta Crystallogr. Section A*, 53:789–798.

Girard, É., Stelter, M., Anelli, P. L., Vicat, J., and Kahn, R. (2003). A new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives: results from SAD experiments with hen egg-white lysozyme and urate oxidase from *Aspergillus flavus*. *Acta Crystallogr. Section D*, 59:118–126.

Gnuplot (2010). `http://www.gnuplot.info`.

Gottardi, W. (1969). Über Bromierungen mit Dibromoisocyanursäure unter ionischen Bedinungen, 2. Mitt.: Perbromierungen. *Monatshefte für Chemie*, 100:42–50.

Hampton Research (2008). Silver Bullets Screen, HR2-096, Aliso Viejo, CA 92656-3317, USA.

Hendrickson, W. A. (1999). Maturation of MAD phasing for the determination of macromolecular structures. *J. Synchrotron. Rad.*, 6:845–851.

Hendrickson, W. A., Horton, J. R., and LeMaster, D. (1991). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.*, 9:1665–1672.

Hendrickson, W. A., Smith, J. L., and Sheriff, S. (1985). Direct phase determination based on anomalous scattering. *Meth. Enzymol.*, 115:41–55.

Holland, D. R., Hausrath, A. C., Juers, D., and Matthews, B. W. (1995). Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Sci.*, 4:1955–1965.

Holmes, M. A. and Matthews, B. W. (1982). Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.*, 160:623–639.

Hübschle, C. B. (2010). Molecoolqt (rev. 171) - a molecule viewer for charge density related science. *University of Göttingen, Germany*, `http://www.molecoolqt.de`.

Hübschle, C. B. and Luger, P. (2006). *MolIso* – a program for colour-mapped iso-surfaces. *J. Appl. Cryst.*, 39:901–904.

Hunter, C. A. and Sanders, J. K. M. (1990). The nature of $\pi - \pi$ interactions. *J. Am. Chem. Soc.*, 112:5525–5534.

Jeffrey, G. A. and Saenger, W. (1991). *Hydrogen Bonding in Biological Structures*. Berlin, Germany: Springer Verlag.

Joyce, M. G., Radaev, S., and Sun, P. D. (2010). A rational approach to heavy-atom derivative screening. *Acta Crystallogr. Section D*, 66:358–365.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, 26:795–800.

Karle, J. (1968). Partial structural information combined with the tangent formula for noncentrosymmetric crystals. *Acta Crystallogr. Section B*, 24:182–186.

Karle, J. (1980). Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int. J. Quant. Chem.*, 18:357–367.

Karle, J. and Hauptman, H. (1956). A theory of phase determination for the four types of non-centrosymmetric space groups $1P222$, $2P22$, $3P_12$, $3P_22$. *Acta Crystallographica*, 9:635–651.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662–666.

Ko, T.-P., Day, J., Greenwood, A., and McPherson, A. (1994). Structures of three crystal forms of the sweet protein thaumatin. *Acta Crystallogr. Section D*, 50:813–825.

Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T., and Ficner, R. (2009). *De novo* sulfur SAD phasing of the lysosomal 66.3kDa protein from mouse. *Acta Crystallogr. Section D*, 65:220–228.

Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, 3:1171–1179.

Larsen, A. A., Moore, C., Sprague, J., Cloke, B., Moss, J., and Hoppe, J. O. (1956). Iodinated 3,5-diaminobenzoic acid derivatives. *J. Am. Chem. Soc.*, 78:3210–3216.

Leiserowitz, L. (1976). Molecular packing modes. Carboxylic acids. *Acta Crystallogr. Section B*, 32:775–802.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, 50:437–450.

Lukat, P., Hoffmann, M., and Einsle, O. (2008). Crystal packing of the c_6-type cytochrome OmcF from *Geobacter sulfurreducens* is mediated by an N-terminal Strep-tag II. *Acta Crystallogr. Section D*, 64:919–926.

Lunin, V. Y. and Woolfson, M. M. (1993). Mean phase error and the map-correlation coefficient. *Acta Crystallogr. Section D*, 49:530–533.

McPherson, A. and Cudney, B. (2006). Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J. Struct. Biol.*, 156:387–406.

Merrit, E. A. (2010). `http://skuld.bmsc.washington.edu/scatter/AS_form.html`.

Miyatake, H., Hasegawa, T., and Yamano, A. (2006). New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr. Section D*, 62:280–289.

Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A., and Weiss, M. S. (2007). On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Crystallogr. Section D*, 63:366–380.

Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A., and Weiss, M. S. (2005). On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. *Acta Crystallogr. Section D*, 61:1263–1272.

Murray, J. and Garman, E. (2002). Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *J. Synchrotron Rad.*, 9:347–354.

Murray, J. W., Garman, E. F., and Ravelli, R. B. G. (2004). X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *J. Appl. Cryst.*, 37:513–522.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Section D*, 53:240–255.

Nagem, R. A., Polikarpov, I., and Dauter, Z. (2003). Phasing on rapidly soaked ions. *Meth. Enzymol.*, 374:120–137.

Nyburg, S. C. (1979). 'Polar flattening': non-spherical effective shapes of atoms in crystals. *Acta Crystallogr. Section A*, 35:641–645.

Nyburg, S. C. and Faerman, C. H. (1985). A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon. *Acta Crystallogr. Section B*, 41:274–279.

Oliéric, V., Rieder, U., Lang, K., Serganov, A., Schulze-Briese, C., Micura, R., Dumas, P., and Ennifar, E. (2009). A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *RNA*, 15:707–715.

Owen, R. L., Rudino-Pinera, E., and Garman, E. F. (2006). Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl. Acad. Sci. U.S.A.*, 103:4912–4917.

Pal, G. P., Kavounis, C. A., Jany, K. D., and Tsernoglou, D. (1994). The three-dimensional structure of the complex of proteinase K with its naturally occurring protein inhibitor, PKI3. *FEBS Lett.*, 341:167–170.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2009). On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. Section D*, 65:1089–1097.

Pannu, N. S. and Read, R. J. (1996). Improved Structure Refinement Through Maximum Likelihood. *Acta Crystallogr. Section A*, 52:659–668.

Pape, T. and Schneider, T. R. (2004). *HKL2MAP*: a graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Cryst.*, 37:843–844.

Parkesh, R., Gowin, W., Lee, T. C., and Gunnlaugson, T. (2006). Synthesis and evaluation of potential CT (computer tomography) contrast agents for bone structure and microdamage analysis. *Org. Biomol. Chem.*, 4:3611–3617.

Pearson, A. R. and Owen, R. L. (2009). Combining X-ray crystallography and single-crystal spectroscopy to probe enzyme mechanisms. *Biochem. Soc. Trans.*, 37:378–381.

Polgar, L. (2005). The catalytic triad of serine peptidases. *Cell. Mol. Life Sci.*, 62:2161–2172.

Politzer, P., Lane, P., Concha, M. C., Ma, Y., and Murray, J. S. (2007). An overview of halogen bonding. *J Mol Model*, 13:305–311.

Politzer, P., Murray, J. S., and Clark, T. (2010). Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys Chem Chem Phys*, 12:7748–7757.

Pompidor, G., Maury, O., Vicat, J., and Kahn, R. (2010). A dipicolinate lanthanide complex for solving protein structures using anomalous diffraction. *Acta Crystallogr. Section D*, 66:762–769.

Prince, E., editor (2004). *International Tables for Crystallography, Volume C: Mathematical, Physical and Chemical Tables*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95–99.

Ravelli, R. B., Leiros, H. K., Pan, B., Caffrey, M., and McSweeney, S. (2003). Specific radiation damage can be used to solve macromolecular crystal structures. *Structure*, 11:217–224.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. Section A*, 42:140–149.

Recacha, R., Carson, M., Costanzo, M. J., Maryanoff, B., DeLucas, L. J., and Chattopadhyay, D. (1999). Structure of the RWJ-51084–bovine pancreatic β-trypsin complex at 1.8 Å. *Acta Crystallogr. Section D*, 55:1785–1791.

Riley, K. E. and Hobza, P. (2008). Investigations into the Nature of Halogen Bonding Including Symmetry Adapted Perturbation Theory Analyses. *J. Chem. Theory Comput.*, 4:232–242.

Rodriguez, D. D., Grosse, C., Himmel, S., Gonzalez, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M., and Usón, I. (2009). Crystallographic ab initio protein structure solution below atomic resolution. *Nat. Methods*, 6:651–653.

Roversi, P., Johnson, S., and Lea, S. M. (2010). With phases: how two wrongs can sometimes make a right. *Acta Crystallogr. Section D*, 66:420–425.

Rupp, B. (2010). *Biomolecular Crystallography - Principles, Practice and Application to Structural Biology*. New York, USA: Garland Science.

Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H., and Watson, H. C. (1978). The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution: comparisons with the structure of alpha-chymotrypsin. *J. Mol. Biol.*, 118:137–208.

Saxena, A. K., Singh, T. P., Peters, K., Fittkau, S., and Betzel, C. (1996). Strategy to design peptide inhibitors: structure of a complex of proteinase K with a designed octapeptide inhibitor N-Ac-Pro-Ala-Pro-Phe-DAla-Ala-Ala-Ala-NH2 at 2.5 Å resolution. *Protein Sci.*, 5:2453–2458.

Schenk, H. (2008). Crystallography Across the Sciences 2. *Acta Crystallogr. Section A*, 64:preface.

Schiltz, M. and Bricogne, G. (2008). Exploiting the anisotropy of anomalous scattering boosts the phasing power of SAD and MAD experiments. *Acta Crystallogr. Section D*, 64:711–729.

Schiltz, M. and Bricogne, G. (2010). 'Broken symmetries' in macromolecular crystallography: phasing from unmerged data. *Acta Crystallogr. Section D*, 66:447–457.

Schneider, G. and Lindqvist, Y. (1994). Ta_6Br_{14} is a useful cluster compound for isomorphous replacement in protein crystallography. *Acta Crystallogr. Section D*, 50:186–191.

Schneider, T. R. and Sheldrick, G. M. (2002). Substructure solution with *SHELXD*. *Acta Crystallogr. Section D*, 58:1772–1779.

Schuermann, J. P. and Tanner, J. J. (2003). MRSAD: using anomalous dispersion from S atoms collected at Cu$K\alpha$ wavelength in molecular-replacement structure determination. *Acta Crystallogr. Section D*, 59:1731–1736.

Schultz, L. W., Quirk, D. J., and Raines, R. T. (1998). His...Asp catalytic dyad of ribonuclease A: structure and function of the wild-type, D121N, and D121A enzymes. *Biochemistry*, 37:8886–8898.

Sheldrick, G. M. (2002). Macromolecular phasing with *SHELXE*. *Z. Kristallogr.*, 217:644–650.

Sheldrick, G. M. (2008a). A short history of *SHELX*. *Acta Crystallogr. Section A*, 64:112–122.

Sheldrick, G. M. (2008b). TWINABS. *University of Göttingen, Germany.*

Sheldrick, G. M. (2009). SADABS. *University of Göttingen, Germany.*

Sheldrick, G. M. (2010). Experimental phasing with *SHELXC/D/E*: combining chain tracing with density modification. *Acta Crystallogr. Section D*, 66:479–485.

Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R., and Usón, I. (2001). *International Tables for Crystallography, Volume F: Crystallography of biological macromolecules*, chapter 16.1: *Ab initio* phasing, pages 247–255. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shiono, M. and Woolfson, M. M. (1992). Direct-space methods in phase extension and phase determination. I. Low-density elimination. *Acta Crystallogr. Section A*, 48:451–456.

Silvaggi, N. R., Martin, L. J., Schwalbe, H., Imperiali, B., and Allen, K. N. (2007). Double-lanthanide-binding tags for macromolecular crystallographic structure determination. *J. Am. Chem. Soc.*, 129:7114–7120.

Sippel, K. H., Robbins, A. H., Reutzel, R., Domsic, J., Boehlein, S. K., Govindasamy, L., Agbandje-McKenna, M., Rosser, C. J., and McKenna, R. (2008). Structure determination of the cancer-associated Mycoplasma hyorhinis protein Mh-p37. *Acta Crystallogr. D Biol. Crystallogr.*, 64:1172–1178.

Spek, A. L. (2009). Structure validation in chemical crystallography. *Acta Crystallogr. Section D*, 65:148–155.

Steiner, T. (2002). The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. Engl.*, 41:49–76.

Sugahara, M., Asada, Y., Ayama, H., Ukawa, H., Taka, H., and Kunishima, N. (2005). *Heavy-atom Database System*: a tool for the preparation of heavy-atom derivatives of protein crystals based on amino-acid sequence and crystallization conditions. *Acta Crystallogr. Section D*, 61:1302–1305.

Sun, P. D. and Radaev, S. (2002). Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures. *Acta Crystallogr. Section D*, 58:1099–1103.

Sun, P. D., Radaev, S., and Kattah, M. (2002). Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. *Acta Crystallogr. Section D*, 58:1092–1098.

Tajima, M., Urabe, I., Yutani, K., and Okada, H. (1976). Role of calcium ions in the thermostability of thermolysin and Bacillus subtilis var. amylosacchariticus neutral protease. *Eur. J. Biochem.*, 64:243–247.

Tavares, F. X., Boucheron, J. A., Dickerson, S. H., Griffin, R. J., Preugschat, F., Thomson, S. A., Wang, T. Y., and Zhou, H.-Q. (2004). *N*-Phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amines as potent and selective inhibitors of glycogen synthase kinase 3 with good cellular efficacy. *J. Med. Chem.*, 47:4716–4730.

Tereshko, V., Uysal, S., Koide, A., Margalef, K., Koide, S., and Kossiakoff, A. A. (2008). Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Sci.*, 17:1175–1187.

Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. Section D*, 56:965–972.

Terwilliger, T. C. and Berendzen, J. (1997). Bayesian Correlated MAD Phasing. *Acta Crystallogr. Section D*, 53:571–579.

Thygesen, J., Weinstein, S., Franceschi, F., and Yonath, A. (1996). The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. *Structure*, 4:513–518.

Tonnessen, L. E., Pedersen, B. F., and Klaveness, J. (1996). Molecular and Crystal Structure of Sodium Diatrizoate, an X-Ray Contrast Agent. *Acta Chem. Scand.*, 50:603–608.

Usón, I. and Sheldrick, G. M. (1999). Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.*, 9:643–648.

Usón, I., Stevenson, C. E. M., Lawson, D. M., and Sheldrick, G. M. (2007). Structure determination of the *O*-methyltransferase NovP using the 'free lunch algorithm' as implemented in *SHELXE*. *Acta Crystallogr. Section D*, 63:1069–1074.

van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des.*, 10:255–262.

Vocadlo, D. J., Davies, G. J., Laine, R., and Withers, S. G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature*, 412:835–838.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods Mol. Biol.*, 364:215–230.

Voth, A. R., Khuu, P., Oishi, K., and Ho, P. S. (2009). Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nat. Chem.*, 1:74–79.

Weiss, M. S. (2001). Global indicators of X-ray data quality. *J. Appl. Cryst.*, 34:130–135.

Weiss, M. S., Palm, G. J., and Hilgenfeld, R. (2000). Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD. *Acta Crystallogr. Section D*, 56:952–958.

Weiss, M. S., Panjikar, S., Nowak, E., and Tucker, P. A. (2002). Metal binding to porcine pancreatic elastase: calcium or not calcium. *Acta Crystallogr. Section D*, 58:1407–1412.

Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Section D*, 57:122–133.

Würtele, M., Hahn, M., Hilpert, K., and Höhne, W. (2000). Atomic resolution structure of native porcine pancreatic elastase at 1.1 Å. *Acta Crystallogr. Section D*, 56:520–523.

Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004). The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.*, 22:1297–1301.

Yano, J., Kern, J., Irrgang, K. D., Latimer, M. J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A., and Yachandra, V. K. (2005). X-ray damage to the Mn_4Ca complex in single crystals of photosystem II: a case study for metalloprotein crystallography. *Proc. Natl. Acad. Sci. U.S.A.*, 102:12047–12052.

Yogavel, M., Gill, J., Mishra, P. C., and Sharma, A. (2007). SAD phasing of a structure based on cocrystallized iodides using an in-house Cu$K\alpha$ X-ray source: effects of data redundancy and completeness on structure solution. *Acta Crystallogr. Section D*, 63:931–934.

Yu, S. B. and Watson, A. D. (1999). Metal-Based X-ray Contrast Media. *Chem. Rev.*, 99:2353–2378.

Zhou, P., Lv, J., Zou, J., Tian, F., and Shang, Z. (2010). Halogen-water-hydrogen bridges in biomolecules. *J. Struct. Biol.*, 169:172–182.

List of abbreviations

α	phase shift between heavy-atom phase and total phase
λ	wavelength
ϕ_A	heavy-atom phase
ϕ_T	total phase
ϕ_{h}	phase angle of reflection $\mathbf{h}$
θ	Bragg angle
μ	micro
B	Debye-Waller factor
CC	correlation coefficient
d	crystallographic resolution
d_{hkl}	spacing between a set of planes h, k, l
E	normalised structure factor
f	atomic scattering factor
f'	dispersive term of the atomic scattering factor
f''	anomalous term of the atomic scattering factor
f^B	atomic scattering factor in the crystal lattice
I_{h}	measured intensity of reflection $\mathbf{h}$
Å	Ångström ($1\,\text{Å} = 10^{-10}\,\mathrm{m}$)
$\mathbf{h}$	a set of planes with Miller indices h, k, l; the index vector of a reflection h, k, l at these planes
$\mathbf{x}_j$	position of atom j in the unit cell

M	molar
a, b, c, α, β, γ	unit cell parameters
a.u.	atomic units
AAS	anisotropy of anomalous scattering
ADA	*N*-(2-acetamido)iminodiacetic acid
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartate
B3C	5-amino-2,4,6-tribromoisophthalic acid
B3M	3-amino-2,4,6-tribromo-5-methoxybenzoic acid
B3O	2,4,6-tribromo-5-hydroxyisophthalic acid
B4C	tetrabromoterephthalic acid
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
CHES	2-(cyclohexylamino)ethanesulfonic acid
Cys	cysteine
Da	Dalton
DBI	dibromoisocyanuric acid
DMSO	dimethylsulfoxide
e	electron
EI	electron ionisation
ESI	electon-spray ionisation
eV	electron volt
EXAFS	extended X-ray absorption fine structure
F^+, F^-	structure factors of a Bijvoet/Friedel pair
F_A	structure factor of the heavy atoms

F_D	structure factor for a isomorphous derivative
F_P	structure factor of the protein
F_T	total scattering factor
F_h	complex structure factor for reflection **h**
FLA	free lunch algorithm
Gln	glutamine
Glu	glutamate
Gly	glycine
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	histidine
I3C	5-amino-2,4,6-triiodoisophthalic acid
Ile	isoleucine
Leu	leucine
Lys	lysine
m	milli
MAD	multi-wavelength anomalous dispersion
MES	2-(*N*-morpholino)ethanesulfonic acid
MIR	multiple isomorphous replacement
NBS	*N*-bromosuccinimide
NMR	nuclear magnetic resonance
PDB	Protein Data Bank
PEG	polyethylene glycol
PMF	Patterson minimum function
Pro	proline
PSMF	Patterson superposition minimum function
SAD	single-wavelength anomalous dispersion
Ser	serine

SIR	single isomorphous replacement
SIRAS	single isomorphous replacement plus anomalous scattering
SLS	Swiss Light Source
Thr	threonine
TLC	thin-layer chromatography
Tris	tris(hydroxymethyl-)aminomethan
Trp	tryptophane
Tyr	tyrosine
Val	valine
XANES	X-ray absorption near-edge fine structure

List of figures

List of tables

Scientific contributions

PUBLICATIONS

Sen, S.S., Kritzler-Kosch, M.P., Nagendra, S., Roesky, H.W., Beck, T., Pal, A., Herbst-Irmer, R. (2010). "Synthesis of Monomeric Divalent Tin(II) Compounds with Terminal Chloride, Amide, and Triflate Substituents." *Eur. J. Inorg. Chem.* 2010, 33:5304–5311.

Enders, D., Seppelt, M., Beck, T. (2010). "Enantioselective Organocatalytic Synthesis of Arylglycines via Friedel-Crafts Alkylation of Arenes with a Glyoxylate Imine." *Adv. Synth. Catal.* 2010, 352:1413–1418.

Beck, T., Gruene, T., Sheldrick, G.M. (2010). "The magic triangle goes MAD: experimental phasing with a bromine derivative." *Acta Crystallogr. Section D*, 66:374–380.

Tietze, L.F., Düfert, A., Lotz, F., Sölter, L., Oum, K., Lenzer, T., Beck, T., Herbst-Irmer, R. (2009). "Synthesis of Chiroptical Molecular Switches by Pd-Catalyzed Domino Reactions." *J. Am. Chem. Soc.*, 131:17879–17884.

Beck, T., da Cunha, C.E., Sheldrick, G.M. (2009). "How to get the magic triangle and the MAD triangle into your protein crystal." *Acta Crystallogr. Section F*, 65:1068–1070. *front cover*

Andrews, P.C., Beck, T., Fraser, B.H., Junk, P.C., Massi, M., Moubaraki, B., Murray, K.S., Silberstein, M. (2009). "Functionalised β-diketonate polynuclear lanthanoid hydroxo clusters: Synthesis, characterisation and magnetic properties." *Polyhedron*, 28:2123–2130.

Justus, K., Beck, T., Noltemeyer, M., Fitjer, L. (2009). "Percycloalkylated cyclohexanes: attempted synthesis of a trispropellane." *Tetrahedron*, 65:5192–5198.

Beck, T., Herbst-Irmer, R., Sheldrick, G.M. (2009). "5-Amino-2,4,6-tribromoisophthalic acid: the MAD triangle for experimental phasing." *Acta Crystallogr. Section C*, 65:o237–o239.

Jana, A., Roesky, H.W., Schulzke, C., Döring, A., Beck, T., Pal, A., Herbst-Irmer, R. (2009). "Facile Access of Stable Divalent Tin Compounds with Terminal Methyl, Amide, Fluoride, and Iodide Substituents." *Inorg. Chem.*, 48:193–197.

Beck, T., Krasauskas, A., Gruene, T., Sheldrick, G.M. (2008). "A magic triangle for experimental phasing of macromolecules." *Acta Crystallogr. Section D*, 64:1179–1182. *Highlighted article in the IUCr Newsletter 2008, Vol. 16, No 4.*

Beck, T., Sheldrick, G.M. (2008). "5-Amino-2,4,6-triiodoisophthalic acid monohydrate." *Acta Crystallogr. Section E*, 64:o1286.

Jiménez-Pérez, V.M., Muñoz-Flores, B.M., Roesky, H.W., Schulz, T., Pal, A., Beck, T., Yang, Z., Stalke, D., Santillan, R., Witt, M (2008). "Monomeric Boron and Tin(II) Heterocyclic Derivatives of 1,8-Diaminonaphthalenes: Synthesis, Characterization and X-ray Structures." *Eur. J. Inorg. Chem.*, 2238–2243.

Andrews, P.C., Beck, T., Forsyth, C.M., Fraser, B.H., Junk, P.C., Massi, M., Roesky, P.W. (2007). "Templated assembly of a μ_6-CO_3^{2-} dodecanuclear lanthanum dibenzoylmethanide hydroxo cluster with concomitant formation of phenylglyoxylate." *Dalton Transactions*, 5651–5654. *front cover*

Andrews, P.C., Beck, T., Fraser, B.H., Junk, P.C., Massi, M. (2007). "Cationic, neutral and hydroxo bridged lanthanoid (Y, La, Gd, Ho, Yb) bis 5-(2-pyridyl)tetrazolate complexes." *Polyhedron*, 26:5406–5413.

CONFERENCE CONTRIBUTIONS - ORAL PRESENTATIONS

Talk: "Structure determination of biological macromolecules with sticky triangles." *12th JCF Spring Symposium*, Göttingen, Germany 2010. *prize for best oral presentation.*

Talk: "Experimental phasing of macromolecules with sticky triangles." *International Research Training Group 1422 Biometals Workshop*, Goslar, Germany 2010.

Invited talk: "Neue Werkzeuge für die experimentelle Phasenbestimmung von Biomolekülen: 'Sticky Triangles'." *GDCh Weihnachtskolloquium*, Göttingen, Germany 2009.

Invited tutorial: "Sample handling and soaking with B3C." *Training in Methods for Macromolecular Crystallography: M2M-9 Practical Course*, EMBL Hamburg, Germany 2009.

Invited talk: "Phasing with triangles: new compounds for heavy-atom derivatization of macromolecules." *Winter School on Soft X-rays in Macromolecular Crystallography*, Berlin, Germany 2009.

Talk: "Triangles for experimental phasing of macromolecules." *International Research Training Group 1422 Biometals Workshop*, Copenhagen, Denmark 2009.

Invited talk: "The magic triangle: a new phasing tool." *CCP4 Study Weekend*, Nottingham, UK 2009.

Talk: "A magic triangle for experimental phasing of macromolecules." *9th International School on the Crystallography of Biological Macromolecules*, Como, Italy 2008.

Talk: "Towards a better understanding of the formation of lanthanoid hydroxo clusters with chelating ligands." *Terrae Rarae*, Bonn, Germany 2007.

Talk: "Lanthanide coordination compounds with chelating tetrazolates." *Inorganic Symposium RACI*, Melbourne, Australia 2006.

Poster: "Sticky triangles for experimental phasing." *European Crystallographic Meeting (ECM 25)*, Istanbul, Turkey 2009. *poster prize "outstanding young scientist"*.

Poster: "Experimental phasing with triangles." *Annual Meeting of the German Society for Crystallography (DGK)*, Hannover, Germany 2009. *poster prize*.

Poster: "Ein magisches Dreieck? Experimentelle Phasenbestimmung mit einem Triiodobenzolderivat." *2nd Göttingen Chemistry Forum*, Göttingen, Germany 2008. *poster prize*.

Poster: "A magic triangle? Experimental phasing of macromolecules with a triiodo benzene derivative." *Annual Meeting of the American Crystallographic Association (ACA)*, Knoxville, USA 2008. *Pauling poster prize*.

Poster: "A magic triangle? Experimental phasing of macromolecules with a triiodo benzene derivative." *Annual Meeting of the German Society for Crystallography (DGK)*, Erlangen, Germany 2008. *poster prize*.

Poster: "A new approach towards Rieske protein analogues." *International Research Training Group 1422 Biometals Workshop*, Lund, Sweden 2007.

Talk and Poster: "Kinetics of hydrogen bonded complexes: reaction of *N,N*-dimethylformamide with OH radicals at low temperature." *Seventh Workshop of the Physical and Chemical Graduate School*, Göttingen, Germany 2006.

Acknowledgments

I would like to thank Prof. George M. Sheldrick for the supervision of this thesis. He provided me with many valuable ideas and also gave me the freedom to try my own ones. The discussions with him were very important for the success of this project. I am very grateful to be a member of his group. I would like to thank Prof. Dr. Oliver Einsle for his advice during the thesis committee meetings and for accepting to review this thesis.

I am very thankful to Dr. Tim Grüne for his support. I have enjoyed the numerous discussions about all aspects of crystallography. I would like to thank him for his suggestions and comments, especially during the compilation of this thesis. I also appreciate his advice in software-related issues and I am grateful for the maintenance of the computer facilities. I would like to thank Dr. Regine Herbst-Irmer for advice regarding small-molecule crystallography and twinning. I am very thankful for her comments on my refined small-molecule structures.

I would also like to thank the other members of the Sheldrick group in Göttingen for interesting discussions: Kathrin Meindl, Christian Große, Andrea Thorn, Kevin Pröpper and Navdeep Sidhu, especially Kathrin for carrying out the Gaussian calculations, Christian for discussions about small-molecule crystallography and Andrea for introductory advice regarding *PYTHON*. I would like to acknowledge the three research students who worked on this project during their laboratory rotation: Andrius Krasauskas, Carlos da Cunha and Kevser Gencalp. I would like to extend my thanks to the former members of the Sheldrick group: Dr. Ina Dix, Dr. Burkhard Heisen, Dr. Stephan Rühl, Dr. Aritra Pal and Dr. Roland Pfoh. I would like to thank the members of the Dittrich group Julian Holstein, Dr. Christian Hübschle and Dr. Birger Dittrich for discussions, especially Christian for last minute changes in *MOLECOOLQT* and personal software support. I would also like to thank Dr. Julian Henn and Prof. Dr. Isabel Usón for helpful discussions. I am very thankful to Helmut Dehnhardt for this technical assistance and the maintenance of the diffractometers. I would like to thank Dalila Griffin for taking care of administrative issues. I would like to thank Tim, Kathrin, Regine, Birger and Isabel for proof reading and valuable discussions during the compilation of this thesis.

I would like to thank the collaborators for providing data and helpful suggestions. Katherine Sippel and Dr. Robert McKenna from the University of Florida, Gainesville, USA, used I3C with their protein, after having seen our poster at the ACA 2008 in Knoxville, USA. It was a great idea of them to publish our methods article back-to-back with their structure, the first novel protein solved with I3C, in the same issue of Acta Crystallogr. Section D. I would like to thank Dr. Christoph Parthier, Ulrike Bräuer and Prof. Dr. Milton Stubbs from the Martin-Luther-Universität in Halle-Wittenberg, Ger-

many, for kindly providing data of an I3C derivatised crystal. I am also thankful for data provided by Jeanine Amacher and Prof. Dr. Dean Madden from Dartmouth Medical School, Hanover, USA. Dr. Scott Lovell from the University of Florida, Gainesville, USA, also kindly provided data from a novel unpublished structure solved with I3C.

I would like to thank Dr. Paul Emsley and Dr. Bernhard Lohkamp for help with scripting in *COOT* and the implementation of the *COOT* feature in the *TRIFIND* script. I am thankful to Prof. Dr. Derek Logan, Renzo Johansson and Dr. Maria Håkansson for their support during my stay in Lund and the experiments at Max-lab. I would like to thank Dr. Vincent Oliéric, SLS, Villigen, Switzerland, and Dr. Manfred Weiss, BESSY, Berlin for support during data collection.

I would like to acknowledge all employees at the institute, including NMR and MS department and the analytical laboratory, especially Matthias Hesse and Martin Schlote for distribution of chemicals and laboratory equipment. I would like to thank Dr. A. Claudia Stückl for her support in administrative issues.

For advice on synthetical questions I would like to thank Karsten Meyenberg, Alexander Düfert and André Nadler, Georg-August-Universität Göttingen and Matthias Seppelt, RWTH Aachen. I would like to acknowledge Peer Lukat, Albert-Ludwigs-Universität Freiburg, for advice with *PYMOL*. I would like to thank all the present and former members of the International Research Training Group 1422 for the interesting seminars and workshops, especially Friederike Fehr and Anja Pomowski for co-organisation of several seminars and workshops. I would like to acknowledge Dr. Jonathan MacLellan for introduction to small-molecule crystallography. I would also like to acknowledge Dr. Manfred Speldrich for introducing me to LaTeX many years ago.

I would like to acknowledge funding from the German National Academic Foundation (Studienstiftung des deutschen Volkes) for a student and a PhD scholarship, the International Centre for Diffraction Data (ICDD) for the Ludo Frevel Scholarship Award (2008 and 2009) and the DFG within the International Research Training Group IRTG 1422 "Metal Sites in Biomolecules: Structures, Regulation and Mechanisms".

Finally, I would like to extend warm thanks to my family for support throughout my studies in Aachen, Melbourne and Göttingen. I truly thank Anne for her loving support during the last years and especially at the final stages of the thesis: Thank you!

Curriculum Vitae

PERSONAL

Name:	Tobias Beck
Date of Birth:	September 6th 1981
Place of Birth:	Hannover
Nationality:	German
Marital Status:	married

EDUCATION

1988 – 1992	Gemeinschaftsgrundschule Roetgen
1992 – 1998	Viktoriaschule Aachen, Gymnasium der evangelischen Kirche im Rheinland
1998 – 1999	Le Mars Community High School, Iowa, USA
1999 – 2001	Viktoriaschule Aachen
05/2001	Allgemeine Hochschulreife
2001 – 2002	Civilian service, Chemisches Landes- und Staatliches Veterinäruntersuchungsamt, Münster

ACADEMIC CAREER

10/2002 – 09/2004	chemistry student, RWTH Aachen
09/2004	Diplomvorprüfung “sehr gut”
10/2004 – 06/2007	chemistry student, Georg-August-Universität Göttingen
09/2006 – 02/2007	Diploma thesis at Monash University, Melbourne, Australia “Lanthanoid hydroxo clusters: Towards a new class of contrast media for X-ray imaging ”, advisors: Peter Junk, Phil Andrews and George M. Sheldrick
06/2007	Diplomprüfung “mit Auszeichnung”
06/2007 – present	PhD student, Georg-August-Universität Göttingen “New tools for experimental phasing of macromolecules” advisor: George M. Sheldrick

HONOURS AND AWARDS

2002 – 2007	fellowship of the German National Academic Foundation
2005	Chemistry Award for outstanding performance in Diplomvorprüfung, RWTH Aachen
2007	Gustav Tammann Award for the best chemistry diploma thesis, Georg-August-Universität Göttingen
2008	Ludo Frevel Scholarship Award, International Centre for Diffraction Data (ICDD), USA
2008 – present	PhD fellowship of the German National Academic Foundation
2008	Pauling poster prize American Crystallographic Association Meeting, Knoxville, USA
2009	Ludo Frevel Scholarship Award, International Centre for Diffraction Data (ICDD), USA
2009	poster prize "outstanding young scientist" European Crystallographic Meeting, Istanbul