X-RAY CRYSTALLOGRAPHIC STUDIES ON TWO CYSTEINE-RICH ANTICARCINOGENIC MICROPROTEINS

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August Universität zu Göttingen

vorgelegt von

Judit Éva Debreczeni

aus Budapest, Ungarn

Göttingen, 2004

D 7

ACKNOWLEDGEMENTS

ABBREVIATIONS

Ala	alanine
ANKA	Angströmquelle Karlsruhe
Arg	arginine
Asn	asparagine
Asp	aspartate
Asx	asparagine or aspartate
BBi	Bowman-Birk inhibitor
CC	correlation coefficient
CHM	chymotrypsin
Cys	cysteine
DESY	Deutsches Elektronen Synchrotron
EMBL	European Molecular Biology Laboratory
FFT	fast Fourier transformation
fom	figure of merit
Gln	glutamine
Glu	glutamate
Glx	glutamine or glutamate
Gly	glycine
His	histidine
HPLC	high performance (pressure/prise) liquid chromatography
HTH	helix-turn-helix
Ile	isoleucine
LBTI	lima bean trypsin-chymotrypsin inhibitor
Leu	leucine
Lys	lysine
MAD	multiple anomalous dispersion
Met	methionine
MR	molecular replacement
NCS	non-crystallographic symmetry
PDB	Protein Data Bank
PEG	polyethylene glycol
Phe	phenylalanine
Pro	proline
PS	phosphatidyl-serine
rmsd	root mean square deviation

SAD	single wavelength anomalous dispersion
SDS PAGE	sodium-dodecylsulfate polyacrylamide gel electrophoresis
Ser	serine
SIR	single isomorphous replacement
sp (spp)	species (Pl.)
ssp	subspecies
Thr	threonine
TRP	trypsin
Trp	tryptophan
Tyr	tyrosine
Val	valine
VT	viscotoxin

MOLECULAR GRAPHICS

All three-dimensional molecular drawings were generated using *MOLSCRIPT* (Kraulis, 1991) or *BOBSCRIPT* (Esnouf, 1997) and rendered with *RASTER3D* (Merritt & Bacon, 1997) unless otherwise indicated.

Figure 2.31 was drawn and rendered with *PYMOL* (DeLano, 2003), while *Figures 2.20, 2.22, 2.30, 3.14* and *3.21* show molecules as displayed by *XtalView* (McRee, 1999). *Figure 2.6* was drawn by hand.

TABLE OF CONTENTS

OVERVIEW

In order to overcome the constant microbial attacks and possible injuries, plants have developed a large number of defence mechanisms. For instance, they produce several antimicrobial substances triggered by infection. Also, their tissues contain a reservoir of toxic compounds as prevention against pathogenic threats. For the pharmaceutical industry, always in search of new potential drugs, compounds of the constitutive plant defence can be of considerable interest because of their usually strong biological activity and their natural abundance.

Both protein classes presented in this thesis were shown to play a role as defence chemicals: the lima bean trypsin-chymotrypsin inhibitor (LBTI) blocks serine proteinases, while viscotoxins (VTs) are able to disrupt cellular membranes. In spite of their fundamentally different way of action, their *in vivo* roles serve as a basis for their joint discussion in the scope of one thesis: both LBTI and the VTs were shown to exhibit anticarcinogenic properties having, however, different action targets on tumour cells. In addition, they occupy similar positions in sciences related to biochemistry, as their literature reflects about 70 years of intensive research on their isolation and characterisation. Interestingly, the biologically most exciting publications in these fields can be regarded as milestones of method development as well: the LBTI was investigated as one of the prototypes of the Bowman-Birk inhibitor family and became later one of the most widely used proteinase inhibitors in protein purification procedures. The VTs and their close relatives, the thionins, provided many excellent subjects for improving NMR and X-ray techniques.

From the structural point of view, one can assume that the common main architecture of these proteins can account for their other similarities. Both of them are of low molecular weight and made rigid and insensitive to environmental impacts by an unusually large number of disulfide bridges. This thesis presents the structure determination and analysis of LBTI, its ternary complex with both of its target enzymes and two viscotoxin isoforms. The high symmetry and solvent content of the uncomplexed LBTI crystals together with the unusually high ratio of cysteines enabled structure solution via the anomalous scattering of sulfurs with data collected on a rotating anode, giving the first example for low resolution in-house sulfur-

SAD phasing. Similarly to LBTI, the first VT-structure was also determined exploiting the weak anomalous dispersion of sulfurs, in this case the failure of classical heavy atom derivatisation procedures incidental to the tight crystal packing and low solvent content clearly demanded the application of this up-to-date method and lead to improvements in heavy atom location.

Apart from the methodological point of interest, the analysis of these structures also shed light on some details of their biological role, showing that structural studies of such small proteinaceous inhibitors combined with method development can be of scientific relevance even in the post-ribosomal era.

I THEORETICAL BACKGROUND

I.1 Crystallisation of proteins

The crystallisation process of proteins and also of low molecular-weight organic compounds can be divided into two major subsequent phases:

1. nucleation: formation of the first ordered aggregates,

2. crystal growth: attachment of new particles to the surface of pre-formed aggregates.

To initiate both, nucleation and crystal growth, a solution of the material to be crystallised is brought to supersaturation. The rate of these processes depends on the degree of supersaturation, as demonstrated in *Figure 1.1*. Since crystallisation experiments aim at the formation of a few but large crystals, a low supersaturation level is desirable, where the nucleation/growth ratio is small and promotes the formation of a small number of nuclei that can grow with a slower rate and reach a maximum degree of order. In the case of proteins, the classical "salting-out"-method with light metal salts or polyethylene-glycols (PEGs) that serve to immobilise water can be regarded as the mildest way of increasing the effective concentration of the protein.
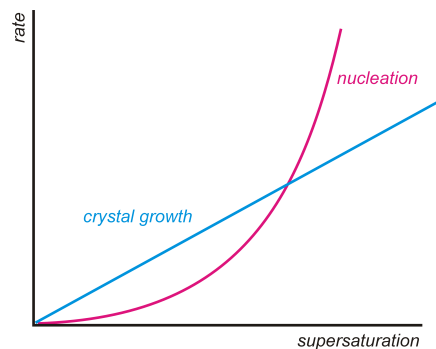
Figure 1.1. The rate of nucleation and crystal growth as function of the supersaturation level.

Although theoretical differences between protein and small molecule crystallisation are only marginal, the tendency of proteins to precipitate, the high solvent content of macromolecular crystals and the usual shortage of protein samples require different crystallisation techniques. However, the number of parameters that exert influence on crystal growth and protein stability is much larger than those of small molecules. These include ─ beyond the supersaturation level ─ temperature, ionic strength and pH of the protein solution,

but other physical factors such as vibration and magnetic field can also contribute to the success of a crystallisation experiment. This complexity nearly excludes the possibility of systematic investigation of all possible crystal growth parameters and as a practical consequence, sparse matrix screens were developed for wide-ranging trial-and-error type search of suitable crystallisation conditions.

In addition, nucleation and crystal growth is affected by the crystallisation technique employed. All methods that are widely used nowadays can be classified as micromethods, as they allow application of small volumes of protein solution and therefore a great number of experiments.

Among these techniques probably the vapour diffusion methods are the most widespread. The principle of different crystallisation techniques based on vapour diffusion is demonstrated in *Figure 1.2*. A droplet containing the protein solution to crystallise and crystallisation agents (buffer, precipitant, additives, etc) is equilibrated against a reservoir containing a solution of the same crystallisation agents at higher concentrations than in the droplet. The crystallisation vessel is sealed hermetically, so volatile species (mostly water and organic solvents) can diffuse only from the droplet into the reservoir until the vapour pressures of the droplet and the reservoir balance out.
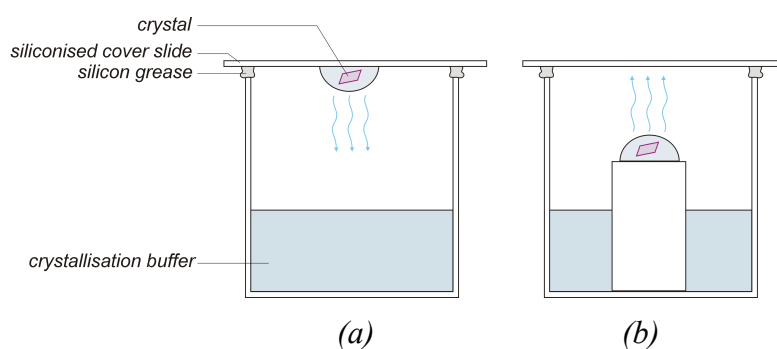
Figure 1.2. Crystallisation with *(a)* the hanging drop *(b)* sitting drop technique.

As a consequence, the concentration of all constituents of the droplet increases and the resulting supersaturation can lead to crystallisation of the protein (*Figure 1.3.a*). The droplets are usually placed onto siliconised cover slides (hanging drop technique) or microbridges (sitting drop technique) in order to avoid the spread of the droplet due to the lower surface tension of solutions containing organic solvents.

Microbatch methods represent a simpler technique for crystallisation and allow the set-up of droplets with volumes in the submicroliter range (*Figure 1.4.a*). In this case the precipitation agent is added to the protein solution instantaneously in such a way that the
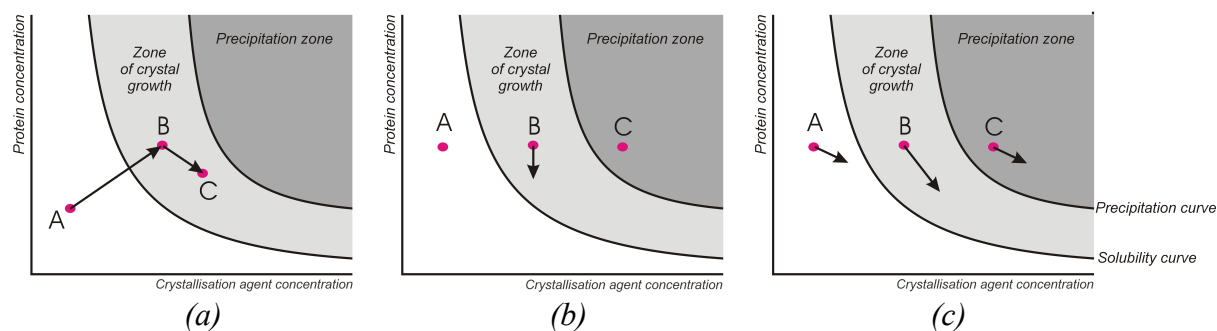
Figure 1.3. Changes in the protein solubility during crystallisation: *(a)* vapour diffusion (**A**: undersaturated state, **B**: supersaturation without crystal growth, **C**: supersaturation after crystal growth) *(b)* microbatch *(c)* microbatch using a paraffin oil-silicon oil mixture (**A**: undersaturated state, **B**: supersaturation with crystal growth, **C**: precipitation).

supersaturated state can be immediately reached (point B in *Figure 1.3.b*). Since the droplet is covered by an oil layer, evaporation of the solvent and thus further concentration of the protein and precipitant are strongly hindered. In less favourable cases the protein may stay in an undersaturated state (point B) or precipitate immediately (point C). Application of silicon oil or silicon oil - paraffin oil mixtures allow the gradual evaporation of water that renders concentration changes possible (*Figures 1.4.b* and *1.3.c*).
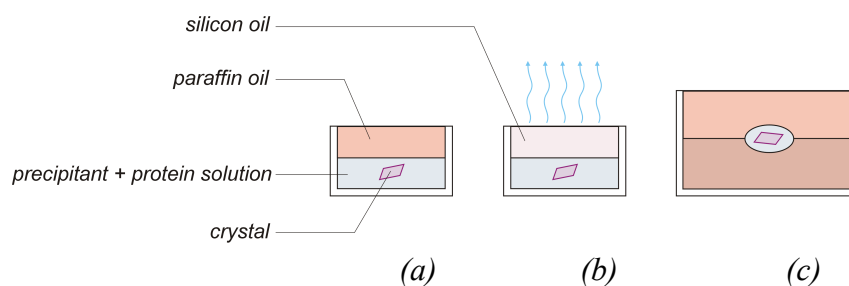
Figure 1.4. Crystallisation under oil: *(a)* microbatch *(b)* microbatch using silicon oil *(c)* "Containerless" crystallisation between two oil layers.

I.2 Methods to solve the phase problem

X-rays are scattered by the electron cloud of atoms constituting the molecule. Since the interaction of X-rays with the scattering material is weak, several molecules are needed to achieve detectable diffraction. In the case of crystalline materials, where a large number of molecules is arranged in an identical orientation, the internal ordering leads to a regular diffraction pattern dependent on the crystal lattice and the contents of the unit cell.

Constructive interference of scattered X-ray photons result in the appearance of reflections in certain directions determined by the Laue conditions of diffraction or the Bragg

equation. A reflection — identified by its Miller indices h, k and l — can be characterised by the structure factor ($\mathbf{F}_{hkl}$) that can be calculated from the electron density distribution (ρ) of the unit cell of the crystal with a Fourier transformation:

$$\mathbf{F}_{hkl} = \int\limits_{cell} \rho(xyz)\exp[2\pi i(hx + ky + lz)]dV \tag{1}$$

where x, y and z represent fractional coordinates. However, the aim of crystallographic studies is the calculation of the electron density distribution — that carries the information about the atom positions — and not the prediction of the diffraction pattern. However, the Fourier transform operation is reversible and therefore the electron density can be calculated as a Fourier transform of the structure factors, too:

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}\mathbf{F}_{hkl}\exp[-2\pi i(hx + ky + lz)] \tag{2}$$

Since $\mathbf{F}_{hkl}$ is a complex function in a general case, it can be written in the following form:

$$\mathbf{F}_{hkl} = |\mathbf{F}_{hkl}|\exp(i\alpha_{hkl}) \tag{3}$$

where $|\mathbf{F}_{hkl}|$ is the modulus of the structure factor, while α_{hkl} is its phase angle. From the two parameters required for electron density calculation only the amplitude $|\mathbf{F}_{hkl}|$ can be derived from the measured reflection intensities (I_{hkl}):

$$|\mathbf{F}_{hkl}|^2 \sim I_{hkl} \tag{4}$$

Unfortunately, it is almost impossible to determine the phase angle α_{hkl} experimentally. The lack of experimental information about the phases is known as the crystallographic phase problem and represents the main obstacle in macromolecular structure solution. The following subsections discuss those techniques of overcoming the phase problem that were successfully applied for the structure solution of protein molecules presented in this thesis.

I.2.1 Direct methods in macromolecular crystallography

Direct methods are statistical techniques that estimate phases from measured intensities. Structure solution via direct methods requires atomic resolution data, since the assumptions used to derive the formulae that constitute the basis of direct methods are good approximations only if the electron density distribution displays a certain degree of atomicity.

In addition, atomic resolution ensures high data-to-parameter ratio and thus the phase problem is overdetermined.

Direct methods employ normalised structure factors:

$$E_{hkl}{}^2 = \frac{\dfrac{F_{hkl}{}^2}{\varepsilon}}{\left\langle \dfrac{F_{hkl}{}^2}{\varepsilon} \right\rangle} \tag{5}$$

where $<>$ is an average over a resolution shell and ε is a statistical factor. Normalised structure factors virtually eliminate the effect of atomic thermal motion and the angle dependence of the atomic scattering factor arising from the finite size of the electron cloud, and correspond to structure factors of a point atom structure.

Direct methods assume that

1. the electron density is never negative, and
2. the structure consists of randomly distributed resolved atoms.

These assumptions lead to probability relations between phases, among which the triple phase relations and negative quartets are the most important phase relations in direct methods. The triple phase relation can be formulated as

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3} \cong 0 \tag{6}$$

when $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$ (where $\mathbf{h}_1$ means (h_1, k_1, l_1)) and $E_1>>1$, $E_2>>1$ and $E_3>>1$. Combination of many triple phase relations gives the tangent formula (Karle & Hauptman, 1956):

$$\tan \varphi_{\mathbf{h}} = \frac{\sum_{\mathbf{h'}} \left|E_{\mathbf{h'}} E_{\mathbf{h}-\mathbf{h'}}\right| \sin\left(\varphi_{\mathbf{h'}} + \varphi_{\mathbf{h}-\mathbf{h'}}\right)}{\sum_{\mathbf{h'}} \left|E_{\mathbf{h'}} E_{\mathbf{h}-\mathbf{h'}}\right| \cos\left(\varphi_{\mathbf{h'}} + \varphi_{\mathbf{h}-\mathbf{h'}}\right)} \tag{7}$$

that (or its modified form) is used by direct methods programs iteratively to refine phases. The negative quartets represent weaker phase relations, but have the advantage that they can provide useful information, as they can discriminate against the uranium atom solution. They can be written as

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3} + \varphi_{\mathbf{h}_4} \cong \pi \tag{8}$$

if $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3+\mathbf{h}_4=0$ and $E_{\mathbf{h}_1}, E_{\mathbf{h}_2}, E_{\mathbf{h}_3}, E_{\mathbf{h}_4} >> 1$ and $E_{\mathbf{h}_1+\mathbf{h}_2}, E_{\mathbf{h}_1+\mathbf{h}_3}, E_{\mathbf{h}_1+\mathbf{h}_4} << 1$.

In classical direct methods procedures, first random phases are assigned to reflections with the highest E values. These phases are then refined using the tangent formula and after convergence figures of merit are calculated. The best phase set is used for calculating an electron density map (with or without phase extension to weaker reflections), which is then peak searched. Larger structures where the number of atoms exceeds about 200 can rarely be solved using classical direct methods because the probability distribution of phase relations becomes weaker with increasing the number of atoms (Cochran, 1955). On the other hand, crystals of macromolecules usually do not diffract to atomic resolution.

A breakthrough was achieved in this field with the introduction of the shake-and-bake technique, also known as dual-space recycling implemented in the programs *SnB* and *SHELXD* (Sheldrick *et al.*, 2001).

The dual-space procedure, as illustrated in *Figure 1.5*, starts each trial with a set of random atoms. The arrangement of the atoms can, however, be constrained in various ways that may improve the success rate. Such constraining parameters are the number and type of starting atoms or restrictions on the interatomic distances. The starting atoms can also be set to be consistent with the Patterson function that is obtained as a Fourier transformation with the squares of structure factor amplitudes as coefficients without phase angles. Phases are then calculated from the starting atoms and improved e.g. by the tangent formula ("shaking").
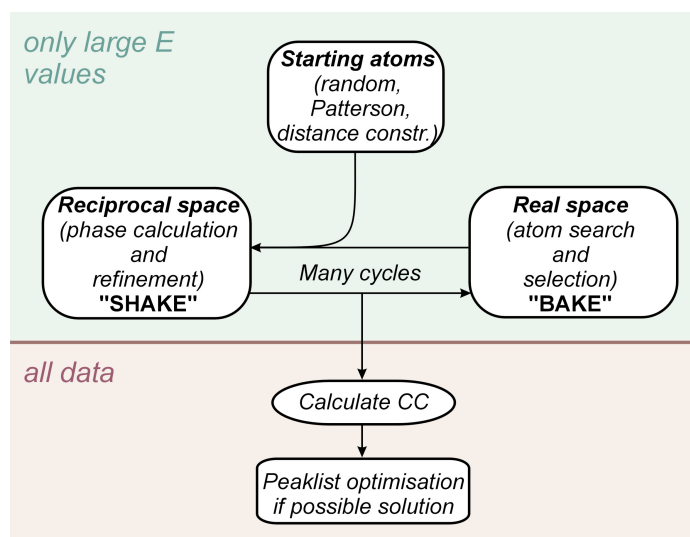
Figure 1.5. The general scheme of dual-space recycling.

With these improved phases and E values derived from the observed intensities an electron density map is calculated. In this map peaks are searched and the strongest peaks are kept as atoms according to the expected number of atoms. In *SHELXD*, a given percentage of

atoms found is deleted at random to increase the search efficiency and a new cycle is started with phases calculated from atoms retained. The real space half of the procedure is also called "baking".

By the improvement achieved through recycling the atoms (found and reasonably selected in real space) to reciprocal space where phases are refined, not only the size of structures solved by direct methods could be expanded. These methods not only allow structure solution up to 1000-2000 atoms in the asymmetric unit but also prove very powerful in substructure solution from e.g. anomalous data.

For the latter case, *SHELXD* provides the option of occupancy refinement for atoms obtained in the last dual space cycle. This procedure turns out to be especially helpful in the case of derivatives prepared by soaking where fractional occupancies are expected, but also enhances the discrimination between real heavy atoms and noise peaks.

The quality of a solution is estimated by the correlation coefficient calculated between the observed and calculated E values (E_o and E_c) (Fujinaga & Read, 1987):

$$\mathrm{CC} = 100 \frac{\sum wE_o E_c \sum w - \sum wE_o \sum wE_c}{\left[\left(\sum wE_o^2 \sum w - \left(\sum wE_o\right)^2\right)\left(\sum wE_c^2 \sum w - \left(\sum wE_c\right)^2\right)\right]^{1/2}} \quad (9)$$

For *ab initio* applications, the solution can be further improved with the peak list optimisation technique that optimises the set of atoms in such a way that the CC increases. This method is not used for improving the heavy atom model because of the unreliability of weak E values.

I.2.2 Single wavelength anomalous scattering

Theoretical background

The scattering of X-rays by electrons of an anomalous scatterer differs from a non-anomalous one and can be described by writing the scattering factor of the anomalous scatterer as follows:

$$f_{\mathrm{total}}(\theta, \lambda) = f(\theta) + f'(\lambda) + if''(\lambda) \quad (10)$$

where f represents the scattering factor for normal elastic scattering that can be described by a real number. Both f', the real part and if'', the imaginary part of the anomalous contribution depend on the wavelength used in the experiment for a certain atom type, unlike the non-anomalous scattering factor f that depends only on the scattering angle. f'' (known also as

"anomalous signal") is proportional to the atomic absorption of X-rays and to their fluorescence, whereas f' ("dispersive signal") is a result of X-ray absorption and immediate reemission and follows the derivative of the f'' function (or more precisely: f' and f'' are related by the Kramers-Kronig transformation). As also demonstrated in *Figure 1.6*, the time delayed photon reemission causes a phase shift of 90°, therefore if'' affects not only the amplitude but the phase as well.
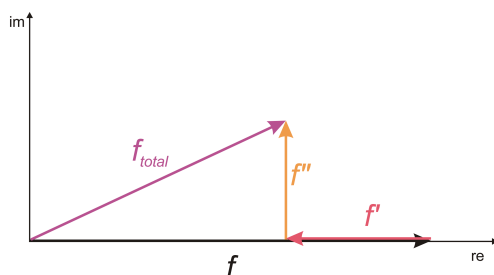
Figure 1.6. Vector representation of the scattering factor of an anomalous scatterer.

The anomalous effect is usually small but increases with increasing wavelength (or decreasing X-ray energy that is inversely proportional to the wavelength) up to the point that corresponds to the absorption edge of the particular element where it diminishes abruptly (*Figure 1.7*). In practice, the imaginary anomalous contribution is normally determined in X-ray fluorescence measurements as a function of the wavelength, whereas f' is calculated as its derivative. The anomalous effect originating from light atoms is usually negligibly small, but careful data acquisition can allow phasing of proteins on the anomalous differences of sulfur or the determination of the absolute structure of organic small molecules exploiting the anomalous signal of oxygen.
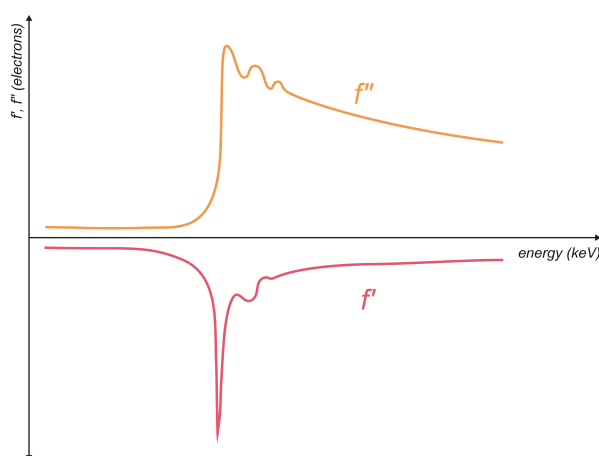
Figure 1.7. f' and f'' as a function of X-ray energy.

As also demonstrated in *Figure 1.8*, the total non-anomalous scattering factor $\mathbf{F}_T$ can be thought of as a sum of $\mathbf{F}_p$, the structure factor of the protein (i.e. of all non-anomalous scatterers) and $\mathbf{F}_A$, the structure factor of the anomalous scatterers (heavy atoms) ignoring their anomalous effects. $\mathbf{F}_A'$ represents the dispersive signal and is applied in the opposite direction to $\mathbf{F}_A$, since f' is usually negative. The imaginary if'' contribution, as mentioned above, corresponds to a 90° phase shift and therefore $\mathbf{F}_A''$ is drawn at 90° anticlockwise to give $\mathbf{F}^+$ and 90° clockwise for $\mathbf{F}^-$ (or more precisely: for its complex conjugate). The vector $\mathbf{a}$ represents the anomalous contribution of the heavy atoms:

$$\frac{f'}{f}\mathbf{F}_A + \frac{if''}{f}\mathbf{F}_A = \mathbf{F}_A' + \mathbf{F}_A'' = \mathbf{a} \tag{11}$$

As a consequence of the anomalous scattering, the structure factors of Friedel mates $\mathbf{F}^+$ and $\mathbf{F}^-$ are no longer equal in length and $\varphi^+ \neq -\varphi^-$.
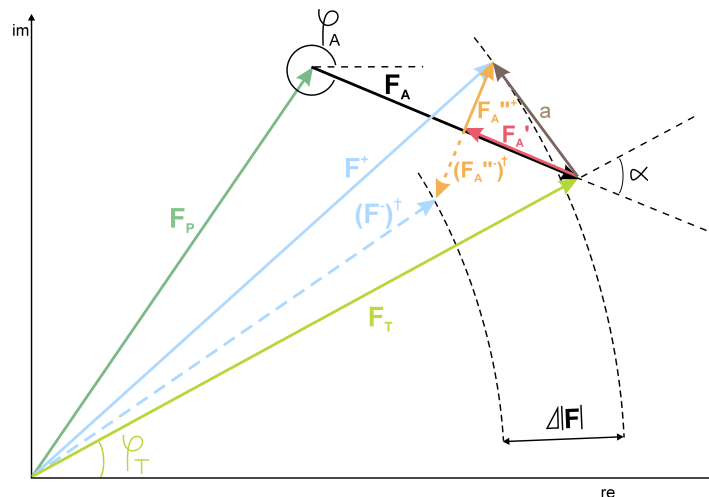
Figure 1.8. Argand diagram representing the presence of anomalous scatterer atoms. $\mathbf{F}_T$ (the total non-anomalous scattering) is constructed as the sum of $\mathbf{F}_P$ (the structure factor of the protein) and $\mathbf{F}_A$ (the non-anomalous contribution of the anomalous scatterer atoms). The measurable intensities are determined by F^+ and F^- that can be derived by the addition of the anomalous components: $\mathbf{F}_A' = \frac{f'}{f}\mathbf{F}_A$ and $\mathbf{F}_A'' = \frac{if''}{f}\mathbf{F}_A$ to $\mathbf{F}_T$.

From the construction shown in *Figure 1.8* the following relations can be drawn. Applying the cosine theorem to the triangle determined by $\mathbf{F}_T$, $\mathbf{F}^+$ and $\mathbf{a}$,

$$\left|\mathbf{F}^+\right|^2 = \left|\mathbf{F}_T\right|^2 + \left|\mathbf{a}\right|^2 - 2\left|\mathbf{F}_T\right|\left|\mathbf{a}\right|\cos\left(\varphi_T + 180 - \varphi_a\right) \tag{12}$$

can be derived (where φ_a is the phase of **a**) and by expressing the length of **a** with $\left|\mathbf{F}_A\right|$, the length of $\mathbf{F}^{\pm}$, i.e. the measured intensities can be written in the general form (Hendrickson *et al.*, 1985):

$$\left|\mathbf{F}^{\pm}\right|^2 = \left|\mathbf{F}_T\right|^2 + \frac{f'^2 + f''^2}{f^2}\left|\mathbf{F}_A\right|^2 + \frac{2f'}{f}\left|\mathbf{F}_T\right|\left|\mathbf{F}_A\right|\cos\alpha \pm \frac{2f''}{f}\left|\mathbf{F}_T\right|\left|\mathbf{F}_A\right|\sin\alpha \qquad (13)$$

where $\alpha = \varphi_T - \varphi_A$. This equation gives the basis of SAD and MAD experiments that aim at the determination of $\mathbf{F}_T$ from the heavy atom structure factor $\left|\mathbf{F}_A\right|$ and the angle α between $\mathbf{F}_A$ and $\mathbf{F}_T$. Two pairs of such equations — corresponding to a two-wavelength MAD measurement — unambiguously determine $\left|\mathbf{F}_T\right|$, $\left|\mathbf{F}_A\right|$ and α, since the number of observables is four (two Friedel mates per wavelength) for three unknowns for each reflection; in the case of three wavelengths the problem is even more overdetermined. The phase of the heavy atom structure factor $\mathbf{F}_A$ can be deduced from the heavy atom substructure that is usually solved using $\left|\mathbf{F}_A\right|$ values calculated from the above equation.

On the contrary, in a SAD experiment we measure only $\left|\mathbf{F}^+\right|$ and $\left|\mathbf{F}^-\right|$ per reflection at one wavelength on one crystal that contains one anomalous scatterer element, therefore approximations are required to solve the equation (13). Since $\left|\mathbf{F}_A\right|$ values are desirable for substructure solution, $\left|\mathbf{F}_T\right|$ and α are approximated as follows.

1. With the assumption that $\left|\mathbf{F}_A\right| << \left|\mathbf{F}_T\right|$ and $f', f'' << f$, i.e. if the anomalous contribution is small in comparison to the non-anomalous scattering, $\left|\mathbf{F}_T\right|$ can be approximated as

$$\left|\mathbf{F}_T\right| \cong \frac{1}{2}\left(\left|\mathbf{F}^+\right| + \left|\mathbf{F}^-\right|\right) \qquad (14)$$

2. The angle α is set to 90° if $\left|\mathbf{F}^+\right| > \left|\mathbf{F}^-\right|$ and 270° if $\left|\mathbf{F}^+\right| < \left|\mathbf{F}^-\right|$.

3. As a consequence of these approximations, $\left|\mathbf{F}_A\right|$ values are estimated using ΔF values, because from (13):

$$\left|\mathbf{F}^+\right|^2 - \left|\mathbf{F}^-\right|^2 = 4\frac{f''}{f}\left|\mathbf{F}_T\right|\left|\mathbf{F}_A\right|\sin\alpha \qquad (15)$$

and after replacing $\left|\mathbf{F}_T\right|$ with its approximate described by (14):

$$\Delta F \cong 2\frac{f''}{f}\left|\mathbf{F}_A\right|\sin\alpha \qquad (16)$$

Note that $\sin\alpha \leq 1$ for the not approximated α and therefore this ΔF is a lower limit estimate of the corresponding $|\mathbf{F}_A|$.

The heavy atom substructure is solved on ΔF data that can be reliably used by direct methods programs in spite of the errors caused by approximations, since only reflections with the largest normalised structure factors (E values) are used and such reflections tend to obey the $\sin\alpha \cong 1$ assumption (see the estimation of weights below). However, errors in smaller ΔF values can disturb Patterson based methods that use all ΔF values to minimise Fourier series truncation errors. As illustrated in *Figure 1.9*, the Patterson function calculated from experimental and theoretical ΔF data can be essentially identical if the dataset was acquired precisely enough, whereas the same Harker section calculated from ideal $|\mathbf{F}_A|$ values can differ significantly from that with ΔF data. Thus, errors originating from the approximation of $|\mathbf{F}_A|$ with ΔF can be more pronounced than measurement errors if the anomalous signal is small. Patterson based methods may thus perform better in heavy atom location on stronger anomalous data.

| (a) | (b) | (c) |
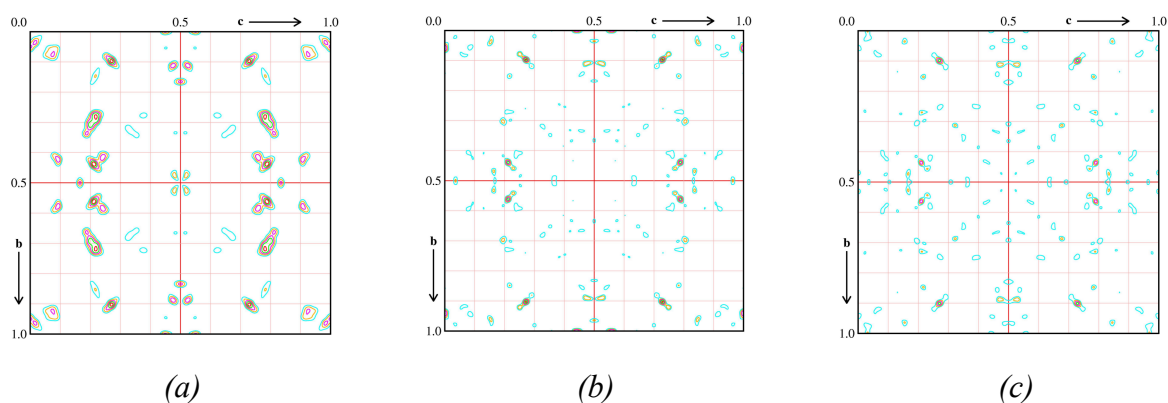
Figure 1.9. Origin removed a=1/2 Harker sections of cubic (Zn-free) insulin calculated from *(a)* ideal $|\mathbf{F}_A|$ data, *(b)* ideal ΔF data, *(c)* experimental ΔF data measured on a Bruker rotating anode generator equipped with Osmic multilayer mirrors, a three-circle goniometer and a Smart6000 4K CCD detector.

Having the anomalous scatterers located, $\mathbf{F}_A$, $\mathbf{F}_A'$ and $\mathbf{F}_A''$ can be calculated. Unfortunately, it is not possible to determine $\mathbf{F}_T$ unambiguously from SAD data, because two $\mathbf{F}_T$ vectors with different lengths and phases can be found to be equally good solutions for each reflection for a particular heavy atom set. This twofold ambiguity can also be illustrated with the Harker construction (*Figure 1.10*). $\mathbf{F}_A$, $\mathbf{F}_A'$ and $\mathbf{F}_A''$ correspond to the heavy atom substructure solution. From the endpoints of the $\mathbf{F}_A''^+$ and $\mathbf{F}_A''^-$ vectors circles can be drawn with radii of the measured $|\mathbf{F}^+|$ and $|\mathbf{F}^-|$. These circles have two intersection points corresponding to the two possible solutions, while point C in *Figure 1.10* represents the

centroid solution. It is instructive to notice that the vector resulting from a 90° rotation of $\mathbf{F_A}$ (that corresponds to $\alpha = 90°$ and $\mathbf{F_A}''$ in *Figure 1.10*) points exactly towards the centroid solution.
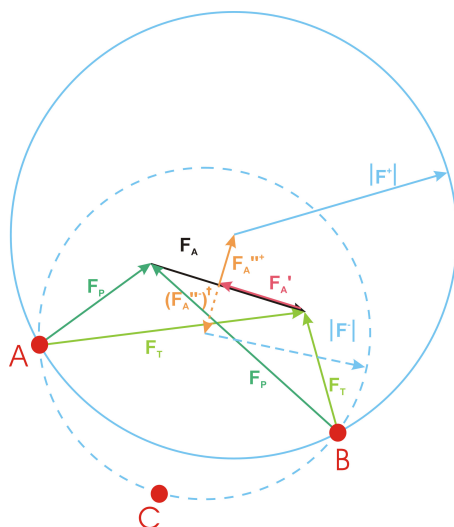
Figure 1.10. Harker construction representing the twofold ambiguity. The two possible solutions are represented by points **A** and **B**, while **C** corresponds to the centroid solution.

There are various methods to resolve the twofold ambiguity. The technique implemented in *SHELXE* (Sheldrick, 2002) can be summarised as follows. First, centroid phases ($\alpha = 90$ or $270°$) are assigned to the reflections as initial phases with weights depending on $|\mathbf{F_A}|$, ΔF and the resolution.

Two special cases can be discerned when evaluating the initial phase: when ΔF is large and when it is small. If ΔF is large (these represent reflections with large E values in the ΔF data) and $|\mathbf{F}^+| >> |\mathbf{F}^-|$, a good approximation of α is $90°$ (*Figure 1.11.a*), whereas if $|\mathbf{F}^+| << |\mathbf{F}^-|$ this value is $270°$ (*Figure 1.11.b*) and a high figure of merit (fom) is assigned to phases of these reflections indicating that the centroid solution does not differ severely from either of the two possible real solutions. On the contrary, setting α to $90°$ or $270°$ is a poor approximation when $|\mathbf{F}^+| \cong |\mathbf{F}^-|$. In such cases in fact it is not possible to decide between $\varphi_T = \varphi_A$ and $\varphi_T = \varphi_A + 180°$ and $\alpha = 90°$ or $270°$ corresponds to a phase error of about $90°$, therefore these reflections are given a fom near zero (*Figure 1.11.c-d*).

In the next step normalised structure factors are calculated both from the observed ΔF data ($|\mathbf{E_A}|$) and from $|\mathbf{F_A}|$ derived from the heavy atom substructure ($|\mathbf{E_H}|$). The correlation coefficient (CC) between $|\mathbf{E_A}|$ and $|\mathbf{E_H}|$ is then calculated as a function of the resolution (equation 9). The weight of a reflection is defined as

$$w = \frac{I_1(X)}{I_0(X)} \tag{17}$$

where I_1 and I_0 are Bessel functions and $X = \frac{|\mathbf{E}_A|CC}{1-CC^2}$.

Using the centroid phases, native $|\mathbf{F}_T|$ amplitudes and weights described above, an electron density map is calculated and a "low density elimination" cycle is performed. In this step only a few percent of the highest density is retained, the most of the electron density is reset to zero. From this map structure factors and phases are generated for centric reflections and reflections with weak $|\Delta F|$ and $|\mathbf{F}_A|$. If the heavy atoms are present in the native structure, their phases are included with sigmaA weights. These phase estimates may be rather poor but can further be improved by density modification methods.
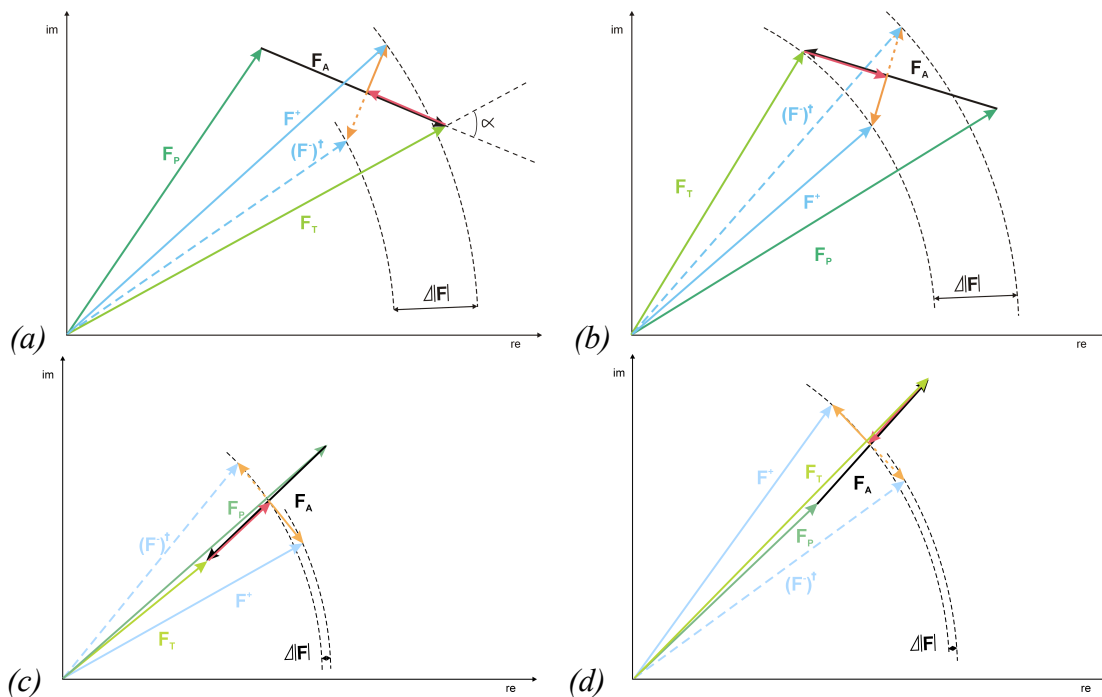
Figure 1.11. Argand diagrams representing two special cases in SAD phasing: *(a-b)* *ΔF* is large, *(c-d)*: *ΔF* is small. (In *(a)* and *(c)*: $|\mathbf{F}^+| > |\mathbf{F}^-|$, while in *(b)* and *(d)*: $|\mathbf{F}^+| < |\mathbf{F}^-|$.)

Heavy atom substructure solution and the importance of data quality

Although it seems straightforward to solve a structure by SAD taking the above considerations into account, this procedure includes the solution of a hidden phase problem, namely the location of the anomalous scatterers in order to calculate reference phases. Since the differences of observed structure factors that provide a basis for substructure solution are usually small, the data used for this purpose are noisy rendering the heavy atom location and

subsequent phase estimation more difficult. Therefore the quality of the anomalous data seems to be one of the most crucial points that determine the success of a SAD experiment. Fortunately, recent developments in cryocrystallography, area detectors, software for data reduction, substructure solution and phasing made the SAD technique viable even for such weak anomalous scatterers as sulfur.

As seen before, the ratio of the anomalous signal is determined by the f'' of the anomalous scatterer present in the structure that depends on the type of anomalous scatterer and on the wavelength used. For a particular element, the wavelength should be chosen where f'' reaches its maximum, i.e. the absorption edge. In the case of sulfur, this would mean a measurement at a wavelength of 5 Å. Although longer wavelengths are available at synchrotron sources or with a chromium anode (where f'' of sulfur is near 1.14 electrons), measurements using soft X-rays can pose several technical difficulties such as air scattering, absorption by the material in the X-ray beam (mother liquor, loop), higher diffraction angles that may strongly limit the resolution, severe radiation damage, instability of the beam position and intensity, etc. Numerous test measurements demonstrated that from this point of view home sources operating at CuKα have even advantages over synchrotrons, as they provide a virtually unlimited measurement time and a stable X-ray beam with a fairly long wavelength where $f''(S) = 0.56$ e.

As demonstrated in following chapters, phasing on the anomalous signal of sulfur measured in-house is indeed feasible, provided the anomalous differences can be measured precisely enough. In practice this requires the acquisition of highly redundant data. Real redundancy implies measurement of equivalent or identical reflections with different paths through the crystal, not just repeated observations; that is expedited by high crystal symmetry and by the use of a three-circle or kappa goniometer. Particular care should be taken in the cryogenic mounting of the crystal to reduce background scattering to a minimum and in the planning of the data collection to ensure a good spread of equivalent reflections on different frames so that residual systematic errors can be well modelled during scaling. S-SAD measurements not necessarily require the acquisition of data to the highest possible resolution; a highly redundant laboratory dataset combined with a low redundancy high resolution synchrotron dataset appears to be the most favourable arrangement, since the higher resolution "native" data can lead to an appreciable improvement in the quality of the experimental maps.

Before attempting to use anomalous differences for heavy atom substructure solution, it is advisable to assess the resolution limit of the anomalous signal in order to exclude resolution shells that would only contribute noise and make the heavy atom search less effective. It appears that the correlation coefficient between the signed anomalous differences of two statistically independent datasets or between experimental and ideal data gives consistent and reliable measures of the strength of the anomalous signal and of its dependence on resolution. In the case of MAD measurements, datasets measured at different wavelengths can be correlated. For SAD, the statistically independent datasets may consist of data collected from two different crystals or two datasets collected using one crystal but in different slicing modes, e.g. φ- and ω-scans. The latter procedure has the advantage that there are no problems in subsequent merging of the datasets to obtain a higher redundancy. After the assessment of the resolution limit of the anomalous signal, all data can be scaled together resulting in a dataset having twice the redundancy of those used for comparison. This explains why the correlation coefficients between observed and ideal data are usually higher than those between φ- and ω-scans. After merging the datasets, as an alternative test, the ratio of ΔF to its estimated standard deviation can be calculated as a function of resolution.

The number of atoms to be located is also an important factor in determining the success of heavy atom substructure solution. The number of sites to be requested should be within 20% of the true value so that the occupancy refinement could work well in *SHELXD* and reveal the real number of heavy atoms. Location of sulfur atoms in disulfide bridges represents a special case in heavy atom location, as the number of sites depends on the resolution of the anomalous data. At low resolution (lower than 2.8 Å) the two sulfur atoms of the disulfide unit coalesce to a super-sulfur atom, therefore half the real number of sulfurs should be searched. On the contrary, high resolution (better than 2.0 Å) anomalous data usually allow the direct resolution of disulfide units to single atoms. Interestingly, modelling of the heavy atom substructure with super-sulfur atoms is not sufficient if the anomalous signal extends to medium resolution and reduces the efficiency of subsequent phasing and phase improvement steps (*Figure 1.12.a*). This problem can be overcome by the direct location of two-atom disulfide units using *SHELXD*. In such cases, first super-sulfur atoms are located that are in turn tested to see if they can be better fitted by a disulfide unit with an S—S distance of 2.06 Å than by a single atom. The positional coordinates of the super-sulfur atoms sometimes correspond to the middle of the disulfide bond, but more often they are closer to one or other of the two sulfur atoms. The best position of the disulfide unit is found

by the local optimisation of its fit starting from twelve systematically chosen orientations. The disulfide unit is preferred to a single atom if the sum of the densities at the two sulfurs is greater than 1.3 times the height of the original super-sulfur peak. Tests on known and previously unknown structures indicated that this innovation — apart from the marked improvement in phasing (*Figure 1.12.b*) — improves the discrimination between correct solutions and noise peaks of the heavy atom substructure and allows the location of "hidden" single sulfur peaks. Although this algorithm is used in the peak-search routine of each dual-space refinement cycle, the extra computer time required is negligible, since only the regions around the highest peaks are analysed.

(a)   *(b)*
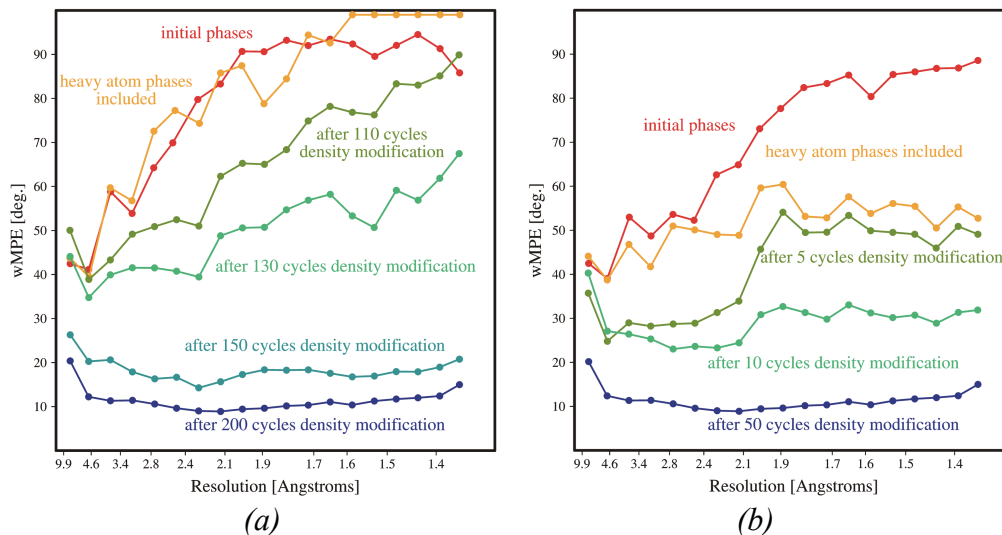
Figure 1.12. Improvement of phasing achieved by the resolution of super-sulfur atoms to single sulfurs of cubic (Zn-free) insulin measured at the PX beamline at ANKA at 1.0 Å wavelength demonstrated with the weighted mean phase error as a function of the resolution at different stages of phase improvement. The anomalous signal extends only to 2.2 Å. *(a)* phasing with super-sulfur atoms and *(b)* with resolved sulfur atoms.

In the case of SAD and MAD phasing, both the heavy atom set and its enantiomorph are in agreement with the anomalous data, therefore both of them need to be tested in subsequent steps, i.e. density modification should be performed for maps calculated using both possible solutions. The heavy atom set with the correct hand usually results in an interpretable electron density map, while the wrong heavy atom solution only gives uninterpretable density. In special SAD cases, when the heavy atoms show intrinsic centrosymmetry, "both enantiomorphs" lead to the same interpretable electron density.

I.2.3 Molecular replacement

The molecular replacement method is based on the observation that proteins exhibiting high sequence homology possess similar three-dimensional structures. Thus, if the structure of a protein that is expected to be similar has already been solved, it can serve as a very first model for further refinement steps. As the number of already determined protein structures with different folds rapidly increased in the last couple of years, molecular replacement has become one of the most effective techniques for macromolecular structure determination. This method has the advantage over other techniques that it requires the acquisition of only one native dataset with an average redundancy and resolution. The model bias that is undoubtedly present in all structures solved with molecular replacement can be mentioned as a disadvantage.

However, the knowledge of atomic coordinates of the search model is not sufficient for structure factor calculation, as it also requires the orientation and position of the model in the asymmetric unit of the unknown structure. Hence, the basic task of molecular replacement is to transfer the known molecular structure from its crystalline arrangement to the crystal of the protein for which the structure is not yet known. This transformation can be described by six variables from which three correspond to a rotation matrix while the other three to a translation vector:

$$x' = [\mathbf{C}]x + \mathbf{t} \tag{18}$$

where x represents the original, x' the transformed coordinates, while $[\mathbf{C}]$ describes the rotation matrix and $\mathbf{t}$ the translation vector. Although current programs (like *EPMR*, Kissinger *et al.*, 1999) calculate $[\mathbf{C}]$ and $\mathbf{t}$ simultaneously in a six-dimensional search, classical molecular replacement programs (like *AMORE*, Navaza, 1994) perform the rotational and translational search in separate steps in order to reduce the computational time.

Both $[\mathbf{C}]$ and $\mathbf{t}$ can be determined by means of the Patterson function. A molecule of n atoms generates a characteristic pattern of $n(n$-1) peaks in the Patterson function corresponding to the self (intramolecular) vectors, i.e. distances between atoms within the molecule. Another identical molecule shows an identical Patterson peak pattern with a rotation corresponding to the rotation between the two molecules. The unit cell — except for P1 — is made up by several molecules connected by symmetry elements and this is reflected in the Patterson function as well, as it also contains peaks corresponding to cross (intermolecular) vectors, which can give information about the position of a molecule in the

unit cell. The self Patterson vectors are relatively short and therefore their end-points appear to be near the origin while the cross vectors are longer and their end-points are further apart from the origin.

The rotational search. The rotation matrix between two identical but differently oriented molecules can thus be derived from the self vectors in the Patterson function. If a Patterson function is superimposed on its rotated image, maximal overlap is obtained when the rotation corresponds to the orientation difference of the molecules. The overlap function R of a Patterson function $P(\mathbf{u})$ is defined as

$$R([\mathbf{C}]) = \int_U P(\mathbf{u}) \times P_r(\mathbf{u}_r) d\mathbf{u} \qquad (19)$$

where $P_r(\mathbf{u}_r)$ is the rotated Patterson function of the search model in a P1 model cell and $[\mathbf{C}]$ is the rotation matrix. The correct selection of the integration volume U is one of the most critical points for the success of the rotational search, as it should only contain the self Patterson vectors. For a molecule of diameter d the intramolecular Patterson vectors occupy the volume of a sphere with a radius of d. A reasonable integration volume therefore corresponds to a sphere with a radius that is smaller than d.

Another point to consider is the resolution range of the data used in the calculation of the rotation function. Low resolution data may be excluded since they are insensitive to rotation and are mostly determined by the solvent. High resolution data are the most discriminating, because they are more sensitive to the model. The exclusion of the high resolution data can therefore help to smear the differences between the search model and the structure to be determined.

The self-rotation function. The non-crystallographic symmetry (NCS) refers to the coexistence of more than one molecules in the asymmetric unit. Unlike crystallographic symmetry, NCS represents only a local symmetry and connects molecules only within the asymmetric unit. The self-vectors for two NCS-related molecules in the Patterson function fulfil the symmetry between the molecules, i.e. maxima found in the rotation function calculated using the Patterson function of the measured dataset and its rotated image correspond to the number and type of symmetry operators of the molecules connected by NCS.

The translational search. Having the orientation of the search model established, the next step is to calculate the translation required to overlap the model onto the structure in

question. A simple way to determine the absolute coordinates of the model in the unit cell is to move the model through the asymmetric unit and evaluate the distinct positions by the agreement of the calculated and measured structure factors. Fortunately, structure factors do not have to be calculated from the new atomic coordinates, since only the phases of the reflections change when moving the molecule relative to the origin. The correlation coefficient (CC) and R-factor (R) can be used as figures of merit:

$$CC = \frac{\sum_{hkl}\left(|\mathbf{F}_o|^2 - \overline{|\mathbf{F}_o|^2}\right) \times \left(|\mathbf{F}_c|^2 - \overline{|\mathbf{F}_c|^2}\right)}{\left[\sum_{hkl}\left(|\mathbf{F}_o|^2 - \overline{|\mathbf{F}_o|^2}\right)^2 \sum_{hkl}\left(|\mathbf{F}_c|^2 - \overline{|\mathbf{F}_c|^2}\right)^2\right]^{1/2}} \qquad (20)$$

$$R = \frac{\sum_{hkl}\left||\mathbf{F}_o| - k|\mathbf{F}_c|\right|}{\sum_{hkl}|\mathbf{F}_o|} \qquad (21)$$

where $\mathbf{F}_o$ is the observed and $\mathbf{F}_c$ the calculated structure factor, k represents a scale factor. The correlation coefficient has the advantage over the R-factor that it is independent of the scale factor.

As an alternative of the trial-and-error method, a more systematic and straightforward technique to determine the position of the model is the calculation of the translation function that uses the cross Patterson vectors of the model structure and the observed Patterson function:

$$T(\mathbf{t}) = \int_V P_{1,2}(\mathbf{u},\mathbf{t}) \times P(\mathbf{u})d\mathbf{u} \qquad (22)$$

where $P_{1,2}(\mathbf{u},\mathbf{t})$ is the set of cross Patterson vectors of two molecules related by crystallographic symmetry, $P(\mathbf{u})$ the observed Patterson function calculated from the measured data. If $\mathbf{t}$ equals an intermolecular vector, the $T(\mathbf{t})$ function reaches a maximum value.

In the P1 space group that does not contain any crystallographic symmetry, the origin of the unit cell can be chosen arbitrarily because it does not affect the absolute value of the structure factors. In such cases the translational search is not necessary. Similarly, in space groups containing a polar axis the translational search can be restricted to two instead of three variables.

I.3 Density modification techniques

In most of the cases the quality of the first electron density map calculated with initial phases is not sufficient to build the first molecular model. This can be attributed to significant phase errors originating from small differences of measured intensities and to the partly resolved twofold ambiguity in the case of SAD and SIR. In addition, data used for substructure solution usually provide phase information to a lower resolution than the native dataset used for the final refinement, so phase extension to higher resolution shells is required. Density modification techniques may be advantageous also in removing the model bias from electron density maps obtained from molecular replacement techniques.

The general scheme of density modification procedures is illustrated in *Figure 1.13*. In each cycle, the experimental $|\mathbf{F}_o|$ values are combined with calculated phases φ_c or with phases that are calculated as an average of experimental (φ) and calculated phases if experimental phases are also available. The resulting electron density is in turn modified in a suitable way and from this density new (calculated) structure factors ($|\mathbf{F}_c|$) and phases (φ_c) are obtained. The calculated structure factors $|\mathbf{F}_c|$ are usually not carried over to the next cycle but can be used indirectly to determine weights for the phase recombination.
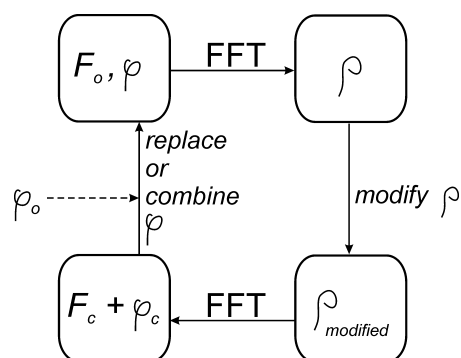
Figure 1.13. General scheme of density modification techniques.

Solvent flattening and flipping techniques

The basic idea behind the solvent flattening and flipping techniques is the general observation that the unit cell of a protein crystal contains two main regions: the region of the protein molecules themselves and large solvent areas with a smooth interface between them (Wang, 1985). The first step of solvent flattening and flipping is to determine which pixels of the electron density map belong to the protein and which to the solvent region. This gives a

raw estimate of the limits of the protein region, in other words: the molecular envelope. The electron density within the protein region is usually higher than that in the solvent which provides a good basis for discrimination between these two regions. A more suitable way of partitioning the cell is based on the determination of the fluctuation, i.e. the standard deviation of the electron density:

$$\overline{\sigma} = \sqrt{\frac{\sum (\rho - \overline{\rho})^2}{n}} \tag{23}$$

where n is a number of grid points within a suitably chosen sphere (e.g. with a radius around 3 Å). Grid points within a strongly fluctuating area are assigned to the protein region, whereas smooth density corresponds to the solvent area.

The grid points are then sorted according to their standard deviation and divided into a solvent and a protein group in such a way that the ratio of protein and solvent pixels gives the expected solvent content that is estimated from the cell volume and the expected number of protein molecules. The density in the solvent region is reset to a constant low value, while the density in the protein region is retained. Small islands of protein in the solvent and solvent cavities in the protein are also removed. The phases calculated from the resulting map are combined with the experimental phases or with those from the previous cycle. Unfortunately, the probability distribution of φ_c is not independent of φ_o.

The correlation between φ_c and φ_o is remarkably reduced by the solvent flipping technique. In this case instead of flattening the solvent, density of the solvent is inverted:

$$\rho_{\text{modified}} = -\gamma \rho + \overline{\rho} \tag{24}$$

where γ is a constant in the $0 < \gamma < 2$ range and is estimated from the solvent content. This algorithm is implemented in *SOLOMON* (Abrahams & Leslie, 1996) and also in *DM* (Cowtan, 1994).

SHELXE is based on a slightly different idea of distinguishing between solvent and protein, as it uses the fluctuation on a spherical surface instead of a spherical volume around the map grid points. It assumes a pixel in the electron density to be part of the protein region if the variance of the electron density

$$V = \overline{(\rho - \overline{\rho})^2} \tag{25}$$

on a spherical surface of radius 2.42 Å is large. The distance defined by this sphere corresponds to the dominant 1,3-bond distance in macromolecules, a high variance at such a distance therefore indicates the presence of an atom on the particular pixel (sphere of influence method). The density at the protein pixels is replaced by

$$\rho_{P,\,modified} = \sqrt{\frac{\rho^4}{g^2\sigma^2(\rho)+\rho^2}} \tag{26}$$

if the density is positive and by zero if it is negative. The solvent density is simply flipped:

$$\rho_{S,\,modified} = -\gamma\rho \tag{27}$$

where g and γ are usually set to 1.0. In addition to the solvent and protein regions, a fuzzy solvent boundary is defined to ensure that the density modification is not locked into an initial poor solvent boundary. In this region both modifications described above are applied and the resulting density is the weighted mean of the two treatments:

$$\rho_{fuzzy,\,modified} = \beta\rho_{P,\,modified} + (1-\beta)\rho_{S,\,modified} \tag{28}$$

where β is derived from the estimated probability of pixels being part of the protein region.

Non-crystallographic symmetry (NCS) averaging

If the asymmetric unit of a crystal contains more than one equivalent copies of the protein molecule, the electron density within the molecular envelopes of the "monomers" can be assumed to be very similar. With the knowledge of the mask and the symmetry operation connecting the molecules, the electron density can thus be averaged at corresponding points or modified in a similar manner, which leads to an improvement of phases (Bricogne, 1974; Schuller, 1996). NCS averaging is a powerful method at low resolution, especially in combination with solvent flattening or flipping. In the case of several copies (high NCS), e.g. in virus structures, it can even serve as an *ab initio* phasing technique.

Histogram matching

Histogram matching is usually applied in combination with solvent flattening and flipping techniques. This procedure utilises the observation that the difference between the magnitude distribution of the electron density for similar structures with the same solvent content at similar resolutions is negligible. In practice, the frequency distribution of certain

density values at grid points of the electron density map does not depend on the protein, only on the resolution and the solvent content of the crystal. On the contrary, distribution of density values can differ strikingly for a high and a low quality electron density map (*Figure 1.14*). Thus, the relative frequency of points carrying a certain electron density is calculated and compared to that of a theoretical histogram. The electron density is then replaced by a value that better corresponds to the theoretical histogram. This method does not require the definition of the molecular envelope, but if used in combination with solvent flattening or flipping, it is only applied in the protein region (Zhang & Main, 1990).
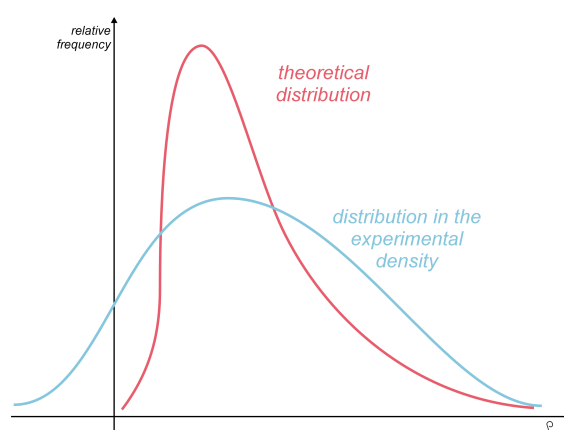
Figure 1.14. Theoretical and experimental frequency distribution of the magnitude of electron density.

Sayre's equation

Sayre's equation relates the electron density and its square:

$$\rho = q\rho^2 \tag{29}$$

where q is a resolution dependent scaling factor. Using the convolution theorem this equation can be written in reciprocal space as

$$\mathbf{F_h} = Q\sum_{\mathbf{h'}}\mathbf{F_h}\mathbf{F_{h-h'}} \tag{30}$$

where $\mathbf{h}$ means (h,k,l) and Q is a function of the resolution and is estimated from a structure with similar resolution and solvent content (Sayre, 1974). Phase improvement methods using Sayre's equation are applied in reciprocal space. The reciprocal space form presents an exact relationship between the structure factors of two different reflections — and as a

consequence, their phases. For N unknown reflections 2N such equations can be written and solved e.g. by least squares. The assumption formulated as $\rho=q\rho^2$ holds only at high resolution where atomicity of the electron density map can be expected, making this approach more powerful at atomic resolutions.

I.4 Structure refinement

Initial models obtained by the interpretation of the experimental electron density map or from molecular replacement are generally only in poor agreement with the observed data, as they can contain coordinate errors or missing parts. Structure refinement as the last step of crystal structure determination aims at the reduction of the disagreement of experimental structure factor amplitudes and those calculated from the model. The crystallographic R-factor (given in equation 21) is widely used to monitor the agreement between measurement and the model.

During the refinement procedure, the R-factor is minimised by repeated manual or automatic model adjustment and completion and parameter refinement steps. Since decrease of the R-factor can be achieved by increasing the number of refined parameters, the structure might be overfitted, especially at low resolution where the data-to-parameter ratio is low. To avoid this, i.e. to give a measure of overfitting, cross-validation of the model was introduced (Brünger, 1992). For this purpose, a small set of reflections is excluded from the refinement (free set) and the model is only adjusted to the rest of reflections (working set). The R_{free} that is calculated as in (21) but only on the free reflection set monitors the fit of the atomic model to reflections not used in the refinement; the deviation of R_{free} from R gives thus a measure of overfitting.

Because of the unfavourable data-to-parameter ratio in the case of macromolecular structure refinement, especially at lower resolution, it is necessary to reduce the number of parameters or increase the number of observations. This is easily achievable by taking chemical information into account, such as bond lengths and angles. Beyond stereochemical information, other observations, like the flat density of the bulk solvent or the similarity of NCS-related molecules can also be incorporated in the refinement process. There are two main ways of handling prior knowledge: restraints and constraints. Restraints allow the variation of the parameter around a standard value controlled by an energy term, whereas constraints keep a parameter at an exactly defined value.

The refinement of parameters requires a defined target function that is minimised or maximised during refinement. It can be e.g. a least squares target if the sum of squared errors is minimised (Cruickshank, 1970):

$$Q = \sum_{hkl} w_{hkl} \left(\left| \mathbf{F}_o \right|^2 - \left| \mathbf{F}_c \right|^2 \right)^2 \tag{31}$$

where Q is the function to minimise and w_{hkl} is a weighting factor. The least squares refinement is adequate for nearly perfect models at higher resolutions (better than 2 Å), since it is fairly sensitive to larger errors in the starting models. Maximum likelihood targets aim at the maximisation of the probability that with the given model the given data would be observed. This can be defined as

$$L = \prod_i P\left(F_{i,o}; F_{i,c} \right) \tag{32}$$

where $P(F_{i,o}; F_{i,c})$ is the probability distribution of the structure factor $F_{i,o}$ given the model structure factor $F_{i,c}$ (Pannu & Read, 1996). Maximum likelihood refinement is particularly useful for incomplete models because it produces residuals that are less biased by the current model than those produced by least squares.

In macromolecular crystallography the most frequently used algorithms to minimise/maximise the target functions are the conjugate gradient algorithm and the simulated annealing approach. In contrast to the conjugate gradient algorithm that is adequate for medium and high resolution refinements, simulated annealing — especially in combination with torsion angle dynamics — is a powerful algorithm for removing gross errors from the model due to its larger convergence radius.

II A BOWMAN-BIRK INHIBITOR FROM THE LIMA BEAN (*PHASEOLUS LUNATUS* L.)

II.1 Introduction

II.1.1 Taxonomy and economic impact of lima bean

The species lima bean (*Phaseolus lunatus* L. or *limensis* L. in other sources) together with other well-known beans such as the common bean (*Phaseolos vulgaris* L.) belongs to the pea family (*Fabaceae*). This family can be classified into the class of dicotyledons (*Magnoliopsida*), among them into the subclass of roses (*Rosidae*). The family *Fabaceae*, called also *Leguminosae*, comprises 30 species that are widely used vegetable plants worldwide. Most of them are grown for their seeds and pods, some of them for tubers or tuberous roots. The seeded pods of plants in this family are among the first nutriments known to humanity, dating back 4000 years, but even nowadays beans provide an important dietary source of carbohydrates and proteins in both industrialised and developing countries, especially where animal protein is scarce or expensive (Hutchinson, 1969).

Biological nitrogen fixation by the symbiotic *Rhizobium* bacteria living in small growths on the lateral roots called nodules is one of the most important biological and physiological properties of plants in the *Leguminosae* family. Nitrogen fixed by bacteria supplies parts or in some cases the whole nitrogen demand of the plant, while the plant contributes a significant amount of photosynthates or other nutritional factors to the bacteria. Since the diverted energy can reach 20-30% of the dry matter accumulation, nitrogen fixation is thus a limiting factor for the economic yield of legumes. Nitrogen fertilisers are therefore often required to suppress nitrogen fixation and enhance the diversion of photosynthates to other plant functions (Smith & Gallon, 1993).

Beans are native to Central and South America, mainly to southern Mexico, Guatemala, Honduras and Costa Rica. The lima bean was named after Lima, the capital of Peru where it was found in the 16th century. Today two main varieties, the Fordhook and the baby lima are cultivated globally in warm climate environments and are the main food legumes in tropical Africa. They are grown for their seeds and picked fresh (i.e. green immature) or dried

(mature). The legume, a dehiscent pod which opens along the dorsal and ventral sutures, is tough and has a rather unpleasant flavour (*Figure 2.1*).

(a)   *(b)*
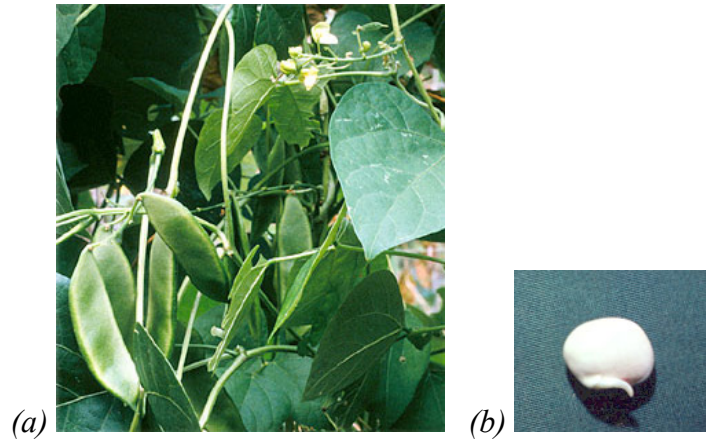
Figure 2.1. *(a)* Immaturate pods and *(b)* a seed of the lima bean.

II.1.2 Serine proteinases

Proteinases can be grouped into four major families based on their principal functional group in their active site: serine-, cysteine- (thiol-), aspartic- (carbonyl-) and metallo-proteinases. Among these the group of serine proteinases is the most numerous, widespread and diverse family. Serine proteinases can further be classified into two groups: the chymotrypsin and the subtilisin family. The chymotrypsin-related proteinases appear in prokaryotes, animals and viruses, while the subtilisin homologues are only represented in bacteria (Barrett, 1986; Lesk & Fordham, 1996). Similarly to these enzymes, the wheat serine carboxypeptidase II seems to have a serine-dependent mechanism, it can be thus regarded as a member of a distinct group of serine proteinases (Liao *et al.*, 1992). The three classes represent three different evolutionary avenues arriving at the same mechanism of action, they are otherwise not related to one other, as they possess different folding patterns.

The class of mammalian serine proteinases — members of the chymotrypsin family — are particularly well studied; they have been objects of wide-ranging research for their key role or involvement in numerous physiological processes. They act both as digestive enzymes and regulators through proteolytic activation of precursor proteins, like during the process of blood clotting (Davie *et al.*, 1991), fertilisation (Baba *et al.*, 1989) and complement activation in the immune response (Reid *et al.*, 1986). These enzymes are mostly synthesised in form of precursors that require the cleavage of the N- and/or C-terminal part for activation (Barrett, 1986).

The chymotrypsin related proteinases consist of two juxtaposed domains of similar structure that is believed to have arisen by gene duplication and divergence (McLachlan, 1979). Both domains fold into six-stranded β-barrels with shear number S=8, which than pack together asymmetrically (Lesk & Fordham, 1996). In most members of the family intradomain disulfide bridges stabilise the three-dimensional structure of the enzyme. The catalytic site lies in the cleft between the two domains (*Figure 2.2*).
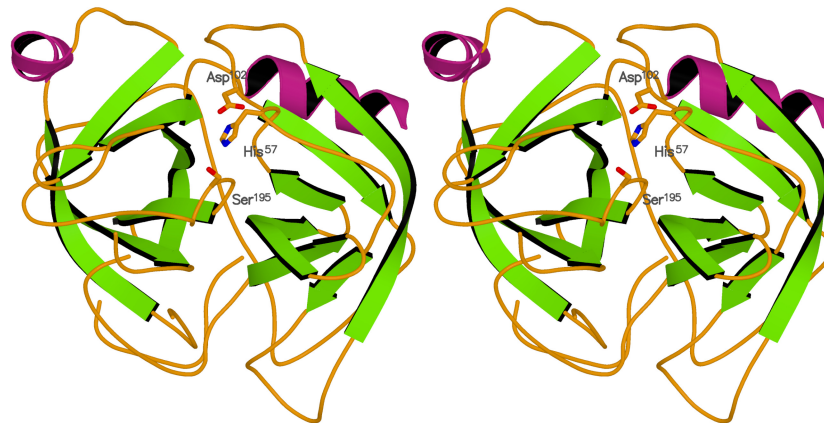
Figure 2.2. Stereo view of the ribbon representation of bovine α-chymotrypsin.

The various enzymes of this family catalyse the cleavage of polypeptide chains with different substrate preference. The accepted system to describe the specificity of proteases and complementary features of the substrate refers to the position of the residue in the substrate relative to the bond that is cleaved, as shown in *Figure 2.3* (Schechter & Berger, 1967). Trypsin favours substrates having Arg or Lys at the P1 position, while chymotrypsin hydrolyses polypeptides with aromatic amino acid residues at position P1. The molecular basis of substrate specificity is mainly determined by the residue at the S1 subsite; the preference of trypsin for Arg and Lys at P1 is well explained by Asp at S1. The P1-S1 interaction is crucial for orienting the substrate properly, but alone not sufficient to achieve maximum catalytic activity, although in enzymes with intensive P1-S1 binding other subsites
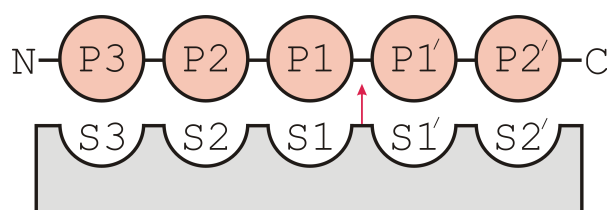
Figure 2.3. Scheme for the terminology of specificity subsites of proteinases and the complementary features of the substrate (Schechter & Berger, 1967). The binding site is composed of subsites (S3-S2'), while the positions of the substrate are represented as P3-P2'. The cleavage site is shown as a red arrow.

have little effect on the rate of catalyis. Three other interactions in the catalytic centre serve to fix the scissile bond of the substrate in a specific orientation. One is a H-bond between the backbone N-atom of P1 and the carbonyl group of the residue occupying the position 214 in the enzyme. Two additional H-bonds connect the carbonyl group of P1 to the backbone N atoms of residues 193 and 195, known as the oxyanion binding site (Hedstrom *et al.*, 1992).

Early studies were aimed primarily at the elucidation of the exact chemical mechanism of catalysis (Polgár, 1989). Three amino acid residues, namely His[57], Asp[102] and Ser[195] (according to chymotrypsin numbering) were found to play an essential role in the peptide cleavage process. These three residues have later come to be known as the charge-transfer relay or the catalytic triad. The enzymatic reaction is of the ping-pong or substituted enzyme type, and can be divided into two symmetric reaction sequence steps (*Figure 2.4*). In the substrate-free state the O-H bond in the hydroxyl group of Ser[195] is strongly polarised due to the proximity of the basic imidazole ring of His[57]. As the first step, the nucleophilic O atom of Ser[195] attacks the carbonyl carbon atom at the scissile bond. It is followed by proton donation from His[57] to the newly formed amine group. After dissociation of the first product and formation of an acyl-enzyme complex, a water molecule attacks the carbonyl oxygen, i.e. the second step occurs via the same mechanism. The resulting acidic component then leaves the catalytic cleft and so the starting state of the active site is restored. In the other serine proteinase families, e.g. in the case of subtilisin, similar constellations of three such residues were found. However, the residues important for the catalysis (Asp[32], His[64] and Ser[221]) occur in different order in the primary structures of the members in the two families proving their independent evolution and later convergence.
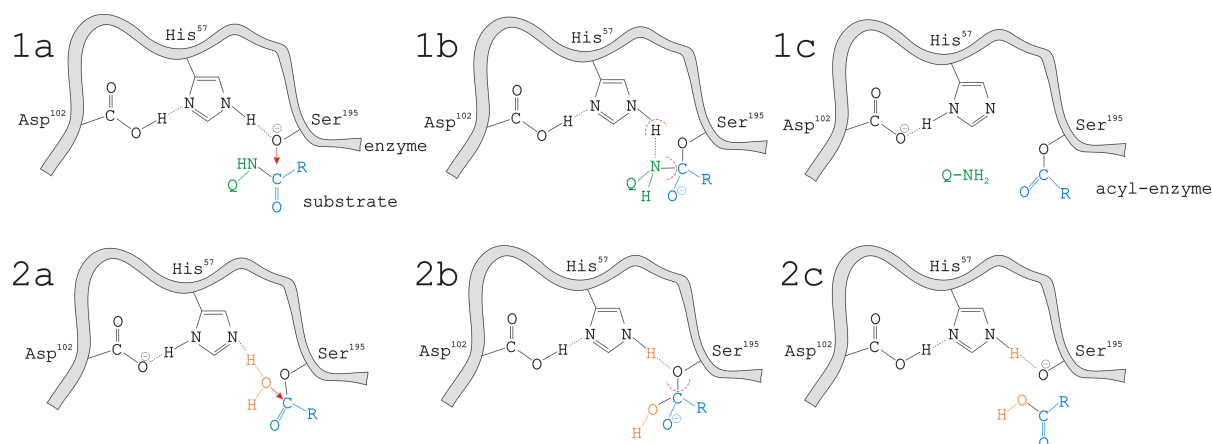
Figure 2.4. The molecular mechanism of serine proteinases. Note that the two consecutive steps (*1a-c* and *2a-c*) occur via the same mechanism.

II.1.3 Proteinaceous inhibitors of serine proteinases

Structure and mechanism

Protein type inhibitors of proteinases are ubiquitous in nature, as they are present in many different forms and even isoforms in almost all plant and animal tissues as well as in microorganisms, where they prevent unintentional proteolysis and can also signal receptor interactions. Besides protein proteinase inhibitors, microorganisms often synthesise small non-proteinaceous inhibitors that block serine proteinases in the host cell. The number of inhibitors characterised as serine proteinase inhibitors far exceeds that of other mechanistic classes of proteinases reflecting the evolutionary success and abundance of this proteinase family (Laskowski & Kato, 1980).

The majority of these inhibitors can be described as small proteins of around 30-200 amino acid residues with some common structural features such as the presence of a characteristic binding loop in a canonical conformation and a generally compact shape. The binding loop exhibits in almost all cases an extended geometry from P3 to P3'. Its shape resembles the main chain conformation of an antiparallel β-strand at P3, a polyproline II at P2 and P1', a 3_{10}-helix at P1 and a parallel β-strand at P2' and P3'. The main chain conformation forces the side chains around the scissile bond to project towards the solvent or the target enzyme (Bode & Huber, 1992).

Apart from the canonical binding loop, the inhibitor molecules possess quite different folding motifs characteristic of the inhibitor class they belong to. However, they all exhibit a very compact packing of secondary elements and have a hydrophobic core that very often consists of cross-connecting disulfide bridges that provide these structures with a high degree of rigidity. The native inhibitors are usually highly thermostable and able to preserve their stability even in solutions containing guanidine hydrochloride at high concentrations. Although the binding loop is extended, it is fixed by many intramolecular H-bonds and often with additional disulfide bridges in such a way that it is virtually spanned between anchoring points of the structure.

The overall minimal mechanism of the enzyme-inhibitor interaction can be written as follows:

$$E+I \rightleftharpoons EI \rightleftharpoons E+I^*$$
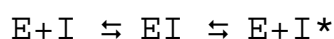
where E is the enzyme and I and I* are the virgin and modified inhibitors, respectively. EI, which represents the stable complex is rapidly formed and then dissociates slowly into a free enzyme and an unaltered or modified inhibitor. The latter, like the product of the successful enzymatic reaction is specifically cleaved at the P1-P1' bond (Finkenstadt *et al.*, 1974; Quast *et al.*, 1978).

The mode of enzyme-inhibitor association is very similar to that of an enzyme-substrate complex: the canonical binding loop serves as a linear functional epitope that is easily recognised by the enzyme. The structural basis of inhibitor specificity also corresponds to the substrate preference of the enzyme. In the case of chymotrypsin-like serine proteinases, a short antiparallel β-strand is formed between the P3-P1 and the S3-S1 (214-216 in chymotrypsin) subsites through main chain - main chain H-bonds, while the C-terminal side of the loop is bound through another H-bridge at P2'. The oxygen atom of the "reactive" carbonyl group always projects into the oxyanion hole and forms two H-bonds with the N atoms of Gly^{193} and Ser^{195}, while the carbonyl carbon makes "sub-van der Waals" contact with the side chain O of Ser^{195} of the catalytic triad. In the overwhelming majority of the inhibitor-proteinase complexes the scissile peptide bond remains intact, the carbonyl oxygen is out-of-plane displaced, though.

All the interactions listed above ensure similar recognition patterns between different inhibitors and proteinases. Since the P1-S1 interaction is one of the most important, energetically determining contact and S1 serves as a molecular filter in the chymotrypsin-like proteinases, the residue occupying the P1 site can be regarded as the primary determinant of the enzyme specificity. An Arg or a Lys residue at P1 confer for example trypsin specificity. The selectivity can be supported by further interactions with proteinase subsites comprising about 10 residues.
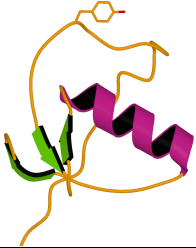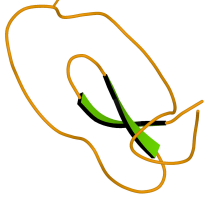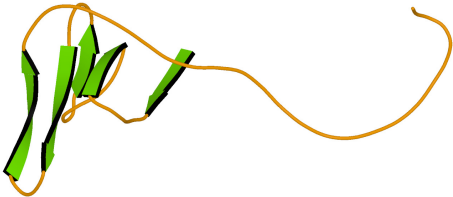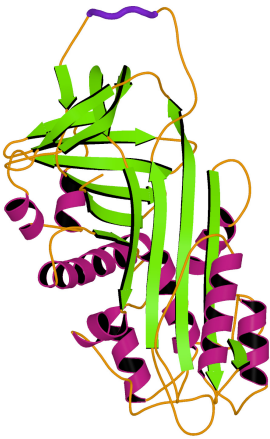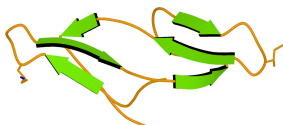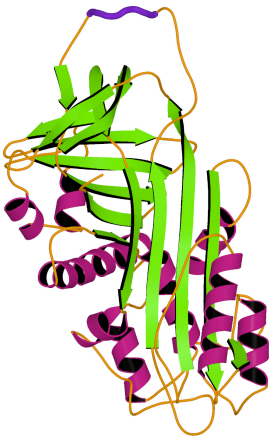
The canonical binding loop thus perfectly mimics the conformation and interaction pattern of a productive substrate. Considering this, the following question arises: how can the inhibitors preserve their binding loop in an intact state? The binding loop of the inhibitor in the complex is tightly packed and extremely rigid. These two features, i.e. the perfect complementarity with the enzyme and the inherent rigidity of the whole inhibitor molecule confer thermodynamic stability to the EI state and permit only a very slow dissociation to either products. The stabilising interactions prevent any conformational changes in the active site that is believed to be essential in the peptide bond cleavage and so slow down the catalytic process considerably (James *et al.*, 1980).

Classification

The serine proteinase inhibitors can be grouped into about 20 families based on their tertiary structure and enzyme preference, but since the list of inhibitors characterised is still expanding, the number of the diverse families can also be expected to grow. Some features of the most important 11 families can be summarised as follows (*Table 2.1*):

(1) Historically, the BPTI (bovine pancreatic trypsin inhibitor)-**Kunitz family** is the oldest one, it was named after the first inhibitor to be isolated, sequenced, crystallised and synthesised. This family comprises inhibitors mostly from the tissues of bovids and caprids. They consist of a three-stranded antiparallel β-sheet and an α-helix, to which the proteinase binding loop is attached.

(2) Members of the **Kazal-family**, like the pancreatic secretory inhibitors or ovomucoids also contain a three-stranded antiparallel β-sheet, an α-helix and a binding loop, but exhibit a different conformation and disulfide bridge arrangement.

(3) Representatives of the **STI** (soybean trypsin inhibitor)-**Kunitz family** were the very first plant inhibitors to be characterised. Their structure contains only β-strands joined by long loops.

(4-5) Members of both potato inhibitor families (**PI-I and PI-II**) possess a β-sheet core that is in the case of the PI-I family an interesting, mixed parallel-antiparallel sheet. The binding loop creates the connection between two of the β-strands. The short α-helix that is missing from the PI-II molecules has a similar function in the PI-I inhibitors.

(6) The **Bowman-Birk** type **inhibitors** are double headed, as their folding shows internal symmetry in the central β-sheet, which can be divided into two similar halves. The proteinase binding segment is also duplicated and as a result, these inhibitors are able to block two proteinases simultaneously.

(7) The **squash seed inhibitors** possess particularly short peptide chains that — apart from the very short central two-stranded β-ladder — have no regular secondary structure.

(8) The *Streptomyces* **inhibitors** are excreted by various species of *Streptomyces* with diverse specificity depending on the species. The inhibitor molecule consists of a central five-stranded β-sheet, a binding loop and a relatively short α-helix attached to that. The S-SI, a representative member of this family forms strong non-covalent dimers and inhibits two enzyme molecules without dissociation.

Table 2.1. A summary of serine proteinase inhibitor families.

structure	family (examples)	structure	family (examples)
	BPTI-Kunitz family • bovine pancreatic trypsin inhibitor • amyloid β-protein precursor inhibitor		Squash seed inhibitor family • *Ecballium elaterium* trypsin inhibitor (EETI) • *Cucurbita maxima* trypsin inhibitor
	Kazal family • pancreatic secretory trypsin inhibitor		SSI-family • *Streptomyces* subtilisin inhibitor • plasminostreptin
	STI-Kunitz family • soybean trypsin inhibitor		Chelonianin family • mucous proteinase inhibitor
	PI-I family • barley chymotrypsin inhibitor • eglin (from leech)		Hirudin
	PI-II family • potato chymotrypsin inhibitor		Serpins • α₁ protease inhibitor • ovalbumin, plakalbumin
	Bowman-Birk family • soybean inhibitor • mung bean inhibitor • peanut inhibitor		(The binding loop that gets cleaved is shown in blue.)

(The views were chosen to present the binding loop on the top in similar orientations.)

(9) Similarly to squash seed inhibitors, no regular secondary structures can be found in the fold of **mucous proteinase inhibitors**, i.e. inhibitors of the chelonianin family, apart from two pairs of short β-strands. Some of these molecules consist of two similarly folded domains, but only one of them has inhibitory activity.

(10) **Hirudin**, the selective inhibitor of the coagulation proteinase thrombin, folds into a globular, compact domain, but its C-terminal tail remains flexible in non-complexed state. This part of the peptide chain embraces the protease molecule upon binding.

(11) The **serpin**s (**ser**ine **p**roteinase **in**hibitors) form a non-canonical serine proteinase family. They are of high molecular weight compared to members of other families having about 400 amino acid residues. In contrast to other inhibitor types, the reactive loop of serpins does get cleaved by their target enzyme. This family also contains proteins that apparently lack inhibitory properties, such as ovalbumin.

II.1.4 The Bowman-Birk inhibitor family

Historical overview

Proteinase inhibitors occurring in the plant kingdom attracted the attention of researchers in many disciplines for their possible effect on the nutritive value of plant proteins, particularly in legume seeds. As early as 1917, Osborne & Mendel observed that soybeans, unless cooked for several hours, would not support the desirable growth of rats and cause pancreatic hypertrophy. This observation was followed by the discovery of a heat labile, acetone soluble protein (the Kunitz type soybean inhibitor) that inhibits mammalian trypsins (Ham & Sandstedt, 1944). The soybean Bowman-Birk inhibitor (BBi) was purified from the acetone insoluble fraction somewhat later, in 1961 (Birk, 1961). During the following decades many BBis were identified, isolated and characterised from other legumina, like from *Phaseolus* spp (Stevens *et al.*, 1974; Wilson & Laskowski, 1975; Ishikawa *et al.*, 1979; Kiyohara *et al.*, 1981; Zhang *et al.*, 1982; Funk *et al.*, 1993).

Activity, structure and evolution of BBis

BBis have molecular weights ranging from 7 to 10 kDa and all contain seven highly conserved disulfide bridges. Sequence alignment of inhibitors of this kind from *Glycine max* (soybean) and *Phaseolus* (bean) spp (*Figure 2.5*) reveals the presence of two independent proteinase binding sites in two homologous regions. Accordingly, most of the legume

(a)

```
binding loop 1.
                         0          10              20          30          40
                         |          |               |           |           |
                                           ----                 *
       gl. max      ----  ------SHDE  PS----ESSEPCCD  SCDCTKSIPP  ECHCANIRLN  SCHSA
       gl. max      ----  -XHHXSSXXX  PS----XSSPPCCX  ICVCTASIPP  QCVCTXIRLX  SCHSA
       gl. max      KLGL  LMKSDHQHSN  DD----ESSKPCCD  QCACTKSNPP  QCRCSDMRLN  SCHSA
       ph. angularis ----  SVHHQDSSDE  PS----ESSHPCCD  LCLCTKSIPP  QCQCADIRLD  SCHSA
       ph. angularis ---M  ELNLFKSDHS  SS--DDESSKPCCD  LCMCTASMPP  QCHCADIRLN  SCHSA
       ph. vulgaris  --MC  ILSFLKSDQS  SSYDDDEYSKPCCD  LCMCTRSMPP  QCSCEDIRLN  SCHSD
       ph. vulgaris  ----  SGHHDETTDE  PS----ESSKPCCD  QCSCTKSMPP  KCRCSDIRLN  SCHSA
       ph. vulgaris  -SGH  RHESXXSTXX  AS----XSSKPCCX  HCACTKSIPP  QCRCSXLRLN  SCHSE
       ph. aureus    KSGH  RHESXXSTXX  AS----XSSKPCCX  HCACTKSIPP  QCRCSXLRLN  SCHSE
       ph. lunatus   ----  SGHHEHSTDX  PS----XSSKPCCX  HCACTKSIPP  QCRCTDLRLD  SCHSA

binding loop 2.
                         50         60             70          80
                         |          |               |           |
                              *
       gl. max      CKSCI  CTRSMPGKCR  CLDTDDFCYK  PCESMDKD--  ---
       gl. max      CKSCM  CTRSMPGKCR  CLXTTXYCYK  SCKSXSGXXX  ---
       gl. max      CKSCI  CALSYPAQCF  CVDITDFCYE  PCKPS-EDDK  EN-
       ph. angularis CKSCM  CTRSMPGQCR  CLDTHDFCHK  PCKSRDKD--  ---
       ph. angularis CDRCA  CTRSMPGQCR  CLDTTDFCYK  PCKSSDEDDD  ---
       ph. vulgaris  CKSCM  CTRSQPGQCR  CLDTNDFCYK  PCKSRDD---  ---
       ph. vulgaris  CKSCA  CTYSIPAKCF  CTDINDFCYE  PCKSSRDDDW  DN-
       ph. vulgaris  CKGCI  CTFSIPAQCI  CTDTNNFCYE  PCKSSHGPXX  NN-
       ph. aureus    CKGCI  CTFSIPAQCI  CTDTNNFCYE  PCKSSHGPX-  NN-
       ph. lunatus   CKSCI  CTLSIPAQCV  CXXIXDFCYE  PCKSSHSDDD  NNN
```

(b)

Figure 2.5. *(a)* Sequence alignment of the BBis isolated from *Glycine* and *Phaseolus* spp. The conserved Cys residues are marked with red, the proteinase binding loops are indicated with green, the residues at position P1 are marked with *. Ambiguous residues were replaced by X. *(b)* Disulfide bridge arrangement in the BBi family.

inhibitors exhibit double action, as they typically inhibit trypsin and chymotrypsin simultaneously in an independent way, therefore they are also called double-headed inhibitors. However, some of these inhibitors were demonstrated to react with other proteinases like elastase (in the case of the kidney bean inhibitor) or to react with the same enzyme at both of the active sites, like the mung bean BBi (Birk, 1985).

The specificity, similarly to other inhibitor families, is mainly determined by the amino acid residue at the P1 position of the reactive loop. Inhibitors with Arg or Lys are able to block trypsin; Leu, Tyr or Phe inhibits chymotrypsin; while Ala is specific for elastase (Ikenaka & Norioka, 1986).

The three-dimensional structures of these inhibitors perfectly explain the independent manner of their double action. These molecules contain two homologous domains that together resemble the shape of a bow-tie. Both domains consist of a three-stranded β-sheet and an extended binding loop in the case of all representatives described so far. (Publications concerning structural studies in BBis are listed in *Table 2.2*.)

Table 2.2. Structural studies on BBis. (TRP=trypsin)

name	method	PDB deposition code	reference
BBi from snail medic seeds	NMR	1MVZ	Catalano *et al.*, 2003
cyclised hairpin from sunflower BBi	NMR	1O8Y-Z, 1GM2	Marx *et al.*, 2003; Brauer *et al.*, 2002
pea seed BBi (dimeric)	X-ray	1PBI	Li de la Sierra *et al.*, 1999
barley BBi	X-ray	1C2A	Song *et al.*, 1999
soybean BBi	X-ray	1K9B	Voss *et al.*, 196
pineapple BBi	NMR	1BI6, 2BI6	Hatano *et al.*, 1996
tracy soybean BBi	X-ray	1PI2	Chen *et al.*, 1992
soybean BBi	NMR	1BBI	Werner & Wemmer, 1992
peanut BBi	X-ray	not deposited	Suzuki *et al.*, 1987
soybean BBi in complex with 2 TRPs	X-ray	1D6R	Koepke *et al.*, 2000
mung bean BBi in complex with a dengue virus proteinase	X-ray	1DF9	Krishna Murthy *et al.*, 2000
synthesised fragment from mung bean BBi in complex with TRP	X-ray	1SMF	Li *et al.*, 1994
mung bean BBi in complex with 2 TRPs	X-ray	not deposited	Lin *et al.*, 1993
adzuki bean BBi TRP-binding region in complex with TRP	X-ray	1TAB	Tsunogae *et al.*, 1986

Odani & Ikenaka (1973) separated the two active regions by splitting the peptide chain between them and showed that both reactive sites remained active even after separation. Based on this finding, the following scheme could be proposed for the evolution of BBis (*Figure 2.6*). The first double-headed inhibitor is hypothesised to originate from an ancestral single-headed inhibitor by gene duplication. Its homologous domains were connected by a single connective chain. A second, this time only partial gene duplication extended the C-terminal part that in turn replaced the N-terminus of the first domain. This resulted in the rearrangement of disulfide bridges and the formation of a new disulfide bond, but did not cause any conformational changes at the reactive sites (Suzuki *et al.*, 1987).

(a)   *(b)*   *(c)*
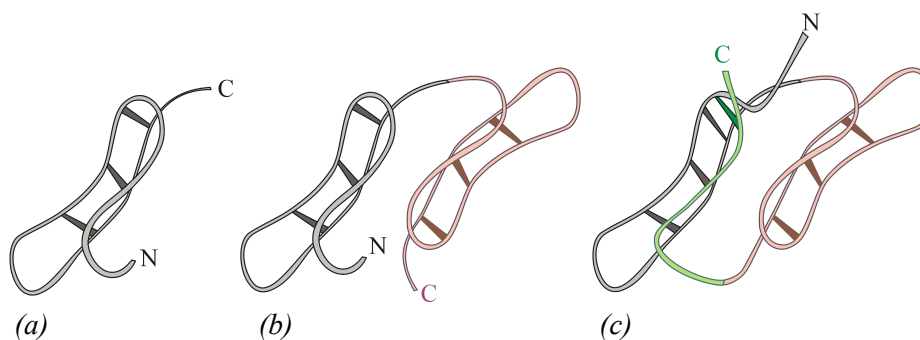
Figure 2.6. Hypothesised evolution steps of BBis: *(a)* single-headed ancestor, *(b)* the first double headed inhibitor after gene duplication, *(c)* double-headed inhibitor after the second (partial) gene duplication and disulfide bridge rearrangement.

II.2 Crystal structure of the lima bean trypsin-chymotrypsin inhibitor

II.2.1 Purification, crystallisation and data collection

Background

The lima bean trypsin inhibitors were purified five years after the isolation of the soybean Kunitz-inhibitor from the acetone insoluble fraction (Tauber *et al.*, 1949) and characterised as heat-stable proteins that crystallise from acetone and inhibit trypsin and chymotrypsin, the latter to a lesser degree, though. Fraenkel-Conrat *et al*. (1952) improved the yield of the purification and the crystal quality by the application of ethanol instead of acetone and separated three inhibitor fractions with different activities. Haynes & Feeney (1967) applied chromatography methods to isolate several active fractions and observed that the lima bean extracts differ in peak pattern depending on the lima bean varieties used as sources. In 1971 two of the four variants were successfully sequenced (Tan & Stevens, 1971a and b; Stevens & Doskoch, 1973). These studies demonstrated that neither of these two variants represents a single homogenous molecular species (*Figure 2.7*) and exists as a mixture of isoforms. Variant VI exhibits variations at positions 10, 13, 20, 23, 35, 37, 62, 63 and 56, while variant I possesses shorter N- and C-termini.

```
domain 1          SGHHEHSTDZ  PSZSSKPCCB   HCLCTKSIPP   QCRCTDLRLD   SCHSA
                                                A                 S  F


domain 2          CKSCI   CTLSIPAQCV   CTBIBDFCYE   PCKSSHSDDD   NNN
                                         B   T
```

Figure 2.7. The amino acid sequence of the lima bean trypsin-chymotrypsin inhibitor (variant IV). Variations in var. IV are shown in yellow. Var. I lacks residues marked with grey.

The inhibitor molecule has an apparent molecular weight of about 9 kDa, but it undergoes concentration-dependent dimerisation. High speed sedimentation equilibrium studies showed that variants I and II do not self associate at moderate pH and protein concentration. On the contrary, variants III and IV display pH-dependent self-association properties. Variant IV is present as a monomer at pH 2.0, but dimerises at pH 7.0 and forms at least trimers at the intermediate pH of 4.65 (Birk, 1976).

The lima bean inhibitor blocks both trypsin and chymotrypsin with a 1:1 stoichiometry. The inhibitory activities of different variants towards trypsin are essentially identical (with a

K_I in the order of 10^{-10}), whereas some differences exist in the inhibition of chymotrypsin. While variant IV is a strong chymotrypsin inhibitor, variants I-III do not achieve more than 60-80% inhibition relative to their antitryptic activity (Kassell, 1976; Birk, 1976).

All variants are extremely stable to heat and extremes of pH, as they can be boiled for several minutes or stored in 10 mM NaOH or HCl at room temperature without loss of activity (Kassell, 1976).

Samudzi *et al.* (1997) reported crystallisation conditions and preliminary X-ray studies on crystals diffracting to 2.0 Å but were unable to solve the structure by molecular replacement. An apparent non-crystallographic twofold axis was observed in the self-rotation map, two molecules were thus predicted to be present in the asymmetric unit.

Assessment of the protein purity and first crystallisation trials

Lima bean trypsin-chymotrypsin inhibitor (LBTI) was purchased from Sigma (product number: T9378) as a crude product. To assess the protein purity, i.e. to establish the ratio of contaminants, the protein sample was submitted to SDS-PAGE analysis. LBTI was dissolved in double distilled water to a concentration of 20 mg/ml and boiled for 10 min in the presence of sodium-dodecylsulfate and dithiothreitol. Different amounts of the prepared samples were then run in a *GenePhor* PAGE-chamber on *ExcelGel Homogenous 25* gels at 300 V and finally silver-stained in a *Hoefer* automated gel stainer. As illustrated in *Figure 2.8*, multiple bands in lanes U1 and U2 corresponding to the crude material at different concentrations indicated the presence of protein type contaminants in a relatively high ratio.

The inhomogeneity of the LBTI-preparate thus demanded further purification steps and the high thermostability of BBis presented itself as the most evident possibility. A 20 mg/ml LBTI solution was subjected to a boiling water bath for 10 min and then quenched on ice for 15 min. Heat-unstable components were removed by centrifugation at 15 000 rpm for 45 min at room temperature. SDS-PAGE analysis of the resulting pellet and supernatant (prepared as described above) showed that the ratio of contaminants could be reduced to a tolerable level by heat-treatment of the protein solution and the heat-unstable components were concentrated in the pellet. This also confirmed the expected molecular weight of LBTI that was found to be around 10 kDa, in contrast to that reported by Samudzi *et al*.

The supernatant fraction was used for the first crystallisation trials. The crystallisation condition described by Samudzi *et al.* (1.0 M K,Na-tartrate, 0.1 M HEPES pH 7.5) was regarded as starting point for optimisation.
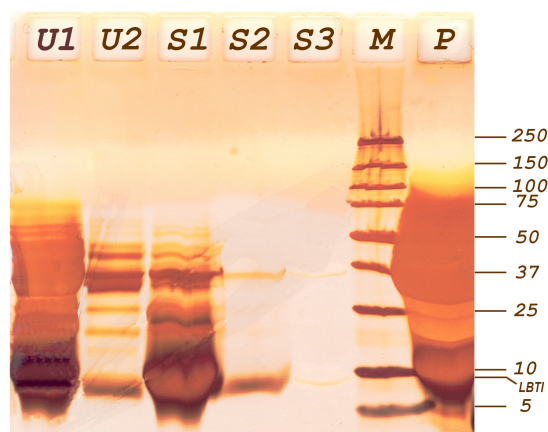
Figure 2.8. Silver stained SDS-PAGE gel of different LBTI-samples showing results of the boiling experiment. The first two lanes (*U1* an *U2*) represent the unboiled LBTI in different concentrations (50 and 5 µg). *S1-3* correspond to 50, 5 and 0.5 µg LBTI in the supernatant fraction after 10 min boiling, while *P* to 1 mm³ of the pellet. *M* shows a molecular weight marker. Note that LBTI is concentrated in the supernatant fraction after boling.

For crystallisation, equal volumes of the protein and precipitant solutions were mixed on siliconised cover slides and equilibrated with 0.5 ml of the precipitant solution in the reservoirs of 24-well tissue culture plates according to the hanging drop method. Fragile cubic crystals of 0.05-0.10 mm could be obtained from 0.8 M K,Na-tartrate, 0.1 M HEPES pH 7.5 in 3-7 days at room temperature (*Figure 2.9*). The crystals did not show birefringence when viewed between polar filters that agreed with their cubic space group described by Samudzi *et al.* (1997). Test measurements were first carried out at room temperature with crystals mounted in capillaries on a Bruker rotating anode equipped with Osmic multilayer focusing mirrors, a three-circle goniometer and a Bruker Smart6000 4K CCD camera. Diffraction properties were examined also under cryoconditions on shock-frozen crystals using various small molecular weight polyethylene-glycols (PEGs) and other alcohols as cryoprotectants. The crystals exhibited poor diffraction properties, as they only diffracted up to ~5 Å. Nevertheless, data collected to 5 Å allowed the determination of the space group. Autoindexing yielded a primitive rhombohedral cell with a=94.33 Å and $\alpha=\beta=\gamma=109.44°$ that could then be converted into an I centered cubic cell with the

$$a_{cubic} = \frac{a_{rhomb.} \cdot 2}{\sqrt{3}} \tag{33}$$

formula. This resulted in the cell constant a=108.92 Å and the space group I2$_1$3 or I23.

In spite of the optimisation of crystallisation and cryoconditions and the application of the macroseeding technique, neither the crystal size nor the diffraction limit could be improved, therefore further investigations aiming at the protein purity were decided.
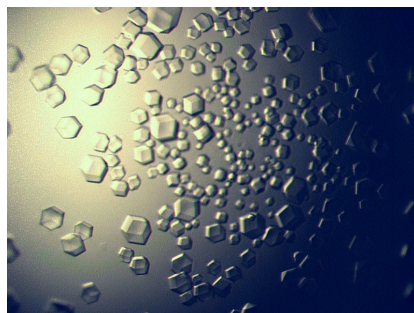
Figure 2.9. Crystals obtained using the supernatant fraction
of the boiled LBTI sample without further purification.

Purification

When the unboiled, thus crude LBTI-solution was subjected to an HPLC analysis on a C18 analytical column using an acetonitrile/water gradient, the peak pattern shown in *Figure 2.10.a* was obtained. In order to identify the peaks corresponding to LBTI this chromatogram was compared to that of the supernatant fraction of the boiled sample. As illustrated in *Figure 2.10.b*, three of the peaks were removed during the boiling process, the remaining two large peaks were therefore assigned as LBTI-peaks.

After optimisation of the acetonitrile/water gradient profile and selection of a more suitable HPLC column, the two LBTI-peaks could be resolved into three peaks, of which the later two slightly overlapped (*Figure 2.10.c*). Unfortunately, only the first peak could be separated from the others with a reasonable yield on a preparative scale.
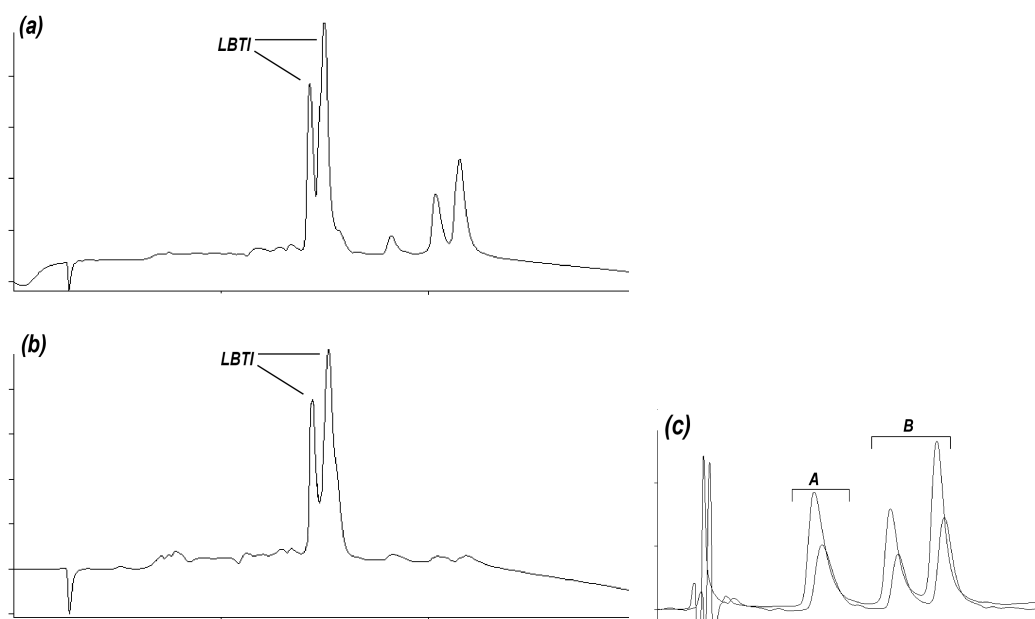
Figure 2.10. HPLC chromatograms of *(a)* unboiled LBTI and *(b)* the supernatant fraction of the boiled LBTI sample. *(c)* Resolution of the LBTI-peaks.

For the preparative-scale HPLC purification experiment a Nucleosil-100 C18 column was used with a 25-34% acetonitrile/water gradient. The eluted protein fractions were then lyophilised and stored at -80°C until crystallisation.

The observed peak pattern raises the question which of the four LBTI variants (or even isoforms of variants) are present in the crude product and in the purified fractions. According to the Sigma information board, the crude LBTI is isolated via the method reported by Tauber *et al.* (1949), so the crude product should contain all of the four variants (and their possible isoforms) and even some Kunitz-type inhibitors as well. This contradicts the information received from the Sigma Kundenservice (personal communication). According to that, the LBTI is purified by the Fraenkel-Conrat-method (1952) that already describes fractionation steps of crude LBTI.

For crystallisation, two types of LBTI sample were thus used: *fraction A*, containing the first peak, and *fraction B*, containing a mixture of the second and third peaks.

Crystallisation of *fraction A*

As crystallisation conditions of the boiled, but unpurified LBTI did not prove reproducible in the case of *fraction A*, new conditions were screened with the Hampton Crystal Screen Kits (Jancarick & Kim, 1991; Cudney *et al.*, 1994) using the hanging drop method at room temperature. Microcrystals could be obtained from conditions containing MPD or small molecular weight PEGs in high concentrations (over 15-20%). These conditions were further optimised with sitting drops to avoid the deformation of droplets originating from the lower surface tension of alcoholic solutions. Crystals typically of two crystal habits grew from 27-30% MPD, 0.02 M $CaCl_2$, 0.1 M Na-acetate/acetic acid buffer pH 4.6 in 3-7 days. Surprisingly, a marginal, only 3% difference in the MPD concentration led to a change of crystal habit as shown in *Figure 2.11*.

| (a) | (b) |
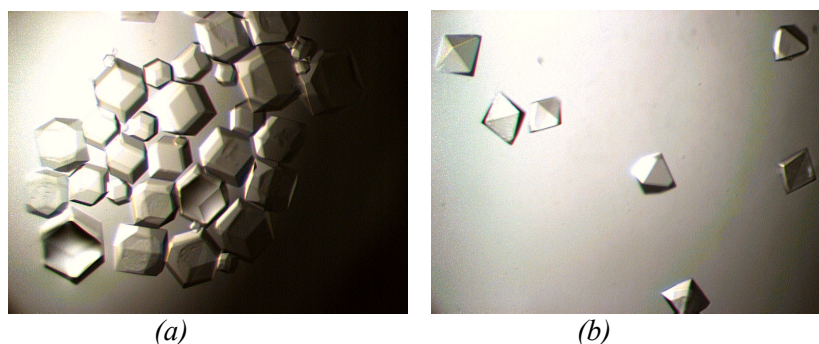
Figure 2.11. Crystals of LBTI *fraction A* grown from *(a)* 27% *(b)* 30% MPD.

Crystallisation of *fraction B*

In contrast to *fraction A*, *fraction B* could be successfully crystallised from the conditions found and optimised for the non-purified sample. However, minor adjustment of the protein and precipitant concentration, drop size and crystallisation technique was necessary for optimal crystal growth. Crystals of about 0.5 mm diameter could be obtained from 0.6 M Na,K-tartrate, 0.1 M HEPES pH 7.5 in 5 days using 6+4 µl droplets (*Figure 2.12*). These crystals were particularly fragile and tended to stick to the cover slide they grew on, rendering the crystal mounting more difficult or making it nearly impossible. Drops were therefore placed between two oil layers using non-miscible different density paraffin oils offered in the Hampton Containerless Crystallisation Kit.
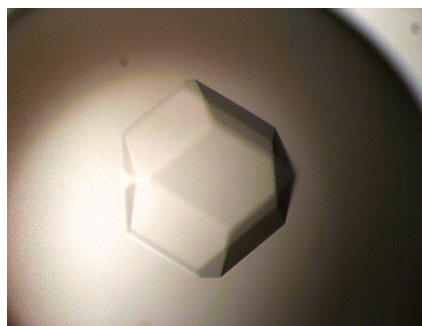
Figure 2.12. A crystal of LBTI *fraction B* with a diameter of 0.5 mm.

Data collection

Since crystals obtained from *fraction A* had very poor diffraction properties that could not be improved by further optimisation, only crystals resulting from *fraction B* were used in subsequent procedures. The apparent contradiction that crystals containing isoform impurity diffract better is accounted for by the biochemical finding that different variants exhibit different self-association properties, as discussed above. This may also explain the difficulties in the purification.

Fraction B crystals gave diffraction pattern to 3 Å, but because of the regular Gaussian reflection profiles and the high symmetry space group (which ensures respectable data redundancy within a reasonable measurement time) these crystals appeared to be perfect candidates for a sulfur-SAD experiment.

As the crystallisation condition, which these crystals were growing from, contained no cryoprotectant, soaking of the crystals was necessary for low temperature measurements. In first trials crystals were soaked directly into a cryobuffer containing 25% glycerol, but they

did not tolerate the high cryoprotectant concentration and cracked even when the soaking time was reduced. The glycerol concentration was therefore increased gradually and the crystals were soaked in several steps. After soaking they were mounted directly onto the goniometer head in loops and flash-cooled to 100 K in the nitrogen stream.

One 360° φ-scan and two 180° ω-scans at different φ angles were collected with 0.2°-slicing on the aforementioned in-house system with an exposure time of 1 min per frame to 3.0 Å resolution and a redundancy of 85.3. Although the respectably high redundancy and the thin slicing mode would suggest an unreasonably long data collection, measurement time did not exceed 2.5 days. Note that the two different scan types were of about the same quality in terms of redundancy, resolution and completeness. The dataset was processed using the Bruker programs *PROTEUM*, *SAINT*, *SADABS* and analysed with *XPREP* (Bruker Nonius, 2002). Subsequent tests indicated that about half the actual redundancy of 85.3 would have in fact sufficed for structure solution.

After the structure had been solved using the in-house data, a higher resolution (2.05 Å) synchrotron dataset was collected at the BW7B beamline at EMBL/DESY with a MAR345 image plate detector at 0.8431 Å wavelength and used for the final model building and refinement. The appreciably large difference in the resolution limit of the two datasets can undoubtedly be accounted for by the strength of the beamline and the conspicuous crystal size. In this case frames were taken in two consecutive passes at two different sample-detector distances and exposure times to avoid overloads in the inner resolution shells. 160 high resolution images were collected at 270 mm distance with 0.5° rotation steps, while the 40° low resolution pass was measured in 1° steps at 290 mm using a smaller read-out mode. These data were indexed and integrated with *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1997).

Data collection statistics are shown in *Table 2.3*.

Table 2.3. Data collection statistics.

	In-house	Synchrotron
Space group	I2$_1$3	I2$_1$3
Cell parameters (Å)	a=b=c=110.17	a=b=c=109.20
Data collected	393286	88491
Data unique	4608	13776
Resolution (Å)	3.00 (3.00-3.10)	2.05 (2.15-2.05)
Completeness (%)	99.7 (96.9)	98.5 (96.1)
Redundancy	85.3 (83.3)	6.3 (3.1)
R$_{int}$ (%)	8.91 (42.27)	5.67 (29.75)
Average I/σ(I)	76.25 (16.65)	12.43 (3.58)

(Values in parentheses refer to the last resolution shell.)

II.2.2 "Heavy atom" substructure solution and phasing

The structure was solved exploiting the weak anomalous signal in the in-house dataset originating from the inherently present sulfur atoms. Firstly, the φ- and ω-scans were scaled separately preserving their independence and the correlation coefficient between the signed anomalous differences for the two scan modes was calculated with *XPREP*. This indicated the anomalous signal to be significant up to 3.8 Å:

```
Inf - 8.0 - 6.0 - 5.0 - 4.6 - 4.4 - 4.2 - 4.0 - 3.8 - 3.6 - 3.4 - 3.2 - 3.0 Å

  94.9  96.0  92.2  80.7  81.8  74.3  64.6  56.0  28.1  13.2  10.3   1.3
```

All data were then scaled together to give the final dataset and to calculate the anomalous $I/\sigma(I)$ that was found to be in agreement with the resolution limit established from the correlation coefficient, as it fell below 1.3σ at 3.7 Å (Schneider & Sheldrick, 2002):

```
Inf - 8.0 - 6.0 - 5.0 - 4.6 - 4.4 - 4.2 - 4.0 - 3.8 - 3.6 - 3.4 - 3.2 - 3.0 Å

  7.25  5.11  4.11  3.44  2.61  2.31  2.28  1.71  1.52  1.34  1.27  1.10
  8.79  6.09  4.79  3.84  2.64  2.18  2.18  1.59  1.34  1.18  1.16  1.15
```

For the purpose of heavy atom location, data were truncated at 3.7 Å. Seven super-sulfur atoms, i.e. heavy atoms that correspond to the disulfide bridges at such low resolution, were searched with *SHELXD* using the following commands:

```
FIND 7              (number of sites to look for)
PATS                (Patterson seeding on)
MIND –3.5           (minimum distance between sites)
SHEL 100 3.7        (resolution)
```

The high mechanical sensitivity of the crystals suggested the presence of only one molecule in the asymmetric unit (with an estimated solvent content of 73%), therefore no translational symmetry was expected, so the Patterson seeding could be switched on with the command **PATS**. *SHELXD* found indeed seven super-sulfur atoms (*Figure 2.13*) and four noise peaks with high CC values and a clear cutoff in the peak-heights that indicated the previous assumption to be correct:

```
CC All/Weak 42.13 / 24.37
PATFOM   5.07

      x        y        z      height
  0.76541  0.51882  0.19047   99.90
  0.88009  0.57641  0.14575   97.93
  0.79163  0.55353  0.20591   91.82
  0.81370  0.58246  0.23161   90.84
  0.88845  0.56083  0.21072   82.85
  0.92895  0.60138  0.27455   74.23
  0.92101  0.58621  0.23861   72.21
  0.62165  0.60984  0.24436   33.58
  0.58220  0.79098  0.21448   29.81
  0.51028  0.83416  0.06059   29.72
  0.62122  0.52394  0.26385   29.06
```
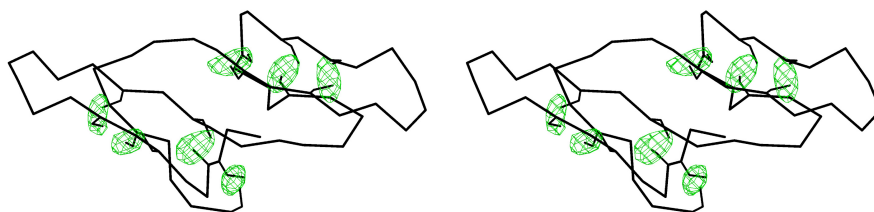
Figure 2.13. Stereo view of the backbone representation of LBTI showing the Cys residues and an anomalous Fourier map calculated with *SHELXE* phases at 3 Å.

At this stage also the correct space group, $I2_13$ could be determined, as solutions could only be obtained in this space group.

The substructure solution was input to *SHELXE* (Sheldrick, 2003) without any editing or heavy atom parameter refinement. Two *SHELXE* jobs were started for the two possible heavy atom enantiomorphs. The *contrast* and *connectivity* figures of merit unambiguously identified the correct hand (1.099 and 0.924, respectively, for the correct heavy atom enantiomorph and 0.201 and 0.750, respectively, for the wrong one). 20 cycles of density modification using the *sphere-of-influence* method implemented in *SHELXE* resulted in a very clear traceable electron density map even though the resolution was limited to 3.0 Å (*Figure 2.14.a*). The high solvent content was undoubtedly a major factor in determining the map quality.

When the synchrotron data became available, the phases were extended to 2.05 Å using *SHELXE*. *Figure 2.14.a-c* shows the progress of phasing.

(a)   *(b)*   *(c)*
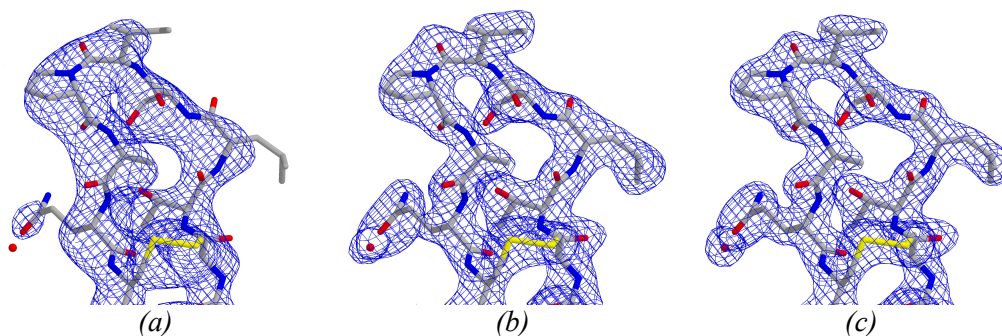
Figure 2.14. Electron density maps around the chymotrypsin binding loop of LBTI: *(a)* a map produced by *SHELXE* using the in-house data at 3 Å resolution *(b)* after phase extension to 2.05 Å *(c)* a $2F_o\text{-}F_c$ map after refinement against the 2.05 Å dataset.

II.2.3 Model building, refinement and validation

The initial model of LBTI was built by hand using the 3.0 Å electron density map output by *SHELXE* that was visualised with *XtalView* (McRee, 1999). After phase extension to 2.05 Å, this model was further adjusted to the electron density and refined against F^2 data of the 2.05 Å synchrotron dataset without resolution cutoff.

Autobuilding of the model was also tried with *ARP/wARP* (Perrakis *et al.*, 1999) and *RESOLVE* (Terwilliger, 1999) using the 2.05 Å map; it did not prove successful in either of the cases probably due to the relatively high number of disulfide bridges that often causes autobuilding programs to misinterpret these regions. Ambiguities concerning the Glu/Gln and Asp/Asn residues were tried to be clarified from the molecular environment of these side chains. Water molecules were added to the model with the program *SHELXWAT* and checked manually. In each refinement cycle *2mF$_o$-DF$_c$* and *F$_o$-F$_c$* type maps were displayed for manual editing of the model with *XtalView* at 1σ and ±3σ levels, respectively.

Throughout the refinement process angle distance, bond distance, chiral volume and planarity restraints were applied. The model was refined to an R-factor of 23.96% and an R$_{free}$ of 24.92%; 5% of the reflections were excluded from the refinement in thin shells as R$_{free}$-reflections using *XPREP*.

The final model contains 57 residues; the first 15 and the last 11 amino acids according to the sequence of variant IV (or the first 7 and the last residues according to the sequence of variant I) cannot be seen in the electron density and therefore they were not modelled. 78 water molecules were located, two of them on special positions. The residue occupying position 23 and reported to be either Ala or Leu was found to be Ser based on its shape in the electron density. Two Pro residues were observed to adopt the unusual *cis* conformation, both of them play essential roles in the shape determination of the reactive loops, as discussed later. The stereochemistry of the model was assessed with *PROCHECK* (Laskowski *et al.*, 1993). 98% of the residues were found to be in the allowed regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968), only Asp[40] appeared to be in the disallowed region having torsion angles Ψ=39° and Φ=73.4°.

A summary of the refinement statistics is shown in *Table 2.4*.

Table 2.4. Model refinement statistics.

Resolution range (Å)	77.19-2.05
Number of protein residues	57
Number of non-H atoms	501
Number of water molecules	78
R-factor/R$_{free}$ (%)	23.02/24.92
rmsd from ideal geometry	
Bond length (Å)	0.009
Angle distances (Å)	0.030
Ramachandran plot (%)	
Allowed region	98
Generously allowed region	0
Disallowed region	2

II.2.4 Structure description

Overall fold, comparison to related structures

The LBTI molecule adopts a "bow-tie"-fold canonical to the Bowman-Birk inhibitor family. It consists of two pseudo-symmetry related domains of three antiparallel β-strands connected by short loop regions and associated in two β-sheets (*Figure 2.15.a*). The first domain comprises residues 1-35 and 63-71, while the second one involves residues 36-62. The structure is strongly constrained by the seven disulfide bridges that are invariant in the BBi family. The disulfide bridges (between residue pairs 18/72, 19/34, 22/68, 24/32, 42/49, 46/61 and 51/59, see also *Figure 2.5.b*) together with a large number of intramolecular H-bonds stabilise the fold in such a manner that all the β-strands and even the reactive site loops lie in one slightly bent plane. This fold is rather conserved and the only difference between the members is the bending angle of the main molecular plane (*Figure 2.15.b*) that is correlated with the relative position of the antitryptic and antichymotryptic domains.

(a)

(b)
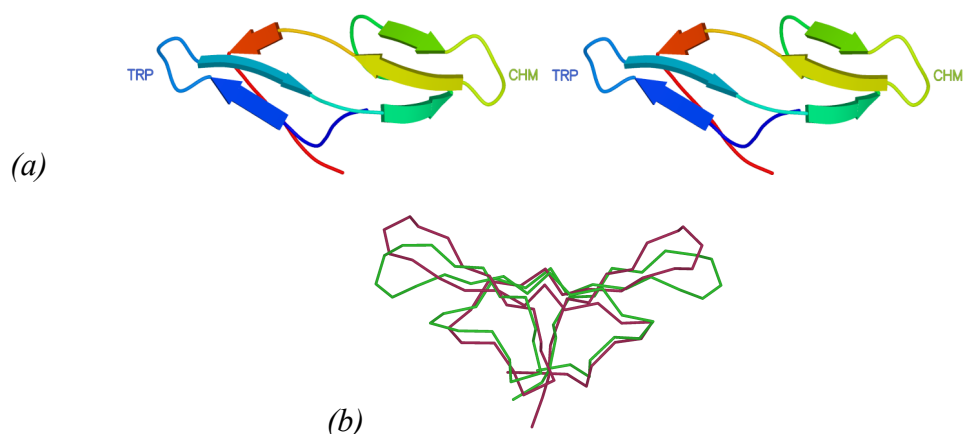
Figure 2.15. (a) Stereo view of the ribbon representation of the LBTI molecule. TRP: antitryptic loop, CHM: antichymotryptic loop. *(b)* Backbone representation of LBTI (green) in superposition with the snail medic seed BBi (brown). The plane defined by the β-sheet is horizontal in this figure.

Interdomain interactions within the molecule are restricted to a few weak solvent mediated H-bonds, though. This confers the molecule a certain degree of flexibility between the two domains whereas the overall conformation of the binding loops is preserved. Interestingly, such a bending of the molecule is not hindered by any of the disulfide bonds, as BBi molecules lack interdomain disulfide bridges. *Table 2.5* shows rmsd values resulting from the least squares superposition of different BBi structures with LBTI. As it can be concluded from the data, the observed differences are significantly higher when all main

chain atoms were used for superposition between the first and last Cys residues (to exclude the two termini that adopt different geometries due to their flexibility) than for only the distinct domains showing that the expected variability in the relative position of the domains can indeed be detected. The bending of the plane defined by the β-sheets was found to be the least marked in the case of the LBTI structure and the most pronounced in the snail medic seed BBi, as suggested by the most conspicuous rmsd values in *Table 2.5* (See also *Figure 2.15.b*).

Table 2.5. Rmsd values resulting from the least squares superposition of LBTI with related structures.

LBTI related structure	PDB ID	all main chain atoms	TRP loop only	CHM loop only
snail medic seed BBi	1MVZ	3.33	0.97	1.35
pea seed BBi	1PBI	1.12	0.39	0.48
tracy soybean BBi	1K9B	1.75	0.81	1.06
soybean BBi	1BBI	1.87	0.79	1.15
mung bean BBi	1DF9	1.25	0.71	0.63
adzuki bean BBi	1TAB	*	0.56	*

Synthesised BBi fragments, not deposited structures and distant relatives (i.e. monocotyledons) were omitted from this comparison. * 1TAB only contains the TRP domain.

As mentioned in *II.1.4*, the domains of BBis are developed through gene duplication, thus they are closely related to each other in structure. Least squares superposition of the main chain atoms of the TRP and CHM domain gives an rmsd of 0.52 Å proving the close structural similarity between them. This also accounts for the apparent twofold NCS in the self-rotation function.

Conformation of the reactive loops

The reactive loops span over nine residues that can be assigned to the Schechter-Berger notation (1967) as follows:

```
Schechter-Berger   P3    P2    P1    P1'   P2'   P3'   P4'   P5'   P6'
TRP domain         C24   T25   K26   S27   I28   P29   P30   Q31   C32
CHM domain         C51   T52   L53   S54   I55   P56   A57   Q58   C59
```

Each binding site loop of LBTI exhibits a heart-like shape (Suzuki *et al.*, 1987) when viewed from the direction shown in *Figure 2.16*. This characteristic structure is maintained by a *cis*-prolin residue at position P3', the disulfide bridge between positions P3 and P6', and a typical H-bonding pattern that involves residues Thr[25], Ser[27], Pro[29] and Gln[31] in the TRP loop and Thr[52], Ser[54], Pro[56] and Gln[58] in the CHM loop, i.e. residues at positions P2, P1', P3' and P5'. Two water molecules also take part in the loop stabilisation connecting P1' and P3'.
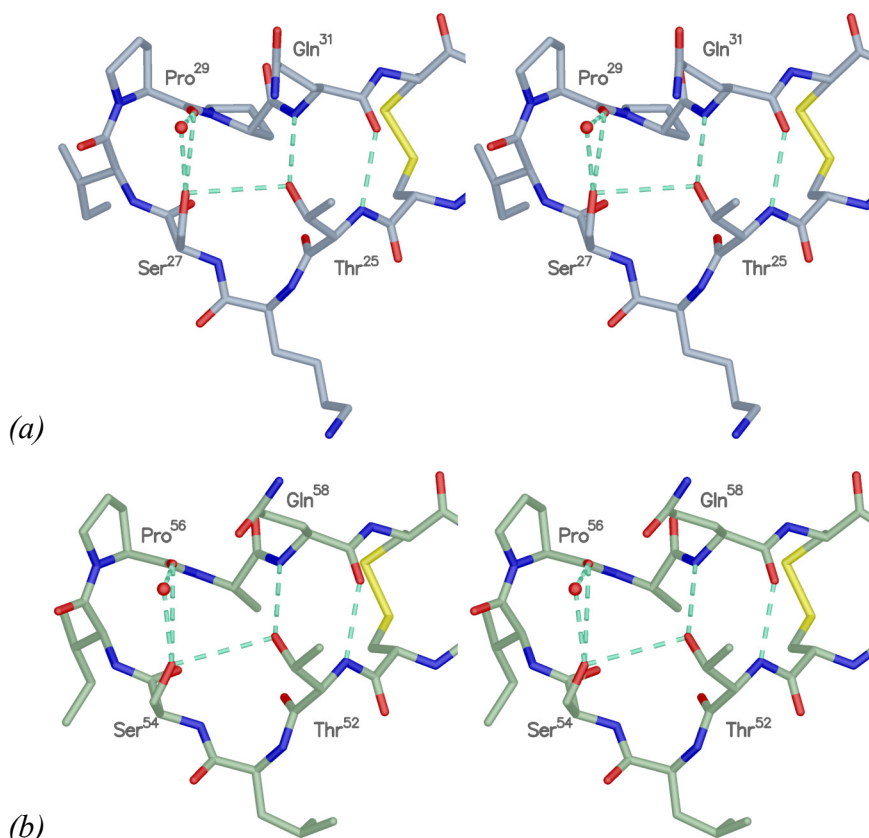
Figure 2.16. Stereo views of the reactive loops in LBTI: *(a)* antitryptic loop
and *(b)* antichymotryptic loop.

The conserved residues Thr at P2 and Ser at P1' play key role in keeping the residues around the scissile bond in an active conformation. The P2-Thr makes direct H-bonds with its O^γ-atom to the main chain N of P5'-Gln and the O^γ-atom of P1'-Ser, while with its N atom to the main chain O of P5'-Gln. In addition to the foregoing H-bridges, the O^γ of P1-Ser is bound to the O of P3'-Pro.

The remarkably high rigidity of both loops is reflected in their overall temperature factor values, as they are not any higher than that of the whole structure (44.0, 42.1 and 44.1 $Å^2$ for the TRP- and the CHM-loop and the whole structure, respectively) in spite of their peripheral position in the molecule.

Buried polar residues, integral solvent particles and the aromatic cluster

There are two buried polar side chains: His[43] and Asp[65] in the interior of the molecule facing each other but not interacting. Along these residues a long hydrophilic channel can be seen filled with water molecules that are well defined in the electron density. Besides these solvent particles, two buried water molecules are present as an integral part of the inhibitor molecule. They lie symmetrically around the intramolecular pseudo-twofold axis that relates

the two domains. Each water molecule donates two H-bonds to two main chain carbonyl groups (to Cys^{34} and Asp^{36}, Cys^{61} and Asp^{63} in the TRP- and CHM-loops, respectively) and accepts one from a peptide NH-group (from Cys^{61} and Cys^{34}) (*Figure 2.17*). Their B-factors, 36.2 and 31.6, are comparable to that of the whole molecule suggesting a tight binding. Similarly arranged water molecules were observed in other BBi structures as well (Voss *et al.*, 1996; Lin *et al.*, 1993; Suzuki *et al.*, 1987) and therefore they seem to be essential in maintaining the tertiary structure of the BBis. Since these water molecules as well as the buried polar residues lie at the interface between the two domains of the molecule, it can be hypothesised that they originate from the surface region of the ancestral single-headed inhibitor and became stabilising particles after the gene duplication but occupied internal positions in the molecule.
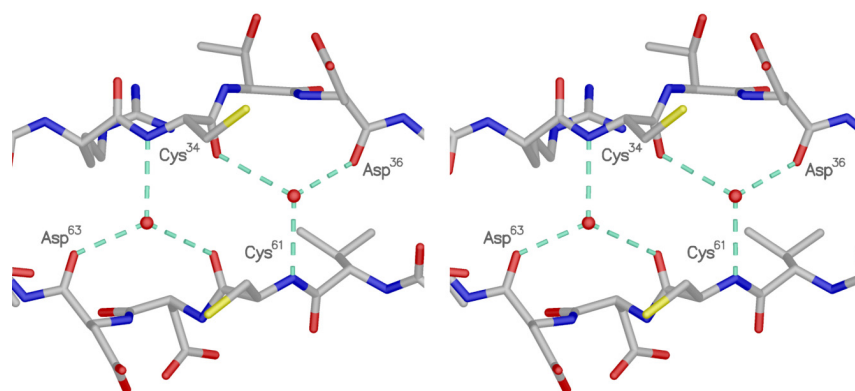
Figure 2.17. Stereo view of the two characteristic buried water molecules.

Another noteworthy feature of the structure is an aromatic cluster formed by His^{43}, Phe^{67} and Tyr^{69} arranged in a typical geometry in which pairs of aromatic side chains are almost orthogonal to one another and with centroids at distances shorter than 7 Å (*Figure 2.18*). The aromatic cluster is present in all the other BBis of known structure. This geometry is known to confer thermostability to many proteins (Burley & Petsko, 1988) in agreement with the observation that BBis are particularly insensitive to heat.
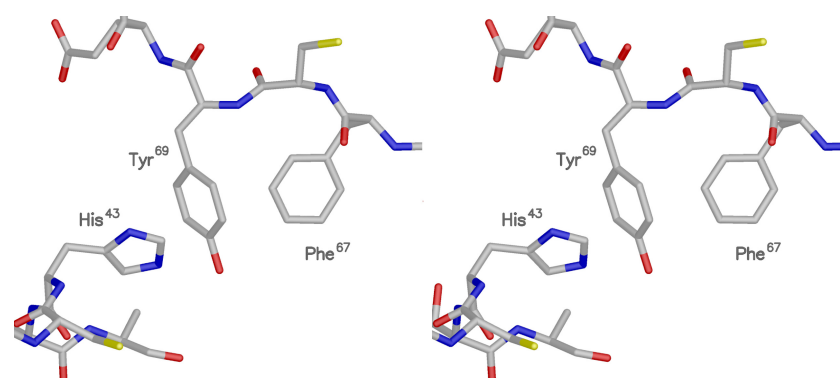
Figure 2.18. Stereo view of the aromatic cluster.

Dimerisation and crystal packing

Two molecules related by a crystallographic twofold axis form a compact globular shaped dimer in the crystal structure (*Figure 2.19*). It is possible that the monomers are in fact two slightly different isoforms (see *II.2.2*), but the resulting heterodimers are disordered about the twofold axis. The interface between the two monomers is formed by their β-sheets, with a rotation angle of 180º between the monomers, so that the binding-site loops are located on the surface of the dimer. The dimeric state is stabilised by a set of hydrogen bonds and a hydrophobic core. The residues His[21], Arg[33], Thr[35], Glu[58], Asp[62] and four water molecules play an essential role in the hydrogen bonding network. The hydrophobic core consists of the side chains of residues Ile[50], Val[60] and the methyl group of Thr[35]. Note that the distribution of hydrophobic and hydrophilic residues on the dimer interface is strongly asymmetric, as hydrophobic side chains concentrate mostly at the CHM domain while the hydrophilic ones at the TRP domain; in addition, interacting residues mostly occupy the surface of the CHM domain providing the dimer with further asymmetry. Since the peripheral hydrophobic residues are all shielded through the dimer formation and the accessible surface drops by 521 Å^2 upon dimerisation, the dimeric state probably corresponds to the physiological unit of free LBTI in solution.
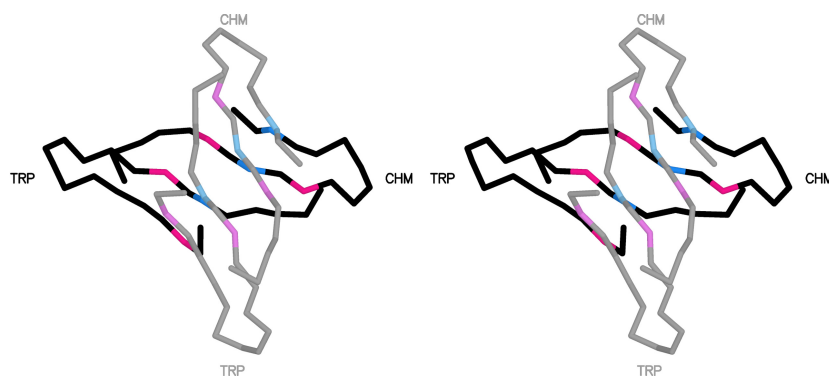
Figure 2.19. Stereo view of the dimeric state of LBTI. Hydrophobic residues participating in the dimer formation are marked with blue, hydrophilic ones are shown in pink. The N- and C-terminal tails and the connecting loop were removed for clarity.

As mentioned in *II.2.2*, about 75% of the unit cell is filled with disordered water constituting large, ca. 70 Å broad solvent channels, therefore the number of intermolecular interactions is rather limited. The binding loops extend from the globular core of the dimer offering the only possibility for further association. The molecules are connected to each other by a hydrophobic interface made up by residues Leu[23], Ile[28], Pro[33], Leu[37], Leu[39], Ile[55] and

Pro[56] in such a way that they associate to units of three dimers around the crystallographic threefold axis (*Figure 2.20*). The molecules in the unit cell could also be imagined as molecular chains parallel to the twofold screw axes. The C- and N-terminal tails of the peptide chain cannot be located in the electron density, as they are disordered pointing towards the large solvent channel and do not participate in intermolecular interactions.
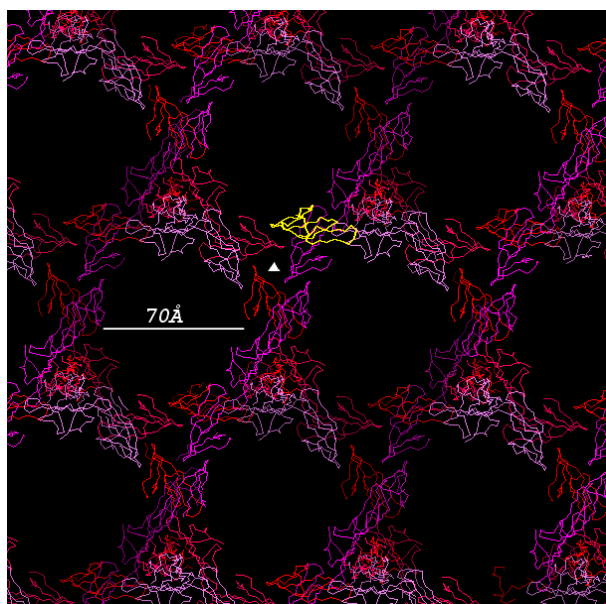
Figure 2.20. Crystal packing of LBTI. The white triangle represents a crystallographic threefold axis.

II.3 Crystal structure of LBTI in ternary complex with its target enzymes

II.3.1 Crystallisation and data collection

Background

Although members of both the BBi and the serine proteinase family exhibit favourable features in terms of stability and availability, only two ternary complex structures have been solved so far: that of the mung bean (*Phaseolus aureus*; Lin *et al.*, 1993) and the soybean (*Glycine max*; Koepke *et al.*, 2000) BBis, as shown also in *Table 2.2*. In both cases the two reactive loops of the inhibitors were found to be bound to trypsin molecules simultaneously. This represents the natural way of complexation for the mung bean BBi, which possesses basic amino acids at both P1 positions but unusual for the soybean BBi having a Leu residue at P1 in the second loop. Although Gaier *et al.* (1981) reported the crystallisation and

preliminary X-ray studies of soybean BBi in ternary complex with trypsin and chymotrypsin, this structure has not yet been published. Crystals described in this study diffracted up to 2.2 Å and were proved to contain all components of the ternary complex by SDS-PAGE.

Since structural basis of the Janus-faced enzyme preference of LBTI could only be roughly derived from the free inhibitor, LBTI was cocrystallised with trypsin and chymotrypsin for structure determination.

Crystallisation

Bovine β-trypsin was purchased from Sigma (product number: T1426), bovine α-chymotrypsin from Fluka (product number: 27270) as lyophilised powder. *Fraction B* of LBTI (see *II.2.1*) was used for cocrystallisation. Proteins were dissolved in double distilled water and used immediately for the subsequent steps. The ternary complex was prepared in two consecutive steps in order to avoid the previously described non-specific complex formation. First, a binary LBTI-TRP complex was formed by mixing equimolar amounts of TRP and LBTI. The solution was then allowed to stand at room temperature for about 15 min to ensure complete reaction. An equimolar amount of CHM was then added and allowed to react at room temperature for 15 min. The total ternary complex concentration was adjusted to 20 mg/ml. The solution was then divided into aliquots that were immediately frozen and stored at -80 °C.

Although crystallisation conditions were also reported for the soybean BBi-TRP-CHM ternary complex and the inhibitor molecule did not appear to be an important factor in crystal packing in the other two crystal structures of ternary complexes, a complete screening of crystallisation conditions was performed with the Hampton Crystal Screen Kits at room temperature using the hanging drop method. 2+2 µl drops were placed on siliconised cover slides and equilibrated with 0.5 ml of the precipitant solution in the reservoirs of 24-well tissue culture plates. Precipitation occurred and microcrystals grew in conditions containing medium molecular weight PEGs and acetate buffer pH 4.6 that interestingly corresponded to the crystallisation conditions reported for the soybean BBi-TRP-CHM complex. These conditions were further optimised by varying the precipitant concentration, drop size and composition using high throughput crystallisation plates available from *Nextal*. Very thin and fragile plate-like crystals of dimensions 0.05x0.2x0.4 mm^3 could be obtained from droplets containing 4 µl protein solution and 2 µl precipitant solution (0.1 M NaAc/HAc pH 4.6, 8% PEG4000) in 5-10 days (*Figure 2.21*).
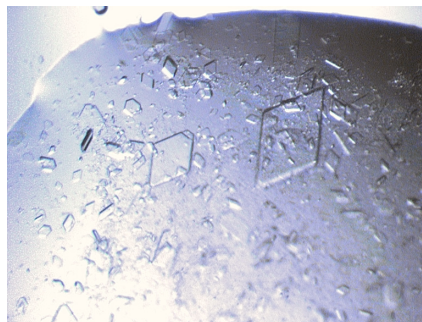
Figure 2.21. Thin plate-like crystals of the LBTI-TRP-CHM complex.

Data collection

Measurements were performed at low temperature and since the crystallisation condition did not contain any cryoprotectant agent, cryoprotection by soaking the crystals into a cryobuffer was required. Various small molecular weight PEGs and alcohols were tried and finally 20% glycerol was found to be optimal for cryoprotection. Crystals were first transferred with a loop into a droplet containing the precipitants in the original concentration and additionally 20% glycerol. After about 1 min soaking time crystals were mounted in loops and shock-frozen in the nitrogen stream to 100 K.

First test images were collected for indexing purposes on a Bruker rotating anode generator equipped with Osmic focusing mirrors, a single ω-axis goniometer and a MAR345 image plate detector. The crystals displayed poor and strongly anisotropic diffraction properties, as they only diffracted to 3.2-3.5 Å with 20 min exposure time and the reflections were severely elongated in certain crystal orientations, as observed in many cases of thin plate-like crystals. Relatively large loops were therefore chosen for crystal mounting in order to avoid any direct contact between the crystal and the fibre that has been shown to cause plate-shaped crystals to diffract even more anisotropically.

Diffraction data were collected at the X11 beamline at EMBL/DESY at 0.8527 Å wavelength on a marCCD detector in two passes. High resolution frames were measured at 210 mm crystal-detector distance in a 100° scan with 0.3° steps, while 300 low resolution images were taken at 250 mm with also 0.3° steps. Unfortunately, the resolution of the dataset was limited to 2.6 Å; reflections at 2.3 Å resolution could be observed in favourable crystal orientations, though. Technical difficulties in positioning the beamstop can account for the fairly low completeness in the inner resolution shells. Data were indexed, integrated with *DENZO* and scaled using *SCALEPACK*. Data collection statistics are summarised in *Table 2.6*.

Table 2.6. Data collection statistics.

Space group	$P2_12_12_1$
Cell parameters (Å)	a=51.30
	b=101.32
	c=108.62
Data collected	129748
Data unique	17274
Resolution (Å)	2.6 (2.70-2.60)
Completeness (%)	95.3 (99.0)
Redundancy	7.16 (7.40)
R_{int} (%)	8.24 (36.95)
Average I/σ(I)	16.88 (7.79)

(Values in parentheses refer to the last resolution shell.)

II.3.2 Structure solution

The previously known structures of bovine β-trypsin (PDB ID: 1BJU) and α-chymotrypsin (PDB ID: 4CHA) were used to solve the phase problem by means of molecular replacement. Higher resolution information that may lower the success of the MR method were excluded by truncating the data at 3.5 Å resolution. Solvent and inhibitor molecules of the TRP and CHM models were also removed. The program *AMORE* (Navaza, 1994) was applied to locate the search models in the asymmetric unit in two main steps. The radius of the integration sphere in the rotational search was set to 20 Å in both cases to avoid intermolecular (cross) Patterson vectors.

First, the TRP molecule was located in the asymmetric unit. The best 20 peaks resulting from the rotational search were input to the translational search and the solutions were then objected to a rough rigid body refinement. The first two peaks of the rigid body fitting function were in sharp contrast to the other peaks in terms of correlation coefficient and R-factor:

```
Solution  Model_no   ALPHA    BETA   GAMMA     TX      TY      TZ    CorrF    RFac

   1         1       148.750  80.450 167.310  0.999   0.377   0.087  33.500   48.100
   2         1        49.060  67.120  30.050  0.425   0.191   0.335  19.600   52.300
   3         1        37.930  49.840  81.290  0.147   0.250   0.033  11.700   54.600
   4         1        68.280  90.000 351.200  0.047   0.414   0.328  11.700   54.200
   5         1        28.200  70.300  42.130  0.315   0.234   0.256  10.800   54.400
   etc.
```

The first peak with a correlation coefficient of 33.5 and an R-factor of 48.1 corresponds to the correct position and orientation of the TRP molecule in the asymmetric unit. Although the second peak possesses an essentially smaller CC and a higher R-factor, it still represents a significantly better solution than the next ones. This pseudo-solution was later identified as the position and orientation of the CHM molecule, and its appearance can be attributed to the high structural similarity of these enzyme molecules.

In the next step, the TRP molecule was fixed and only the CHM molecule was searched separately. Solutions from the translation search together with the TRP model located in the first step were then rigid body fitted that yielded a solution pair with a CC of 63.6 for the CHM molecule far above the first non-solutions following it:

Solution	Model_no	ALPHA	BETA	GAMMA	TX	TY	TZ	CorrF	RFac
1	2	147.870	80.570	166.140	1.000	0.378	0.088	35.900	47.200
2	1	53.320	69.170	196.070	0.414	0.179	0.334	63.600	36.900
3	2	147.870	80.570	166.140	1.000	0.378	0.088	35.900	47.200
4	1	40.310	69.330	309.060	0.296	0.985	0.345	26.500	50.400
5	2	147.870	80.570	166.140	1.000	0.378	0.088	35.900	47.200
6	1	177.360	71.430	192.450	0.764	0.867	0.755	25.700	51.000
etc.									

The high CCs and the at least in the case of an initial model acceptable R-values suggested that the conformation of the enzyme molecules remained unchanged upon complexation and therefore it was also confirmed that they could be used as models for the first refinement step without major rebuilding.

Finally, the LBTI molecule was tried to be located in the asymmetric unit containing both TRP and CHM molecules on fixed positions and using the LBTI model described in *II.2* after the removal of the solvent molecules. In this case the rotational search proved unsuccessful and so the subsequent translational search was not performed. Considering the fact that the LBTI molecule constitutes only about 12% of the protein content of the asymmetric unit, the failure of the LBTI location can be regarded as to be expectable. Fortunately, the LBTI molecule was clearly visible in the sigmaA-weighted electron density map calculated from the solution containing both TRP and CHM molecules (*Figure 2.22*). LBTI was thus fitted to the density by hand using *XtalView*.

(a)   **(b)**
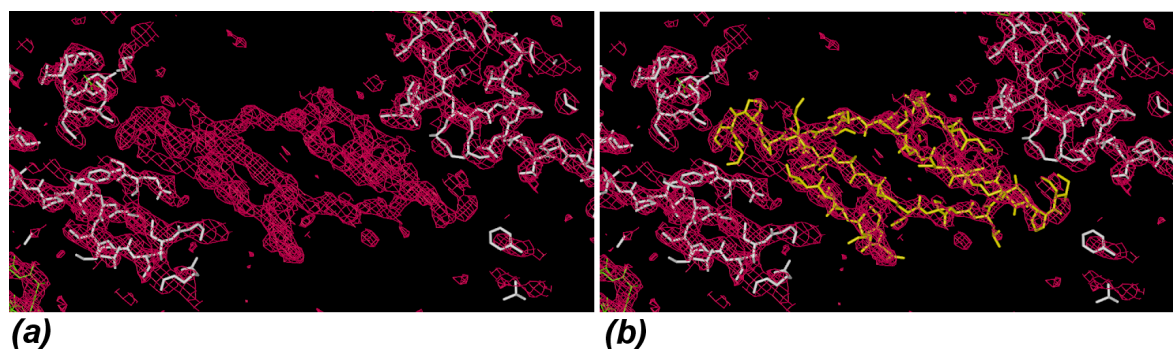
Figure 2.22. A section of the SigmaA weighted electron density map calculated directly after structure solution with the TRP and CHM molecules (white) showing the position of LBTI in the ternary complex as displayed in *XtalView*, ***(a)*** with and ***(b)*** without the LBTI molecule (yellow) fitted to the density.

II.3.3 Model building, refinement and validation

Since the structure could be successfully solved by molecular replacement and even the LBTI molecule was located unambiguously in the density, the model building step could be narrowed down to the completion of the solvent model. In order to improve the LBTI position, all three molecules were subjected to a rigid body refinement using the *CNS* software package (Brünger *et al.*, 1998).

In subsequent refinement cycles, simulated annealing was applied with torsion angle dynamics using *CNS* to reduce the number of parameters and the degree of overfitting of the fairly low resolution data. The annealing was started at 2500 K and the temperature was gradually reduced to 300 K in 25 K steps. The resulting model was then further refined with the CGLS algorithm implemented in *CNS* using F data between 40 and 2.6 Å resolution. Ordered water molecules were located with the automated water pick option in *CNS* and later examined manually using *XtalView*. Finally, B-factors were optimised by refining two B-factors for each residue: one for the main chain and one for the side chain atoms. In all refinement cycles, $2F_o$-F_c and F_o-F_c maps were displayed with *XtalView* at 1σ and $\pm3\sigma$ levels, respectively, for manual adjustment of the model. After several refinement cycles the R-factor reached 20.69%.

The final model includes one TRP, one CHM and one LBTI molecule, i.e. 555 amino acid residues and additionally 119 water molecules. Most of the non-Gly residues showed torsion angles typical for the allowed regions of the Ramachandran-plot, only one residue of LBTI was found in the disallowed region, but its torsion angles could be confirmed based on the uncomplexed structure (*Table 2.7*).

Table 2.7. Model refinement statistics.

Resolution range (Å)	40.0-2.60
Number of protein residues	555
Number of non-H atoms	3919
Number of water molecules	119
R-factor/R_{free} (%)	20.69/26.64
rmsd from ideal geometry	
Bond length (Å)	0.006
Bond angles (°)	1.65
Ramachandran plot (%)	
Allowed region	99.6
Generously allowed region	0.2
Disallowed region	0.2

II.3.4 Structure description

Structure overview in comparison with related structures

The ternary complex is arranged approximately as a 90 Å long ellipsoid consisting of a TRP and a CHM molecule bridged by the double-headed LBTI molecule in the middle. The complex displays a pseudo-twofold axis inherent to LBTI. This assembly is stabilised by the canonical enzyme-inhibitor interactions and additional enzyme-enzyme contacts. The binding loops of LBTI are embedded into the active site of the corresponding enzyme molecule in such a way that the N-terminal half of the β-hairpin of LBTI forms an extension of the β-sheet consisting of the β-strands 180-184, 210-217 and 225-230 (according to CHM numbering, which will be used throughout this chapter) in the C-terminal β-barrel of the enzymes. The enzyme-inhibitor interactions far extend this primary contact region, as both LBTI domains are encompassed by the β-hairpin 91-93, the loop 54-64 and 174-178 of the enzymes (*Figure 2.23.a*).

The two P1 residues, Lys^{26} and Leu^{53}, are ca. 34 Å apart allowing the two reactive site loops to bind and inhibit two proteinase molecules simultaneously and independently. A distance of 39 Å between the P1 residues was, however, predicted to be the minimum required for simultaneous binding (Li de la Sierra *et al.*, 1999). In the LBTI-TRP-CHM complex the relatively small bending angle of the main molecular plane of LBTI enables the binding of both proteinases in spite of the short P1-P1 distance. Although the binding loops are comparatively close to one another, their relative orientation allows the enzyme molecules to bind without overlapping.

Interestingly, enzyme-enzyme interactions could also be detected: H-bonds were found between the side chain amide group of Asn^{95}(CHM) and the main chain oxygen of Gly^{174}(TRP) and between the side chain oxygen of Thr^{174}(CHM) and the side chain nitrogen of Asn^{95}(TRP) (*Figure 2.23.b*).

As described by Hirono *et al.* (1984) and later by Koepke *et al.* (2000), three regions can be discerned in the enzyme-inhibitor interface. The primary contact region comprises residues of the binding loops around the P1 position (P4-P2'), while the secondary contact region consists of side chains around the P8'-Cys projecting out from the three-stranded β-sheets of the LBTI domains. Koepke *et al.* also mention residues corresponding to Leu^{37}(LBTI) and Ile^{64}(LBTI) as the tertiary contact region. These enzyme-inhibitor contact regions will be discussed in the following subsections.

(a)

(b)
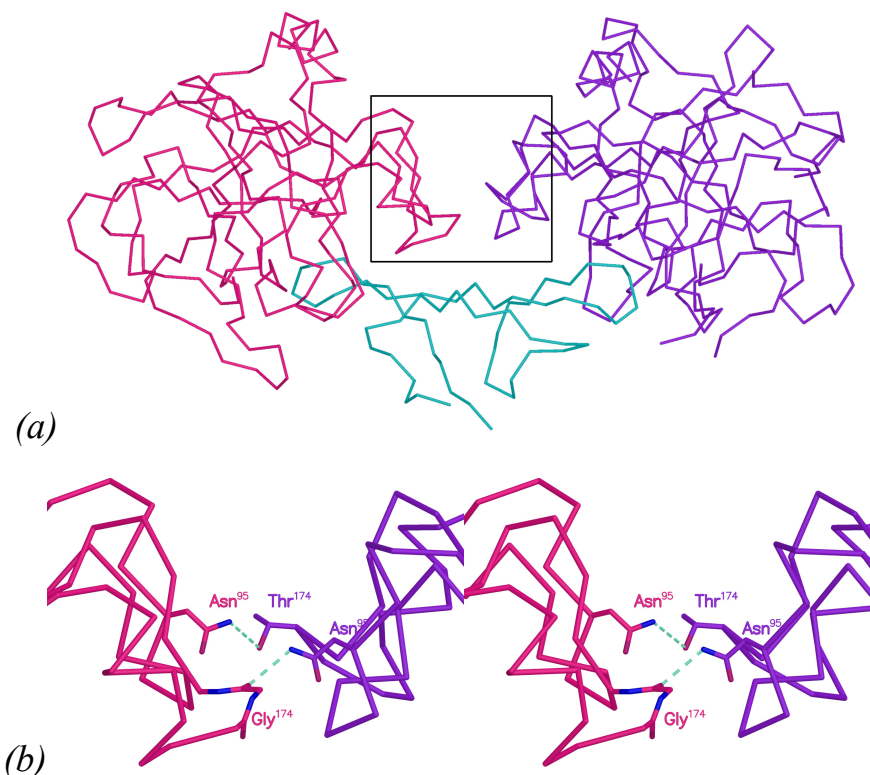
Figure 2.23. *(a)* Backbone representation of the LBTI-TRP-CHM ternary complex. The marked area is shown enlarged in *(b)*. ***(b)*** Stereo view of enzyme-enzyme interactions.

LBTI-TRP interactions

The primary contact region. The contacts between TRP and the antitryptic loop of LBTI are dominated by the interaction of the P1-Lys[26] with TRP residues forming the specificity site. The long side chain of Lys[26] is deeply encapsulated in the active pocket and interacts with the side chain of S1-Asp[189](TRP) through a water molecule and with the main chain oxygen of Ser[190](TRP). The carbonyl oxygen of the P1 residue projects into the oxyanion hole and is fixed by two H-bonds to the N-atoms of Gly[193](TRP) and Ser[195](TRP), as expected for a substrate-like binding mode. The distance between the carbonyl carbon atom at the scissile bond and the side chain oxygen of Ser[195](TRP) is only 2.7 Å that refers to the "sub-van der Waals" contact mentioned in *II.1.7* (*Figure 2.24*). Unfortunately, the low resolution of the dataset does not allow further interpretations concerning the typical out-of-plane displacement of the O-atom of the P1 residue.

The H-bridge pattern between the P1-P3 and S1-S3 stretches resembles that of an antiparallel β-sheet. Three such H-bonds could be observed, namely between the atom pairs N-P1-Lys[26](LBTI)─O-Ser[214](TRP), O-P3-Cys[24](LBTI)─N-Gly[216](TRP) and N-P3-Lys[24](LBTI)─
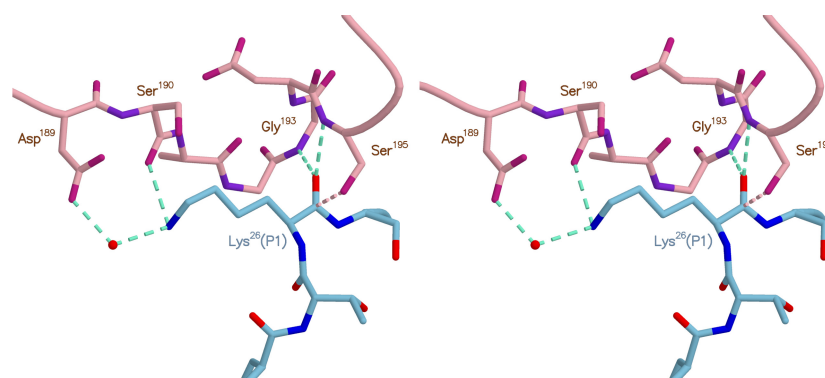
Figure 2.24. Stereo view of residues interacting at the P1 residue of the TRP binding domain. (TRP: pink, LBTI: blue)

O-Gly^{216}(TRP). In addition to these, the P4-Ser^{23}(LBTI) forms a H-bond by means of its O$^{\gamma}$ atom with the side chain N of Gln^{175}(TRP). The P2 residue Thr^{25}(LBTI) is not involved in this H-bonding network, it is, however, in hydrophobic contact with Leu^{99}(TRP) through its methyl group. As an extension of the primary region, the carbonyl oxygen of P5-Cys^{22}(LBTI) binds the O$^{\gamma}$ atom of Ser^{217}(TRP) (*Figure 2.25*).
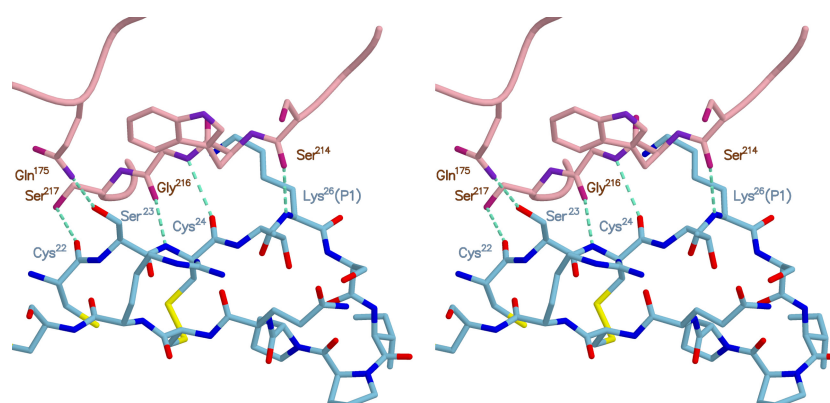
Figure 2.25. Stereo view of residues interacting at the N-terminal part of the TRP binding domain. (TRP: pink, LBTI: blue)

On the C-terminal side of the scissile bond only P2'-Ile^{28}(LBTI) makes a contact to the TRP molecule, as its N atom is within a H-bonding distance to the O atom of Phe^{41}(TRP).

The secondary and tertiary contact regions. Not all of the side chains of residues assigned as secondary and tertiary contact regions are compelled by intermolecular forces to point towards the enzyme molecule, in contrast to those in the primary region. Therefore the number of interactions requiring the involvement of side chains in the secondary and tertiary contact regions can be expected to be lower than in the primary region. Indeed, only four side chains: Gln^{31}, Arg^{33}, Thr^{35} and Ile^{64} in LBTI were observed to be in an interacting orientation, from which only Gln^{31} and Arg^{33} are involved in LBTI-TRP contacts. The side chain amide N

of Gln[31](LBTI) is within H-bonding distance to the O of His[57](TRP), while its side chain carbonyl oxygen makes interactions with the carbonyl oxygen of Ser[99](TRP) and the OH group of Tyr[94](TRP) through water molecules. The guanidinium group of Arg[33](LBTI) forms H-bonds with the main chain O atom of Ser[96](TRP) as well as Asn[97](TRP) (*Figure 2.26*).
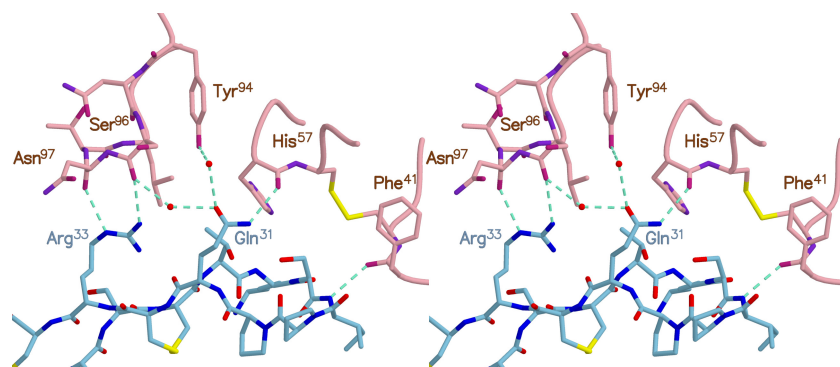
Figure 2.26. Stereo view of residues interacting at the C-terminal part of the TRP binding domain. (TRP: pink, LBTI: blue)

LBTI-CHM interactions

The primary contact region. As both TRP and CHM belong to the same proteinase family and are closely related in structure, the principles of inhibitor binding are essentially identical for TRP and CHM in the primary region. Similarly to TRP-LBTI contacts, P1-P3 residues form H-bonds with S1-S3 subsites that bear a resemblance to the H-bond pattern of an antiparallel β-sheet. This includes contacts between the following atom pairs: N-P1-Leu[53](LBTI)─O-Ser[214](CHM), O-P3-Cys[51](LBTI)─N-Gly[216](CHM) and N-P3-Cys[51](LBTI)─O-Gly[216](CHM) (*Figure 2.27*).
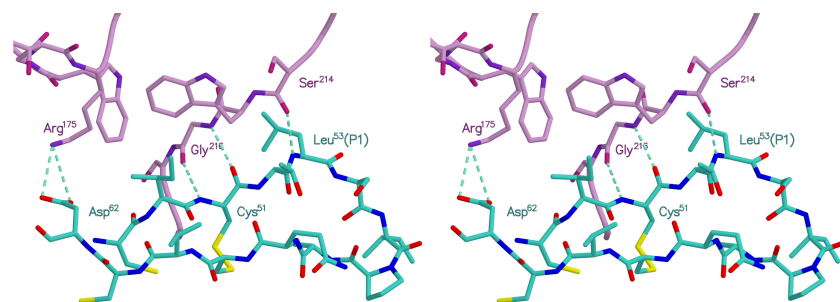
Figure 2.27. Stereo view of residues interacting at the N-terminal part of the CHM binding domain. (CHM: violet, LBTI: turquoise)

The side chain of P1-Leu^{53} is embedded into the active pocket of CHM surrounded by the residues Met^{192}, Gly^{193}, Val^{213} and the disulfide bridge between Cys^{191}-Cys^{220} forming a hydrophobic binding site in CHM. The carbonyl O of P1-Leu^{53} points towards the oxyanion hole making H-bonds to the N atoms of Gly^{193} and Ser^{195} of CHM, corresponding to the canonical inhibitor binding. The OH group of Ser^{195}(CHM) of the catalytic triad, similarly to that in TRP, is in "sub-van der Waals" contact with the carbonyl carbon at the scissile bond, as the distance between these two atoms is only 2.6 Å (*Figure 2.28*).
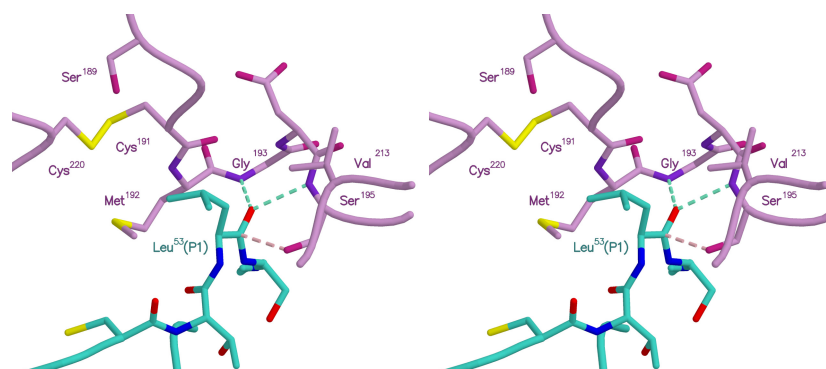
Figure 2.28. Stereo view of residues interacting at the P1 residue of the CHM binding domain. (CHM: violet, LBTI: turquoise)

The secondary and tertiary contact regions. The interacting surface of the LBTI antichymotryptic domain is characterised by hydrophobic side chains, such as Leu^{37}, Ile^{50} and Val^{60}. These residues together with the methyl group of Thr^{52} make hydrophobic contact with the area consisting of side chains of Leu^{97}, Ile^{99}, Trp^{172} and Trp^{215} in the CHM molecule. The long hydrophobic carbon chain of Lys^{175}(CHM) is involved in these hydrophobic interactions, whereas its guanidinium group participates in salt bridge formation with Asp^{62}(LBTI). Interestingly, Gln^{58}(LBTI) that corresponds to Gln^{31}(LBTI) in the TRP-binding domain, does not play any role in CHM-binding.

Strctural changes upon complexation

It is very instructive to compare the free LBTI molecule with that in the TRP-CHM-complex, as this comparison may reveal possible side chain or even main chain reorientations upon binding to the target enzyme. The rmsd resulting from the superposition of all main chain atoms was only 0.49 Å showing small but not negligible changes in the geometry of the inhibitor molecule.

Significant side chain reorientation was observed to occur in the case of residues that are involved in either the dimer formation or the enzyme binding. More pronounced differences were detected at the TRP-binding domain where the numerous H-bonds required precise positioning of partner atoms. For example, the OH group of Ser[23](LBTI) points towards Gln[175](TRP) but displays a different conformation in the uncomplexed structure. Similarly, the guanidinium group of Arg[23](LBTI) that is in a H-bond with Thr[35](LBTI) in the LBTI-dimer gets displaced in order to build two H-bridges with Asn[97](TRP) and Ser[96](TRP). A further example is provided by the side chain of Lys[26](LBTI), i.e. the P1 residue of the antitryptic loop. It possesses rather high B values in the uncomplexed structure showing its high flexibility. In the complex, however, it is kept in a different orientation by electrostatic interactions with Asp[189](TRP) and displays temperature factors similar to its neighbouring residues. Careful examination of the CHM-domain reveals variations in the conformation of side chains at the enzyme-inhibitor interface. The most striking difference was detected in the orientation of Asp[62](LBTI) side chain, which was rotated around its C^{α}-C^{β} bond in favour of a closer interaction with Lys[175](CHM). Also, conformation of side chains of hydrophobic residues, such as Ile[50](LBTI), Leu[53](LBTI), Ile[55](LBTI) and Val[60](LBTI) vary significantly, to a lesser extent, though.

Surprisingly, in spite of the well-known rigidity of the structure, not only side chain reorientation occurs upon complex formation. The LBTI molecule undergoes a slight distortion, as the TRP-binding loop moves towards the N- and C-terminal tail of the molecule whereas the CHM-binding loop gets displaced in the other direction, i.e. away from the connecting loop 39-48, so that the planarity of the β-sheets remains preserved. The difference between the free and complexed binding loops was found to be ~1.5-2.0 Å at the *cis*-Pro residue.

This variance in the overall molecular geometry can probably be accounted for by the different contact surface areas of the two LBTI domains. Residues involved in enzyme-

(a)   *(b)*
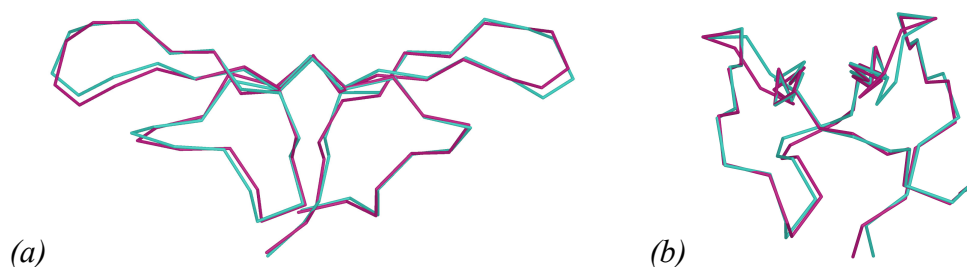
Figure 2.29. Backbone representation of LBTI in the LBTI-TRP-CHM complex (green) superimposed with the model of the free LBTI molecule (purple). ***(b)*** shows the molecules after 90° rotation of ***(a)*** around the vertical axis.

inhibitor interactions are more concentrated around the active loop in the case of the TRP-domain, although the number of interacting residues is nearly the same as in the CHM domain. The binding interface therefore corresponds to a larger area in the CHM-domain. The distortion in the CHM-binding loop may thus be in favour of a better fit to the LBTI-CHM complementary regions. Unlike this, the TRP-binding residues occupy a smaller area and thus bending of the reactive loop in the opposite direction may be required for a better enzyme-inhibitor positioning.

Crystal packing

Unlike the other structures presented in this thesis, crystals of the LBTI-TRP-CHM complex possess a solvent content of about ~50%, typical of protein crystals. The asymmetric unit contains only one complex that touches the symmetry equivalents at two main points of contact in such a way that the LBTI molecule plays only a minor role in lattice formation. On one hand, the hydrophobic side chains of Leu^{163}(TRP), Tyr^{184}(TRP) and Leu^{185}(TRP) are in contact with Ala^{185}(CHM) and Gly^{187}(CHM). This interaction is further strengthened by H-bonds between residue pairs Glu^{186}(TRP)—Thr^{222}(CHM) and Thr^{134}(TRP)—Tyr^{171}(CHM). By means of these contacts the molecules build chains along the *a* axis. The chains are cross-connected at a rather small LBTI-CHM interface in *y* and by one single H-bond between Ser^{19}(TRP) and Ser^{244}(TRP) in the *z* direction. Two H-bonds and a salt bridge could be located between CHM and a symmetry equivalent LBTI molecule: Cys^{68}(LBTI) and Glu^{70}(LBTI) form H-bridges with Asn^{204}(CHM), while Asp^{66}(LBTI) is in electrostatic interaction with the N-terminal Cys^{1}(CHM). Similarly to the uncomplexed LBTI structure, the N- and C-termini of LBTI extend to the solvent cavity and due to their flexibility cannot be seen in the electron density (*Figure 2.30*).
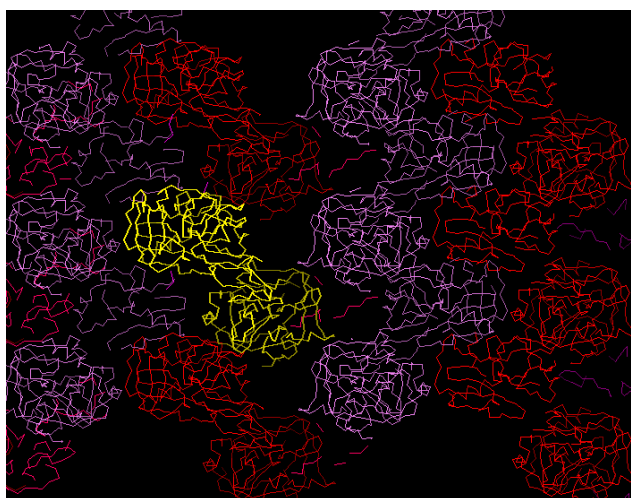
Figure 2.30. Crystal packing of the LBTI-TRP-CHM complex.

II.4 Biological implications

In vivo role and medical application of members of the BBi family

Members of the BBi family function as natural protective agents against insect proteinases (Sakal *et al.*, 1989) as suggested by wound-induced expression of BBi genes in plants (Eckelkamp *at al.*, 1993). However, other studies demonstrated that these inhibitors play a role in sulfur-storage, in blocking endogenous proteinases (Tan-Wilson *et al.*, 1985) as well as in the neutralisation of reduction equivalents (Kobrehel *et al.*, 1991).

On the other hand, BBis are of considerable interest in medical practice (Flecker, 1993), as they were shown to exert anticarcinogenic and immunstimulating effects. BBi extracts of soybean have already reached phase II in clinical trials in the treatment of oral premalignant lesions and proved effective against oral leukoplakia (Armstrong *et al.*, 2003). Furthermore, BBis are capable of suppressing chemical-induced carcinogenesis in rats, mice and hamsters, but in human organ systems as well, like in colon, liver, lung and oesophagus epithelium and connective tissues (Kennedy, 1998a). They were also demonstrated to improve radiotherapy of various tumours by selective stimulation of DNA-repair mechanism in non-carcinogenous cells (Dittman *et al.*, 2003) and reduction of the carcinogen-induced gene amplification to normal levels (Kennedy, 1994) as well as DNA-damage caused by superoxides (Tolcher *et al.*, 2001). BBis prevent tumour invasion and metastasis (Ware *et al.*, 1999) and potentiate the cisplatin-induced cytotoxicity in human ovarian and lung cancer cells (Zhang *et al.*, 1999).

Side effects can, however, be obstructive in clinical use of BBis, as they can generate immune response when administered intravenously in high concentrations (Maki *et al.*, 1994). BBis were found to cause pancreatic hypertrophy and the inhibition of growth in rats, but later this was attributed to the phytic acid and saponin content of the soybean extracts used (Liener, 1995).

Fortunately, half of the BBi is taken up into the blood stream when administered orally, and the incidence of provoking immune responses is rare (Kennedy, 1998a). Low molar concentrations of BBi extracts were shown to be sufficient as chemopreventive agent that does not have to be present continuously to suppress carcinogenesis. BBis are thus of high potential for medical applications (Kennedy, 1998b).

Yavelov *et al.* (1993) hypothesised that the anticarcinogenic effect of BBis originates from inhibition of serine proteinases with chymotrypsin-like substrate preference; Koepke *et al.* (2000), however, did not exclude the involvement of tryptic enzymes either. It is therefore

instructive to examine more closely the structural basis of enzyme specificity of members of the BBi family, with special regard to LBTI.

Structural factors in determining enzyme preference of BBis

It is common consensus that the enzyme preference of serine proteinase inhibitors is primarily determined by the amino acid residue occupying position P1; in other words, the geometry and electrostatic properties of the S1 pocket encode for the specificity of proteinases that are otherwise very similar in structure. The buried Asp^{189} of the enzyme carrying a negative charge is responsible for trypsin-like activity, i.e. for the preference of a basic residue at P1. In chymotrypsin-like proteinases the position 189 is occupied by an uncharged Ser that allows the binding and cleavage at aromatic residues, such as Phe, Tyr and Trp. As the example of LBTI clearly demonstrates, smaller hydrophobic residues like Leu, can also bind to the S1 pocket of CHM. Residues constituting the specificity pocket around the P1 side chain can be regarded as additional factors of specificity. In elastase, the bulkier Val^{16} and Thr^{226} restrict the size of the P1 residue to that of an Ala, while in thrombin Ala^{190} allows the encapsulation of the sterically more demanding Trp side chains.

However, replacement of the $S1-Asp^{189}$ of TRP with a Ser residue — that should have resulted in an active pseudo-chymotrypsin — yielded only a poor non-specific protease (Gráf *et al.*, 1988), although the binding pockets of these two enzymes are highly conserved and homologous and thus the mechanism of action would not change by a D189S mutation. Further studies demonstrated that changes in loop1 (189-195) and loop2 (214-228) and the Y172W mutation improves the catalytic efficiency towards CHM-like substrates. Specificity codeterminants occupying distant segments of the protein chain were thus proved to be involved in precise positioning of the scissile bond that can be attained with the TRP→CHM[S1+L1+L2+Y172W] hybrid enzyme.

Similarly, enzyme specificity of a serine proteinase inhibitor is mainly determined by the P1 residue. A single mutation at a P1 site, however, does not necessarily result in complete loss of inhibitory activity. As an example, the LP1R mutation in ovomucoid-type inhibitors decreases antichymotryptic activity only modestly (Lu *et al.*, 1997). This finding strongly hints that amino acids beyond the P1 sites may also be responsible for enzyme binding and act as codeterminants in enzyme preference.

As presented in *Table 2.8*, in spite of high structural similarity of the two LBTI domains, their binding to the corresponding proteinase enzymes displays striking differences. The number of interacting residues is of the same order of magnitude, but whereas contacts between TRP and the inhibitor are overwhelmingly hydrophilic, interactions with CHM are mostly hydrophobic. The pattern of hydrophobic and hydrophilic patches exhibited by LBTI binding domains can be regarded as perfectly complementary to the interacting surface of the enzymes. However, as discussed in *II.3.4*, binding to the enzyme residues at positions 214 and 216 is essentially identical in the case of both LBTI inhibitory loops. On the other hand, similarities in the primary region are restricted to P1 and P3 residues, as P4 and P2' side chains play different roles in complex formation within the distinct binding loops. The hydrophilic character of the TRP domain and the large hydrophobic surface at the CHM domain can be thus hypothesised to be the main specificity factor besides the P1 residue (*Figure 2.31*).

Table 2.8. Enzyme-inhibitor interactions.

TRP-domain					CHM-domain				
Residue	Position	region	interaction partners	mode of interaction	Residue	Position	region	interaction partner	mode of interaction
Cys^{22}			Ser^{217}	H-bonds					
Ser^{23}	P4		Gln^{175}	H-bonds	Ile^{50}	P4		hydrophobic cluster **1**	hydrophobic
Cys^{24}	P3		Gly^{216}	H-bonds	Cys^{51}	P3		Gly^{216}	H-bonds
Thr^{25}	P2	**P**	Leu^{99}	hydrophobic	Thr^{52}	P2	**P**	hydrophobic cluster **1**	hydrophobic
Lys^{26}	P1		Ser^{214}	H-bonds	Leu^{53}	P1		Ser^{214}	H-bonds
Ser^{27}	P1'				Ser^{54}	P1'			
Ile^{28}	P2'		Phe^{41}	H-bonds	Ile^{55}	P2'		hydrophobic cluster **2**	hydrophobic
Pro^{29}	P3'				Pro^{56}	P3'			
Pro^{30}	P4'				Ala^{57}	P4'		hydrophobic cluster **2**	hydrophobic
Gln^{31}			His^{57}, Ser^{96}, Tyr^{94}	H-bonds	Gln^{58}				
Cys^{32}					Cys^{59}				
Arg^{33}		**Q**	Ser^{96}, Asn^{97}	H-bonds	Val^{60}		**Q**	hydrophobic cluster **1**	hydrophobic
Cys^{34}					Cys^{61}				
Thr^{35}					Asp^{62}			Lys^{175}	salt bridge
Asp^{36}					Asx^{63}				
Ile^{64}		**R**			Leu^{37}		**R**	hydrophobic cluster **1**	hydrophobic

Canonical P1-S1 interactions are not listed. **P**: primary, **Q**: secondary, **R**: tertiary contact regions. The hydrophobic clusters in chymotrypsin comprise residues Leu^{97}, Ile^{99}, Trp^{172}, Trp^{215} and Lys^{175} for hydrophobic cluster **1**, and Phe^{39}, Phe^{41}, Leu^{143} and Met^{192} for hydrophobic cluster **2**.
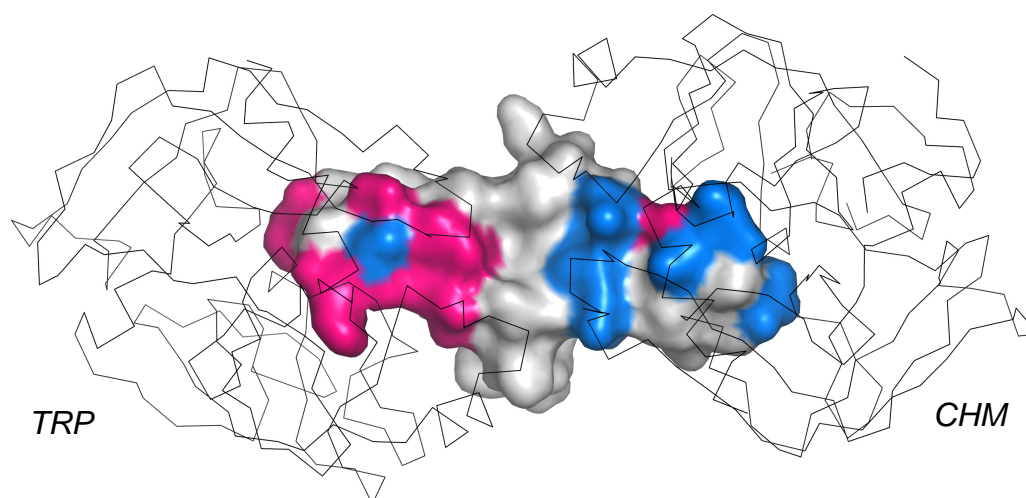
Figure 2.31. Binding surface of LBTI. Hydrophobic residues are coloured blue, hydrophilic residues are pink. TRP and CHM molecules are represented as C^{α}-traces.

Considering the differences between the binding mode and the distribution of interacting residues (*Table 2.8*), a sequence-based division of the binding loops to primary, secondary and tertiary binding regions suggested by Hirono *et al.* (1984) and later by Koepke *et al.* (2000) may not be the adequate description of the proteinase binding sites. Unfortunately, out of a scarce number of CHM-BBi complex structures, a more precise but still sufficiently general categorisation of interacting sites may also be problematic.

Rigidity vs. motility

As discussed in *II.1.3*, the respectable rigidity of the enzyme binding loops provides the inhibitor with resistance against enzymatic cleavage, as it prevents conformational changes that might be important for the catalytic processing. This is supported by the finding that the complex formation is accompanied by only slight rearrangements in side chain conformation, i.e. that the reactive loop inherently possesses a geometry that is perfectly complementary to the enzyme binding site. In addition, the canonical loop conformation is further stabilised by tight intermolecular interactions that also originate from the complementarity of the interacting surfaces. The enzyme-inhibitor complex can thus be regarded as a thermodynamic sink where the activation energy of conformational transitions that would lead to an enzymatic reaction is markedly increased. The barrier of the nitrogen-inversion step between the tetrahedral and acyl-enzyme intermediate is raised by the unusual rigidity that is further supported by interactions between the backside of the scissile bond and the enzyme.

Furthermore, the rigid β-turn of the (theoretical) leaving group can also disfavour the dissociation of the leaving group.

The comparison of the free and complexed inhibitor, however, revealed significant differences in the relative position of the two inhibitor domains. As expected from the evolution and disulfide bridge arrangement of BBis, these molecules exhibit a certain degree of flexibility at the interdomain interface that can be assumed to be necessary for a better enzyme-inhibitor fit, but the observed motility could also allow a precise adjustment of the proteinase molecules upon simultaneous binding and the formation of the ternary complex without steric clashes.

It is noteworthy that previous crystallographic studies on TRP-CHM-BBi ternary complex structures did not provide a satisfactory basis for the examination of such variations in the geometry of the inhibitor. In the case of both complex structures determined by Lin *et al.* (1993) and Koepke *et al.* (2000) only the proteinase molecules obeyed the crystallographic symmetry, the inhibitor molecules were lying on a twofold crystallographic axis that coincided with their inherent pseudo-twofold axis, therefore the inhibitor molecules were disordered. Although the slightly higher resolution of these structures would have allowed observations pertaining to such conformational changes, the uncertainties introduced by the inhibitor disorder precluded precise interpretation of the electron density map.

Dimerisation and inhibitory activity

Titration experiments on the soybean BBi demonstrated that the resulting TRP-BBi ratio depends on the BBi concentration. For the micromolar concentration range 1:1 ratio was borne out for both TRP and CHM (Odani & Ikenaka, 1978), while a 2:1 ratio was reported for TRP when applying nanomolar concentrations (Koepke *et al.*, 2000). This seems to imply that the CHM subdomain is shielded at higher concentrations, i.e. in dimeric state only the TRP loop is left free for interactions and so inhibitor variants with high CHM-inhibitory activity can be expected to associate less readily.

Precise comparison of residues taking part in both interactions reveals that Arg[33](LBTI) is the only amino acid residue on the TRP binding site of the molecule that is involved in dimerisation as well as binding the TRP molecule. On the contrary, three out of four residues interacting with the other monomer in the CHM-domain participate in enzyme-inhibitor interactions. The asymmetric distribution of involved side chains thus seems to support the aforementioned assumption.

The self-association properties, however, appear to be in correlation with the antichymotryptic activity of the inhibitor molecule, as discussed in *II.2.1*. This would contradict to the hypothesised decrease of inhibitory activity with increasing inclination for dimerisation, as variant IV was found to be the strongest chymotrypsin inhibitor as well as to form dimers or trimers unlike the other variants. This discrepancy may be resolved by taking the hydrophobic character of the CHM domain into account. As shown in *Figures 2.19* and *2.31*, similar forces play role in both the dimer formation and enzyme binding: the hydrophobic side chains concentrate on the surface of the CHM domain, while the TRP domain promotes the dimerisation with mostly hydrophilic residues. Since hydrophobic patches were found to be essential for both dimer and complex formation, dimerisation therefore does not preclude the possibility of a stronger binding to CHM, but can result in an apparent decrease of affinity when the inhibitor is present at high concentration.

III Viscotoxins from the European mistletoe (*Viscum album* L.)

III.1 Introduction

III.1.1 Taxonomy and nomenclature of mistletoe species

Taxonomically, the European mistletoe (*Viscum album* L.) belongs to the class of dicotyledons (*Magnoliopsida*), among them to the subclass of roses (*Rosidae*) and to the polyphyletic order of santals (*Santalales*). This order comprises the families of *Viscaceae*, *Loranthaceae*, *Santalaceae*, *Olacaceae* and *Eremolepidaceae*; its various members adopted the semiparasitic lifestyle independently (Schmeil & Fitschen, 1993). The *Viscaceae* family can be further divided into seven genera with altogether 400 species worldwide. The genus *Viscum* is represented by about hundred species distributed mostly over tropical areas, typically in South Africa, Madagascar and Asia. The European mistletoe grows natively only in Europe and Asia; it occurs, however, in California as a result of cultivation (Kirkup *et al.*, 2000a).

Based on their different host specificity and biochemistry, three subspecies can be separated: *V. album* L. ssp *album*, ssp *abietis* (Wiesb.) Abrom. and ssp *austriacum* (Wiesb.) Vollm. The subspecies *album* (called also *platyspermum* Kell. in certain sources) grows exclusively on deciduous trees and shrubs, such as poplar (*Populus* spp), apple (*Malus* spp), maple (*Acer* spp), birch (*Betula* spp), ash (*Fraxinus* spp) and linden (*Tilia* spp), while ssp *austriacum* (=*laxum* Fiek.) can only be found on conifera, most commonly on pine trees (*Pinus* spp), but sometimes on spruce (*Picea* spp) and larch (*Larix* spp) as well. The ssp *abietis* is the specific semiparasite of fir (*Abies* spp). On the contrary, mistletoe rarely attacks oak (*Quercus* spp) and elm (*Ulmus* spp) trees; moreover, beech (*Fagus* spp) and bux (*Buxus* spp) show high resistance against *Viscum* spp. Some tree species appear resistant to mistletoe, like eucalyptus (*Eucalyptus* spp), ginkgo (*Ginkgo biloba*) and cedar (*Cedrus* spp) (Kelly *et al.*, 1997). To date only one plant, the grey broom (*Genista cinerea*), has been described as a host of both ssp *album* and ssp *austriacum* (von Tubeuf, 1923; Grazi & Zemp, 1985).

The name "mistletoe" originates from "mistel-tan" in Anglo-Saxon that is a composite word of the German diminutive for dung (Mist) and the Anglo-Saxon twig (tan), probably

because it was observed early that a crucial role in distributing the seeds of mistletoe is played by birds (Durant, 1976). The name "*Viscum*" is the Latin equivalent of bird-lime, the sticky substance that was made of mistletoe berries and smeared on branches in order to catch smaller birds. "*Album*" means white and describes the usual colour of the plant's fruit (Jaeger, 1994).

III.1.2 Biology and economic impact

Mistletoe is an evergreen semiparasite of decidua and conifera, it takes up water and inorganic substances from the host, but it has the capability of producing its own carbohydrates by photosynthesis (Schulze & Ehleringer, 1984). The mistletoe plants are attached to the host through their haustoria (modified roots). They have dichasially branched stems with nodal constrictions and opposite, simple, entire leaves. Plants often develop a round form up to 50 cm in diameter. Mistletoes are dioecious, thus either female plants producing berries or male ones producing pollen only. Their flowers grow without petals and are inconspicuous, yellowish-green. The fruit is a white-yellow berry containing one or two seeds and a gluey fluid (*Figure 3.1*).

(a)   *(b)*

Figure 3.1. *(a)* Part of a mistletoe branch. Stems are dichasially branched carrying simple, entire leaves. Mistletoe berries are usually yellow-white. *(b)* Mistletoe flower, growing at the nodal constrictions without petals.

The mistletoe seeds are covered with viscin instead of a seed coat. The major means of dissemination is the deflection of the seeds by birds; although some birds, like the blackcap (*Sylvia atricapilla* L.), eat only the viscid pulp. The distributed seeds bind to the branches of trees through viscin, which is a hygroscopic mass of cellulotic strands in a pectic matrix. They germinate usually in the first couple of months of the year.

Two phases can be distinguished in the life cycle of mistletoes. The first, non-parasitic phase includes spreading and germination of seeds and the first period of fastening to the host. The subsequent parasitic phase begins with permanent adhesion and rooting into the xylem of the host tree. The first leaves grow only one year after the germination, while the blossoming and fruiting occurs only in the fourth-to-fifth year (von Tubeuf, 1923).

Mistletoes can cause damage in forests, orchards, plantations and ornamentals, since there are no ideal methods of controlling or preventing mistletoe infections. The use of plant growth regulators, such as ethephon, provides only a temporary control. Severely infested branches need to be cut entirely in order to remove embedded and hidden haustoria completely and prevent further seed distribution, as the infection rapidity of newly planted trees is correlated with the proximity and severity of already established infestations. Mistletoes were found to be responsible for the reduction of growth, wood quality and fruiting as well as for predisposing trees to attacks by fungi and insects (Kelly *et al.*, 1997; Hawksworth, 1980).

III.1.3 Folklore and medicine

From the earliest of times mistletoe has been considered to be sacred, magical and mysterious, a bestowal of life and fertility. It was commonly used in medicine as well, it appeared in the herbals of whole Europe, China and the Middle East. Several authors, e.g. Discoroides and Pliny, mentioned mistletoe and its extracts as a cure of numerous diseases, such as infections, epilepsy and sterility. According to the Christian legend, mistletoe used to be a forest tree that shrunk to its current growth form out of shame after it was used to make the Cross. Mistletoe is therefore still called "Herbe de la Croix" in parts of Brittany. In the Middle Ages and later until the 18th century mistletoe was used as a panacea, a remedy for a broad variety of illnesses (Kirkup *et al.*, 2000b). *Table 3.1* shows the history of mistletoe as a medicinal plant. In today's medicine, especially in Germany, several conditions, including hypertension and cancer are treated with mistletoe preparations. The use of mistletoe extracts

in anticancer therapy dates back to 1920 when the anthroposoph Rudolf Steiner suggested the parentheral injection of aqueous mistletoe extracts, based on the observed similarities in tumour and mistletoe growth (Steiner, 1961). Several standardised mistletoe extracts are offered in the market (European Iscador, Helixor, Eurixor, Iscucin, Isorel/Vysorel, Lektinol, Plenosol, Abnobaviscum, T4GEN etc.) and used as part of complementary cancer therapy, i.e. as a companion to radiation- and chemotherapy. These usually contain mistletoe lectins and VTs in a specific ratio and concentration. To prevent the biologically active constituents from digestion, mistletoe extracts are administered mostly subcutaneously (Kienle & Kiene, 2003).

Table 3.1. History of mistletoe as a medicinal plant.

period	the use of mistletoe
5th century BC	against kidney diseases
1st century AD	used as cure-all and antidote; treatment of infertility
12th century AD	against epilepsy, kidney, spleen diseases and infertility
16th century AD	against epilepsy, liver and spleen diseases
18th century AD	against epilepsy, weak heart oedema, lung and liver diseases
19th century AD	used as cure-all
20th century AD	against high blood pressure, cancer and arthrosis

III.1.4 Viscotoxins

The structure of viscotoxins

The scientific search for the biologically active components of the mistletoe dates back to 1881, when Park described a digitalis-like effect on a mistletoe extract (Park, 1881). Until the turn of the century mainly aqueous extracts were studied, the first isolation of a mistletoe alkaloid was only presented in 1907 (Gaultier & Chevalier, 1907). The first publication reporting the isolation of viscotoxin is linked with the name of Winterfeld (1942); he coined the new term "viscotoxin" in 1948 (Winterfeld & Bjil, 1948). In these early studies VT was described as a toxic material that consists of amino acids. Based on sequence homology, S-S bridge arrangement (Samuelsson *et al.*, 1968; Samuelsson & Pettersson, 1971a; Schrader & Apel, 1991) and its biological role (Florack & Stiekema, 1994) VT was classified in later studies into the α- and β-thionin family of basic, cystein rich and toxic microproteins, which are abundant throughout the plant kingdom. The α- and β-thionins usually consist of 46 amino acid residues and have an approximate molecular weight of 5 kDa (Florack & Stiekema, 1994). A list of the currently known α- and β-thionins is given in *Table 3.2.*

Table 3.2. Summary of the most important α- and β-thionins.

name	source	common name	references
viscotoxin A1	*Viscum album* ssp *album*	European mistletoe	Schaller *et al.*, 1996
viscotoxin A2	*Viscum album* ssp *album*	European mistletoe	Olson & Samuelsson, 1971
viscotoxin A3	*Viscum album* ssp *album*	European mistletoe	Samuelsson *et al.*, 1968
viscotoxin B	*Viscum album* ssp *album*	European mistletoe	Samuelsson & Pettersson, 1971
viscotoxin C	*Viscum album* ssp *coloratum ohwi*	Asiatic mistletoe	Romagnoli *et al.*, 2003
viscotoxin 1-Ps	*Viscum album* ssp *austriacum*	European mistletoe	Samuelsson & Jayawardene, 1974
phoratoxin A	*Phoradendron tomentosum*	Californian mistletoe	Mellstrand & Samuelsson, 1974
phoratoxin B	*Phoradendron tomentosum*	Californian mistletoe	Thunberg, 1983
denclatoxin	*Dendrophtora clavata*	Columbian mistletoe	Samuelsson & Pettersson, 1977
ligatoxin A	*Phoradendron liga*	Argentinian mistletoe	Thunberg & Samuelsson, 1982
ligatoxin B	*Phoradendron liga*	Argentinian mistletoe	Thunberg & Samuelsson, 1982
α1-purothionin	*Triticum aestivum*	wheat	Ohtani *et al.*, 1977
α2-purothionin	*Triticum aestivum*	wheat	Ohtani *et al.*, 1977
β-purothionin	*Triticum aestivum*	wheat	Mak & Jones, 1976
α-hordothionin	*Hordeum vulgare*	barley	Hernández-Lucas *et al.*, 1986
β-hordothionin	*Hordeum vulgare*	barley	Hernández-Lucas *et al.*, 1986
Pyrularia thionin	*Pyrularia pubera*	rabbitwood/buffalonut	Vernon *et al.*, 1985
leaf thionin BTH6	*Hordeum vulgare*	barley	Bohlmann & Apel, 1987
leaf thionin DB4	*Hordeum vulgare*	barley	Bohlmann & Apel, 1987
leaf thionin DG3	*Hordeum vulgare*	barley	Bohlmann & Apel, 1987
α-avenothionin	*Avena sativa*	oat	Békés & Lásztity, 1981
β-avenothionin	*Avena sativa*	oat	Békés & Lásztity, 1981
secalethionin	*Secale cereale*	rye	Békés *et al.*, 1982
hellethionin	*Helleborus purpurascens*	lenten rose	Milbradt *et al.*, 2003
crambin A	*Crambe abyssinica*	Abessinian cabbage	Teeter *et al.*, 1981
crambin B	*Crambe abyssinica*	Abessinian cabbage	Teeter *et al.*, 1981

To date at least six different isoforms of VT from ssp *album* have been described, namely A1 (Schaller *et al.*, 1996; Orrú *et al.*, 1997) , A2 (Olson & Samuelsson, 1972), A3 (Samuelsson *et al.*, 1968), B (Samuelsson & Pettersson, 1971b), 1-PS (Samuelsson & Jayawardene, 1974) and U-PS (Ribéreau-Gayon *et al*, 1986). Except for U-PS, the sequences of all isoforms are available. As shown in *Fig. 3.2*, the primary structure of VTs is highly conserved, three-quarters of the sequence are identical amongst the different isoforms, but it also shows a high degree of sequence homology when compared to other members of the α- and β-thionin family, such as α- and β-purothionin (Ohtani *et al.*, 1977; van Campenhout *et al.*, 1998), hordothionin (Ozaki *et al.*, 1980), crambin (Teeter *et al.*, 1981), phoratoxin (Mellstrand & Samuelsson, 1974) and hellethionin (Milbradt *et al.*, 2003). All of these microproteins are stabilised by three disulfide bridges between Cys^3/Cys^{40}, Cys^4/Cys^{32} and Cys^{16}/Cys^{26}, defining a structural motif denoted as "concentric motif" (Orrú *et al.*, 1997; *Figure 3.3*).

```
                                  10          - 20          30          40
                                   |           - |           |           |
viscotoxin a1            KSCCPSTTGR  NIYNTCR-LTG  SSRETCAKLS  GCKIISASTC  PSNYPK
viscotoxin a2            KSCCPNTTGR  NIYNTCR-FGG  GSRQVCASLS  GCKIISASTC  PSDYPK
viscotoxin a3            KSCCPNTTGR  NIYNACR-LTG  APRPTCAKLS  GCKIISGSTC  PSDYPK
viscotoxin b             KSCCPNTTGR  NIYNTCR-LGG  GSRERCASLS  GCKIISASTC  PSDYPK
viscotoxin 1ps           KSCCPNTTGR  NIYNTCR-FGG  GSREVCARIS  GCKIISASTC  PSDYPK

viscotoxin consensus     KSCCPnTTGR  NIYNtCR-:gG  gsRe:CA:lS  GCKIISaSTC  PSdYPK


phoratoxin A             KSCCPTTTAR  NIYNTCR-FGG  GSRPVCAKLS  GCKIISGTKC  DSGWNH
phoratoxin B             KSCCPTTTAR  NIYNTCR-FGG  GSRPICAKLS  GCKIISGTKC  DSGWNH
denclatoxin              KSCCPTTAAR  NQYNICR-LPG  TPRPVCAALS  GCKIISGTGC  PPGYRH
ligatoxin A              KSCCPSTTAR  NIYNTCR-LTG  TSRPTCASLS  GCKIISGSTC  XSGWXH
ligatoxin B              KSCCPSTTAR  NIYNTCR-LTG  ASRSVCASLS  GCKIISGSTC  DSGWNH


alpha1 purothionin       KSCCRSTLGR  NCYNLCR--AR  GAQKLCAGVC  RCKISSGLSC  PKGFPK
alpha2 purothionin       KSCCRTTLGR  NCYNLCR--SR  GAQKLCSTVC  RCKLTSGLSC  PKGFPK
beta   purothionin       KSCCKSTLGR  NCYNLCR--AR  GAQKLCANVC  RCKLTSGLSC  PKDFPK
alpha  hordothionin      KSCCRSTLGR  NCYNLCR--VR  GAQKLCAGVC  RCKLTSSGKC  PTGFPK
beta   hordothionin      KSCCRSTLGR  NCYNLCR--VR  GAQKLCANAC  RCKLTSGLKC  PSSFPK
alpha avenotoxin         KSCCRDTLGR  DCYDLCR--SR  GAPKLCATLC  RCKISSGLSC  PKDFPK
beta  avenotoxin         KSCCRDTLGR  DCYDLCR--AR  GAPKLCSTLC  RCKITSGLSC  PKDFPK
secaletoxin              KSCCKSTLGR  DCYDLCR--GR  GAEKLCAELC  RCKITSGLSC  PKDFPK
leaf thionin BTH6        KSCCKDTLAR  NCYNTCRF-AG  GSRPVCAGAC  RCKIISGPKC  PSDYPK
leaf thionin DB4         KSCCKDTLAR  NCYNTCHF-AG  GSRPVCAGAC  RCKIISGPKC  PSDYPK
leaf thionin DG3         KSCCKNTTGR  NCYNACRF-AG  GSRPVCATAC  GCKIISGPTC  PRDYPK
pyrularia thionin        KSCCRNTWAR  NCYNVCRLPGT  ISREICAKKC  DCKIISGTTC  PSDYPK
hellethionin             KSCCRNTLAR  NCYNACRF-TG  GSQPTCGILC  DCIHVTTTTC  PSSHPS

toxic thionin consensus  KSCCk:TlgR  nCY:lCr  ::  g::klC:::C  :Cki:::::C  P:dfPK


crambin A                TTCCPSIVAR  SNFNVCR-LPG  TPEALCATYT  GCIIIPGATC  PGDYAN
crambin B                TTCCPSIVAR  SNFNVCR-LPG  TSEAICATYT  GCIIIPGATC  PGDYAN

crambin consensus        TTCCPSIVAR  SNFNVCR-LPG  T:EA:CATYT  CGIIIPGATC  PDGYAN
```

Figure 3.2. Sequence alignment and consensus sequences in the α- and β-thionin family. The conserved residues, corresponding to the viscotoxin consensus sequence are marked with red.

```
                          -20         -10          0
                           |           |           |
signal sequence          MEVVRG  SSLVLLVLLL  GALLVSQVES


                          10          20          30          40
                           |           |           |           |
Matured viscotoxin a3    KSCCPNTTGR  NIYNACRLTG  APRPTCAKLS  GCKIISGSTC  PSD


                          50          60          70          80
                           |           |           |           |
acidic polypeptide       FYCT  LGCESSQCAT  NSNGDAEAVR  CKTACSDLCQ  DVDDA
```
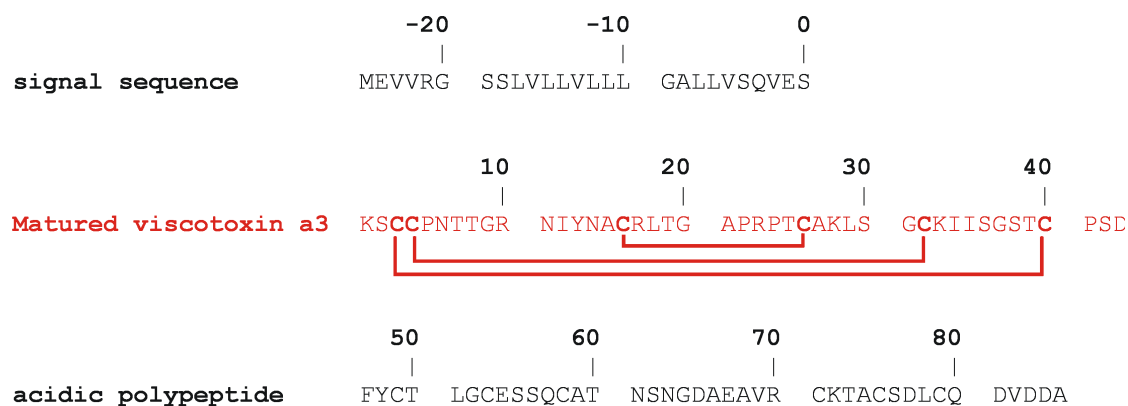
Figure 3.3. The primary structure of the VT A3 prepropeptide. The arrangement of the disulfide bridges corresponds to the "concentric motif". The N-terminal signal sequence and the C-terminal acidic polypeptide are cleaved during the maturation of the protein.

The three-dimensional structure of α- and β-thionins resembles a capital letter L (or the Greek capital Γ, as mentioned in some sources, e.g. by Hendrickson & Teeter, 1981). The longer arm of the letter L is constituted by a pair of amphipathic α-helices running in opposite directions, while the shorter arm is defined by two short antiparallel β-sheets and the C-terminal coiled part of the protein chain (*Figure 3.4*). The biologically active members of this family display a high positive electrostatic potential, especially around the pocket between the two arms of the letter L (Romagnoli *et al.* 2000). Publications concerning structural studies on α- and β-thionins are listed in *Table 3.3*.
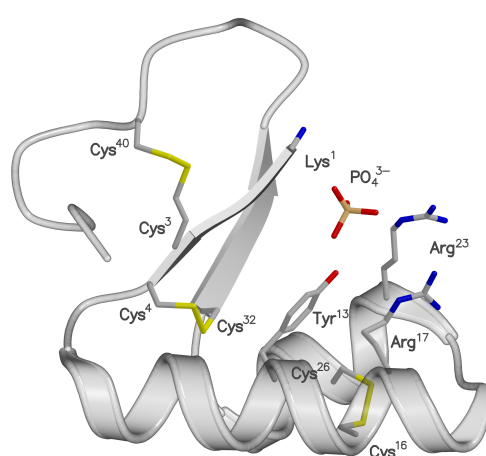

Figure 3.4. Three-dimensional structure of VT A3 that resembles a capital letter L.

Table 3.3. Structural studies on α- and β-thionins.

name	method	PDB deposition code	reference
hellethionin	NMR	1NBL	Milbradt *et al.*, 2003
ligatoxin B	molecular modelling	not deposited	Li *et al.*, 2002
viscotoxin A2	NMR	1JMN	Coulon *et al.*, 2003
viscotoxin A3	NMR	1ED0	Romagnoli *et al.*, 2000
viscotoxin B	NMR	1JMP	Coulon *et al.*, 2003
viscotoxin C	NMR	1ORL	Romagnoli *et al.*, 2003
crambin*	X-ray	not deposited	Hendrickson & Teeter, 1981
crambin*	X-ray	1CNR	Yamano & Teeter, 1994
crambin*	NMR	not deposited	Xu *et al*, 1999
α-hordothionin	NMR	not deposited	Han *et al.*, 1996
α_1-purothionin	NMR	not deposited	Clore *et al.*, 1986, 1987a
α_1-purothionin	X-ray	2PLH	Rao *et al.*, 1995
β-purothionin	X-ray	1BHP	Stec *et al.*, 1995
phoratoxin B	NMR	not deposited	Kleywegt *et al.*, 1991; Clore *et al.*, 1987b

(*Only some selected publications are listed for crambin.)

Biosynthesis and occurrence of viscotoxins

The viscotoxin molecule is synthesised in form of a precursor, similarly to other thionins, like hordothionin or leaf thionins (Schrader *et al.*, 1991). The prepropeptide with a molecular weight of 15 kDa contains an N-terminal signal leader sequence of 26 residues and an acidic polypeptide chain of 39 amino acids at the C-terminus besides the 46 residue long central thionin sequence (*Figure 3.3*). Both the signal and the acidic part are cleaved off during the processing of VT. As reported by Schrader & Apel (1991), the acidic part of the prepropeptide might neutralise the toxic effects of VT within the plant cells.

VTs occur mostly in the peripheral organs of the mistletoe, like in its leaves, stems and berries. The mistletoe plants are especially rich in VTs during summer, while in winter the VT-content strongly decreases. *Viscum album* contains 0.1-1.0% VT on average, depending on the host tree it is growing on: e.g. poplar mistletoes produce nearly twice as much VTs as apple mistletoes and about four times as much as pine mistletoes (ssp *austriacum*). The VT-content of fir mistletoes (ssp *abietis*) corresponds to that of decidua mistletoes (ssp *album*). A host-dependent variation in the isoform pattern was also described: mistletoes growing on decidua exhibit a higher ratio of VT A2 and VT A3 besides smaller amounts of VT A1 and VT B, but contain no VT 1-PS and VT U-PS. On the contrary, mistletoes from pine trees produce predominantly VT 1-PS and VT U-PS. In fir mistletoes all VTs are present, these contain mostly VT A3, though (Schaller *et al.*, 2000).

III.2 Crystal structure of viscotoxin A3

III.2.1 Purification, crystallisation and data collection

Isolation and purification

The damp climatic conditions of Göttingen – where the European mistletoe is thriving and gives a nearly infinite natural source for isolation even in large quantities – are especially suitable for research on VTs. For their relatively high VT-content, as mentioned in *III.1.4*, poplar trees were chosen as the mistletoe source. Mistletoe bushes were collected in April 2001 in Weende (Göttingen), and immediately dried at 60˚C in order to avoid the enzymatic degradation of its microproteins.

The dried plant material was ground, the powder was then extracted with 2% acetic acid and the resulting extract was concentrated and passed through a cation exchange column (SP Sephadex C-25, elution with 1 M ammonium acetate). The raw product was further purified using MPLC (LiChroprep RP-8, elution with 40% acetonitril, in the presence of 0.1% TFA). Viscotoxin isoforms were separated by HPLC (Nucleosil 100 C8 and Phenomenex C18 5 μ, elution with a water-acetonitril gradient in the presence of 0.1% TFA). It should be noted that no phosphate containing chemicals were used during isolation and purification. (For more details see Girmann, 2003.) 100 g mistletoe powder yielded 5 mg VT A1, 8 mg VT A2, 15 mg VT A3, 14 mg VT B, 7 mg VT 1-PS and 0.2 mg of a new VT isoform: VT B2. Viscotoxin samples were stored at 4°C as lyophilised powder.

Crystallisation

For crystallisation, VT A3 was dissolved in double distilled water to a concentration of 15 mg/ml and centrifuged at 8000 rpm for 5 minutes to remove solid particles. Crystallisation was carried out applying the hanging drop technique (Hampel *et al.*, 1968) at room temperature; equal volumes of the protein and precipitant solution were mixed on siliconised cover slides and inverted over the reservoirs of 24-well tissue culture plates. Crystallisation conditions were screened using the Hampton Crystal Screen Kits (Jancarik & Kim, 1991; Cudney *et al.*, 1994). Bunches of thin needle crystals with two crystal habits could be obtained in conditions containing high molecular weight polyethylene-glycols (PEGs) and ammonium-sulfate at high concentrations. Conditions yielding microcrystals were optimised by varying the pH and the concentration of precipitants systematically in order to achieve better crystal quality. Thin and particularly fragile, but measurement quality crystals of approximate dimensions 0.1x0.1x1.0 mm grew from 2+2 μl droplets over 0.5 ml of the reservoir solution containing 0.15 M $(NH_4)_2SO_4$, 0.05 M cacodylate buffer pH 6.5 and 30% PEG8000 in 1-2 weeks. In order to introduce anomalous scatterers and to improve the quality of these needle crystals, 15 mM $HgCl_2$ was used as additive (*Figure 3.5*).

Data collection

Measurements were performed under cryoconditions, therefore cryoprotection of the crystals was indispensable. In spite of the relatively high concentration of PEG8000, the presence of $(NH_4)_2SO_4$ in the mother liquor required the addition of 25% glycerol as cryoprotectant. Crystals were transferred into a droplet of the cryo buffer, soaked for
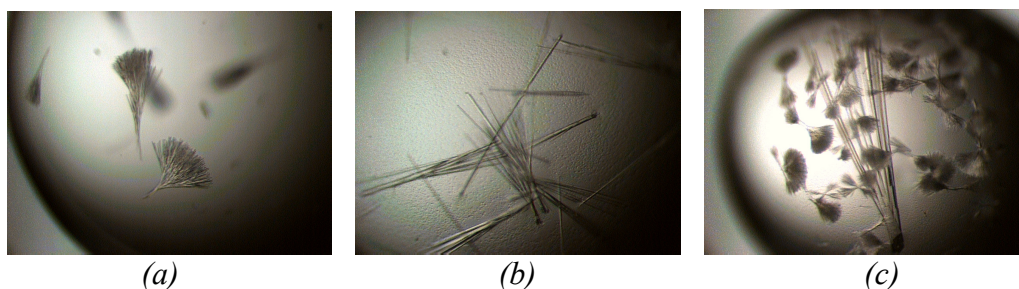
(a) *(b)* *(c)*

Figure 3.5. (a-b) Two crystal habits of of VT A3 crystals (zoom: ~40x). **(b)** Optimised VT A3 crystals (zoom: ~20x).

2-5 minutes and then mounted in a loop onto the goniometer head in an arbitrary orientation and flash-cooled in a nitrogen stream to 100 K.

Since VT A3 was isolated from a natural source, contains no methionine and all my attempts at molecular replacement (using the deposited thionin structures listed in *Table 3.3*) and classical heavy atom derivatisation failed, I decided to carry out a sulfur-SAD (single-wavelength anomalous diffraction) experiment to take advantage of the weak anomalous scattering of the sulfur atoms in the three disulfide units per molecule. To ensure high data redundancy and precision, four 360º φ-scans and nine 180º ω-scans were collected with thin slicing (0.2º, over 15000 frames!) on a Bruker rotating anode generator equipped with Osmic focusing mirrors, a three-circle goniometer and a Bruker Smart6000 4K CCD detector. Scans were arranged in such a way that data resulting from φ- and ω-scans were of approximately equal redundancy, resolution and completeness. Data were processed using *PROTEUM*, *SAINT* and *SADABS*, and analysed with *XPREP* (Bruker Nonius, 2002).

For phase extension and refinement purposes, a 1.75 Å synchrotron dataset was also measured at the X31 beamline at EMBL/DESY. In this case 280 images were collected in 0.5° steps on a MAR345 image plate detector and processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Data collection statistics are shown in *Table 3.4*.

Table 3.4. Data collection statistics.

	In-house	Synchrotron
Space group	P2$_1$2$_1$2	P2$_1$2$_1$2
Unit cell parameters (Å)	a=47.697	a=47.995
	b=68.773	b=68.593
	c=25.346	c=25.264
Data collected	333468	50787
Data unique	4490	8938
Resolution (Å)	2.20 (2.30-2.20)	1.75 (1.75-1.85)
Completeness (%)	96.8 (75.3)	100.0 (99.7)
Redundancy	71.87 (4.86)	5.25 (4.42)
R$_{int}$ (%)	4.85 (10.85)	8.92 (38.66)
Average I/σ(I)	93.85 (12.61)	14.31 (3.89)

(Values in parentheses refer to the last resolution shell)

III.2.2"Heavy atom" substructure solution and phasing

After measurement and integration of the in-house dataset, scaling was performed separately for the two different types of scans in order to preserve their independence. The correlation coefficient between the signed anomalous differences derived separately from the two half datasets were then calculated with *XPREP* in order to confirm the presence of significant anomalous signal. It is generally recommended that for substructure solution the data should be truncated at the resolution where this correlation coefficient falls below 30% (Schneider & Sheldrick, 2002); the signal in subsequent shells is not reliable and inclusion of these would only increase the noise and hinder the heavy atom search. In this case the correlation coefficient test indicated the presence of significant anomalous signal up to 2.4 Å:

```
Inf - 8.0 - 6.0 - 5.0 - 4.0 – 3.6 – 3.4 – 3.2 – 3.0 – 2.8 – 2.6 – 2.4 – 2.2 Å
    96.2  96.7  91.5  92.1  78.8  73.1  64.8  67.1  63.9  37.5  30.0    2.5
```

In the next step, all scans were scaled together in a single *SADABS* run to improve anomalous signal quality further by taking advantage of all available redundancy; statistical independence of the datasets was necessary for an unbiased data assessment, but this had already been performed and there was no need to keep the two types of scans separate any further. Anomalous differences used in subsequent substructure solution and phasing steps were calculated from this merged dataset and confirmed the resolution limit established with the correlation coefficient, since the anomalous $\Delta F/\sigma(\Delta F)$ dropped below 1.3 (Schneider & Sheldrick, 2002) at the same resolution:

```
Inf - 8.0 - 6.0 - 5.0 - 4.0 – 3.6 – 3.4 – 3.2 – 3.0 – 2.8 – 2.6 – 2.4 – 2.2 Å
    5.35  6.22  4.68  4.11  3.05  2.64  2.55  2.48  2.36  1.84  1.44  1.18
    5.82  6.50  4.73  4.27  3.15  2.79  2.67  2.47  2.36  1.91  1.51  1.30
```

Although the anomalous signal was significant to 2.4 Å resolution, the data were truncated at 3.0 Å in order to promote the coalescence of disulfide bridges into single super-sulfur atoms. As this resulted in a great reduction of reflections used in substructure solution, it seemed necessary to weaken the amplitude criterion in the selection of reflections, and thus the following settings were employed in *SHELXD* (Sheldrick *et al.*, 2001):

```
FIND 6              (number of sites to look for)
MIND –3.5           (minimum distance between sites)
SHEL 100 3.0        (resolution shell)
ESEL 1.4            (selection of smaller E-values)
```

Six super-sulfurs with good figures of merit and a sharp cutoff in site occupancies were found by *SHELXD* corresponding to six disulfide bridges and indicating the presence of two molecules in the asymmetric unit with an estimated solvent content of 32%:

```
CC All/Weak 30.59 / 10.48
```

```
PATFOM    22.95

        x          y          z       height
   0.422720   0.757589   0.266453   1.0000
   0.643856   0.619290   0.457746   0.8691
   0.716744   0.574620   0.773041   0.7998
   0.725876   0.440945   0.824074   0.7383
   0.857356   0.398658   0.962270   0.7249
   0.458417   0.658519   0.400175   0.5563
   0.619024   0.590883   0.576724   0.2814
   0.355361   0.746407   0.350031   0.2552
   0.793291   0.672993   0.267973   0.2535
```

The heavy atom sites could be divided into two groups obeying twofold non-crystallographic symmetry, and distances between peaks were in agreement with those of disulfide bridges in known thionin structures, and also allowed the derivation of the NCS operator. However, the map after *SHELXE* (Sheldrick, 2002) phasing and density modification proved not to be interpretable, and NCS averaging with the NCS operator and an NCS mask generated from the heavy atoms did not improve it at all (*Figure 3.7.a*). After successful structure solution, a *post-mortem* analysis showed that there was very little initial phase information, apparently not enough to define the molecular envelope adequately.

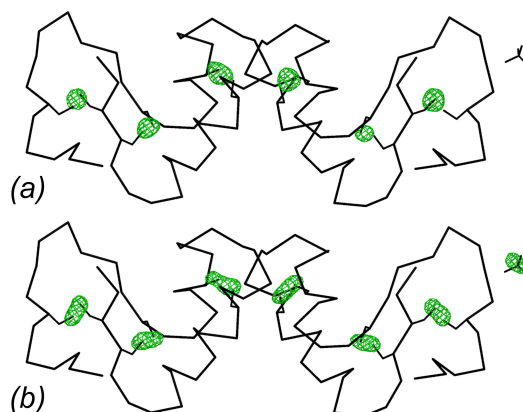Previous experience gained from measurements on test crystals suggested that the

Figure 3.6. Anomalous Fourier maps (phases from *SHELXE* after density modification) showing *(a)* six super-sulfur atoms around the disulfide bridges and *(b)* resolved sulfur atoms and a sulfate anion.

precision of the heavy atom model was insufficient to obtain acceptable phases and the low solvent content hindered the phase improvement stage. To improve the heavy atom model, *SHELXD* was modified by its author to introduce a little chemical information, namely that the super-sulfur atoms are actually disulfides. The new algorithm enables the direct location and resolution of disulfide units even when applying it for medium resolution anomalous data. For disulfide location, data were thus truncated at somewhat higher resolution than for the

super-sulfur search, i.e. at 2.5 Å, as indicated by the statistics in *XPREP*. The modified *SHELXD* applying the command

```
        DSUL 6
```

found all of the six disulfide bridges and one putative sulfate anion with appreciably higher correlation coefficients between E_o and E_c than for the super-sulfurs alone, indicating substantial improvement of the heavy atom substructure (*Figure 3.6.b*):

```
CC All/Weak 37.85 / 16.00
PATFOM   7.72

     x        y        z      height
0.44790 0.33872 0.12316  99.90
0.70462 0.41914-0.25153  98.93
0.71950 0.56653-0.35220  95.60
0.90390 0.38982-0.38536  93.98  ← SO₄²⁻ ion
0.86529 0.59903-0.43330  92.81
0.72043 0.55250-0.28041  90.90
0.84032 0.61037-0.49205  90.62
0.41659 0.24246 0.28427  90.28
0.64555 0.37179 0.07058  83.11
0.46808 0.35016 0.05836  73.67
0.72723 0.42584-0.31836  69.63
0.43642 0.23838 0.21292  67.33
0.64408 0.39273 0.01255  66.52
0.76481 0.70607-0.33368  47.43
```
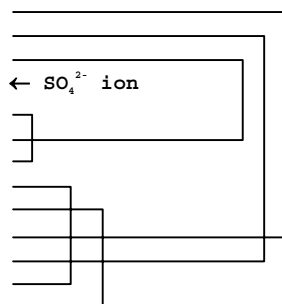
With these atoms as input, without further refinement of their coordinates, *SHELXE* gave a good discrimination between the two heavy atom enantiomorphs and a map with a map correlation coefficient of 0.71 relative to the final refined structure that was almost certainly traceable by hand (but not by *ARP/wARP* using default settings). After NCS averaging with *DM* (Cowtan, 1994) using an NCS mask generated from the sulfur positions with *MAPMASK* (CCP4, 1994) the map correlation coefficient rose to 0.84 and the map allowed autobuilding with *ARP/wARP* (Perrakis *et al.*, 1999). *Figures 3.7* and *3.8* show the progress of phase improvement.

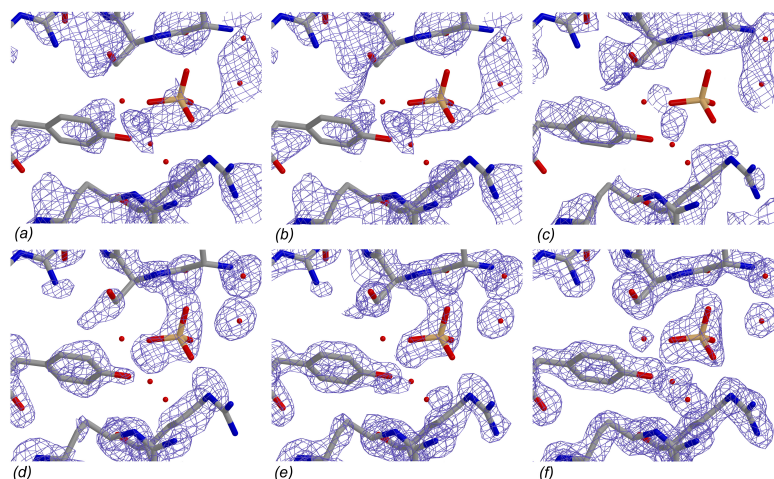*Figure 3.7.* Electron density maps around the tetrahedral anion found in the positively charged binding pocket, calculated at different stages of phase improvement: *(a)* initial phases and *(b)* after density modification in *SHELXE*, disulfides not resolved; *(c)* initial phases and *(d)* after density modification in *SHELXE*, disulfides resolved; *(e)* after NCS averaging in *DM*; *(f)* final refinement.
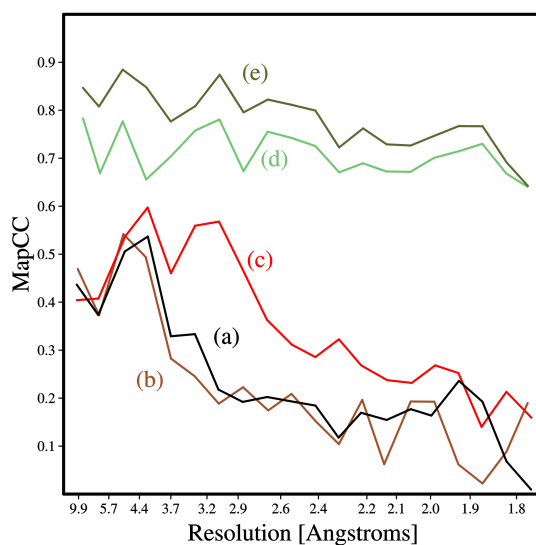
Figure 3.8. The map correlation coefficient as function of resolution in different stages of phase improvement. The labels correspond to those of *Figure 3.7*.

III.2.3 Model building, refinement and validation

An initial main chain trace was created automatically by *ARP/wARP*; it built all residues except for the four termini in a standard run of 100 building cycles. After manual addition of the terminal amino acids, side chains were docked using the known sequence of VT A3. Solvent molecules were added with *SHELXWAT* (Sheldrick & Schneider, 1997) and manually. The refinement of this model was carried out with *SHELXL* (Sheldrick & Schneider, 1997) against all F^2 data between 40 and 1.75 Å resolution using bond length, bond angle, chiral volume and planarity restraints to a final R-factor of 16.74%. *XPREP* was used to mark 5% of the reflections with R_{free}-flag in thin resolution shells. Throughout the refinement, *2mF_o-DF_c* and *F_o-F_c* type maps were displayed for manual editing of the model with *XtalView* (McRee, 1999) at 1σ and ±3σ levels, respectively.

The final model contains two independent peptide chains with altogether 92 residues, 94 water molecules and 3 tetrahedral solvent particles refined as phosphate and sulfate ions.

The stereochemistry of the final model was assessed with *PROCHECK* (Laskowski *et al.*, 1993); all amino acid residues were found to be in the allowed regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968). A summary of the refinement statistics is given in *Table 3.5*.

Table 3.5. Model refinement statistics.

Resolution range (Å)	40.00-1.75
Number of protein residues	92
Number of non-H atoms	780
Number of water molecules	94
R-factor/R_{free} (%)	16.74/23.91
rmsd from ideal geometry	
Bond length (Å)	0.006
Angle distances (Å)	0.021
Ramachandran plot (%)	
Allowed region	100
Generously allowed region	0
Disallowed region	0

III.2.4 Structure description

Overall fold

The overall fold of VT A3 resembles the canonical architecture of α- and β-thionins, a capital letter L (Hendrickson & Teeter, 1981). The long arm of this letter is an antiparallel pair of helices, while the short arm is defined by two antiparallel β-strands (*Figure 3.4*). Residues 6-18 and 22-30 are involved in the formation of the helices that are stabilised by a number of interhelical H-bonds and a disulfide bridge between the residues 16 and 26. The β-strands comprise the residues 1-5 and 32-37, linked by four H-bonds and a disulfide bridge between residues 4 and 32. The C-terminal part of the protein chain forms a coil fixed by mainly hydrophobic interactions and a disulfide bridge between residues 3 and 40.

The two arms of the letter L are connected by a H-bond and salt-bridge network involving the highly conserved residues 2, 4, 10 and 46: the C-terminal carboxyl group binds the guanidinium group of Arg^{10} with a salt bridge and the N-atom of Cys^4 through a H-bond, while the hydroxyl and carbonyl O-atoms of Ser^2 form H-bonds to the guanidinium group of Arg^{10}.

This pattern of stabilising forces is rather conserved among the α- and β-thionins and provides the molecule with a high degree of rigidity. However, flexibility of the molecule can be expected at the hinge region in spite of the stabilising quartet of residues 2, 4, 10 and 46, since no covalent bonds fix the relative position of the two arms in the letter L.

Two nearly equivalent molecules were found in the asymmetric unit, connected by a non-crystallographic twofold axis forming a hydrophobic type dimer (referred to as the first and second molecule). Least squares superposition of the molecules gave an rmsd of 0.41 Å for all main chain atoms proving the high conformational similarity.

Comparison with the NMR models

The crystal structure of the VT A3 shows an rmsd of 1.17-1.42 Å when superimposed on the NMR models (Romagnoli *et al.*, 2000; PDB ID: 1ED0). As illustrated in *Figure 3.9*, models superimposed between the residues 8-28 indicate clear differences in the relative positions of the two major domains. The short β-sheets are displaced towards the helices in the crystal structure, therefore the entire molecule adopts a closed conformation. The C-terminal end and the first few residues connected to it are less precisely determined in the NMR structures, but it is worthy of note that the conformational variance amongst the ten deposited NMR structures is smaller than the structural difference between the X-ray and NMR models. Similar results were obtained when VT A3 was compared to the recent hellethionin-D structure (Milbradt et al., 2003; PDB ID: 1NBL).
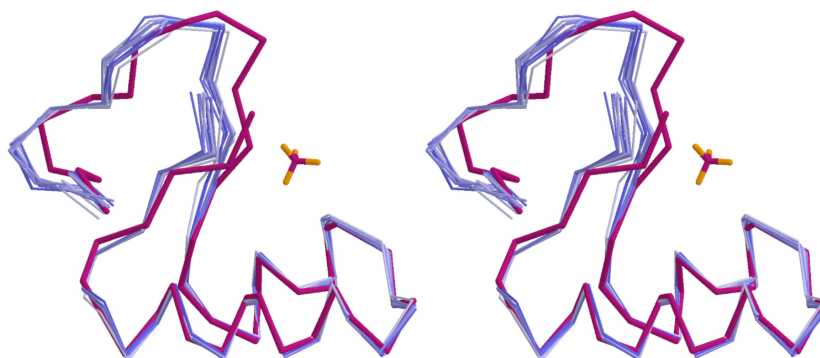

Figure 3.9. Stereo view of the backbone representation of VT A3: the X-ray structure (purple) superimposed on the NMR models (blue).

Phosphate binding sites

A tetrahedral anion bound to the positively charged groove between the two domains of the L-structure can account for the closed conformational state of both molecules in the asymmetric unit. These anions were refined as phosphates based on the following evidences (Girmann, 2003). Mass spectra indicated an additional VT A3 peak with 98 Da mass difference corresponding to a phosphate adduct. ^{31}P-NMR spectra with a single peak at 2.03 ppm relative to phosphoric acid also showed the presence of non-covalently bound phosphate in the purified sample, although no phosphate containing chemicals were used during isolation and purification. Photometric phosphate-determination indicated phosphate in the sample after hydrolysing the protein. The fact that the sulfate anion found additionally to the twelve sulfur sites from anomalous differences does not coincide with the position of

either of these anions is also in agreement with the chemical data. Disorder or extensive thermal motion could be responsible for the absence of the anomalous density around the phosphate ion in the second molecule. On the other hand, the B-factor of the anion bound to the first molecule refines to a value similar to that of the sulfate anion involved in lattice formation, and should therefore be visible in the anomalous map if the assumption of its being a sulfate were correct (*Figure 3.6.b*).

The coordination of the phosphate ion can be summarised as follows. One of the oxygen atoms of the anion is hydrogen-bonded directly to the N^ε atom of Arg^{23} while the interaction with the N^ε atom of Arg^{17} is mediated by a water molecule. Another phosphate oxygen forms H-bonds with the OH groups of Tyr^{13} and of Ser^2. The backbone N atoms of Lys^1 and Ser^2 are also involved in the anion binding (*Figure 3.10*).
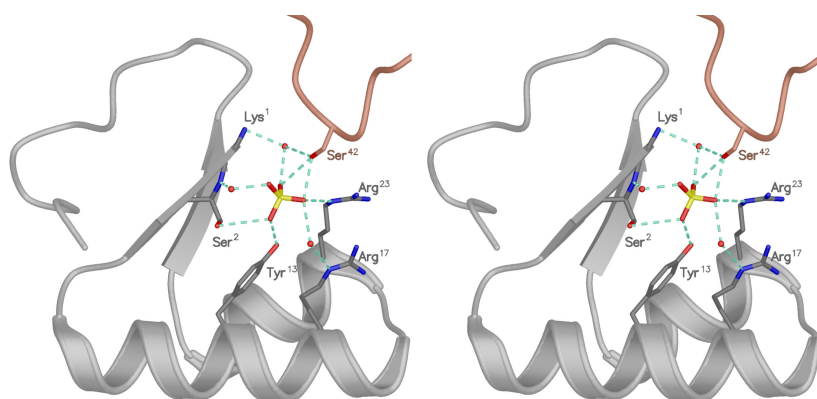

Figure 3.10. Stereo view of residues involved in phosphate binding in VT A3. (The colouring of molecules corresponds to that in *Figure 3.13.*)

The arrangement of the coordinating side chains is essentially identical between the two monomers; the binding pockets of the monomers differ in two features, though. $Ser^{42}(2)$ of a symmetry equivalent molecule forms connections with the oxygen atoms of the phosphate in the first molecule, while this interaction is missing in the second case. As a result, the phosphate anion of the second molecule is disordered, and is partially displaced towards the solvent region.

Dimerisation and crystal packing

As a consequence of the extremely low solvent content, there are several interactions between symmetry equivalent molecules in the crystal structure. These mostly involve H-bonds mediated by a number of ordered structural water molecules. There are, however, two intermolecular interactions that can be considered to play the major role in the crystal

lattice formation. Firstly, as mentioned above, the two molecules in the asymmetric unit build a hydrophobic type dimer. The apolar residues Ala^{15}, Cys^{16}, Ala^{21} and Leu^{29} together with the hydrophobic groups of residues Thr^{19}, Thr^{25} and Lys^{28} form the dimer interface between the two monomers (*Figure 3.11*). The hydrophobic surfaces of the amphipathic helices are therefore protected from direct contact with the solvent.
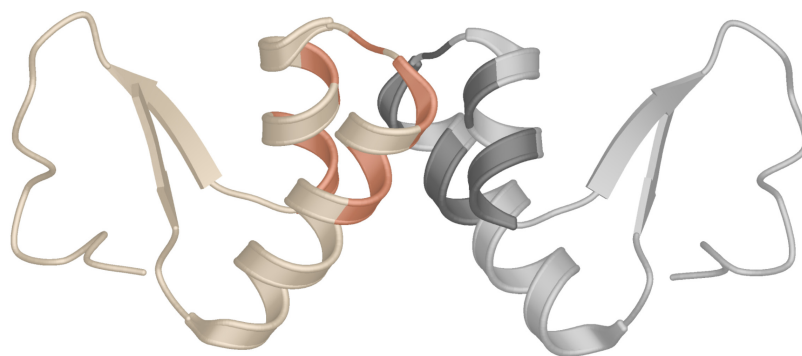

Figure 3.11. Ribbon diagram of the hydrophobic VT A3 dimer. Residues constituting the dimer interface are represented in darker colours.

In addition to the hydrophobic type of molecular association, strong hydrophilic contacts are also present. The strongest one of these involves the binding of a sulfate anion and leads to a tight packing of four molecules around the crystallographic twofold axis (*Figure 3.12-13*). As shown in *Figure 3.12*, the sulfate ion is coordinated by three molecules: by the first molecule of the asymmetric unit and by two symmetry equivalents of the second one. The backbone N atoms of $Thr^{39}(1)$ and $Ser^{38}(1)$ of the first molecule and the guanidinium groups of $Arg^{17}(2)$ and $Arg^{23}(2)$ in one of the symmetry mates of the second one form a pocket with relatively high positive electrostatic potential where the sulfate ion is bound. This pseudo-tetrameric association of the molecules is strengthened by the contact between the sulfate ion and the side chain oxygens of $Thr^{39}(1)$ and $Ser^{36}(2)$.
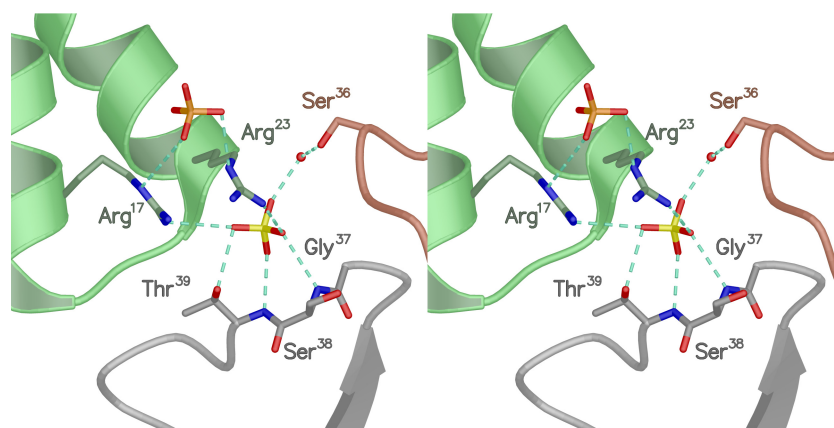

Figure 3.12. Stereo view of the sulfate binding residues. (The colouring of molecules corresponds to that in *Fig. 3.13*.)
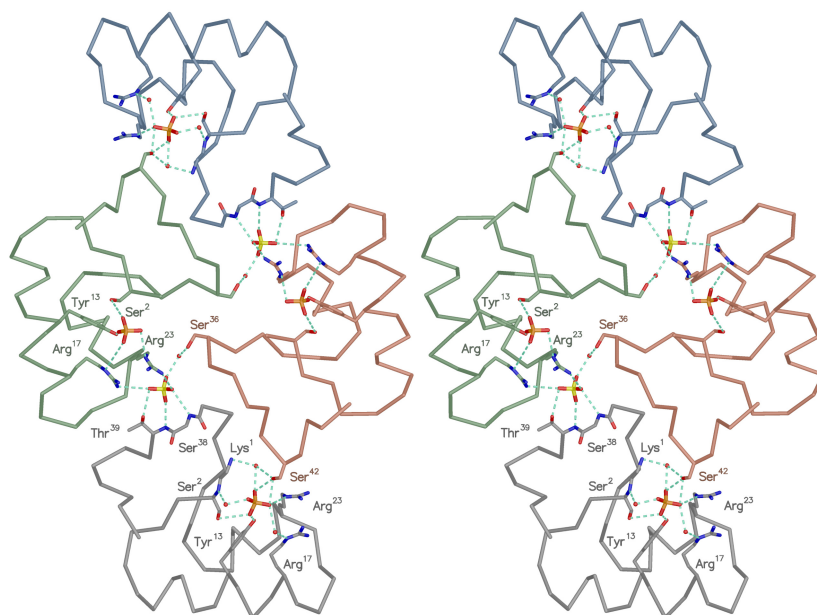
Figure 3.13. Stereo view of the tetrameric association of four VT A3 molecules around the twofold axis. The first molecule of the asymmetric unit and its symmetry equivalent are drawn in grey and blue, while the second one and its symmetry mate are coloured brown and green. The sulfate ion is denoted by yellow, the phosphate by orange.

The crystal packing of VT A3 resembles that of α- and β-purothionin (Rao *et al.*, 1995; Stec *et al.*, 1995). All of these crystals contain tetrameric associations of molecules that constitute layers (*Figure 3.14*). Unlike purothionins, further association of VT A3 tetramers to zigzag-like layers is attained by the hydrophobic dimerisation of the VT A3 molecules, the contact between the tetramers in the plane of the layers is thus hydrophobic. In contrast, the interactions between the layers can be characterised as rather weak, as they only involve weak H-bonds mediated by solvent molecules.

Figure 3.14. Crystal packing of the VT A3 molecules. The molecules represented in yellow correspond to the asymmetric unit, the yellow box shows the position of a tetramer.

III.3 Crystal structure of viscotoxin B2

III.3.1 Purification, crystallisation and data collection

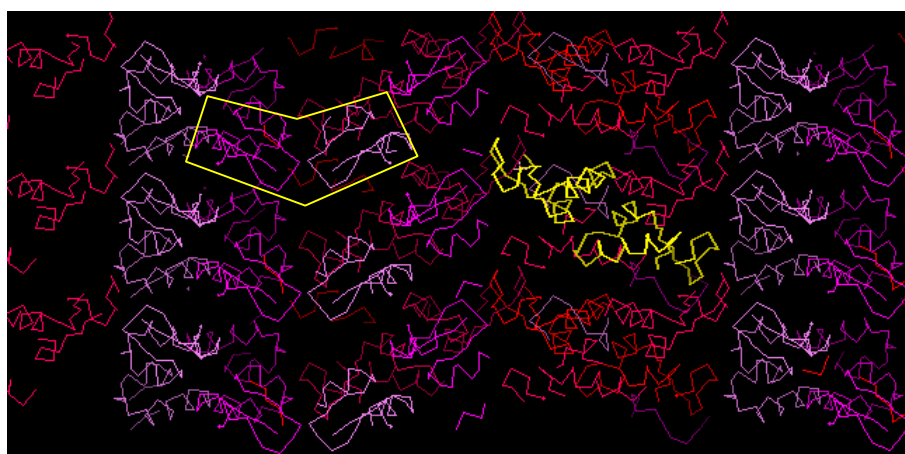### Isolation and purification

HPLC analysis of the viscotoxin raw product containing all the six previously known isoforms showed the presence of an unknown microprotein (*Figure 3.15*). This component was classified to the viscotoxins for its similar behaviour in ESI-MS experiments (formation of multiply charged ions, molecular weight and its presence as a phosphate adduct). Since its retention time in HPLC purification steps appeared to be similar to that of VT B, it was later termed as VT B2. The similarity of their amino acid sequence was proved in MS-MS experiments. (Nevertheless, as some uncertainties in the sequence concerning the Asp/Asn and Glu/Gln residues existed at the time of model building and refinement, these were tried to be clarified based on the molecular environment of these residues in the crystal structure.) The protocol established for the other VT isoforms was used for isolation and purification of this new compound.
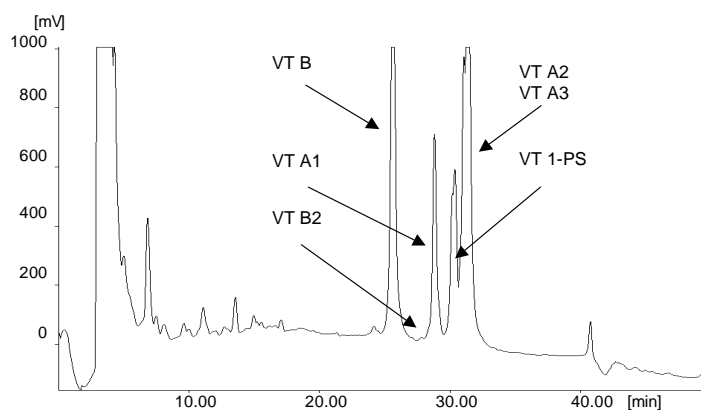
Figure 3.15. HPLC chromatogram of the VT row product showing the presence of a previously unknown isoform (Girmann, 2003).

Crystallisation

The poor yield of VT B2 (0.2 mg from 100 g dried mistletoe powder) demanded the application of submicroliter volumes of the protein solution during the crystallisation screens, and first crystallisation trials were therefore performed with the microbatch method. Droplets were set up by mixing 0.3-0.4 µl of both the protein and the precipitant solution in 72 well

plates under a 1:1 mixture of silicon and paraffin oil that allows the gradual evaporation of water.

The suitable protein and precipitant concentrations were thus established in mirobatch experiments and later adjusted to the hanging drop technique. In the final crystallisation experiments 1+1 µl droplets were used. Diffraction quality crystals (*Figure 3.16.*) with approximate dimensions of 0.2x0.2x0.5 mm^3 grew in 2-3 months (!) from droplets containing equal volumes of protein solution (30 mg/ml VT B2 in double distilled water) and precipitant (0.20 M $(NH_4)_2SO_4$, 0.08 M cacodylate buffer pH 6.5 and 35% PEG 8000 or 25% PEG20000). These crystals, unlike those of VT A3, were unsensitive to mechanical impacts and behaved rather like salt crystals.
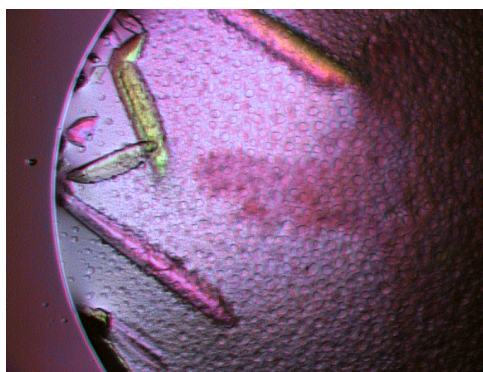
Figure 3.16. VT B2 crystals grown in 2-3 months.

Data collection

VT B2 crystals were mounted in loops and flash cooled in a nitrogen stream to 100 K using 25% glycerol as cryoprotectant, like in the case of VT A3. Two datasets were then measured at EMBL/DESY: one for refinement purposes and another one for S-SAD phasing.

The excellent diffraction properties of these crystals allowed data collection up to 1.05 Å at the X13 beamline using 0.98 Å wavelength; high intensity reflections could be observed even at the edge of the detector. The resolution of the dataset was virtually determined by the small diameter of the MarCCD camera used in this experiment and the sample-detector distance minimum. As usual in the case of high resolution synchrotron data collections, the dataset was measured in two consecutive passes at two different sample-detector distances in order to avoid overloads in the inner resolution shells. 1080 high resolution images were taken with 0.5° steps at 75 mm distance, while the 180° low resolution pass was collected with 1.0° oscillations at 210 mm.

For phasing purposes a longer wavelength was chosen to achieve a stronger anomalous signal from sulfur atoms. However, at 1.7 Å wavelength the increased diffraction angles limited the resolution to 1.80 Å, although a MAR345 detector was used in the largest readout mode and was positioned to the minimum distance. The weak beam intensity of the beamline X31 and the unfortunately limited measurement time allowed the collection of only 398 frames with 1° oscillations.

Both the high resolution and the anomalous datasets were indexed and integrated with *DENZO* and scaled using *SADABS*. The correct space group, $P2_12_12_1$, was determined with *XPREP*. Data collection statistics for both datasets are listed in *Table 3.6*.

Table 3.6. Data collection statistics.

	λ = 0.98 Å (X13)	λ = 1.70 Å (X31)
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a=39.819	a=40.186
	b=40.392	b=40.986
	c=44.694	c=45.198
Data collected	732 985	105 782
Data unique	34 317	7 331
Resolution (Å)	1.05 (1.15-1.05)	1.80 (1.90-1.80)
Completeness (%)	99.8 (99.6)	99.8 (99.0)
Redundancy	21.32 (21.32)	14.40 (13.97)
R_{int} (%)	6.61 (22.13)	7.14 (38.76)
Average I/σ(I)	25.84 (12.64)	24.10 (6.97)

(Values in parentheses refer to the last resolution shell)

III.3.2 Structure solution

Similarly to VT A3, the phase problem could not be solved with molecular replacement, even when using the crystal structure of VT A3 or its fragments as search models probably due to the extremely low solvent content. After successful application of the S-SAD on VT A3, this method was expected to be feasible in the case of this isoform as well.

In spite of the larger f'' value of sulfur at 1.70 Å wavelength, the anomalous signal of the low resolution dataset did not prove sufficient for heavy atom location, as the anomalous $\Delta F/\sigma(F)$ statistics output by *XPREP* showed the signal to be nearly random:

```
Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 - 1.8 A
    1.42  1.45  1.35  1.48  1.95  1.67  1.79  1.79  1.83  1.47  1.10  0.80
    1.81  1.88  1.80  1.45  1.44  1.23  1.34  1.31  1.32  1.22  1.11  1.03
```

This can probably be accounted for by the fairly low data redundancy and the infamous unreliability of the beamline X31 that had been observed during previous measurements as well. As a result, sulfur atoms could not be located based on these data.

The failure of S-SAD directed my attention to the atomic resolution dataset. The unit cell parameters suggested that only two molecules, i.e. about 800 atoms were present in the asymmetric unit offering a challenging target for *ab initio* direct methods. Two different structure solution strategies were tried: one *SHELXD* job was started to search for all atoms directly, another one was setup to locate first the sulfur atoms only and then to expand the structure with peak list optimisation performed in several steps. The first *SHELXD* run did not give any solution in a reasonably short time (one week). In the second one a lower success rate was expected, therefore only those solutions were fed into the peak list optimisation cycle that had a CC/CCweak above a pre-defined limit, in order to speed up the process. This lead to a clear solution (with a final CC of 78.28) in 2.5 days on a 2GHz Xeon machine. As a test, *SHELXE* was applied starting from the sulfur atoms located by *SHELXD* ─ practically used in "*ACORN* mode" ─ for phase determination and expansion from the sulfur positions. This also resulted in a very clear, high quality map (*Figure 3.17*).
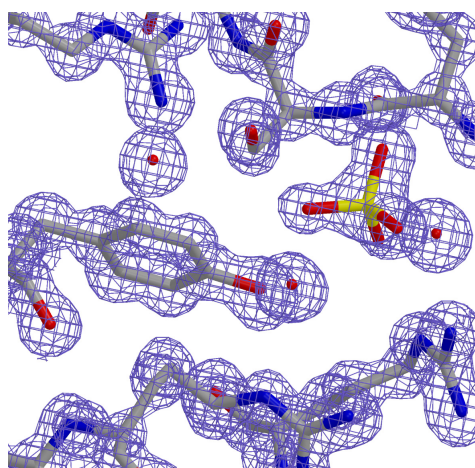
Figure 3.17. Electron density map around the phosphate ion in the VT B2 structure. Phases were calculated with *SHELXE* from the sulfur positions using only the atomic resolution dataset.

III.3.3 Model building, refinement and validation

Solely out of practical reasons, i.e. to avoid manual model building, the map output by *SHELXE* was traced by *ARP/wARP* in 100 building cycles. The initial model contained all residues except the four termini that could easily be built manually. Side chains were also docked using *ARP/wARP*. Water molecules were located in the density with the program *SHELXWAT* and checked manually. This structure was refined against all F^2 data with no resolution cutoff using the program *SHELXL* to an R-factor of 12.66%. As R_{free} reflections,

5% of the data were excluded from the refinement. Throughout the refinement angle distance, bond distance, chiral volume and planarity restraints were applied. All non-hydrogen atoms were refined anisotropically with suitable similarity and rigid bond restraints.

The model was visualised and manually adjusted to the electron density with *XtalView* using $2mF_o$-DF_c and F_o-F_c type maps contoured at 1σ and $\pm 3\sigma$ level, respectively. 13 disordered residues were identified and refined in multiple discrete conformations. The final model contains two independent peptide chains with altogether 92 amino acid residues, 96 water molecules and 5 tetrahedral solvent particles refined as phosphate and sulfate ions. The stereochemistry of the final model was assessed with *PROCHECK*. Similarly to the final model of VT A3, all residues showed φ and ψ angles typical for the allowed region of the Ramachandran plot. Refinement statistics are listed in *Table 3.7*.

Table 3.7. Model refinement statistics.

Resolution range (Å)	40.00-1.05
Number of protein residues	92
Number of non-H atoms	818
Number of water molecules	96
R-factor/R_{free} (%)	12.66/17.00
rmsd from ideal geometry	
Bond length (Å)	0.020
Angle distances (Å)	0.034
Ramachandran plot (%)	
Allowed region	100
Generously allowed region	0
Disallowed region	0

III.3.4 Structure description

Overall fold

The asymmetric unit of the crystal strucutre of VT B2 consists of two nearly identical molecules related by non-crystallographic symmerty and five tetrahedral anions. Two of these are bound to the positively charged groove of the VT molecules and were identified as phosphate ions on the analogy of VT A3, while three of them were found to take part in other intermolecular interactions.

The three-dimensional structures of VT B2 and A3 exhibit the same fold, even the angle between the two branches of the letter L is identical in the two molecules; least squares superposition of the models gave an rmsd of 0.07 Å for all main chain atoms. However, conformational differences can be observed between the residues Cys[40] and Lys[45], i.e. at the

C-terminal loop due to the different intermolecular interactions located in this region (*Figure 3.19.a*). Similarly to VT A3, superposition of the X-ray and NMR models (PDB ID: 1JMP, Coulon *et al.*, 2003) showed that the molecule adopts a more closed conformation in the crystal structure than in solution, as discussed in *III.2.4*.

Phosphate binding sites

A phosphate ion, like in the case of the VT A3 structure, could be located between the two arms of the letter L in both of the VT molecules of the asymmetric unit. As shown in *Figure 3.18*, the highly conserved residues Ser[2], Tyr[13], Arg[17] and Arg[23] take part in the phosphate binding. The orientation of their side chains is practically identical between the two molecules of the asymmetric unit. Moreover, only marginal differences could be detected between the isoforms A3 and B2. The N atoms of the positively charged guanidinium groups of Arg[17] and Arg[23] form three H-bonds with the O atoms of the phosphate ion, the side chain oxygens of Ser[2] and Tyr[13] and the N atom of Ser[2] are also involved in phosphate binding.
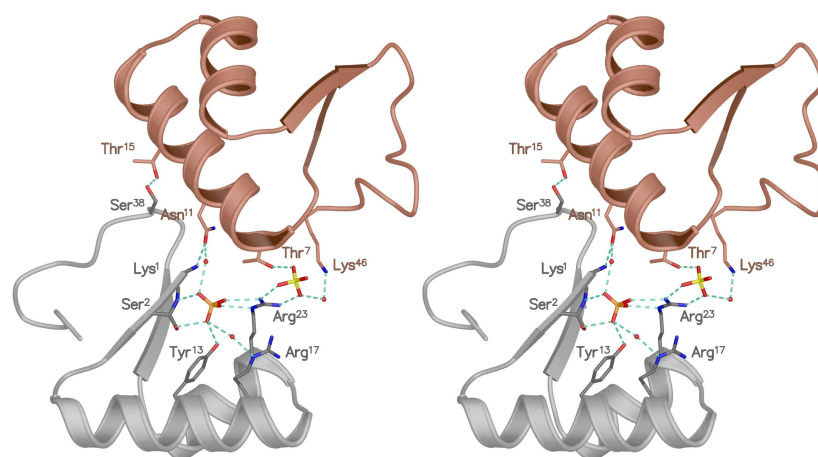
Figure 3.18. Stereo view of the phosphate binding residues in VT B2. (The first molecule of the asymmetric unit is coloured grey, while a symmetry equivalent of the second one is shown in brown.)

Dimerisation and crystal packing

Due to the very tight crystal packing a high number of intermolecular interactions can be detected in the crystal structure of VT B2. An intricate H-bonding network stabilises the molecular lattice involving a relatively large amount of well ordered solvent molecules. Two intermolecular interactions are particularly worthy of highlighting: the formation of a hydrophobic type dimer and the binding of three sulfate ions at the dimer interface and near

the binding pocket of the first molecule of the asymmetric unit. Like in the case of VT A3, the two molecules in the asymmetric unit associate to a dimer, as shown in *Figure 3.19.a*. The binding interface is made up by the hydrophobic groups of the residues Ile[12], Cys[16] and Lys[29] of both molecules.

Two tetrahedral anions, most probably sulfate ions originating from the crystallisation buffer, could be located at the turns of the helix-turn-helix-(HTH)-pair (*Figure 3.19.b*) where the residue Arg[25] of both molecules provides a positively charged environment for these ions. The monomers are crosslinked by the side chains of both Arg[25] residues, since they point towards the other molecule and make H-bonds with the sulfate ions. The Arg[25](1) of the first molecule takes part in H-bonds with the sulfate ion and the O atom of Gly[19](2) from the second molecule, while the Arg[25](2) of the second one forms H-bridges with the sulfate near the first one. Gly[20](1) and Ser[22](1) of the first molecule and Gly[19](2) and Ser[22](2) of the second one complete the coordination of the sulfates and further strengthen the dimerisation with seven H-bonds.
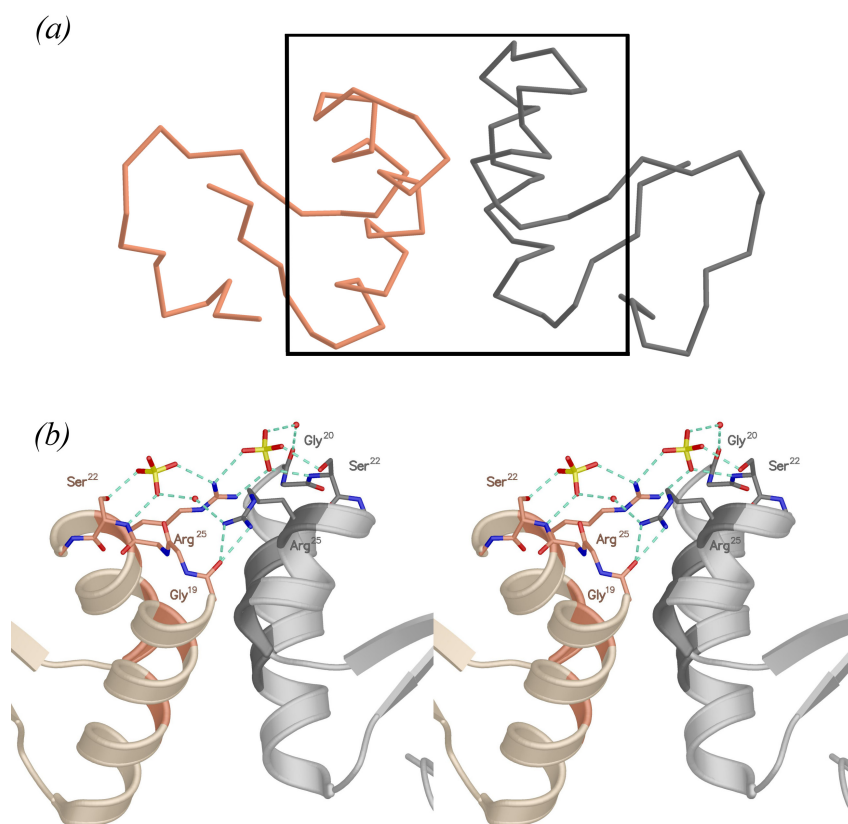

(a)

(b)

Figure 3.19. *(a)* Backbone representation of the VT B2 dimer. The marked area is shown enlarged in *(b)*. *(b)* Stereo view of the residues taking part in the dimer formation. Residues constituting the hydrophobic interface are shown in darker colours.

A third sulfate anion is bound near the phosphate ion of the first VT molecule. The residue Arg[23](1) that coordinates the phosphate ion contributes with two H-bonds to the coordination of the sulfate. A symmetry equivalent of the second molecule is also involved in the sulfate binding with the residues Thr[7](2) and Lys[46](2). As depicted in *Figure 3.18*, this interaction between the two molecules is made stronger by two H-bridges: one of them can be found between the side chain oxygens of Ser[38](1) and Thr[15](2), while the other one between the main chain N atom of Lys[1](1) and the side chain O of Asn[11](2).

Superposition of the dimers of VT A3 and B2 reveals striking differences in the relative position of the monomers, as illustrated in *Figure 3.20*. In the case of isoform A3, a nearly perfect closed NCS connects the molecules in the asymmetric unit. A 178° rotation brings the first molecule into an overlapping position with the second one and the angle between the rotation axis and the translation vector was found to be 88° (i.e. near 90°) that suggests a pure rotation. On the contrary, the monomers of VT B2 represent an improper NCS with a rotation angle of 168° and an angle of 103° to the translation vector. Variations of the primary structure in the region responsible for dimerisation offer a plausible explanation for the marked difference observed in the mode of dimerisation. The sterically demanding Arg[25] in B2 prevents the approach of Thr[19], which would result in a dimer interface corresponding to that of VT A3 and imposes a distorted dimer geometry.

Figure 3.20. Backbone representation of the VT dimers. The first monomer of VT A3 (green) has been superimposed on the first monomer of VT B2 (grey).

Both VT isoforms crystallise in orthorhombic space groups, VT A3 in P2₁2₁2, while B2 in P2₁2₁2₁. The VT B2 crystals therefore lack the twofold axis, which was demonstrated to be important for the tetramer formation (see *III.2.4*). This fact is interrelated with the finding that

the VT B2 structure is devoid of tetramers and as a consequence, stacking layers of tetramers as well. The VT B2 molecules connected by 2_1 axes form a rather isotropic lattice with about the same number of intermolecular interactions in all directions (*Figure 3.21*) what explains the more favourable physical and diffraction properties of the VT B2 crystals. Unlike these, VT A3 crystals were less resistant to mechanical impacts, had the tendency to break during soaking and mounting, which could be attributed to the weak intermolecular forces in direction *c*.
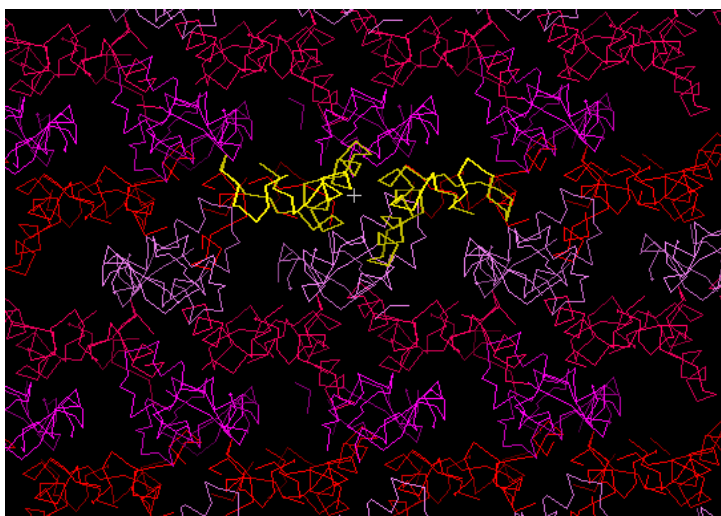
Figure 3.21. Crystal packing of the VT B2 molecules. The molecules drawn in yellow correspond to the asymmetric unit (and the hydrophobic type dimer).

III.4 Biological implications

The biological role and activity of viscotoxins

The *in vivo* function of VTs and other thionins is still unclear; it is, however, assumed that they play a role in the defence mechanism of plants and thus protect plant tissues, especially the endosperm, against microbial infections (Florack & Stiekema, 1994).

Hordothionin, the antifungal thionin from barley (*Hordeum vulgare*) seeds, was reported to cause permeabilisation of fungal hyphae to the non-metabolite α-aminoisobutyric acid and to alter the electrical properties of artificial lipid bilayers (Thevissen *et al*, 1996). Transgenic lines of *Arabidopsis thaliana* L. overexpressing their own endogenous thionin exhibited enhanced resistance against *Fusarium oxosporum* (Epple *et al.*, 1997). Similarly, *Arabidopsis*

lines with high VT A3 levels were more resistant to infection by *Plasmodiophora brassicae* (Holtorf *et al.*, 1998).

Thionins were shown to have a broad spectrum of biological activity, as they are toxic to Gram positive or Gram negative bacteria, fungi, yeast, and various mammalian cell types (Caleya *et al.*, 1972; Ewans *et al*, 1989). The cytotoxicity of VTs and mistletoe extracts was investigated in detail on different tumour cell lines, such as human myeloid cell line K562, human T-cell leukaemia cell line Molt4, Yoshida sarcoma cells and human granulocytes and lymphocytes (Jung *et al.*, 1990; Urech *et al.*, 1995). The Yoshida cell assay appeared to be the best indicator for cytotoxicity originating from VTs. Comprehensive toxicity studies showed considerable differences among the biological activity of VT isoforms (*Table 3.8*): VT A3 was found to be the most cytotoxic, whereas VT B the least potent (Schaller *et al.*, 1996).

In addition to their toxicity, VTs exhibit immunomodulating effects in non-toxic concentrations. They stimulate the phagocytosis of *E. coli* and CD8+ cells by human granulocytes (Stein *et al.*, 1999a, 1999b), and enhance the natural killer cell mediated and antibody redirected lysis of tumour cells (Tabiasco *et al.*, 2002).

Table 3.8. Comparison of the cytotoxic activity of the different VT isoforms on Yoshida sarcoma cells.

VT isoform	ED_{50}* (µg/ml)
VT A1	0.87
VT A2	1.08
VT A3	0.31
VT B	4.58
VT 1-PS	0.44
VT U-PS	4.04

(*ED_{50}: concentration of the substance inhibiting 50% of the incorporation of ^{3}H-thymidin.)

In spite of extensive studies on the biological activity of these microproteins, their mechanism of action has not been clarified definitively. In the early 80s VTs were proved to protect DNA against thermal denaturation through ionic binding (Woynarowski & Konopa, 1980). Although thionins indeed possess a helix-turn-helix (HTH) like structural motif occurring in the structure of many DNA-binding proteins, superposition of HTH-DNA complex structures with the NMR-models of VT A3 clearly demonstrated severe sterical hindrance between viscotoxin and DNA (Romagnoli *et al.*, 2000). Subsequent experiments

showed that thionins are able to permeabilise the cell membrane of various cell types (Carrasco *et al.*, 1981). Two different hypotheses accounting for the lytic effects of thionins: ion channel formation (Hughes *et al.*, 2000) and membrane disruption through electrostatic forces (Coulon *et al.*, 2002) were proposed as possible explanations. Structural studies together with electrostatic calculations also suggested interactions with negatively charged molecular targets driven by electrostatic forces (Romagnoli *et al.*, 2000). Activity measurements on chemically modified thionins indicated that the proper folding of the molecule as well as the residue occupying the position 13 are essential for the toxic activity (Evans *et al.*, 1989; Wada *et al.,* 1982).

At the cellular level, VTs act as initiators of necrotic cell death, and were reported to induce membrane permeabilisation, swelling of mitochondria with loss of their christae and degradation of cytoplasm and chromatin (Büssing *et al.*, 1999). Less severe membrane affections lead to apoptotic cell death as a response to distinct death signals. According to Coulon *et al.* (2002), phosphatidylserine (PS) molecules are required in the outer leaflet of model membrane vesicles for successful VT A3-membrane interactions. This finding is further strengthened by the fact that tumourigenic cells exhibit an excess of PS in the outer leaflet compared to non-tumourigenic cells (Utsugi *et al.*, 1991; Connor *et al.*, 1989). This may also be related to the lower hemolytic activity of VTs (Lankisch & Vogt, 1971).

Structure vs. function

Previous studies on the cytotoxic effect of thionins and the related viscotoxins proved that these microproteins are able to disrupt the membrane integrity of many types of cells and revealed a strong correlation between the cytotoxicity of the different toxins and their capability for membrane permeabilisation (Thevissen *et al.*, 1996). Since the α- and β-thionins and viscotoxins possess very similar three-dimensional structures with little conformational variance, the sometimes striking differences in their biological activity should therefore be accounted for by specific positions in their amino acid sequence.

The primary structure of viscotoxins is highly conserved, three-quarters of the sequence are identical amongst the different isoforms. Most of the variable amino acids can be found in the central region of the protein chain, between positions 12 and 30. Interestingly, all conserved residues in this region either have structure stabilising functions or are involved in the phosphate binding.

The residue Tyr[13] was previously identified as a key residue for the toxicity. The diiodination or nitration of Tyr[13] in purothionins led to the complete loss of their activity (Wada *et al.*, 1982; Evans *et al.*, 1989). This finding suggested that this sole Tyr residue is essential for membrane binding, which was later supported by X-ray crystallographic studies: in the molecular structure of these toxins a glycerol molecule is bound in the pocket around Tyr[13] (Rao *et al.*, 1995; Stec *et al.*, 1995). On the contrary, the highly conserved Tyr[13] of the VT A3 and B2 is involved in the phosphate binding, suggesting a different mechanism of action.

Electrostatic forces were hypothesised to play a role in the interaction with the charged head groups of membrane components, which was supported by the finding that their charge is strongly correlated with their toxicity. Both VT A3 and β-purothionin display an overall high positive electrostatic potential, especially between the two arms of the letter L. In addition, VT A3 shows a larger difference in the potential between the two faces of the molecule, i.e. around the binding pocket and the dimer interface. However, a polycationic structure alone, like a poly-Lys, poly-Arg or the unfolded VT A3, would probably not be sufficient to trigger cell lysis (Büssing *et al.*, 1999). It is therefore supposed that the positively charged binding pocket is an indispensable condition for the interaction with the negatively charged groups of the membrane, but there may be additional structural factors that can influence molecular interactions with membrane components and determine the toxicity. The residues constituting the binding pocket belong to the most conserved amino acids in the viscotoxin sequences, thus they cannot account for the differences in the toxic activity.

The hydrophobic residues constituting the dimer interface in the VT A3 structure vary considerably among the different isoforms. As described by Büssing *et al.* (1999), variations in the side chain of residue at position 25 (Thr in VT A3, Val in VT A2 and Arg in VT B), i.e. exactly in the non-conserved region, can enormously influence the cytotoxic potential. In VT A3 this position is a Thr residue that takes part in the dimer formation. Its substitution by a larger and positively charged Arg introduces steric and electrostatic changes in the dimer interface. As described by Coulon *et al.* (2003) as well, the average accessible surface of hydrophobic residues located underneath the two α-helices — not taking the dimerisation into account — is significantly higher for VT A3 than for B2, that perfectly corresponds to the differences in their activity. In both isoforms, however, around half of the hydrophobic surfaces lie upon the plane formed by the pair of α-helices (*Figure 3.22*). It is therefore supposed that besides the presence of an intact (not iodinated) and positively charged binding

pocket a sterically accessible hydrophobic face is necessary for membrane interactions. The variability of side chains at the dimer interface may allow the specific binding of different membrane constituents and offer various patterns for molecular recognition. This conclusion is further supported by the observed flexibility of the molecule at the hinge region between the two arms of the letter L, which is illustrated by the conformational differences between the NMR and X-ray models (*Figure 3.9*).
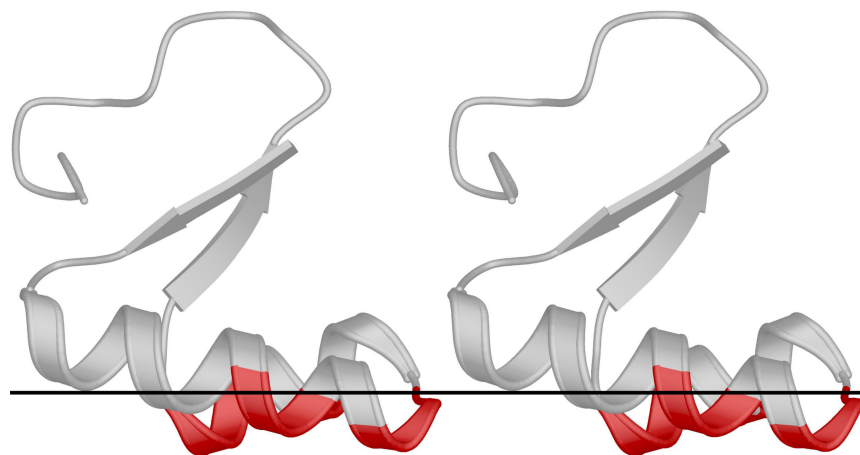
Figure 3.22. Hydrophobic residues in the variable region (drawn in red). These amino acids occupy the positions under the plane defined by the amphipathic α-helix-pair (represented by a black line).

The questions as to whether the dimer connected by hydrophobic interactions or the tetramer-like associate correspond to the physiological unit, and whether they play any role in membrane disruption, remain open. There are, however, indications that thionins associate at higher concentrations (Ohtani *et al.*, 1977, Milbradt *et al.*, 2003). A higher stability of the hydrophobic dimer can be assumed in the case of viscotoxins, since most of the hydrophobic residues are protected from the solvent through this association.

SUMMARY

Two cysteine-rich, anticarcinogenic microproteins belonging to distinct protein classes were studied with X-ray crystallographic methods.

1. The **lima bean trypsin-chymotrypsin inhibitor** (LBTI) can be classified into the Bowman-Birk serine proteinase family that comprises several double-headed serine proteinase inhibitors typical of leguminous plants. The unusually high cysteine content and the high symmetry space group enabled the structure solution of the free inhibitor by means of the anomalous signal of sulfur atoms measured in house, and the high solvent content of the crystals made the subsequent phase improvement step extremely powerful. The LBTI structure resembles the bow-tie fold canonical in the Bowman-Birk family. The inhibitor molecules associate to dimers in the crystal structure in such a manner that the reactive site loops that display a highly conserved geometry remain unshielded.

2. Structure of the ernary **complex of LBTI with its both target enzymes** could be easily solved by molecular replacement. The structure shed light on important factors that can influence the enzyme preference of inhibitors in the Bowman-Birk family, such as the asymmetric distribution of hydrophobic and hydrophilic patches on the enzyme-inhibitor interface. Comparison of the complexed and free LBTI molecules revealed conformational changes that are hypothesised to ease the precise adjustment of the two enzyme molecules upon simultaneous binding.

3. Viscotoxins (VTs) belong to the family of α- and β-thionins, a group of low molecular weight plant proteins that were showed to exert cytotoxic activity and be able to disrupt biological membranes. Crystal structure of **the VT isoform A3** was also solved exploiting the weak anomalous signal of sulfur atoms measured in house. In this case the medium resolution anomalous signal required an improvement in disulfide location for successful phasing. VT A3 molecules adopt the canonical α- and β-thionin fold that is similar to a capital letter L. A positively charged binding pocket was located between the two branches of the L structure, where a putative phosphate ion is bound. The binding site is assumed to be essential for successful membrane interactions. Molecules of the asymmetric unit were observed to form dimers and to associate to tetramers.

4. Favourable diffraction properties of crystals of **the viscotoxin isoform B2** enabled data collection to atomic resolution and structure solution with *ab initio* direct methods. Differences in the relative position of the monomers in the dimer and sequence alignment analysis suggested that hydrophobic residues responsible for high cytotoxic activity are located on the external surface of the amphipathic α-helices constituting the longer arm of the letter L.

REFERENCES

Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in structure determination of bovine mitochondrial F1 ATPase. *Acta Cryst.* **D52**, 30-42.

Armstrong, W. B., Wan, X. S., Kennedy, A. R., Taylor, T. H. & Meyskens, F. L. Jr. (2003). Development of the Bowman-Birk inhibitor for oral cancer chemoprevention and analysis of Neu immunohistochemica staining intensity with Bowman-Birk inhibitor concentrate treatment. *Laryngoscope* **113**, 1687-1702.

Baba, T., Watanabe, K., Kashiwabara, S. & Arai, Y. (1989). Primary structure of human proacrosin deduced from its cDNA sequence. *FEBS Lett.* **27**, 296-300.

Barrett, A. J. (1986). An introduction to proteinases. In: Proteinase inhibitors. (A. J. Barrett & G. S. Salvesen eds.) Elsevier Science Publishers (Biochemical Division).

Békés, F. & Lásztity, R. (1981). Isolation and determination of amino acid sequence of avenothionin, a new purothionin analogue from oat. *Cereal Chem.* **58**, 360-361.

Békés, F., Smied, I. & Lásztity, R. (1982). Investigations of thionins of wheat, barley, rye and oat. *Devel. Food Sci.* **5**, 163-168.

Birk, Y. (1961). Purification and some properties of a highly active inhibitor of trypsin and alpha-chymotrypsin from soybeans. *Biochim. Biophys. Acta* **54**, 378-381.

Birk, Y. (1976). Lima bean trypsin inhibitors. *Methods Enzymol.* **45**, 707-709.

Birk, Y. (1985.) The Bowman-Birk inhibitor. *Int. J. Peptide Protein Res.* **25**, 113-131.

Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interactions with proteinases. *Eur. J. Biochem.* **204**, 433-453.

Bohlmann, H. & Apel, K. (1987). Isolation and characterization of cDNAs coding for leaf-specific thionin closely related to the endosperm-specific hordothionin of barley (*Hordeum vulgare* L.) *Mol. Gen. Genet.* **207**, 446-454.

Brauer, A. B. E., Kelly, G., Matthews, S. J. & Leatherbarrow, R. J. (2002). The ^{1}H-NMR solution structure of the antitryptic core peptide of Bowman Birk inhibitor proteins: a minimal 'canonical loop'. *J. Biomol. Struct. Dyn.* **20**, 59-70.

Bricogne, G. (1974). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Cryst.* **A30**, 395-405.

Bruker Nonius (2002). *PROTEUM, SAINT, SADABS* and *XPREP* computer programs.

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

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nigles, M., Pannu, M. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905-921.

Burley, S. K. & Petsko, G. A. (1988). Weakly polar interactions. *Advanced Protein Chem.* **39**, 125-189.

Büssing, A., Stein, G. M., Wagner, M., Wagner, B., Schaller, G., Pfüller, U. & Schietzel, M. (1999). Accidental cell death and generation of reactive oxygen intermediates in human lymphocytes induced by thionins from *Viscum album* L. *Eur. J. Biochem.* **262**, 79-87.

Carrasco, L., Vázquez, D., Hernández-Lucas, C., Carbonero, P. & García-Olmedo, F. (1981). Thionins: plant peptides that modify membrane permeability in cultured mammalian cells. *Eur. J. Biochem.* **116**, 185-189.

Catalano, M., Ragona, L., Molinari, H., Tava, A. & Zetta, L. (2003). Anticarcinogenic Bowman Birk inhibitor isolated from snail medic seeds (*Medicago scutellata*): solution structure and analysis of self-association behaviour. *Biochemistry* **42**, 2836-2846.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Chen, P., Rose, J., Love, R., Wei, C. & Wang, B. C. (1992). Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *J. Biol. Chem.* **267**, 1990-1994.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986). The three-dimensional structure of α_1-purothionin in solution: combined use of nuclear resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5**, 2729-2735.

Clore, G. M., Sukumaran, D. K., Gronenborn A. M., Teeter, M. M. & Whitlow, M. (1987a). Nuclear magnetic resonance study on the solution structure of α_1-purothionin. *J. Mol. Biol.* **193**, 571-578.

Clore, G. M., Sukumaran, D. K., Nilges, M. & Gronenborn A. M. (1987b). Three-dimensional structure of phoratoxin in solution. Combined use of nuclear resonance, distance geometry and restrained molecular dynamics. *Biochemistry* **26**, 1732-1745.

Cochran, W. (1955). Relations between the phases of structure factors. *Acta Cryst.* **8**, 473-478.

Connor, J., Bucana, C., Fidler, I. J. & Schroit, A. J. (1989). Differentiation-dependent expression of phosphatidylserin in mammalian plasma membranes: qualitative assessment of outer-leaflet lipid by prothrombinase complex formation. *Proc. Natl. Acad. Sci.* **86**, 3184-3188.

Coulon, A., Berkane, E., Sautereau, A. M., Ureck, K., Rougé, P. & Lopez, A. (2002). Modes of membrane interaction of a natural cystein-rich peptide:viscotoxin A3. *Biochim. Biophys. Acta* **1559**, 145-159.

Coulon, A., Mosbah, A., Lopez, A., Sautereau, A. M., Schaller, G., Urech, K., Rougé, P. & Darbon, H. (2003). Comparative membrane interaction study of viscotoxins A3, A2 and B from mistletoe (*Viscum album*) and connections with their structures. *Biochem. J.* **374**, 71-78.

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.

Cruickshank, D. W. J. (1970). Least-squares refinement of atomic parameters. *Crystallographic computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 187-196. Munksgaard, Copenhagen.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. *Acta Cryst.* **D50**, 414-423.

Davie, E. W., Fujikawa, K. & Kisiel, W. (1991). The coagulation cascade: initiation, maintenance and regulation. *Biochemistry* **30**, 10363-10370.

de Caleya, R. F., Gonzalez-Pascual, B., García-Olmedo, F. & Carbonero, P. (1972). Susceptibility of phytopathogenic bacteria to wheat purothionins in vitro. *Appl. Microbiol.* **20**, 998-1000.

DeLano, W. L. (2003). The PyMOL molecular graphics system. *DeLano Scientific LLC*, San Carlos, CA, USA.

Dittmann, K. H., Mayer, C. & Rodemann, H. P. (2003). Radioprotection of normal tissue to improve radiotherapy: the effect of the Bowman-Birk protease inhibitor. *Curr. Med. Chem. Anticanc. Agents* **3**, 360-363.

Durant, M. (1976). Who named the Daisy? Who named the Rose? Dodd, Mead and Co., NY

Eckelkamp, C., Ehmann, B. & Schopfer, P. (1993). Wound-induced systemic accumulation of a transcript coding for a Bowman-Birk trypsin inhibitor-related protein in maize (*Zea mays* L.) seedlings. *FEBS Lett.* **323**, 73-76.

Epple, P., Apel, K. & Bohlmann, H. (1997). Overexpression of endogenous thionin enhances resistance of *Arabidopsis* agains *Fusarium oxosporum*. *Plant Cell* **9**, 509-520.

Esnouf, R. M. (1997). An extensively modified version of molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 132-134.

Evans, J., Wang, Y., Shaw, K. P. & Vernon, L. P. (1989). Cellular responses to Pyrularia thionin are mediated by Ca^{2+} influx and phospholipase A_2 activation and are inhibited by thionin tyrosin iodination. *Procl. Natl. Acad. Sci.* **86**, 5849-5853.

Finkenstadt, W. R., Hamid, M. A., Mattis, I. A., Schrode, I., Sealock, R. W., Wang, D. & Laskowski M. Jr. (1974). Kinetics and thermodynamics of the interaction of proteinases with protein inhibitors In: Proteinase inhibitors (H. Fritz, H. Tschesche, L. J. Greene & E. Truscheit eds.) (Bayer-symp. V), Springer-Verlag, Berlin, pp. 389-341.

Flecker, P. (1993). In: Proteinase inhibitors as cancer chemopreventive agents. (W. Troll & A. R. Kennedy eds.) pp 161-176, Plenum Press, New York.

Florack, D. E. & Stiekema, W. J. (1994). Thionins: properties, possible biological roles and mechanisms of action. *Plant Mol. Biol.* **26**, 25-27.

Fraenkel-Conrat, H., Bean, R. C., Ducay, E. D. & Olcott, H. (1952). Isolation and characterization of a trypsin inhibitor from lima beans. *Arch. Biochem. Biophys.* **37**, 393-407.

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

Funk, A., Weder, J. K. & Belitz, H. D. (1993). Primary structures of proteinase inhibitors from *Phaseolus vulgaris* var. *nanus* (cv. *Borlotto*). *Z. Lebensm. Unters. Forsch.*, **196**, 343-350.

Gaier, J. A., Tulinsky, A. & Liener, I. E. (1981). Formation, crystallisation, and preliminary crystallographic data of the ternary complex of α-chymotrypsin, β-trypsin, and the Bowman-Birk inhibitor. *J. Biol. Chem.* **256**, 11417-11419.

Gaultier, R. & Chevalier, J. (1907). *C. R.* **145**, 941. Cited in: Winterfeld & Bjil, 1948.

Girmann, B. (2003). Beiträge zur Isolierung und Strukturaufklärung der Viscotoxine aus *Viscum album* L. sowie von cytotoxischen Komponenten der Gilvocarcin-Reihe aus Streptomyceten. PhD thesis, University of Göttingen, Germany

Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, G., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradszky, K. & Rutter, W. J. (1988). Electrostatic complementarity within the substrate binding pocket of trypsin. *Proc. Natl. Acad. Sci.* **85**, 4961-4965.

Grazi, G., Zemp, M. (1985). Über den Ginster (*Genista cinerea*) als Sammelwirt für Laubholzmistel und Kiefernmistel, *8. Symp. Morphologie, Anatomie, Systematik, Hamburg*

Ham, W. E. & Sandstedt, R. M. (1944). A proteolytic inhibiting substance in the extract from unheated soybean meal. *J. Biol. Chem.* **154**, 505-506.

Hampel, A., Labanauskas, M., Connors, P. G., Kirkegard, L., RayBhandary, U. L., Sigler, P. B. & Bock, R. M. (1968). Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's. *Science* **162**, 1384-1387.

Han, K. H., Park, K. H., Yoo, H. J., Cha, H., Suh, S. W., Thomas, F., Moon, T. S. & Kim, S. M. (1996). Determination of the three dimensional structure of the hordothionin-α by nuclear magnetic resonance. *Biochem. J.* **313**, 885-892.

Hatano, K., Kojima, M., Tanokura, M. & Takahashi, K. (1996). Solution structure of bromelian inhibitor VI from pineapple stem: structural similarity with Bowman-Birk trypsin/chymotrypsin inhibitor from soybean. *Biochemistry* **35**, 5379-5384.

Hawksworth, F. G. (1980). Crop loss assessment. Proc. E. C. Stakman Commemorative Symp. Univ. Wmn. Exp. Stn. Misc. Publ. 7.

Haynes, R. & Feeney, R. E. (1967). Fractionation and properties of trypsin and chymotrypsin inhibitors from lima bean. *J. Biol. Chem.* **242**, 5378-5385.

Hedstrom, L., Szilágyi, L. & Rutter, W. J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**, 1249-1253.

Hendrickson, W. A. & Teeter, M. M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**, 107-113.

Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). Direct phase determination based on anomalous scattering. *Methods Enzymol.* **115**, 41-54.

Hernández-Lucas, C., Royo, J., Paz-Ares, J., Ponz, F., García-Olmedo, F. & Carbonero, P. (1986). Polyadenylation site heterogenity in mRNA encoding the precursor of the barley toxin α-hordothionin. *FEBS Lett.* **200**, 103-105.

Hirono, S., Akagawa, H., Mitsui, Y. & Iitaka, Y. (1984). Crystal structure of the 2.6 Å resolution complex of subtilisin BPN' with Streptomyces subtilisin inhibitor. *J. Mol. Biol.* **178**, 389-413.

Holtorf, S., Ludwig-Müller, J., Apel, K. & Bohlmann, H. (1998). High level expression of a viscotoxin in *Arabidopsis thaliana* gives enhanced resistance against *Plasmodiophora brassicae*. *Plant Mol. Biol.* **36**, 673-680.

Hughes, P., Dennis, E., Whitecross, M., Llewellyn, D. & Gage, P. (2000). The cytotoxic plant protein, β-purothionin, forms ion channels in lipid membranes. *J. Biol. Chem.* **275**, 823-827.

Hutchinson, J. (1969). Evolution and phylogeny of flowering plants. Academic Press, London, New York.

Ikenaka, T. & Norioka, S. (1986). Bowman-Birk family serine proteinase inhibitors. In: Proteinase inhibitors. (A. J. Barrett *&* G. S. Salvesen eds.) Elsevier Science Publishers (Biochemical Division), Chapter 9.

Ishikawa, C., Nakamura, S., Watanabe, K. & Takahashi K. (1979). The amino acid sequence of adzuki bean proteinase inhibitor I. *FEBS Lett.* **99**, 97-100.

Jaeger, E. C. (1994). A source-book of Biological names and terms. Charles C, Thomas, Springfield, IL

James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C. A. (1980). Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. A model for serine protease catalysis. *J. Mol. Biol.* **144**, 43-88.

Jancarik, J. & Kim, S. H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst*. **24,** 409-411.

Jung, M. L., Baudino, S., Ribéreau-Gayon, G. & Beck, J. P. (1990). Characterization of cytotoxic proteins from mistletoe (*Viscum album* L.). *Cancer Lett.* **51**, 103-108.

Karle, J. & Hauptman, H. (1956). A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, 3P$_1$2, 3P$_2$2. *Acta Cryst.* **9**, 635-651.

Kassell, B. (1976). Trypsin inhibitors from other legumes. *Methods Enzymol.* **19**, 862-871.

Kelly, P., Reid, N. & Davies, I. (1997). Effects of experimental burning, defoliation and pruning on survival and vegetative respoting in mistletoe. *Int. J. Plant Sci.* **158**, 856-861.

Kennedy, A. R. (1994). Prevention of carcinogenesis by protease inhibitors. *Cancer Res.* **54**, 1999-2005 (suppl.)

Kennedy, A. R. (1998a). The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. Am. *J. Clin. Nutr.* **68**, 1406-1412 (suppl.).

Kennedy, A. R. (1998b). Chemopreventive agents: protease inhibitors. *Pharmacol. Ther.* **78**, 167-209.

Kienle, G. S. & Kiene, H. (2003). Die Mistel in der Onkologie, Schattauer Verlag, Stuttgart

Kirkup, D. W., Polhill, R. M. & Wiens, D. (2000a). Mistletoe, the genus *Viscum*., editor: Büssing, A., Harwood Academic Publishers, Singapore, Amsterdam, pp 1-6.

Kirkup, D. W., Polhill, R. M. & Wiens, D. (2000b). Mistletoe, the genus *Viscum*., editor: Büssing, A., Harwood Academic Publishers, Singapore, Amsterdam, pp 7-31.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Cryst.* **D55**, 484-491.

Kiyohara, T., Yokota, K., Masaki, Y., Matsui, O., Iwasaki T. & Yoshikawa M. (1981). The amino acid sequences of proteinase inhibitors I-A and I-A' from adzuki beans. *J. Biochem.* **90**, 721-728.

Kleywegt, G. J., Boelens, R., Cox, M., Llinas, M. & Kaptein, E. (1991). Computer assisted assignment of 2D ^{1}H-NMR spectra of proteins: basic algorithms and aplication to phoratoxin B. *J Biomol. NMR* **1**, 23-47.

Kobrehel, K., Yee, B. C. & Buchanan, B. B. (1991). Role of the NADP/thioredoxin system in the reduction of alpha-amylase and trypsin inhibitor proteins. *J. Biol. Chem.* **266**, 16135-16140.

Koepke, J., Ermler, U., Warkentin, E., Wenzl, G. & Flecker, P. (2000). Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 Å resolution. Structural basis of Janus-faced serine protease inhibitor spacificity. *J. Mol. Biol.* **298**, 477-491.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Krishna Murthy, H. M., Judge, K., DeLucas, L. & Padmanabhan, R. (2000). Crystal structure of the dengue virus NS3 protease in complex with a Bowman-Birk inhibitor: implications for flaviviral polyprotein processing and drug design. *J. Mol. Biol.* **301**, 759-767.

Lankisch, P. G. & Vogt, W. (1971). Potentiation of hemolysis by the combined action of phospholipase A and a basic peptide containing S-S bonds (viscotoxin B). *Experientia*, **27**, 122-123.

Laskowski, M. Jr. & Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593-626.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Lesk, A. M. & Fordham, W. D. (1996). Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *J. Mol. Biol.* **258**, 501-537.

Li de la Sierra, I., Quillien L., Flecker, P., Gueguen, J. & Brunie, S. (1999). Dimeric crystal structure of a Bowman-Birk protease inhibitor from pea seeds. *J. Mol. Biol.* **285**, 1195-1207.

Li, S. S., Gullbo, J., Lindholm, P., Larsson, R., Thunberg, E., Samuelsson, G., Bohlin, L. & Claeson, P. (2002). Ligatoxin B, a new cytotoxic protein with a novel helix-turn-helix DNA-binding domain from the mistletoe *Phoradendron liga*. *Biochem. J.* **366**, 405-413.

Li, Y., Huang, Q., Zhang, S., Liu, S., Chi, C. & Tang, Y. (1994). Studies on artificial trypsin inhibitor peptide derived from the mung bean trypsin inhibitor: chemical synthesis, refolding and crystallographic analysis of its complex with trypsin. *J. Biochem.* **116**, 18-25.

Liao, D., Breddam, K., Sweet, R. M., Bullock, T. & Remington, S. J. (1992). Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry* **31**, 9796-9812.

Liener, I. E. (1995). Possible adverse effects of soybean anticarcinogens. *J. Nutr.* **125**, 744-750 (suppl.)

Lin, G., Bode, W., Huber, R., Chi, C. & Engh, R. A. (1993). The 0.25-nm X-ray structure of the Bowman-Birk-type inhibitor from mung bean in ternary complex with porcine trypsin. *Eur. J. Biochem.* **212**, 549-555.

Lu, W., Apostol I., Quasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. & Laskowsky, M. A. Jr. (1997). Binding of amino acid side chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441-461.

Mak, A. S. & Jones, B. L. (1976). The amino acid sequence of the wheat β-purothionin. *Can. J. Biochem.* **22**, 835-842.

Maki, P. A., Paterson, Y. & Kennedy, A. R. (1994). Studies related to the potential antigenicity of the Bowman-Birk inhibitor, an anticarcinogenic protease inhibitor isolated from soybeans. *Nutr. Cancer* **22**, 185-193.

Marx, U. C., Korsinczky, M. L. J., Schirra, H. J., Jones, A., Condie, B., Otvos, L. Jr. & Craik, D. J. (2003). Enzymatic cyclization of a potent Bowman-Birk protease inhibitor, sunflower trypsin inhibitor-1, and solution structure of an acyclic precursor peptide. *J. Biol. Chem.* **287**, 21782-21789.

McLachlan, A. D. (1979). Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* **121**, 49-79.

McRee, D. E. (1999). XtalView/Xfit — a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.

Mellstrand, S. T. & Samuelsson, G. (1974). Phoratoxin, a toxic protein from the mistletoe *Phoradendron tomentosum* ssp *macrophyllum* (*Lorathaceae*). The amino acid sequence. *Acta Pharm. Suec.* **11**, 347-360.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Milbradt, A. G., Kerek, F., Moroder, L. & Renner, C. (2003). Structural characterisation of hellethionins from *Helleborus purpurascens*. *Biochemistry* **42**, 2404-2411.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Odani, S. & Ikenaka, T. (1973). Scission of soybean Bowman-Birk proteinase inhibitor into two small fragments having either trypsin or chymotrypsin activity. *J. Biochem.* **74**, 857-860.

Odani, S. & Ikenaka, T. (1978). Studies on soybean trypsin inhibitors. XIII. Preparation and characterisation of active fragments from Bowman-Birk proteinase inhibitor. *J. Biochem.* **83**, 747-753.

Ohtani, S., Okada, T., Yoshizumi, H. & Kagamiyama, H. (1977). Complete primary structures of two subunits of purothionin A, a lethal protein for brewer's yeast. *J. Biochem.* **82**, 753-767.

Olson, T. & Samuelsson, G. (1972). The amino acid sequence of viscotoxin A2 from the Europen mistletoe (*Viscum album* L., *Loranthaceae*). *Acta Chem Scand.* **26**, 585-595

Orrú, S., Scaloni, A., Giannattasio, M., Urech, K., Pucci, P. & Schaller, G. (1997). Amino acid sequence, S-S bridge arrangement and distribution in plant tissues of thionins from *Viscum album*. *Biol. Chem.* **378**, 989-996.

Osborne, T. B. & Mendel, L. B. (1917). The use of soybeans as food. *J. Biol. Chem.* **32**, 369-387.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Ozaki, Y., Wada, K., Hase, T., Matsubara, H., Nakanishi, T. & Yoshizumi, H. (1980). Amino acid sequence of purothionin homolog from barley flour. *J. Biochem.* **87**, 549-555.

Pannu, N. S. & Read, R. J. (1996). Improved refinement through maximum likelihood. *Acta Cryst.* **A52**, 659-668.

Park (1881). *The practioner* **271**, 348. Cited in: Winterfeld & Bjil, 1948.

Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463.

Polgár, L. (1989). Structure and function of serine proteinases. In: Mechanisms of proteinase action. Boca Raton, Florida: CRC Press. Chapter 3.

Quast, V., Engel, J., Steffen, E., Tschesche, H. & Kupfer, S. (1978). Kinetics of the interaction of alpha-chymotrypsin with trypsin kallikrein inhibitor (Kunitz) in which the reactive-site peptide bond Lys-15--Ala-16 is split. *Eur J Biochem.* **86**, *353-60.*

Ramachandran, G. N. & Shasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Prot. Chem.* **28**, 283-437.

Rao, U., Stec, B. & Teeter, M. M. (1995). Refinement of purothionins reveals solute particles important for lattice formation and toxicity. Part I. α_1-purothionin revisited. *Acta Cryst.* **D51**, 904-913.

Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. D., Sim, R. B., Kristensen, T. & Tack, B. F. (1986). Complement system proteins which interact with C3B or C4B. *Immunol. Today* **7**, 230-234.

Ribéreau-Gayon, G., Jung, M. L., Baudino, S., Salle, G. & Beck, J. P. (1986). Effect of mistletoe (*Viscum album* L.) extract on cultured tumor cells. *Experientia* **42**, 594-599.

Romagnoli, S., Fogolari, F., Catalano, M., Zetta, L., Schaller, G., Urech, K., Giannattasio, M., Ragona, L. & Molinari, H. (2003). NMR solution structure of viscotoxin C1 from *Viscum album* species *Coloratum ohwi*: toward a structure-function analysis of viscotoxins. *Biochemistry* **42**, 12503-12510.

Romagnoli, S., Ugolini, R., Fogolari, F., Schaller, G., Urech, K., Giannattasio, M., Ragona, R. & Molinari, H. (2000). NMR structural determination of viscotoxin A3 from *Viscum album* L. *Biochem. J.* **350**, 569-577.

Sakal, E., Applebaum, S. W. & Birk, Y. (1989). Purification and characterization of trypsins from the digestive tract of *Locusta migratoria*. *Int. Peptide Protein Res.* **34**, 498-505.

Samudzi, C., Schroeder, S., Griffith, S., Chen, X. & Quinn, T. P. (1997). Crystallization and preliminary studies of lima bean trypsin inhibitor. *Proteins* **27**, 311-314.

Samuelsson, G. & Jayawardene, A. L. (1974). Isolation and characterization of viscotoxin 1-PS from *Viscum album* L. ssp *austriacum* (Wiesb.) Vollmann, growing on Pinus silvestris. *Acta Pharm. Suec* **11**, 175-184.

Samuelsson, G. & Pettersson, B. (1971a). The disulfide bonds of viscotoxin A3 from the European mistletoe (*Viscum album* L., *Loranthaceae*). *Acta Chem. Scand.* **25**, 2048-2054.

Samuelsson, G. & Pettersson, B. (1971b). The amino acid sequence of viscotoxin B from the European mistletoe (*Viscum album* L., *Loranthaceae*). *Eur. J. Biochem.* **21**, 86-89.

Samuelsson, G. & Pettersson, B. (1977). Toxic proteins from the mistletoe *Dendrophtora clavata* II. The amino acid sequence of denclatoxin B. *Acta Pharm. Suec.* **14**, 245-254.

Samuelsson, G., Seger, L. & Olson, T. (1968). The amino acid sequence of oxidized viscotoxin A3 from the European mistletoe (*Viscum album* L., *Loranthaceae*). *Acta Chem. Scand.* **22**, 2624-2642.

Sayre, D. (1974). Least squares phase refinement II. High resolution phasing of a small protein. *Acta Cryst.* **A30**, 180-184.

Schaller, G., Urech, K. & Giannattasio, M. (1996). Cytotoxicity of different viscotoxins and extracts from the European supspecies of *Viscum album* L. *Phytother. Res.* **10**, 473-477.

Schaller, G., Urech, K. & Giannattasio, M. (2000). Viscotoxin in den drei Unterarten von *Viscum album* L. *Mistilteinn* **2000/1**, 32-40.

Schechter, I. & Berger, A. (1967). On the size of the active site in proteinases. *Biochem. Biophys. Res. Commun.* **27**, 157-162.

Schmeil, O. & Fitschen, J. (1993). Flora von Deutschland und angrenzender Länder, 89. neubearbeitete und erweiterte Auflage, Quelle & Meyer Verlag, Heidelberg, Wiesbaden.

Schneider, T. R & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Cryst.* **D58**, 1772-1779.

Schrader, G. & Apel, K. (1991). Isolation and characterization of cDNAs encoding viscotoxins from mistletoe (*Viscum album*). *Eur. J. Biochem.* **198**, 549-553.

Schuller, D. J. (1996). MAGICSQUASH: More versatile non-crystallographic averaging with multiple constraints. *Acta Cryst.* **D52**, 425-434.

Schulze, E. D. & Ehleringer, J. R. (1984). The effect of nitrogen supply on growth and water-use efficiency of xylem-tapping misltetoes. *Planta* **162**, 268-275.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high resolution refinement. *Methods. Enzymol.* **277**, 319-343.

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

Sheldrick, G. M., Hauptmann, H. A., Weeks, C. M., Miller, M. & Usón, I. (2001). *Ab initio* phasing. *International Tables for crystallography*, Vol. *F*, edited by E. Arnold & M. Rossmann, pp. 333-351. Dodrecht: Kluwer Academic Publishers.

Smith, R. J. & Gallon, J. R. (1993). Nitrogen fixation. In: Plant biochemistry and molecular biology. (P. J. Lea & R. C. Leegood eds.) pp 129-153. John Wiley and Sons, New York

Song, H. K., Kim, Y. S., Yang, J. K., Moon, J., Lee, J. Y. & Suh, S. W. (1999). Crystal structure of a 16 kDa double-headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *J. Mol. Biol.* **293**, 1133-1144.

Stec, B., Rao, U. & Teeter, M. M. (1995). Refinement of purothionins reveals solute particles important for lattice formation and toxicity. Part II. Structure of β-purothionin at 1.7Å. *Acta Cryst.* **D51**, 914-924.

Stein, G. M., Schaller, G., Pfüller, U., Schietzel, M. & Büssing, A. (1999b). Thionins from *Viscum album* L.: influence of the viscotoxins on the activation of granulocytes. *Anticancer Res.* **19**, 1037-42.

Stein, G. M., Schaller, G., Pfüller, U., Wagner, M., Wagner, B., Schietzel, M. & Büssing, A. (1999a). Characterisation of granulocyte stimulation by thionins from European mistletoe and from wheat. *Biochim. Biophys. Acta* **1426**, 80-90.

Steiner, R. (1961). Geisteswissenschaft und Medizin, 4. Auflage, Rudolf Steiner Nachlaßverwaltung, Dornach

Stevens, F. C. & Doskoch, E. (1973). Lima bean proteinase inhibitor: reduction and reoxidation of the disulfide bonds and their reactivity in the trypsin inhibitor complex. *Can. J. Biochem.* **51**, 1021-1028.

Stevens, F. C., Wuerz, S. & Krahn J. (1974). Structure-function relationships in lima bean protease inhibitor. In: Proteinase inhibitors (H. Fritz, H. Tschesche, L. J. Greene & E. Truscheit eds.) (Bayer-symp. V), Springer-Verlag, Berlin, pp. 344-354.

Suzuki, A., Tsunogae, Y., Tanaka, I., Yamane, T., Ashida, T., Norioka, S., Hara, S. & Ikenaka, T. (1987). The structure of the Bowman-Birk type protease inhibitor A-II from peanut (*Arachis hypogaea*) at 3.3 Å resolution. *J. Biochem.* **101**, 267-274.

Tabiasco, J., Pont, F., Fournié, J. J. & Vercellone, A. (2002). Mistletoe viscotoxins increase natural killer cell-mediated cytotoxicity. *Eur. J. Biochem.* **269**, 2591-2600.

Tan, C. G. L. & Stevens, F. C. (1971a). Amino acid sequence of the lima bean protease inhibitor component IV. 1. Isolation and sequence determination of the tryptic peptides. *Eur. J. Biochem.* **18**, 503-514.

Tan, C. G. L. & Stevens, F. C. (1971b). Amino acid sequence of the lima bean protease inhibitor component IV. 2. Isolation and sequence determination of the chymotryptic peptides and the complete amino acid sequence. *Eur. J. Biochem.* **18**, 514-523.

Tan-Wilson, A. L., Hartl, P. M., Delfel, N. E. & Wilson, K. A. (1985). Differential expression of Kunitz and Bowman-Birk soybean proteinase inhibitors in plant and callus tissue. *Plant Physiol.* **78**, 310-314.

Tauber, H., Kershaw, B. B. & Wright R. D. (1949). Studies on the growth inhibitor fraction of lima beans and isolation of a crystalline heat-stable trypsin inhibitor. *J. Biol. Chem.* **179**, 1155-1161.

Teeter, M. M., Mazer, J. A. & L'Italien J. J. (1981). Primary structure of the hydrophobic plant protein crambin. *Biochemistry* **2**, 5437-5443.

Terwilliger, T. C. (1999). Reciprocal space solvent flattening. *Acta Cryst.* **D55**, 1863-1871.

Thevissen, K., Ghazi, A., De Samblanx, G. W., Brownlee, C., Osborn, R. W. & Brockaert W. F. (1996). Fungal membrane responses induced by plant defensins and thionins. *J. Biol. Chem.* **271**, 15018-15025.

Thunberg, E. & Samuelsson, G. (1982). The amino acid sequence of ligatoxin A from the mistletoe *Phoradendron liga*. *Acta Pharm. Suec.* **19**, 447-456.

Thunberg, E. (1983). Phoratoxin B, a toxic protein from the mistletoe *Phoradendron tomentosum* subsp. *macrophyllum* (*Loranthaceae*). *Acta Pharm. Suec.* **20**, 115-122.

Tolcher, A. W., Kennedy A. R., Padley, J. R., Majeed, N., Pollak, M. & Kantoff, P. W. (2001). Other novel agents: rationale and current status as chemopreventive agents. *Urology* **57**, 86-89.

Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S. & Takahashi, K. (1986). Structure of the trypsin-binding domain of Bowman-Birk type protease inhibitor and its interaction with trypsin. *J. Biochem.* **100**, 1637-1646.

Urech, K., Schaller, G., Ziska, P. & Giannattasio, M. (1995). Comparative study on the cytotoxic effect of viscotoxin and mistletoe lectin on tumour cells in culture. *Phytother. Res.* **9**, 49-55.

Utsugi, T., Schroit, A. J., Connor, J., Bucana, C. D. & Fidler I. J. (1991). Elevated expression of phosphatidylserin in the outer membrane leaflet of human tumour cells and recognition by human blood monocytes. *Cancer Res.* **51**, 3062-3066.

van Campenhout, S., Sagi, L., Vander Stappen, J. & Volckaert, G. (1998). Characterisation of the type-I thionin loci from the A, B, D and R genomes of wheat and rye. *Theor. Appl. Genet.* **96**, 80-86.

Vernon, L. P., Evett, G. E., Zeikus, R. D. & Gray, W. R. (1985). A toxic thionin from *Pyrularia puberea*: purification, properties and amino acid sequence. *Arch. Biochem. Biophys.* **238**, 18-29.

von Tubeuf, K. (1923). Monographie der Mistel, Oldenbourg Verlag, München.

Voss, R. H., Ermler, U., Essen, L. O., Wenzl, G., Kim, Y. M. & Flecker, P. (1996). Crystal structure of the bifunctional soybean Bowman-Birk inhibitor at 0.28-nm resolution. Structural peculiarities in a folded protein conformation. *Eur. J. Biochem.* **242**, 122-131.

Wada, K., Ozaki, Y., Matsubara, H. & Yoshizumi, H. (1982). Studies on purothionin by chemical modifications. *J. Biochem.* **91**, 257-263.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90-112.

Ware, J. H., Wan, X. S. & Kennedy, A. R. (1999). Bowman-Birk inhibitor suppresses production of superoxide anion radicals in differentiated HL-60 cells. *Nutr. Cancer* **33**, 174-177.

Werner, M. H. & Wemmer, D. E. (1992). Three-dimensional structure of soybean trypsin-chymotrypsin Bowman-Birk inhibitor in solution. *Biochemistry* **31**, 999-1010.

Wilson, K. A. & Laskowski, M. Sr. (1975). The partial amino acid sequence of trypsin inhibitor II from garden bean, *Phaseolus vulgaris*, with location of the trypsin and elastase-reactive sites. *J. Biol. Chem.* **250**, 4261-4267.

Winterfeld, K. & Bjil, L. H. (1948). Viscotoxin, ein neuer Inhaltstoff der Mistel (*Viscum album* L.). *Liebig's Annales* **561**, 107-111.

Winterfeld, K. (1942). *Pharmaz. Ind.* **1942**, 37-41. Cited in: Winterfeld & Bjil, 1948.

Woynarowski, J. M. & Konopa, J. (1980). Interaction between DNA and viscotoxins. Cytotoxic, basic polypeptides from *Viscum album* L. *Hopple-Seyler's Phys. Chem.* **361**, 1535-1545.

Xu, Y., Wu, J., Gorenstein, D. & Braun, W. (1999). Automated 2D NOESY assignment and structure calculation of crambin (S22/I25) with selfcorrecting distance geometry based NOAH/DIAMOND programs. *J. Magn. Resonance* **136**, 76-85.

Yamano, A. & Teeter, M. M. (1994). Correlated disorder of the pure Pro^{22}/Leu^{25} form of crambin at 150 K refined to 1.05 Å resolution. *J. Biol. Chem.* **269**, 13956-13965.

Yavelow, J., Schepis, L. T., Nickels, J. Jr. & Ritchie, G. (1993). In: Proteinase inhibitors as cancer chemopreventive agents. (W. Troll & A. R. Kennedy eds.) pp 217-2256, Plenum Press, New York.

Zhang, K. Y. J. & Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta Cryst.* **A46**, 377-381.

Zhang, L., Wan, X. S., Donahue, J. J., Ware, J. H. & Kennedy, A. R. (1999). Effects of the Bowman-Birk inhibitor on clonogenic survival and cisplatin- or radiation-induced cytotoxicity in human breast, cervical, and head and neck cancer cells. *Nutr. Cancer* **33**, 165-173.

Zhang, Y., Luo, S., Tan, F., Qi, Z., Xu, L. & Zhang, A. (1982). Complete amino acid sequence of mung bean trypsin inhibitor. *Sci. Sin., Ser. B, Chem. Biol. Agric. Med. Earth Sci.* **25**, 268-277.

PUBLIKATIONEN IN FACHZEITSCHRIFTEN

[1]   L. Farkas, **J. Debreczeni**, V. Harmat, K. Kohama, A. Nakamura & L. Nyitray
*Structure of Ca^{2+}-bound regulatory domain of Physarum myosin II at 1.8 Å resolution:
    Ca^{2+} is bound to a conventional EF-hand that is in a closed lobe conformation*
J. Muscle Res. Cell Motil. **22** (2001) 591

[2]   L. Fitjer, R. Gerke, J. Wieser, G. Bunkóczi & **J. É. Debreczeni**
Helical primary structures of four-membered rings: (M)-trispiro[3.0.0.3.2.2]tridecane
Tetrahedron **59** (2003) 4443

[3]   C. Lehmann, **J. É. Debreczeni**, G. Bunkóczi, M. Dauter, Z. Dauter, L. Vértesy & G. M.
    Sheldrick
Structures of four crystal forms of decaplanin
Helvetica Chimica Acta, **86** (2003) 1478

[4]   **J. É. Debreczeni**, G. Bunkóczi, B. Girmann & G. M. Sheldrick
*In-house phase determination of the lima bean trypsin inhibitor: a low resolution
    sulfur-SAD case*
Acta Cryst. **D59** (2003) 393

[5]   **J. É. Debreczeni**, G. Bunkóczi, Q. Ma, H. Blaser & G. M. Sheldrick
In-house measurement of the sulfur anomalous signal and its use for phasing
Acta Cryst. **D59** (2003) 688

[6]   **J. É. Debreczeni**, B. Girmann, A. Zeeck, R. Krätzner & G. M. Sheldrick
Structure of viscotoxin A3: disulfide location from weak SAD data
Acta Cryst. **D59** (2003) 2125

KONFERENZBEITRÄGE

[1]   **J. Debreczeni**, G. Bunkóczi, V. Harmat, L. Farkas & L. Nyitray
Structure of the regulatory domain of Physarum polycephalum myosin
Poster at the Sixth International School on the Crystallography of Biological
    Macromolecules, Como, Italy, May 2001

[2]   **J. É. Debreczeni**, G. Bunkóczi, Q. Ma & G. M. Sheldrick
In-house measurement of the anomalous signal of sulfur and its use for phasing
Poster at the XIX. Congress and General Assembly of the International Union of
    Crystallography, Geneva, Switzerland, August 2002

[3]   **J. É. Debreczeni**, G. Bunkóczi, B. Girmann & G. M. Sheldrick
Novel structures solved using in-house S-SAD
Poster at the Workshop on Phasing with Soft X-rays, Bressanone/Brixen, Italy,
    February 2003

[4]   **J. É. Debreczeni**
Solving unknown structures with SHELXD/E
Oral presentation at the SHELX Workshop, Göttingen, Germany, September 2003

LEBENSLAUF

PERSÖNLICHE DATEN

Name	Judit É. Debreczeni
Geburtsdatum	19. Januar 1977
Geburtsort	Budapest, Ungarn
Staatsangehörigkeit	ungarisch
Familienstand	verheiratet

SCHULBILDUNG

1991-1995	Fazekas Mihály Gimnázium, Budapest, mit Abschluss Abitur

HOCHSCHULSTUDIUM

1995-2000	Eötvös Loránd Universität Budapest
1999-2000	Diplomarbeit am Lehrstuhl für Theoretische Chemie im Arbeitskreis von Prof. G. Náray-Szabó zum Thema: „Kristallisation und röntgenkristallographische Untersuchungen an der Regulationsdomäne des *Physarum polycephalum* Myosins"

PROMOTION

2000-2004	Dissertation am Lehrstuhl für Strukturchemie der Georg-August-Universität Göttingen im Arbeitskreis von Prof. G.M. Sheldrick zum Thema: „X-ray crystallographic studies on two cysteine-rich anti-carcinogenic microproteins"

Göttingen, den 16.03.2004

Judit É. Debreczeni