Structure determination of Viscotoxin A1, Tendamistat and Tri Peptidyl Peptidase – I

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August Universität zu Göttingen

vorgelegt von

Aritra Pal

aus Ranchi, Indien

Göttingen, 2008

D7

Anleiter: Prof. George M. Sheldrick

Referent: Prof. Dr. Ralf Ficner

Korreferent: Jr. Prof. Dr. Oliver Einsle

Tag der mündlichen Prüfung:24.01.2008

Acknowledgements

First of all I would like to thank **Prof. George M. Sheldrick** for giving me the opportunity to work in a competitive but at the same time relaxing environment. His concern, constant encouragement and appreciation have helped me in accomplishing my targets. Not only he is a great source of knowledge and inspiration, he is also the best supervisor one could ever get. I feel privileged to be one of his students.

I am grateful to **Prof. Dr. Ralf Ficner** for accepting to be referent of my thesis. Also I would like to thank **Jr. Prof. Dr. Oliver Einsle** for accepting to be co-referent of my thesis.

I am indebted to **Dr. Regine Herbst-Irmer** for her valuable guidance in small molecule crystallography. She provided me with an outlook to deal with twinning and disorders in small molecule crystallography and to refine the structures to perfection.

I am indebted to **Dr. Tim Grüne** and **Dr. Madhumati Sevvana** for teaching me the experimental skills and for constant support, encouragement and guidance during the entire course of my thesis work. Special thanks to **Tim** for helping me overcome various roadblocks during the structural analysis of Tripeptidyl Peptidase-I.

I express my gratitude to **PD Dr. Dr. Robert Steinfeld** and **Dr. Ralf Krätzner** for the fruitful collaboration with this group involving the structural analysis of Tripeptidyl Peptidase-I. I thank them for providing me with the protein samples and various useful discussions.

I am thankful to **Helmut Dehnhardt** for his technical assistance and keeping the diffractometers in a working state.

I thank **Dr. Stephan Rühl**, **Burkhard Heisen**, **Kathrin Meindl**, **Dr. Christine Schlicker** and **Dr. Marianna Biadene** for their continuous support during the course

List of Abbreviations

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BESSY	Berlin elektronen synchrotron
CC	Correlation coefficient
CCD	Coupled charge device
Cys	Cysteine
DESY	Deustches elektronen synchrotron
DNA	Deoxyribonucleic acid
EMBL	European molecular biology laboratory
Fom	Figure of merit
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HPLC	High performance liquid chromatography
Ile	Isoleucine
kDa	kiloDalton
KSCP	Kumamolysin serine crboxyl protease
Leu	Leucine
Lys	Lysine

MAD	Multiwavelength anomalous diffraction
Met	Methionine
MIR	Multiple isomorphous replacement
MPD	2-Methyl, 2,4-pentanediol
NCS	Non-Crystallographic Symmetry
PDB	Protein databank
PEG	Polyethylene glycol
Phe	Phenylalanine
Pro	Proline
PSCP	Pepstatin-insensitive carboxyl protease
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SAD	Single wavelength anomalous diffraction
SDS PAGE	Sodium-dodecylsulfate polyacrylamide gel electrophoresis
Ser	Serine
SIR	Single isomorphous replacement
SLS	Swiss Light Source
Tds	Tendamistat
Thr	Threonine
TPP-I	Tripeptidyl Peptidase-I
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Molecular Graphics

All three dimensional figures were drawn with CHIMERA (Petterson *et al.*, 2004) or drawn and rendered in PYMOL (DeLano, 2003).

Contents

Overview

The current work reports the crystal structures of three different proteins: Viscotoxin A1, Tendamistat and Tri Peptidyl Peptidase-I. While Viscotoxin A1 and Tendamistat can be used as a basis to design anticancer drugs and α –amylase inhibitors respectively, the structure of Tri Peptidyl Peptidase-I can be used as a template for further functional analysis of the enzyme and to study the structural aspects of the defective enzyme.

Viscotoxin A1 is a plant defense protein and is known to disrupt cellular membranes. Also it exhibits anticarcinogenic properties. The structure might shed some light on the mechanism of cellular membrane disruption by Viscotoxins and variance in potency to disrupt membranes among different isoforms of Viscotoxins.

Tendamistat is an α –amylase inhibitor from *Streptomyces tendae*. This structure may provide insight to design novel α –amylase inhibitor peptides using the conformation of the residues involved in target binding as a starting point in creating drugs to counter diabetes mellitus.

Viscotoxin A1 and Tendamistat can also serve as excellent examples for developing X-Ray techniques and to study protein folding and dynamics.

Tri Peptidyl Peptidase-I (TPP-I) is a member of the newly characterized sedolisin class of serine protease with a unique Ser-Glu-Asp catalytic triad and is the first of this family from mammalian source. Defects in the functioning of TPP-I leads to neurodegenerative disorder and diminishes the survival of affected patients. The structure of TPP-I reported here might serve as a structural basis to understand the mechanism of its action and the structural differences in the defective enzymes.

1. Theoretical Background

1.1 Introduction

The living cells contain numerous macromolecules such as proteins, nucleic acids and carbohydrates. Among the nucleic acids DNAs contain the cellular blueprints, proteins are the working molecules executing a wide range of activities. The structures of these macromolecules are important to understand the various cellular processes as well as evolutionary relationships.

The three dimensional structures of macromolecules can be determined using X-ray crystallography. A prerequisite for structure determination is the availability of well diffracting crystals. X-rays have a wavelength of 0.1-100 Å. Resolution of the inter-atomic bond distances in molecules (covalent bond distances of 1-2 Å, polar interactions and hydrogen bond distances of 2.5-3.5 Å) is possible using X-ray diffraction of the crystals of the target molecules, as X-rays have a wavelength comparable to the inter-atomic bond distances.

Scattering is the reemission of light by a particle (electron) upon interaction with an incident photon. Diffraction occurs as a result of constructive interference of several scattered X-ray photons. Molecules have surrounding electron clouds and these electron clouds interact with X-rays to emit radiation of the same frequency (elastic or Thomson scattering) or a different frequency (inelastic or Compton scattering).

The atomic scattering factor f_a of a single atom can be described by the equation:

$$f_a = \int_0^\infty \lambda r \rho_a(r) \frac{\sin\left(2\pi r \frac{2\sin\theta}{\lambda}\right)}{\sin\theta} dr, \quad 1.1$$

where λ is the X-ray wavelength, ρ_a is the electron density distribution around the atomic core, r is distance from the nucleus and θ is scattering angle.

Taking into account the correction for atomic and molecular vibrations, the corrected atomic scattering factor f_a' is written as:

$$f_a' = f_a e^{\left(-B\frac{\sin^2\theta}{\lambda^2}\right)}, \qquad 1.2$$

where $B = 8\pi^2 u^2$ is the atomic displacement parameter (isotropic or anisotropic) and u is the square mean displacement of the atom.

Constructive interference of the scattered X-ray photons gives rise to reflections in certain directions. In accordance to the crystal planes in which the diffraction originates, the diffracted rays are assigned Miller indices h, k and l. The structure factors of the diffracted rays are:

$$F_{hkl} = \sum_{i=1}^{n} f_a' e^{\left[2\pi i\left(hx_i + ky_i + lz_i\right)\right]}, \qquad 1.3$$

where the summation is over all the atoms in the unit cell and

$$F_{hkl} = | F_{hkl} | e^{(i\alpha_{hkl})}, \qquad 1.4$$

where $| F_{hkl} |$ is the structure factor amplitude of reflection, α_{hkl} is its phase.

The electron density ρ_{xyz} in a crystal can thus be calculated from the following Fourier summation:

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} |F_{hkl}| e^{\left[-2\pi i(hx+ky+lz)+i\alpha_{hkl}\right]}, \quad 1.5$$

where x, y and z are the coordinates and V is the volume of the unit cell.

In an X-ray diffraction experiment only reflection amplitudes $|F_{hkl}|$ can be obtained from measured intensities I_{hkl} using the following relation:

$$I_{hkl} = |F_{hkl}|^2 \qquad\qquad 1.6$$

However, no information is obtained for the phase angles α_{hkl} and therefore the electron density distribution in the unit cell cannot be calculated. This is known as the *phase problem* in protein X-ray crystallography. There are four principle techniques to solve the phase problem:

- Isomorphous Replacement
- Anomalous Dispersion
- Molecular Replacement
- Direct Methods

Typically these methods lead to an initial approximation of the electron density distribution in the crystal, which can be improved iteratively, eventually converging at a reasonable structural protein model.

1.2   Isomorphous Replacement

The Isomorphous Replacement method is based on the heavy atoms present in one (Single Isomorphous Replacement) or more heavy atom derivative crystals (Multiple Isomorphous Replacement). The heavy atoms serve as strong scattering centers by binding tightly to one or few specific sites without disturbing the unit cell and symmetry of the native crystal (hence the term Isomorphous Replacement). However isomorphism in practice is not perfect as the heavy atoms distort or change the surrounding protein structure in the crystal.

The structure factors, F_{PH}, of derivative crystals can be written as:

$$F_{PH} = F_H + F_P \qquad\qquad\qquad 1.7$$

where F_P is the structure factor for native protein and F_H is the structure factor for the heavy atom. The intensities of native and one or more heavy atom derivative crystals are collected. The differences in the native and derivative crystal/s are employed to determine the heavy atom positions. From the differences, the heavy atom positions can be derived through examination of the Patterson maps.

Figure 1.1 shows the case of multiple isomorphous replacement method, where the phases can be derived unambiguously compared to single isomorphous replacement method, where there exists a two fold phase ambiguity.
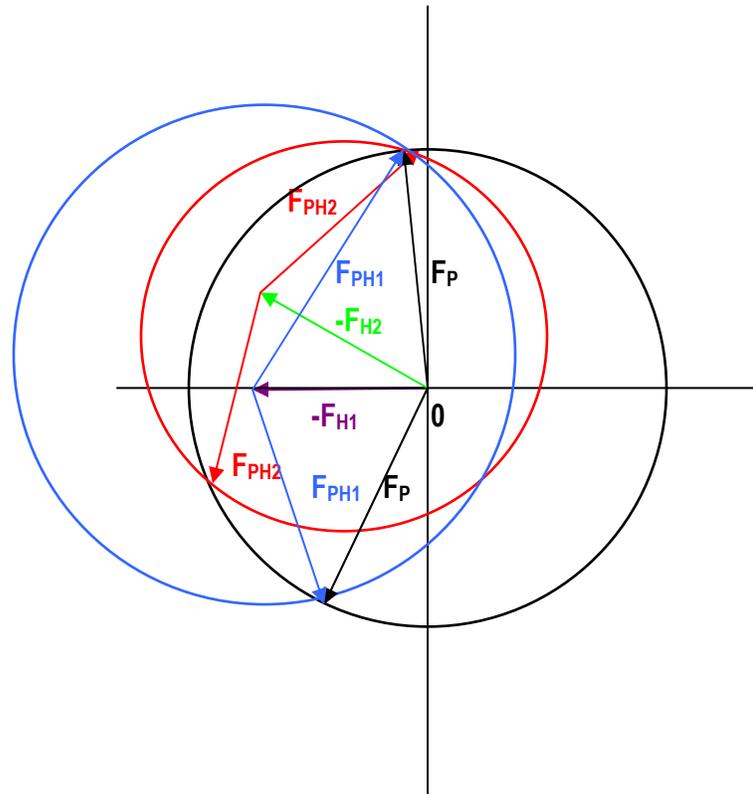
Figure 1.1: Harker construction of the Multiple Isomorphous Replacement (MIR) method for protein phase determination. The **black** circle with radius F_P indicates the native crystal with unknown phases. The **blue** circle with radius F_{PH1} represents the first derivative with unknown phases and the **red** circle with radius F_{PH2} represents the second derivative. From the Patterson map, Vectors F_{H1} and F_{H2} are calculated from the heavy atom position for first and second derivatives respectively. Two possible values for the phase of the native crystal are defined by the intersection of the **black** and **blue** circles (as in the Single Isomorphous Replacement). Using the second derivative defined by the **green** circle (Multiple Isomorphous Replacement), all the three circles intersect at a single point, thus defining the only possible value of the native crystal phase.

1.3 Anomalous Scattering (Dispersion)

Anomalous scattering is due to the fact that electrons in a crystal cannot be regarded as completely free electrons. Anomalous scattering takes place when the X-ray wavelength is near the absorption edge of the atom. This effect is wavelength dependent and is stronger for heavier atoms in the periodic table.

Friedel's law states that the intensities of diffraction and structure factor amplitudes of orders *hkl* and *-h-k-l* are identical:

$$| F_{hkl} |=| F_{-h-k-l} |$$ 1.8

This is applicable to normal diffraction. As a consequence of the anomalous scattering by heavy atoms present in the protein structure, the amplitudes of the reflection *hkl* and its Bijvoet mate *-h-k-l* are no longer equal. Thus anomalous scattering results in the breakdown of the Friedel's law:

$$| F_{hkl} |\neq| F_{-h-k-l} |$$ 1.9

The scattering factor of the atom giving rise to anomalous scattering can be expressed as:

$$f_{total} = f_0 + \Delta f^{'} + i f^{''} = f^{'} + i f^{''},$$ 1.10

where f_0 is the normal scattering factor far from the absorption edge, and $\Delta f^{'}$ and

$f^{''}$ are the correction terms which arise due to the dispersion effects. The imaginary part is responsible for the phase shift. A graphical representation of this case is illustrated in Figure 1.2.
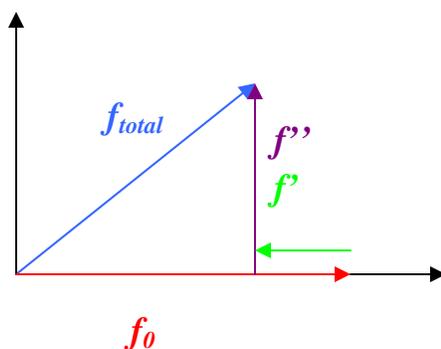
Figure 1.2: Vector representation of the structure factor of an anomalously scattering atom.

While f' has appreciable values only near the absorption edge, f'' values are appreciable only at energies equal or higher than the absorption edge (example for Se K-absorption edge in Figure 1.3). The absorption edges achievable by synchrotrons used in macromolecular crystallography is possessed only by elements in the fourth or lower periods in the periodic table.
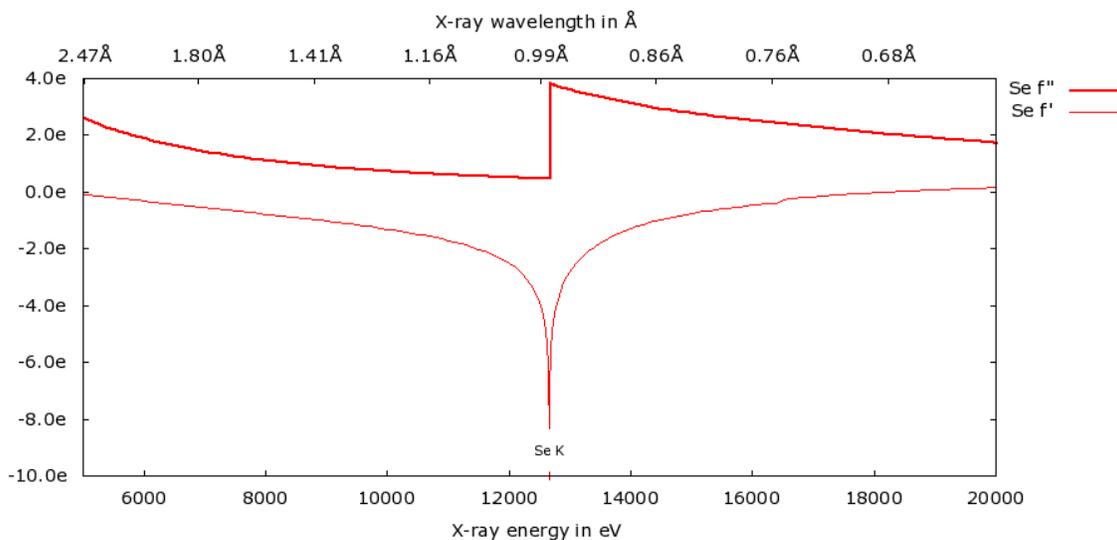
Figure 1.3: Theoretical Selenium K-absorption edge and values of f' and f'' as a function of X-ray energy (http://www.bmsc.washington.edu/scatterer).

The structure factor amplitudes of an anomalous scatterer can be summarized as (Hendrickson *et al.*, 1985):

$$|F^{\pm}|^2 = |F_T|^2 + \frac{f''^2 + f'^2}{f_0^2}|F_A|^2 + 2\frac{f'}{f_0}|F_T||F_A|\cos\alpha \pm 2\frac{f''}{f_0}|F_T||F_A|\sin\alpha, \qquad 1.11$$

$$\alpha = \phi_T - \phi_A$$

$|F_A|$ and $|F_T|$ are the derivative structure factor amplitudes without anomalous contributions and α is the phase angle between $|F_A|$ and $|F_T|$. On the basis of this equation (1.11), SAD (Single Wavelength Anomalous Dispersion) and MAD (Multiple Wavelength Anomalous Dispersion) phasing can be done. Using four or six observables from a two or four wavelength MAD experiment respectively, the quantities $|F_A|$, $|F_T|$ and α can be obtained from this equation. All data can be measured from a single crystal; consequently there is no isomorphism problem. In a SAD experiment data are collected using one crystal and single wavelength. Through some approximations to equation 1.11 SAD phasing can be done. The resulting phases are generally of poor quality and need to be improved by density modification to render an interpretable electron density map.

1.4   Molecular Replacement

In this method a homologous search model can be used to provide the initial estimate of the unknown structure phases. Two three dimensional searches are performed: a rotation search followed by a translation search. Recently, using brute force methods on faster computers, a six dimensional search is performed to find the best fit of the search model to the target model. The initial map obtained from molecular replacement can be improved iteratively. The search model should have a sequence homology of at least 40% to that of the target model. A main drawback in this method of phasing is model bias towards the search model.

1.5   Direct Methods

Direct methods assume that phase information is included in the intensities and do not require prior knowledge of any atomic positions. These methods assume that electron density is never negative and that atoms are randomly distributed in the unit cell. A prerequisite to solve a structure by direct methods is that the diffraction data should be complete to a resolution of 1.2Å. At atomic resolution there is high data to parameter ratio and also the atomic peaks are well resolved in the electron density map. Direct methods employ structure factors normalized over resolution shells. One of the most important relationships in this method is the triple phase relation:

$$\phi_H = \phi_{H'} + \phi_{H-H'} \quad , \quad\quad\quad 1.12$$

where *H* represents *hkl* and *H'* represents *h'k'l'*. Combination of all relevant triplets gives the tangent formula (Karle & Hauptman, 1956):

$$\tan(\phi_H) = \frac{\sum_{H'} | E_{H'} E_{H-H'} | \sin(\phi_{H'} + \phi_{H-H'})}{\sum_{H'} | E_{H'} E_{H-H'} | \cos(\phi_{H'} + \phi_{H-H'})}, \quad\quad 1.13$$

which is used to iteratively refine the phases. Using the tangent formula or its improved versions, conventional direct methods can solve structures of up to 200 unique atoms. However for structures larger than 200 unique atoms, the success rate of structure solution using conventional direct methods is low as the probability distribution of phase relations becomes weaker with increasing number of atoms (Cochran *et al.*, 1955).

For the solution of macromolecular structures dual space refining is used (Miller *et al.*, 1993), which starts each trial with a set of random atoms. Phases are then calculated from the starting atoms and subsequently refined and an electron density map is generated. This process is repeated for several cycles (Figure 1.4). Only the strongest reflections in each resolution shell are used in the calculation and

correlation coefficients between normalized structure factors E_O and E_C are used to identify the best solution over many tries.
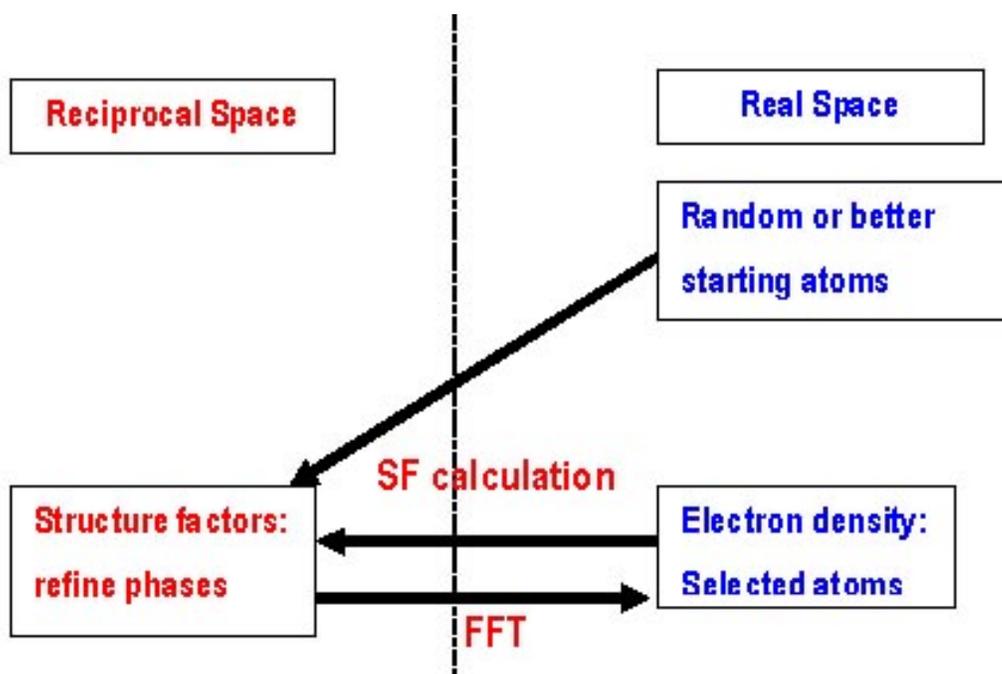
Figure 1.4: General scheme of dual space recycling.

1.6 Refinement

In most cases the initial model obtained either from experimental phasing or from molecular replacement are in poor agreement with the observed data and might contain coordinate errors or missing parts. The refinement process aims to increase the agreement between structure factor amplitudes obtained from experimental data $|F_{obs}|$ and those calculated from the model $|F_{calc}|$. This agreement is expressed as the R-factor:

$$R = \frac{\sum_{hkl} ||F_{obs}| - k |F_{calc}||}{\sum_{hkl} |F_{obs}|} \qquad 1.14$$

In macromolecular crystallography, to cross validate the refinement process, an additional index R_{free} is used (Brünger, 1992). In this process a random set of (~5%) reflections called the free set of reflections are excluded from refinement and the model is adjusted to the rest of the reflections (working set). The difference between R and R_{free} serves as a measure of overfitting or model bias.

The refinement process involves automated optimization (using least- squares or maximum likelihood algorithms) alternating with manual corrections, thereby improving the electron density maps and convergence of the observed and calculated structure factors.

Often for macromolecular structure refinement the data to parameter ratio is low, especially in low resolution data. Reducing the number of parameters or increasing the observations may therefore help in the refinement process. A few ways of achieving this is by taking chemical and stereochemical information into account, flattening the bulk solvent density or taking into account the similarity of molecules in a crystal related by non-crystallographic symmetry (NCS).

2. Viscotoxin A1 from *European Mistletoe*

2.1 Introduction

2.1.1 Overview of Viscotoxins

Viscotoxins are a group of low molecular weight basic proteins found predominantly in the European mistletoe subspecies of *Viscum Album Loranthaceae* (Romagnoli *et al.,* 2003).These species have been scientifically studied for their extracts for a long time, dating back to 1881 when a digitalis-like effect was attributed to it by Park (Park *et al.*, 1881).Mistletoe extracts have long been used in treatment of tumours (Friess *et al.*, 1996; Heiny *et al.*,1998; Steuer-Vogt *et al.*, 2001; Zarkovic *et al.*,2001). The first reports on the purification of Viscotoxins was published by Winterfeld in 1942 and subsequently also the term "Viscotoxin" was coined (Winterfeld & Bjil, 1948). All the Viscotoxins are basic cystein rich proteins exhibiting a net positive charge and containing three conserved disulfide bridges, to stabilize their three dimensional structures (Romagnoli *et al.,* 2000). Viscotoxins also display a high degree of similarity to the plant thionins and hence classified in the family of thionins. Thionins constitute an important family of plant toxins that are heat stable proteins implicated in antifungal and antibacterial activities (Bohlmann &Apel, 1991). Thionins are classified into five different types based on the amino acid sequence homology (García-Olmedo *et al.*, 1992). Type I and II thionins have four disulfide bridges and have a net positive charge. Type III thionins have also a net positive charge but have three disulfide bridges. Type IV thionins are neutral thionins with four disulfide bridges. Type V thionins are different from the others with respect to the primary

21

structure and arrangement of disulfide bridges. Viscotoxins belong to the type III thionins. To date, seven different isoforms of Viscotoxin are known, namely Viscotoxin A1 (Schaller *et al.*, 1996; Orrù et al., 1997), Viscotoxin A2 (Olson & Samuelsson, 1972), Viscotoxin A3 (Samuelsson *et al*., 1968), Viscotoxin B (Samuelsson & Petterson, 1971), Viscotoxin 1-PS (Samuelsson and Jayawardene, 1974), Viscotoxin U-PS (Ribéreau-Gayon et al., 1986) and Viscotoxin C1 (Romagnoli *et al.*, 2003). As shown in Fig. 1.2, the Viscotoxins not only show a high degree of sequence homology (~75%) among the various isoforms, they also exhibit a high degree of sequence homology to some other members of the thionin family. They are stabilized by three disulfide bridges between residues Cys^3/Cys^{40}, Cys^4/Cys^{32} and Cys^{16}/Cys^{26}. Three dimensional structures of the Viscotoxins are very similar to that determined for several other members of the thionin family, represented by a global Γ fold (resembling the Greek capital letter gamma (Γ) or English capital L), consisting of an arm and a stem with a substantial groove in between, which displays a high positive electrostatic potential in the biologically potent members of the family as shown in Figure 2.1 (Hendrickson & Teeter, 1981). The stem is formed by two antiparallel α-helices, while the arm is defined by two short antiparallel β- sheets and the C-terminal coil. Also the aromatic residue at position 13 is quite conserved. Thionins lacking a tyrosine residue at position 13 display weak ability to inhibit fungal growth (Terras *et al.*, 1996). However the Viscotoxins are proposed to be more potent than the other six-Cysteine type antimicrobial peptides (Giudici *et al.*, 2004). A structural overview of the various thionins is illustrated in Figure 2.2.

The thionins and Viscotoxins are found to be cytotoxic to various cell types and their mechanisms are discussed in details in the later sections.

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

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 2.1: Sequence alignment and consensus sequences in the thionin family. The conserved residues, corresponding to the Viscotoxin consensus sequence are marked in red, the similar residues are marked in black and variable residues are marked as '**:**'.
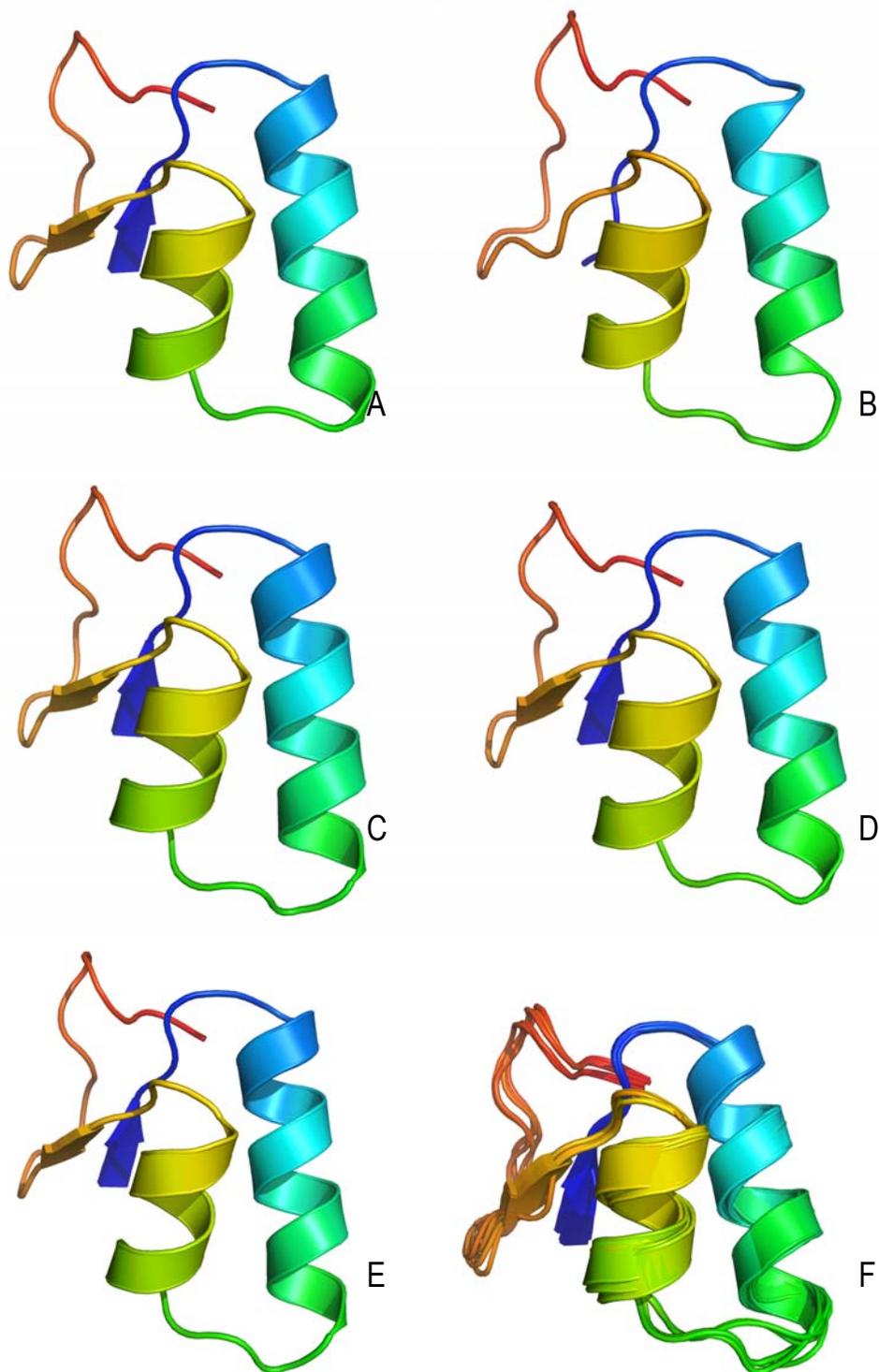
Figure 2.2: A structural overview of the related thionins (A) β- purothionin (type II thionin; PDB id: 1BHP) (B) Crambin (Type IV thionin; PDB id: 1EJG) (C) Viscotoxin A3 (type III thionin; PDB id: 1OKH) (D) β- hordothionin (type II thionin; PDB id: 1WUW) (E) Viscotoxin B2 (type III thionin; PDB id: 2VB9) (F) Superposition of all the above.

2.1.2 Biological significance and mechanism of action

Viscotoxins are considered to play a role mainly in plant defense against microbial agents (García-Olmedo *et al.*, 1998) as well as storage proteins (Schrader-Fischer and Apel, 1993). However the exact significance of Viscotoxins in plant cells is still unclear. They are toxic to a broad spectrum of cell types at micro molar concentrations. γ -Hordothionin from barley (*Hordeum vulgareum*) was reported to have α- Amylase inhibitory activity (Mendez *et al.*, 1990) and also causes permeabilization of fungal hyphae to α – aminoisobutyric acid and altered the electrical properties of artificial lipid bilayers (*Thevissen et al.*, 1996). Other similar thionins exhibiting anti-fungal activity are MsDef1 from *Medicago sativa* (Spelbrink *et al.*, 2004), MsDef2 from *Medicago truncatula* (Spelbrink *et al.*, 2004), SPE10 from *Pachyrrizus erosus* (Song *et al.,* 2005). Antibacterial activity was exhibited by Pseudo-thionin (Pth-St1) from *Solanum tuberosum* to name a few. γ1 - Purothionin from *Triticum turgidum* showed insecticidal activity (Colilla *et al.*, 1990). *Arabidopsis* cell lines expressing high Viscotoxin levels were more resistant to infection by *Plasmodiophora brassicae* (Holtorf *et al.*, 1998).

2.1.3 Mechanism of Antifungal Activity exerted by Viscotoxins

 Recent biochemical studies by various groups (Aerts *et al.*, 2007; Broekaert *et al.*, 1995; Coulon *et al.*, 2002, 2003; Florack & Stiekema, 1994; García-Olmedo *et al.*, 1998; Giudici *et al.*, 2003, 2004, 2006; Hughes *et al.*, 2000; Llanos *et al.*, 2004; Thevissen *et al.*, 1996, 1999.; Tabiasco *et al.,* 2002) proposed the following mechanism of action, which explains the toxicity of Viscotoxins towards fungi. Most of the experiments were performed using Viscotoxin A3 and to a lesser extent, Viscotoxin B and Viscotoxin A2 and other related proteins of Thionin family against phytopathogenic fungi of agricultural relevance (Fusarium *solani*, *Sclerotinia*

sclerotiorum and *Phytophthora infestans*, *Fusarium culmorum*) and also to human pathogenic fungi (*Candida albicans*). Viscotoxin A3 is reported to be the most potent of all the Viscotoxin isoforms and Viscotoxin B showed the least toxicity (Schaller *et al.*, 1996).

The first step towards the toxic effects exerted by Viscotoxin is interaction with membranes, causing rapid permeabilization of membranes, subsequently forming ion channels or pores. However, the channel like gating activity was transient, depending on the concentration of Viscotoxin, following which an abrupt membrane leakage occurred, followed by accumulation of Viscotoxin inside the cells. Cell membrane leakage/disruption was found to depend on lipid composition and inhibited by divalent cations. In the next step, Viscotoxin induces H_2O_2 production, which is dependent on the incubation time and the concentration of the Viscotoxin. Furthermore, Viscotoxin induces an increase in internal Ca^{2+} concentration (probably from the internal stores such as vacuoles, endoplasmic reticulum, etc.) and generation of Reactive Oxygen Species and loss of mitochondrial internal membrane potential (by destabilization of the mitochondrial membrane). Since the two antiparallel α- helices of a single Viscotoxin molecule are shorter than the bilayer thickness, it is assumed that Viscotoxin molecules subsequently assemble to form a transmembrane complex forming ion channels or pores in membranes.

Thus, Viscotoxins exhibit two important properties which might explain their cytotoxicity. Firstly they have the ability to form ion channels in cell membranes, as exhibited by β purothionin (Hughes *et al.*, 2000). Secondly they can bind to the membrane surface and disturb the membrane organization, as observed in α-hordothionin and wheat α-thionin (Thevissen *et al.*, 2003; Caaveiro *et al.*, 1998). Thus, at low concentrations Viscotoxins may bind to membranes and form channels and at high concentrations might lyse the membrane just as a detergent.

2.1.4 Effects on Mammalian cells

It is reported that Viscotoxins are cytotoxic towards human myeloid cell line K562, human T- cell leukemia cell line Molt4, Yoshida sarcoma cells and human granulocytes (Jung *et al.*, 1990; Urech *et al.*, 1995). Also their toxic effect against KB and HeLa cell lines has been determined and depending on the dose and isoforms, shown to reduce the viability of these cells (Konopa *et al.*, 1980). In addition, Viscotoxins increase natural killer cell mediated cytotoxicity towards tumour cells but spare non-target cells (Tabiasco *et al.*, 2002).Viscotoxins exhibit toxicity towards human lymphocytes, and resulted in altered membrane permeability, generation of reactive oxygen species and cell death (Büssing *et al.*, 1999), which is quite similar to the cytotoxic mechanisms against fungi as discussed above.

2.1.5 Biosynthesis and Occurrence of Viscotoxins

Similar to other plant thionins, the Viscotoxins are synthesized as a prepropeptide having a molecular weight of 15KDa, containing an N-terminal signal peptide of 26 residues and a C-terminal acidic polypeptide chain, along with the 46 residue long central thionin sequence (Schrader *et al.*, 1991) as shown in Figure 2.3. Subsequently during post translational modifications, the N- & C- terminal sequences are cleaved off. The C– terminal part of the prepropeptide, owing to the acidic nature may help to neutralize the toxic effects of Viscotoxins within plant cells (Schrader & Apel, 1991).

Viscotoxins are found in the leaves stem and berries of mistletoe plants. However the composition of the various isoforms obtained from extracts varies depending on the host tree for that particular plant and also on the seasons (Schaller *et al*., 2000).

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


                            10 -          20         30          40
                             |            |          |           |
Mature Viscotoxin A1       KSCCPSTTGR  NIYNTCRLTG  SSRETCAKLS  GCKIISASTC  PSNYPK


                      50 -          60         70          80
                       |            |          |           |
Acidic polypeptide   FCYT   LGCESSQCAT   NSNGDAEAVR   CKTACSDLCQ   DVDDA
```
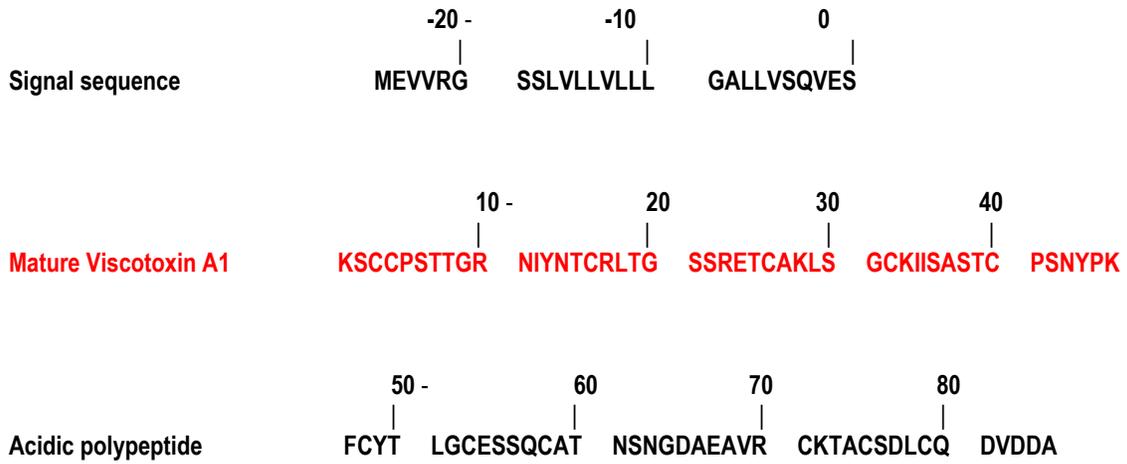
Figure 2.3: The primary structure of the VT A3 prepropeptide. The N-terminal signal sequence and the C-terminal acidic polypeptide are cleaved during the maturation of the protein.

2.2   Materials and Methods

2.2.1 Purification

The isolation and purification of Viscotoxins was performed by Beatrix Girmann in the group of Prof. Dr. Axel Zeeck, Institute für Organische Chemie, Göttingen. Mistletoe bushes were collected during April from poplar trees as they are rich in the Viscotoxin content normally and even higher in summer (Schaller *et al*., 2000).

The harvested plant was minced and then dried at 60°C for five days to prevent enzymatic degradation of the target proteins. The dried material was finally ground further to a fine powder. 100g of this powder was solubilized by suspending in 2.5L, 2% acetic acid (pH 4.0) at 4°C for 24 h and then sonicated for 30 min. using an ultrasound bath. Subsequently the suspension was filtered, the filtrate was concentrated and then passed through a cation exchange column (SP Sepadex C-25, pre-equilibrated with 0.02 M ammonium acetate). The column material was rinsed with 700 ml of 0.02 M ammonium acetate until the column material decolorized from brown to light fawn. Finally the material was eluted with 1 M ammonium acetate. The eluate was then lyophilized. To remove salts, the supernatant was purified using MPLC (LiChroprep RP-8). After charging the sample to the column, it was washed, first with 450ml of double distilled H_2O, then with 450 ml of 20 % acetonitrile and 0.1% Trifruoroacetic acid. The Viscotoxin raw product was eluted with 450 ml of 40% acetonitrile and 0.1 % Trifruoroacetic acid. Viscotoxin isoforms were separated by HPLC (Nucleosil 100 C 8 and Phenomenex C 18 5 µ, elution with water-acetonitrile gradient along with 0.1% Trifruoroacetic acid. 100 g of dry mistletoe powder finally yielded 5mg Viscotoxin A1, 8mg Viscotoxin A2, 15mg Viscotoxin A3, 14mg Viscotoxin B, 7mg Viscotoxin 1-PS and 0.2mg of the newly

characterized isoform: Viscotoxin B2. All the Viscotoxin samples were stored at 4°C as lyophilized powder (detailed description in Girmann, 2003, Schaller *et al.*, 1996, Orrù et al., 1997).

2.2.2 Crystallization

For crystallization, Viscotoxin A1 was dissolved in double distilled water to a concentration of 15 mg/ml and centrifuged at 14000 RPM for 2 minutes to remove insoluble particles. Viscotoxin A1 was crystallized by hanging drop vapor diffusion method (Hampel *et al.*, 1968; McPherson, A.., 1992) at room temperature by equilibrating a drop containing 2 µl of protein solution and 2 µl of reservoir solution. Crystallization conditions were screened using Hampton Crystal Screen Kits (Jancarick & Kim, 1991; Cudney *et al.*, 1994). Crystals having needle like morphology were obtained which were further optimized (Figure 2.4). Crystals with good diffraction quality were obtained from reservoir solution containing 0.1 M HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5 and 1.4 - 1.6 M tri-Sodium citrate dihydrate. All solutions used for crystallization were sterile filtered and contained 0.03% sodium azide.

Figure 2.4: Needle like crystals of Viscotoxin A1.

2.2.3 Data Collection

Before measurement, the crystals were first soaked in a cryoprotectant solution containing 10 % glycerol and the rest as reservoir solution, then picked with a suitable size nylon loop and finally flash frozen under liquid nitrogen stream. Data were collected to a resolution of 1.7 Å on a Bruker rotating anode generator with Osmic focusing mirrors, a three circle diffractometer and a Smart 6000 4K CCD detector. To ensure high data precision and redundancy, four 360 ° φ scans were collected at 0.18 ° oscillation width and nine 180° ω scans at 0.2 ° oscillation width were collected. Data were collected, processed, scaled and analyzed using *PROTEUM, SAINT, SADABS* and *XPREP* (Bruker Nonius, 2002).

For extension of phases, a dataset to 1.25 Å was measured at the BW7B beamline at EMBL/DESY. 45 images at an oscillation width of 1° were collected on a MAR345 image plate detector and processed with *HKL2000* (Otwinowski *et al.*, 1997).

(A)   (B)
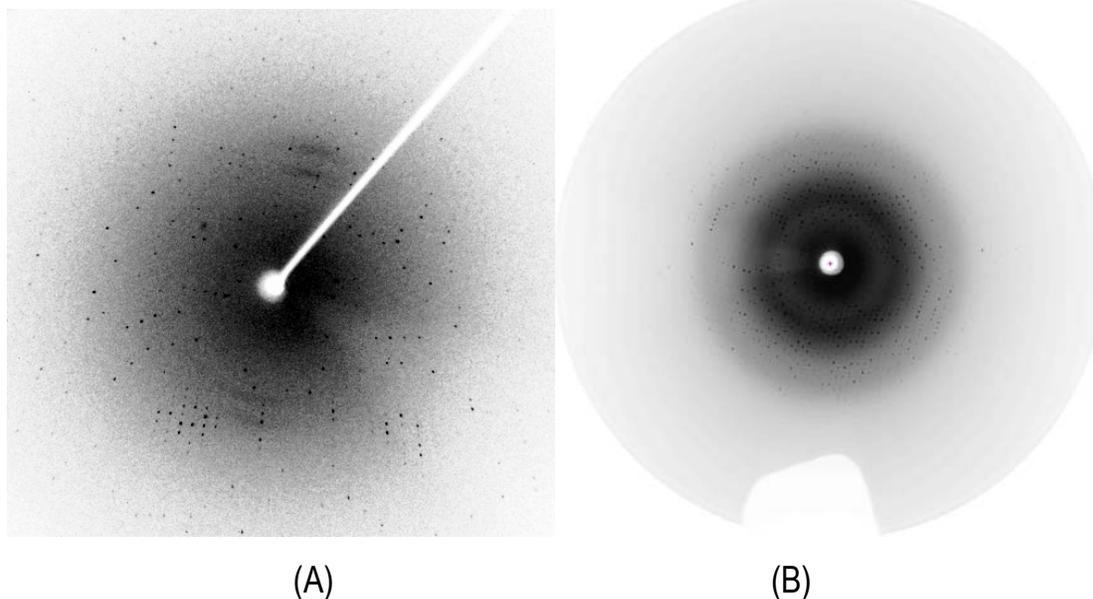
Figure 2.5: Diffraction pattern photograph of (A) In-house data collected on SMART 6000 4K CCD detector (B) Data collected on BW7B beamline at EMBL/DESY on a MAR345 image plate.

Table 2.1: Data collection statistics of Viscotoxin A1 in-house on a CCD and at BW7B beamline. Values in parenthesis are for the outer resolution shell.

Data Statistics	In-house	BW7B
Wavelength (Å)	1.5418	1.000
X-ray source	Cu Kα rotating anode	DESY-BW7B
Detector	SMART 6000 CCD	MAR345 image plate
Space group	$P4_32_12$	$P4_32_12$
a (Å)	65.56	65.73
b (Å)	65.56	65.73
c (Å)	47.08	47.16
Resolution (Å)	1.70 (1.79-1.70)	1.25 (1.35-1.25)
Unique Reflections	11828	28947
Redundancy	46.44 (4.51)	3.99 (3.96)
Completeness (%)	100 (100)	99.1 (100)
Mean I/σ (I)	43.68 (3.44)	15.82 (2.90)
R_{int} (%)	7.47 (33.36)	4.77 (41.83)

2.2.4 Structure Solution and Phasing

The data processing and scaling for the ω and φ scans for the in-house data were performed separately, to preserve their independence. To determine the anomalous signal, the correlation coefficient between the signed anomalous differences from the two datasets was calculated using *XPREP*. For substructure solution, the data should be truncated at the resolution where the correlation coefficient falls below 30% (Schneider & Sheldrick, 2002), the inclusion of data in the subsequent shells

might not be reliable and just include noise and in turn might prevent the heavy atom search. In this case the correlation coefficient test indicated the presence of anomalous signal up to 2.6 Å:

Resolution: **Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.2 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 Å**

Correlation coefficient:91.1   92.8   85.7   75.5 66.5   56.7 44.4 43.3 30.8 20.1 26.2   4.9

In the subsequent step, all data from the ω and φ scans were scaled together using *SADABS* to improve the anomalous signal quality, taking advantage of the redundancy. Anomalous differences used in the subsequent steps were calculated using the dataset merged from the ω and φ scans. For the merged dataset, the anomalous $\Delta F/\sigma$ (ΔF) values against resolution are as follows (Schneider & Sheldrick, 2002). The first line is based on resolution, the second line on input sigma values and the third line on variances of F+ and F- :

Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.5 - 2.3 - 2.1 - 1.9 - 1.7 - 1.5 Å

4.46 4.96 3.81 2.75 2.54 1.94 1.77 1.57 1.47 1.13 1.03 0.97

4.42 4.84 3.63 2.68 2.37 1.83 1.64 1.45 1.37 1.09 0.97 0.92

The structure was solved using *SHELXD* (Sheldrick *et al.*, 2001) using the following settings:

SHEL 999 2.6
PATS
MIND -2.5
FIND 6
DSUL 6

All the sulfurs corresponding to the six disulfide bridges were located using *SHELXD*, indicating a presence of two molecules in the asymmetric unit and a solvent content

of ~50%. The sulfur atom positions pertaining to the disulfide bridges in Viscotoxin A1 are shown in Figure 2.6.

CC All/ Weak   52.39   /   28.12   PATFOM 7.17

	X	Y	Z	height
S001	0.825319	0.451560	0.011393	1.0000
S002	0.848383	0.582847	-0.362615	0.9775
S003	0.797495	0.651412	-0.466825	0.9440
S004	0.838609	0.441665	0.048571	0.8537
S005	0.836458	0.597068	-0.019506	0.8285
S006	0.776781	0.375968	0.149455	0.8081
S007	0.755640	0.387122	0.121058	0.8011
S008	0.816431	0.592147	0.013499	0.7990
S009	0.937250	0.695754	-0.083486	0.7170
S010	0.931680	0.671966	-0.049657	0.6933
S011	0.839937	0.572456	-0.323041	0.6602
S012	0.776193	0.638564	-0.440084	0.5008
S013	0.807832	0.537131	0.176089	0.3823

With the first twelve atoms as input and using the high resolution dataset from DESY for phase extension, *SHELXE* (Sheldrick, 2002) gave a good discrimination between the two heavy atom enantiomorphs using the contrast and connectivity figures of merit. A value of about 70% for the pseudo-free correlation coefficient indicates an interpretable map and is a good indication as to when phase refinement has converged. The pseudo-free correlation coefficient is based on the comparison of E_O and E_C for 10% of the data left out at random in the calculation of a map that is density modified and Fourier back transformed (Sheldrick, 2002).

Mean weight and estimated mapCC as a function of resolution for the original map:
        d   inf - 3.33 - 2.61 - 2.27 - 2.06 - 1.91 - 1.79 - 1.70 - 1.62 - 1.56 - 1.51
        <wt>   0.747 0.741 0.799 0.813 0.775 0.759 0.733 0.673 0.646 0.555
        <mapCC> 0.920 0.894 0.943 0.956 0.946 0.943 0.950 0.931 0.910 0.865
        d is resolution in Å, wt is weight, mapCC is map co-relation coefficient

        Pseudo-free CC = 76.14 %

Mean weight and estimated mapCC as a function of resolution for the inverted map

d   inf - 3.33 - 2.61 - 2.27 - 2.06 - 1.91 - 1.79 - 1.70 - 1.62 - 1.56 - 1.51

<wt>   0.668 0.597 0.620 0.616 0.622 0.593 0.574 0.562 0.521 0.438

<mapCC> 0.871 0.790 0.831 0.833 0.859 0.835 0.834 0.842 0.792 0.732

d is resolution in Å, wt is weight, mapCC is map co-relation coefficient

Pseudo-free CC = 61.90 %

The maps were well interpretable and automatic model building was possible. Figure 2.7 shows examples of initial and the final electron density maps.
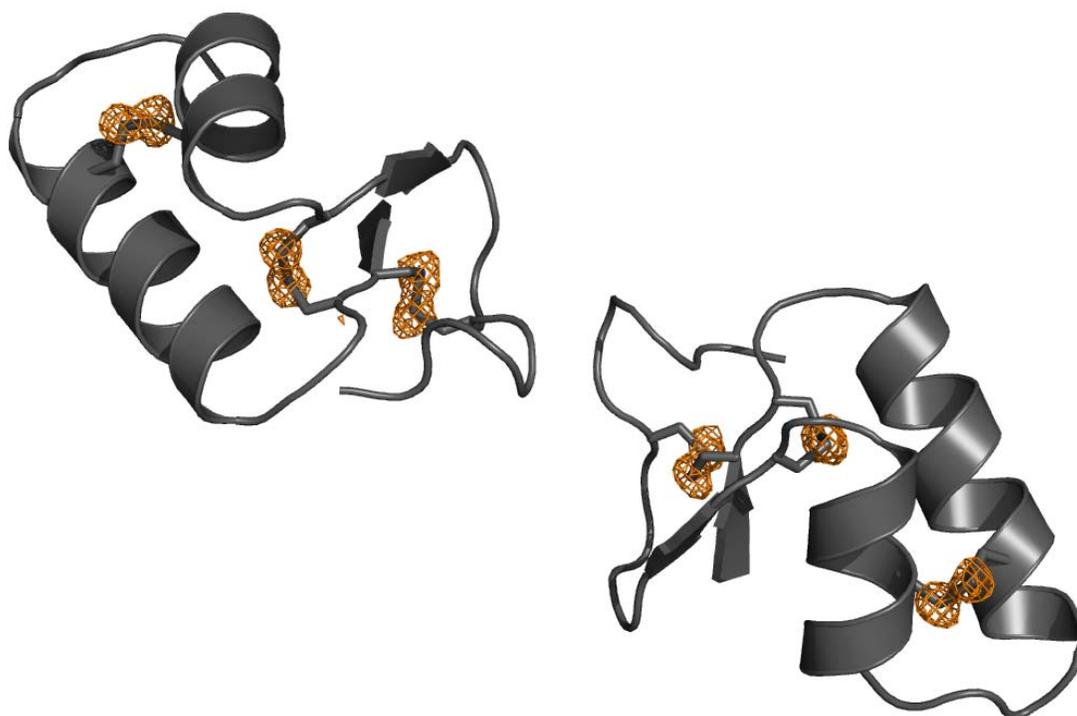
**Figure 2.6**: Anomalous Fourier map using the phases from *SHELXE* after density modification, showing the resolved sulfur atoms involved in disulfide bridge formation.
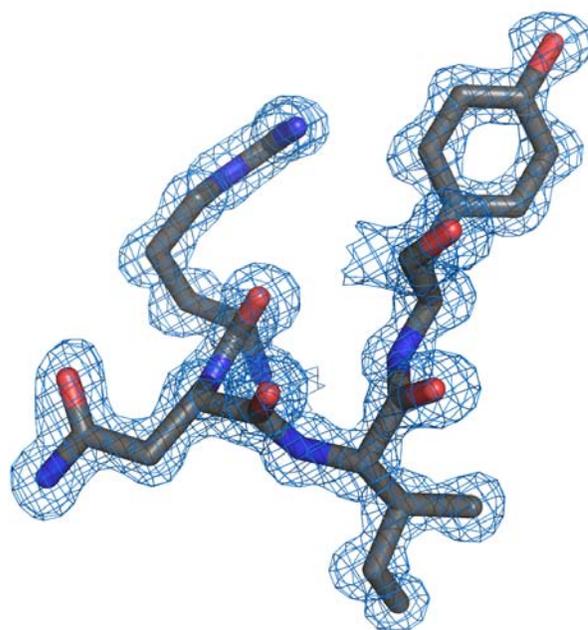
(A)
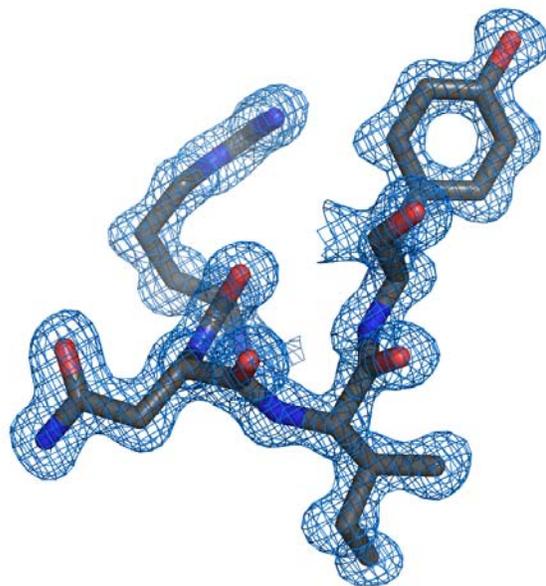
(B)

Figure 2.7: Examples of electron density maps around residues Arg[10], Asn[11], Ile[12] and Tyr[13] showing (A) map produced by *SHELXE* after phase extension to 1.25 Å (B) final refinement map using the 1.25 Å dataset. Maps are contoured at 1.5σ.

2.2.5 Model Building, Refinement and Structure Validation

The map from *SHELXE* was traced completely and all the side chains docked using *ARP/wARP* (Perrakis *et al.*, 1999) and the known sequence of Viscotoxin A1, in a standard run of 50 building cycles. Solvent molecules were added using *SHELXWAT* (Sheldrick & Schneider, 1997) and also *COOT* (Emsley & Cowtan, 2004). The structure was refined using *SHELXL* (Sheldrick & Schneider, 1997) alternating with model building using *COOT*.

The final model comprises of two independent peptide chains containing 92 residues and 206 water molecules. The final model was validated using the *COOT* Ramachandran plot validation tool (Ramachandran & Sasisekharan, 1968) and *MOLPROBITY* (Lovell *et al.*, 2003). All amino acid residues were found to be in the allowed regions of the Ramachandran plot as shown in Figure 2.8. The refinement statistics are summarized in Table 2.2.

Table 2.2: Refinement statistics of Viscotoxin A1.

Refinement Statistics	Viscotoxin A1
R-Factor (%)	12.42
Free R-Factor (%)	16.71
R.m.s. Deviations from ideal geometry	
Bond Lengths (Å)	0.016
Bond angles (°)	1.512
No. Of Protein Residues	92
No. of Water Molecules	206
Ramachandran Plot (%)	206
Residues in most favored regions	98.86
Residues in additionally allowed regions	1.14
Residues in disallowed regions	0
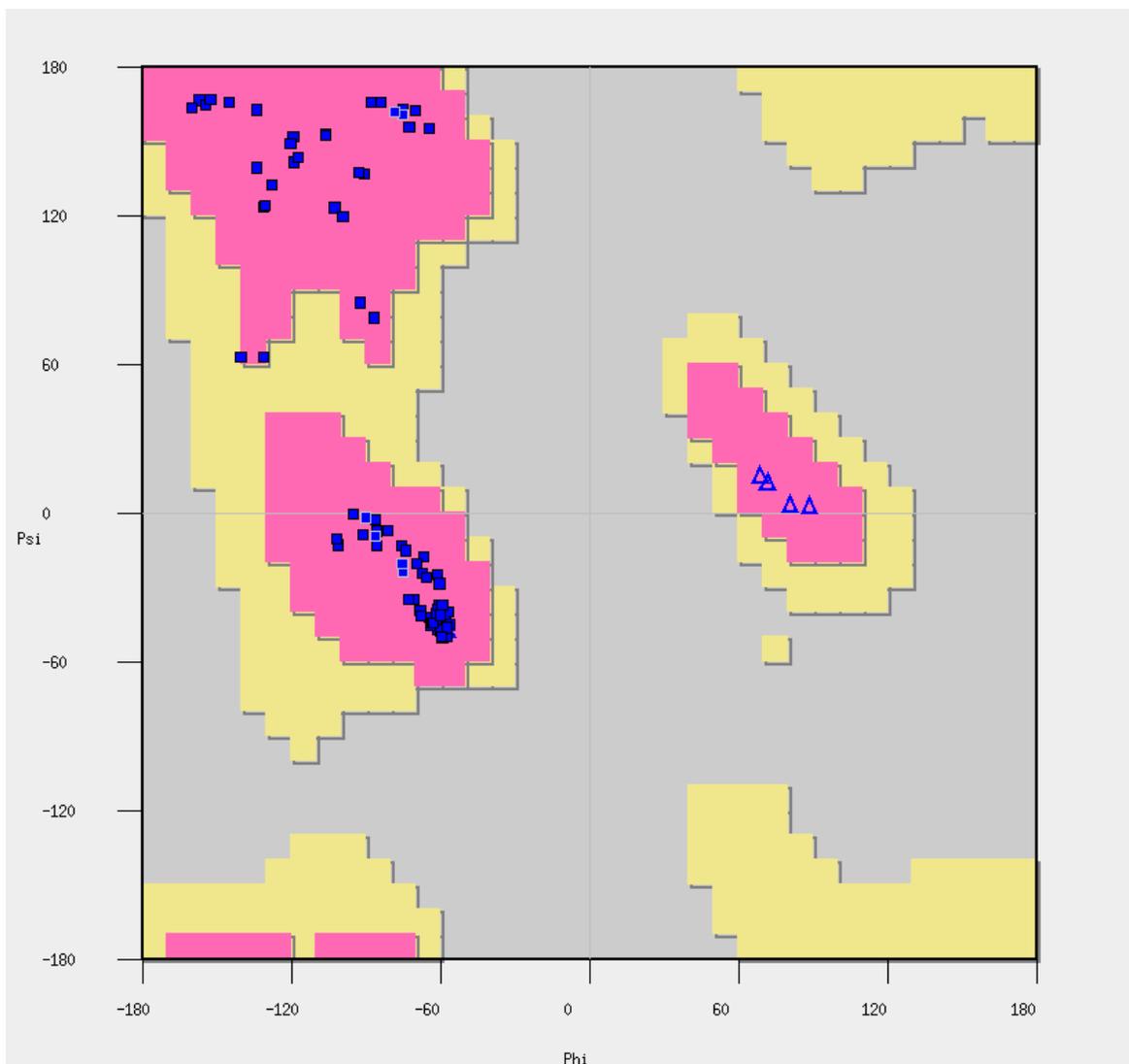
Figure 2.8: Ramachandran Plot for the structure of Viscotoxin A1.

2.3   Structure Description

2.3.1 Overall Fold

Two similar molecules, forming a dimer, related by non- crystallographic symmetry, are observed in the asymmetric unit (Figure 2.9). Least squares superposition of the two molecules give an r.m.s. deviation of 0.20 Å for Cα positions and 0.45 Å for all atoms, signifying high conformational similarity. The overall fold and secondary structure is the same as other Viscotoxins and pertains to the global Γ fold of other related thionins (Hendrickson and Teeter, 1981), signified by an arm and a stem. The arm comprises of a pair of antiparallel β- strands along with the C- terminal coil and a stem is defined by a pair of antiparallel α- helices. The structure is cross linked by three disulfide bridges (residues Cys^3 and Cys^{40}, Cys^4 and Cys^{32} and Cys^{16} and Cys^{26}), as observed in the other Viscotoxins. Two helices are observed in ViscotoxinA1, residues 6 - 18 forming the first helix (termed helix H1) and residues 22 - 30 forming the second helix (termed helix H2).These helices are stabilized by a disulfide bridge between residues 16 and 26, in addition to the interhelical H-bonds. Two short stretches of antiparallel β- strands are constituted by residues 1 - 5 (forming strand S1) and residues 32 – 37 (forming strand S2). These strands are linked by hydrogen bonds; strand S1 has a disulfide bridge between residues 3 and 40. The C- terminal part of the protein chain adopts a coil stabilized by hydrophobic interactions and a disulfide bridge between residues 3 and 40 (Figure 2.10).

As observed in the other Viscotoxins, the arm and the stem are connected by H-bond and salt – bridge network involving the conserved residues Ser^2, Cys^4, Arg^{10} and Lys^{46}. The C-terminal carboxyl group of Lys^{46} forms H-bond to N- atom of Cys^4 and a salt bridge with the guanidinium group of Arg^{10}, while the hydroxyl and

carbonyl O-atoms of Ser^2 form H-bonds to the guanidinium group of Arg^{10} (Figure 2.11).

The stabilizing forces mentioned above are conserved among the Viscotoxins and related thionins, thus making the molecule extremely rigid. However, the molecule might be rendered flexible, as none of the stabilizing quartet of residues Ser^2, Cys^4, Arg^{10} and Lys^{46} are covalently bonded and might be susceptible to external destabilizing forces.
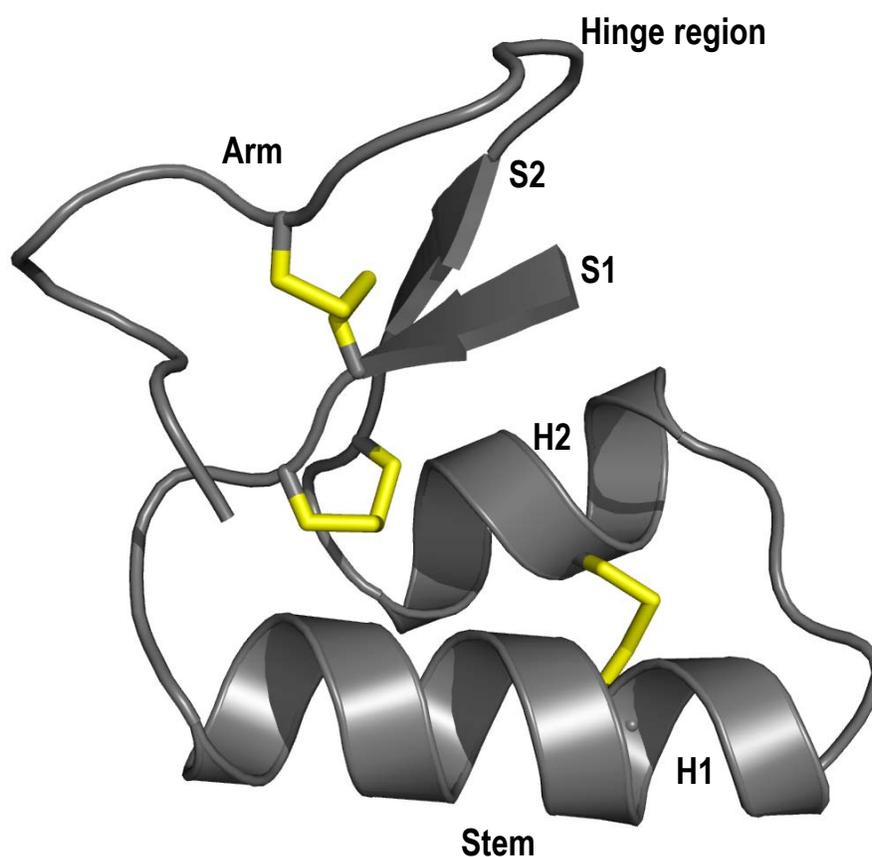
Figure 2.10: Each molecule of Viscotoxin A1 is comprised of two helices (**H1** and **H2**) and two antiparallel ß- strands (**S1** and **S2**) which are stabilized by three disulfide bridges. Viscotoxin has a global Γ fold similar to other related Viscotoxins, composed of and arm made up of strands **S1** and **S2** by an arm and a stem comprised of helices **H1** and **H2**, linked by a hinge region.
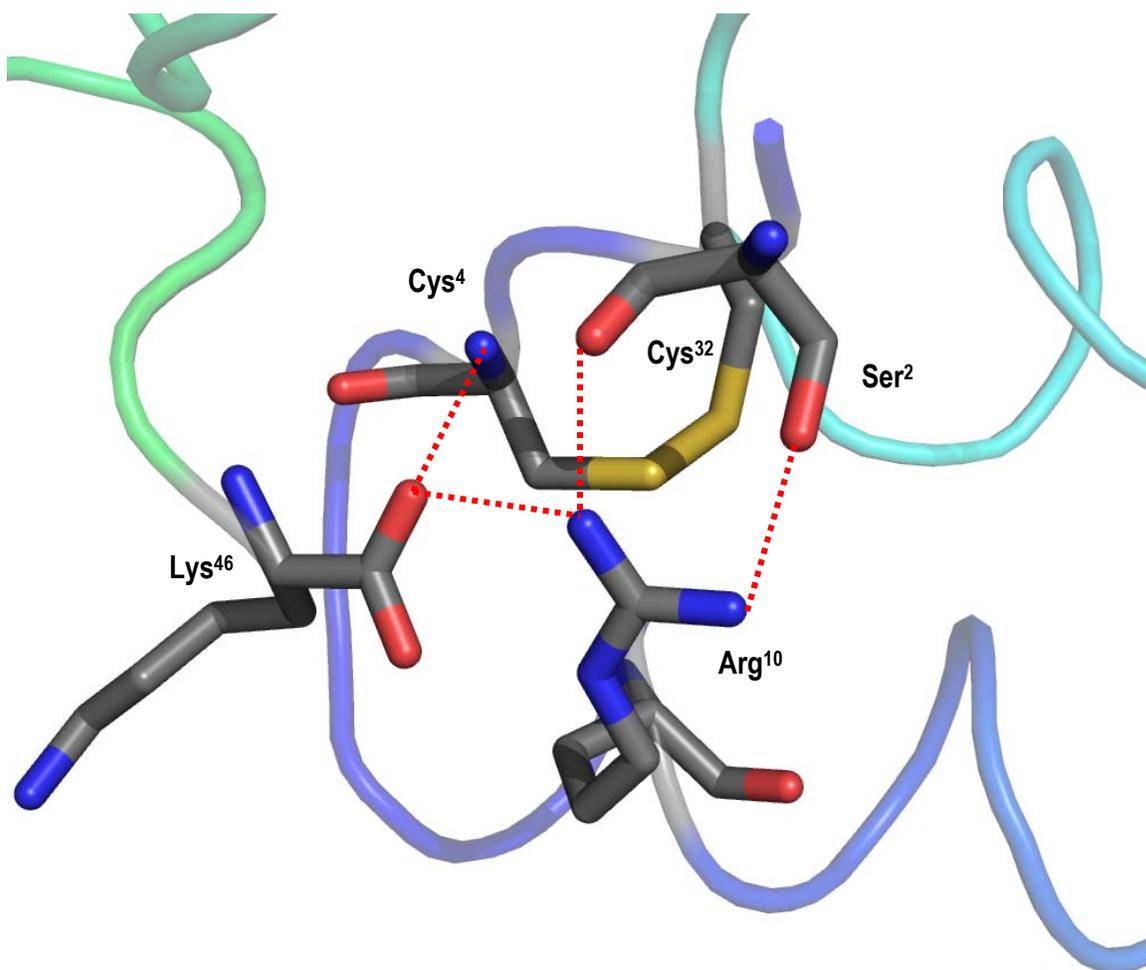
Figure 2.11: Stabilizing forces in the hinge region observed in Viscotoxin A1 and other Viscotoxins. The C-terminal carboxyl group of Lys[46] forms H-bond to N- atom of Cys[4] and a salt bridge with the guanidinium group of Arg[10], while the hydroxyl and carbonyl O - atoms of Ser[2] form H-bonds to the guanidinium group of Arg[10].

2.3.2 Comparison with other Viscotoxin and related thionin models

The r.m.s deviations of the crystal structure of Viscotoxin A1 compared to the other Viscotoxins and related thionins are listed in 2.3 and Figure 2.12.

These findings suggest the robust nature of the overall Γ fold, observed in the thionin family of proteins.

A major difference to the other Viscotoxin and related thionin structures is the absence of any sulfate ions, phosphate ions or glycerol molecules bound to the molecule (which might have survived the purification process) as reported in the previously determined structures (Debreczeni 2004; Debreczeni *et al.*, 2003; Johnson *et al.*, 2005; Stec *et al.*, 1995), in spite of presence of glycerol as a cryoprotectant. Furthermore, the crystallization condition contained citrate and HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)), none of which were observed to bind to the protein at the previously mentioned sites.

Table 2.3: Listing of the r.m.s. deviations (in Å) of the current reported Viscotoxin A1 structure with respect to Viscotoxin B2 (PDB id: 2VB9), Viscotoxin A3 (PDB id: 1OKH), Crambin (PDB id: 1EJG), ß- hordothionin (PDB id: 1WUW) and ß- purothionin (PDB id: 1BHP).

	ViscotoxinA1 molecule A	Viscotoxin A1 molecule B
Viscotoxin B2 molecule A	0.621	0.593
Viscotoxin B2 molecule B	0.494	0.489
Viscotoxin A3 molecule A	0.443	0.480
Viscotoxin A3 molecule B	0.488	0.519
Crambin	0.632	0.689
ß- hordothionin molecule A	2.782	2.796
ß- hordothionin molecule B	2.704	2.717
ß- Purothionin	2.698	2.714
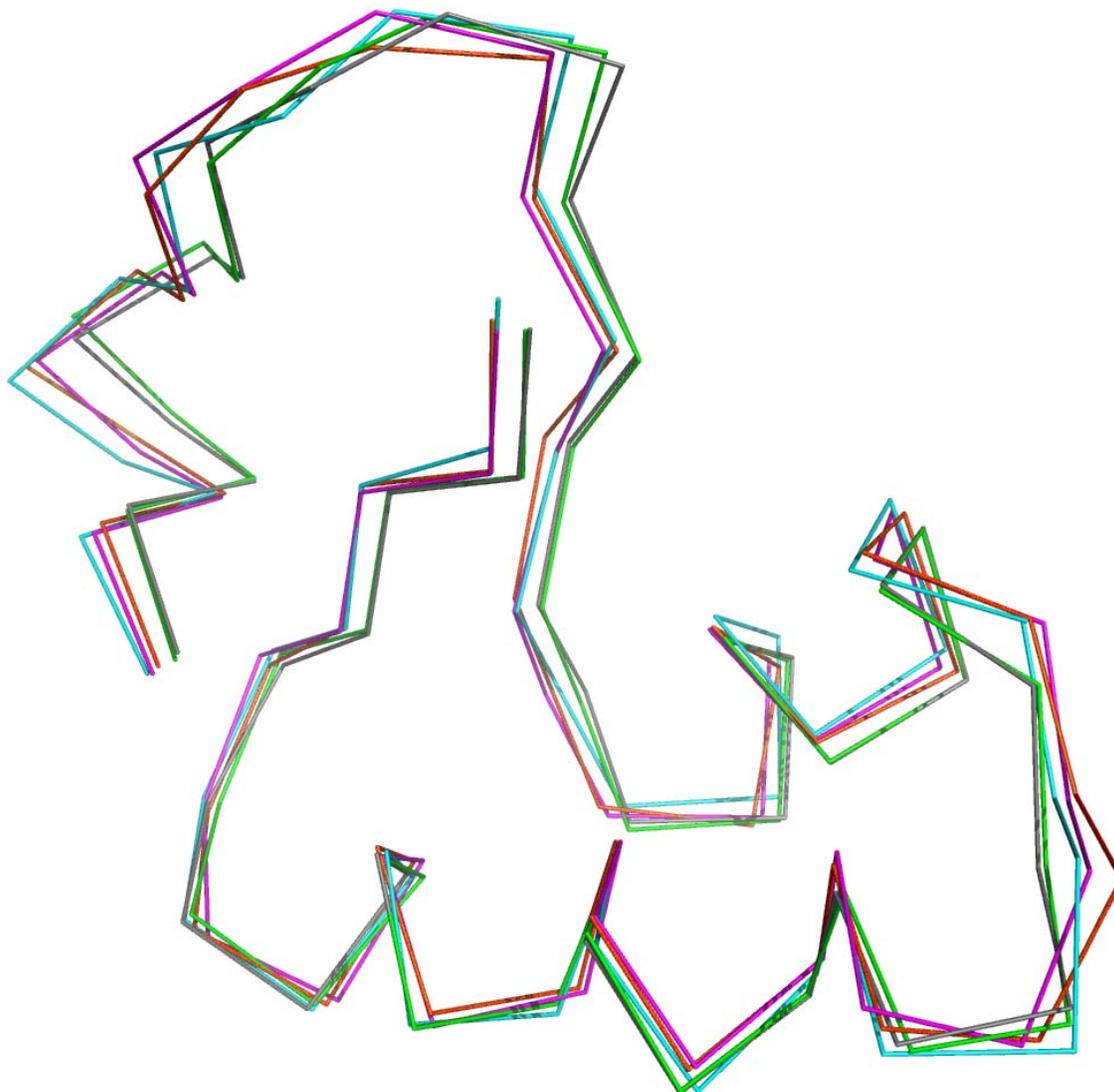
Figure 2.12: Superposition of the Viscotoxin molecule A (colored in **red**), Viscotoxin B2 molecule A (PDB id: 2V9B, colored in **orange**), Viscotoxin A3 molecule A (PDB id: 1OKH, colored in **magenta**), Crambin (PDB id: 1EJG: colored in **cyan**), ß-hordothionin molecule A (PDB id: 1WUW, colored in **grey**) and ß- Purothionin (PDB id: 1BHP, colored in **green**).

2.3.3 Dimerization and crystal packing

Packing of the crystal lattice might provide information about the native aggregation states of the molecules. Despite the high similarities among the sequences of related structures of Viscotoxins, there is considerable difference in the crystal packing. While Viscotoxin A3 and Viscotoxin B2 crystallized in space group $P2_12_12$ and $P2_12_12_1$ respectively and with a solvent content of ~30 %, Viscotoxin A1 crystallized in space group $P4_32_12$ and a solvent content of ~50 %. The low solvent content in the case of the Viscotoxins A3 and B2 results in larger number of interactions between the symmetry related molecules in contrast to the Viscotoxin A1. In the asymmetric unit two molecules associate to form a dimer. The previously reported structures of Viscotoxins and other related thionins exhibit a tetrameric association of molecules that constitute a layer, whereas the Viscotoxin A1 structure associates to form channels in the unit cell. The main interactions in Viscotoxin A3 and Viscotoxin B2 structures were through the 2 sulfate ions forming the dimer (Figure 2.14A and 2.14B) apart from the hydrophobic interactions between the hydrophobic interfaces of the helices. Here the main interactions involved in the dimer formation are between residues A/Ser^{42} - B/Asn^{43} and A/Pro^{41} - B/Ser^{42} (Figure 2.13B). Also the dimer formation in the asymmetric unit is different to that of the Viscotoxin A3 and Viscotoxin B2 structures as observed in Figures 2.14A and 2.14B. However Viscotoxin A1 interacts with the symmetry related molecules through the hydrophobic interactions between the hydrophobic interfaces of the helices and forms a similar aggregate (Figure 2.14C). The main interactions between the molecules which are important for crystal packing are listed in Table 2.4. The ~50% solvent content is filled with disordered water constituting 23 Å solvent channels (Figure 2.15 and Figure 2.16). The molecules are associated in such a way that they

form tetramers around the crystallographic four fold axis (Figure 2.16). The molecules could be imagined as molecular chains parallel to the fourfold screw axis. The residues Arg^{17}, Ser^{22}, Arg^{23}, Ser^{36}, Ser^{38} and Thr^{39}, which are supposed to bind the sulfate or phosphate ions in the other Viscotoxin structures, are pointed towards these channels (Figure 2.15). The hydrophobic regions are thus shielded from the solvent channel through hydrophobic interactions with each other. These interactions might be similar to that involved in binding of Viscotoxins to the membranes where the hydrophobic parts can interact with the membranes and the hydrophilic regions can form channels, thereby disrupting the membranes.

Table 2.4: Main interactions between residues involved in crystal packing in Viscotoxin A1 (A and B denote molecules A and B respectively, symm denotes the symmetry equivalents, W denotes the water molecules).

A/Thr^8 - A/Asn^{14} (symm)	A/Leu^{18} – B/Leu^{29} (symm)	A/Asn^{11} – W^{43} - B/Asn^{11} (symm)
A/Asn^{14} – A/Thr^8 (symm)	A/Cys^{16} – B/Thr^{19} (symm)	A/Leu18 – A/Ile^{12} (symm)
A/Ser^{21} – B/Thr^{19} (symm)	A/Leu^{29} – B/Gly^{31} (symm)	A/Cys^{40} – B/Thr^{39} (symm)
B/Thr^8 - B/Asn^{14} (symm)	B/Asn^{11} – W^{50} - A/Asn^{11} (symm)	A/Ile^{12} – A/Leu^{18} (symm)
B/Cys^{16} – A/Thr^{19} (symm)	B/Leu^{18} – B/Ile^{12} (symm)	B/Ser^{21} – A/Thr^{19} (symm)
B/Leu^{29} – A/Gly^{31} (symm)	B/Ser^{38} – W^{39} – Ser^{42} (symm)	B/Cys^{40} – B/Thr^{39} (symm)

(A)

Ser4^{42}

Pro^{41}

Asn^{43}

Ser^{42}
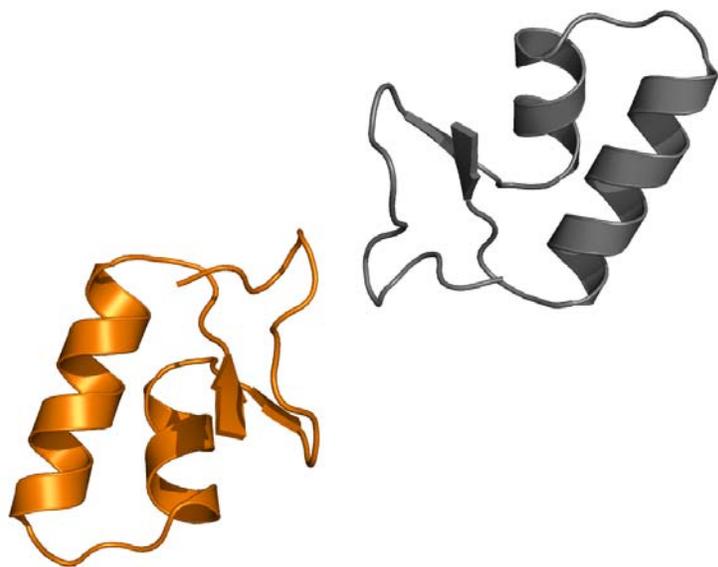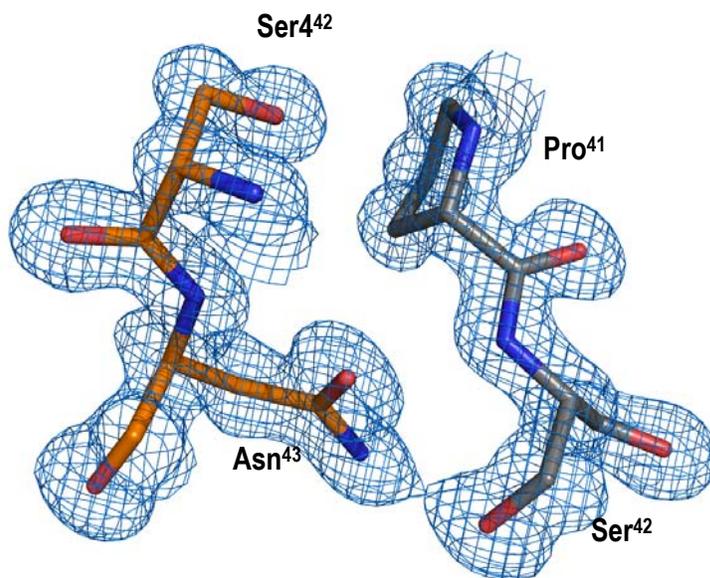
(B)

Figure 2.13: (A) Representation of the Viscotoxin A1 dimer in the asymmetric unit (B) Residues involved in formation of the Viscotoxin A1 dimer in the asymmetric unit. Molecule A is colored in **orange**, while molecule B is colored in **grey**.

(A)   (B)

(C)
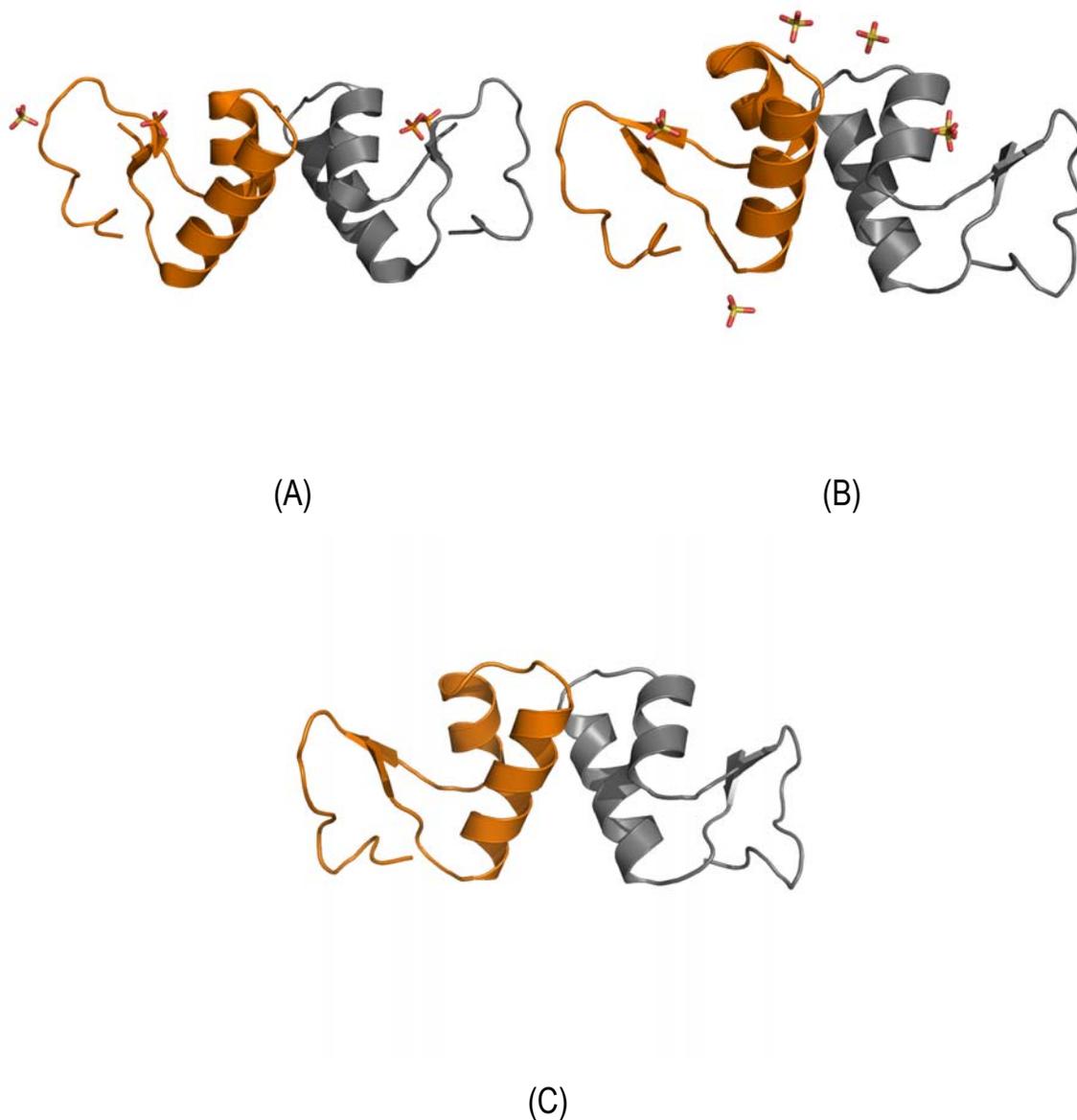
Figure 2.14: Dimer formation in the asymmetric unit in (A) Viscotoxin A3 (PDB id: 1OKH) (B) Viscotoxin B2 (PDB id: 2V9B) (C) Formation of similar aggregate in Viscotoxin A1 with symmetry molecule. The sulfates or phosphates bound to the Viscotoxins A3 and B2 molecules (absent in Viscotoxin A1) are shown in yellow and orange respectively. The interactions between the hydrophobic interfaces of the helices are common in Viscotoxin structures and might be important for the aggregation of Viscotoxin molecules.
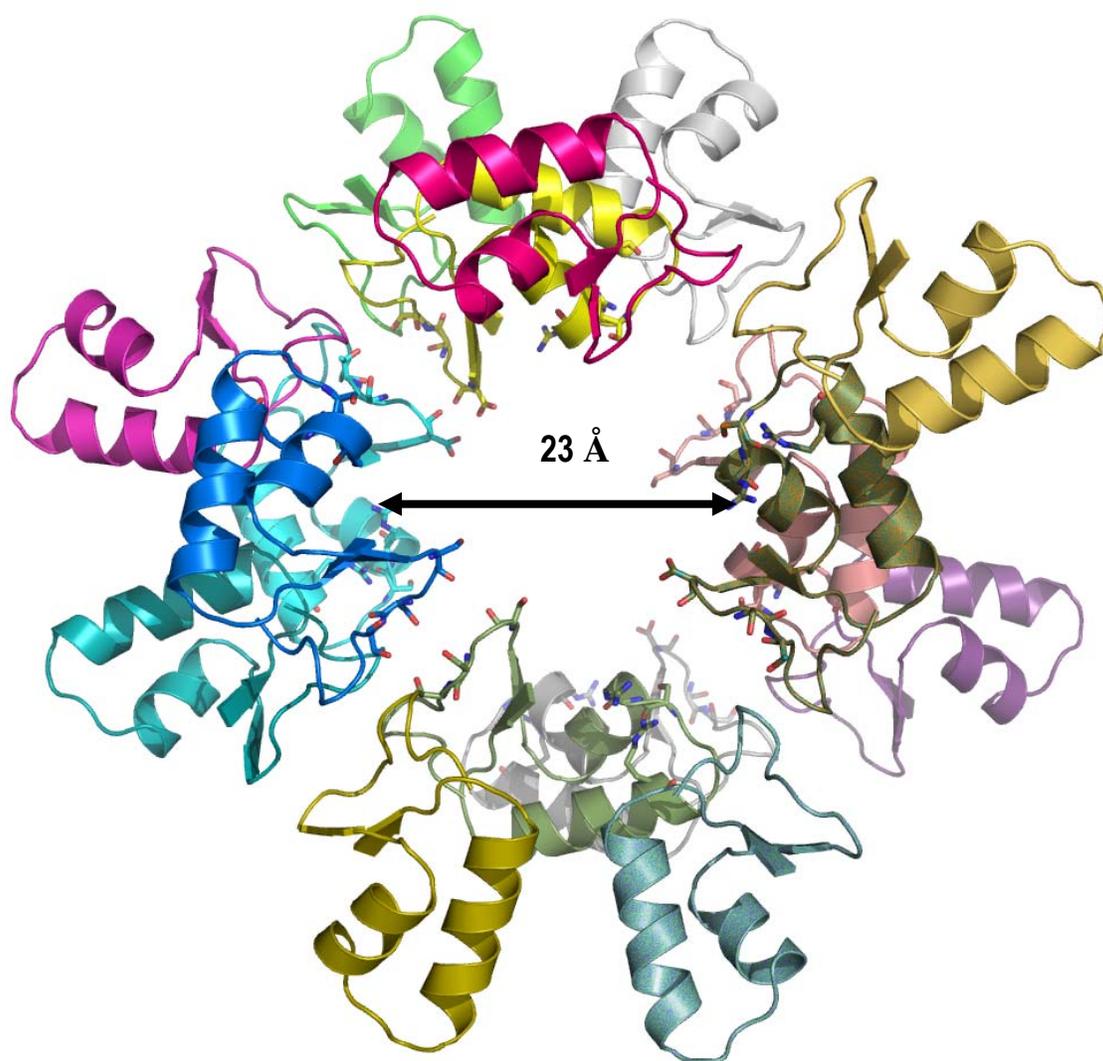
Figure 2.15: Tetrameric association of the dimers of Viscotoxin A1 form 23 Å solvent channels, with the charged residues pointing towards the channels and the hydrophobic helices facing away from the channels and interact with other hydrophobic helices.
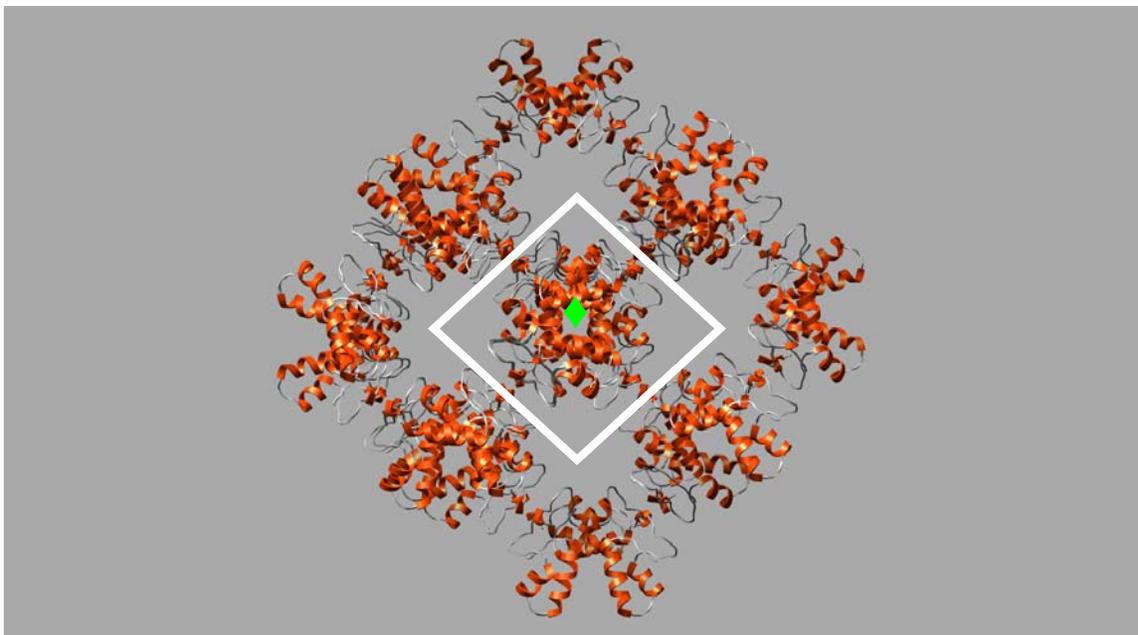
(B)

Figure 2.16: Crystal packing of Viscotoxin A1 molecules in the unit cell (The white box shows the unit cell and ♦ shows the crystallographic four fold axis).

2.4   Conclusion and Perspectives

Structural studies together with electrostatic calculations suggested Viscotoxin interactions with negatively charged molecular targets driven by electrostatic forces (Romagnoli *et al.*, 2000). Tyr^{13} is identified as a key residue for toxicity of Viscotoxins and other active members of the thionin family (Terras *et al.*, 1996). Purothionins completely lose their activity, when the Tyrosine residue was diiodinated or nitrated (Evans *et al*, 1989; Wada *et al.*, 1982). Crystallographic studies (Debreczeni *et al.*, 2003; Johnson, 2005) demonstrated the binding of negatively charged ligands around Tyr^{13} residues, proposing varied membrane binding mechanisms. The most conserved residues in Viscotoxins are in the binding pocket, so they alone cannot account for their varied toxicity (Debreczeni *et al.*, 2003). The hydrophobic residues in the helical regions vary considerably among the different isoforms. The variation in side chain at a particular non conserved position, namely residue 25 (Thr in Viscotoxin A1 and A3, Val in Viscotoxin A2, Arg in Viscotoxin B) can also influence the cytotoxic potential enormously (Büssing *et al.*, 1999; Debreczeni *et al.*, 2003). Therefore the presence of a positively charged active binding pocket along with a sterically accessible hydrophobic surface might just compliment each other in the binding of Viscotoxins to membranes. The variation in the hydrophobic interface at the helices might influence the specific binding of different membrane constituents and offer various patterns of molecular recognition (Debreczeni *et al.*, 2003). Although the formation of channels in crystals is common, the structure and the molecular association of Viscotoxin A1 reported in this thesis supports the view that Viscotoxins might be able to form ion or solvent channels or pores in cell membranes as proposed previously (Llanos *et al.*,2004; Hughes *et al.*,2000 and Giudici *et al.*, 2006).

3. Tendamistat from *Streptomyces tendae*

3.1 Introduction

3.1.1 Overview of α- amylase inhibitors

α- amylases (α- 1, 4-glucan-4-glucanohydrolase) are a group of enzymes widely distributed in microorganism, plant and animal secretions, catalyzing the hydrolysis of (α-1,4) glycosidic linkages in starch and various oligosaccharides, transforming polysaccharides into mono- and disaccharides (Grossi-de-Sá and Chrispells, 1997; Franco *et al.*, 2002; Pelegrini *et al.*, 2006). The hydrolysis of starch by α- amylases constitutes the first step in the enzymatic degradation of polysaccharides by α-amylases and is essential for carbohydrate assimilation. However inhibition of α-amylases may prove to be important in a number of cases: firstly control of insect pests and pathogens (fungi, bacteria), which severely destroy agricultural crops and secondly inhibition of α-amylases reduces the postprandial glucose peaks, particularly important in diabetes patients. The α-amylase inhibitors are used as a strategy to control pests as it reduces insect growth by reducing carbohydrate absorption (Yamada *et al.*, 2001).Inhibitors from bean were proposed to be applied as starch blockers to control obesity and diabetes owing to their effects on human and animal nutrition (Garcia-Olmedo *et al.*, 1987, P. Layer *et al.*, 1985).

Some of the α- amylase inhibitors show strict target enzyme specificity (R.J. Weselake *et al.*, 1983; O.L. Franco *et al.*, 2000), whereas others have high affinity for both mammalian and insect α- amylases (Franco *et al.*, 2002). However most of

the plant enzymes are insensitive to α- amylase inhibitors (Mundy *et al.*, 1983; Weselake *et al.*, 1983). Taking into account sequence similarity and three-dimensional structures, proteinaceous α- amylase inhibitors have been classified into seven classes (Franco *et al.*, 2002; Garcia-Olmedo *et al.*, 1987; Ho *et al.*, 1994; Svensson *et al.*, 2002): microbial, lectin-like, knottin-like, cereal-like, Kunitz-like, γ-purothionin like and thaumatin-like inhibitors.

Tendamistat belongs to the family of microbial α- amylase inhibitors. The proteins constituting this family have approximately 75 amino acids and are purified from various *Streptomyces* species. Tendamitat (Vértesy *et al.*, 1984), Haim (Murao *et al.*, 1980), Paim (Murao *et al.*, 1983) belong to this family and display approximately 30 % sequence identity and conserved disulfide topology. Sequence comparison of Tendamistat, Haim, and Paim (Figure 3.1) shows that the residues 18-20 are conserved in the three sequences. These residues were suggested to be involved in the inhibition mechanism (Vértesy *et al.*, 1984; Wiegand *et al.*, 1995).

Figure 3.1: showing sequence comparison between the three related microbial proteinaceous α- amylase inhibitors.

3.1.2 Brief Description of Tendamistat

Tendamistat is a 74 amino acid protein, produced by *Streptomyces tendae*. Tendamistat has been found to have strong inhibitory activity towards mammalian α-amylases, though the exact role of this inhibitory protein in the producing organism is unclear. Use of tendamistat to control postprandial glucose peaks in humans was

proposed but was unsuccessful owing to immunogenic effects (Aschauer *et al.*, 1983; Meyer *et al.*, 1984).Previously the crystal structure of tendamistat has been determined to 2 Å at room temperature (Pflugrath *et al.*,1986; PDB id: 1HOE), to 0.93 Å at 100K (König *et al.*, 2003; PDB id: 1OK0) and also as a part of complex with porcine pancreatic α- amylase to 2.5 Å (Wiegand *et al.*, 1995 PDB id: 1BVN).The complex of tendamistat to porcine pancreatic α- amylase involves 1330 Å^2 of surface area of the free tendamistat molecule resulting in steric blockage of the active site of the enzyme (Wiegand *et al.*, 1995).

Tendamistat has been used as a model protein for molecular dynamics simulations (Doruker *et al.*, 2000; Bonvin & van Gunsteren, 2000), for methods development for NMR techniques (Gorler & Kalbitzer, 1997; Scarselli et al., 1999), to study protein folding and stability (Schönbrunner *et al.*, 1997; Bachmann & Kiefhaber, 2001; Pappenberger et al., 2000, 2003; Balbach et al., 1998; Zscherp *et al.*,2003) and also as a template to design inhibitors(Ono *et al.*, 1998)

3.2   Materials and Methods

3.2.1 Purification

Tendamistat was purified from cultures for *Streptomyces tendae* 4158, ATCC 31210, by absorption resin and ion-exchange chromatography and crystallization at pH 5.0 (Vértesy *et al.*, 1984).

3.2.2 Crystallization

For crystallization, tendamistat was dissolved in double distilled water to a concentration of 30 mg/ml and centrifuged at 14000 RPM for 2 minutes to remove insoluble particles. Tendamistat was crystallized by hanging drop vapor diffusion

method (Hampel *et al.*, 1968; McPherson, A., 1992) at room temperature by equilibrating a drop containing 2 µl of protein solution and 2 µl of reservoir solution. Crystallization conditions were screened using Hampton Crystal Screen Kits (Jancarick & Kim, 1991; Cudney *et al.*, 1994). Block-like crystals (Figure 3.2A) of tendamistat were obtained initially using the screens, which were further optimized to obtain well diffracting crystals. In optimized conditions, crystals appeared in 4 days. Native crystals were obtained from the reservoir condition being 0.02 M Magnesium chloride hexahydrate, 0.2 M HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)) pH7.5, and 16-20 % Polyacrylic acid 5100 Sodium salt. All solutions used for crystallization were sterile filtered and contained 0.03% sodium azide.

3.2.3 Data Collection

Before measurement, the crystals were first soaked in a cryoprotectant solution containing 20 % glycerol and the rest as reservoir solution (in an increasing gradient of 5% for glycerol), then picked with a suitable size mitegen loop and finally flash frozen under liquid nitrogen stream and data were collected on a Bruker Rotating anode generator with Osmic focusing mirrors, a three circle diffractometer and a Smart 6000 4K CCD detector. Tendamistat crystals diffracted to 2 Å and due to the high symmetry space group, I $4_1$32 (which ensures respectable data redundancy within a reasonable measurement time) these crystals appeared to be tentative candidates for a sulfur-SAD experiment.

To ensure high data precision and redundancy, one 360 ° φ scan was collected at 0.18 ° oscillation width and four 180° ω scans at 0.2 ° oscillation width were collected. Data were collected, processed, scaled and analyzed using *PROTEUM*, *SAINT*, *SADABS* and *XPREP* (Bruker Nonius, 2002).

For the extension of phases, a dataset to 1.8 Å was measured at the PX2 beamline at Swiss Light Source. 360 images (Figure 3.2B) at an oscillation width of 1° were collected on a MAR225 mosaic CCD and processed with *XDS* (Kabsch, 1993).

To further improve the diffraction limit of the crystal, the crystals were annealed (Harp *et al.*, 1998, 1999; Juers *et al.*, 2007). The crystals were fished out using a Mitegen loop from the cryoprotectant solution, flash frozen under liquid nitrogen stream, taken out of the liquid nitrogen stream, put in the cryoprotectant solution for 30 seconds and finally shock frozen. Data were collected on this crystal to 1.5 Å at the PX2 beamline at Swiss Light Source. 900 high resolution images were collected at 150 mm distance with 0.4° rotation steps using a wavelength of 1 Å on a MAR225 mosaic CCD and processed with *XDS* (Kabsch, 1993). The crystal annealing was reproducible and consistently increased the resolution by 0.2 - 0.3 Å for other similar tendamistat crystals. The data collection statistics are summarized in Table 3.1.
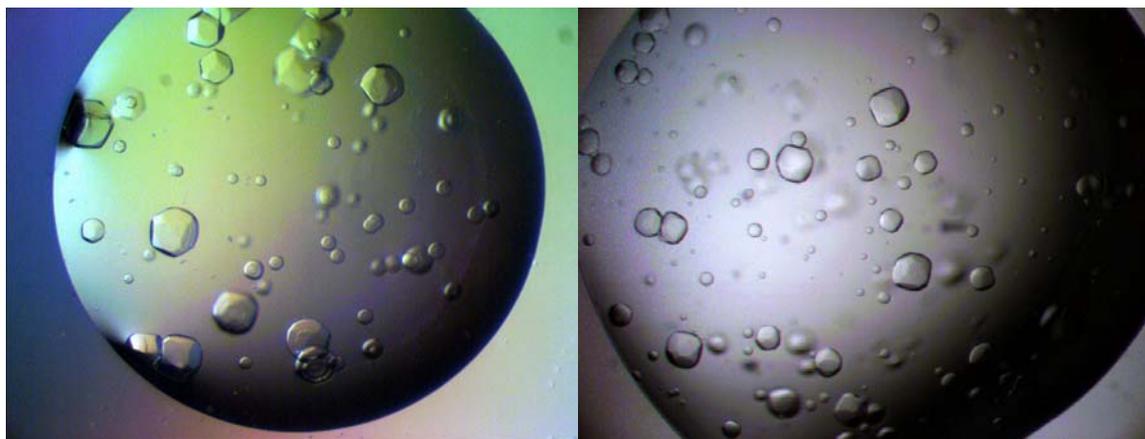
(A)
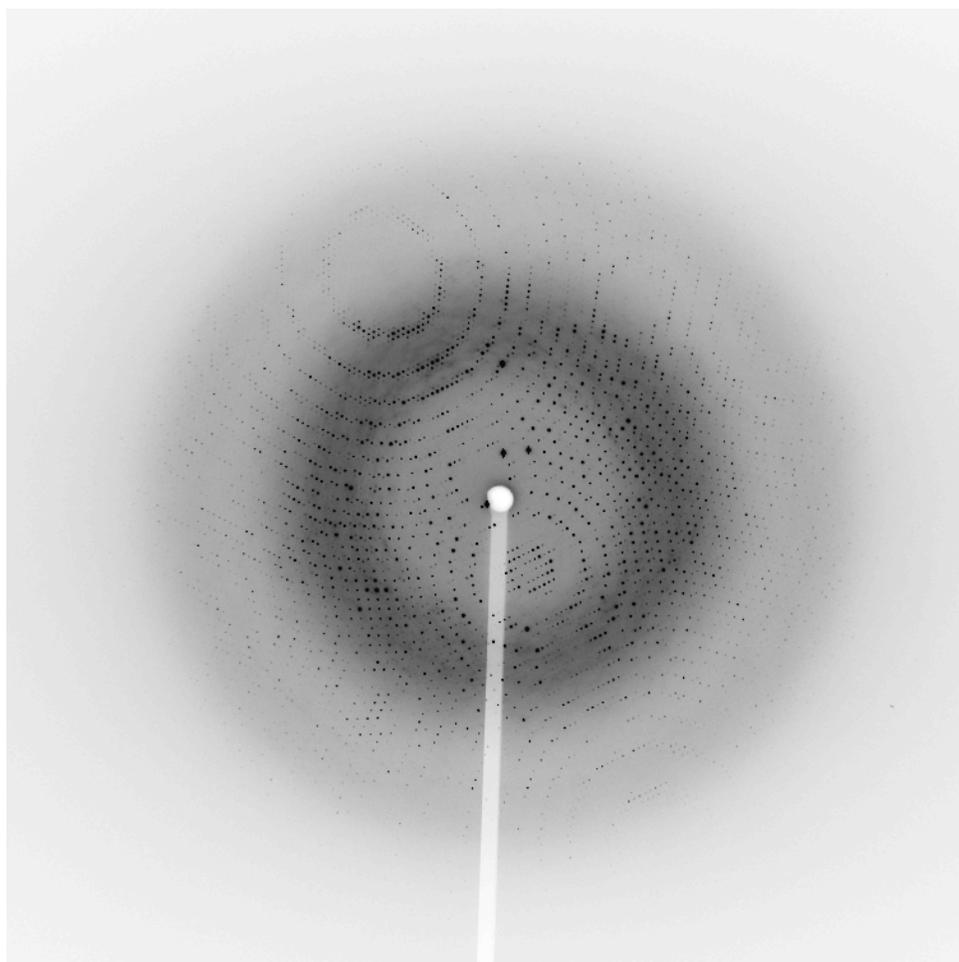
(B)

Figure 3.2: (A) Cubic crystals of Tendamistat. (B) Diffraction pattern at SLS PX2 beamline.

Table 3.1: Data collection statistics of Tendamistat in-house and at SLS-PX2 beamline. Values in parenthesis are for the outer resolution shell.

Data Statistics	In-house data	Synchrotron data
Wavelength (Å)	1.54	1.00
X-ray source	Cu Kα rotating anode	SLS-PX2
Detector	SMART6000 CCD	MARCCD
Space group	I 4_1 3 2	I 4_1 3 2
a (Å)	162.72	162.62
b (Å)	162.72	162.62
c (Å)	162.72	162.62
Resolution (Å)	2.00 (2.10-2.00)	1.50(1.60-1.50)
Unique Reflections	25155	52339
Redundancy	38.9 (17.5)	80.69 (49.89)
Completeness (%)	99.9 (100)	99.8 (99.5)
Mean I/σ(I)	75.22 (5.80)	41.16 (3.61)
R_{int}(%)	1.51 (17.13)	1.30 (24.75

3.2.4 Structure Solution and Phasing

The data processing and scaling for the ω and φ scans for the in-house data were performed separately, to preserve their independence.

All data from the ω and φ scans were scaled together using *SADABS* to improve the anomalous signal quality, taking advantage of the redundancy. Anomalous differences used in the subsequent steps were calculated using the dataset merged from the ω and φ scans. For the merged dataset, the anomalous $\Delta F/\sigma$ (ΔF) values (not shown) suggested a very weak anomalous signal and a cut off limit at around 3.0-3.5 Å (Schneider & Sheldrick, 2002).

The structure was solved using *SHELXD* (Sheldrick *et al.*, 2001) using the following settings:

SHEL 999 3.5
WEED 0.3
MIND -2.5
FIND 8
MIND -3.5 2.0
NTRY 10000
SEED 1

All the eight super sulfurs with good figures of merit and sharp cutoff were found using *SHELXD*, corresponding to the eight disulfide bridges (Figure 3.3). At 3.5 Å resolution, the sixteen sulfurs coalesce into eight super sulfurs, corresponding to the eight disulfide bridges, indicating a presence of four molecules in the asymmetric unit and a solvent content of ~54%.

CC All/ Weak	36.99	/	18.67	
	X	**Y**	**Z**	**height**
S001	0.250870	0.019806	0.078128	1.0000
S002	0.460640	-0.155312	0.068957	0.9813
S003	0.479126	-0.103943	0.117557	0.9681
S004	0.266479	-0.022041	0.046237	0.9559
S005	0.555992	-0.053497	0.140208	0.8731
S006	0.319756	0.016693	0.125056	0.8156
S007	0.314362	0.030296	-0.001851	0.8006
S008	0.405731	-0.164642	0.070804	0.7241
S009	0.352379	0.086998	0.110919	0.2616
S010	0.507919	-0.182701	0.154582	0.2376
S011	0.554886	-0.037086	0.072871	0.2305

Using the first eight atoms as input for *SHELXE* (Sheldrick, 2002), a clear distinction could be made between the two heavy atom enantiomorphs (using the contrast and connectivity figures of merit). A value of about 70% for the pseudo-free correlation coefficient indicates an interpretable map and is a good indication as to when phase refinement has converged (Sheldrick, 2002).

Mean weight and estimated mapCC as a function of resolution for original map

```
 d    inf - 3.28 - 2.59 - 2.26 - 2.05 - 1.90 - 1.79 - 1.69 - 1.62 - 1.56 - 1.50
 <wt>   0.640  0.598  0.554  0.553  0.537  0.460  0.364  0.312  0.223  0.221
 <mapCC> 0.847  0.821  0.766  0.766  0.734  0.639  0.480  0.395  0.274  0.265
d is resolution in Å, wt is weight, mapCC is map co-relation coefficient
 Pseudo-free CC = 51.39 %
```

Mean weight and estimated mapCC as a function of resolution for inverted map

```
 d    inf - 3.28 - 2.59 - 2.26 - 2.05 - 1.90 - 1.79 - 1.69 - 1.62 - 1.56 - 1.50
 <wt>   0.349  0.323  0.330  0.344  0.298  0.235  0.192  0.149  0.101  0.090
 <mapCC> 0.484  0.461  0.490  0.504  0.433  0.322  0.250  0.187  0.119  0.106
d is resolution in Å, wt is weight, mapCC is map co-relation coefficient
```

 Pseudo-free CC = 29.07 %

However, the map using *SHELXE* (Sheldrick, 2002) extended to 1.5 Å dataset proved weakly interpretable owing to the low pseudo-free correlation coefficient value. The extremely weak anomalous signal coupled with the low resolution cutoff for structure solution hindered the development of interpretable electron density

maps, though the positions of the super sulfurs were precise as compared to the positions of the disulfide bonds obtained from molecular replacement solution as tried later.

Ultimately the tendamistat structure could be solved using molecular replacement using brute force methods in *PHASER* (McCoy *et al.*, 2005), first performing a rotational search and then a translational search around this solution, where the previously determined structure of tendamistat (König *et al.*,2003; PDB id: 1OK0) was used as a search model. The side chains and alternate conformations for the residues of the search model were removed and a polyalanine model was used as a search model. Examples of the initial map and the final map are given in Figure 3.4.

(A)   (B)
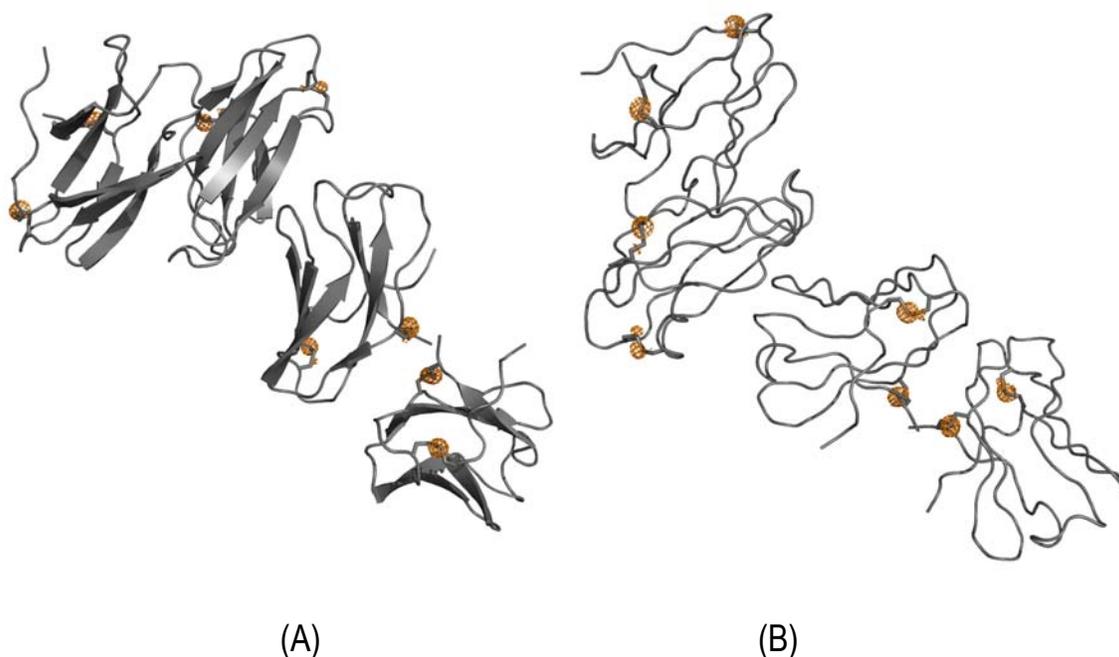
Figure 3.3: View of the (A) domain and (B) backbone representation of Tendamistat tetramer showing the sulfur atoms of disulfide linkages superimposed on the anomalous Fourier map calculated with S*HELXE*. The map was contoured at 3.0 Å.

(A)

(B)
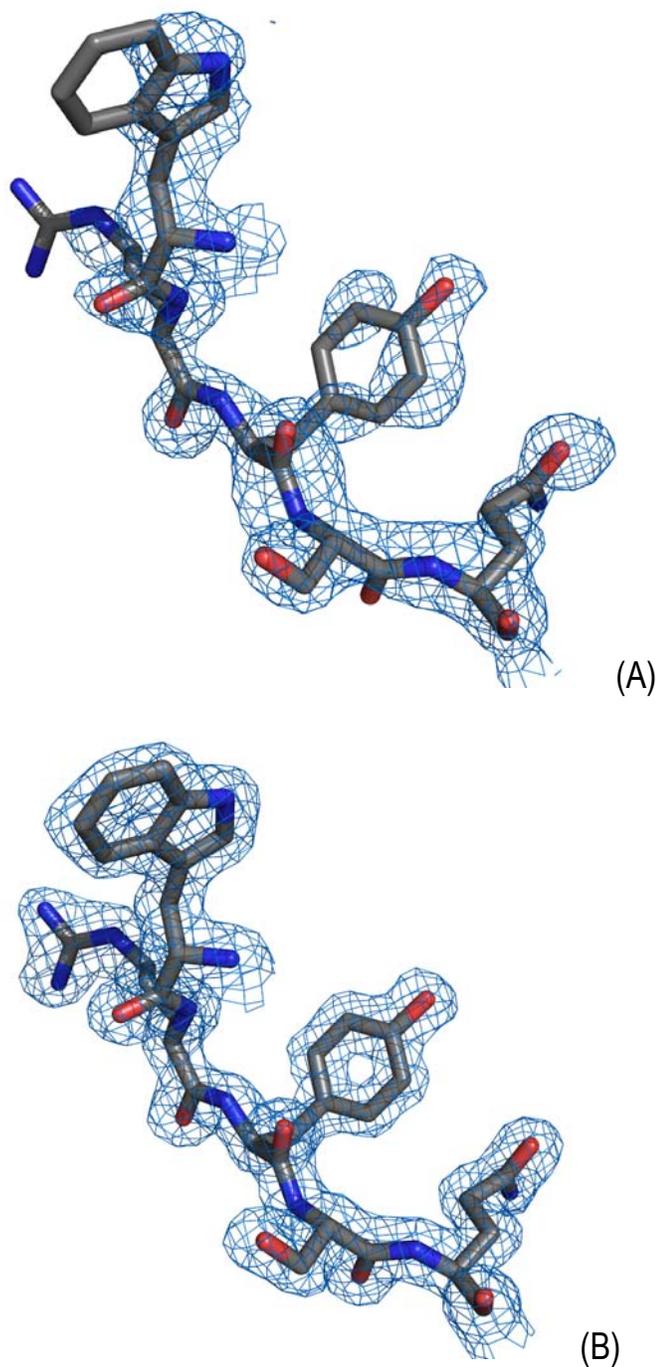
Figure 3.4: Examples of electron density maps around residues Trp[18], Arg[19], Tyr[20], Ser[21] and Gln[22] from the final refinement model for monomer D superimposed on (A) map produced by *SHELXE* after phase extension to 1.5 Å (B) final refinement map using the 1.5 Å dataset using the molecular replacement solution as the starting point. Maps are contoured at 1.5σ.

3.2.5 Model Building, Refinement and Structure Validation

The map from *PHASER* was traced completely and all the side chains docked using *ARP/wARP* (Perrakis *et al.*, 1999) and the known sequence of Tendamistat, in a standard run of 50 building cycles. Solvent molecules were added using *SHELXWAT* (Sheldrick & Schneider, 1997) and also *COOT* (Emsley & Cowtan, 2004).The structure was refined using Refmac_5.2.0019 (Murshudov *et al.*, 1997), alternating with model building using *COOT*. The first three residues, pertaining to the N-terminus were not visible in the electron density map for all the four molecules in the asymmetric unit and hence, were not modeled.

The final model comprises of four independent peptide chains containing 284 residues and 251water molecules. The final model was validated with *PROCHECK* (Laskowsky et al., 1993) and *MolProbity* (Lovell *et al.*, 2003). All amino acid residues were found to be in the allowed regions of the Ramachandran plot (Ramachandran & Sasisekharan, 1968) as shown in Figure 3.5. The refinement statistics have been summarized in Table 3.4.
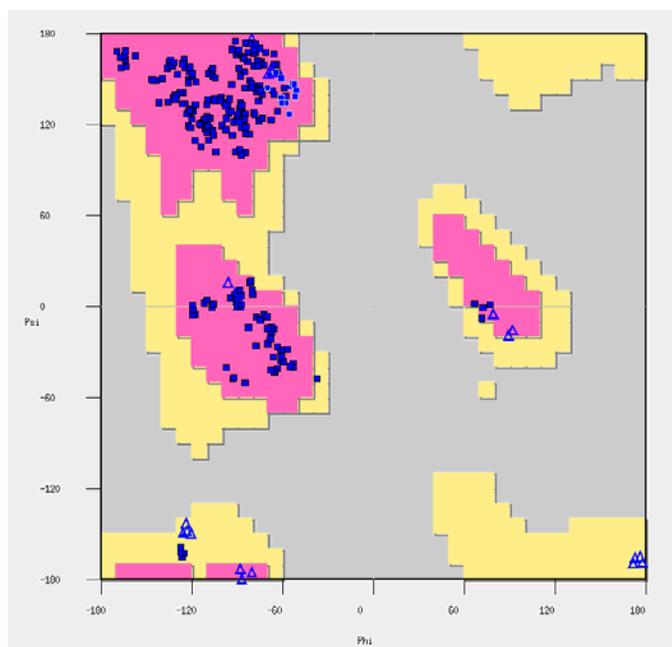
Figure 3.5: Ramachandran analysis plot for the structure of Tendamistat.

Table 3.4: Refinement statistics for Tendamistat.

Refinement Statistics	Tendamistat
R-Factor (%)	21.224
Free R-Factor (%)	24.572
R.m.s. Deviations from ideal geometry	
Bond Lengths (Å)	0.016
Bond angles (°)	1.600
No. Of Protein Residues	284
No. of Water Molecules	318
Ramachandran Plot (%)	
Residues in preferred regions	95.44
Residues in allowed regions	4.56
Residues in disallowed regions	0

3.3   Structure Description

3.3.1 Overall Fold

Four similar molecules, forming a tetramer, related by non-crystallographic symmetry, were observed in the asymmetric unit. Tendamistat has an immunoglobulin-like fold, constituted by two three stranded anti-parallel ß- sheets, stabilized by two disulfide bridges. Each monomer consists of 74 amino acid residues, out of which the first three residues at the N-terminus cannot be located in the electron density, as they are disordered. The overall folding of the protein is shown in Figure 3.6. The main structural elements are a pair of three antiparallel ß-strands. The first of the pair of three antiparallel ß-strands is constituted of residues $Cys^{11}$to Ser^{17} (S1), Tyr^{20} to Cys^{27} (S2) and $Gly^{51}$to Gly^{57} (S5), stabilized by disulfide bridge between residues Cys^{11} and Cys^{27} .There is a ß- turn between residues 17 to 20, containing the triplet Trp^{18}, Arg19 and Tyr^{20}.This triplet is conserved in all inhibitors and is suggested to be essential in the formation of inhibitor α-amylase complexes and hence its inhibitory activity (Hofmann *et al.*, 1985; Arai *et al.*, 1985; Wiegand et al., 1995). The other three stranded antiparallel ß- sheet is formed by residues Val^{31} to Tyr^{37} (S3), Leu^{44} to Ala^{49} (S4) and Arg^{68} to Arg^{72} (S6), with a disulfide bond between residues Cys^{45} and Cys^{73}.The rest of the residues form loops.
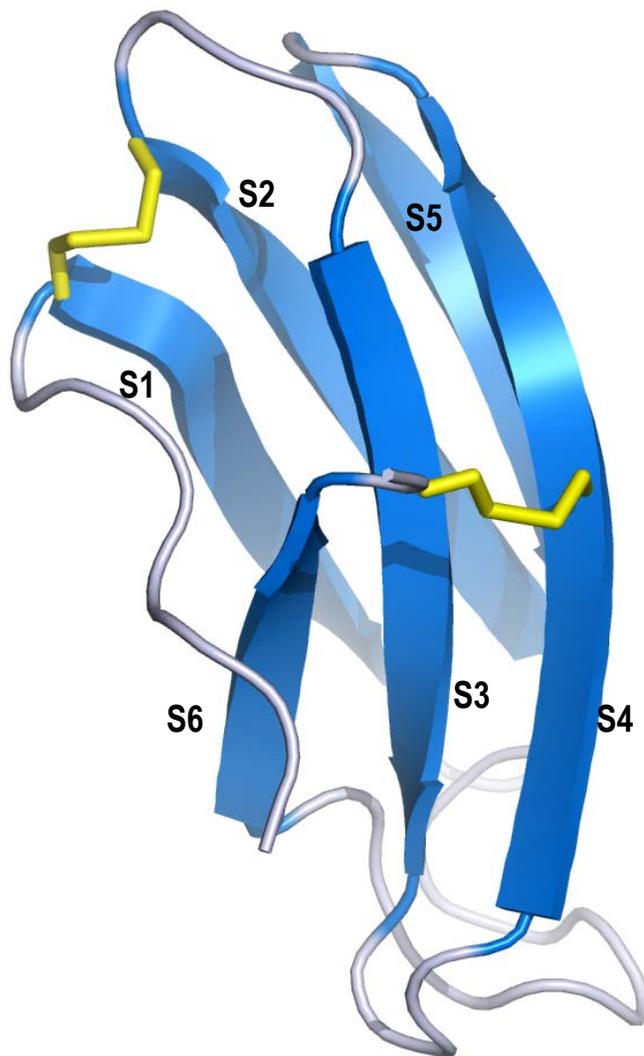
Figure 3.6: Overall fold of the tendamistat molecule showing the six ß- strands and disulfide bridges' locations in the Tendamistat molecule.

3.3.2 Tetramerization and Crystal Packing

The unit cell contains a total of 192 individual molecules constituting 46 % of the volume whereas solvent content of the crystal is about 54%.Four molecules form a linear tetramer in the asymmetric unit as shown in Figure 3.7 and are denoted by TdsA, TdsB, TdsC, TdsD for the respective chains. Table 3.2 lists the main

66

interactions involved in the tetramer formation and crystal packing. Most of the interactions are common to al the molecules except for the interactions between the terminal carboxyl group of Leu[74].The terminal carboxyl group of residues Leu[74] of TdsB and TdsD interact with the positive side chain residues Lys[34] of TdsD and TdsB respectively, whereas the Leu[74] residues of TdsA and TdsC interact with the symmetry equivalents of residues Arg[72] of TdsC and TdsA respectively.
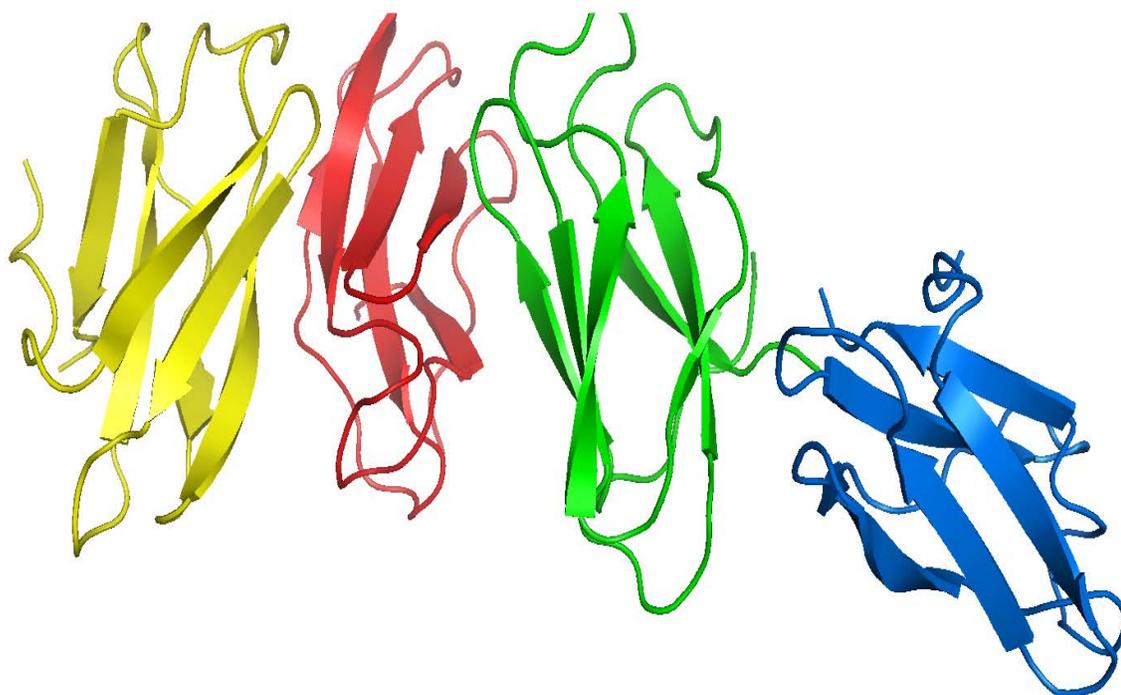
Figure 3.7: Linear tetrameric association of Tendamistat molecules in the asymmetric unit.

Table 3.2: Main interactions between residues involved in the tetramer formation and also involved in crystal packing (TdsA, TdsB, TdsC and TdsD represent monomers A, B, C and D respectively; symm denotes the symmetry equivalents).

TdsA/Tyr^{15}- TdsA/His^{66}(symm)	TdsA/Trp^{18}-TdsA/Tyr^{20}(symm)	TdsA/Arg^{19}-TdsC/Gln^{22}
TdsA/Tyr^{20}- TdsA/Trp^{18}(symm)	TdsA/Tyr^{46}-TdsC/Tyr^{46}	TdsA/His^{66}- TdsA/Tyr^{15}(symm)
TdsB/ Tyr^{15}- TdsC/His^{66}	TdsB/ Trp^{18}-TdsC/Tyr^{20}	TdsB/ Arg^{19}-TdsB/Gln^{22}(symm)
TdsB/Tyr^{20}- TdsA/Trp^{18}	TdsB/Tyr^{46}-TdsB/Tyr^{46}(symm)	TdsB/His^{66}- TdsC/Tyr^{15}
TdsC/Tyr^{15}- TdsB/His^{66}	TdsC/ Trp^{18}-TdsB/Tyr^{20}	TdsC/Arg^{19}-TdsA/Gln^{22}
TdsC/Tyr^{20}- TdsB/Trp^{18}	TdsC/Tyr^{46}-TdsA/Tyr^{46}	TdsC/His^{66}- TdsB/Tyr^{15}
TdsD/Tyr^{15}- TdsD/His^{66}(symm)	TdsD/Trp^{18}-TdsD/Tyr^{20}(symm)	TdsD/Arg^{19}-TdsD/Gln^{22}(symm)
TdsD/Tyr^{20}- TdsD/Trp^{18}(symm)	TdsD/Tyr^{46}-TdsD/Tyr^{46}(symm)	TdsD/His^{66}-TdsD/Tyr^{15}(symm)
TdsB/Lys^{34}-TdsD/Leu^{74}	TdsD/Lys^{34}-TdsB/Leu^{74}	TdsA/Agr^{72}-TdsC/Leu^{74}(symm)
TdsC/Leu^{74}-TdsA/Arg^{72}(symm)		

3.3.3 Comparison with other crystal structures of tendamistat

The r.m.s deviations for the C^{α} atoms of four individual tendamistat molecules of the current structure and the other reported crystal structures (Tendamistat to 0.93 Å at 100K (König *et al.*, 2003; PDB id: 1OK0), complex with porcine pancreatic α-amylase to 2.5 Å (Wiegand *et al.*, 1995 PDB id: 1BVN) and are listed in Table 3.3 and shown in Figure 3.8.

The current molecule was crystallized at pH 7.5, whereas 1Ok0 was crystallized at pH 1.6 and 1BVN at pH 9.0. It can be emphasized that (within experimental error), the backbone conformation of the residues are similar. However the main difference is in the conformation of residues Tyr[15], Trp[18], Arg[19], Tyr[20], which are considered to be important in the interaction with α- amylases. In the structure of complex with porcine pancreatic α- amylase to 2.5 Å (1bvn) and 0.93 Å structure (1ok0), the hydrophobic part of residue Arg[19] is stacked between the stacked aromatic side chains of Trp[18] and Tyr[20], with the guanidinium group sticking out into the solvent, whereas in the current structure, the conformation of these specified side chains is different, and Arg[19] is not stacked between the stacked aromatic side chains of Trp[18] and Tyr[20] as visible from Figure 3.9.
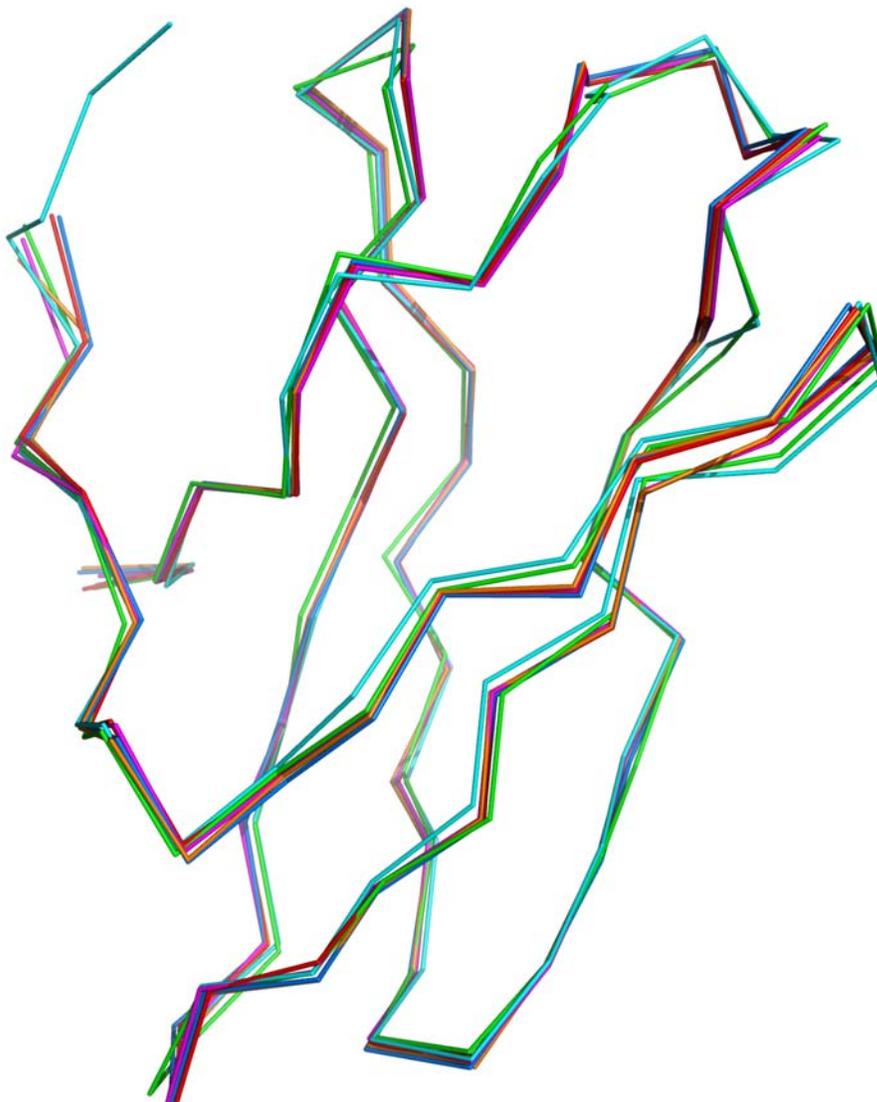
Figure 3.8: Superposition of Tendamistat structure (**TdsA**, **TdsB**, **TdsC**, **TdsD** for respective molecules), Tendamistat structure at 0.93 Å resolution (**1ok0**) and structure of Tendamistat in complex with Pig pancreatic α- amylase (**1bvn**).

Table 3.3: showing the r.m.s. deviations among current reported Tendamistat structure (**TdsA**, **TdsB**, **TdsC**, **TdsD** for respective molecules), Tendamistat structure at 0.93 Å resolution (**1ok0**) and structure of Tendamistat in complex with Pig pancreatic α- amylase (**1bvn**) for C^{α} atoms.

	1ok0	1bvn	TdsA	TdsB	TdsC	TdsD
1ok0	0	0.429	0.605	0.654	0.543	0.582
1bvn	0.429	0	0.566	0.602	0.504	0.558
TdsA	0.605	0.566	0	0.226	0.279	0.301
TdsB	0.654	0.602	0.226	0	0.315	0.260
TdsC	0.543	0.504	0.279	0.315	0	0.242
TdsD	0.582	0.558	0.301	0.260	0.242	0
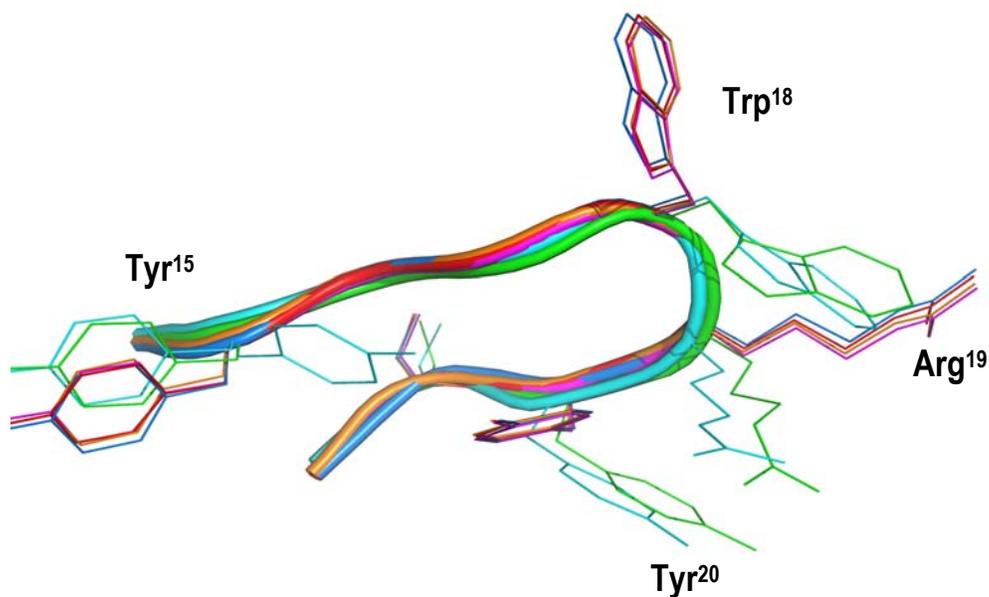
Figure 3.9: Superposition of residues Tyr[15], Trp[18], Arg[19] and Tyr[20], showing the difference in side chain conformations of residues Trp[18], Arg[19] and Tyr[20] for the chains denoted by colors **TdsA**, **TdsB**, **TdsC**, **TdsD**, **1ok0** and **1bvn**. Coloring scheme is same as in Table 3.3.

3.4   Conclusion and Perspectives

The structure of tendamistat could be solved using the weak anomalous signal of sulfur atoms, though the map could not be traced automatically using the existing automatic programs. Because of the high symmetry space group in which Tendamistat crystallizes in the current reported structure, it can also be used as a test crystal to assess experimental equipments: Firstly a fast data collection can be done on the crystal to full completeness and used to identify problems in complex hardware and software infrastructure of modern crystallographic X-ray equipments or for the calibration of the same. Secondly the low anomalous signal can be used as a benchmark to calibrate the precision of crystallographic X-ray equipments.

The overall fold of Tendamistat is similar to that of the previously determined structures of the same, though the previous structures crystallize at different pH values (pH 1.6 for 1Ok0 and pH 9.0 for 1BVN). This suggests the rigidity of Tendamistat. Differences are observed in the side chain conformations of residues Tyr[15], Trp[18], Arg[19] and Tyr[20]. Incidentally, these residues are involved in inhibition of α-amylase. The conformational flexibility of these residues could be on one hand basis of broad target specificity and on the other hand to aid in optimization of binding by conformational adjustments (König *et al.*, 2003; PDB id: 1OK0). This flexibility of the binding residues can be capitalized to design specific α-amylase inhibitors.

4.   Tripeptidyl Peptidase from *Homo sapiens*

4.1   Introduction

4.1.1 Overview of Proteases

Proteases (proteinases, peptidases) are degradative enzymes which catalyze protein breakdown by hydrolyzing peptide bonds. They are involved in many important diverse intracellular and extracellular physiological as well as pathophysiological processes and are important targets for drug design. Approximately 2% of the genes in humans and mice code for proteases (Puente *et al.*, 2003; Rawlings *et. al.*, 2006). Proteases have evolved, adapted to variances in pH, reductive environment, to name a few of the varied range of conditions found in simple and complex organisms. Furthermore, they are specialized to perform certain specific functions. Proteases can be categorized into six different groups depending on the active site catalytic residue as serine, threonine, cysteine, aspartic, glutamic and metalloproteases, as annotated in the MEROPS database (Rawlings *et al.*, 2006). Most of these different groups of proteases are further classified into families and subfamilies depending on the three dimensional folds, the catalytic residues and sequence homology. A brief overview of the overall folding patterns of various proteases and their active sites, from recent literature (Tyndall *et al.*, 2005) and from MEROPS database (Rawlings *et. al.*, 2006) is discussed.

Serine proteases cleave the peptide bond by nucleophilic attack of the serine hydroxyl group on the scissile carbonyl bond of the substrates. Serine proteases

adopt either a trypsin-like (Figure 4.1, Conners *et al.*, 2007) or subtilisin-like structural fold (Figure 4.2, Radisky *et al.*, 2004). Subtilisin-like serine proteases are further subdivided into a large number of subfamilies, having a low sequence homology and often limited to regions surrounding the catalytic triad (Siezen *et al,* 2007; Istavn and Wlodawer, 2007). Recently, some new serine protease structures have been characterized and determined, which do not have the classical His-Asp-Ser triad, but may have a Ser-Lys dyad as in case of Lon Protease B (Botos *et al.*, 2005) or a Ser-Glu-Asp catalytic triad as in sedolisins (Wlodawer *et al.*, 2001) or a catalytic dyad composed of Ser and His as in case of Rhomboid GlpG protease (Wang *et al.*, 2006). A representation of the active sites of these few distinct serine proteases is shown in Figure 4.3.

(A)   (B)
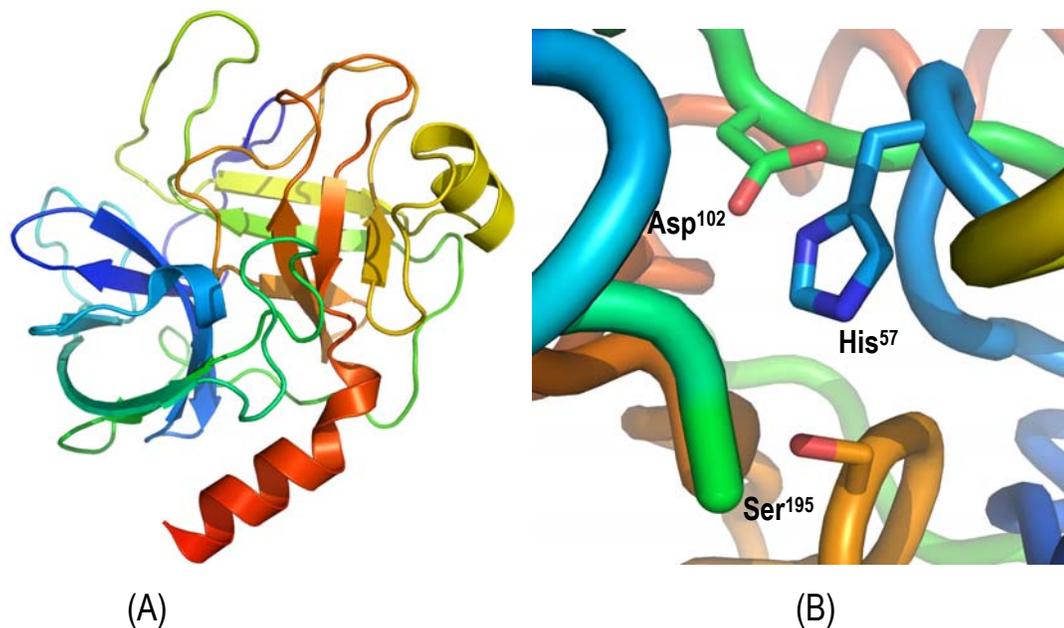
Figure 4.1: (A) Structural fold of trypsin-like serine protease and (B) the active site residues His[57], Asp[102] and Ser[195] (PDB id: 2CMY).
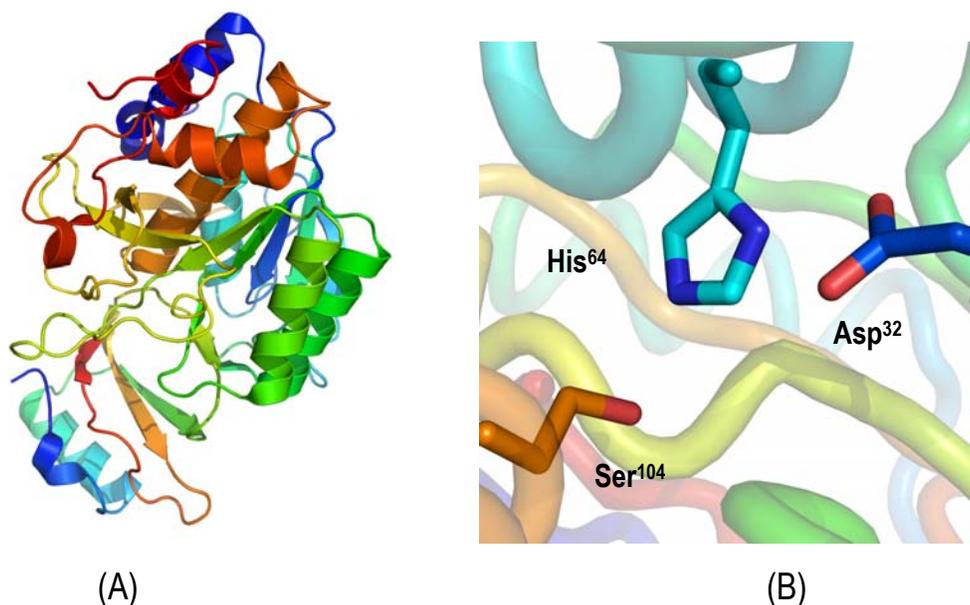
Figure 4.2: (A) Secondary structure representation of subtilisin-like serine protease and (B) active site residues Asp^{32}, His^{64} and Ser^{104} (PDB id: 1TO2).
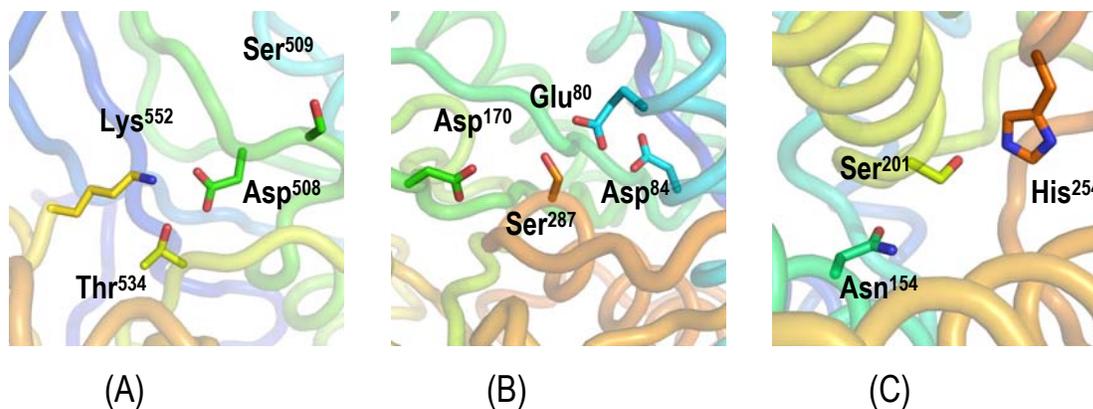
Figure 4.3: (A) Catalytic dyad residues Ser^{509} and Lys^{552} in proteolytic domain of B-type Lon protease (PDB id: 1Z0W). In absence of substrate, instead of Ser^{509}, Asp^{508} interacts with Lys^{552}. (B) Catalytic triad residues Glu^{80}, Asp^{84} and Ser^{287} in sedolisins, additionally Asp^{170} is also required for catalysis and is structurally equivalent to the Asn^{155} in subtilisin (PDB id: 1GA6). (C) Rhomoid protease catalytic dyad formed by Ser^{201} and His^{254}, the Asn^{154} creates the oxyanion hole (PDB id: 2IC8).

Threonine proteases catalyze the cleavage of peptide bonds in which the nucleophile threonine is the primary catalytic residue. Members of this new class of protease include the 20S archaean proteasome (Seemuller *et al.*, 1995; Coux *et al.*, 1996) and Hs1V component of Hs1UV peptidase (Rawlings *et al.*, 2006). The overall fold is shown in Figure 4.4 (Löwe *et al.*, 1995).

(A)   (B)
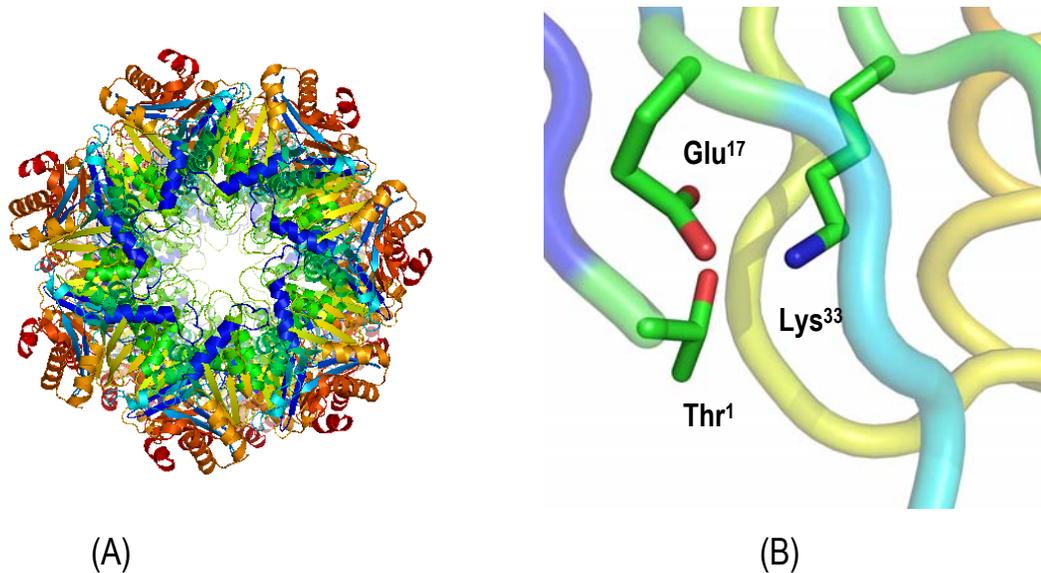
Figure 4.4: (A) overall fold of the structure of archaean 20S proteasome, having a threonine protease domain and (B) the active site residues Thr^1, Glu^{17} and Lys^{33} (PDB id: 1PMA).

Cysteine proteases use the sulfur-hydrogen bond of a cysteine residue to catalyze peptide bond cleavage. Three main structural folds are observed for cysteine proteases: viral-like, papain-like or caspase-like. A Trypsin-like fold is observed in the viral-like cysteine proteases like tobacco etch virus (Figure 4.5, Phan *et al.*, 2002). Papain-like fold is adopted by the papain like proteases which include papain and cathepsinB (Figure 4.6, Kamphuis *et al.*, 1985). The caspase like cysteine proteases adopt a subtilisin-like fold, as observed in caspase1 and gingipain (Figure 4.7, Eichinger *et al.*, 1999).

(A)   (B)
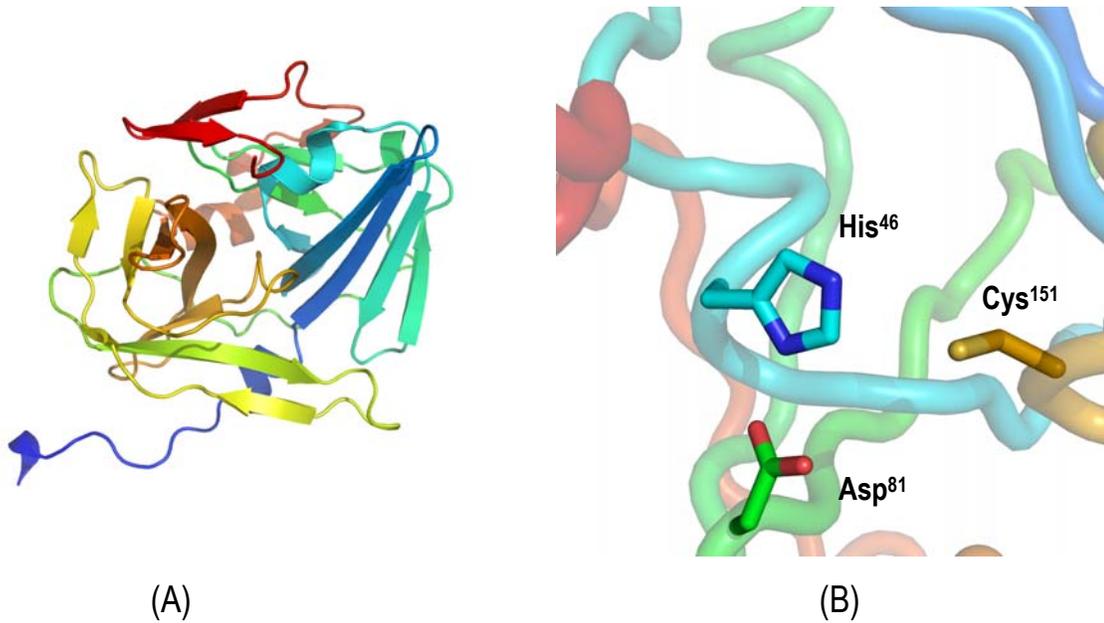
Figure 4.5: (A) A viral like cysteine protease showing a trypsin like fold and (B) active site residues His[46], Asp[81] and Cys[151] (PDB id: 1LVM).

(A)   (B)
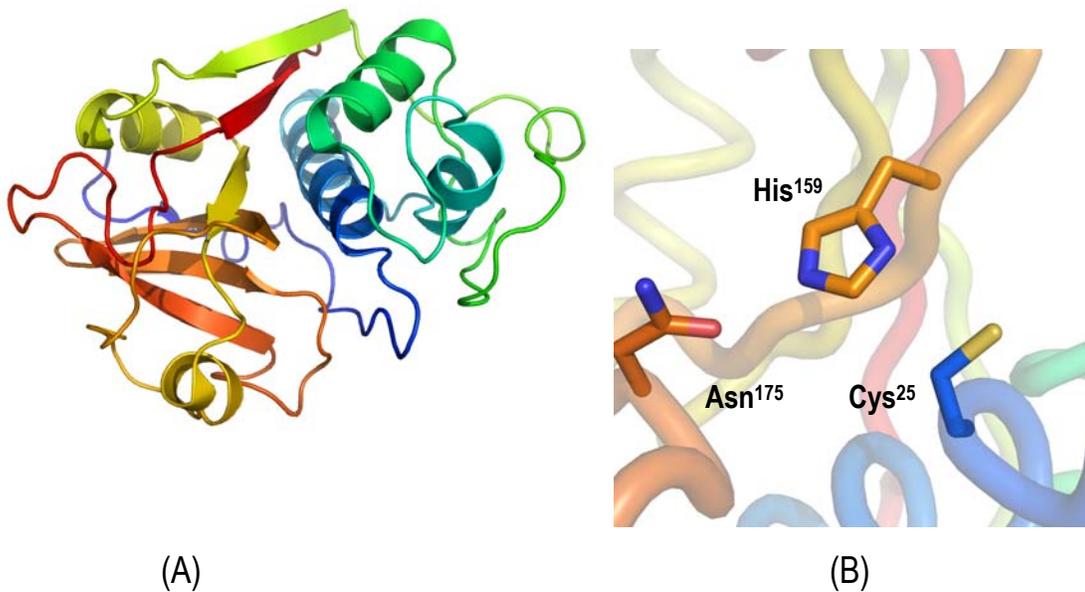
Figure 4.6: (A) Papain like cysteine protease showing α-ß complex fold and (B)Catalytic residues Cys[25], His[159] and Asn[175] (PDB id: 9PAP).

(A)   (B)
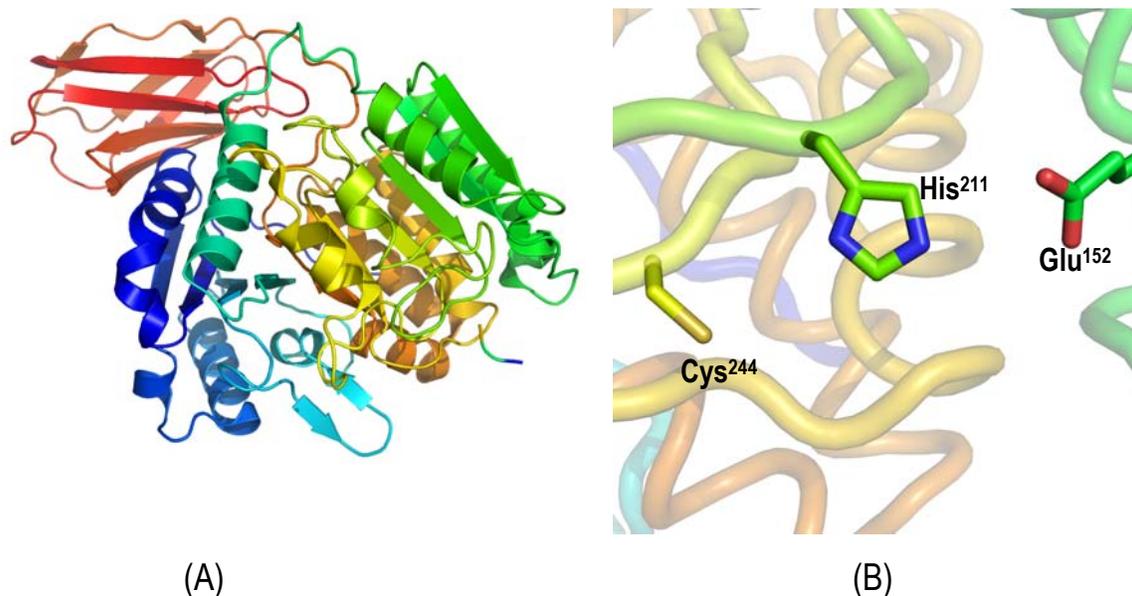
Figure 4.7:(A) Caspase-like cysteine proteases adopt a Rossman fold and (B) Catalytic residues His[211], Cys[244], and third proposed residue, Glu[152] (PDB id: 1CVR).

Aspartic proteases catalyze peptide bond cleavage using the acidic carboxyl groups of two aspartate residues. These proteases adopt either pepsin-like fold or retropepsin-like fold. Members exhibiting the pepsin fold are Cathepsin D and memapsin. They commonly have a bilobal structure with two similar ß-barrel domains, each contributing an aspartate residue to the catalytic site (Figure 4.8, Fujinaga *et al.,* 1995; Tang *et al.*, 1978; Davies, 1990; Blundell *et al.*, 1991) Retropepsin-like fold is observed in members of this family which are generally of viral origin and are structurally different to the pepsin-like aspartic proteases, having two ß-barrel domains forming an active homo-dimer (Figure 4.9, Swain *et al.*, 1990; Dunn *et al.*, 2002; Wlodawer *et al.*, 2000).

(A)   (B)
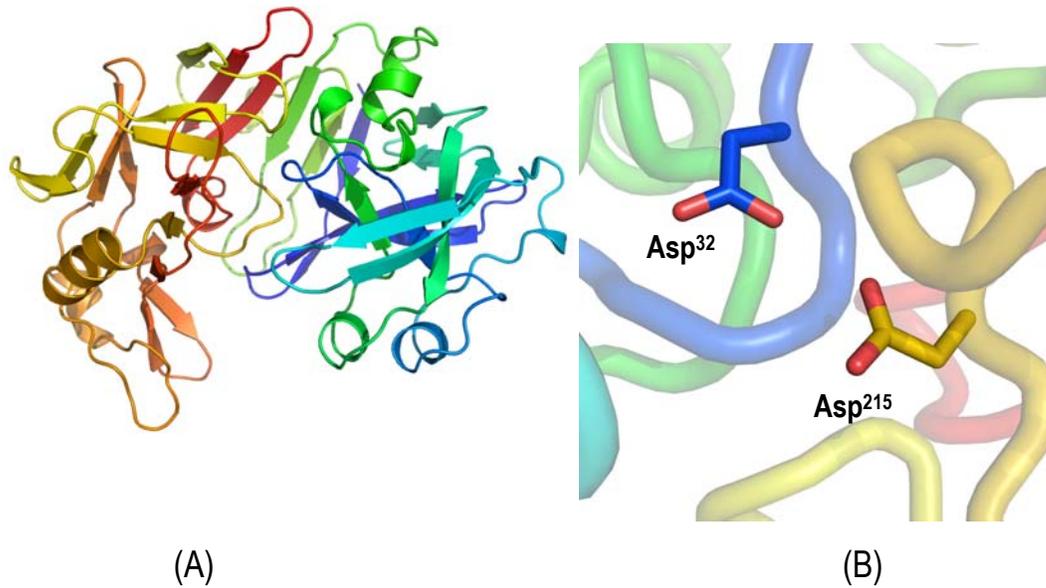
Figure 4.8: (A) Pepsin like-fold in Aspartic proteases as observed in Human Pepsin and (B) catalytic residues Asp[32] and Asp[215] (PDB id: 1PSO).

(A)   (B)
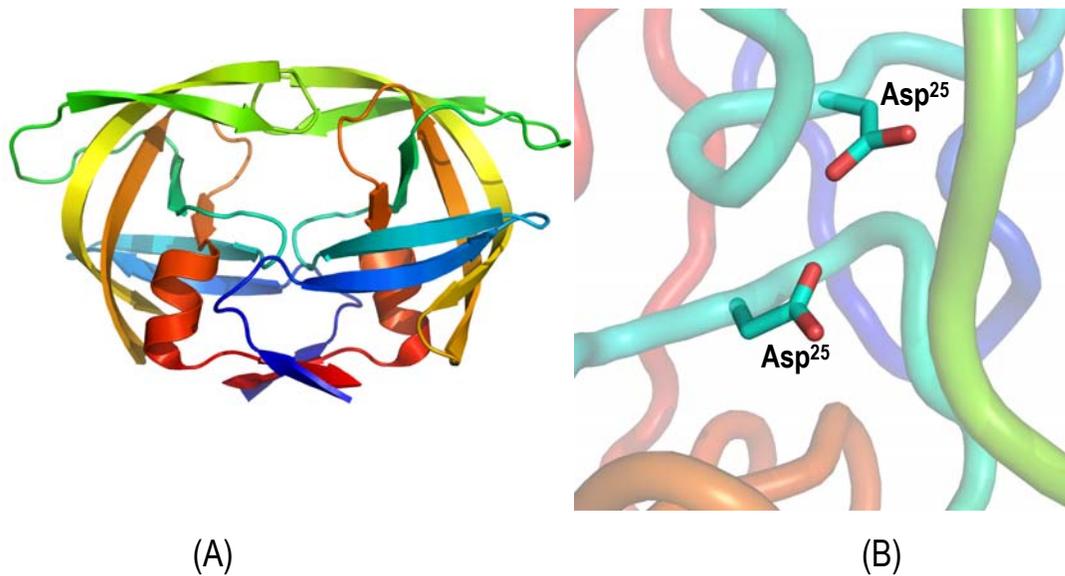
Figure 4.9: (A) Retropepsin aspartic protease ß-barrel in Human Immunodeficiency Virus and (B) two catalytic Asp[25] residues, contributed by each of the chains to form the active homo-dimer (PDB id: 7HVP).

Glutamic proteases have a catalytic dyad of glutamate and glutamine residues and have an unusual structure composed of a pair of seven-stranded antiparallel ß-sheets. Scytalidoglutamic peptidase-B, SCP-B (figure 4.10, Fujinaga *et al.*, 2004; Pillai *et al.,* 2007) is the first discovered member of this sixth family of peptidases.

(A)   (B)
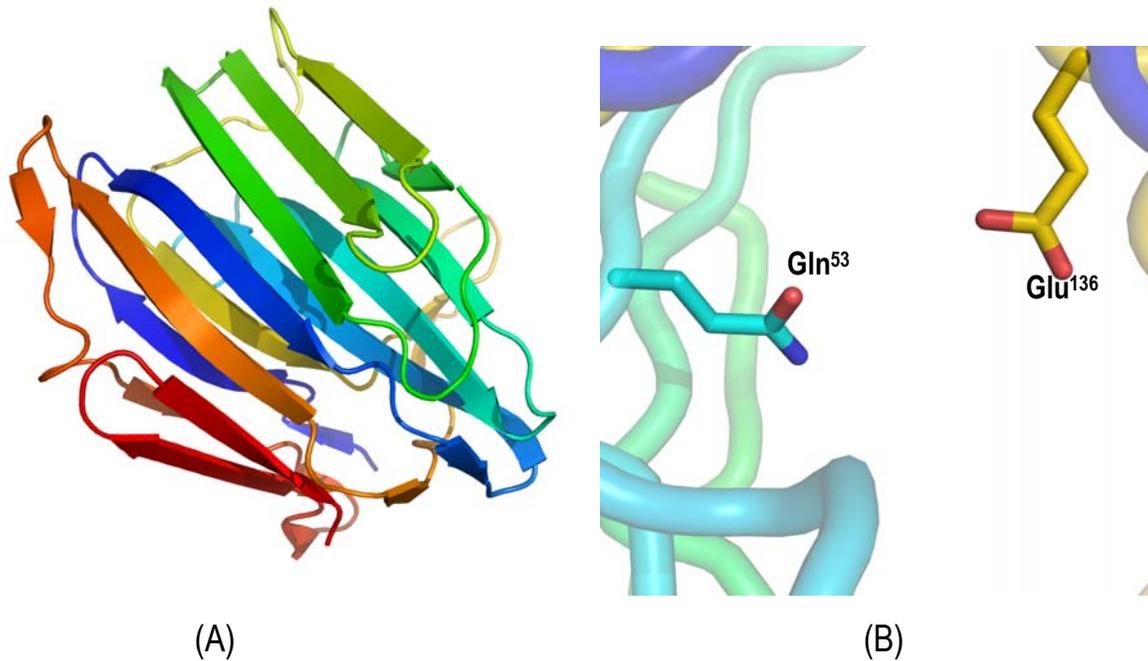
Figure 4.10: (A) An unusual fold exhibited by the only determined structure of the Glutamic protease, Scytalidoglutamic peptidase-B and (B) catalytic dyad residues Gln^{53} and Glu^{136} (PDB id: 2IFR)

Metalloproteases contain a catalytic thermolysin domain, in addition to several other domains and also have a tightly bound zinc atom to the catalytic centre composed of two histidines (Figure 4.11, Bode *et al.*, 1993; Hooper, 1994). Thremolysin-like proteases constitute the largest clan of metalloproteases. Angiotensin-converting enzyme peptidase unit 1 and Thermolysin are one of the many members of this family.

(A)   (B)
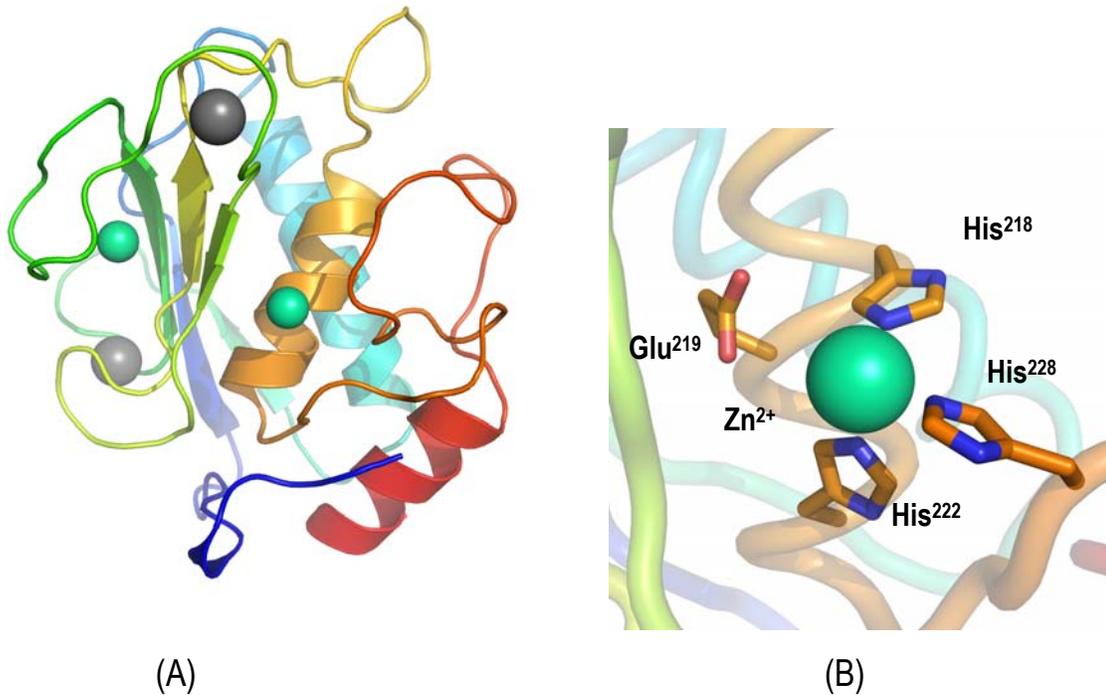
Figure 4.11: (A) Thermolysin like catalytic domain adopted by most metalloproteases as also in human Matrilysin here. The zinc ions are shown in cyan and calcium ions in grey and (B) Active site residues His[218], Glu[219], His[222] and His[228] coordinating a zinc ion, which is essential for the activity of a third of these type of proteases (PDB id: 1MMQ).

4.1.2 Tripeptidyl-peptidase- I

Tripeptidyl-peptidase-I (TPP-I) is a lysosomal aminopeptidase that cleaves off tripeptides from the free N-termini of polypeptides and also shows minor endopeptidase activity (Tomkinson, 1999; Ezaki *et al.,* 2000) but its exact physiological function is not known. It belongs to the S53 family of serine proteases, also called the sedolisins (Rawlings *et. al.*, 2006). Mutations in TPP-I are associated with a fatal neurodegenerative disease called the classical late-infantile neuronal ceroid lipofuscinosis, also abbreviated as LINCL or CLN2 (Sleat *et al.*, 1997, 1999). Patients with naturally occurring mutations in TPP-I develop seizures and cognitive decline by 2-4 years of age, followed by visual loss and death in the second decade of their life (Wisniewski *et al.*, 2001). Furthermore, there is also accumulation of autofluorescent material in the various tissues and organs with severe damage incurring in the brain and retinae (Kida *et al.*, 2001). Loss of TPP-I activity is usually associated with accumulation of hydrophobic subunit c of ATP synthase (Ezaki *et al.,* 1995, 1996, 1997, 1999, 2000). TPP-I also plays a major role in the physiological degradation of cholecystokinin-8 in mouse (Warburton and Bernardini, 2002; Bernardini and Warburton, 2002). Also it was shown that *in vitro* TPP-I degrades several neuropeptides, peptide hormones, synthetic amyloid ß-peptides 1-42 and 1-28, subunit C of mitochondrial ATP synthase and is also involved in degradation of bone collagen (Vines *et al.*, 1998; Du *et al.*, 2001, Junaid *et al.*, 2000; Warburton and Bernardini, 2002; Ezaki *et al.*, 1999; Page *et al.*, 1993). Hence, proper functioning of TPP-I is crucial to various biological processes, especially to the neuronal cells of the central nervous system (Golabek *et al.*, 2005).

TPP-I is synthesized as an inactive precursor, which upon arrival to the lysosomes, are proteolytically cleaved to yield the mature active enzymes (Golabek *et al.,* 2004). The human TPP-I in the preproenzyme form has 563 amino acid

residues in total. The first 19 residues form a signal peptide, which is cleaved off cotranslationally or when the protein is translocated into the endoplasmic reticulum lumen. Furthermore, a 176-residue propeptide is removed during the maturation process to yield a mature enzyme of 368 amino acid residues (Sleat *et al.*, 1997; Lin *et al.*, 2001; Liu *et al.*, 1998).

The proenzyme has a mass of 66 kDa and the mature enzyme has a molecular mass of 46-48 kDa as demonstrated by SDS-PAGE (Figure 4.24). It has five potential N-glycosylation sites at positions Asn^{210}, Asn^{222}, Asn^{286}, Asn^{313} and Asn^{443} (Liu *et al.*, 1998) and it was shown that all of these N-glycosylation sites are utilized and the glycosylation contributes ~10 kDa of the total weight of TPP-I (Golabek *et al.*, 2003; Wujek *et al.*, 2004). Furthermore, it was demonstrated that elimination of the glycosylation site at Asn^{286} dramatically affected the folding and translocation of the enzyme as well as the enzymatic activity, whereas the other glycosylation sites also play a complimentary role in the folding process (Steinfeld *et al.*, 2002, 2004; Tsiakas *et al.* 2004; Wujek *et al.*, 2004). Also the oligosaccharides at Asn^{210} and Asn^{286} are preferentially phosphorylated in TPP-I and phosphorylation of TPP-I at these two positions is needed for efficient sorting of TPP-I in the trans-Golgi compartment (Wujek *et al.*, 2004). Polyanionic glycosaminoglycans not only increase the rates of activation but also allow for activation of the proenzyme at higher pH, to pH 6.0. Mature TPP-I is an acidic protease that is unstable at alkaline and neutral pH. Also these bound glycosaminoglycans improved the thermal stability of TPP-I and protected the enzyme against alkaline pH induced denaturation (Golabek *et al.*, 2005; Wujek *et al.*, 2004).

The TPP-I proenzyme autoactivates at acidic pH (Lin *et al.*, 2001) and the mature enzyme is generated. It was demonstrated that *in vitro*, autoprocessing of TPP-I proenzyme can be carried out at a wide pH range, from ~ 2.0 to 6.0, although with different efficiencies depending on the pH and the type of buffer. However, the

acquisition of enzymatic activity in the same buffer took place in a narrower pH range of 3.6–4.2 (Golabek *et al.*, 2004). CD spectroscopy revealed that autoprocessing was not associated with any major change in the secondary structure of the proenzyme (Golabek *et al.*, 2004). After autoactivation at higher pH (4.5 and 5.0), two major cleavage sites were identified, after Glu^{189} and Ser^{181}, generating mature TPP I with N-terminal extensions containing 6 and 14 amino acid residues, respectively. At pH 5.0, the yields of both polypeptides were nearly equimolar, whereas at pH 4.5, the polypeptide starting at Val^{190} predominated. In addition, a minor polypeptide with molecular mass of ~60 kDa, thus of intermediate size between pro-TPP I and the mature enzyme that was generated at pH 4.5 and 5.0 had an N-terminal sequence starting at Thr^{98}. These data suggest that the lack of tripeptidyl peptidase activity of mature TPP I generated during pro-TPP I processing *in vitro* at higher pH may be caused by the presence of small N-terminal extensions on enzymes generated in this manner (Golabek *et al.*, 2004).

The active site is comprised of catalytic residues Ser^{475}, Asp^{360}, Glu^{272}, Asp^{276} and Asp^{327} (Walus *et al.*, 2005; Lin *et al.*, 2001). The catalytic triad is made up of Ser^{475}, Asp^{360}, Glu^{272}, in which Ser^{475} acts as the nucleophile and the protonated Asp^{360} is supposed to stabilize the electronegative oxygen atom of nucleophilic Ser^{475} during the formation of the tetrahedral intermediate (Walus *et al.*, 2005; Lin *et al.*, 2001). Also it was proposed that although Asp^{327} is not a catalytic residue of the active site, it plays an important role in the catalytic mechanism of TPP-I, which may include hydrogen bonding with the substrate or orienting another catalytic group in the active site (Walus *et al.*, 2005). It was also demonstrated that the catalytic residues of TPP-I are topologically similar to that of other members of the sedolisins family such as Pepstatin insensitive serine-carboxyl protease (PSCP), Xanthomonas-serine carboxyl protease (XSCP) and Kumamolysin-serine carboxyl

protease (KSCP), with TPP-I as the first mammalian member of this family (Walus *et al.*, 2005, Comellas-Bigler *et al.*, 2002).

It was shown that mammalian enzymes homologous to Human TPP-I, from macaque, dog, mouse, rat and cow form a family of sedolisins with highly conserved sequences (Wlodawer *et al.*, 2003). The overall sequence identity for these enzymes was reported to be 81% and similarity to be 92%. Comparison between human and mouse enzymes yielded 88% identity and 94% similarity (Wlodawer *et al.*, 2003).

A brief overview of TPP-I sequence is schematized in Figure 4.12.

Figure 4.13 shows the sequence alignment and comparison of mammalian TPP-I like enzymes from Human, mouse, rat, chimpanzee and dog.

1   **20**

MGLQACLLGLFALILSGKCSYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQ

▲   **99**

NVERLSELVQAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAGAQKCH

▲

SVITQDFLTCWLSIRQAELLLPGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVD

182   **190**   **196**   **210**   **222**

FVGGLHRFPPTSSLRQRPEPQVTGTVGLHLGVTPSVIRKRYNLTSQDVGSGTSNN

▲   ▲   ▲   **272**   **276**

SQACAQFLEQYFHDSDLAQFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDV

286   **313**   **327**

QYLMSAGANISTWVYSSPGRHEGQEPFLQWLMLLSNESALPHVHTVSYGDDEDSL

360

SSAYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSPYVTT

443

VGGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLSSSPHLPPSSYFNA

475

SGRAYPDVAALSDGYWVVSNRVPIPWVSGTSASTPVFGGILSLINEHRILSGRPPLG

FLNPRLYQQHGAGLFDVTRGCHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFP

563

ALLKTLLNP

Figure 4.12: Sequence of TPP-I preproenzyme showing the signal peptide and the prodomain. The cleave points between the signal peptide, the prodomain (between residues **Cys[19]** and **Ser[20]**) and the mature enzyme (between residues **Gly[195] and Leu[196]**) are marked by ▲ and ▲ respectively. Also shown are the N-glycosylation sites (residues **Asn[210], Asn[222], Asn[286], Asn[313]** and **Asn[443]**) and the catalytic residues (residues **Ser[475], Asp[360], Glu[272], Asp[276]** and **Asp[327]**).The alternative cleavage points observed at higher pH (between residues His[97] and Thr[98], Ser[181] and Leu[182] and Glu[189] and Val[190]) are indicated by ▲ .

86

Figure 4.13: Sequence comparisons of the mammalian TPP-I preproenzyme. The residues showing N-glycosylation and catalytic residues are marked by **G** and **C** respectively above the residues (which are conserved among selected species). The signal peptide cleavage site is marked by ▲, the main maturation and alternative maturation cleavage sites are marked ▲ and ▲ respectively. The conserved residues are shown in red boxes and the consensus sequence is also shown.

4.2   Materials and Methods

4.2.1 Purification

The recombinant protein (native and selenomethionine derivative TPP-I) was overexpressed in human embryonic kidney 293 cells and purified from the cell culture supernatant using nickel affinity chromatography by the group of PD Dr. Dr. Robert Steinfeld, Universitäts Medizin Göttingen.

4.2.2 Crystallization

Crystallization was carried by the hanging drop vapor diffusion method (Hampel *et al.*, 1968; McPherson, 1992) by equilibrating a drop containing a mixture of 2µl of 10mg/ml protein and 2 µl of reservoir containing 7% PEG4000, 0.02 M Zinc sulfate, 0.1 M Sodium acetate pH 5.0 and 0.1 M Ammonium sulfate. The final pH of the reservoir solution was 4.9. Block-like crystals (Figure 4.14) with good diffraction quality appeared after seven days. Both the native and selenomethionine derivative crystals crystallized in the same reservoir conditions. All solutions used for crystallization were sterile filtered and contained 0.03% Sodium azide.
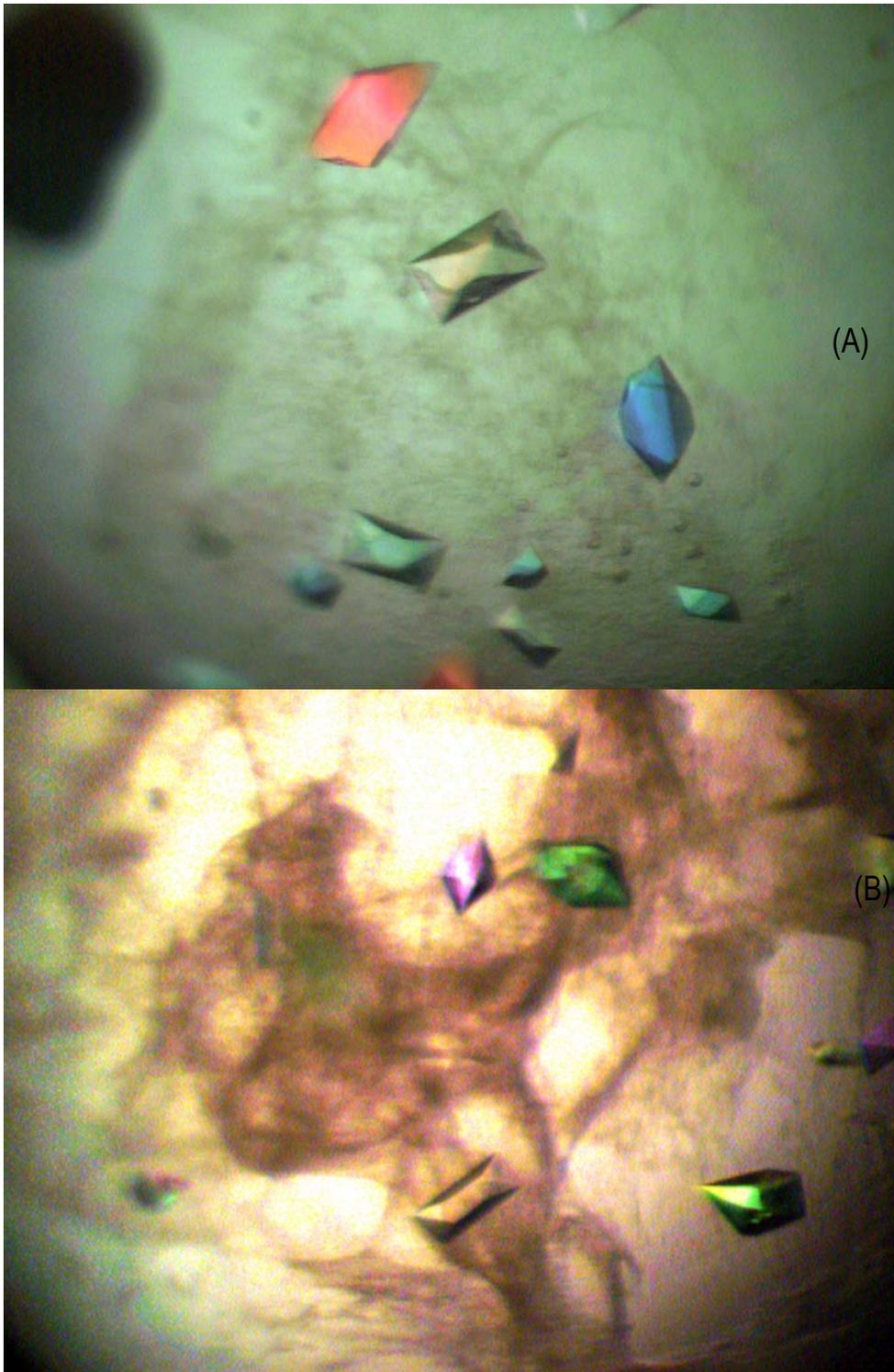
Figure 4.14: (A) Crystals of native TPP-I (B) Crystals of selenomethionine derivative of TPP-I.

4.2.3 Data Collection

The crystals (both native and selenomethionine derivative) were flash frozen using 25% glycerol and 5% MPD (in steps of 5% increasing gradient) using Mitegen loops. A native dataset to 2.35 Å was measured at the PX2 beamline at the Swiss Light Source. 360 images (Figure 4.15) with an oscillation width of 1° were collected on a MAR225 mosaic CCD detector and processed with *XDS* (Kabsch, 1993). The selenomethionine derivative datasets were collected on a MAR165 CCD detector at the beamline 14.2 at the BESSY, Berlin. Three different selenomethionine derivative crystals were used to collect data, to increase the redundancy of the derivative data. 360 images with an oscillation width of 1° were collected for the peak wavelength (Figure 4.16A), as well as for the inflection point wavelength (Figure 4.16B), for each of the first two crystals. 180 images each for peak and inflection point at an oscillation of 1° were collected for the third crystal. Crystals diffracted to 2.9 Å, 2.7 Å and 2.7 Å for the peak and inflection point wavelengths and processed with *XDS* (Kabsch, W, 1993). The peak wavelength data from the three crystals were then merged using *XPREP* and same procedure was followed for the inflection point data. The data collection statistics for the peak wavelength and inflection point wavelength are summarized in table 4.1 and 4.2 respectively. The merged data statistics for the peak and inflection point data are summarized in table 4.3 along with the data statistics for the native data.
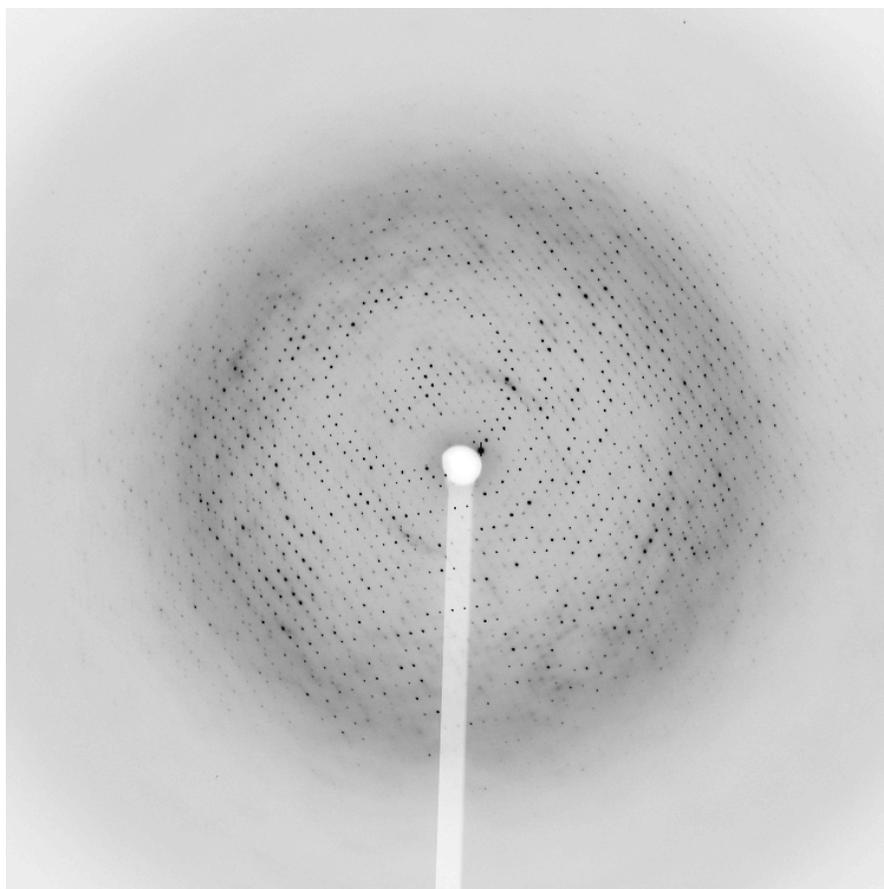
Figure 4.15: Diffraction image of the native crystal of TPP-I at SLS-PX2 beamline.

(A)   (B)
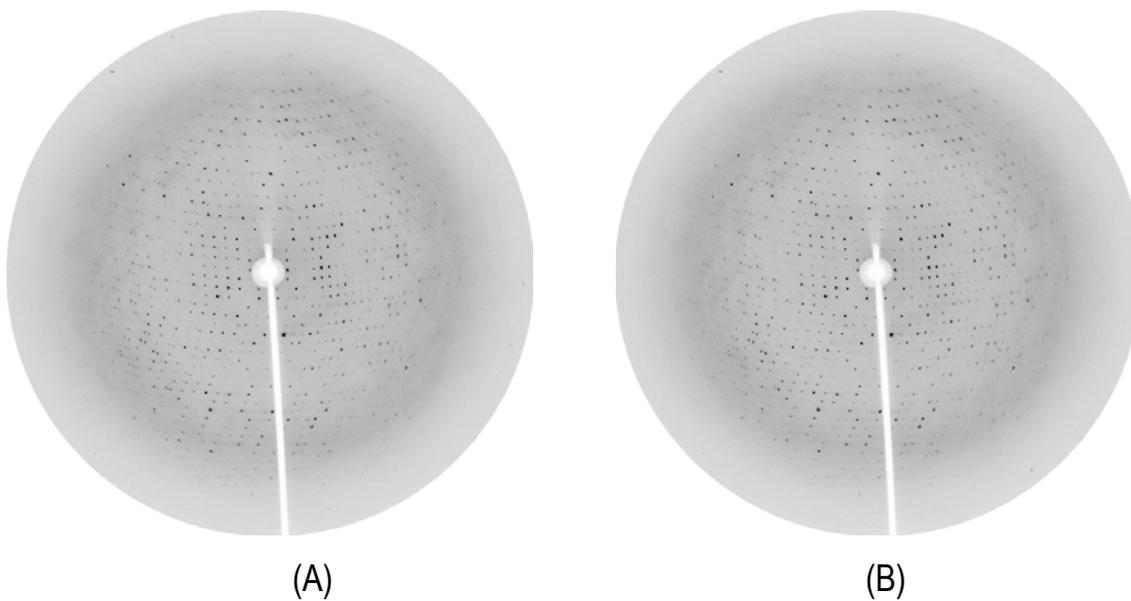
Figure 4.16: Diffraction images of the selenomethionine derivate crystals at (A) peak wavelength and (B) inflection point wavelength collected at BESSY-14.2 beamline.

Table 4.1: Data collection statistics of TPP-I at BESSY-14.2 beamline at peak wavelength for three different crystals coded as PK1, PK2 and PK3. Values in parenthesis are for the outer resolution shell.

Data Statistics	PK1	PK2	PK3
Wavelength (Å)	0.97971	0.97971	0.97971
X-ray source	BESSY-14.2	BESSY-14.2	BESSY-14.2
Detector	MARCCD	MARCCD	MARCCD
Space group	P 2_12_12	P 2_12_12	P 2_12_12
a (Å)	100.44	100.43	100.30
b (Å)	113.34	113.26	113.18
c (Å)	128.65	128.40	128.17
Resolution (Å)	2.99 (3.09-2.99)	2.70 (2.79-2.70)	2.69 (2.79-2.69)
Unique Reflections	30021	40634	40664
Redundancy	14.55 (13.53)	14.62 (13.48)	7.28 (6.58)
Completeness (%)	99.0 (94.5)	99.0 (93.5)	99.0 (93.0)
Mean I/σ (I)	23.7 (6.86)	22.46 (4.47)	12.27 (2.42)
R_{int} (%)	9.18 (35.80)	8.77 (55.51)	11.75 (66.30)

Table 4.2: Data collection statistics of TPP-I at BESSY-14.2 beamline at peak wavelength for three different crystals coded as IP1, IP2 and IP3. Values in parenthesis are for the outer resolution shell.

Data Statistics	IP1	IP2	IP3
Wavelength (Å)	0.97987	0.97987	0.97987
X-ray source	BESSY-14.2	BESSY-14.2	BESSY-14.2
Detector	MAR225CCD	MAR165CCD	MAR165CCD
Space group	P 2_12_12	P 2_12_12	P 2_12_12
a (Å)	100.50	100.35	100.33
b (Å)	113.38	113.11	113.20
c (Å)	128.70	128.27	128.20
Resolution (Å)	2.98(3.08-2.98)	2.69(2.79-2.69)	2.70(2.79-2.70)
Unique Reflections	30112	40733	40667
Redundancy	14.5 (12.78)	14.62 (13.40)	7.28 (6.64)
Completeness (%)	98.7 (91.0)	98.9 (92.9)	99.0 (93.4)
Mean I/σ (I)	25.53 (5.95)	25.16 (4.42)	14.05 (2.40)
R_{int} (%)	9.06 (40.72)	8.35 (56.53)	11.04 (66.23)

Table 4.3: Data collection statistics for TPP-I for native, peak and inflection point data. The peak data here were generated by merging datasets PK1, PK2 and PK3, and inflection point data was also generated by merging datasets IP1, IP2 and IP3. Values in parenthesis are for the outer resolution shell.

Data Statistics	Native data	Peak data	Inflection data
Wavelength (Å)	1.00000	0.97971	0.97987
X-ray source	SLS-PX2	BESSY-14.2	BESSY-14.2
Detector	MARCCD	MARCCD	MARCCD
Space group	P 2_12_12	P 2_12_12	P 2_12_12
a (Å)	100.50	100.43	100.35
b (Å)	113.45	113.26	113.11
c (Å)	128.93	128.40	128.27
Resolution (Å)	2.35 (2.45-2.35)	2.69(2.79-2.69)	2.69(2.79-2.69)
Unique Reflections	61829	40695	40750
Redundancy	7.18 (6.11)	32.62 (19.89)	32.63 (19.92)
Completeness (%)	99.5 (97.2)	99.0 (93.6)	99 (93.1)
Mean I/σ (I)	17.20 (3.15)	31.19 (5.04)	34.49 (4.99)
R_{int} (%)	5.91 (41.55)	11.34 (61.89)	11.07 (62.51)

4.2.4 Structure Solution and Phasing

XPREP was used to prepare the peak and the selenomethionine derivative datasets for structure solution. The correlation coefficient between the peak and the inflection datasets suggested a cutoff margin between 3.5 - 3.7 Å, taking into consideration that data should be truncated at the resolution where the correlation coefficient falls below 30% (Schneider & Sheldrick, 2002):

Resolution: Inf - 8.0 - 6.0 - 5.0 - 4.3 - 4.1 - 3.9 - 3.7 - 3.5 - 3.3 - 3.1 - 2.9 - 2.7 Å

Correlation coefficient: 93.0 85.6 76.1 61.4 45.5 37.5 30.6 20.7 13.9   5.7   4.4   3.7

For the peak dataset, the anomalous signal/noise ratios against the resolution are as follows (Schneider & Sheldrick, 2002).The first line is based on resolution, the second line on input sigma values and the third line on variances of F+ and F- :

Inf - 8.0 - 6.0 - 5.0 - 4.3 - 4.1 - 3.9 - 3.7 - 3.5 - 3.3 - 3.1 - 2.9 - 2.7 Å

```
 4.97  4.68  3.61  2.70  2.17  2.04  1.80  1.64  1.59  1.45  1.37  1.31
 5.22  4.01  3.12  2.33  1.81  1.67  1.44  1.32  1.29  1.25  1.24  1.19
```

The structure was solved using *SHELXD* (Sheldrick *et al.*, 2001) using the following settings:

```
SHEL 999   3.50
PATS
FIND   8
MIND -3.5   2.2
NTRY 1000
```

Eight heavy atom peaks were found using *SHELXD*, corresponding to the eight expected selenomethionine positions, indicating a presence of two immature

molecules in the asymmetric unit and a solvent content of ~62%, if the ~10kDa glycosylation is not taken into account and ~55% solvent, taking into account the ~10kDa glycosylation. Figure 4.17 shows the selenomethionine positions in one TPP-I catalytic domain.

CC 41.74   CC (weak) 32.81   PATFOM 11.44

	X	Y	Z	Peak height
Se01	0.165573	0.678139	0.328315	1.0000
Se02	0.279572	0.410950	0.430053	0.9900
Se03	0.264069	0.549255	0.065330	0.9676
Se04	0.308876	0.282867	0.511985	0.9596
Se05	0.259911	0.680794	-0.015946	0.9030
Se06	0.288963	0.765915	-0.170853	0.8929
Se07	0.290443	0.795624	-0.226937	0.8582
Se08	0.169289	0.648232	0.268798	0.7891

With these atoms as input and using the high resolution native dataset from SLS for phase extension, *SHELXE* (Sheldrick, 2002), gave a good discrimination between the two heavy atom enantiomorphs using the contrast and connectivity figures of merit. A value of about 70% for the pseudo-free correlation coefficient indicates an interpretable map and is a good indication as to when phase refinement has converged (Sheldrick, 2002). Examples of initial and current maps are shown in Figure 4.18.

Mean weight and estimated mapCC as a function of resolution for original map

```
 d    inf - 5.22 - 4.14 - 3.61 - 3.28 - 3.04 - 2.86 - 2.71 - 2.59 - 2.49 - 2.40
 <wt>    0.651 0.720 0.722 0.719 0.702 0.676 0.675 0.668 0.671 0.644
 <mapCC> 0.808 0.875 0.893 0.907 0.903 0.880 0.866 0.854 0.846 0.817
d is resolution in Å, wt is weight, mapCC is map co-relation coefficient
 Pseudo-free CC = 71.17 %
```

Mean weight and estimated mapCC as a function of resolution for inverted map

```
 d    inf - 5.22 - 4.14 - 3.61 - 3.28 - 3.04 - 2.86 - 2.71 - 2.59 - 2.49 - 2.40
 <wt>    0.535 0.621 0.648 0.684 0.677 0.639 0.633 0.636 0.642 0.601
 <mapCC> 0.721 0.818 0.852 0.897 0.897 0.856 0.837 0.842 0.833 0.790
 d is resolution in Å, wt is weight, mapCC is map co-relation coefficient
 Pseudo-free CC = 65.99 %
```

(A)   (B)
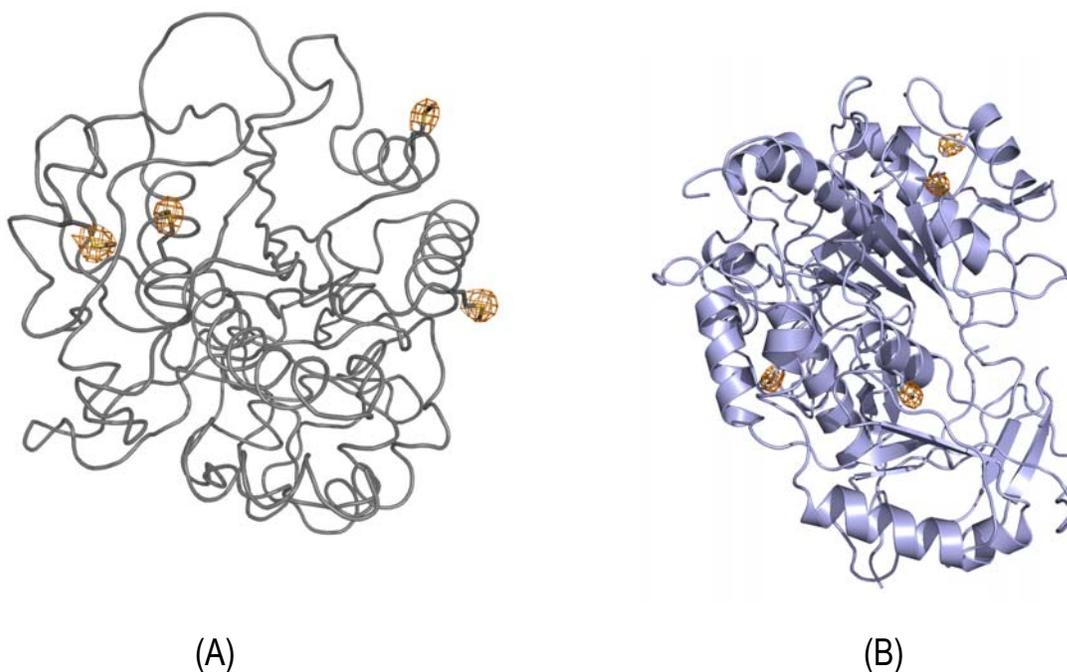
Figure 4.17: View of the (A) backbone representation and (B) domain representation of TPP-I showing Selenomethionine residues superimposed on the anomalous Fourier map calculated with *SHELXE* phases at 3.3 Å.

(A)   (B)
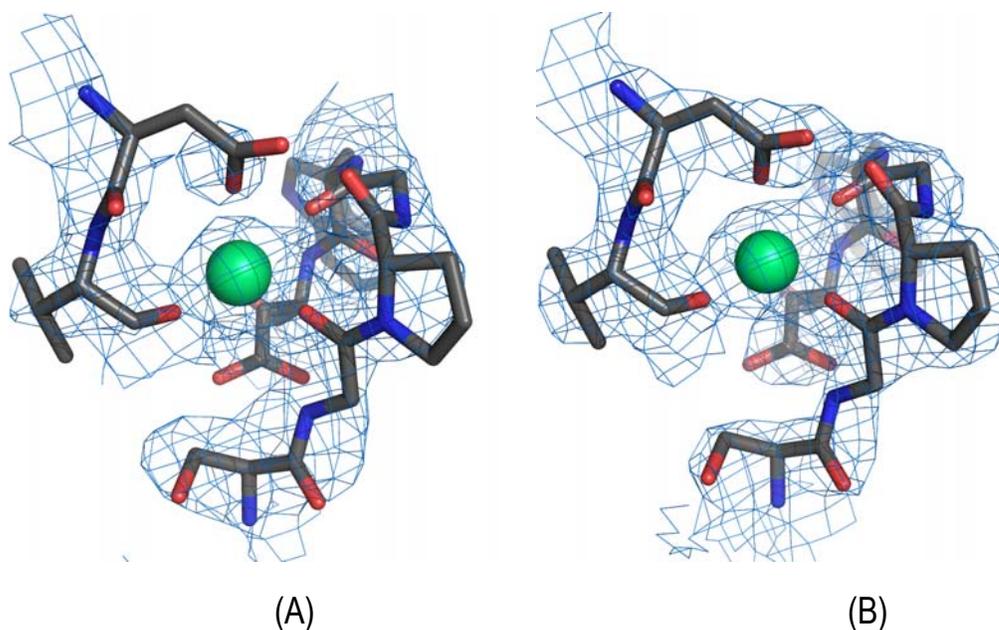
Figure 4.18: Examples of electron density maps superimposed on the TPP-I calcium binding region: (A) initial map produced by *SHELXE* after phase extension to the native data at 2.35 Å and (B) map produced by the current refined model. The map is contoured at 1.2σ.

4.2.5 Model Building and Refinement

Although the value of Pseudo-free CC obtained from *SHELXE* indicated an interpretable map and also maps looked reasonably good, they could not be traced automatically using the existing auto-tracing programs. An alpha test version of *SHELXE* (Sheldrick G.M., 2002) was used to find some helices and using the selenomethionine positions and other bulky side chains as a guideline, the model was built manually using *COOT* (Emsley & Cowtan, 2004). *Refmac_5.2.0019* (Murshudov *et al*., 1997) was used for refinement of the structure.

The current model has four independent peptide chains(two copies each for residues Ser[20] to Ser[180] and residues His[197] to Arg[564]) containing 1054 residues, 108 water molecules, 8 molecules of N-Acetyl-D-Glucosamine, 2 Calcium, 8 Zinc and 2 sulfates. The presence of zinc ions are likely to be unphysiological and a result of the crystallization procedure. Two independent peptide chains (residues Ser[20] to Ser[180] and residues His[197] to Arg[564]) are grouped as one unit. One such unit is discussed as molecule A and the other as B in this thesis for simplicity. The quality of the current model was assessed using the Ramachandran plot (Ramachandran & Sasisekharan, 1968) validation tool using *COOT* (Emsley & Cowtan, 2004) and *MOLPROBITY* (Lovell *et al.*, 2003). The refinement statistics of the current model are summarized in table 4.4.The Ramachandran plot (Figure 4.19A) shows 92.7% and 5.7% residues in the preferred and allowed regions, whereas 1.6% of the total residues are present in the disallowed regions. The Kleywegt plot (Kleywegt and Jones, 1996) shows differences between a few equivalent residues of the two NCS related TPP-I units (Figure 4.19B).

Table 4.4: Refinement statistics of TPP-I.

Refinement Statistics	TPP-I
R-Factor (%)	21.987
Free R-Factor (%)	25.686
R.m.s. Deviations from ideal geometry	
Bond Lengths (Å)	0.008
Bond angles (°)	1.250
No. Of Protein Residues	1054
No. of Sugar Molecules	8
No. of Calcium ions	2
No. of Zinc ions	8
No. of Sulfates	2
No. of Water Molecules	108
Ramachandran Plot (%)	
Residues in preferred regions	92.7
Residues in allowed regions	5.7
Residues in disallowed regions	1.6

(A)

(B)
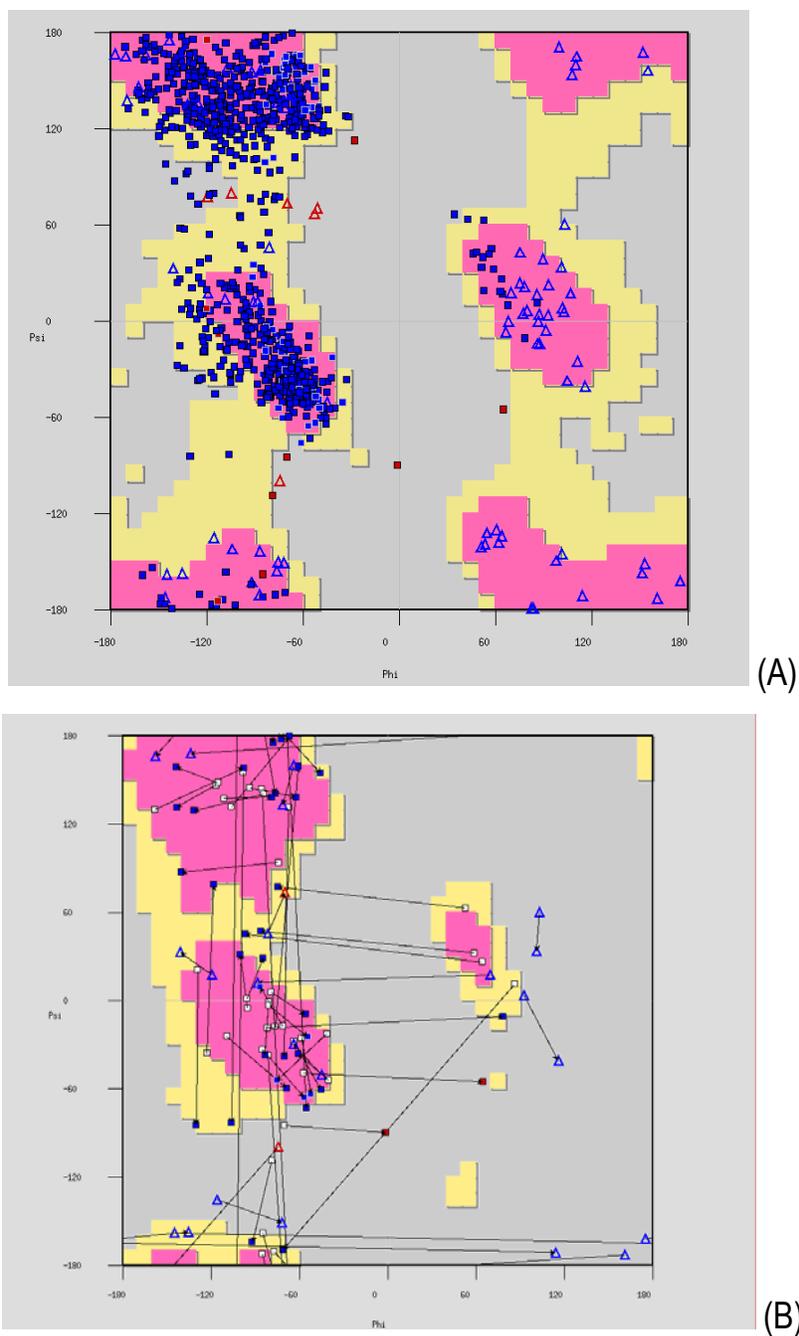
Figure 4.19: (A) Ramachandran plot of TPP-I (Percentage residues in preferred regions: 92.6; allowed regions: 5.6; disallowed regions: 1.6). (B) Kleywegt plot of TPP-I; points representing equivalent residues in the NCS related molecules are represented by lines, showing the differences between torsion angles of equivalent residues.

4.3 Structure Description

4.3.1 Overall Structure and Folding

Two similar units of TPP-I, related by non-crystallographic symmetry were observed in the asymmetric unit (Figure 4.32). Each unit is a complex of two chains (Figure 4.20): the first chain constituted by the prodomain and the second chain containing the catalytic domain. The residues Ser^{20} to Ser^{180} form the prodomain and residues His^{197} to Arg^{564} form the catalytic domain for unit A as well as unit B. TPP-I is globular and has an overall fold similar to that of subtilisin and other similar molecules of the S53 family of sedolisins such as PSCP and KSCP.
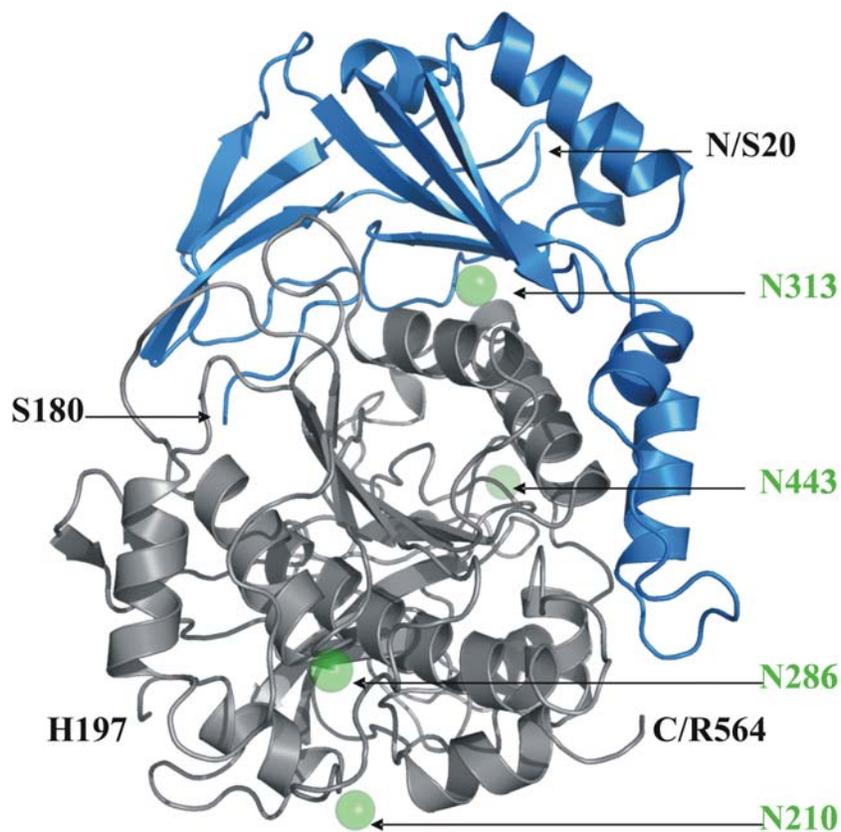
Figure 4.20: Single TPP-I unit exists as a complex of the prodomain (blue) and catalytic domain (grey). Glycosylation sites in green and the termini in **black**.

101

4.3.2 Prodomain

The prodomain spans residues Ser[20] to Ser[180] and has a calculated molecular mass of ~17 kDa as also observed by SDS PAGE shown in Figure 4.24. The prodomain (Figure 4.21) has four helices formed by residues Val[58] to Ser[69] (H1), Leu[82] to Val[89] (H2), Pro[93] to Leu[103] (H3), and Ile[126] to Leu[132] (H4). Leu[47] to Ala[52] (S2), Lys[110] to Val[114] (S3) and Phe[119] to Leu[124] (S4), form one set of antiparallel ß-strands, while Val[36] to Arg[40] (S1), Phe[138] to Val[142] (S5) and Pro[145] to Val[150] (S6) constitute the other set of antiparallel ß- strands. The rest of the residues form loops and turns. Also a disulfide bond is observed between residues Cys[111] and Cys[122].
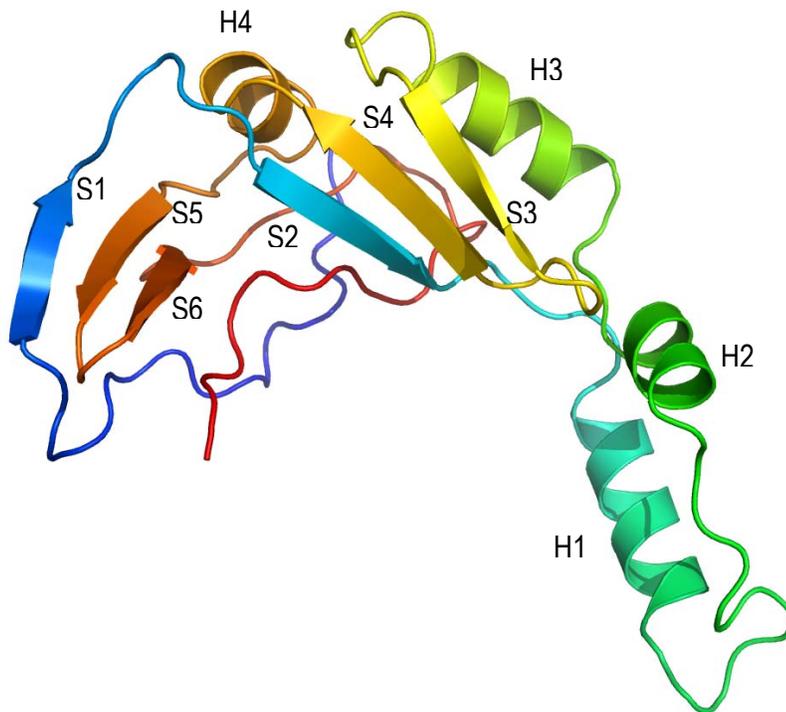
Figure 4.21: The prodomain of TPP-I consists of four helices H1, H2, H3 and H4. Also there is a pair of three antiparallel ß- strands: S2 – S4 – S3 forms one pair while S1 – S5 – S6 forms the other pair.

4.3.3 Catalytic Domain

The catalytic domain contains residues His^{197} to Arg^{564} and a calculated molecular mass of 29.5 kDa for the amino acid residues and taking the glycosylation into account the molecular weight is ~48kDa as observed by SDS PAGE shown in Figure 4.24. The catalytic domain (Figure 4.22) contains ten helices and 18 ß- strands. The helices are formed by residues Pro^{202} to Tyr^{209} (H5), Ser^{238} to Leu^{246} (H6), Ile^{271} to Ala^{283} (H7), Pro^{303} to Ser^{312} (H8), Glu^{329} to Ser^{331} (H9), Ser^{334} to Arg^{350} (H10), Ser^{420} to Ser^{432} (H11), Thr^{474} to Leu^{494} (H12), Asn^{504} to Gln^{510} (H13) and Phe^{553} to Leu^{561} (H14). A set of seven all- parallel ß-strands is formed by residues Ser^{224} to Gln^{229} (S7), Arg^{259} to Val^{261} (S8), Thr^{289} to Ser^{293} (S9), Val^{320} to Ser^{324} (S10), Thr^{353} to Ala^{357} (S11) Thr^{386} to Phe^{393} (S14) and Val^{452} to Leu^{455} (S16). Gly^{364} to Ser^{367} (S12) and His^{372} to Arg^{375} (S13) form antiparallel ß- strands, whereas Trp^{460} to Asn^{464} (S17) and Val^{466} to Trp^{470} (S18) are also antiparallel to each other. Residues Gln^{402} to Val^{404} (S18) form a strand that is antiparallel to strand S14. There are two disulfide linkages, the first from residue Cys^{365} to Cys^{526} and the second from Cys^{522} to Cys^{537}. The seven-stranded all-parallel ß-sheets are flanked by four helices H7, H8, H10 and H12 at four corners.

The prodomain and catalytic domain of each unit interacts with each other through residues Arg^{28}-Asp^{328}, Arg^{28}-Asp^{330}, Thr^{50}-Gln^{306}, Gln^{55}-Thr^{342}, Asp^{118}-Lys^{346}, Phe^{119}-Met^{309}, Thr^{121}-Gln^{306} and Phe^{169}-Tyr^{336}. The overall secondary structure topology is schematized in Figure 4.23.
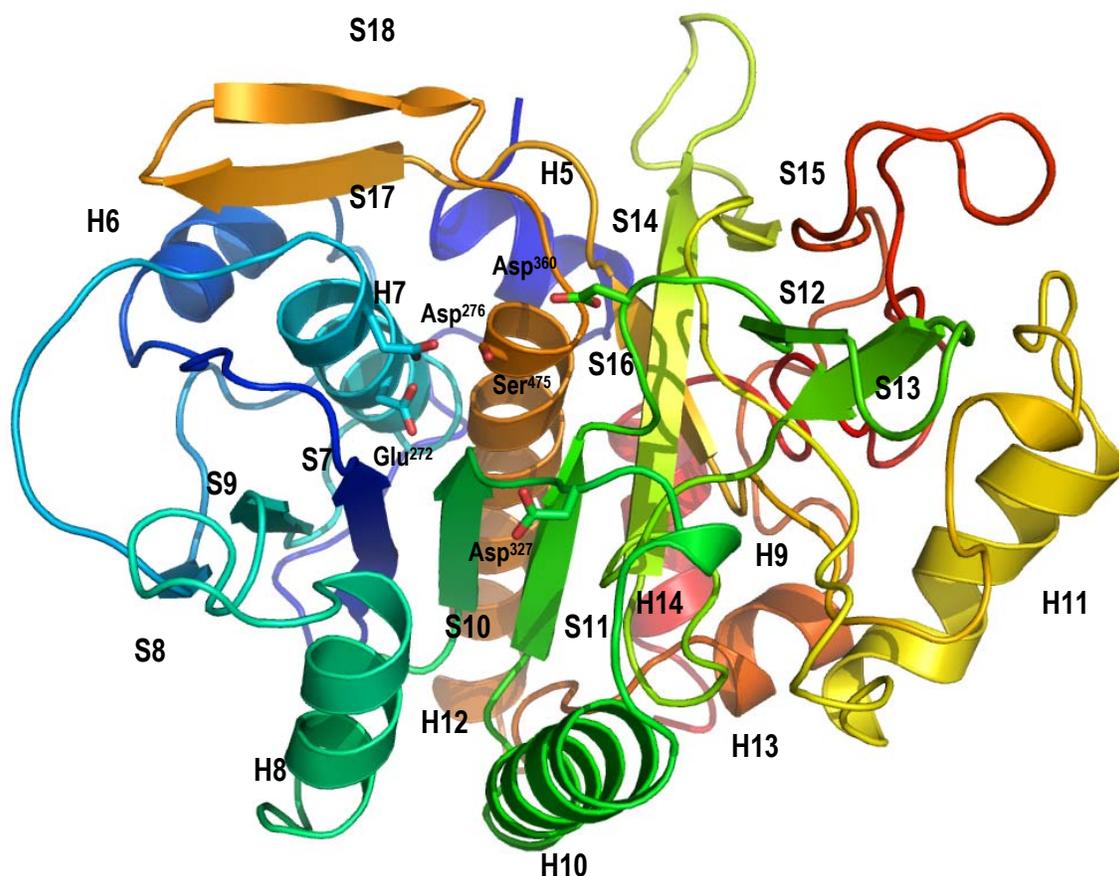
Figure 4.22: Overall fold of the catalytic domain of TPP-I showing the helices, β-strands and side chains of catalytic residues Ser[475], Asp[360] and Glu[272] along with residues Asp[276] and Asp[327]. A set of seven all- parallel ß-strands S8-S9-S7-S10-S11-S14-S16 is surrounded by helices H7, H12, H8 and H10 and form the infrastructure for the active site.

|20 |30 |40 |50 |60
SYSPEPDQRRTLPPGWVSLGRADPEEELSLTFALRQQNVERLSELV

S1   S2   H1

|70 |80 |90 |100 |110 |120 |130
QAVSDPSSPQYGKYLTLENVADLVRPSPLTLHTVQKWLLAAGAQKCHSVITQDFLTCWLSIRQAEL

H1   H2   H3   S3   S4   H4

|140 |150 |160 |170 |180
LLPGAEFHHYVGGPTETHVVRSPHPYQLPQALAPHVDFVGGLHRFPPTS

S5   S6

Prodomain   **Catalytic Domain**

|200 |210 |220 |230 |240 |250
HLGVTPSVIRKRYNLTSQDVGSGTSNNSQACAQFLEQYFHDSDLAQFMRLFGGNFAHQASVAR

H5   S7   H6   S8

|260 |270 |280 |290 |300 |310 |320
VVGQQGRGRAGIEASLDVQYLMSAGANISTWVYSSPGRHEGQEPFLQWLMLLSNESALPHVHTV

S8   H7   S9   H8   S10

|330 |340 |350 |360 |370 |380
SYGDDEDSLSSAYIQRVNTELMKAAARGLTLLFASGDSGAGCWSVSGRHQFRPTFPASSPYVTTV

H9   H10   S11   S12   S13   S14

|390 |400 |410 |420 |430 |440 |450
GGTSFQEPFLITNEIVDYISGGGFSNVFPRPSYQEEAVTKFLSSSPHLPPSSYFNASGRAYPDVAAL

S14   S15   H11   S16

|460 |470 |480 |490 |500 |510 |520
SDGYWVVSNRVPIPWVSGTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDVTRG

S17   S18   H12   H13

|530 |540 |550 |560
CHESCLDEEVEGQGFCSGPGWDPVTGWGTPNFPALLKTLLNPR

H14

Figure 4.23: Topology diagram of the structure of TPP-I showing the prodomain and the catalytic domain. Helices (**H-**) are represented by cylinders and sheets (**S-**) are represented by arrows. Catalytic residues, glycosylated residues and disulfide forming cysteine residues are marked as **D**, **N** and **C** respectively.

4.3.4 Linker Region

The residues Ser181 to Leu196 could not be modeled as no residual density is observed for these residues. The crystallographic data do not provide enough crystallographic evidence either supporting or disproving the presence of both the mature and the immature forms of TPP-I in the crystal. However, SDS PAGE analysis of the crystals (Fig 4.24) supports the presence of the mature and the immature forms. Consistent with the literature (Golabek *et al.*, 2004, 2008), the appearance of crystals after 7 days at pH 4.8-5.0 might also support the observation that the crystals indeed contain both immature and mature TPP-I, in which the prodomain binds tightly to the catalytic domain even after activation. Autoactivation at higher pH (>4.5) results in three atypical cleavage sites (apart from the usual site after Gly195), namely after Thr98, Glu189 and Ser181 (Golabek *et al.*, 2004). It has also been recently demonstrated that at pH of >4.5, a reversible tight complex is formed between the prodomain and catalytic domain of TPP- I , and its stability is a consequence of a slow dissociation rate of the prosegment held to the enzyme by strong ionic interactions (Golabek *et al.*, 2008), as also observed in the current crystal structure. Therefore it can be proposed that the current model contains both the mature and the immature forms. The inability to model residues Ser181 to Leu196 is contributed to by two reasons: firstly the variation of the cleavage sites at higher pH values after residues Ser181, Glu189 and Gly195 in different mature TPP-I molecules in the crystal and secondly the putative presence of these residues in a disordered state in some of the immature TPP-I molecules in the crystal. However, no other disorder was found in any of the regions near the active site. This explanation may emphasize that autoprocessing is not associated with any major change in the secondary structure of the immature TPP-I as also revealed by CD spectroscopy (Golabek *et al.*, 2004).
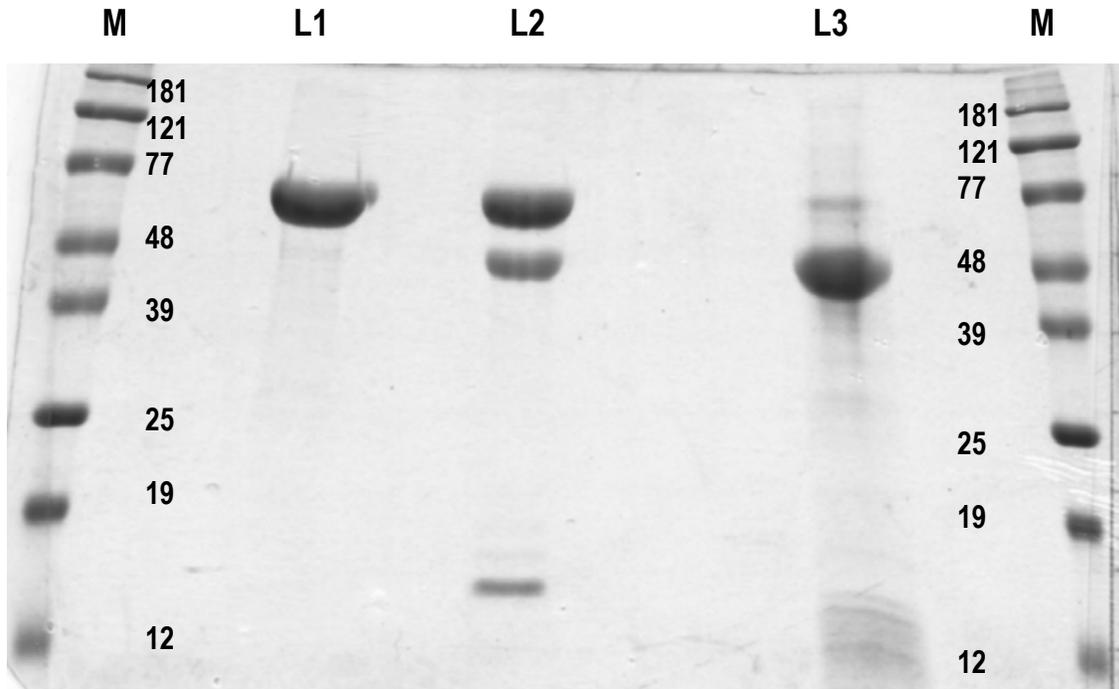
Figure 4.24: SDS PAGE gel of TPP-I. The first and the last lanes (M) from top to bottom show the molecular weight markers (181, 121, 77, 48, 39, 25, 19 and 12 kDa). Lane L1 corresponds to the immature form, Lane L2 corresponds to the crystals and Lane L3 shows the mature form. Lane L2 shows that both the mature and immature forms are present along with the prodomain (residues Ser[20] to His[197]) in the crystals of TPP-I

4.3.5 Catalytic Residues and Catalysis

TPP-I has the characteristic sedolisin features such as the Ser^{475}-Glu^{272}-Asp^{360} catalytic triad. The catalytic residues Ser^{475}, Asp^{360} and Glu^{272} along with residues Asp^{276} and Asp^{327} are shown in Figure 4.26. The Ser^{475} residue resides on helix H12, whereas residue Glu^{272} along with Asp^{276} lies on helix H7 (Figure 4.27). The Asp^{360} residue lies on a loop connecting the strands S11 and S12, while Asp^{327} is observed on the loop connecting helix H9 to strand S10 (Figure 4.27). The Ser^{475} acts as a nucleophile, while Asp^{360} side chain creates part of the oxyanion hole, stabilizing the tetrahedral intermediate of the reaction. Glu^{272} is the general base catalyst. The side chain of Glu^{272} is quiet flexible and supports its multifunctional role during catalysis as proposed for His^{57} in serine proteases. The Asp^{360} would need to be protonated in order to serve its function of creating an oxyanion hole. This might be one of the prominent features responsible for the observation that TPP-I activity is optimal at low pH values (3 - 4.5)

During catalytic process, the Ser^{475} hydroxyl group becomes nucleophilic by the catalytic apparatus, by releasing the O_γ proton. The only possible proton acceptors are the carboxylic oxygen atoms of Glu^{272}. Upon approach of a polypeptide substrate, the scissile peptide bond will be oriented towards Ser^{475} O_γ with its carbonyl group inserting into the oxyanion hole, fixed by subsite interactions of both flanking peptide moieties. This would allow the nucleophilic Ser^{475} to O_γ to attack the polarized carbonyl group of the scissile peptide bond to form a tetrahedral intermediate and to pass the transition state, under simultaneous transfers from the acidic Ser^{475} oxonium cation to the Glu^{272} carboxylic group to Asp^{276}. In the second acylation step, this proton would be transferred back to the leaving group of nitrogen of the substrate allowing the cleavage of the scissile peptide bond, formation of

intermediate acyl ester and dissociation of the C-terminal substrate fragment. Following this the deacylation step would occur in an inverted manner, with a fixed water molecule occupying the position of the leaving nitrogen. This mechanism of TPP-I is derived and schematized based on the other similar members of the sedolisin family (Walus *et al.*, 2005, Comellas-Bigler *et al.*, 2002; Guo *et al.*, 2006) in Figure 4.25.

Also it can be emphasized that in spite of having only 20-30 % overall sequence homology with other sedolisins such as KSCP and PSCP, the active site region is conserved (Figure 4.28 and 4.29), having a fully conserved catalytic triad and by presence of Asp[360] in the oxyanion hole and this indicates a convergent evolution of these proteases.
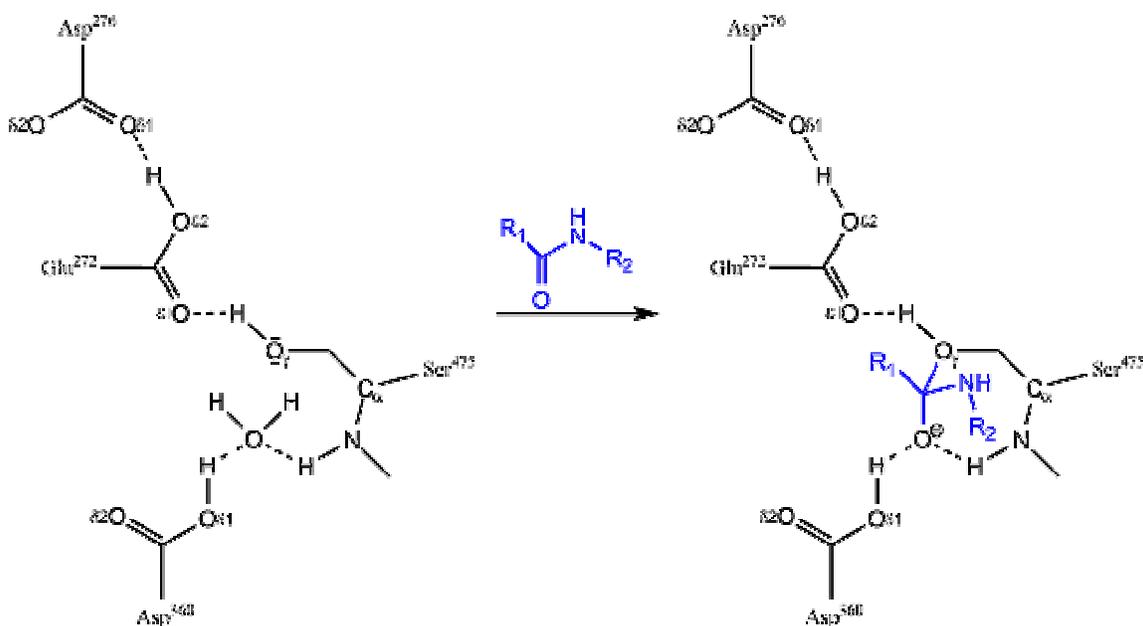

Figure 4.25: Schematic diagram of the suggested process of nucleophilic attack by Ser[475] during the formation of tetrahedral adduct complex in the active site of TPP-I. TPP-I residues are shown in black, while the peptide substrate is colored in **blue**.

(A)

(B)

(C)

(D)
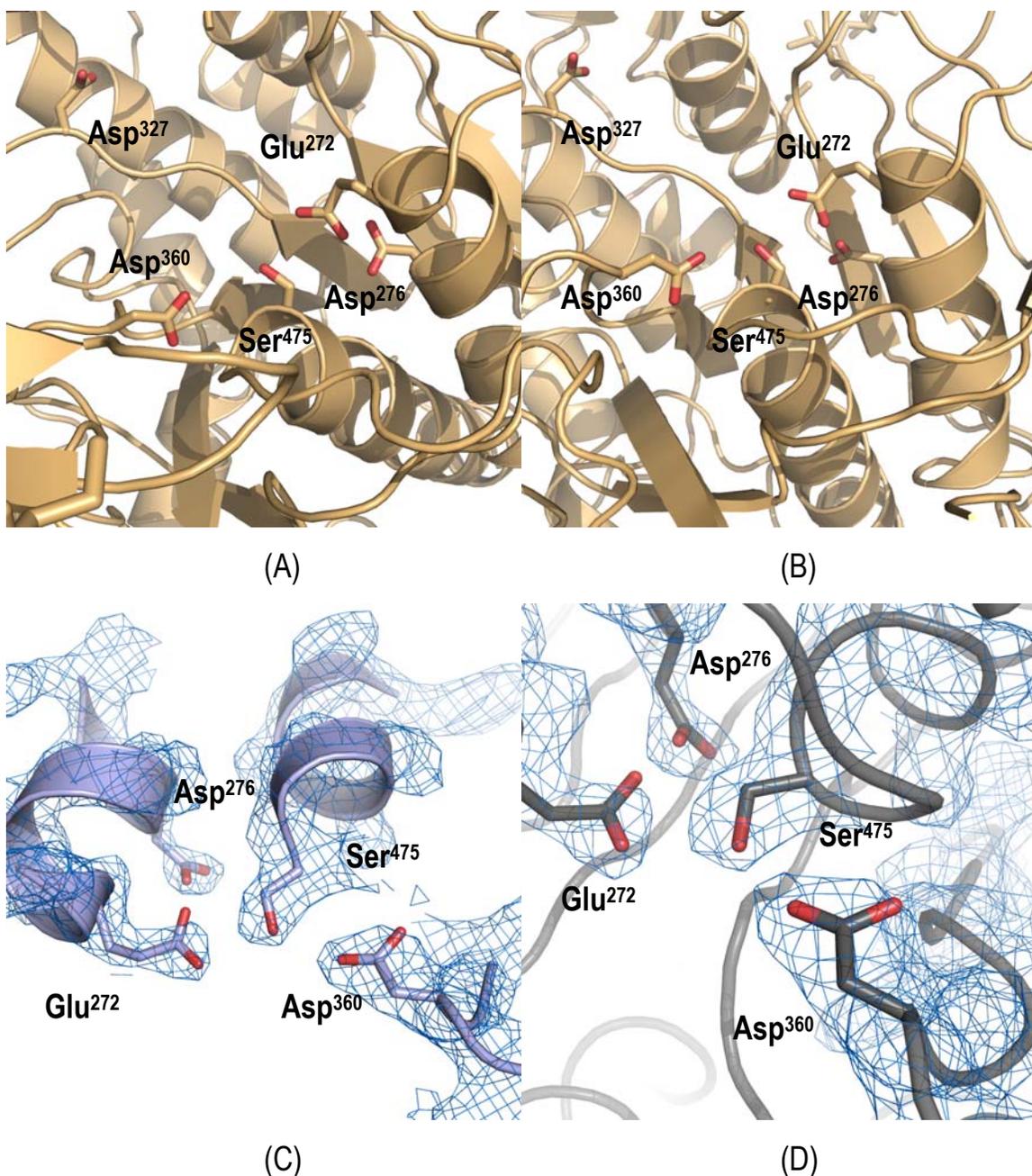
Figure 4.26: Catalytic triad residues Ser[475] - Asp[360] - Glu[272] and supporting residues Asp[276] and Asp[327] for (A) unit A and (B) unit B, shown in different orientations. Asp[327] is located distantly to the active site and is unlikely to play a catalytic role but might be involved in hydrogen bonding to substrate or substrate specificity. (C) and (D) Superposition of the catalytic residues and surrounding chains on the current state electron density map contoured at 1.2σ.
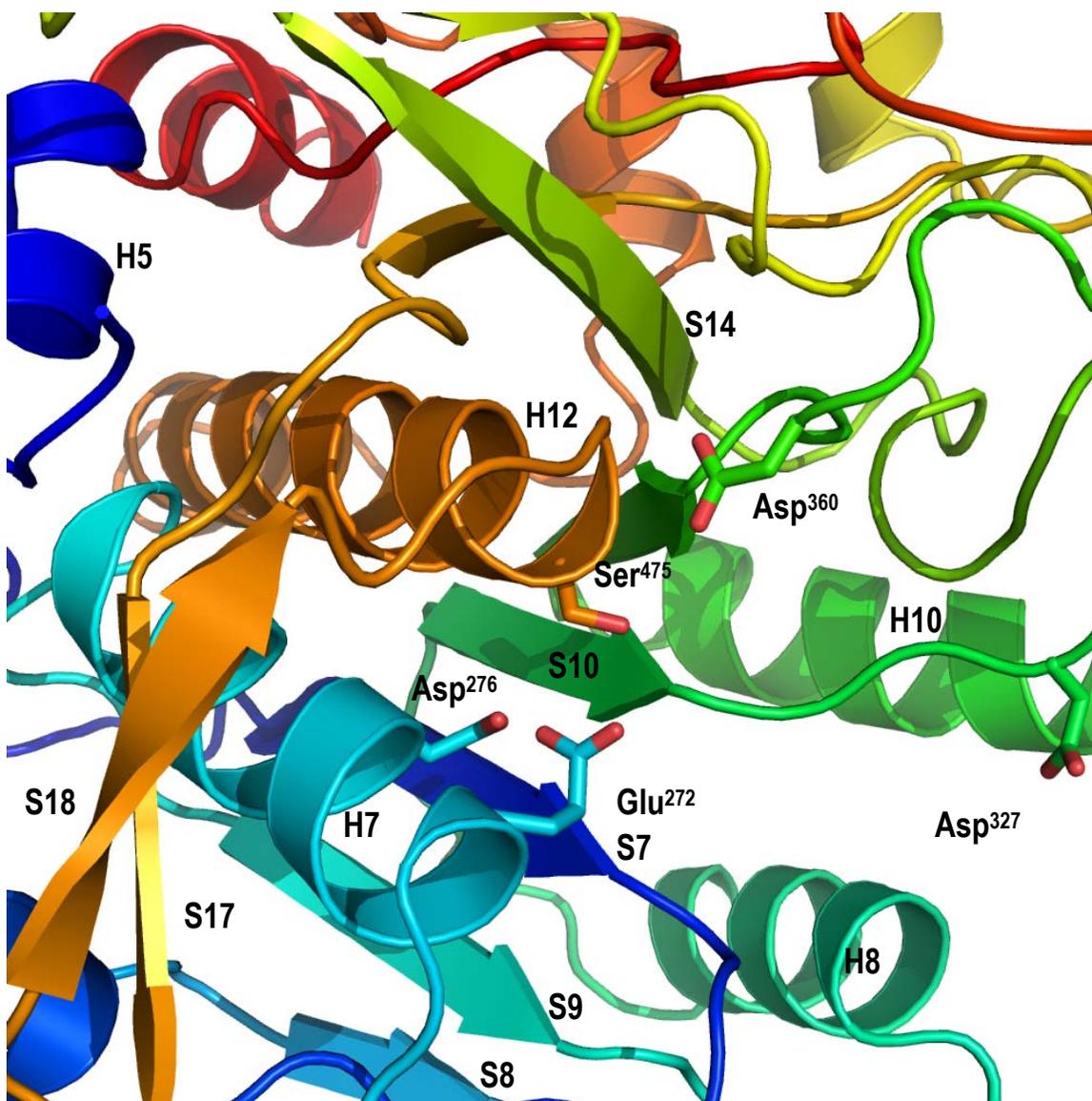
Figure 4.27: The catalytic triad of Ser^{475} - Asp^{360} - Glu^{272} and supporting residues Asp^{276} and Asp^{327}. Ser^{475} is located on helix H12, Glu^{272} and Asp^{276} are residing on helix H7. The residue Asp^{360} lies on a loop connecting the strands S11 and S12, while Asp^{327} is observed on the loop connecting helix H9 to strand ß10.
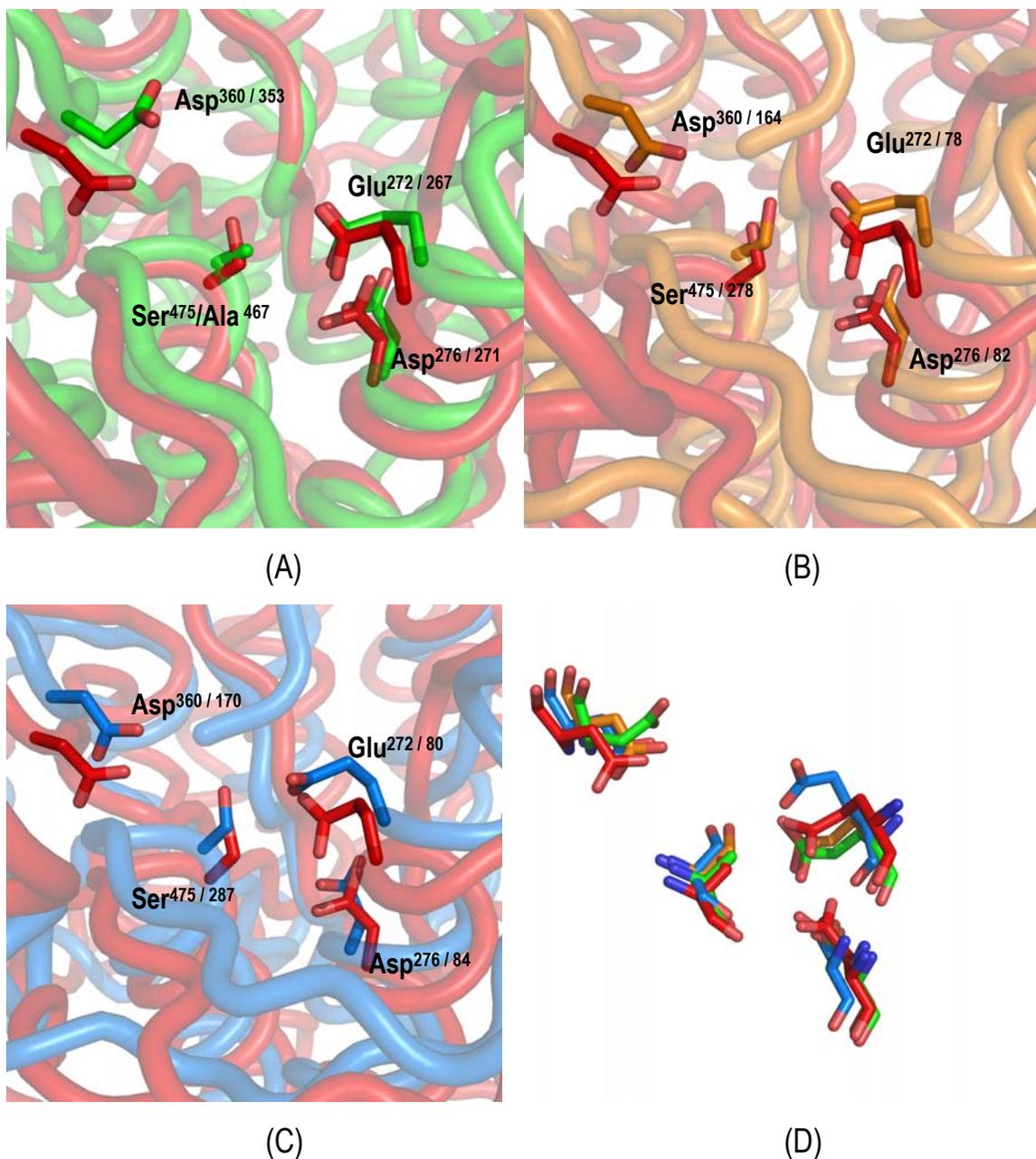
Figure 4.28: Superposition of the active site residues Ser^{475} - Asp^{360} - Glu^{272} and Asp^{276} (**red**) of TPP-I to (A) Pro-Kumamolysin inactive mutant catalytic residues Ala^{467} - Asp^{353} – Glu^{267} and Asp^{271} colored in **green** (PDB id: 1T1E), (B) Mature active Kumamolysin catalytic residues Ser^{278} – Asp^{164} – Glu^{78} and Asp^{82} colored in **orange** (PDB id: 1GTJ), (C) Pseudomonas Serine Carboxyl Protease active site residues Ser^{287} – Asp^{170} – Glu^{80} and Asp^{84} colored in **blue** (PDB id: 1KDV) and (D) superposition of the catalytic residues from all the above with the same color coding.

112

```
TPP-I   DLAQFMRLFGGNFAHQASVARVVGQQGRGRAGIEASLDVQYLMSAGAN--ISTWVYSSPG 296
1KDV    QFTSANGLASVNTQTIQTGSSNGDYSDDQQGQGEWDLDSQSIVGSAGGAVQQLLFYMADQ 106
1T1E    YFASLGVSAPQVVSVSVDGATNQPTGDPNGPDGEVELDIEVAGALAPG--AKIAVYFAPN 290
1GTJ    YFASLGVSAPQVVSVSVDGATNQPTGDPNGPDGEVELDIEVAGALAPG--AKIAVYFAPN 102

TPP-I   RHEGQEPFLQWLMLLSNE-SALPHVHTVSYGDDEDSLSSAYIQRVNTELMKAAARGLTLL 355
1KDV    SASGNTGLTQAFNQAVSDNVAKVINVSLGWC-EADANADGTLQAEDRIFATAAAQGQTFS 165
1T1E    TDAG---FLNAITTAVHDPTHKPSIVSISWGGPEDSWAPASIAAMNRAFLDAAALGVTVL 347
1GTJ    TDAG---FLNAITTAVHDPTHKPSIVSISWGGPEDSWAPASIAAMNRAFLDAAALGVTVL 159

TPP-I   FASGDSGAGCWSVSG---RHQFRPTFPASSPYVTTVGGTSFQEP--FLITNEIVDYIS-- 408
1KDV    VSSGDEGVYECNNRGYPDGSTYSVSWPASSPNVIAVGGTTLYTTSAGAYSNETVWNEGLD 225
1T1E    AAAGDSGSTDGEQDG-----LYHVDFPAASPYVLACGGTRLVAS-AGRIERETVWNDGPD 401
1GTJ    AAAGDSGSTDGEQDG-----LYHVDFPAASPYVLACGGTRLVAS-AGRIERETVWNDGPD 213

TPP-I   --------GGGFSNVFPRPSYQEEAVTKFLSSSPHLPPSSYFNASGRAYPDVAALSDG-- 458
1KDV    SNGKLWATGGGYSVYESKPSWQSVVS--------GTPGR-------RLLPDISFDAAQGT 270
1T1E    GG----STGGGVSRIFPLPSWQERAN--------VPPSANPGAGSGRGVPDVAGNADPAT 449
1GTJ    GG----STGGGVSRIFPLPSWQERAN--------VPPSANPGAGSGRGVPDVAGNADPAT 261

TPP-I   YWVVSNRVPIPWVSGTSASTPVFGGILSLINEHRILSGRPPLGFLNPRLYQQHGAGLFDV 518
1KDV    GALIYNYGQLQQIGGTSLASPIFVGLWARLQSANSN----SLGFPAASFYSAISS-TPSL 325
1T1E    GYEVVIDGETTVIGGTAAVAPLFAALVARINQKLGK----PVGYLNPTLYQLP----PEV 501
1GTJ    GYEVVIDGETTVIGGTSAVAPLFAALVARINQKLGK----PVGYLNPTLYQLP----PEV 313
```

Figure 4.29: Sequence alignment of TPP-I with other reported structures of the sedolisin family, spanning fragments containing the catalytic residues. 1KDV represents pepstatin insensitive carboxyl proteinase, 1T1E represents the intact mutant inactive pro- kumamolysin and 1GTJ represents the active kumamolysin. The catalytic residues are marked in red and underlined.

4.3.6 Glycosylation Sites

Glycosylation is observed at the residues Asn[210], Asn[286], Asn[313] and Asn[443] as shown in Figures 4.20 and 4.30 for glycosylation sites at residues Asn[210] and Asn[443]. The sugars modeled at the glycosylation sites are N-Acetyl-D-Glucosamine. However, no electron density is observed for the glycosylation site at residue Asn[222] in both the units and therefore no sugar moiety was added to that specific residue.


(A)   (B)
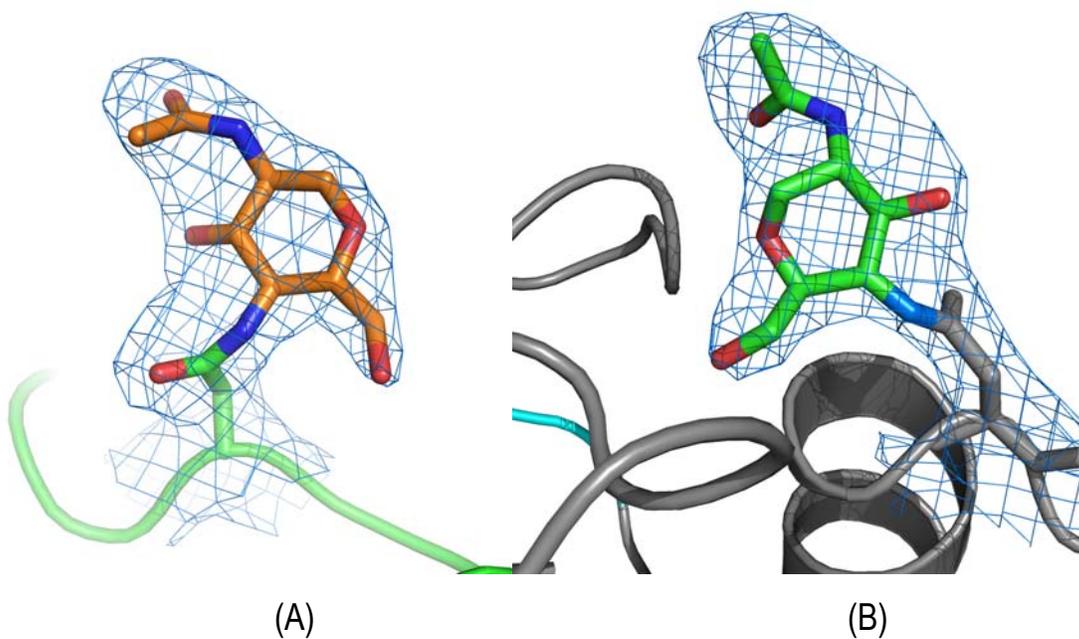
Figure 4.30: Superposition of the model on the electron density map showing the modeled sugars at residues (A) Asn[443] and (B) Asn[210]. The sugar modeled here is N-Acetyl-D-Glucosamine. The maps are contoured at 1.2σ.

4.3.7 Calcium Coordination Site

A calcium ion was observed to coordinate to the side chain of residue Asp[517] in both the units. Furthermore the calcium ion also coordinated to other atoms as visualized in Figure 4.31. This residue (Asp[517]) is proposed to be involved in the folding and stability of TPP-I (Walus *et al.*, 2005). This supports the notion that indeed Asp[517] might play an important role in folding and stability of TPP-I. The calcium ion is coordinated to one carboxylate oxygen of Asp[517] and Asp[543] each and the main chain carbonyl groups of Val[518], Gly[539] and Gly[541]. A water molecule coordinates to the calcium ion completing the coordination.
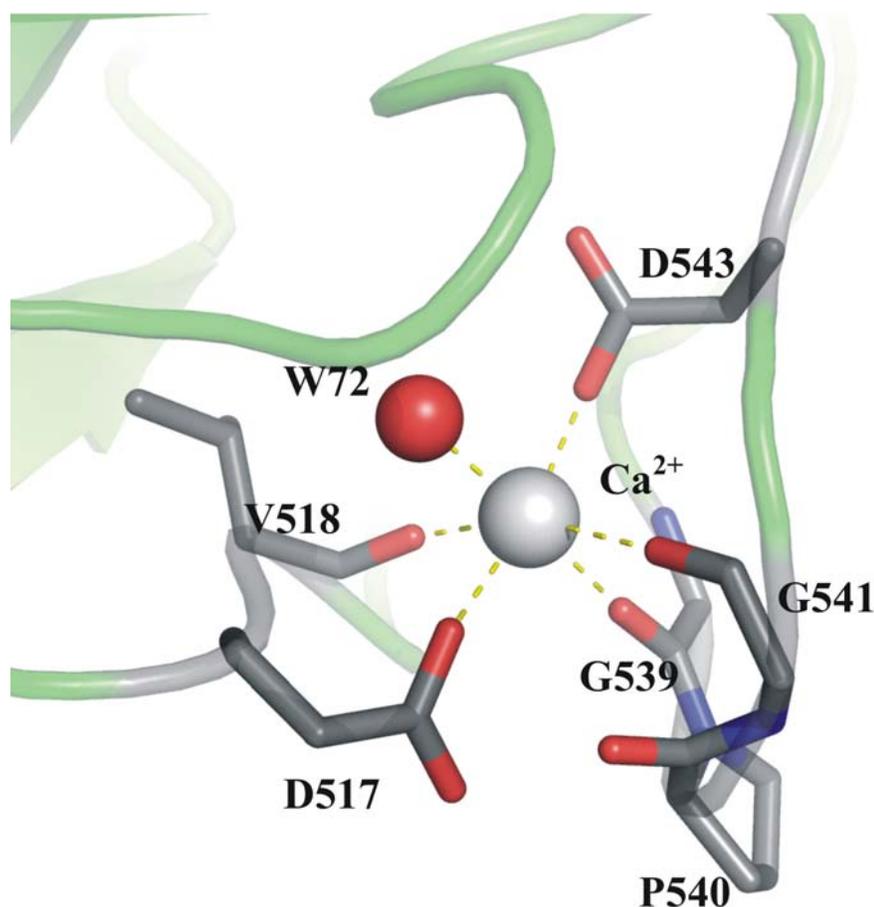

Figure 4.31: The calcium binding site showing the carboxylate oxygen atoms of Asp[517] and Asp[543] each and the main chain carbonyl groups of Val[518], Gly[539] and Gly[541] coordinating to the calcium ion.

4.3.8 Tetramerization and Crystal Packing

The unit cell contains a total of 16 individual units and solvent content of ~62%, if the ~10kDa glycosylation is not taken into account and ~55% solvent, taking into account the ~10kDa glycosylation. Each unit contains two molecules. The two units are related by non-crystallographic symmetry and form a globular dimer in the asymmetric unit (Figure 4.32). Considering each unit as a hetero-dimer, there is a hetero-tetramer formation in the asymmetric unit. The interface between the two catalytic domains of each unit is formed by interaction between residues Glu^{532}, Tyr^{406} and Glu^{529}, Asp^{405}. The interface between the prodomain and the catalytic domain is stabilized by a number of interactions and covers approximately 21% and 15% of the total solvent accessible surface areas of the prodomain and the catalytic domain respectively. The catalytic domain and the prodomain contribute 56 and 65 residues respectively to form the interface between the two domains. Salt bridges are formed predominantly between the guanidium group of Arg^{339} and the carboxylate groups of Asp^{118}, the ε-amino group of Lys^{346} and carboxylate groups of Asp^{118}, carboxylate groups of Glu^{302} and guanidium group of Arg^{175}. Apart from this there are several hydrogen bonds stabilizing the complex between the prodomain and the catalytic domain.
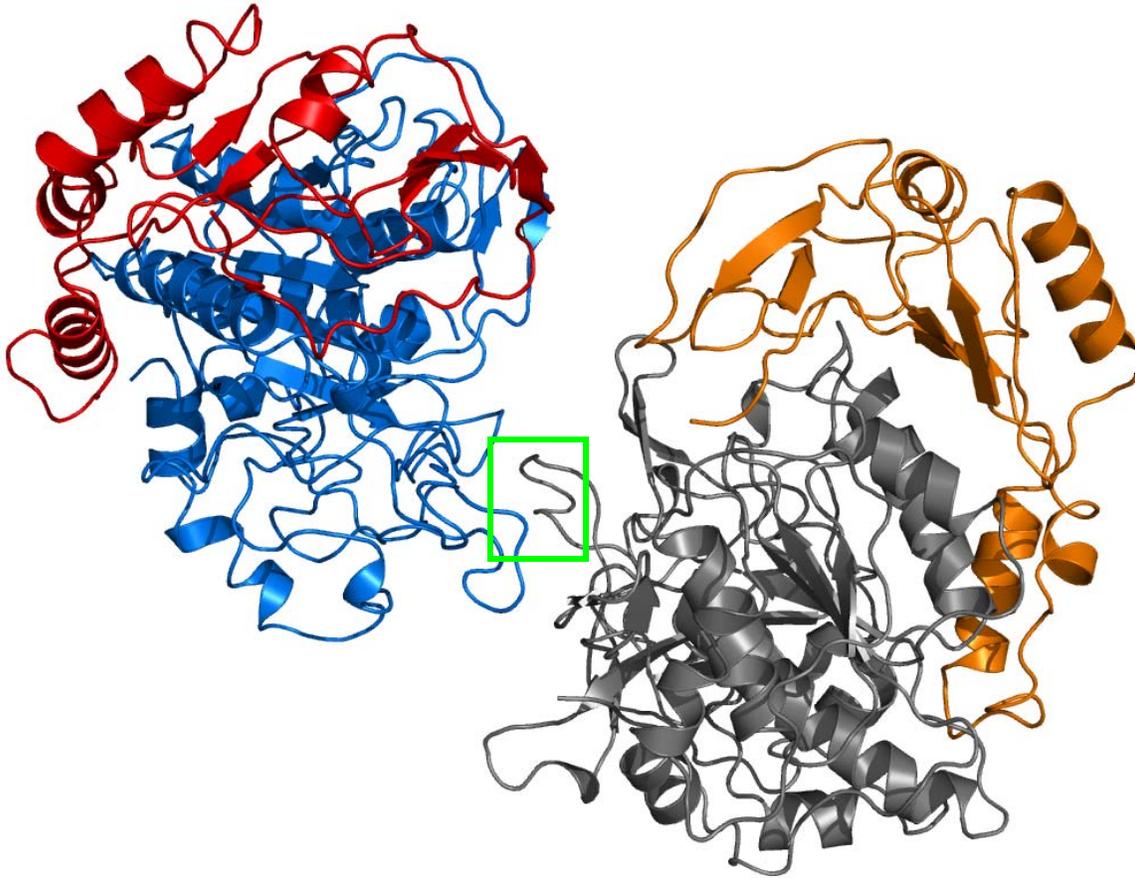
Figure 4.32: TPP-I units in the asymmetric unit. Each unit of TPP-I is a complex of the prodomain (colored in **red** and **orange**) and the catalytic domain (**blue** and **grey**). The catalytic domains of each unit interact through residues Glu^{532}, Tyr^{406} and Glu^{529}, Asp^{405} located on loops shown in the **green** box and form a dimer of the two units the asymmetric unit or a hetero-tetramer of the two prodomains and catalytic domains.

4.4   Discussion and future Perspectives

The current structure of TPP-I is the first structure from the S53 sedolisin family of serine proteases from a mammalian source. Incidentally this is also the only protease identified in humans and numerous mammals that shows significant homology to the Serine-carboxyl Protease family of prokaryotes and lower eukaryotes. The current work has given an insight of the catalytic residues and the active site. Since both the mature and immature forms are most likely to be present in the crystal without any significant disorder in the active site region and other structurally important domains, it can be inferred that there is minimal rearrangement of the domains during the transition from immature to the mature form.. Mutation of the catalytic residues might be exploited to study the resulting defects in the enzyme as well as to improve the resolution and data quality of the structure. This might have an impact on the linker region which is disordered in the current structure owing to cleavage in some molecules in the crystal and being intact in other molecules. This might result in an ordered conformation of the linker region and therefore better diffraction data, perhaps to a better resolution can be obtained. Further experiments might be done to study the structural effects of deglycosylation of the N-linked glycosylation sites, especially its effects on folding and structure of the TPP-I. Furthermore, there is a need to determine the structure of complexes of TPP-I with its inhibitors as well as natural substrates to ascertain the specific binding pockets, which might be responsible for its specificity. Knowing the structure and the active site of TPP-I might also help in modeling its natural neuropeptide substrates and subunit C of ATPase to learn about the mode of action of TPP-I.

References

Aerts, A. M., Francois, I. E., Meert, E. M., Li, Q. T., Cammue, B. P. & Thevissen K (2007). *J Mol Microbiol Biotechnol.* **13(4)**, 243-247.

Arai, M., Oouchi, N. & Murao, S. (1985). *Agric. Biol. Chem.* **49**, 987–991.

Aschauer, H., Vértesy, L., Nesemann, G. & Braunitzer, G. (1983). *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1347-1356.

Bachmann, A. & Kiefhaber, T. (2001). *J. Mol. Biol.* **306**, 375-386.

Balbach, J., Seip, S., Kessler, H., Scharf, M., Kashani-Poor, N. & Engels, J. W. (1998). *Proteins*, **33**, 285-294.

Bernardini, F. & Warburton, M. J. (2002). *Biochem. J.* **366**, 521- 529.

Blundell, T. L., Cooper, J. B., Sali, A. & Zhu, Z. Y. (1991). *Adv. Exp. Med. Biol.* **306**, 443-453.

Bode, W., Gomis-Ruth, F. X. & Stockler, W. (1993). *FEBS Lett.* **331**, 134-140.

Bohlmann, H. & Apel, K. (1987). *Mol. Gen. Genet.* **207**, 446-454.

Botos I. & Wlodawer A. (2007). *Current opinion in Structural Biology*. **17**, 1-8.

Botos, I., Melnikov, E. E., Cherry, S., Kozlov, S., Makhovskaya, O. V., Tropea, J. E., Gustchina, A., Rotanova, T. V. & Wlodawer A.(2005). *J. Mol. Biol.* **351**, 144–157.

Broekaert, W. F., Terras, F. R., Cammue, B. P. & Osborn, R. W. *Plant Physiol.* (1995). **(4)**, 1353-1358.

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

Brünger, A. T. (1992). *Nature* **355**, 472-475.

Büssing, A., Stein, G. M., Wagner, M., Wagner, B., Schaller, G., Pfüller, U. & Schietzel, M. (1999). *Eur. J. Biochem.* **262**, 79-87.

Caaveiro, J. M., Molina, A., Rodríguez-Palenzuela, P., Goñi, F. M. & González-Mañas, J. M. (1998). *Protein Sci.* **7(12)**, 2567-2577.

Cochran, W. (1955). *Acta Cryst.* **8**, 473-478.

Colilla, F., J., Rocher, A. & Mendez, E. (1990). *FEBS Lett.* **270(1-2)**, 191-194.

Comellas-Bigler, M., Fuentes-Prior, P., Maskos, K., Huber, R., Oyama, H., Uchida, K., Dunn, B.M., Oda, K. & Bode, W. (2002). *Structure.* **10**, 865-876.

Conners, R., Konarev, A. V., Forsyth, J., Lovegrove, A., Marsh, J., Joseph-Horne T., Shewry, P., & Brady R. L. (2007). *J Biol Chem.* **282(38)**, 27760-27768.

Coulon, A., Berkane, E., Sautereau, A. M., Ureck, K., Rougé, P. & Lopez, A. (2002). *Biochim. Biophys. Acta* **1559**, 145-159.

Coulon, A., Mosbah, A., Lopez, A., Sautereau, A. M., Schaller, G., Urech, K., Rougé, P. & Darbon, H. (2003). *Biochem. J.* **374**, 71-78.

Coux, O., Tanaka, K. & Goldberg, A. L. (1996). *Annu. Rev. Biochem.* **65**, 801-847.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* **D50**, 414-423.

Davies, D. R. (1990). *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 189-215.

Debreczeni J. É. (2004). PhD thesis, University of Göttingen, Germany.

Debreczeni, J. É., Girmann, B., Zeeck, A., Krätzner, R. & Sheldrick, G. M. (2003). *Acta Cryst*. **D59**, 2125-2132.

DeLano, W. L. (2003). The PYMOL molecular graphics system. DeLano Scientific

Doruker, P., Atilgan, A. R. & Bahar, I. (2000). *Proteins*. **40**, 512-524.

Du, P.-G., Kato, S., Li, Y.-H., Maeda, T., Yamane, T., Yamamoto, S., Fujiwara, M., Yamamoto, Y., Nishi, K. & Ohkubo, I. (2001) *Biol. Chem.* **382**, 1715–1725.

Dunn, B. M., Goodenow, M. M., Gustchina, A. & Wlodawer, A. (2002). *Genome Biol.* **3(4)**.

Eichinger, A., Beisel, H., Jacob, U., Huber, R., Medrano, F., Banbula, A., Potempa, J., Travis, J., & Bode, W. (1999). *The EMBO Journal*. **18(20)**, 5453–5462.

Emsley, P. & Cowtan, K., (2004). *Acta Cryst.* **D60**, 2126-2132.

Evans, J., Wang, Y., Shaw, K. P. & Vernon, L. P. (1989). *Procl. Natl. Acad. Sci.* **86**, 5849-5853.

Ezaki, J., Takeda-Ezaki, M. & Kominami, E. (2000). *J Biochem* **128**, 509-516.

Ezaki, J., Takeda-Ezaki, M., Oda, K., and Kominami, E. (2000) *Biochem. Biophys. Res. Commun.* **268,** 904–908.

Ezaki, J., Tanida, I., Kanehagi, N. & Kominami, E. (1999) *J. Neurochem.* **72,** 2573-2582.

Ezaki, J., Wolfe, L. S. & Kominami, E. (1996). *J Neurochem* **67**, 1677-1687.

Ezaki, J., Wolfe, L. S. & Kominami, E. (1997). *Neuropediatrics*. **28**, 53-55.

Ezaki, J., Wolfe, L. S., Higuti, T., Ishidoh, K. & Kominami, E. (1995). *J Neurochem* **64**, 733-741.

Florack, D. E. & Stiekema, W. J. (1994). *Plant Mol. Biol.* **26**, 25-27.

Franco, O. L., Ridgen, D. J., Melo, F. R., Bloch, C. Jr., Silva, C. P. & Grossi de Sa, M. F. (2000). *Eur J. Biochem.* **267**, 2166 - 2173.

Franco, O. L., Rigden, D. J., Melo, F. R. & Grossi-de-Sá, M. F. (2002). *Eur. J. Biochem.* **269**, 397-412.

Friess H., Beger H. G., Kunz J., Funk N., Schilling M. & Büchler M. W. (1996). *Anticancer Res*. **16 (2)**, 915-920.

Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann S. C. & James, M. N. G. (1995). *Protein Science*. **4**, 960-972.

Fujinaga, M., Cherney, M. M., Oyama, H., Oda, K. & James, M. N. G. (2004). *Proc. Natl. Acad. Sci.* **101**, 103364-103369.

García-Olmedo, F., Carmona, M. J, López-Fando, J. J, Fernández, J. A, Castagnaro, A., Molina A., Hernández-Lucas, C. & Carbonero, P. (1992). *Plant Gene Research. Springer-Verlag, Wien New York*, 283-302.

García-Olmedo, F., Molina, A., Alamillo, J. M. & Rodriguez-Palenzuela, P. (1998). *Biopolymers* **47**, 479-491.

García-Olmedo, F., Molina, A., Alamillo, J. M. & Rodriguez-Palenzuela, P. (1998). Biopolymers **47**: 479-491.

Garcia-Olmedo, F., Salcedo, G., Sanchez-Monge, R., Gomez, L., Royo, J. & Carbonero, P. (1987). *Oxford Surv. Plant Mol. Cell Biol.* **4**, 275–334.

Girmann, B. (2003). PhD thesis, University of Göttingen, Germany

Giudici, A. M., Regente, M.C., Villalaín J, Pfüller K, Pfüller U. & De La Canal L. (2004) *Physiol Plant.* **121(1)**, 2-7.

Giudici, M., Pascual, R., de la Canal, L., Pfüller, K., Pfüller, U. & Villalaín, J. (2003).*Biophys J.* **85(2)**, 971-981.

Giudici, M., Poveda, J. A., Molina, M. L., de la Canal, L., González-Ros, J. M., Pfüller, K., Pfüller, U. & Villalaín, J. (2006). *FEBS J.* **273(1)**, 72-83.

Golabek AA, Dolzhanskaya N, Walus M, Wisniewski KE, Kida E.(2008). *J Biol Chem.*, **283(24)**, 16497-504.

Golabek, A. A., Wujek, P., Walus, M., Bieler, S., Soto, C., Wisniewski, K. E. & Kida E. (2004). *J Biol Chem.* **279(30)**, 31058-31067.

Golabek, A. A., Kida, E., Walus, M., Wujek, P., Mehta, P. & Wisniewski, K. E. (2003). *J Biol Chem.* **278(9)**, 7135–7145.

Golabek, A. A., Walus, M., Wisniewski, K. E. & Kida, E. (2005). *J Biol Chem.* **280(9)**, 7550–7561.

Gorler, A. & Kalbitzer, H. R. (1997). *J. Magn. Reson.* **124**, 177-188.

Grossi-de-Sá, M. F. & Chrispeels, M. J. (1997). *Insect Biochem. Mol. Biol.* **27**, 271-281.

Guo, H., Wlodawer, A., Nakayama, T., Xu Q. & Guo H. (2006). *Biochemistry*, **45**, 9129-9137.

Hampel, A., Labanauskas M., Connors P. G., Kirkegard, L., RayBhandary, U. L., Sigler, P. B. & Bock, R. M. (1968). *Science* **162**, 1384-1387.

Harp, J. M., Hanson, B. L., Timm, D. E. & Bunick G. J. (1999). *Acta Cryst.* **D55**, 1329-1334.

Heiny B. M., Albrecht V. & Beuth J. (1998). *Anticancer Res.* **18(1B)**:583-586.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature* **290**, 107-113.

Hendrickson, W. A., Smith, J. L. and Sheriff, S. (1985). *Methods Enzymol.* **115**, 41-55.

Ho, M.F., Yin, F.F., Lajolo, F., & Whitaker, J. R. (1994). *Protein Structure– Function Relationships in Foods, Blackie Academic & Professional, London,* 89– 119.

Hofmann, O., Ve´rtesy, L. & Braunitzer, G. (1985). *Biol. Chem. Hoppe-Seyler* **366**, 1161–1168.

Holtorf, S., Ludwig-Müller, J., Apel, K. & Bohlmann, H. (1998). *Plant Mol. Biol.* **36**, 673-680.

Hooper, N. M. (1994). *FEBS Lett.* **354**, 1-6.

Hughes, P., Dennis, E., Whitecross, M., Llewellyn, D. & Gage, P. (2000). *J Biol Chem* **275**, 823-827.

Jancarik, J. & Kim, S. H. (1991). *J. Appl. Cryst*. **24,** 409-411.

Joel M. Harp, David E. Timma & Gerard J. Bunick, (1998). *Acta Cryst*. **D54**, 622-628.

Johnson, K. A., Kim, E., Teeter, M. M., Suh, S. W. & Stec, B. (2005). *FEBS Lett*. **579(11)**, 2301-2306.

Juers, D. H., Lovelace, J., Bellamy, H. D., Snell, E. H., Matthews, B. W. & Borgstahl, G. E. O.(1997). *Acta Cryst.* **D63**, 1139-1153.

Junaid, M. A., Wu, G. & Pullarkat, R. K. (2000) *J. Neurochem.* **74,** 287–294.

Jung, M. L., Baudino, S., Ribéreau-Gayon, G. & Beck, J. P. (1990). *Cancer Lett.* **51**, 103-108.

Kabsch W. (1993). *J.App. Cryst.* **26**, 795-800.

Kamphuis, I. G., Drenth, J. & Baker, E. N. (1985).J.Mol.Biol. **182(2)**, 317-329.

Kida, E., Wisniewski, K. E. & Golabek, A. A. (2001). *Adv. Genet.* **45,** 35–68.

Kleywegt, G. J. and Jones, T. A. (1996). *Structure.* **4(12)**, 1395-1400.

Konopa, J., Woynarowski, J. M. & Lewandowska-Gumieniak, M. (1980). *Hoppe Seylers Z Physiol Chem.* **361(10)**, 1525-1533.

Layer, P., Carlson, G. L. & DiMagno E. P. (1985). *Gastroenterology* **88**, 1895–1902.

Lin, L., Sohar, I., Lackland, H., and Lobel, P. (2001) *J. Biol. Chem.* **276,** 2249–2255.

Liu, C.-G., Sleat, D. E., Donnelly, R. J., and Lobel, P. (1998) *Genomics* **50,** 206–212.

Llanos, P., Henriquez, M., Minic, J., Elmorjani, K., Marion, D., Riquelme, G., Molgo, J. & BenoitE (2004). *Eur Biophys J.* **33**, 283-284.

LLC, San Carlos, CA, USA.

Löwe J., Stock D., Jap B., Zwickl P., Baumeister W. & Huber R. (1995). *Science, New Series.* **268(5210)**, 533-539.

Lu, S. Deng, P., Liu, X., Luo, J., Han, R., Gu, X., Liang, S., Wang, X., Li, F., Lozanov, V., Patthy, A. & Pongor, S. (1999). *J. Biol. Chem.* **274**, 20473– 20478.

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni L.C. & Read, R.J(2005). *Acta cryst* **D61**, 458-464.

McPherson, A. (1992). *J. Crystal. Growth.* **122**, 161-167.

Mendez, E., Moreno, A., Colilla, F., Pelaez, F., Limas, G. G., Mendez, R., Soriano, F., Salinas, M. & de Haro C. (1990). *Eur J Biochem*. **194(2)**, 533-539.

Meng, E. C. & Ferrin, T. E. (2004). J. Comput. Chem. **25**, 1605-1612.

Meyer, B. H., Muller, F. O., Kruger, J. B. & Grigoleit, H. G., (1984). *S. Afr. Med. J.* **65**, 287-288.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. and Hauptman, H. A.. (1993). *Science.* **259**, 1430-1433.

Mundy, J., Svendsen, I., & Hejgaard, J. (1983). *Carlsberg Res. Commun.* **48**, 81–90.

Murao, S., Goto, A., Matsui, Y. & Ohyama, K. (1980). *Agric. Biol. Chem*. **44**, 1679-1681.

Murao, S., Oouchi, N., Goto, A. & Arai, M. (1983). *Agric. Biol. Chem.* **47**, 453–454.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240-255.

Olson, T. & Samuelsson, G. (1972). *Acta Chem Scand.* **26**, 585-595.

Orrú, S., Scaloni, A., Giannattasio, M., Urech, K., Pucci, P. & Schaller, G. (1997). *Biol. Chem.* **378**, 989-996.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.

Page, A. E., Fuller, K., Chambers, T. J. & Warburton, M. J. (1993) *Arch. Biochem. Biophys.* **307,** 354–359.

Pappenberger, G., Bachmann, A., Müller, R., Aygün, H., Engels, J.W. & Kiefhaber, T. (2003). *J. Mol. Biol.* **326**, 235-246.

Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 17-22.

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

Pelegrini, P. B., Murad, A. M., Grossi-de-Sa, M. F., Mello, L. V., Romeiro, L. A., Noronha, E. F., Caldas, R. A. & Franco, O. L. (2006). *Arch. Insect Biochem. Physiol.* **61**, 77-86.

Petterson, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Pflugrath, J. W., Wiegand, G., Huber, R. & Vértesy, L. (1986). *J. Mol. Biol*. **189**, 383-386

Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters III, H. K., Kapust, R. B., Li, M., Wlodawer, A. & Waugh, D. S. (2002). *J Biol Chem*. **277(52)**, 50564–50572.

Pillai, B., Cherney, M. M., Hiraga, K., Takada, K., Oda, K., & James, M. N. G. (2007). *J. Mol. Biol.* **365**, 343–361.

Puente, X. S., Sanchez, L. M., Overall, C. M. & Lopez- Otin, C. (2003). *Nature Rev. Genet.* **4**, 544–558.

Radisky, E. S., Kwan, G., Lu C. K. & Koshland Jr., D. E. (2004). *Biochemistry*. **43***, 13648-13656.

Ramachandran, G. N. & Sasisekharan V. (1968). *Adv Protein Chem.* **23**, 283-438.

Rawlings, N. D., Morton, F. R. & Barrett, A. J. (2006). *Nucleic Acids Res.* **34**, 270-272.

Rawlings, N. D., O'Brien, E. & Barrett, A. J. (2002). *Nucleic Acids Res. 30*, 343.

Ribéreau-Gayon, G., Jung, M. L., Baudino, S., Salle, G. & Beck, J. P. (1986). *Experientia* **42**, 594-599.

Romagnoli, S., Fogolari, F., Catalano, M., Zetta, L., Schaller, G., Urech, K., Giannattasio, M., Ragona, L. & Molinari, H. (2003). *Biochemistry* **42**, 12503-12510.

Romagnoli, S., Ugolini, R., Fogolari, F., Schaller, G., Urech, K., Giannattasio, M., Ragona, R. & Molinari, H. (2000). *Biochem. J.* **350**, 569-577.

Samuelsson, G. & Jayawardene, A. L. (1974). *Acta Pharm. Suec.* **11**, 175-184.

Samuelsson, G. & Pettersson, B. (1971). *Eur. J. Biochem.* **21**, 86-89.

Samuelsson, G., Seger, L. & Olson, T. (1968). *Acta Chem. Scand.* **22**, 2624-2642.

Scarselli, M., Bernini, A., Segoni, C., Molinari, H., Esposito, G., Lesk, A. M., Laschi, F., Temussi, P. & Niccolai, N. (1999). *J. Biomol. NMR*. **15**, 125-133.

Schaller, G., Urech, K. & Giannattasio, M. (1996). *Phytother. Res.* **10**, 473-477.

Schaller, G., Urech, K. & Giannattasio, M. (2000). *Mistilteinn* **2000 (1)**, 32-40.

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

Schonbrunner, N., Pappenberger, G., Scharf, M., Engels, J. & Kiefhaber, T. (1997). *Biochemistry*. **36**, 9057-9065.

Schrader, G. & Apel, K. (1991). *Eur. J. Biochem.* **198**, 549-553.

Schrader-Fischer, G. & Apel, K. (1993). *Plant Physiol.* **101**: 745-749.

Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R. & Baumeister, W. (1995). *Science*. **268**, 579-582.

Sheldrick G. M. (2002). *Z. Kristallogr.* **217**, 644-650.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319-343.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods.Enzymol.* **277**, 319-343.

Siezen, R. J., Renckens, B. & Boekhorst, J. (2007). *Proteins*, **67**, 681-694.

Simon C. Lovell, Ian W. Davis, W. Bryan Arendall III, Paul I. W. de Bakker, J. Michael Word, Michael G. Prisant, Jane S. Richardson & David C. Richardson (2003) *Proteins: Structure, Function, and Genetics.* **50**, 437-450.

Sleat, D. E., Donnelly, R. J., Lackland, H., Liu, C. G., Sohar, I., Pullarkat, R. K. & Lobel, P. (1997). *Science* **277**, 1802–1805.

Sleat, D. E., Gin, R. M., Sohar, I., Wisniewski, K. E., Sklower Brooks, S., Pullarkat, R., Palmer, D. N., Lerner, T. J., Boustany, R. M., Uldall, P., Siakotos, A. N., Donnelly, R. J. & Lobel, P. (1999). *Am. J. Hum. Genet.* **64**, 1511–1523.

Song, X., Wang, J., Wu, F., Li, X., Teng, M. & Gong W. (2005). *Plant Mol Biol*. **57(1)**, 13-20.

Spelbrink, R. G., Dilmac, N., Allen, A., Smith, T. J., Shah, D. M. & Hockerman, G. H. (2004). *Plant Physiol.* **135(4)**, 2055-2067.

Stec, B., Rao, U. & Teeter, M. M. (1995). *Acta Cryst.* **D51**, 914-924.

Steinfeld R, Steinke HB, Isbrandt D, Kohlschütter A, Gärtner J.(2004). *Hum Mol Genet.* **13(20)**, 2483-2491.

Steinfeld, R., Heim, P., von Gregory, H., Meyer, K., Ullrich, K., Goebel, H. H., and Kohlschütter, A. (2002). *Am J Med Genet.* **112(4)**, 347-354.

Steuer-Vogt, M. K., Bonkowsky, V., Ambrosch, P., Scholz, M., Neiss, A., Strutz, J., Hennig, M., Lenarz, T. and Arnold, W. (2001) *Eur J Cancer*. **37(1)**, 9-11.

Svensson, B., Fukuda, K., Nielsen, P. K. & Bønsager, B. C. (2004). *Biochimica et Biophysica Acta* **1696**, 145– 156.

Swain, A. L., Miller, M. M., Green, J., Richo, D. H., Schneider, J., Kent, S. B. H. & Wlodawer, A. (1990). *Proc. Natl. Acad. Sci.* **87**. 8805-8809.

Tabiasco, J., Pont, F., Fournié, J. J. & Vercellone, A. (2002). *Eur J Biochem.* **269(10)**, 2591-600.

Tang, J.; James, M. N.; Hsu, I. N.; Jenkins, J. A.; Blundell, T. L. (1978). *Nature.* **271**, 618.

Terras, F. R. G., Torrekens S., Van Leuven, F. & Broekaert, W. F. (1996) *Plant Physiol Biochem* **34**, 599-603.

Thevissen, K., Ferket, K. K., Francois, I. E. & Cammue, B. P. (2003). *Peptides.* **24(11)**, 1705-1712.

Thevissen, K., Ghazi, A., De Samblanx, G. W., Brownlee, C., Osborn, R. W. & Brockaert W. F. (1996). *J. Biol. Chem.* **271**, 15018-15025.

Thevissen, K., Terras, F. R. & Broekaert, W. F. (1999). *Appl Environ Microbiol.* **65(12)**, 5451-5458.

Tomkinson, B. (1999) *Trends Biochem. Sci.* **24,** 355–359

Tsiakas K, Steinfeld R, Storch S, Ezaki J, Lukacs Z, Kominami E, Kohlschütter A, Ullrich K, Braulke T.(2004). *Glycobiology*. **14(4)**, 1C-5C.

Tyndall, J. D. A., Nall, T. & Fairlie D. P. (2005) *Chem. Rev.* **105**, 973-999.

Urech, K., Schaller, G., Ziska, P. & Giannattasio, M. (1995). *Phytother. Res.* **9**, 49-55.

Vértesy, L., Oeding, V., Bender, R., Zepf, K. & Nesemann G. (1984). *Eur. J. Biochem.* **141**, 505– 512.

Vines, D. & Warburton, M. J. (1998) *Biochim. Biophys. Acta* **1384,** 233–242.

Wada, K., Ozaki, Y., Matsubara, H. & Yoshizumi, H. (1982). *J. Biochem.* **91**, 257-263.

Walus, M., Kida, E., Wisniewski, K. E. & Golabek A. A. (2005). *FEBS Letters*. **579**, 1383–1388.

Wang, Y., Zhang Y. & Ha Y. (2006). *Nature.* **444(9)**, 179-183.

Warburton, M. J. & Bernardini, F. (2002) *Neurosci Lett.* **331,** 99–102.

Weselake, R. J., MacGregor, A. W. & Hill, R.D. (1983). *Plant Physiol.* **72**, 809-812.

Wiegand, G., Epp, O. & Huber, R. (1995). *J. Mol. Biol.* **247**, 99–110.

Winterfeld, K. & Bjil, L. H. (1948). *Liebig's Annales* **561**, 107-111.

Wisniewski, K. E., Kida, E., Golabek, A. A., Kaczmarski, W., Connell, F. & Zhong, N. (2001). *Adv. Genet.* **45**, 1–34.

Wlodawer, A. & Gustchina, A. (2000). *Biochim. Biophys. Acta*. **1477(1-2)**, 16-34.

Wlodawer, A., Durell, S. R., Li, M., Oyama, H., Oda K. & Dunn, B. M. (2003). *BMC Structural Biology*, **3:8**.

Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M. & Oda, K. (2001). *Nature structural biology*. **8(5)**, 442-446.

Wujek, P., Kida, E., Walus, M., Wisniewski, K. E. & Golabek A. A. (2004). *J Biol Chem*. **279(13)**, 12827–12839.

Yamada, T., Hattori, K. and Ishimoto, M. (2001). *Phytochem.* **58**, 59-66.

Zarkovic N., Vukovic T., Loncaric I., Miletic M., Zarkovic K., Borovic S., Cipak A., Sabolovic S., Konitzer M. and Mang, S. (2001).*Cancer Biother Radiopharm.* **16(1)**, 55-62.

Zscherp, C., Aygün, H., Engels, J. W. & Mäntele, W. (2003). *Biochimica et Biophysica Acta*. **1651**, 139-145.

Publications

1. **Pal A**, Debreczeni JE, Sevvana M, Gruene T, Kahle B, Zeeck A, Sheldrick GM. *Structures of viscotoxins A1 and B2 from European mistletoe solved using native data alone. Acta Crystallogr D Biol Crystallogr.* 2008 Sep; 64 (Pt 9):985-92.

2. Sarish S, Nembenna S, Nagendran S, Roesky HW, **Pal A**, Herbst-Irmer R, Ringe A, Magull J. *A reactivity change of a strontium monohydroxide by umpolung to an acid. Inorg Chem.* (2008) Jul 7;47(13):5971-7.

3. Gurubasavaraj PM, Roesky HW, Nekoueishahraki B, **Pal A**, Herbst-Irmer R. *From unstable to stable: half-metallocene catalysts for olefin polymerization.* Inorg Chem. (2008) Jun 16; 47(12):5324-31.

4. Lyashenko G, Herbst-Irmer R, Jancik V, **Pal A**, Mösch-Zanetti NC. *Molybdenum oxo and imido complexes of beta-diketiminate ligands: synthesis and structural aspects. Inorg Chem.* (2008) Jan 7;47(1):113-20.

5. Singh, S., Chai, J., **Pal**, **A.**, Jancik, V., Roesky, H. W. & Herbst-Irmer, R. *Base free lithium-organoaluminate and the gallium congener: potential precursors to heterometallic assemblies. Chem. Commun.*, 2007, 4934 – 4936.

6. So, C-W., Roesky, H. W., Oswald, R. B., **Pal, A.** & Jones, P. G. *Synthesis and characterization of [{PhC(NBut)2}Si(S)SBut]: a silicon thioester analogue with the Si (:S)-S-skeleton. Dalton Transactions* (2007*).* (45), 5241-5244

7. Widjaja, T., Fitjer, L., **Pal, A.**, Schmidt, H-G., Noltemeyer, M., Diedrich, C. & Grimme S. *Pseudohelical and Helical Primary Structures of 1,2-Spiroannelated Four- and Five-Membered Rings: Syntheses and Chiroptical Properties. Journal of Organic Chemistry* (2007). 72(24), 9264-9277.

8. Sachse, A., Moesch-Zanetti, N. C., Lyashenko, G., Wielandt, J. W., Most, K., Magull, J., Dall'Antonia, F., **Pal, A.** & Herbst-Irmer, R. *Rhenium(V) Oxo*

Complexes with Acetylacetone Derived Schiff Bases: Structure and Catalytic Epoxidation. Inorganic Chemistry (2007). 46(17), 7129-7135.

9. Tripathi, P., **Pal, A.**, Jancik, V., Pandey, A. K., Singh, J. & Singh, N. K. *Metal-assisted transformation of N-benzoyldithiocarbazate to 5-phenyl-1,3,4-oxadiazole-2-thiol in the presence of ethylenediamine, and its first row transition metal complexes. Polyhedron* (2007). 26(12), 2597-2602.

10. Gurubasavaraj, P. M., Roesky, H. W., Sharma, Prabhuodeyara M. V., Oswald, R. B., Dolle, V., Herbst-Irmer, R. & **Pal, A.** *Oxygen Effect in Heterobimetallic Catalysis: The Zr-O-Ti System as an Excellent Example for Olefin Polymerization. Organometallics* (2007). 26(14), 3346-3351.

11. Nehete, U. N., Roesky, H. W., Jancik, V., **Pal, A.** & Magull, J., *Polyhedral antimony(III) and bismuth(III) siloxanes: Synthesis, spectral studies, and structural characterization of [Sb(O3SiR)]4 and [Bi$_{12}$(O$_3$SiR)$_8$(μ_3-O)$_4$Cl$_4$(THF)$_8$](R=(2,6-iPr$_2$C$_6$H$_3$)N(SiMe$_3$)). Inorganica Chimica Acta* (2007). 360(4), 1248-1257.

12. Lyashenko, G., Saischek, G., **Pal, A.,** Herbst-Irmer, R. & Moesch-Zanetti, N. C. *Molecular oxygen activation by a molybdenum(IV) monooxo bis(ß - ketiminato) complex. Chemical Communications* (2007). (7), 701-703.

13. Gurubasavaraj, P. M., Mandal, S. K., Roesky, H. W., Oswald, R. B., **Pal, A.** & Noltemeyer, M. *Synthesis, Structural Characterization, Catalytic Properties, and Theoretical Study of Compounds Containing an Al-O-M (M = Ti, Hf) Core. Inorganic Chemistry* (2007). 46(4), 1056-1061.

14. Singh, S., **Pal, A.**, Roesky, H. W. & Herbst-Irmer, R. *Adducts of Cp3Ln with LGa(Me)OH, syntheses and x-ray crystal structures of LGa(Me)HO$^{\rightarrow}$ LnCp3 {Ln = Sm, Nd, Yb; L = HC[C(Me)N(2,6-iPr$_2$C$_6$H$_3$)]$_2$}. European Journal of Inorganic Chemistry* (2006). (20), 4029-4032.

15. Nembenna, S., Roesky, H. W., Mandal, S. K., Oswald, R. B., **Pal, A,** Herbst-Irmer, R., Noltemeyer, M. & Schmidt, H.-G. *Soluble Molecular Compounds with the Mg-O-Al Structural Motif: A Model Approach for the Fixation of Organometallics on a MgO Surface. Journal of the American Chemical Society* (2006). 128(40), 13056-13057.

16. Singh, N., Jabeen, T., **Pal, A.**, Sharma, S., Perbandt, M., Betzel, C. & Singh, T. P. *Crystal structures of the complexes of a group IIA phospholipase A2 with two natural anti-inflammatory agents, anisic acid, and atropine reveal a similar mode of binding. Proteins: Structure, Function, and Bioinformatics* (2006). 64(1), 89-100.

17. Lyashenko, G., Jancik, V., **Pal, A.**, Herbst-Irmer, R. & Moesch-Zanetti, N. C. *Dioxomolybdenum(VI) and dioxotungsten(VI) complexes supported by an amido ligand. Dalton Transactions* (2006). (10), 1294-1301.

18. Singh, S., Ahn, H.-J., Stasch, A., Jancik, V., Roesky, H. W., **Pal, A.**, Biadene, M., Herbst-Irmer, R., Noltemeyer, M. & Schmidt, H.-G. *Syntheses, Characterization, and x-ray Crystal Structures of ß -Diketiminate Group 13 Hydrides, Chlorides, and Fluorides. Inorganic Chemistry* (2006), 45(4), 1853-1860.

19. **Pal**, **A.**, Pradhan, P., Banerji, A. & Singh, T. P. *Crystal structure of pimolin, $C_{26}H_{20}O_8$. Z. Kristallogr.* (2003). NCS 218, 337–338.

Manuscripts under preparation / in press

1. **Pal**, **A.**, Grüne, T. & Sheldrick, G. M. *Crystal Structure analysis of Tendamistat* (Manuscript under preparation).

2. **Pal**, **A.**, Grüne, T., Steinfeld, R., Krätzner R. & Sheldrick, G. M. *The structure of tripeptidyl peptidase I unravels the molecular basis of late neuronal ceroid lipofuscinosis* (Manuscript under preparation).

Lebenslauf

Persönliche Daten

Name	Aritra Pal
Geburtsdatum	13th December, 1981
Geburtsort	Ranchi
Staatsangehörigkeit	Indisch

Schulbildung

1985-1996	Primary School (class I – IX)
1997	Indian Certificate Secondary Examination (Class X)
1997-1999	All India Senior School Certificate Examination (Class XI- XII)

Hochschulstudium

1999-2002	**B.Sc. (hons) in Human Biology** with Honors in Biophysics, All India Institute of Medical Sciences, New Delhi, Indien
2002-2004	**M.Sc. in Biophysics**, All India Institute of Medical Sciences, New Delhi, Indien

Promotion

2004-2008	Dissertation am Lehrstuhl für Strukturchemie der Georg-August Universität Göttingen im Arbeitkreis von Prof. George M. Sheldrick zum Thema: "Structure determination of Viscotoxin A1, Tendamistat and Tri Peptidyl Peptidase - I"