Structural Studies of
Australian Snake Venom Compounds

A thesis submitted for the degree of Doctor of Philosophy
at the University of Queensland in November 2007
Emma-Karin I. Millers
School of Molecular and Microbial Sciences
Statement of Originality

The work presented in the current document was undertaken in the School of Molecular and Microbial Sciences at The University of Queensland, Brisbane, Australia, between the dates of August 2004 and November 2007 under supervision of Dr Luke W. Guddat. I hereby declare that the work presented in this document is, to the best of my knowledge and belief, original and has not been submitted either in whole or in part for another degree at The University of Queensland or any other institution. The assistance or work of others has either been referenced as it appears in the document or has been mentioned in the “Acknowledgements” section.

Emma-Karin Millers
November 2007
Acknowledgements

This thesis is part of the Venomics Research Program, established in 2002 to study the components of Australian snake venoms. The research is a collaboration between The University of Queensland (UQ), Queensland Institute of Medical Research (QIMR), Princess Alexandra Hospital (PAH) and QRx Pharma Pty Ltd and was supported by an ARC Linkage grant.

There are many people who have helped me and given me guidance and support throughout the course of this project. Firstly, I wish to thank my supervisors at The University of Queensland; Professor John de Jersey, Dr Luke Guddat, and Dr Manuela Trabi for their efforts to keep me on the right track and their support. In particular I am indebted to my principal supervisor Dr Luke Guddat for the encouragement, assistance and patience he has manifested from the beginning and throughout the research.

At School of Molecular and Microbial Science (UQ), I am thankful to Chris Wood for carrying out the mass spectroscopy experiments and Lyle Carrington for assisting me with the analytical centrifugation experiments.

At QIMR I have had the privilege to get valuable assistance and I am thankful to Professor Martin Lavin and Dr Geoff Birrell for their guidance during the expression of microplasmin. I also would like to thank Hayley Snelling (QIMR) for producing ACII-4 for my studies.

At PAH I am thankful to Dr Paul Masci, one of the originators of the project, for his support and useful ideas and for his help with purification of microplasmin and the rTxIn-1 microplasmin complex.

The X-ray experiments were carried out at UQ at the Institute for Molecular Bioscience (IMB) and I am thankful to Dr Karl Byriel for his assistance and support during the measurement of X-ray data.
I sincerely appreciate the help of Professor Chai Zhang (Oklahoma Medical Research Foundation, USA) who kindly donated the plasmid containing human microplasmin and Dr Nancy Wakeham who assisted in the microplasmin production via email correspondence.

The following people have been helpful and deserve recognition for their assistance in their areas of expertise; Dr Gregor Guncar (UQ) for his valuable assistance regarding various computer programs used during the process of solving protein structures by X-ray crystallography, Serge Cohen (ARP/wARP) for looking at my data set of ACII-4 and correctly predicting that it was twinned, Dr Ronan Keegan (CCP4) for his suggestions and for introducing me to the MrBump computer program to solve the structure of ACII-4, Professor Lars Liljas (Uppsala University, Sweden) for previous teaching and helpful correspondence regarding molecular replacement, Professor Rolf Hilgenfeld (University of Lübeck, Germany) for valuable ideas about the rTxln-1-microplasmin complex and James Letts (The Rockefeller University, USA) for useful suggestions regarding protein structures.

It has been a pleasure to be a member of the Venomics group and to participate in the friendly cooperation among the members. In addition to those members already mentioned above I want to express my gratitude to the group members; Drs Stephen Earl, Liam St Pierre, Simone Flight, Igor Filippovich, Natasha Sorokina, Renee Stirling and the students Jonathan Robson, Ann Liao, Cindy Yan, Sharon Du.

It had not been possible to complete this work without support and love of my family and friends. I thank all of you for your support and encouragement.
Publications by the Candidate relevant to the Thesis

Some of the crystallization work presented in this thesis has been reported in the following journal article.

Crystallization and preliminary X-ray analysis of a Kunitz-type inhibitor, textilinin-1 from *Pseudonaja textilis textilis*.

Four further publications are expected. These will be on crystal structures of;

- free textilinin-1 from *Pseudonaja textilis*
- textilinin-1 in complex with trypsin
- textilinin-1 in complex with plasmin
- ACII-4 from *Pseudechis australis*
Abstract

The Australian elapids are some of the most deadly snakes in the world. Their venoms contain a cocktail of chemicals including peptides and proteins with a broad range of biological activities including pre- and post-synaptic neurotoxins, anti- and pro-coagulants, cardiotoxins, cytotoxins, hemorrhagic toxins and myotoxins. To date, however, a systematic approach aimed at the annotation and characterization of all components of all Australian snake venoms has not been undertaken. Therefore, the overall aim of our “Venomics” group is to isolate and evaluate a large number of venom proteins from Australian snakes with a view to finding biologically active molecules that could ultimately have application in a therapeutic setting. A key component of the overall effort is the understanding of structure-function relationships for many of these molecules. This thesis focuses on the structure determination and analysis of two proteins, textilinin-1 from the Australian Common Brown snake, *Pseudonaja textilis*, and ACII-4 a phospholipase from the King Brown snake, *Pseudechis australis*.

Textilinin-1 is a small protein of 59 amino acids. A BLAST analysis showed this molecule has a sequence identity of 45% to the well known Kunitz inhibitor, aprotinin (bovine pancreatic trypsin inhibitor). The most interesting property of textilinin-1 is that it has potent anti-fibrinolytic activity, most likely because it is a nM inhibitor of plasmin. For this reason, textilinin-1 is being considered as an alternative to aprotinin in surgery as an anti-bleeding agent. In this study, I determined the three-dimensional structure of recombinant textilinin-1 as the free inhibitor to 1.63 Å resolution, and its structure in complex with trypsin and microplasmin (the catalytic domain of plasmin) to 1.64 Å and 2.78 Å resolution, respectively. An unusual feature of the structure of free textilinin-1 is that its canonical loop, the region which predominantly interacts with and inhibits the protease, can adopt multiple conformations. In one snapshot taken from the crystal structure this loop is inverted such that the critical protease binding residue Val18 becomes partially buried. In complex with trypsin or microplasmin, textilinin-1 binds such that its P1 residue, Arg17, inserts into the specificity pocket. In these complexes the catalytic serine hydroxyl oxygen of the protease is in closer-than-van der Waals
contact with the carbonyl carbon of the scissile peptide bond of textilinin-1. In the microplasmin complex, His603 adopts an altered non-catalytically competent position such that its side-chain has swiveled around its χ₁ bond and out of its classical catalytic triad location. To compensate, a water molecule is observed bridging the serine and aspartate side-chains. The reason for this appears to be due to the close approach of Val18 from textilinin-1 forcing the histidine to move. An interesting observation from the trypsin-textilinin-1 complex when compared to the structure of trypsin in complex with aprotinin is that the relative docking angles of these two inhibitors differs by ~25°. The reason for this difference is the variations in the sequences of the two inhibitors. At the P1' and P3' sites, two bulkier amino acids, valine and phenylalanine are observed in textilinin-1 as compared to alanine and isoleucine which are present in aprotinin. As a result the textilinin-1 molecule is unable to dock into trypsin in the same orientation as aprotinin.

ACII-4 has a number of important biological activities including its ability to increase prothrombin time in citrated plasma and to inhibit the conversion of prothrombin to thrombin. ACII-4 has also been shown to inhibit platelet aggregation by collagen, arachionic acid and ADP, and inhibit the conversion of factor X to factor Xa. As a first step to understanding structure-function relationships of this molecule its X-ray crystal structure has been determined to 1.56 Å resolution. ACII-4 has a characteristic PLA₂ fold stabilized by seven disulfide bonds. In the crystal structure, a Ca²⁺ ion is observed bound to the polypeptide, coordinated to seven ligands Tyr28O, Gly30O, Gly32O, the side-chain oxygens of Asp49, and two water molecules. The active site consisting of the catalytic triad of His48, Asp92, and a water molecule are visible in the structure. A polyethylene glycol molecule from the crystallization buffer is positioned in the active site where the fatty acid substrate would be found. The side-chain of Trp31 is situated in the opening to the active site and forms a hydrophobic contact with a second polyethylene glycol molecule. Two possible dimer formations of ACII-4 can be deduced from the crystal lattice contacts. One is a hydrophilic interface with a buried surface of 1141 Å² and the other is a hydrophobic interface with a buried surface of 686 Å². The implications of the structures determined in this thesis for the functions of textilinin-1 and ACII-4 as venom components and as potential pharmaceuticals are discussed.
## Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Twinning fraction or phase angle</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Phi, main-chain dihedral angle</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Chi, side-chain dihedral angle</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>Psi, main-chain dihedral angle</td>
</tr>
<tr>
<td>Å</td>
<td>Ångström, the unit denoting $10^{-10}$ metre</td>
</tr>
<tr>
<td>$B$</td>
<td>Thermal parameter</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization mass spectrometry</td>
</tr>
<tr>
<td>$F, F$</td>
<td>Structure factor; complex number and absolute value</td>
</tr>
<tr>
<td>$F_{\text{calc}}, F_\text{c}$</td>
<td>Calculated structure factor</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>$F_{\text{obs}}, F_\text{o}$</td>
<td>Observed structure factor</td>
</tr>
<tr>
<td>h, k, l</td>
<td>Miller indices</td>
</tr>
<tr>
<td>I</td>
<td>Intensity of X-ray diffraction beam</td>
</tr>
<tr>
<td>$k_1, k_{-1}, k_2, k_{-2}$</td>
<td>Rate constants</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>$K_i$</td>
<td>Dissociation constant for inhibitor binding</td>
</tr>
<tr>
<td>LLG</td>
<td>Log-likelihood gain</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple anomalous dispersion</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption ionization-time of flight</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td>$M_w$</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein databank, Brookhaven PDB</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PLA$_2$</td>
<td>Phospholipase A$_2$</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>$R_{\text{factor}}$</td>
<td>Residual index</td>
</tr>
<tr>
<td>$R_{\text{free}}$</td>
<td>Residual index</td>
</tr>
<tr>
<td>rmsd</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>rTxln-1</td>
<td>Recombinant textilinin-1</td>
</tr>
<tr>
<td>SAK</td>
<td>Staphylokinase</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>STK</td>
<td>Streptokinase</td>
</tr>
<tr>
<td>t$_{1/2}$</td>
<td>Half-life</td>
</tr>
<tr>
<td>$V_M$</td>
<td>Matthews coefficient</td>
</tr>
</tbody>
</table>
Table of Contents

Statement of Originality .................................................................................................................. iii

Acknowledgements ............................................................................................................................ iv

Publications by the Candidate relevant to the Thesis ................................................................. vi

Abstract ............................................................................................................................................... vii

Abbreviations ....................................................................................................................................... ix

Table of Contents ................................................................................................................................ xi

Table of Figures .................................................................................................................................... xvii

List of Tables ....................................................................................................................................... xxvi

Chapter 1 .............................................................................................................................................. 1

Introduction .......................................................................................................................................... 1

1.1 Background ..................................................................................................................................... 1

1.2 Australian snake venoms and their components ..................................................................... 1

1.3 Blood clotting ............................................................................................................................... 3

1.4 Serine proteases .......................................................................................................................... 4

1.5 Mechanism of inhibition of serine proteases ........................................................................... 9

1.6 The properties of plasmin and its role in blood clotting ...................................................... 9

1.7 Aprotinin in therapeutic use to regulate blood loss ................................................................. 13

1.8 The properties of textilinin-1 ............................................................................................... 14

1.8.1 Textilinin-1's mechanism of inhibition .............................................................................. 16

1.9 Phospholipases .......................................................................................................................... 17

1.10 The properties of ACII-4 ........................................................................................................ 21

1.11 Aim and synopsis of this thesis ............................................................................................... 21

Chapter 2 ........................................................................................................................................... 23

Preparation of Proteins for Crystallization ................................................................................... 23

2.1 Introduction .................................................................................................................................... 23

2.2 Materials and methods ............................................................................................................... 24
2.3 Cloning, expression and purification of recombinant textilinin-1 ...........25
2.4 Preparation of free recombinant textilinin-1 for crystallization experiments .....................................................................................................................26
  2.4.1 Concentration and SDS-PAGE of rTtxn-1 ........................................26
  2.4.2 Mass spectral analysis of rTtxn-1......................................................27
2.5 Preparation of the recombinant textilinin-1-trypsin complex for crystallization experiments ...............................................................29
  2.5.1 Mass spectrometry analysis, SDS-PAGE and activity assay of bovine trypsin .........................................................................................29
  2.5.2 Mass spectrometry analysis of the rTtxn-1-trypsin complex ..............32
  2.5.3 Purification by gel filtration ...............................................................32
  2.5.4 Superdex 200 column .......................................................................33
  2.5.5 Mass spectral analysis of Superdex 200 fractions .............................35
  2.5.6 Activity assay of the rTtxn-1-trypsin complex ..................................35
  2.5.7 Superdex 75 column .........................................................................36
  2.5.8 Concentration of rTtxn-1-trypsin complex .......................................37
  2.5.9 Additional samples of rTtxn-1-trypsin for crystallization .................37
2.6 Cloning, expression and purification of human microplasminogen ........37
  2.6.1 Extraction of plasmid ........................................................................37
  2.6.2 Expression and isolation of human microplasminogen .....................38
  2.6.3 Gel filtration of refolded human microplasminogen ..........................41
  2.6.4 Preparation of rTtxn-1-microplasmin complex .................................42
2.7 Purification and characterization of native ACII-4 ...............................44
2.8 Preparation of ACII-4 for crystallization experiments .........................46
  2.8.1 Concentration of ACII-4 .................................................................46
2.9 Discussion ............................................................................................46
2.10 Summary .............................................................................................47

Chapter 3 ........................................................................................................48

Crystallization of proteins .............................................................................48

  3.1 Introduction ..........................................................................................48
  3.2 Methods and Materials .........................................................................50
  3.3 Crystallization experiments with free recombinant textilinin-1 ..........51
     3.3.1 Cryoprotection of free rTtxn-1 crystals ........................................54
3.3.2 Data collection and processing of free rTxln-1 crystals ................. 54
3.3.3 Space group determination of free rTxln-1 crystals ......................... 55
3.4 Crystallization experiments with recombinant textilinin-1-trypsin .......... 56
  3.4.1 Cryoprotection of the rTxln-1-trypsin crystals ................................ 57
  3.4.2 Data collection for the best rTxln-1-trypsin crystals ....................... 58
  3.4.3 Space group determination of the rTxln-1-trypsin complex crystals .... 58
3.5 Crystallization experiments for the recombinant textilinin-1- microplasmin complex .................................................................................................................. 59
  3.5.1 Data collection and processing of the rTxln-1-microplasmin crystals .. 60
3.6 Crystallization experiments with ACII-4 ............................................... 62
  3.6.1 Data collection of ACII-4 crystals .................................................. 62
3.7 Discussion ............................................................................................... 64
3.8 Summary ................................................................................................. 66

Chapter 4 ........................................................................................................ 67
Crystal structure of free recombinant textilinin-1 ............................................. 67
  4.1 Introduction ............................................................................................ 67
  4.2 Introduction to solving the phase problem of the rTxln-1 crystal ............ 69
  4.3 Phase determination of free rTxln-1 ...................................................... 70
  4.4 Initial model building of the rTxln-1 crystal structure .......................... 72
  4.5 Refinement of model of free rTxln-1 .................................................... 73
  4.6 Refinement statistics and quality of free rTxln-1 structure ..................... 73
  4.7 Crystal structure of free rTxln-1 .......................................................... 77
  4.8 Crystal contacts ..................................................................................... 87
  4.9 Comparison between textilinin-1 and aprotinin .................................. 96
  4.10 Comparison between textilinin-1 and proteins with highest amino acid sequence identity .......................................................... 106
  4.11 Comparison between the six amino acid sequences of textilinin-1 to 6.. 113
  4.12 Summary ............................................................................................ 115

Chapter 5 ........................................................................................................ 116
Crystal structure of the rTxln-1-trypsin complex ............................................. 116
  5.1 Introduction ............................................................................................ 116
  5.2 Phase determination, model building and refinement of rTxln-1-trypsin 118
  5.3 Refinement of the rTxln-1-trypsin complex crystal structure ............... 119
Chapter 5 ..................................................................................................................120
5.4 Refinement statistics and quality of rTxln-1-trypsin complex crystal structure ........................................................................................................120
5.5 Crystal structure of the rTxln-1-trypsin complex ........................................125
5.6 The interface between rTxln-1 and trypsin................................................126
5.7 Crystal contacts..........................................................................................130
5.8 Comparison between rTxln-1 in the trypsin complex and free rTxln-1 ...133
5.9 Comparison of the rTxln-1 trypsin complexes with an aprotinin trypsin complex ....................................................................................................137
5.10 Summary...................................................................................................145

Chapter 6 ..................................................................................................................146
Crystal structure of the rTxln-1 recombinant human microplasmin complex........146
6.1 Phase determination of rTxln-1-microplasmin complex .........................148
6.2 Refinement of the rTxln-1-microplasmin crystal structure ......................148
6.3 Refinement statistics and quality of rTxln-1-microplasmin crystal structure ........................................................................................................150
6.4 Crystal structure of the rTxln-1 microplasmin complex ..........................155
6.5 The interface between rTxln-1 and microplasmin....................................157
6.6 Crystal contacts..........................................................................................167
6.7 Comparison between rTxln-1-microplasmin and existing microplasminogen and microplasmin structures .....................................................174
6.8 Comparison between rTxln-1 in the microplasmin complex, in the trypsin complex and free rTxln-1 .................................................................185
6.9 Comparison between the rTxln-1-microplasmin complex, the rTxln-1-trypsin complex and an aprotinin-trypsin complex .................................188
6.10 Discussion of rTxln-1 dimer .....................................................................198
6.11 Summary...................................................................................................200

Chapter 7 ..................................................................................................................202
Crystal structure of native ACII-4 from Australian King Brown snake (Pseudechis australis) venom.................................................................202
7.1 Introduction to hemihedral crystal twinning .............................................203
7.2 Phase determination of free ACII-4.........................................................204
7.3 Refinement of model of PLA₂ crystal structure .......................................209
7.4 Refinement statistics and quality of ACII-4 crystal structure ..................210
Table of Figures

Figure 1-1  Blood coagulation and fibrinolysis pathways ............................................. 4
Figure 1-2  Nomenclature for the description of protease subsites ................................. 6
Figure 1-3  The catalytic mechanism of the serine proteases ....................................... 8
Figure 1-4  Enzymes and pathways involved in fibrinolysis ....................................... 10
Figure 1-5  Domain structure of plasminogen .............................................................. 11
Figure 1-6  Human microplasmin aligned with bovine trypsin .................................. 13
Figure 1-7  Txln-1 aligned with aprotinin ................................................................. 14
Figure 1-8  Sequences of Txln-1 to 6 .......................................................................... 15
Figure 1-9  Phospholipid hydrolysis by phospholipase A2 ......................................... 18
Figure 1-10  The catalytic mechanism of PLA2 .......................................................... 20
Figure 1-11  Sequence alignment of PA11, PA3 and PA15 .......................................... 21
Figure 2-1  Non-reduced 4-20% gradient SDS-PAGE of purified rTxln-1 ................. 26
Figure 2-2  MALDI-TOF mass spectrum of rTxln-1 ................................................... 28
Figure 2-3  Sedimentation equilibrium experiments of rTxln-1 performed in the analytical ultracentrifuge ................................................................. 29
Figure 2-4  MALDI-TOF mass spectrum of trypsin ...................................................... 31
Figure 2-5  Reduced 4-20% gradient SDS-PAGE of the trypsin sample used in MALDI-TOF mass spectrometry ................................................................. 31
Figure 2-6  MALDI-TOF mass spectrum of the rTxln-1-trypsin complex ................. 32
Figure 2-7  The elution profile of the rTxln-1-trypsin complex by Superdex 200 size exclusion chromatography ................................................................. 33
Figure 2-8  Non reduced 15% SDS-PAGE of fractions collected from Superdex 200 column .................................................................................................. 34
Figure 2-9  Calibration graph of the Superdex 200 column ........................................ 34
Figure 2-10  MALDI-TOF mass spectrum of the rTxln-1-trypsin complex ............... 35
Figure 2-11  The elution profile of the rTxln-1-trypsin complex by Superdex 75 size exclusion chromatography ................................................................. 36
Figure 2-12  Calibration graph of the Superdex 75 column ........................................ 37
Figure 2-13  Analysis of microplasminogen expression .............................................. 39
Figure 2-14  Reduced 12% SDS-PAGE of the purification of inclusion bodies of microplasminogen ................................................................. 40
Figure 2-15 Elution profile of refolded microplasminogen from Sephacryl 300 chromatography..............................................................42
Figure 2-16 Non reduced 15% SDS-PAGE of the activation of plasminogen with urokinase........................................................................................................43
Figure 2-17 Non reduced 15% SDS-PAGE of the rTxlN-1-microplasmin complex.44
Figure 2-18 Sequence of ACII-4 as determined by ms/ms or tryptic digestion alone. ........................................................................................................................45
Figure 3-1 Crystals of free rTxlN-1. ........................................................................52
Figure 3-2 Diffraction pattern for rTxlN-1..............................................................55
Figure 3-3 Crystals of the rTxlN-1-trypsin complex...............................................57
Figure 3-4 Diffraction pattern for rTxlN-1-trypsin complex .......................................58
Figure 3-5 Diffraction pattern for rTxlN-1-microplasmin complex............................61
Figure 3-6 An ACII-4 crystal cluster grown from Hampton crystal screen I condition 17 ...................................................................................................................................62
Figure 3-7 Diffraction pattern for ACII-4. ...............................................................63
Figure 4-1 The crystal packing of rTxlN-1. ...............................................................72
Figure 4-2 Stereoview of the 2Fo-Fc electron density map contoured at 1σ for residues 16-20 and residue 40 of molecule A of rTxlN-1...............................73
Figure 4-3 B-values (Å²) for rTxlN-1 molecules A, B and C. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values. ............................75
Figure 4-4 Ramachandran plot of the crystal structure of free rTxlN-1. ......................76
Figure 4-5 Wiring diagram of the secondary structure of rTxlN-1.............................77
Figure 4-6 Stereo view of the overall fold of molecule A of rTxlN-1. .........................78
Figure 4-7 Connolly surface with electrostatic surface potentials overlaid for molecule A of rTxlN-1. ..................................................................................78
Figure 4-8 Stereo view of superimposition of the three free rTxlN-1 molecules within the asymmetric unit.................................................................80
Figure 4-9 Visualization of the main-chain B-values for the three rTxlN-1 molecules. .................................................................................................81
Figure 4-10 Stereoview of interactions stabilizing the conformation of the canonical loop in molecule C of rTxlN-1.................................................................81
Figure 4-11 Superimposition of the three free rTxlN-1 molecules. .............................83
Figure 4-12 Two views, rotated by 90°, of the Connolly surface with electrostatic potentials overlaid for molecule A of rTxlN-1........................................85
Figure 4-13  Three conserved buried water molecules.............................................. 86
Figure 4-14  Stereo diagram of the hydrogen bond formed between an internal water molecule and Thr13O, Cys16N and Cys40O in molecule A of rTxln-1 ...... 87
Figure 4-15  Crystal contacts for molecule A, B and C of rTxln-1 ........................... 89
Figure 4-16  Stereoview of the packing of the three rTxln-1 molecules in the asymmetric unit.............................................................................................. 89
Figure 4-17  Interactions between molecule A and C in the free rTxln-1 structure. . 91
Figure 4-18  Visualization of the main-chain B-values for the three molecules of rTxln-1 in the asymmetric unit ................................................................. 92
Figure 4-19  Stereoview of the canonical loop in molecule C located between two A molecules. ........................................................................................................ 94
Figure 4-20  Expanded view of the crystal packing of the free rTxln-1 molecules. . 95
Figure 4-21  Stereoview of the superimposition of the three rTxln-1 molecules and aprotinin......................................................................................................... 98
Figure 4-22  Stereoview of the region between residues 40-45 (rTxln-1 numbering) after superimposition of molecule A of rTxln-1 and aprotinin............... 98
Figure 4-23  The Connolly surface and electrostatic potentials for molecule A of rTxln-1 and aprotinin................................................................. 100
Figure 4-24  Internal water molecules in rTxln-1 and aprotinin............................. 101
Figure 4-25  Amino acid sequence alignment of Txln-1, aprotinin and two aprotinin mutants (8PTI and 1LD6). The cysteine residues are highlighted in orange and the mutations in aprotinin in yellow. .............................................. 102
Figure 4-26  Stereo view of the position of the side-chain of Tyr37 in molecule A of rTxln-1 and Tyr35 in aprotinin................................................................. 104
Figure 4-27  Structural comparison of rTxln-1 and two aprotinin mutants.......... 105
Figure 4-28  Sequence alignment of rTxln-1 and the proteins mentioned in Table 4-7. ..................................................................................................................... 107
Figure 4-29  Structural comparison of rTxln-1 and four other proteins with the aprotinin-like fold. .......................................................................................... 109
Figure 4-30  Location of six conserved residues.................................................... 111
Figure 4-31  Superimposition of rTxln-1 and dendrotoxin K............................... 112
Figure 4-32  Replacement of one internal water molecule, (in aprotinin), by an asparagine side-chain in rTxln-1 and dendrotoxin K.............................. 112
Figure 4-33 Amino acid sequence alignment of the six textilinin molecules; Txln-1 to 6 .........................................................114
Figure 5-1 The crystal packing of the rTxln-1-trypsin complex ..................120
Figure 5-2 B-values (Å$^2$) for rTxln-1 in the complex. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values ..................................................122
Figure 5-3 B-values (Å$^2$) for trypsin in the complex. Glycine residues have no side-chain B-value .........................................................123
Figure 5-4 Visualization of the B-values for the rTxln-1-trypsin complex.......124
Figure 5-5 Ramachandran plot of the rTxln-1-trypsin complex .................125
Figure 5-6 Stereodiagram of the overall fold of the rTxln-1-trypsin complex..126
Figure 5-7 Interactions between rTxln-1 and trypsin ................................127
Figure 5-8 Stereo view of rTxln-1 bound in the active site of trypsin ...........129
Figure 5-9 Stereo view of the 2Fo-Fc electron density map contoured at 1σ for the rTxln-1-trypsin complex ......................................................130
Figure 5-10 The interface between rTxln-1 and trypsin ..............................130
Figure 5-11 Crystal contacts between a symmetry related rTxln-1 molecule and trypsin ...............................................................132
Figure 5-12 Two of the largest interfaces in the crystal packing of rTxln-1-trypsin complex .........................................................................................132
Figure 5-13 Stereo view of the superposition of free rTxln-1 and trypsin bound rTxln-1 .........................................................................................134
Figure 5-14 Stereo view of the canonical loop region of free and trypsin bound rTxln-1 .........................................................................................135
Figure 5-15 B-values (Å$^2$) for rTxln-1 bound to trypsin and the three free rTxln-1 molecules ...............................................................................136
Figure 5-16 Visualization of the B-values for rTxln-1 bound to trypsin...........137
Figure 5-17 Superposition of rTxln-1 and aprotinin bound to trypsin ..........138
Figure 5-18 Superimposition of trypsin in the rTxln-1-trypsin complex and the aprotinin-trypsin complex (2PTC) .....................................................139
Figure 5-19 Residues P3-P4' in rTxln-1 and aprotinin .................................140
Figure 5-20 The P1'-P4' residues in rTxln-1 and aprotinin ..........................141
Figure 5-21 The P4-P1 residues in rTxln-1 and aprotinin ............................142
Figure 5-22 rTxln-1 and aprotinin bound to trypsin .................................143
Figure 5-23 Variation in the location of the secondary loop .........................144
Figure 6-1 The crystal packing of the rTxln-1-microplasmin complex
Figure 6-2 Visualization of the main-chain B-values in the rTxln-1-microplasmin
Figure 6-3 B-values (Å²) for molecule A (blue) and B (pink) of rTxln-1 in complex
with microplasmin
Figure 6-4 B-values (Å²) for molecule A (blue) and B (pink) of microplasmin mol A
and B from respective complexes
Figure 6-5 Ramachandran plot of the rTxln-1-microplasmin complex
Figure 6-6 Stereodiagram of the overall fold of complex A of rTxln-1-microplasmin
Figure 6-7 Comparison of the binding modes of the two molecules of rTxln1 in the
asymmetric unit of the rTxln-1-microplasmin complex
Figure 6-8 Interactions between the canonical loop and secondary loop of rTxln-1
and the active site of microplasmin
Figure 6-9 Stereo view of the catalytic triad of microplasmin and interface residues
of rTxln-1 in complex A
Figure 6-10 Stereo view of the interaction between His603 and Glu606 of
microplasmin in complex A
Figure 6-11 The interface between rTxln-1 and microplasmin in complex A
Figure 6-12 Stereo view of the superimposition between complex A and B
Figure 6-13 Comparison of the binding modes of the two molecules of rTxln1 in the
asymmetric unit of the rTxln-1-microplasmin complex
Figure 6-14 Crystal contacts for complex A and B of rTxln-1-microplasmin
Figure 6-15 Crystal contacts between rTxln-1 in complex A and surrounding
molecules
Figure 6-16 Crystal contacts between rTxln-1 in complex B and surrounding
molecules
Figure 6-17 Crystal contact between Arg19 of rTxln-1 and Asp28 of rTxln-1 from
complex B
Figure 6-18 Crystal contacts between Glu41 and Arg644
Figure 6-19 Largest crystal contact between two rTxln-1 molecules
Figure 6-20 Difference in crystal contacts between rTxln-1 from complex A and B

xxi
Figure 6-21 Difference in crystal contacts between rTxln-1 from complex A and B ......................................................................................................................174
Figure 6-22 Superposition of microplasmin and microplasminogen. .........................176
Figure 6-23 Formation of the oxyanion hole in microplasmin........................................177
Figure 6-24 Formation of the specificity entrance of microplasmin. ..........................178
Figure 6-25 Complex of staphylokinase-microplasmin and rTxln-1-microplasmin. ..............................................................179
Figure 6-26 Complex of streptokinase-microplasmin and rTxln-1-microplasmin. 180
Figure 6-27 Stereo view of the different positions of the side-chain of the catalytic histidine. ..................................................................................................................181
Figure 6-28 Stereo view of a water molecule in the position of catalytically functional His603. ........................................................................................182
Figure 6-29 Stereo view of catalytic histidine forming an ionic bond with Glu606. ..................................................................................................................182
Figure 6-30 Stereo view of different positions of Asp646. ........................................183
Figure 6-31 Stereo view of a water molecule in the position of the catalytically functional Asp646. ...........................................................................................................183
Figure 6-32 Stereo view of inhibitor in the active site of microplasmin.......................184
Figure 6-33 Stereo view of interaction between Glu606 and bound molecule. ........184
Figure 6-34 Superimposition of all rTxln-1 structures................................................186
Figure 6-35 Visualization of the B-values in rTxln-1. .......................................................187
Figure 6-36 B-values (Å^2) for all rTxln-1 structures ....................................................187
Figure 6-37 Amino acid sequence alignment of microplasmin and trypsin.................189
Figure 6-38 Superposition of microplasmin and trypsin..............................................189
Figure 6-39 Differences between microplasmin and trypsin.......................................190
Figure 6-40 Interface between rTxln-1 and microplasmin or trypsin. .......................191
Figure 6-41 Interface between rTxln-1 and microplasmin or trypsin. .......................191
Figure 6-42 The docking angles between the bound rTxln-1 molecules when superimposing trypsin onto microplasmin. ................................................192
Figure 6-43 The difference in position of the secondary loop of rTxln-1 when superimposing trypsin onto microplasmin. ................................................193
Figure 6-44 The positions of the catalytic triads and rTxln-1 when microplasmin and trypsin are superimposed. ..............................................................................193
Figure 6-45 Comparison of bovine trypsin and Asn102 Rattus norvegicus trypsin mutant. .......................................................... 195
Figure 6-46 Comparison of bovine trypsin and porcine pancreatic elastase. .......... 196
Figure 6-47 Stereo view of the environment around the catalytic histidine. .......... 196
Figure 6-48 The location of aprotinin when superposing the trypsin molecule from the aprotinin-trypsin complex onto microplasmin. ........................................... 197
Figure 6-49 Two rTxln-1 molecules and one microplasmin molecule. ................. 198
Figure 6-50 Residues in interface between complex A and symmetry related rTxln-1. .................................................................................................................. 199
Figure 6-51 Residues in interface between complex B and symmetry related rTxln-1. ..................................................................................................................... 199
Figure 6-52 Possible dimer binding to microplasmin. ............................................. 200
Figure 7-1 Cumulative distribution of H. ............................................................... 206
Figure 7-2 The two ACII-4 molecules in the asymmetric unit. ......................... 209
Figure 7-3 Visualization of the B-values for the two molecules of ACII-4 in the asymmetric unit. .......................................................................................... 211
Figure 7-4 B-values (Å²) for molecule A and B of ACII-4. ..................................... 212
Figure 7-5 Ramachandran plot of the crystal structure of ACII-4. ....................... 213
Figure 7-6 Wiring diagram of the secondary structure of ACII-4. ....................... 215
Figure 7-7 Stereo view of the overall fold of molecule A of ACII-4 represented as a cartoon drawing. ....................................................................................... 215
Figure 7-8 Superposition of the two ACII-4 molecules. ..................................... 216
Figure 7-9 Stereo view of the seven ligands that coordinate to the Ca²⁺ ion form a distorted pentagonal bipyramid. ................................................................. 217
Figure 7-10 Stereo view of the two disulfide bonds securing the correct relative orientation of the Ca²⁺ binding loop. .......................................................... 217
Figure 7-11 The calcium ion bound to the Ca²⁺ loop. ........................................... 218
Figure 7-12 The Ca²⁺ binding loop. .................................................................... 219
Figure 7-13 The Ca²⁺ binding loop in ACII-4. ....................................................... 219
Figure 7-14 Ca²⁺ ions bound to PLA₂. ................................................................. 220
Figure 7-15 PLA₂ with Lys49. ............................................................................ 220
Figure 7-16 The polyethylene glycol molecules in ACII-4. ................................. 221
Figure 7-17 The hydrophobic channel leads into the active site in ACII-4........ 222
Figure 7-18 Hydrophobic entrance to the active site of ACII-4. ......................... 222
Figure 7-19 The 2Fo-Fc electron density map contoured at 1σ for His48, Asp49 and Asp92, two water molecules, one polyethylene glycol molecule and one Ca^{2+} ion for ACII-4. 223

Figure 7-20 The polyethylene glycol molecules and the catalytic diad. 223

Figure 7-21 The catalytic diad and the catalytic water molecule. 224

Figure 7-22 Stereo view of the two tyrosine residues stabilizing Asp92. 224

Figure 7-23 Comparison of active sites of ACII-4 and PLA\textsubscript{2} from porcine pancrease. 225

Figure 7-24 Comparison of active sites of ACII-4 and PLA\textsubscript{2} from Common Indian krait venom. 226

Figure 7-25 Comparison of bound molecules in the active site. 227

Figure 7-26 Comparison of the active site entrance. 228

Figure 7-27 The amino acid sequence of ACII-4. 228

Figure 7-28 Interfacial binding surface. 229

Figure 7-29 Sequence alignment of PLA\textsubscript{2} to show the elapid loop. 230

Figure 7-30 Comparison of group IA and group II PLA\textsubscript{2}. 230

Figure 7-31 Sequence alignment of PLA\textsubscript{2} proteins homologous to ACII-4. 231

Figure 7-32 Comparison of the structures of ACII-4 and the six top hits from a BLAST search. 232

Figure 7-33 Differences between ACII-4 and notexin. 234

Figure 7-34 Differences between ACII-4 and notexin. 235

Figure 7-35 Structural differences between ACII-4 and the BLAST top hits belonging to group IA. 235

Figure 7-36 The elapid loop in PLA\textsubscript{2} belonging to group IA and the pancreatic loop in PLA\textsubscript{2} belonging to group IB. 236

Figure 7-37 Structural differences between ACII-4 and the BLAST top hits belonging to PLA\textsubscript{2}s from group II. 237

Figure 7-38 Contacts between the surrounding atoms and molecule A of ACII-4. 238

Figure 7-39 The amino acid sequence of ACII-4. 238

Figure 7-40 Region 61-64 and Trp31 of ACII-4 participating in a hydrophobic interface. 239

Figure 7-41 Crystal contacts for ACII-4. 240

Figure 7-42 Crystal contacts for ACII-4. 240

Figure 7-43 Crystal contacts for ACII-4. 241
Figure 7-44 Crystal contacts for ACII-4................................................................. 241
Figure 7-45 Four views of molecule A of ACII-4......................................................... 242
Figure 7-46 Four views of molecule B of ACII-4............................................................. 242
Figure 7-47 AB dimer formation.................................................................................... 243
Figure 7-48 AB' dimer formation. .................................................................................. 243
Figure 7-49 Interactions in the AB interface .................................................................... 244
Figure 7-50 Stereo view of the interactions in the AB interface. ..................................... 245
Figure 7-51 Stereo view of the interactions in the AB' interface. ................................. 245
Figure 7-52 Superposing the PLA₂ from Common Indian krait (1TC8) onto each monomer of the two ACII-4 dimer formations shows the fitting of the fatty acid.............................................................. 246
Figure 7-53 Superposing the PLA₂ from Russell’s viper (1JQ9) onto each monomer of the two ACII-4 dimer formations shows the fitting of the protein inhibitor. ............................................................................................................ 246
Figure 7-54 Sequence alignment of PLA₂ with anticoagulant activity. ......................... 248
Figure 7-55 Comparison of residues with possible anticoagulant effect...................... 251
Figure 7-56 Sequence alignment of PLA₂ with anticoagulant activity. ......................... 252
Figure 7-57 Comparison of residues with possible anticoagulant effect...................... 254
Figure 7-58 Sequence alignment of PLA₂₅ with anticoagulant activity......................... 255
Figure 7-59 Comparison of the electrostatic surface potentials. ................................... 256
Figure 8-1 Two orthogonal views of the surface of the putative ACII-4 dimer............. 259

Figure C- 1 Phase diagram for protein solubility. ....................................................... 267
Figure C- 2 Protein crystallization vessels. ................................................................. 268
List of Tables

Table 1-1 Existing crystal structures of plasminogen/plasmin structural domains... 12
Table 1-2 Kinetic constants for the inhibition of plasmin by.................................17
Table 3-1 Experimental conditions that led to the successful crystallization of rTxln-1. .................................................................53
Table 3-2 Crystal parameters and data collection statistics for free rTxln-1...........56
Table 3-3 Experimental conditions that led to the successful crystallization of the rTxln-1-trypsin complex ..................................................57
Table 3-4 Crystal parameters and data collection statistics for the rTxln-1-trypsin complex .................................................................59
Table 3-5 Experimental conditions that led to the successful crystallization of the rTxln-1-microplasmin complex...........................................60
Table 3-6 Crystal parameters and data collection statistics for the microplasmin-rTxln-1 complex ..............................................................61
Table 3-7 Crystal parameters and data collection statistics for ACII-4. ..............64
Table 4-1 Molecular replacement solution: orientation of a single molecule ..........71
Table 4-2 Molecular replacement solution: orientations of two molecules ..........71
Table 4-3 Molecular replacement solution: orientations of three molecules ..........71
Table 4-4 Refinement statistics for rTxln-1 ..........................................................74
Table 4-5 Mean B-values (Å²) for the three rTxln-1 molecules in the asymmetric unit .........................................................................................75
Table 4-6 Ramachandran plot statistics for the three free rTxln-1 molecules....76
Table 4-7 Hits from BLAST search versus the PDB ...........................................106
Table 5-1 Refinement statistics for the rTxln-1-trypsin complex .......................121
Table 5-2 Mean B-values (Å²) for the rTxln-1-trypsin complex .........................122
Table 5-3 Ramachandran plot statistics for the rTxln-1-trypsin complex ..........124
Table 6-1 Refinement statistics of the rTxln-1-microplasmin complexes ..........150
Table 6-2 Mean B-values (Å²) for the rTxln-1-microplasmin complexes in the asymmetric unit ........................................................................152
Table 6-3 Ramachandran plot statistics for the two rTxln-1-microplasmin complexes ........................................................................154
Table 6-4 Table of rmsd values between the rTxln-1 structures, Cα (all atoms) ....185
Table 7-1  PDB entries with the best E-values ..................................................... 208
Table 7-2  Prepared search models with scores .................................................... 208
Table 7-3  Refinement statistics of ACII-4 crystal structure ................................. 210
Table 7-4  Mean B-values (Å²) for the two ACII-4 molecules. ............................... 211
Table 7-5  Ramachandran plot statistics for the two ACII-4 molecules ................. 212
Table 7-6  Hits from BLAST search versus the PDB .......................................... 231

Table A-1  Family Elapidae ................................................................................ 263
Chapter 1
Introduction

1.1 Background

The research described in this thesis covers the determination of the three-dimensional structures by X-ray crystallography of two proteins isolated from Australian snake venoms; textilinin-1 (Txln-1), a plasmin inhibitor from the venom of the Australian Common Brown snake *Pseudonaja textilis*, and ACII-4, an anticoagulant phospholipase A\textsubscript{2} from the venom of the Australian King Brown snake, *Pseudechis australis*. The structure of recombinant Txln-1 (rTxln-1) was determined as the free molecule and in complex with trypsin and microplasmin. The structure of ACII-4 was determined as the free molecule. No crystal structures of any Australian snake venom component have been reported to date thus this study opens the way to rational structural investigations of the pharmacological effects of these molecules.

This chapter includes considerations of Australian snake venom components, blood coagulation and fibrinolysis pathways, especially the role and structures of plasmin. It incorporates a review of serine proteases and their mechanism of action and inhibition by canonical serine protease inhibitors from the Kunitz/BPTI family. A discussion on the mechanism of phospholipase A\textsubscript{2} molecules is presented.

1.2 Australian snake venoms and their components

There are seven families of Australian snakes: file snakes, pythons, colubrids, elapids, sea snakes, sea kraits and blind snakes (Tu, 1977; C.-Y. Lee, 1979; Tu, 1991; Swan, 1998). Members of three of these families, elapids (*Elapidae*), sea snakes (*Hydrophiidae*) and sea kraits (*Laticaudidae*) are considered venomous. A total of 81 out of the 143 species of snakes in Australia belong to the elapid family, with another 99 elapid species existing worldwide (Swan, 1998).

Australian elapid snakes are amongst the most venomous in the world (Broad *et al.*, 1979; Sutherland and Tibballs, 2001). Much of their toxicity arises from the presence of powerful prothrombin-activating proteins, and potent pre- and post-synaptic
neurotoxins. Other components present in Australian snake venoms include phospholipase A₂ (PLA₂) proteins, serine proteinase inhibitors, L-amino acid oxidase and natriuretic peptides (Fry, 2005; St Pierre et al., 2005). Several of these are pharmacologically important and known to affect the mammalian blood hemostatic mechanism via coagulant, anticoagulant, and fibrinolytic actions. A unique feature of the procoagulant properties of venoms from Australian elapid snakes is that they are limited to prothrombin activation and contain no thrombin-like enzymes. These prothrombin activators contain a component that is structurally and functionally similar to that of mammalian coagulation factor Xa (Chester and Crawford, 1982). The prothrombin activators can produce disseminated intravascular coagulation (DIC) on envenomation of a human or prey species. The activated prothrombin is converted to thrombin, and thrombin in turn converts fibrinogen to fibrin initiating widespread clot formation. Extensive activation of thrombin can lead to depletion of fibrinogen, (defibrinogenation) with the net effect of blood loss (Fry, 1999; Kini et al., 2001; Sutherland and Tibballs, 2001).

The Australian Common Brown snake, *Pseudonaja textilis*, is native to Australia and is found primarily in the eastern parts of the mainland. Its lethality is caused by both neurotoxic and procoagulant compounds. The main component in its venom that causes the rapid coagulation of blood is a potent prothrombin activator complex named pseutarin C (Masci et al., 1988; Rao and Kini, 2002). This protein is a complex of two main components and constitutes about 30% of the total venom protein. One of the components is a large multisubunit non-enzymatic protein similar to factor V found in mammalian blood (Rao et al., 2003), and the other is a factor Xa-like enzyme that cleaves prothrombin to form thrombin (Masci et al., 1988). Txln-1 which is a Kunitz-type serine protease inhibitor with a high specificity for plasmin (Willmott et al., 1995) represents approximately 2% of the total venom protein, and was isolated by Dr Paul Masci (Masci, 2000).

Phospholipases are found within most snake venoms regardless of genera and display homologous primary sequences to their mammalian counterparts which are involved in inflammatory responses, cancer metastasis, cell growth and signaling cascades. In contrast to mammalian enzymes, phospholipases from venom may be highly toxic to other animals, having a plethora of physiological effects. The neurotoxic, haemotoxic
and myotoxic effects of Australian elapid envenomations are often the result of the activity of a number of phospholipases within the venom. This variation in activity, even within multiple isoforms from one snake, is observed despite the high degree of conservation in their primary sequence, and in some cases appears to be independent of the phospholipase catalytic activity of the protein (Fry, 1999). ACII-4 is a phospholipase isolated from the Australian King Brown snake, *Pseudechis australis*, and has shown to exhibit a strong anticoagulant activity but low phospholipase activity (Masci, 2000).

1.3 Blood clotting

Cross-linked fibrin clots are formed by a complex enzymatic cascade, in which the activated form of one protein catalyzes the activation of the next. Blood coagulation can be activated both by the “extrinsic” pathway and the “intrinsic” pathway. As shown in Figure 1-1, both extrinsic and intrinsic pathways lead to the formation of factor Xa. Factor Va form a complex with factor Xa (the prothrombinase complex) which cleaves prothrombin to produce thrombin. Thrombin is the enzyme that converts soluble fibrinogen to insoluble fibrin clots which are subsequently covalently crosslinked by the action of factor XIIIa. Blood coagulation is vital for the containment of bodily fluids. Equally important is the reversal of this process, fibrinolysis, which must take place after tissue healing. The processes of coagulation and fibrinolysis are tightly controlled, to achieve haemostasis (Biggs and Rizza, 1984; Furie and Furie, 1992).

Of major importance in fibrinolysis is plasminogen. Plasminogen, a normal component of blood plasma, is deposited on the fibrin strands within a blood clot. The activators of this protein diffuse into the clot and release plasmin, a serine protease. Plasmin then breaks down the fibrin meshwork, and as a result the clot dissolves (Rang *et al.*, 1999; Castellino, 2004).

There is a fine balance between hemorrhage and thrombosis, and the clotting process must be precisely regulated to not produce unrestrained excess of active enzymes that can cause these pathological conditions. Inhibitors of the enzymes in the blood coagulation cascade control hemorrhage and the termination of clotting.
1.4 Serine proteases

Serine proteases are proteases with an active site serine residue and a catalytic mechanism that includes an acyl-enzyme intermediate (Blow et al., 1969; Robertus et al., 1972; Warshel et al., 1989; Liao and Remington, 1990). Serine proteases have evolved to function in many biological processes. These processes in mammals include fertilization, several pathways of differentiation and development, digestion of protein, blood clotting, fibrinolysis and the complement system in humoral immunity. Their activities are related to many diseases including emphysema, asthma, arthritis, skin disorders and cancerous tumor growth (Unkeless et al., 1973; Morioka et al., 1987; He et al., 1989).
Serine proteases are often synthesized as somewhat larger inactive precursors known as zymogens. Trypsin is a serine protease found in animals as a digestive enzyme secreted by the exocrine pancreas into the small intestine (Kühne, 1876; Northrop and Kunitz, 1931; Neurath and Zwilling, 1986). The pH optimum for activity is approximately 8. Bovine trypsin is a 223 amino acid residue long protein. Trypsinogen is the inactive precursor of trypsin which is activated by hydrolysis of a hexapeptide at the N-terminus by specific cleavage between lysine 15 and isoleucine 16 (chymotrypsinogen numbering)(Maroux et al., 1971). This activation results in a conformational change and formation of the oxyanion hole and specificity pocket which are both important for catalytic activity (Kossiakoff et al., 1977; Marquart et al., 1983; Halfon et al., 2004).

The serine proteases exhibit different substrate specificities, which are related to amino acid substitutions in the various enzyme subsites interacting with substrate residues. The nomenclature to describe the interaction of a substrate with a protease considers the amino acid residues of the polypeptide substrate, P, binding to enzyme subsites, S, of the active site. The amino acid residues on the N-terminal side of the scissile bond are numbered P3, P2, P1 and those residues on the C-terminal side are numbered P1', P2', P3' (Figure 1-2). The P1 and P1' residues are those residues located either side of the scissile bond (Schechter and Berger, 1967). Substrate specificity is critically dependent on the shape and size of the S1 site (Schechter and Berger, 1967; Perona and Craik, 1995; Perona and Craik, 1997).
For serine proteases a common structural theme has evolved independently on several separate occasions (Robertus et al., 1972; Liao and Remington, 1990; Ollis et al., 1992). Three residues, one serine, one histidine and one aspartate, form the “catalytic triad” are required to carry out catalysis. The side-chain of the aspartic acid, with a pKa of ~4, is deprotonated and negatively charged. In general the imidiazole ring of histidine has a pKa of ~6 and is not a strong base. The hydroxyl group of the side-chain of the serine normally has a pKa of ~13 and is protonated. A basic mechanistic scheme for the catalysis reaction has generally been agreed upon (Blow et al., 1969; Warshel et al., 1989). Initially a non-covalent Michaelis complex is formed when the serine protease recognizes an appropriate proteolytic cleavage site in a substrate. The imidazole ring deprotonates the hydroxyl group of the serine side-chain. The negatively charged side-chain of the aspartic acid stabilizes the positively charged histidine side-chain or accepts a proton from the imidazolium ion (Warshel et al., 1989). The deprotonated serine is a strong nucleophile and attacks the carbonyl carbon on the P1 residue forcing the carbonyl oxygen to accept an electron. Consequently the hybridization state of the carbonyl carbon changes from sp² to sp³ resulting in the formation of a tetrahedral intermediate. The carbonyl oxygen can accommodate the electron pair of the carbon-oxygen double bond since the carbonyl oxygen of the scissile peptide moves deeper into the active site to occupy a previously unoccupied enzyme subsite with stabilizing interactions,
designated the oxyanion hole (Robertus et al., 1972). Thus, formation of the high energy tetrahedral intermediate is achieved not only because of the covalent bond formed between the serine side-chain oxygen in the catalytic triad and the carbonyl carbon in the substrate, but also because of the substrate destabilization by the oxyanion hole (Stryer, 1996; Voet et al., 1998). During the second covalent stage or deacylation, the acyl-enzyme intermediate is hydrolyzed by a water molecule to release the peptide and to restore the hydroxyl of serine in the enzyme (Barrett et al., 1998). The general mechanism of catalysis is given in Figure 1-3.
Figure 1-3 The catalytic mechanism of the serine proteases.
The first stage in the basic mechanism of serine proteases is an acylation reaction;
(a) Michaelis complex, (b) tetrahedral intermediate and (c) acyl-enzyme intermediate.
The second stage is a deacylation reaction; (d) Water attacks the acyl-enzyme intermediate,
(e) tetrahedral intermediate and (f) the restoration of the catalytic triad and the departure of
the second product
1.5 Mechanism of inhibition of serine proteases

Three classes of protein serine protease inhibitors have evolved: canonical, non-canonical and serpin inhibitors (Krowarsch et al., 2003). A canonical inhibitor mimics the substrate but no product formation is observed during reaction time (Bode and Huber, 1992). Most serine protease inhibitors are canonical inhibitors acting on their protease by what is known as the standard mechanism (M. J. Laskowski and Kato, 1980). All standard mechanism inhibitors have an exposed substrate-like canonical inhibitory loop region with a single reactive site peptide bond (M. J. Laskowski and Qasim, 2000). The inhibitory loops in the different canonical inhibitors possess a common main-chain conformation, though, besides this loop, each family of canonical inhibitors has its own overall three-dimensional structure (Bode and Huber, 1992). There are at least 18 non-homologous families of canonical inhibitors (Bode and Huber, 1991; M. J. Laskowski and Qasim, 2000; Ibrahim and Pattabhi, 2004). For example the inhibitor, aprotinin (bovine pancreatic trypsin inhibitor, BPTI), belongs to the Kunitz/BPTI family. Aprotinin has a structure complementary to the active site of trypsin. The fit is tight with a strong interaction and a long half-life for the complex (months). The Kunitz/BPTI family of serine protease inhibitors including Txln-1 is expected to act on their proteases by the standard mechanism. The carbonyl carbon of the scissile peptide bond in a canonical inhibitor is distorted about half way towards the tetrahedral intermediate (Read and James, 1986). The carbonyl carbon is much further away from the side-chain oxygen of the catalytic serine residue of the enzyme than a covalent single carbon-oxygen bond, but is shorter than a van der Waals contact. Contributing to the partial tetrahedral distortion is the attraction between the carbonyl oxygen of the inhibitor and the oxyanion hole of the enzyme. The carbonyl carbon in a substrate is fully configured as a tetrahedral intermediate and positioned for a nucleophilic attack by the side-chain oxygen of the catalytic serine residue, but in the stable inhibitor-enzyme complex the attack has not yet taken place (M. J. Laskowski and Kato, 1980).

1.6 The properties of plasmin and its role in blood clotting

Blood clots are formed by a series of zymogen activations. Several of the activated enzymes in the blood coagulation cascade are serine proteases, e.g. activated factor X, thrombin and plasmin. Fibrinolysis is initiated to balance the overall blood
clotting system, when the coagulation cascade is activated. Plasmin, the enzyme central to fibrinolysis, breaks down the fibrin clot formed by the coagulation cascade of enzymes.

**Figure 1-4** Enzymes and pathways involved in fibrinolysis

Plasminogen is synthesized in the liver, secreted into the plasma and bonds to fibrin strands within a blood clot. The activators of this enzyme diffuse into the clot and cleave plasminogen to release plasmin. Plasmin breaks down the fibrin meshwork, and as a result the clot dissolves (Rang *et al.*, 1999; Castellino, 2004). Plasmin is also involved in a number of processes unrelated to clot lysis, such as the inflammatory response, cell recruitment, and extracellular matrix remodeling (Unkeless *et al.*, 1973; Mullins and Rohrlich, 1983; He *et al.*, 1989).

Plasminogen is a single-chain protein of 791 amino acids. It contains seven structural domains: one N-terminal growth factor like domain, five homologous kringle domains of approximately 80 residues each, and a C-terminal serine protease catalytic domain (542-791) (Figure 1-5). Its catalytic domain is referred to as microplasminogen when inactive and microplasmin once activated.
Figure 1-5  Domain structure of plasminogen.
It consists of one N-terminal domain (N), five kringle domains (K1-K5) and a catalytic serine protease domain.

Plasminogen activation by tissue plasminogen activator (tPA) on fibrin and by urine plasminogen activator (uPA) on cell surfaces are the major mechanisms of defense against thrombosis in human. Both of these proteins are serine proteases that cleave the Arg561 - Val562 peptide bond in plasminogen to generate plasmin.

Two plasminogen activators have been isolated from bacteria, staphylokinase produced by *Staphylococci*, and streptokinase from hemolytic strains of *Streptococci*. These plasminogen activators are not enzymes themselves but instead are cofactors, form 1:1 complexes with either plasminogen or plasmin and acquire the ability to activate plasminogen. Staphylokinase (SAK) is the smallest known single domain plasminogen activator, consisting of 136 residues with a single domain folded into a compact and flattened structure. Streptokinase (STK) is a 414 residue protein with the structure made up of three domains.

No crystal structure of the entire plasminogen or plasmin molecule has been reported. However the structures of a number of the domains and complexes have been determined. These include kringle domains 1, 2, 4, 5, microplasminogen by itself, microplasminogen in complex with the α domain of staphylokinase, microplasmin in complex with staphylokinase and microplasmin in complex with streptokinase (Table 1-1).
Table 1-1  Existing crystal structures of plasminogen/plasmin structural domains.

<table>
<thead>
<tr>
<th>Domain</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kringle 1</td>
<td>Wu et al., 1994; Mathews et al., 1996</td>
</tr>
<tr>
<td>Kringle 2</td>
<td>Marti et al., 1999 (NMR)</td>
</tr>
<tr>
<td>Kringle 4</td>
<td>Mulichak et al., 1991; Stec et al., 1997</td>
</tr>
<tr>
<td>Kringle 5</td>
<td>Chang et al., 1998</td>
</tr>
<tr>
<td>Human microplasminogen</td>
<td>X. Wang et al., 2000</td>
</tr>
<tr>
<td>Human microplasminogen in complex with the α domain of staphylokinase</td>
<td>Wakeham et al., 2002</td>
</tr>
<tr>
<td>Human microplasmin in complex with staphylokinase</td>
<td>Parry et al., 1998</td>
</tr>
<tr>
<td>Human microplasmin in complex with streptokinases</td>
<td>X. Wang et al., 1998</td>
</tr>
</tbody>
</table>

Kringles are sequences of about 80 amino acids arranged in a triple-loop tertiary structure rigidly stabilized by three disulfide bridges. Kringle domains 1 and 4 of plasminogen contain a lysine binding site that mediates plasminogen binding to lysine residues of fibrin and cell membrane proteins. On binding to fibrin and cells, conversion of plasminogen to plasmin is facilitated. Thus, the kringle domains have an important role in regulation of this system and considered to play a factor in e.g. extracellular matrix degradation (Carmeliet and Collen, 1998), stimulate the release of inflammatory mediators (Syrovets et al., 2001) and cell migration (Tarui et al., 2002).

The crystal structure of microplasmin shows the conformational change of the catalytic domain upon activation (X. Wang et al., 1998; X. Wang et al., 2000). This takes place when the proteolytic target bond Arg561 – Val562 is cleaved. Two polypeptide chains, one 229 amino acids and one 20 amino acids in length, connected by two disulfide bonds are produced as the result of cleavage. After cleavage the released Val562 moves about 12 Å inwards the enzyme. Structural rearrangements take place to form the functional specificity pocket and oxyanion hole (Castellino, 2004).

Microplasmin has a similar overall structure to bovine trypsin with an amino acid sequence alignment of these two proteins showing 39% identity (Figure 1-6). Like
trypsin, plasmin cleaves peptide bonds directly after arginine or lysine side-chains. The overall structure of the catalytic domain of serine proteases consists of two six stranded β-barrels with the active site in a cleft between the barrels (Voet et al., 1998).

| Human microplasmin: AAPSFDCGPVEPKCPGVRVGGCVAHPHSWPQWQSLRTGFHMFCG |
|---------------------------|---------------------------|
| Bovine trypsin: IVGYTCGANTVQPQWQSLNS--GYHCFCG |

GTLISPEWVLTAHECLEKSPSSYYKVLGHEVNLEPHQIEIEVSRLEPT------RKDIALLK

GSLINSQVVSAHCYK------SGIQVRLEDNINVEGNEQFISASKSIIVHPSNNTNNDMLIK

LSSPAVITDKVIPACLSKPVYVVRADTECFITGWGETQ--GFAGLLEKAEQLPVIEKVCNYEFLN

LKSAAASLVASISLPTSCASAG--TQCLISGWGNTKSSGTSYPDVGLKCLKAPILSDSCK--SAYP

GRVQSTELCAGHLAGTSCQGSSGPVLFECDFKYYILQGVTSWGLGCARPNKPGVYVRVRFVTWIE

GVMRNN

QTIASN

**Figure 1-6** Human microplasmin aligned with bovine trypsin.
The catalytic triad of His57, Asp102, and Ser195 in bovine trypsin and the catalytic triad of His603, Asp646 and Ser741 in microplasmin are shadowed. These two proteins show 39% sequence identity.

### 1.7 Aprotinin in therapeutic use to regulate blood loss

Aprotinin (trademark name Trasylol) belongs to the Kunitz/BPTI family and is a serine protease inhibitor which is used as a therapeutic agent to prevent blood loss and reduce the need for blood transfusion in patients undergoing surgical operations to replace a section of coronary artery (Filippovich et al., 2002). Using blood from another donor involves risks with incompatibility and transfusion of transmitted infections. Aprotinin is a broad-spectrum serine protease inhibitor, and when in use in humans inhibits plasma kallikrein, an enzyme involved in inflammation and blood coagulation, and then inhibits plasmin with tight binding kinetics. There is a risk of hypersensitivity resulting in severe allergic reaction and a drop in blood pressure and breathing difficulties in patients who are exposed to aprotinin. Vein-graft occlusion associated with myocardial infarction and thrombosis can also occur following the use of this inhibitor (Engles, 2005; Sodha et al., 2006).

An alternative kinetically distinct inhibitor of plasmin is rTxln-1 isolated from the venom of the Australian Common Brown snake, *Pseudonaja textilis*. This molecule
could potentially have fewer side effects than aprotinin since it is a more specific inhibitor and binds to plasmin with reversible binding kinetics. This inhibitor is being studied as a potential anti bleeding agent in the Venomics research group at the University of Queensland.

1.8 The properties of textilinin-1

Txln-1 is a 59 amino acid protein isolated from the venom of the Australian Common Brown snake, *Pseudonaja textilis*. This inhibitor has 45% amino acid identity with aprotinin. The six cysteine residues in Txln-1 and aprotinin can be aligned (Figure 1-7).

![Figure 1-7](image)

**Figure 1-7** Txln-1 aligned with aprotinin. The cysteine residues are shadowed.

Txln-1 has been extensively studied due to its anti-fibrinolytic properties, which were discovered when applied in a mouse tail vein bleeding model experiment (Filippovich *et al.*, 2002). With this in mind Txln-1 was cloned and the recombinant protein was produced. The structure of recombinant Txln-1 (rTxln-1) was pursued to determine its binding mode and its structural features that lead to its specificity.

When comparing rTxln-1 and aprotinin, differences exist in their amino acid composition, overall charge, binding specificity, rate of inhibition, and reversibility of inhibition. The number of amino acids only differs by one residue with rTxln-1 having two extra amino acids at the amino terminus, but one fewer at the carboxy-terminus. rTxln-1 is an acidic molecule with a pI of 4.4 (Filippovich *et al.*, 2002) while aprotinin is basic with a pI of 8.9 (Gebhard *et al.*, 1986). The reason for this difference is that there are eleven negatively charged and seven positively charged amino acids in rTxln-1 while in aprotinin there are only four negatively charged and ten positively charged amino acids. Only four of the charged amino acids are in identical positions in the two inhibitors. There are six changes from a charged amino acid in rTxln-1 to one that is hydrophobic in aprotinin, while two other changes are from charged to polar side-chains. Moreover three polar and three hydrophobic amino acids in rTxln-1 are changed to six charged amino acids in aprotinin. In two
cases there is a complete change of charge between the proteins. The lysine residue in aprotinin that binds to the S1 pocket in trypsin is aligned with an arginine in rTxln-1.

A total of six textilinin (Txln-1 to Txln-6) sequences (i.e. cDNAs with close sequence homology to textilinin-1) have been identified by cloning from venom gland cDNA (Filippovich et al., 2002). All have a highly conserved propeptide, serine/lysine propeptide cleavage site and can also be aligned to match up with the characteristic six cysteine residues found in Kunitz/BPTI-type serine protease inhibitors (Figure 1-8). The propeptide sequence is hydrophobic, which suggests that it facilitates secretion of the protein prior to cleavage. Conservation of a serine lysine sequence at the cleavage site suggested that all six peptides are processed by the same protease, releasing the twenty-four amino acid propeptide.

All six mature textilinin peptides are 59 amino acids in length (Masci et al., 2000) have moderately conserved sequences throughout their mature peptide regions. Their sequence identities range from 67 to 93%, with the identity between Txln-1 and Txln-2 being the highest. Txln-2 was found in Pseudonaja textilis venom by monitoring plasmin inhibitory activity but the other textilinins were not found by this test (Filippovich et al., 2002). Hence, they were either not present (or present only in minute amounts) or have different inhibitory specificity to Txln-1 and Txln-2.

**Figure 1-8** Sequences of Txln-1 to 6.
The 24 amino acid pro-peptide, the 59 amino acid mature peptide, the cysteine residues and the conserved regions are shadowed. The expected disulfide bond connectivities in Txln-1 are also shown. The sequence identities between Txln-1 and 2, 3, 4, 5 and 6 (aligned with BLAST (blastp) and ClustalW) are 93%, 67%, 71%, 67% and 71%, respectively.

The greatest variation across the six sequences is observed in the segment between amino acids 17 and 21, the region principally responsible for making direct contacts.
to serine proteases (by comparison with aprotinin). An arginine residue in position 17 (the residue that binds in the S1 pocket in serine proteases) is common to Txln-1 and Txln-2, while the other four textilinins have an asparagine, lysine, glutamate or aspartate in this position.

1.8.1 Textilinin-1's mechanism of inhibition

Txln-1 was isolated from the Australian King Brown snake because of its known plasmin inhibitory properties (Willmott et al., 1995; Filippovich et al., 2002). In an earlier study, aprotinin was shown to inhibit plasmin with a slow-onset inhibition consistent with the two-stage reversible mechanism described as (Morrison and Walsh, 1988);

\[
\begin{align*}
E + I & \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} EI & & \overset{k_2}{\underset{k_{-2}}{\rightleftharpoons}} EI^* \\
\end{align*}
\]

where enzyme E and inhibitor I form complex EI with a rate constant \(k_1\). The complex EI can go in two different directions. It can dissociate to E and I with a rate constant \(k_{-1}\), or it can proceed to form the product EI* with a rate constant \(k_2\). EI* may dissolve with the rate constant \(k_{-2}\).

Between aprotinin and plasmin, formation of the initial complex EI is fast, with a relatively loose binding, compared to the slow transition from EI to the very stable complex EI* (Willmott et al., 1995; Filippovich et al., 2002; Flight et al., 2005; Flight et al., 2007). Recent kinetic studies in our laboratory have shown that rTxln-1 inhibition of plasmin also follows the slow tight binding mechanism of equation 1 and reduces blood loss when injected intravenously in an animal model (mouse tail vein excision) (Masci et al., 2000; Flight et al., 2007). Indeed aprotinin is greater than 15 times more efficient than rTxln-1 in inhibiting plasmin, and has a broader specificity compared to rTxln-1 (Table 1-2) (Masci et al., 2000; Filippovich et al., 2002; Flight et al., 2005; Flight et al., 2007). The rate of recovery of plasmin is 18 times faster with rTxln-1 than aprotinin (Flight et al., 2007). Weaker binding of plasmin by rTxln-1 could result in greater clinical efficacy through a reduction in the incidence of vein-graft occlusion and thrombosis compared to aprotinin when used in vascular surgery (Masci et al., 2000).
Table 1-2  Kinetic constants for the inhibition of plasmin by aprotinin and rTxln-1.

<table>
<thead>
<tr>
<th></th>
<th>Aprotinin</th>
<th>rTxln-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_i$ nmol/L</td>
<td>9 ± 3</td>
<td>14 ± 3</td>
</tr>
<tr>
<td>$K_i^*$ nmol/L</td>
<td>0.032 ± 0.006</td>
<td>0.49 ± 0.02</td>
</tr>
<tr>
<td>$k_2/K_i$ s⁻¹ M⁻¹</td>
<td>2.3 ± 0.18 ·10⁶</td>
<td>1.47 ± 0.10 ·10⁶</td>
</tr>
<tr>
<td>$k_2$ 1/s</td>
<td>0.021 ± 0.004</td>
<td>0.020 ± 0.006</td>
</tr>
<tr>
<td>$k_{-2}$ 1/s</td>
<td>4.80 ± 0.54 ·10⁻⁵</td>
<td>8.50 ± 0.11 ·10⁻⁴</td>
</tr>
<tr>
<td>$t_{\frac{1}{2}}$ (on) Sec</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>$t_{\frac{1}{2}}$ (off) Min</td>
<td>240</td>
<td>14</td>
</tr>
</tbody>
</table>

In Table 1-2, $K_i$ is the dissociation constant for the initial loose complex EI, and $K_i^*$ is the dissociation constant for the final tight complex EI*. The $k_2$ and $k_{-2}$ are the rate constants as defined by Equation 1.1. The half-life; $t_{\frac{1}{2}}$ (on) is the half-life for inhibition and $t_{\frac{1}{2}}$ (off) is the half-life for recovery after removal of inhibitor calculated as $-\ln(1/2)/k_{-2}$.

The $k_2$ constant for rTxln-1 is 18 times larger than the $k_2$ for aprotinin, and reflects that the EI* complex of aprotinin and plasmin holds tighter together than the same complex of rTxln-1 and plasmin, and that rTxln-1 has a much higher release. The forward reactions are the same in rTxln-1 and aprotinin. The similarities in the forward reactions are reflected in similar $k_2/K_i$ constant.

1.9 Phospholipases

PLA₂s are widely distributed in nature and can be found both extracellularly and intracellularly. The presence of PLA₂s in pancreatic juices and in cobra venom was discovered around the beginning of the twentieth century and the roles of these secreted PLA₂s in digestion and envenomation were investigated extensively (Bókay, 1877; Kyes, 1903; Ogawa, 1936). Evidence began to appear in the last twenty years that PLA₂s are present in most types of cells and that they are involved in many different cellular functions including lipid metabolism and membrane homeostasis (Dennis, 2000).
Generally two different PLA₂ molecules have evolved: low molecular weight extracellular PLA₂s (13–16 kDa) and high molecular weight cytosolic PLA₂s (>40 kDa) (Dennis, 1994; Mounier et al., 2001; Kini, 2006). Division of PLA₂s into several sub-classes has been suggested based on criteria that include phospholipase activity, Ca²⁺ dependence and structural features. These structural features comprise for example the number of disulfide bonds and loop extensions (Dennis, 1994, 1997, 2000). The extracellular PLA₂s occur abundantly in mammalian pancreatic juice and in the venoms of snakes and insects (Roberts, 1996).

In humans, PLA₂s are important in systems such as signal transduction and eicosanoid production. Through these systems, the actions of PLA₂s affect a wide range of human physiological functions and diseases including asthma and allergy, the initiation and maintenance of parturition, blood clotting, atherosclerosis, sepsis, asthma, inflammatory bowel disease, arthritis and other inflammatory diseases (Dennis, 2000; Mounier et al., 2000). It is mainly the human cytosolic PLA₂s, which are often membrane associated, that are involved in platelet aggregation and generation of inflammatory mediators (Dennis, 1994; Nalefski et al., 1994; Perisic et al., 1998).

PLA₂s hydrolyse the sn-2 ester bond of phospholipids and release lysophospholipids and fatty acids (van Deenen and de Haas, 1964) (Figure 1-9).

![Figure 1-9](image)

Figure 1-9 Phospholipid hydrolysis by phospholipase A₂. R₁ and R₂ represent fatty acyl moieties. X represent polar head groups.

It has been established in bovine pancreatic PLA₂s (Dijkstra et al., 1981a; Yu et al., 1993; Arni and Ward, 1996; Steiner et al., 2001) and porcine pancreatic PLA₂s (Dijkstra et al., 1983; van den Bergh et al., 1989; Finzel et al., 1991) that calcium is an essential cofactor for catalytic active PLA₂. The Ca²⁺ ion is bound to a region
called the Ca\(^{2+}\) binding loop. The catalytic mechanism of PLA\(_2\)s has been well characterized (Fohlman \textit{et al.}, 1979; Dijkstra \textit{et al.}, 1981a; Keith \textit{et al.}, 1981; Brunie \textit{et al.}, 1985; Scott \textit{et al.}, 1990; White \textit{et al.}, 1990; Janssen \textit{et al.}, 1999). The mechanism behind the phospholipase activity is similar to the mechanism of the serine protease catalytic triad but instead of a catalytically active serine residue, PLA\(_2\) has a water molecule acting as a nucleophile and a “catalytic diad” composed of a histidine which accepts/donates a proton and an aspartate which stabilizes the charge developed on the imidazolium ion (Volwerk \textit{et al.}, 1974; Kumar \textit{et al.}, 1994). The carboxylate group of the aspartate is stabilized by two tyrosine residues phenolic hydroxyls (Scott \textit{et al.}, 1990). A PLA\(_2\) substrate binds in a hydrophobic channel where the active site is situated. The Ca\(^{2+}\) ion is the counterpart of the oxyanion hole in serine proteases aiding in the catalysis and also stabilizing the negative charge that arises during the tetrahedral formation at a stage in the process of the catalytic reaction (Figure 1-10) (Verheij \textit{et al.}, 1980b; Thunnissen \textit{et al.}, 1990b; Li and Tsai, 1993; Kumar \textit{et al.}, 1994; Berg \textit{et al.}, 2001).
Figure 1-10 The catalytic mechanism of PLA₂
(a) The “catalytic diad” and a water molecule acting as a nucleophile activates a second water molecule to attack the scissile carbonyl carbon. (b) The Ca²⁺ stabilizes the tetrahedral intermediate. (c) The tetrahedral intermediate decomposes to yield the products.

The PLA₂s from venom possess, in addition to their phospholytic activity, a wide variety of pharmacological activities such as (presynaptic and/or postsynaptic) neurotoxicity, myotoxicity, initiation and/or inhibition of platelet aggregation, haemolytic, anticoagulant, convulsant, hypotensive, cardiotoxic and oedema-inducing effects (Kini and Evans, 1989; Arni and Ward, 1996). An individual PLA₂ can exhibit one or more of these effects.
1.10 The properties of ACII-4

There is a high abundance of phospholipases in the venom of Australian elapids. These molecules have a large number of actions apart from the enzymatic cleavage of phospholipids. Two examples of actions are binding to molecules essential to the coagulation process, thereby acting as anticoagulants (Fry, 1999), or acting as a neurotoxin by binding to the extracellular face of calcium channels (Brown et al., 1987).

Three phospholipases from the Australian King Brown snake, *Pseudechis australis*, have been isolated by our group. They are speculated to be the same molecules as the previously isolated PA11, PA3 and PA15 (Takasaki et al., 1989; Takasaki et al., 1990b). From a sequence alignment of PA11, PA3 and PA15 and other phospholipases it is not clear what region corresponds to the specific activity. One of these molecules is called ACII-4, which is confirmed to correspond to PA11, and has a strong anticoagulant activity and low phospholipase activity. The other two molecules exhibit different activities, one has very strong phospholipase activity and the second has strong haemolytic activity.

![Sequence alignment of PA11, PA3 and PA15.](image)

The conserved regions are shadowed.

1.11 Aim and synopsis of this thesis

Txln-1 and ACII-4 have been selected from purified Australian snake venom components to be structurally characterized using X-ray crystallography. The Txln-1 molecule was chosen because of its potential as an alternative to aprotinin when used in humans to reduce blood loss during heart surgery. Txln-1 has already been extensively characterized kinetically and is available as a recombinant protein in large quantities for crystallization. The aim was to solve the structure of rTxln-1 as the free molecule and also bound to the active site of both human plasmin and trypsin. ACII-4 was chosen because of its anticoagulant properties and elucidating its
structure was part of a larger project focused on clarifying the mechanism by which the PLA₂ exerts its potent anticoagulant effect. The aim was to solve the structure of ACII-4 as the free molecule.

The structures of rTxln-1 can be compared directly with that of the crystal structures of aprotinin as the free molecule and aprotinin in complex with trypsin. It is of interest to map the changes in structure that occur to rTxln-1 upon forming complexes with trypsin and microplasmin, and also to observe the stereochemistry of the catalytic triad when rTxln-1 is docked to the active site.

The functions of staphylokinase and streptokinase have already been investigated by X-ray crystallography and their structures in complex with microplasmin solved. These structures will be compared with microplasmin in complex with rTxln-1.

Detailed three dimensional pictures of the interactions made between rTxln-1 and the active site of plasmin can make it possible to determine what the key interactions between rTxln-1 and plasmin are that make it such a tight inhibitor. It can suggest modifications to the structure of the rTxln-1 that might lead to the discovery of a better or altered inhibitor which could have enhanced therapeutic value.

The specific aims are to give a structure based explanation of the mechanism of inhibition, inhibitory efficiency and specificity of rTxln-1. From these studies it will be possible to perform correlations between structure and inhibition mechanism. One goal was to explain why rTxln-1 binds reversibly to plasmin and has a higher off rate compared to aprotinin. From the complexes it would also be of significance to draw conclusions about inhibition of plasmin since no structure with a Kunitz/BPTI-type inhibitor in complex with plasmin exists.

Analysis of the structure of ACII-4 will allow us to determine the molecular basis for the high anticoagulant activity and low phospholipase activity. The interest in this study lies in the high sequence similarity of this molecule with the two other similar molecules isolated from Pseudechis australis, yet all three have different biological activities.
Chapter 2
Preparation of Proteins for Crystallization

2.1 Introduction
One of the initial aims of the project was to determine the crystal structure of free recombinant textilinin-1 (rTxln-1). Its study was of particular interest because of its potential as a therapeutic agent to prevent blood loss in the case of patients undergoing major surgery. The most likely target for its activity is to inhibit plasmin (Filippovich et al., 2002). After six months working on the project a gene construct of human microplasmin (the catalytic domain of plasmin) became available allowing me the prospect of determining the structure of rTxln-1 in complex with this fragment of the enzyme. For comparative studies and to understand the reasons for the specificity of binding, the structure determination of a complex between trypsin and rTxln-1 was also an aim for the project.

As a part of the overall effort of the Venomics group to discover and characterize proteins from Australian snake venoms a phospholipase A2 referred to as ACII-4 with potent anticoagulant activity was isolated and purified from the venom of the Australian King Brown snake, Pseudechis australis. Determination of the three-dimensional structure of this molecule and comparison with the structure of other phospholipase A2s which possess different activities should begin to explain the reasons for the potent anticoagulant activity.

To determine the crystal structure of a protein requires that the protein of interest is purified to high degree, usually greater than 95%, that at least one but generally tens of milligrams be available and that the concentration of the protein to be in the range of 1- 40 mg/mL. This chapter describes the experiments that were required to obtain the proteins in the amounts and purity necessary to be able to undertake the X-ray crystallographic studies.

One of the most significant advances in accelerating the rate of structure determination by X-ray crystallography has been the revolution in molecular biology and the development of recombinant protein technologies. To produce sufficient
quantities of textilinin-1, recombinant protein was required. This protein was supplied by BresaGen (Section 2.3). Since the complete plasmin molecule consists of a number of domains that are connected by flexible linkers, recombinant technologies were necessary to produce a form of plasmin that was amenable to structural investigation. Cloning made it possible to produce a smaller form of plasmin, referred to as microplasmin, which contains only the serine protease domain (X. Wang et al., 1998). One of the challenges was to purify, solubilize, fold and activate the recombinant human microplasmin in sufficient quantities for crystallization experiments.

One of the difficulties with the preparation of the microplasmin is that it is expressed as inclusion bodies. As a result, the protein must be completely unfolded and then refolded again. The question arises as to whether or not the protein has folded into the correct three-dimensional structure. There are several methods including dynamic light scattering (DLS) and circular dichroism (CD) to test that the protein is correctly folded. However, neither of these is as predictive as measuring the enzymatic activity of the sample using spectrophotometric assays. Enzymatic activity assays were used to test that the complexes were formed and that the enzymes were completely inhibited.

As mentioned earlier, the purity of the protein is critical to the outcome of the crystallization experiment. Impurities can prevent crystals from forming, disturb the formation of a lattice, or produce crystal defects that lead to a reduction in diffraction quality. To monitor the purity of the samples SDS-PAGE, analytical ultracentrifugation, and MALDI-TOF mass spectrometry experiments were carried out.

2.2 Materials and methods
Preparation of rTxln-1, rTxln-1-trypsin complex, rTxln-1-microplasmin complex and ACII-4 are described in the next four sections. rTxln-1 was produced by BresaGen Ltd, Adelaide, South Australia. Bovine trypsin was purchased from Sigma-Aldrich, and the plasmid containing human microplasmin was kindly donated by Professor Chai Zhang, Oklahoma Medical Research Foundation, USA (Loy et al., 2001). ACII-4 was supplied as native protein by Dr Paul Masci. Chemicals for SDS-PAGE
were purchased from Bio-Rad. Prepacked Superdex 75 and HiLoad Superdex 200 (Prep Grade) columns were purchased from Amersham Biosciences. The Sephacryl 300 for gel filtration and CNBr-Sepharose were purchased from GE Healthcare. Chromogenic substrate benzoyl-l-arginyl p-nitroanilide (BApNA) was purchased from Sigma-Aldrich, and S-2251 was purchased from Chromogenix. All chemicals used were of analytical grade.

2.3 Cloning, expression and purification of recombinant textilinin-1

The cloning, expression and purification of rTxln-1 (NCBI accession number AAK95519) was carried out by Dr Stan Bastiras (BresaGen Ltd, Adelaide, South Australia). The cloned construct, abbreviated to “TVWT” consisted of a proprietary N-terminal expression-enhancer sequence (TV), an internal tryptophan residue (W) and the 59-amino-acid mature Txln-1 sequence (T). The protein was expressed as inclusion bodies in Escherichia coli MM294 cells. The inclusion bodies were dissolved at pH 11.7 and at a concentration of 2 mg/mL. Refolding of the fusion protein was initiated by lowering the pH to 9.5 with 1 M HCl. The progression of refolding was assessed by RP-HPLC. After 2 h, cleavage of the TVW moiety was initiated by the addition of N-chlorosuccinimide (Lischwe and Sung, 1977). The released rTxln-1 was purified by a combination of phenyl-Sepharose high performance hydrophobic interaction chromatography and Q-Sepharose high performance anion-exchange chromatography (Amersham Biosciences). The purified rTxln-1 containing fractions from the Q-Sepharose high performance column were pooled and applied onto a Sephadex G25M column equilibrated with 25 mM Tris–HCl pH 7.8. The rTxln-1 was stored at 4°C in 25 mM Tris–HCl pH 7.8, 50 mg/mL mannitol and 0.01% Tween-20.

Dr Simone Flight titrated a trypsin solution with p-nitrophenyl p-guanidinobenzoate (Chase and Shaw, 1967) in the presence and absence of a sub-equimolar amount of rTxln-1, which gave an estimate of the molar concentration of rTxln-1. This was in excellent agreement with the molar concentration calculated from the UV spectrum using a Hitachi U-2800 UV recording spectrophotometer (Hitachi). An $\varepsilon_{280} = 4560$ M$^{-1}$ cm$^{-1}$ was estimated by the method of Gill and von Hippel (Gill and von Hippel, 1989).
2.4 Preparation of free recombinant textilinin-1 for crystallization experiments

2.4.1 Concentration and SDS-PAGE of rTxln-1

Free rTxln-1 was supplied in a formulation buffer given above. I buffer exchanged the protein into 25 mM Tris–HCl pH 7.8 at 4°C prior to crystallization. Samples were concentrated up to 40 mg/mL using an Amicon stirred cell (Millipore) with a 1000 Da molecular weight cut-off membrane (Millipore). The concentration was determined from the absorbance maximum at 280 nm. Its purity and molecular weight was assessed by 4-20% SDS-PAGE by the method of Laemmli (Laemmli, 1970). Protein samples were mixed with loading buffer consisting of a final concentration of 50 mM Tris-HCl pH 6.8, 0.05% (w/v) bromophenol blue, 10% (v/v) glycerol, 2% (w/v) SDS (Sigma-Aldrich). Some samples were prepared under reducing conditions in the presence of 100 mM β-mercaptoethanol. All samples were boiled for five minutes before gel loading. The samples were electrophoresed through the stacking gel and separating gel at 80 V and 200 V, respectively, in a Mini-Protean II vertical slab electrophoresis system. Gels were stained overnight in 0.1% (w/v) Coomassie Brilliant Blue R-250, 40% (v/v) methanol, 10% (v/v) acetic acid and destained until clear in 40% (v/v) methanol 10% (v/v) acetic acid (Sigma-Aldrich). The SDS-PAGE showed one band equivalent to approximately the same molecular weight, as aprotinin (Figure 2-1).

Figure 2-1 Non-reduced 4-20% gradient SDS-PAGE of purified rTxln-1. The gel was stained with Coomassie Brilliant Blue R-250. Lane 1 Protein standards from (Bio-Rad). Lanes 2-4 were rTxln-1 at increasing concentrations.
2.4.2 Mass spectral analysis of rTxln-1

Purified rTxln-1 was analysed using a VOYAGER-DE STR MALDI TOF mass spectrometer (Applied Biosystems, Voyager Instrument Control Panel v 5.10) operating in linear positive mode with an accelerating voltage of 25 kV, grid voltage 93% and delay time of 1100 nsec, with the help of Chris Wood at the University of Queensland. The machine was externally calibrated with bovine insulin (5734 Da), thioredoxin (11674 Da), and *E. coli* myoglobin (16952 Da) (Sigma-Aldrich). rTxln-1 in 25 mM Tris–HCl pH 7.8 was mixed 1:1 with matrix. The matrix was 10 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid, which had been dissolved in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid (Sigma-Aldrich) in water. The purpose of the matrix material was to absorb energy from the laser light preventing the molecules from decomposition and to isolate the molecules from one another. The low pH ensured that the molecules were positively charged (Hillenkamp *et al.*, 1991; Bahr *et al.*, 1992; Lewis *et al.*, 2000). The sample was mixed with the matrix solution and allowed to air dry before being subjected to ionization and analysis by the time-of-flight analyzer. A major peak was detected at 6690 Da (Figure 2-2), which is within an accepted tolerance of 5 Da of the molecular weight of rTxln-1 determined by BresaGen Ltd and within 4 Da of the theoretical molecular weight of the amino acid sequence ($M_w = 6686$ Da). Smaller fragments of 5054 Da, 6031 Da, and 6493 Da were visible which could have been produced by the break down of rTxln-1 when subjected to the ionization laser source. Another species was observed at 13385 Da suggesting that rTxln-1 can also dimerize. This dimer must have strong enough interactions to be sustained in the ionization laser source.
Figure 2-2 MALDI-TOF mass spectrum of rTxln-1. Sample buffer: 25 mM Tris–HCl pH 7.8; matrix: 10 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water. The major peak is at 6690 Da which is within 5 Da of the expected molecular mass for rTxln-1. The smaller peaks are likely fragmentation products of rTxln-1 produced by the ionization laser source. A peak at 13385 Da suggests the presence of a rTxln-1 dimer.

2.4.2.1 Analytical ultracentrifugation of rTxln-1

The degree of self association of rTxln-1 in 25 mM Tris–HCl pH 7.8 was assessed by analytical ultracentrifugation. Sedimentation equilibrium experiments were carried out to determine the molecular weight of rTxln-1 using a Beckman XL-I analytical ultracentrifuge (Beckman Coulter). 200 µL of 81 µM or 0.54 mg/mL rTxln-1 in 25 mM Tris–HCl pH 7.8 was loaded into one of the cell chambers and as a reference 200 µL of 25 mM Tris–HCl pH 7.8 was loaded into the second cell chamber. The cells were placed in an An-60 Ti rotor with absorption optics at 280 nm. The sample was run at 33000 rpm at 20°C for sixteen hours. The experiments provided data to plot ln Absorbance versus the square of the radial distance. The linear relationship (Figure 2-3) indicates a normal behavior for a globular protein.

The molecular weight of the sample can be calculated using the equation

\[ M_w = \frac{2RT}{\omega^2(1-\nu')p} \lambda \]

where \( M_w \) is the molecular weight, \( R \) is the gas constant, \( T \) is the absolute temperature, \( \omega \) is the angular velocity, \( \nu' \) the partial specific volume of the sample, \( p \) the buffer density and \( \lambda \) is the determined slope (Tinoco et al., 1995; Hensley, 1996).

The partial specific volume was determined to be 0.71 mL/g, and the buffer density 1.02 g/mL. Using \( \omega^2 = 11.9422 \times 10^6 \) (33000 rpm), \( T = 293 \) K, \( \lambda = 4862 \) cm\(^2\) and \( (1-\nu'p) = 0.2764 \), a molecular weight of 7181 ± 500 Da was obtained (see Appendix B for theory behind the equation used). Sedfit analysis (Schuck, 1998) gave a
numerical value of the radius parameters of $r^0 = 1.27$ nm (vertical radius), and $r^h = 1.51$ nm (horizontal radius), and a molecular weight of $7200 \pm 500$ Da.

![Figure 2-3](image)

**Figure 2-3** Sedimentation equilibrium experiments of rTxln-1 performed in the analytical ultracentrifuge. Sample buffer: 25 mM Tris–HCl pH 7.8. Data points are plotted as the ln absorbance as a function of square radial distance. The data are consistent with a straight line indicating a globular protein with a molecular weight in accordance with that of a monomer of rTxln-1.

### 2.5 Preparation of the recombinant textilinin-1-trypsin complex for crystallization experiments

#### 2.5.1 Mass spectrometry analysis, SDS-PAGE and activity assay of bovine trypsin

Bovine trypsin (Product No. T1426, Sigma-Aldrich,) was purchased as a salt-free lyophilized powder. The molecular weight specified by Sigma was 23.8 kDa. The MALDI-TOF spectrum of the sample dissolved in water showed one large peak at 23324 Da and several lower molecular weight species. The major peak corresponds to residues 16 to 245 of trypsin (according to chymotrypsin numbering), which gives an expected molecular mass of 23293 Da (Figure 2-4). The MALDI ionisation method results predominantly in the generation of singly charged molecules, although instead of the sample being ionised by the addition of a proton $H^+$, it can be accompanied by salt adducts e.g. $Na^+$ giving rise to a value that is 23 Da greater than the expected molecular mass (Hillenkamp *et al.*, 1991; Bahr *et al.*, 1992). If there are trace amounts of sodium ions present this could explain the slightly higher than expected mass value. The lower molecular weight species can appear due to autolysis.

29
or fragmentation by the ionization laser source. The commercial sample of trypsin (β-trypsin) is produced when a N-terminal hexapeptide of trypsinogen is cleaved at the Lys6 – Ile7 peptide bond, and other species may also form as a result of further proteolysis of the intact chain. An SDS-PAGE showed one strong band corresponding to intact trypsin between the 31 and 21.5 kDa, and several weaker bands at about 20, 14, 10 and 7 kDa, which could be matched with peaks in the mass spectrum (Figure 2-5). A cleavage in trypsin between Lys131 – Ser132 produces an active two-chain form referred to as α-trypsin, forming one 13.1 kDa and one 10.2 kDa polypeptide chain. A third, three-chain active form of trypsin has an additional peptide bond cleaved between Lys176 and Asp177, forming three polypeptides of 13.2 kDa, 4.7 kDa, and 5.5 kDa molecular mass. These fragments bind substrate poorly compared to β-trypsin. Another two-chain form exists which has a cleaved peptide bond between Arg105 and Val106 giving two species with molecular mass of 10.6 kDa and 12.7 kDa. It is therefore understandable that these polypeptides are represented as peaks in the mass spectrum. One of the lower molecular weight species appears to be a doubly charged positive ion of trypsin 23324/2 = 11662 Da (~11647 Da).
Figure 2-4 MALDI-TOF mass spectrum of trypsin. Trypsin dissolved in water; matrix: 10 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water. The major peak at 23323.75 Da corresponds to an enzymatically cleaved fragment of trypsin. A doubly charged positive ion with the same molecular weight, 23324/2 = 11662 Da (~11647 Da) is also observed. Other smaller trypsin fragments are present due to autolysis or fragmentation products formed by the ionization laser source.

Figure 2-5 Reduced 4-20% gradient SDS-PAGE of the trypsin sample used in MALDI-TOF mass spectrometry. The gel was stained with Coomassie Brilliant Blue R-250. Lane 1 Protein Standards (Bio-Rad), and Lane 2 trypsin dissolved in water. The most intense band corresponds to activated trypsin with a molecular weight of 23.3 kDa. The other bands represent autolysis products.

The trypsin activity was assessed by using the chromogenic substrate benzoyl-l-arginyl p-nitroanilide (BApNA) as a substrate. Four replicates were performed with a reaction mixture consisting of 620 µL of 0.2 M Tris-HCl pH 7.6, 0.03 M CaCl₂, 330 µL 1.84 mM BApNA, and 50 µL of 28 µM trypsin solution. Data were obtained by monitoring release of p-nitroaniline at 410 nm at 25°C for six minutes. Under these
conditions, the rate of reaction was linear when plotted against time. The specific activity of trypsin was calculated to be 0.108 µmol mg⁻¹ min⁻¹.

2.5.2 Mass spectrometry analysis of the rTxln-1-trypsin complex
To form the rTxln-1–trypsin complex the lyophilized trypsin was dissolved in a solution containing concentrated rTxln-1 in 25 mM Tris–HCl pH 7.8, in the molar ratio 1:1.1. This mixture was analyzed by MALDI-TOF mass spectrometry. Peaks at 6698 and 23312 Da, corresponding to rTxln-1 and trypsin were visible (Figure 2-6). One peak at 17897 Da, which was also one of the smaller peaks visible in the mass spectrum of trypsin alone, may be a product from autolysis or degradation by the ionization laser source. Several peaks corresponding to low molecular masses may be degradation of rTxln-1 by the ionization laser source or autolysis products of trypsin.

![MALDI-TOF mass spectrum of the rTxln-1-trypsin complex.](image)

**Figure 2-6** MALDI-TOF mass spectrum of the rTxln-1-trypsin complex. Sample buffer: 25 mM Tris–HCl pH 7.8; 10 mg/mL matrix: 3,5-dimethoxy-4-hydroxycinnamic acid in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water. Peaks in agreement with the molecular mass of rTxln-1 and bovine trypsin, are clearly visible. An autolysis product of trypsin is observed at 17899 Da.

2.5.3 Purification by gel filtration
Both Superdex 200 and 75 were tested as media for purification of the complex. Both were successful as fractions could be pooled containing the complex after analysis by SDS-PAGE.
2.5.4 **Superdex 200 column**

A HiLoad Superdex 200 column (Prep Grade), 2.6 x 60.0 cm was equilibrated with 25 mM Tris–HCl pH 7.4 and 50 mM NaCl (Ajax Chemicals), and calibrated with standard proteins of known molecular weight (Figure 2-9). Fractions were monitored by absorbance at 280 nm. The \( r\)Txln-1-trypsin sample was spun at 3000 rpm using a Beckman model TJ-6 centrifuge for five minutes at 25°C prior to being loaded onto the column and eluted at a flow rate of 2 mL/min with the same solution that was used to equilibrate the column (Figure 2-7). A 15% SDS-PAGE under non-reducing conditions showed a smear of weaker bands between the 31 and 21 kDa (Figure 2-8). Fractions that contained both trypsin and \( r\)Txln-1, lane 6-9 (Figure 2-8) were pooled.

![Figure 2-7](image)

**Figure 2-7**  The elution profile of the \( r\)Txln-1-trypsin complex by Superdex 200 size exclusion chromatography. The major peak appears to be the complex, with tail consisting of autolysis fragments and free \( r\)Txln-1.
Figure 2-8 Non reduced 15% SDS-PAGE of fractions collected from Superdex 200 column. The gel was stained with Coomassie Brilliant Blue R-250. Lane 1 Protein Standard (Bio-Rad), Lane 2 fraction D7 collected from 222-224 mL, Lane 3 fraction D5 collected from 226-228 mL, Lane 4 fraction D3 from 230-232 mL, Lane 5 fraction D2 collected from 232-234 mL, Lane 6 fraction E1 collected from 236-238 mL, Lane 7 fraction E3 collected from 240-242 mL, Lane 8 fraction E5 collected from 244-246 mL, Lane 9 fraction E7 collected from 250-252 mL, and Lane 10 fraction E8 collected from 252-254 mL. The strong band between 31 and 21 kDa is trypsin and the weaker band at 6.5 kDa is rTxln-1.

Figure 2-9 Calibration graph of the Superdex 200 column. Both the complex and the free trypsin would be expected to elute at around 236-240 mLs of column volume. With a molecular mass of 6.7 kDa free rTxln-1 would be expected to elute at 290 mL.
2.5.5 Mass spectral analysis of Superdex 200 fractions

The pooled sample was analyzed by MALDI-TOF mass spectrometry. The peaks at 6700 Da and 23346 Da correspond to rTxln-1 and bovine trypsin (Figure 2-10). Peaks at 13395 Da, 30046 Da and 36768 Da correspond to a rTxln-1 dimer, the rTxln-1-trypsin complex and a complex of two rTxln-1 molecules and one trypsin molecule, respectively. A peak at 4797 Da could correspond to the polypeptide consisting of residues 18-59 of rTxln-1.

![Figure 2-10 MALDI-TOF mass spectrum of the rTxln-1-trypsin complex. Sample buffer: 25 mM Tris–HCl pH 7.8; matrix: 10 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid in water. Peaks with molecular masses corresponding to free rTxln-1 and trypsin, a dimer of rTxln-1, the complex, and trypsin complexed with two rTxln-1 molecules were all observed.](image)

2.5.6 Activity assay of the rTxln-1-trypsin complex

Enzymatic activity of the rTxln-1-trypsin complex was measured using BApNA as the substrate. From two assays, one of trypsin alone and the second with the complex, it was possible to calculate the percentage of inhibition \((A_0 - A_i) / A_0\) where \(A_0\) is the absorbance per minute of the uninhibited reaction, and \(A_i\) is the absorbance per minute of the inhibited reaction. The sample showed 93% and 95% inhibition before and after, respectively, purification by Superdex 200 chromatography. This batch of material was then subjected to crystallization trials (Chapter 3).
2.5.7 Superdex 75 column

A second batch of the rTxln-1-trypsin complex was purified by FPLC on a Superdex 75 column, 1 x 30 cm. The column was equilibrated with 25 mM Tris–HCl pH 7.4 and 50 mM NaCl, and calibrated with standard proteins of known molecular weight (Figure 2-12). Samples were eluted at a flow rate of 0.8 mL/min with the same buffer that was used to equilibrate the column. The fractions were monitored by absorbance at 280 nm. Aprotinin, rTxln-1 and trypsin where profiled by loading them onto the same column. Surprisingly, rTxln-1 eluted at 13.20 mL, which corresponds to a molecular weight of 27.5 kDa or four rTxln-1 molecules (Figure 2-12). Trypsin eluted as expected at 13.56 mL corresponding to a molecular weight of 23.3 kDa. The complex of rTxln-1-trypsin eluted at 11.52 mL. The fractions were pooled and a non reduced 12% SDS-PAGE showed one strong band near 31 kDa and one band above the 6.5 kDa marker in the ladder, corresponding to trypsin and rTxln-1 respectively. The samples showed 97% inhibition both before and after purification. This batch of material was subjected to crystallization trials but no crystals were obtained (Chapter 3).

![Figure 2-11](image_url) The elution profile of the rTxln-1-trypsin complex by Superdex 75 size exclusion chromatography. Sample buffer: 25 mM Tris-HCl pH 7.8. One peak eluted at 11.5 mL followed by a tail of smaller polypeptides.
2.5.8 Concentration of rTxln-1-trypsin complex
Sodium chloride was removed by buffer exchange and the complex concentrated to 10 mg/mL using an Amicon stirred cell with a 1000 Da molecular weight cut-off membrane. The protein concentration was determined by measuring the $A_{280}$ value using an $\varepsilon_{280} = 40480\text{ M}^{-1}\text{ cm}^{-1}$ for the complex (Gill and von Hippel, 1989).

2.5.9 Additional samples of rTxln-1-trypsin for crystallization
Three batches were made for crystallization trials by directly dissolving lyophilized trypsin into a solution containing concentrated rTxln-1 in 25 mM Tris–HCl pH 7.8, at a molar ratio of 1:1.1. All batches were >95% inhibited as tested by the activity assay.

2.6 Cloning, expression and purification of human microplasminogen
2.6.1 Extraction of plasmid
A plasmid containing the gene for human microplasminogen (residues 542-791) was a gift from Professor Chai Zhang, Oklahoma Medical Research Foundation, USA. The plasmid was a construct of a pET-11a vector (Novagen), carrying a N-terminal
T7 epitope tag. The plasmid (~1.8 µg DNA), stored on blotting paper was dissolved in 10 µL of Milli-Q water for two minutes, and spun at 3000 rpm in an Eppendorf Centrifuge 5415 D (Eppendorf) at 25°C for 1 min. An A260/A280 of 1.7 signified that the aqueous sample was relatively free of protein contamination. The A260 value equated to a concentration of 138 ng/µL.

2.6.2 Expression and isolation of human microplasminogen

Recombinant microplasminogen was expressed in E. coli BL21(DE3) cells. The DNA was transformed into BL21(DE3) competent cells (Invitrogen) by electroporation 150 ng of the plasmid per 50 µL cells. The cells were grown at 37°C and 220 rpm on a Bioline 4000 Serie shaking incubator for 1 hr in 1 mL of Luria Bertani broth (LB) [10 g tryptone, 5 g yeast, 10 g NaCl dissolved in a total of 1 L water]. Agar plates containing ampicillin were streaked with aliquots of 40 µL of the cells and incubated at 37°C overnight. A single colony was picked and grown overnight in 5 mL of LB broth containing 100 µg/mL ampicillin for immediate use. Microplasminogen expression was confirmed by conducting a small scale expression, 12% SDS-PAGE under reducing conditions (Figure 2-13), and Western blot analysis with a monoclonal anti-T7 tag antibody (Novagen) (Figure 2-13). Western blot analysis (Sambrook et al., 1989) was carried out by transfer of the proteins from SDS-PAGE to a nitrocellulose membrane at 100 V for one hour at 4°C. The transfer buffer consisted of 100 mM Tris-HCl, 40 mM glycine, 0.036% SDS, and 20% methanol. The membrane was blocked against non-specific interactions by immersion in 5% skim milk in phosphate buffered saline (PBS) containing 0.05% Tween 20 (Merck). Milk protein attached to the membrane where the target protein was not attached. The membrane was incubated overnight with a monoclonal anti-T7 antibody. After washing in PBS and 0.05% Tween 20, the membrane was probed with a secondary antibody conjugated to horseradish peroxidase (Chemicon) and incubated for one hour. After washing the membrane in PBS and 0.05% Tween 20, the membrane was developed using enhanced chemiluminescence reagent (Western Lightning) and visualized with a Fujilmager 3000 chemiluminescence detector. The theoretical molecular weight of microplasminogen with the T7•tag is 28371 Da. The major band on SDS-PAGE and Western blot was observed in this region.
Figure 2-13 Analysis of microplasminogen expression
(a) Reduced 12% SDS-PAGE stained with Coomassie Brilliant Blue R-250. (b) Western blot with a monoclonal anti-T7 antibody. Lane 1 uninduced over-expression of microplasminogen. Lane 2 induced over-expression of microplasminogen. Lane 3 Prestained Protein Ladder (Fermentas). Microplasminogen appears as a strong band at about 28 kDa. Some expression of the gene was also present in the uninduced sample.

Once it was confirmed that microplasmin had been transformed into the BL21 cells, a large scale expression was carried out. Two batches (3L and 5L of media) of microplasminogen were produced by the following procedure. Overnight cultures were grown at 37°C and 225 rpm, in LB media containing 100 µg/mL ampicillin. 45 mL of these cultures were added to 1 L of LB media. When the optical density (OD) at 600 nm had reached 0.6, the cells were induced to start transcription of the fusion construct by adding 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG, Sigma-Aldrich). The cells were incubated for three hours at 37°C, 225 rpm after which the cells were pelleted at 6,600 rpm for 5 min at 4°C. Each pellet from 1 L of cell culture was resuspended in 50 mL of 50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 250 µL of one complete protease inhibitor tablet (Roche) dissolved in 1.5 mL Milli-Q water. The cells were lysed by adding 2.5 mg/mL lysozyme per 50 mL resuspended pellet solution and placed on ice for 30 min. 1.25 mL of 20% Triton X-100 (BDH Chemicals) was added per 50 mL lysate before freezing the solution at -20°C for overnight storage. The frozen solutions were thawed in 200 units of DNase (Roche) per 1 L lysate, 2 mM MgCl₂ and 2 mM MgSO₄ (Sigma-Aldrich) and stirred at 4°C for one hour. The high level of expression of human microplasminogen resulted in the formation of inclusion bodies. The inclusion bodies were spun at 11,500 rpm for 15 min at 4°C using a Beckman J2-M1 centrifuge and washed twice with 0.5%
Triton X-100, twice with 0.1% Triton X-100 made up in Tris-HCl buffer pH 8.0, and once with buffer alone to remove soluble hydrophobic proteins and membranes. The inclusion bodies were then dissolved in 100 mM Tris-HCl buffer pH 8.0 containing 2 mM EDTA, 8 M urea, 1 mM glycine, and 10 mM β-mercaptoethanol (“solubilization buffer”), and spun in 5 mL centrifuge tubes at 22,000 rpm for 30 min at 4°C using a Beckman L8-70 Ultracentrifuge and a SW55Ti rotor. This was to remove any unsolubilized protein. Samples taken after each step were analysed by reduced 12% SDS-PAGE (Figure 2-14).

Figure 2-14 Reduced 12% SDS-PAGE of the purification of inclusion bodies of microplasminogen.
The gel was stained with Coomassie Brilliant Blue R-250. Lane 1 Supernatant. Lanes 2-6 wash with decreasing amount of Triton X-100. Lane 7 solubilized inclusion bodies. Lane 8 Protein Marker, Broad Range (BioLabs).

The A_{280} of the solution was adjusted to 5.0 with the addition of solubilization buffer. The microplasminogen was refolded by slow titration with 2 L of 20 mM Tris-HCl buffer containing 2 mM EDTA, 4% glycerol, 0.3 M urea and 0.1 mM β-mercaptoethanol, for every 3 L of starting LB. The final pH was adjusted to 8.0 with HCl and then stirred at 4°C for 72 h. The solution was then spun at 8,000 rpm for 20 min at 4°C, and filtered (0.45 μm) before being concentrated using an Ultrasette tangential flow device (Pall Life Sciences) with a 10 kDa cut off membrane. The filter was prewashed with 0.1 M NaOH before being equilibrated with the refolding buffer. The refolded and concentrated protein (2.7 mg/mL) was spun at 3000 rpm at 25°C for five minutes to remove insoluble material from the solution.
To monitor expression and refolding, samples were taken from uninduced, induced, supernatant, washes, solubilized, refolded and concentrated batches. Activity assays were carried out by adding tissue plasminogen activator (tPA) (Actylase, Genentech) in a 1:10 (human microplasmin:tPA) molar ratio and incubating the mixture for sixty minutes at 37°C, allowing the microplasminogen to be converted microplasmin. For the activity assay, the chromogenic substrate H-D-valyl-L-leucyl-L-lysyl-p-nitroaniline (S-2251; 3mM) was added in a 1:1000 (human microplasmin: chromogenic substrate) molar ratio in the reaction buffer consisting of 0.05 M Tris-HCl pH 7.4. The release of p-nitroaniline was detected by measuring the increase of absorbance at 405 nm at 25°C. The plot of absorbance against time was linear for six minutes. The specific activity of microplasmin from the refolded and concentrated solution was 0.135 µmol mg⁻¹ min⁻¹.

2.6.3 Gel filtration of refolded human microplasminogen

The concentrated solution of microplasminogen was loaded onto a Sephacryl 300, 2.5 x 95.0 cm, size exclusion column. The column was equilibrated with 20 mM HEPES (AppliChem) pH 7.0, 0.4 M urea, and 0.01% azide (Sigma-Aldrich). 10 mL fractions were collected at a flow rate of 50 mL/h. The protein was eluted with the same solution and the fractions were monitored by absorbance at 280 nm. Several peaks eluted from the Sephacryl 300 column (Figure 2-15). Each fraction was tested for plasmin (like) activity using tPA to convert microplasminogen to microplasmin. The first peak had low activity and was therefore assumed to contain high molecular weight aggregates. The third peak had the highest plasmin activity and eluted at a time that would correspond to individual monomers of microplasmin. Fractions from this peak were pooled and used for the preparation of the rTxln-l-microplasmin complex.
2.6.4 Preparation of rTxln-1-microplasmin complex

Active human microplasmin is susceptible to self-digestion, therefore it was necessary to prevent degradation upon activation by forming the rTxln-1 inhibited enzyme complex as quickly as possible. In the large scale production of microplasmin, the microplasminogen was prepared for activation by adding 25% glycerol as well as the inhibitor to prevent microplasmin autoproteolysis. Urokinase-type plasminogen activator (Mitsubishi Pharma Corporation) bound to sepharose was used to convert microplasminogen to active microplasmin; a small scale test was first carried out and the protein solutions loaded onto 15% SDS-PAGE (Figure 2-16) to determine efficiency.
After successful activation of microplasminogen in the small scale test, the urokinase-type plasminogen activator was covalently coupled to CNBr-activated Sepharose 4B beads for larger scale production of the complex and easy separation of the urokinase-type plasminogen activator. The amount of 1 g of freeze-dried CNBr-activated Sepharose 4B beads was swollen for 15 min in 1 mM HCl and washed on a sintered glass filter with the same solution using a total of 200 mL 1 mM HCl. 0.5 mL of 1000 U/mL urokinase-type plasminogen activator concentrated to 2.7 mg/mL (as determined by the BCA assay) was pipetted into 9.5 mL of 0.1 M NaHCO₃ (Sigma-Aldrich) pH 8.3 and 0.5 M NaCl. This solution was then mixed with the swollen beads and centrifuged at 3000 rpm for 15 min. The supernatant was then decanted. 10 mL of blocking agent consisting of 0.2 M glycine pH 8.0 was mixed for four hours with the gel to block the remaining active groups, and then centrifuged at 3000 rpm for 15 min and the supernatant decanted off. The gel was next washed with 50 mM Tris-HCl pH 7.4 and again centrifuged at the same speed for 10 min.

The microplasminogen was preincubated in 25% glycerol and a five-fold molar excess of rTxln-1. 20 mL of this solution was then mixed with the activated beads on a rotating wheel overnight. The mixture was centrifuged at 3000 rpm at 4°C for 15 min to separate the beads with bound immobilized urokinase-type plasminogen activator from the rTxln-1-microplasmin complex. The supernatant containing the complex was shown to be >99% pure on both reduced and non reduced 15% SDS-PAGE.
PAGE. It was then concentrated using an Amicon Ultra centrifugal filter device with a 10 kDa molecular weight cut-off membrane. rTxln-1 was added to the solution during and after concentration to ensure that microplasmin was inhibited.

Two batches of rTxln-1-microplasmin complex were produced, one of 500 µL concentrated to 3 mg/mL and a second of 500 µL concentrated to 6 mg/mL. The concentration of the complex was determined by measuring absorbance at 280 nm using an $\varepsilon_{280}$ of 45460 M$^{-1}$cm$^{-1}$ for the complex. The purity was analyzed by 15% SDS-PAGE (Figure 2-17), and the inhibition of microplasmin were assayed with S-2251 by monitoring release of p-nitroaniline at 405 nm at 25°C.

![Figure 2-17](image)

**Figure 2-17** Non reduced 15% SDS-PAGE of the rTxln-1–microplasmin complex. The gel was stained with Coomassie Brilliant Blue R-250. Lane 1 Microplasminogen. Lane 2 rTxln-1. Lane 3 the microplasmin rTxln-1 complex. Lane 4 Precision Plus Protein Standards (Bio-Rad).

### 2.7 Purification and characterization of native ACII-4

The purification and characterization of native ACII-4 from Australian King Brown, *Pseudechis australis*, snake was carried out by Dr Paul P. Masci in three successive chromatography steps. *Pseudechis australis* venom (1 g) (Venom supplies) was loaded onto a heparin-Sepharose 6B column, 5.0 x 25.0 cm, equilibrated with 0.1 M ammonium acetate pH 6.5 at 4°C. The anticoagulant activity bound to the column and a peak identified as ACII was eluted by a salt gradient (0.0 – 0.5 M NaCl). The ACII peak was loaded onto a CM-Sepharose CL-6B column, 5.0 cm x 10.0 cm, equilibrated with 0.05 M ammonium acetate pH 6.5. Four peaks were eluted with an increasing concentration again by a salt gradient (0.0 – 0.5 M NaCl). A peak identified as ACII-4 (peak 4) which possessed potent anticoagulant activity was then loaded onto a Sephacryl 100 column, 2.5 cm x 95.0 cm, equilibrated with 0.05 M
ammonium acetate pH 6.4. The elution profile showed a single peak of anticoagulant and phospholipase A₂ activity. Phospholipase activity was measured by adding ACII-4 to egg yolk substrate suspension (Fohlmann et al., 1977) except that Triton X-100 was used instead of sodium deoxycholate (Tan and Tan, 1988). The anticoagulant activity was established by measurement of prothrombin time (Ware and Seegers, 1949) using a Hyland-Clotek machine, magnetic steel ball and light time monitoring system at 37°C, both with and without added heparin. Heparin enhanced the anticoagulant activity. The prothrombin time clotting assay measures the time taken for a clot formation via the extrinsic pathway. The anticoagulant activity of ACII-4 citrated plasma increased with increased Ca²⁺ concentration. The addition of platelin (as a source of phospholipids) to the prothrombin clotting assay diminished under reducing and non-reducing conditions the anticoagulant effect but did not completely overcome it. SDS-PAGE with and without β-mercaptoethanol established molecular weights of 14 kDa and 28 kDa suggesting that ACII-4 is a dimer. Electrospray mass spectrometry (ESI, Applied Biosystem) showed a peak at 13 kDa in agreement with a monomer but no sign of the dimer was observed (Masci, 2000).

Subsequently, Miss Hayley Snelling prepared another batch of ACII-4 following the procedure of Dr Paul Masci. This batch was used both for sequence determination by tandem mass spectrometry (ms/ms) of polypeptide fragments of ACII-4 from trypsin digestion carried out by Dr Geoff Birrell, and crystallization experiments. The analysis confirmed 72% of the amino acids as corresponding to the NCBI deposited PA11 sequence (NCBI accession number P04056) (Figure 2-18).

![Figure 2-18 Sequence of ACII-4 as determined by ms/ms or tryptic digestion alone.](Picture generated by Dr Geoff Birrell)
2.8 Preparation of ACII-4 for crystallization experiments

2.8.1 Concentration of ACII-4

Pure ACII-4 from the venom of the Australian King Brown, *Pseudechis australis*, snake was supplied by Miss Hayley Snelling at a concentration of ~0.8 mg/mL in 0.05 M ammonium acetate pH 6.5. The protein was concentrated to 2 mg/mL using an Amicon stirred cell (Millipore) with a 1000 M\textsubscript{w} cut off membrane (Millipore Corporation, PLAC). This value was determined by UV spectroscopy using an A\textsubscript{280} value of 21180 M\textsuperscript{-1} cm\textsuperscript{-1} (Gill and von Hippel, 1989).

2.9 Discussion

rTxln-1 is extremely stable, pure (> 95%), and soluble in 25 mM Tris–HCl pH 7.8, up to 40 mg/mL, making it ideal for crystallization experiments. Complexes with trypsin or microplasmin were confirmed to have been formed by SDS-PAGE and measurement of activity. The complexes were highly pure, stable and soluble, and therefore ideal candidates for structural investigations by X-ray crystallography. However, it was of concern that the lyophilized trypsin showed some evidence of previous formed degradation products on SDS-PAGE, possible due to the lyophilization process or autolysis.

An interesting observation was that under the conditions of sedimentation analysis rTxln-1 existed as a monomeric species, but under the conditions of mass spectrometry, a significant amount of a dimeric species was present. rTxln-1 is a very acidic molecule with a pI of 4.4, and mixed with matrix, which is acidic, makes it energetically favorable to form a dimer as seen in the mass spectrum, or a reaction is taking place in the ionization source. The self association into a dimer may therefore be driven by an increasing negative charge in the surrounding solvent.

The faster than expected elution of rTxln-1 on the Superdex 75 column could be caused by the formation of intermolecular disulfide bonds between rTxln-1 molecules with a change in weight and a change in the shape to a more elongated protein that would elute faster than a small spherical shaped protein. It may have also been caused by a possible tetramer formation of the rTxln-1 dimer observed in the mass spectrum. The faster elution of the rTxln-1-trypsin complex could be caused by a change of shape since a trypsin molecule has the dimension of roughly 30 x 45 x 35
Å and rTxln-1 which is presumable comparable to aprotinin in structure with the
dimension of roughly 35 x 20 x 15 Å together would create a heavier molecule than
trypsin alone and with a more elongated “mushroom” shape.

Further experimental analysis regarding the oligomeric state of rTxln-1 and the
association of rTxln-1 with trypsin could have been undertaken by dynamic light
scattering (DLS), electrospray ionization mass spectrometry (ESI), and nuclear
magnetic resonance (NMR) but given the high purity of samples it was decided to
proceed with crystallization trials.

ACII-4 was >95% pure as determined by SDS-PAGE and after concentration to 2
mg/mL was ready for crystallization experiments. Previous studies on ACII-4
showed that its anticoagulant and phospholipase activites were Ca$^{2+}$ dependent, but
that the individual activities were not dependent on each other (Masci, 2000). The
observation suggests that the different activities are located in separate regions of the
molecule. The earlier studies also stated that the ACII-4 molecule was possibly a
dimer. Anticoagulant phospholipase A$_2$ molecules have been reported to show a
correlation between stronger anticoagulation potency with higher-order oligomeric
forms and increasing positively charged surface (Lok et al., 2005). The self
association of ACII-4 into a dimer may also be an important factor for anticoagulant
activity of ACII-4.

2.10 Summary
The rTxln-1, its complexes with bovine trypsin and human microplasmin, and the
phospholipase A$_2$ (ACII-4) were prepared for crystallization experiments. All
proteins were acquired as pure proteins, in high concentrations, and in large
quantities. The human microplasmin was produced as a recombinant protein. This
procedure included over-expression, purification, solubilization, and folding of the
stable zymogen precursor protein (microplasminogen) from E. coli inclusion bodies,
followed by post-refolding purification before activation with urokinase plasminogen
activator to the less stable but active protease. This was carried out prior to the
formation of a pure rTxln-1-microplasmin complex.
Chapter 3
Crystallization of proteins

3.1 Introduction

The previous chapter described the processes involved in obtaining the proteins required for this study. Txln-1 and plasmin were obtained using recombinant technology, while the phospholipase was obtained from the native snake venom and trypsin was supplied commercially. All of the proteins were extensively analysed to ensure they were of high purity and maximal activity. Under those conditions the best prospects for successful crystallization and structure determination are achieved. This chapter describes growing the high-quality crystals necessary to allow data collection from the X-ray diffraction experiment. To date, no prescriptive and completely general method has been developed that will automatically lead to the successful crystallization of a protein. To find the successful crystallization conditions requires a trial and error approach where combinations of buffers, salts, pH values, precipitating agents, temperatures and additives are all tried to initiate and propagate crystal growth.

The first crystal structure of a protein was determined by John Kendrew and his colleagues (Kendrew et al., 1958). This structure determination took around twenty-five years to complete. Since that time there have been major advances in techniques aimed at improving the prospects of obtaining highly diffracting protein crystals. Perhaps the most important amongst these have been the molecular biology revolution which has lead to the successful crystallization of many proteins that are naturally produced only in minute quantities. The introduction of affinity tags at either the amino or carboxy termini of the polypeptide has greatly enhanced our ability to purify such proteins. Nowadays, companies such as Hampton and Emerald supply crystallization kits based on the sparse screening matrices that were initially developed by Jancarik and Kim in the early 1990s (Jancarik and Kim, 1991). In previous eras researchers were required to make up their own buffers and precipitating agents in ways that were often not highly reproducible. The purity and sources of such reagents often lead to variability in the success rate of the crystallization experiment. Advances in recording instrumentation of the diffraction
data such as the use of synchrotron radiation and the use of imaging plate devices and charge coupled detectors have meant that crystals of a greatly reduced size can produce interpretable three dimensional structures. Today, crystals less than 0.05 mm in all three dimensions can be used in structure determination studies.

Crystallization experiments start with a protein solution which can be concentrated anywhere in the range from 1 mg/mL to as high as 200 mg/mL. This solution is then combined and incubated with other solutions that can promote crystal growth of the protein of interest. Traditionally the size of the solution for crystal formation is in the range of 1 to 10 µL, but robotic pipetting devices now allow drop sizes as low as tens of nanolitres. As described previously, there are many factors that can lead to the successful crystallization of a protein. These include the pH of the solution the protein concentration and whether or not ligands, substrates, inhibitors, cofactors, metal ions or other additives are required. The choice of temperature is also another major factor in the success of the crystallization experiment. At low temperatures proteins are less susceptible to proteolytic cleavage but at high temperatures the protein may be more soluble.

Once a drop containing the protein has been made it must be kept in an environment such that the water can evaporate, concentrating the protein until it reaches a supersaturated state. If the conditions are appropriate, crystal nucleation may then take place. In some instances crystals may begin to appear within hours while others can take several months or years to appear.

The most common method for obtaining protein crystals is by supersaturation using the hanging drop vapor diffusion technique. In this technique the protein is suspended on a coverslip above a well solution that contains an appropriate crystallization solution. The hanging drop method was the approach I used for most of my crystallization experiments.

For producing the first crystals of a newly purified protein the sparse matrix screening method has been most widely adopted (Bergfors, 1999). The commercial screening kits from Hampton and Sigma were used to obtain the initial crystals for all of the proteins that I studied. Once initial crystals were obtained I set up focused
grid screens to begin to optimize the growth conditions. The concentration of protein, precipitant agent, buffer, pH, temperature or the incorporation of additives were all tried as variables.

Data collection can be undertaken at room temperature with the crystal mounted in a capillary. Alternatively, data can be collected at 100 K with the crystal mounted in a nylon loop. At room temperature most protein crystals are highly susceptible to damage through the production of free radicals upon exposure to X-rays. At temperatures around 100 K the formation of free radicals is greatly reduced thus minimizing the effects of radiation. To achieve maximum resolution for data collection at 100 K it is necessary to find suitable conditions whereby a crystal can be cryocooled and yet still maintain its integrity. It is also highly desirable that the cryocooling solution freezes as a vitreous glass so as to prevent the formation of ice crystals which would interfere with the X-ray diffraction image of the protein.

For more detailed description of the methods involved in protein crystallization and X-ray data collection the reader is referred to Appendix C.

### 3.2 Methods and Materials

All chemicals used for crystallization were of analytical grade. The crystallization plates, coverslips and individual reagents were purchased from Hampton Research. Hampton Crystal Screens 1 and 2, and Sigma Crystallization Cryo Kit were used for all of the proteins, except for ACII-4 where only the Hampton Screens were used. Generally, each drop consisted of 1 µl of well solution and 1 µl of protein solution, and the wells were filled with 200 µl crystallization solution, using VDX48 plates from Hampton Research. X-rays were produced by a Rigaku FR-E Super Bright generator operating at 45 kV and 45 mA, and the X-rays focused with confocal X-ray optics Osmic mirrors. Images were recorded on a Rigaku R-Axis IV++ image plate detector.
3.3 Crystallization experiments with free recombinant textilinin-1

Most of the screens for the crystallization trials were set up using the vapor diffusion hanging drop technique. VDX48 trays from Hampton Research were used for most of the experiments. Some sitting drops utilizing the 96 well Intelli-Plate from Hampton Research were also set up. All of the experiments were undertaken at 17°C. Successful crystallization conditions are shown in Table 3-1. Initial attempts to crystallize free rTxln-1 at a concentration of 5 mg/mL were unsuccessful, with virtually all of the experiments yielding clear solutions after several weeks of incubation. Subsequent trials, increasing the protein concentration to 15 mg/mL were also disappointing, with no crystals obtained after three months, and most (> 95%) drops clear of precipitation. To further encourage supersaturation of the protein 200 mM NaCl was added to the well solutions but again this did not yield any crystals.

A survey of the literature showed that adding phenol to the crystallization media was effective in promoting the crystallization of α-dendrotoxin, (Skarzynski, 1992), which is a Kunitz-type molecule isolated from the Green mamba, Dendroaspis angusticeps, venom. Phenol was therefore added to both the drops and the wells such that the final concentration was 100 and 200 mM respectively. Phase separation was observed and ascribed to the presence of phenol. This phenomenon was believed to be the reason for a localized concentration of the protein thereby leading to supersaturation and crystallization of the sample. Precipitation was also observe and recognized as the characteristic of phenol as a denaturing agent forcing the protein concentration closer to supersaturation in some of the conditions. With this approach, crystals began to appear in some of the drops. At about the same, crystals also begun to appear in some of the initial experiments that were set up (in the absence of either sodium chloride or phenol). These crystals took around six months to appear. These initial crystals were tested for diffraction on the beam either by mounting them in capillaries at room temperature or in nylon loops for irradiation at 100 K. However, none of these crystals produced strong diffraction to enable high resolution data to be obtained. All of the crystals mounted at room temperature were not stable enough to collect any significant diffraction data while the crystals in the nylon loop did show diffraction to ~5 Å.
In attempts to further optimize the crystallization conditions (Table 3-1) hanging drop experiments were set up with rTxln-1 concentrated to 15 mg/mL, using Hampton Crystal Screen I and II and the Sigma CryoKit screen but 10% (v/v) 1,4-butanediol was added to all drops. The theory behind the use of 1,4-butanediol is that it can interpose itself between two protein molecules and mediate as a crosslinker by forming electrostatic interactions and hydrogen bonds with its two hydroxyl groups to neighboring protein molecules. Thereby the protein molecules come closer to one another, stabilizing the crystal and promoting growth. In a separate experiment, the concentration of rTxln-1 was increased to 40 mg/mL and Hampton Crystal Screen I and II and the Sigma CryoKit were all surveyed. No additives (i.e. sodium chloride, phenol or 1,4-butanediol) were included in these screens. Both strategies grew large single crystals which produced X-rays that diffracted to at least 2 Å resolution. The rate of crystallization also improved from months to weeks.

Figure 3-1  Crystals of free rTxln-1. These grow from condition (a) HCSI-9 with addition of 200 mM phenol to the drop and 200 mM NaCl to the well solution (size 0.25 x 0.5 x 0.5 mm) (b) HCSI-39 with addition to the drop and 200 mM NaCl to the well solution (size 0.15 x 0.05 x 0.05 mm) and (c) HCSII-14 with the addition of 10% (v/v) butanediol to the drop (size 0.6 x 0.07 x 0.07 mm).
Table 3-1: Experimental conditions that led to the successful crystallization of rTxln-1.

Legend: HD = hanging drop; SD = sitting drop; SCK = Sigma cryo kit; HCSI = Hampton crystal screen 1; HCSII = Hampton crystal screen II; ND = not determined

<table>
<thead>
<tr>
<th>Protein (mg/mL)</th>
<th>Screen solution</th>
<th>Buffer, salt, ppt agent</th>
<th>Comments</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>SCK-13</td>
<td>0.1 M Tris-HCl pH 8.5, 0.2 M Na-citrate 30% (v/v) PEG 400 200 mM phenol</td>
<td>HD. Crystals grew after six months</td>
<td>2.9 Å.</td>
</tr>
<tr>
<td>15</td>
<td>SCK-13</td>
<td>0.1 M Tris-HCl pH 8.5, 0.2 Na-citrate 30% (v/v) PEG 400</td>
<td>HD. Crystals grew after three weeks.</td>
<td>~5 Å. See Figure 3-1(a).</td>
</tr>
<tr>
<td>15</td>
<td>HCSI-9</td>
<td>0.1 M tri-sodium citrate dihydrate pH 5.6, 0.2 M ammonium acetate 30% (w/v) PEG 4000 200 mM phenol</td>
<td>SD. 200 mM NaCl in the well solution. Crystals grew within a month.</td>
<td>~5 Å. See Figure 3-1(b).</td>
</tr>
<tr>
<td>15</td>
<td>HCSI-9</td>
<td>0.1 M tri-sodium citrate dihydrate pH 5.6, 0.2 M ammonium acetate 30% (w/v) PEG 4000</td>
<td>HD. Crystals grew after six months.</td>
<td>ND, did not collect a data set.</td>
</tr>
<tr>
<td>15</td>
<td>HCSI-39</td>
<td>0.1 M HEPES Na pH 7.5 2% (v/v) PEG 400 2.0 M ammonium sulfate 200 mM phenol</td>
<td>SD. Crystals grew within six months.</td>
<td>1.63 Å.</td>
</tr>
<tr>
<td>15</td>
<td>HCSI-39</td>
<td>0.1 M HEPES Na pH 7.5 2% (v/v) PEG 400 2.0 M ammonium sulfate</td>
<td>SD. Crystals grew within a month.</td>
<td>ND, did not collect a data set.</td>
</tr>
<tr>
<td>15</td>
<td>HCSII-13</td>
<td>0.1 M sodium acetate trihydrate pH 4.6 0.2 M ammonium sulfate 30% (w/v) PEG monomethyl ether 2000 200 mM phenol</td>
<td>SD 200 mM NaCl added to the well. Crystals grew within a month.</td>
<td>~5 Å.</td>
</tr>
<tr>
<td>15</td>
<td>HCSII-13</td>
<td>0.1 M sodium acetate trihydrate pH 4.6 0.2 M ammonium sulfate 30% (w/v) PEG monomethyl ether 2000</td>
<td>HD. Crystals grew after six months.</td>
<td>ND, did not collect a data set.</td>
</tr>
<tr>
<td>15</td>
<td>HCSII-14</td>
<td>0.1 M tri-sodium citrate dihydrate pH 5.6 0.2 M potassium sodium tartrate tetrahydrate 2.0 M ammonium sulfate</td>
<td>HD. Crystals grew after four months.</td>
<td>Dissolved in cryoloop</td>
</tr>
<tr>
<td>15</td>
<td>HCSII-14</td>
<td>0.1 M tri-sodium citrate dihydrate pH 5.6 0.2 M potassium sodium tartrate tetrahydrate 2.0 M ammonium sulfate 10% (v/v) 1,4 -butanediol</td>
<td>HD. Crystals grew within a month.</td>
<td>ND, did not collect a data set. See Figure 3-1(c).</td>
</tr>
<tr>
<td>15</td>
<td>HCSII-32</td>
<td>0.1 M HEPES Na pH 7.5 0.1 sodium chloride 1.6 ammonium sulfate</td>
<td>HD. Crystals grew after four months.</td>
<td>2.7 Å.</td>
</tr>
</tbody>
</table>
3.3.1 Cryoprotection of free rTxln-1 crystals
Crystals grown from Hampton Crystal Screen II condition 14 were soaked in cryoprotectant made up of 70% (v/v) well solution, and 30% (v/v) glycerol, but dissolved when mounted in nylon loops. Four conditions that produced free rTxln-1 crystals contained sufficient PEG 400, PEG 4000 or PEG monomethyl ether 2000 to act as cryoprotectant. Therefore these crystals were directly transferred from the drop inside a cryoloop to the cryostream. However, the best diffraction data were obtained from a crystal that was from a drop that was incubated for six months in Hampton Crystal Screen I condition 39. The results of the data processing for this crystal are described in the next section. One condition, Hampton Crystal Screen II condition 14, grew crystals that produce good diffraction without the need to add cryoprotectant.

3.3.2 Data collection and processing of free rTxln-1 crystals
X-ray data were collected using the oscillation method. All of the images had an oscillation width of 0.5° and an exposure time of 240 s (Figure 3-2). The crystal to detector distance was 100 mm. Data were integrated, scaled and merged with HKL2000 (Otwinowski and Minor, 1997). The data collection statistics are presented in Table 3-2.
Figure 3-2  Diffraction pattern for rTxln-1. A 0.5° oscillation image of a cryocooled crystal of rTxln-1. Diffraction data were observed to 1.63 Å. The zoomed square is centered on the 1.76 Å resolution ring.

3.3.3  Space group determination of free rTxln-1 crystals
Crystals of free rTxln-1 belong to the space-group I422 with unit cell parameters of $a = b = 99.9$ Å, $c = 77.5$ Å, $\alpha = \beta = \gamma = 90^\circ$. The number of protein molecules per asymmetric unit can be estimated according to the Matthews coefficient, $V_M$ (Matthews, 1968) and the solvent content. If two molecules of rTxln-1 are present in the asymmetric unit, the solvent content is calculated to be 65%, with a corresponding $V_M = 3.5$ Å$^3$ Da$^{-1}$. If three rTxln-1 molecules are present in the asymmetric unit solvent content is 49% and $V_M = 2.42$ Å$^3$ Da$^{-1}$. These values are within the normally expected range for both solvent content and Matthews
coefficient, thus based on this argument alone it is impossible to unambiguously assign two or three molecules of rTxln-1 to the asymmetric unit.

### Table 3-2 Crystal parameters and data collection statistics for free rTxln-1

<table>
<thead>
<tr>
<th>Protein</th>
<th>rTxln-1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal Parameters:</strong></td>
<td></td>
</tr>
<tr>
<td>Unit cell lengths (Å)</td>
<td>$a = b = 99.92$ $c = 77.71$</td>
</tr>
<tr>
<td>Unit cell angles (°)</td>
<td>$\alpha = 90.0$ $\beta = 90.0$ $\gamma = 90.0$</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>775,841</td>
</tr>
<tr>
<td>Space group</td>
<td>$I422$</td>
</tr>
<tr>
<td>Crystal dimensions (mm)</td>
<td>0.5 x 0.05 x 0.05</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.572</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Diffraction Data:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.54178</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.00 – 1.63</td>
</tr>
<tr>
<td>No of observations</td>
<td>224,218</td>
</tr>
<tr>
<td>No of unique reflections</td>
<td>24,804</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>99.6 (98.8)</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>0.075 (0.360)</td>
</tr>
<tr>
<td>$&lt;I/\sigma(I)&gt;$</td>
<td>16.9 (7.2)</td>
</tr>
</tbody>
</table>

Values in brackets refer to the highest resolution shell (1.69 – 1.63 Å).

$R_{merge} = \sum_{hkl} \sum_i |I_i| - <|I>| / \sum_{hkl} \sum_i |I_i|$ where $I_i$ is the intensity of the $i$th measurement of reflection hkl and $<|I>|$ is the average value over multiple measurements. $<I/\sigma(I)>$ is the intensity to noise ratio, where $I$ is intensity and $\sigma(I)$ is the standard deviation of the measurements.

### 3.4 Crystallization experiments with recombinant textilinin-1-trypsin

Crystallization trials on the rTxln-1-trypsin complex were carried out at 4°C or 17°C in VDX48 trays by the hanging drop method using a protein concentration of 10 mg/mL. Successful crystallization conditions are shown in Table 3-3. The first crystals that were obtained grew from Hampton Crystal Screen I condition 38 as small thin rods attached to precipitate. Subsequent crystals grew in condition Hampton Crystal Screen I condition 39 as bundles of rods or needles, also with a rod like morphology, from a clear solution. A grid screen (Bergfors, 1999) was carried out for Hampton Crystal Screen I condition 27 but failed to produce better crystals. Temperature was an important factor for Hampton Crystal Screen I condition 38. Changing from 17°C to 4°C resulted in obtaining larger crystals and improved resolution of data.
Table 3-3  Experimental conditions that led to the successful crystallization of the rTxln-1-trypsin complex.
HCSI = Hampton crystal screen 1

<table>
<thead>
<tr>
<th>Screen solution</th>
<th>Buffer, salt, ppt agent</th>
<th>Comments</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCSI-38</td>
<td>0.1 M HEPES-Na pH 7.5 1.4 M tri-sodium citrate dehydrate</td>
<td>Crystal formed on the side of precipitation at 17°C, Crystals grew in clusters as needles 4°C. See Figure 3-3 (a).</td>
<td>Crystals grown at 17°C Diffracted to 2.5 Å Crystals grown at 4°C diffracted to 1.63 Å.</td>
</tr>
<tr>
<td>HCSI-27</td>
<td>0.1 M HEPES-Na pH 7.5 0.2 M tri-sodium citrate dihydrate 20% (v/v) iso-propanol</td>
<td>Grown at 17°C. Crystals grew in clusters as needles. See Figure 3-3 (b)</td>
<td>Diffracted to 2.0 Å.</td>
</tr>
<tr>
<td>HCSI-39</td>
<td>0.1 M HEPES-Na pH 7.5 2% (v/v) PEG 400 2.0 M ammonium sulfate</td>
<td>Grown at 17°C. Crystals grew in clusters as rods. See Figure 3-3(c)</td>
<td>Diffracted to 2.3 Å.</td>
</tr>
</tbody>
</table>

Figure 3-3  Crystals of the rTxln-1-trypsin complex. These grew from condition (a) HCSI-38 (size 0.4 x 0.04 x 0.04 mm), (b) HCSI-27 (size 0.2 x 0.02 x 0.02 mm) and (c) HCSI-39 (size 0.3 x 0.03 x 0.03 mm).

3.4.1  Cryoprotection of the rTxln-1-trypsin crystals
All crystals of the trypsin-rTxln-1 complex were transferred from their growth drop to a new drop containing 70% (v/v) well solution and 30% (v/v) glycerol. The crystals were then soaked for 5 min in this solution before being transferred to the cryostream.
3.4.2 Data collection for the best rTxln-1-trypsin crystals

For data collection, oscillation images of 0.5° with exposure times of 360 seconds were obtained. A representative image of the data is shown in Figure 3-4. The crystal to detector distance was 100 mm for all the images allowing a complete data set to a 1.63 Å resolution to be measured. All of the data were processed with the program HKL2000 (Otwinowski and Minor, 1997). The data collection statistics are presented in Table 3-4.

![Figure 3-4](image)

Figure 3-4  Diffraction pattern for rTxln-1-trypsin complex. A 0.5° oscillation image of a cryocooled crystal of the rTxln-1-trypsin complex. Diffraction data were observed to 1.63 Å.

3.4.3 Space group determination of the rTxln-1-trypsin complex crystals

All of the crystals examined belonged to space group P3121, and had similar unit-cell parameters, \(a = b = 79.8\) Å, \(c = 107.4\) Å, \(\alpha = \beta = 90^\circ\), \(\gamma = 120^\circ\). Based on solvent content and Matthews coefficient arguments it is most likely that there is one
complex in the asymmetric unit. For one complex, the solvent content would be 59% and \( V_M = 3.0 \text{ Å}^3 \text{ Da}^{-1} \).

### Table 3-4 Crystal parameters and data collection statistics for the rTxln-1-trypsin complex.

<table>
<thead>
<tr>
<th>Protein</th>
<th>rTxln-1–trypsin complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal Parameters:</strong></td>
<td></td>
</tr>
<tr>
<td>Unit cell lengths (Å)</td>
<td>( a = 79.84 ) ( b = 79.84 ) ( c = 107.39 )</td>
</tr>
<tr>
<td>Unit cell angles (°)</td>
<td>( \alpha = 90.0 ) ( \beta = 90.0 ) ( \gamma = 120.0 )</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>592,841</td>
</tr>
<tr>
<td>Space group</td>
<td>( P3_{1}21 )</td>
</tr>
<tr>
<td>Crystal dimensions (mm)</td>
<td>0.4 x 0.04 x 0.04</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.432</td>
</tr>
</tbody>
</table>

| **Diffraction Data:**   |                         |
| Wavelength (Å)          | 1.54                    |
| Temperature (K)         | 100                     |
| Resolution range (Å)    | 50.00 – 1.63            |
| No of observations      | 192,531                 |
| No of unique reflections| 49,994                  |
| Data completeness (%)   | 97.1 (90.7)             |
| \( R_{\text{merge}} \)  | 0.055 (0.494)           |
| \( <I/\sigma(I)> \)     | 22.2 (11.47)            |

Values in brackets refer to the highest resolution shell (1.69 – 1.63 Å).

\( R_{\text{merge}} = \sum_{hkl} |I_i - \langle I \rangle| / \sum_{hkl} |I_i| \) where \( I_i \) is the intensity of the \( i \)th measurement of reflection \( hkl \) and \( \langle I \rangle \) is the average value over multiple measurements.

\( <I/\sigma(I)> \) is the intensity to noise ratio, where \( I \) is intensity and \( \sigma(I) \) is the standard deviation of the measurement deviations.

### 3.5 Crystallization experiments for the recombinant textilinin-1-microplasmin complex

Two batches of purified complex were produced for crystallization trials, one at 3 mg/mL, the other at 6 mg/mL. Crystallization trials were set up using the hanging drop technique employing VDX48 trays from Hampton, at 17°C. For a summary of the successful crystallization conditions see Table 3-5. Initial crystals grew from the 3 mg/mL protein batch and had a feather like morphology. A grid screen was carried out around this condition, but without any success. However this exhausted all of the material for further crystallization trials. For the next batch of experiments the complex was concentrated to 6 mg/mL and another round of crystallization screening trials was attempted. New crystals were obtained that had the same shape, but were larger, easier to handle and diffracted to higher resolution. From one of these crystals a data set to 2.8 Å resolution was obtained. It was not necessary to soak the crystals.
in cryosolution since they already had enough PEG 8000 or glycerol to act as cryoprotectant.

**Table 3-5** Experimental conditions that led to the successful crystallization of the rTxln-1-microplasmin complex.

<table>
<thead>
<tr>
<th>Protein (mg/mL)</th>
<th>Screen solution</th>
<th>Buffer, salt, ppt agent</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 mg/mL</td>
<td>HCSI - 28</td>
<td>0.1 M sodium cacodylate pH 6.5, 0.2 M sodium acetate trihydrate, 30% (w/v) PEG 8000</td>
<td>X-ray data measured to 3 Å, but crystal twinned</td>
</tr>
<tr>
<td>6 mg/mL</td>
<td>SCK - 31</td>
<td>0.17 M ammonium sulfate, 25.5% (v/v) PEG 4000, 15% (v/v) glycerol</td>
<td>2.8 Å data were collected</td>
</tr>
</tbody>
</table>

### 3.5.1 Data collection and processing of the rTxln-1-microplasmin crystals

For data collection, oscillation images of 0.5° with exposure times of 180 seconds were obtained (Figure 3-5). The crystal to detector distance was 150 mm. The data set to 2.8 Å resolution was integrated, scaled and merged with the program CrystalClear 1.3.6 (Pflugrath, 1999). The data collection statistics are presented in Table 3-6. The complex crystallized in the space group $P2_1$ with unit cell parameters $a = 81.9$ Å, $b = 48.0$ Å, $c = 82.6$ Å, $\alpha = 90^\circ$, $\beta = 102.1^\circ$, $\gamma = 90^\circ$. Based on the assumption of two molecules per asymmetric unit, the solvent content and Matthews number were 48% and $V_M = 2.35 \text{ Å}^3 \text{ Da}^{-1}$ respectively.
Figure 3-5 Diffraction pattern for rTxln-1-microplasmin complex.
A 0.5° oscillation image of a cryocooled crystal of the rTxln-1-microplasmin complex.
Diffraction data were measured to 2.8 Å.

Table 3-6 Crystal parameters and data collection statistics for the microplasmin-rTxln-1 complex.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Human microplasmin-rTxln-1 complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal Parameters:</strong></td>
<td></td>
</tr>
<tr>
<td>Unit cell lengths (Å)</td>
<td>$a = 81.40$  $b = 48.52$  $c = 82.63$</td>
</tr>
<tr>
<td>Unit cell angles (°)</td>
<td>$\alpha = 90.0$  $\beta = 102.1$  $\gamma = 90.0$</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>319,115</td>
</tr>
<tr>
<td>Space group</td>
<td>$P2_1$</td>
</tr>
<tr>
<td>Crystal dimensions (mm)</td>
<td>0.4 x 0.02 x 0.01</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.754</td>
</tr>
<tr>
<td><strong>Diffraction Data:</strong></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.541</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>41.60 – 2.78</td>
</tr>
<tr>
<td>No of observations</td>
<td>53,569</td>
</tr>
<tr>
<td>No of unique reflections</td>
<td>15,858</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>99.8 (99.3)</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>0.117 (0.284)</td>
</tr>
<tr>
<td>$&lt;I/\sigma(I)&gt;$</td>
<td>7.9 (3.6)</td>
</tr>
</tbody>
</table>

Values in brackets refer to highest resolution shell (2.90 – 2.80 Å).

$R_{merge} = \sum_{hkl} \sum_{i} |I_i - <I>| / \sum_{hkl} \sum_{i} |I_i |$ where $I_i$ is the intensity of the $i$th measurement of reflection $hkl$ and $<I>$ is the average value over multiple measurements. $<I/\sigma(I)>$ is the intensity to noise ratio, where $I$ is intensity and $\sigma(I)$ is the standard deviation of the measurement deviations.
3.6 Crystallization experiments with ACII-4

Crystallization trials were carried out by vapor diffusion using hanging drops in VDX48 trays from Hampton incubated at 17°C with a protein concentration of 2 mg/mL. ACII-4 crystals had a trigonal morphology. The best crystallization condition was Hampton Crystal Screen I condition 17 which consisted of 0.2 M lithium sulfate monohydrate 0.1 M Tris-HCl pH 8.5 and 30% (w/v) PEG 4000. The crystals grew as clusters which could be separated from each other by applying pressure with a needle. The 30% (w/v) PEG 4000 in the crystallization buffer was sufficient to act as a cryoprotectant.

Figure 3-6 An ACII-4 crystal cluster grown from Hampton crystal screen I condition 17
The largest crystal is 0.2 x 0.05 x 0.05 mm.

3.6.1 Data collection of ACII-4 crystals

For data collection, oscillation images were 0.5° with exposure times of 180 s (Figure 3-7). The crystal to detector distance was 90 mm. Data was collected to 1.56 Å resolution and was integrated, scaled and merged with the program HKL2000 (Otwinowski and Minor, 1997). The data collection statistics are presented in Table 3-7. The data successfully processed in the space group $P3_1$ with unit cell parameters of $a = b = 82.38$ Å, $c = 28.90$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 120^\circ$. Based on the assumption of two molecules per asymmetric unit, the solvent content of the crystals was calculated to be 43%, with a corresponding Matthews coefficient of $V_M = 2.17$ Å$^3$ Da$^{-1}$ (Matthews, 1968).
Figure 3-7 Diffraction pattern for ACII-4. A 0.5° oscillation image of a cryo-cooled crystal of ACII-4. Diffraction data were observed to 1.56 Å.
Table 3-7 Crystal parameters and data collection statistics for ACII-4.

<table>
<thead>
<tr>
<th>Protein</th>
<th>ACII-4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal Parameters:</strong></td>
<td></td>
</tr>
<tr>
<td>Unit cell lengths (Å)</td>
<td>(a = 82.38) (b = 82.38) (c = 28.90)</td>
</tr>
<tr>
<td>Unit cell angles (°)</td>
<td>(\alpha = 90.0) (\beta = 90.0) (\gamma = 120.0)</td>
</tr>
<tr>
<td>Unit cell volume (Å³)</td>
<td>169,847</td>
</tr>
<tr>
<td>Space group</td>
<td>(P\bar{3}1)</td>
</tr>
<tr>
<td>Crystal dimensions (mm)</td>
<td>0.2 x 0.05 x 0.05</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.347</td>
</tr>
<tr>
<td>No of molecules in asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td><strong>Diffraction Data:</strong></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54178</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.00 – 1.56</td>
</tr>
<tr>
<td>No of observations</td>
<td>101,238</td>
</tr>
<tr>
<td>No of unique reflections</td>
<td>31,296</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>96.2 (88.9)</td>
</tr>
<tr>
<td>(R_{\text{merge}})</td>
<td>0.043 (0.183)</td>
</tr>
<tr>
<td>(&lt;I/\sigma(I)&gt;)</td>
<td>23.4 (8.4)</td>
</tr>
</tbody>
</table>

Values in brackets refer to the highest resolution shell (1.62 – 1.56 Å).

\(R_{\text{merge}} = \sum_{hkl} \sum_{i} |I_i - <I>| / \sum_{hkl} \sum_{i} |I_i|\) where \(I_i\) is the intensity of the \(i\)th measurement of reflection \(hkl\) and \(<I>\) is the average value over multiple measurements. \(<I/\sigma(I)>\) is the intensity to noise ratio, where \(I\) is intensity and \(\sigma(I)\) is the standard deviation of the measurement deviations.

3.7 Discussion

Diffraction quality crystals were obtained for rTxln-1, its complexes with bovine trypsin and human microplasmin, and for ACII-4. The most challenging crystals to obtain were those of free rTxln-1. Determination of appropriate conditions for crystallization of free rTxln-1 free structure was first pursued by using a low protein concentration (i.e. ~5 mg/mL) and no additives. With an increase of protein concentration to ~15 mg/mL small, delicate crystals that showed weak or low resolution data could be obtained. Once the protein concentration was raised to 40 mg/mL or upon the addition of organic solvents phenol and 1,4 butanediol high quality crystals of free rTxln-1 could be obtained. The addition of phenol to the crystallization trials of free rTxln-1 was employed because it had been shown previously to aid in the crystallization of another small multiply disulfide bonded venom peptide, α-dendrotoxin (Skarzynski, 1992; Chayen, 1998). Phenol was also previously successful as an additive for crystallization of both porcine and human insulin (Ciszak et al., 1995; Whittingham et al., 1995). The phase separation upon addition of phenol caused the protein to concentrate leading to supersaturation and
subsequently result in crystal growth. Phenol also acts as a denaturing agent forcing
the protein concentration closer to supersaturation in some of the conditions, leading
to precipitation of the protein. Crystal growth from this precipitation also occurred. It
is hypothesized that 1,4–butandiol may help create and stabilize protein crystals by
interposing themselves between proteins and form intermolecular crosslinks through
non-covalent bonds between 1,4-butanediol two hydroxyl groups and two proteins.
These crossbridges may involve both electrostatic interactions and hydrogen bonding
arrangements (McPherson, 1999; McPherson and Cudney, 2006). The pH of the
condition that grew crystals of free rTxln-1 ranged from 5.6 to 8.5. All of the crystals
that were X-rayed had the same space group and unit cell parameters suggesting that
all contained the interlocking of the individual molecules in a specific organization
which is highly favored across a range of pH values and ionic strengths.

The best crystals obtained for the rTxln-1-trypsin were obtained at 4 °C, this is not
surprising given the protease is likely to be subjected to more degradation at higher
temperatures. It was somewhat alarming to find that the trypsin purchased from
Sigma contained a significant amount of degraded enzyme, but for the most part
these impurities did not impede the crystallization process. Perhaps, if these
impurities could have been removed then better quality crystals could have been
obtained.

The preparation of the rTxln-1-microplasmin complex was technically challenging in
that enzyme had to be activated by urokinase plasminogen activator but at the same
time be inhibited by rTxln-1 so as to prevent autolysis. Fortunately, crystals of the
complex appeared within a few days of setting up the crystallization experiments and
these crystals were able to diffract sufficiently to obtain a 2.8 Å resolution data set.

The crystallization of ACII-4 did not pose any major technical difficulties. Purified
native protein at concentration of 2 mg/mL and the use of Hampton Crystal Screen
were the only requirements to produce high quality crystals of this venom component.
3.8 Summary

Free rTxln-1, and its complexes with bovine trypsin and human microplasmin, as well as the phospholipase, ACII-4 were all crystallized. A range of different techniques were used to promote crystallization including the hanging drop and sitting drop methods. Screening of suitable crystallization conditions was by the incomplete factorial approach. Once a suitable concentration of protein was found, conditions leading to crystal growth could be obtained using the commercial kits. In some instances optimization of the conditions resulted in improved X-ray data.

Sometimes a major bottleneck in the structure determination of the proteins by X-ray crystallography is to find a suitable cryoprotection protocol. There were some initial difficulties in finding such a protocol for the free rTxln-1 crystals but once larger crystals were obtained there was no significant problem. Crystals of the rTxln-1-trypsin complex could easily be transferred to cryosolution before being flash cooled, while the rTxln-1-microplasmin complex and the phospholipase grew in condition with enough cryoprotectant. Complete data sets were collected to 1.63 Å, 1.63 Å, 2.8 Å, and 1.56 Å resolution for the rTxln-1, rTxln-1-trypsin complex, rTxln-1-microplasmin complex, and ACII-4 respectively.
Chapter 4
Crystal structure of free recombinant textilinin-1

4.1 Introduction
rTxIn-1 is a trypsin and plasmin inhibitor, consisting of a single polypeptide chain of 59 amino acid residues with a molecular weight of 6.8 kDa. The amino acid sequence of rTxIn-1 suggests it has an aprotinin-like fold. A search for similar sequences to rTxIn-1 in the Protein Data Bank using the ‘Basic Local Alignment Search Tool’ (BLAST) (Schäffer et al., 2001) shows the highest identity, 57%, with a chymotrypsin inhibitor from the Banded krait (*Bungarus fasciatus*) and 45% sequence identity to aprotinin. Six cysteine residues forming three disulfide bonds are in conserved locations in these molecules. Their shape is of an elongated ellipsoid and contains two antiparallel β-strands with connecting loops and two helices (one N-terminal and one C-terminal).

The aprotinin-like fold is found to exist as an independent, single domain, e.g. aprotinin. There are also proteins that consist of multiple aprotinin-like domains. For example bikunin found in human blood serum and urine (Y. Xu et al., 1998) and ornithodorin isolated from the blood sucking soft tick *Ornithodoros moubata* (van de Locht et al., 1996) contain two such domains, while the tissue factor pathway inhibitor contains three such domains (Burgering et al., 1997; Mine et al., 2002). Aprotinin-like domains are also found within larger structures where they are covalently linked to another protein e.g. human type VII and Type VI (α 3 chain) collagens, which contain an aprotinin-like inhibitor domain at their C-termini (Zweckstetter et al., 1996; Krowarsch et al., 2003). The β₁-bungarotoxin from Indian Common krait, *Bungarus caeruleus* (Sharma et al., 1999) and β₂-bungarotoxin from Taiwanese Banded krait, *Bungarus multicinctus* (Kwong et al., 1995) consist of an aprotinin-like domain linked by a disulfide bond to a phospholipase A₂ domain.

Many proteins with an aprotinin-like fold are serine proteinase inhibitors. Their protruding protease inhibitory loop is referred to as the canonical loop and is associated with a canonical substrate-like inhibition. The canonical loop features a residue which projects into the specificity pocket of serine proteases. This residue is
a crucial determinant of the specificity for different serine proteases. Many important physiological and pathological processes are under the control of serine proteinases such as blood coagulation, fibrinolysis, the kinin-kallikrein system, the complement system, and fertilization. Their regulation by protease inhibitors could lead to the establishment of new methods to treat diseases associated with these processes (Daci et al., 1999; Busuttil et al., 2004; Chun et al., 2004).

Snake venoms are a rich source of molecules that have an aprotinin-like fold (M. J. Laskowski and Kato, 1980). For example, α-dendrotoxin (Skarzynski, 1992) and calcicludine (Gilquin et al., 1999) from green mamba, Dendroaspis angusticeps, dendrotoxin I (Lancelin et al., 1994) and dendrotoxin K (Berndt et al., 1993) from Black mamba, Dendroaspis polylepis polylepis, β₁-bungarotoxin and β₂-bungarotoxin. Many of these molecules act as protease inhibitors but also exhibit targeting capabilities seemingly unrelated to this function (Dufton, 1985; Pritchard and Dufton, 1999). For example, dendrotoxin K is a K⁺ channel blocker (Smith et al., 1997) and calcicludine is a Ca²⁺ channel blocker (Schweitz et al., 1994).

Six textilinin molecules have been identified from the Australian Common Brown snake, Pseudonaja textilis. All have a similar size and the six cysteine residues are in conserved locations (Figure 1-8). Thus all are predicted to have an aprotinin-like fold. However, based on their sequence differences particularly in the canonical loop they are likely to have different functionalities and specificities. Thus there appears to be broad potential for these molecules as drug candidates for a range of pathological processes.

rTxln-1 and aprotinin are both trypsin and plasmin inhibitors but they have very different sequences in their canonical loops. In rTxln-1, the sequence in the canonical loop is Pro-Cys-Arg-Val-Arg-Phe, while in aprotinin it is Pro-Cys-Lys-Ala-Ile. Furthermore, their overall sequences differ to such an extent that their pI values are 4.4 and 8.9 for rTxln-1 and aprotinin, respectively (Masci et al., 2000). Therefore, it can be expected that the exposed surface of these two molecules and hence their activities are very different even though their overall folds are conserved.
This chapter describes the steps used to solve the structure of rTxln-1, and compares the structure of this molecule with aprotinin and other aprotinin-like molecules. Finally, I speculate about the structure and functions of the other five textilinins from the Australian Common Brown snake, *Pseudonaja textilis*.

### 4.2 Introduction to solving the phase problem of the rTxln-1 crystal

Chapter 3 described the experimental procedures and step required to produce X-ray quality crystals of rTxln-1. That chapter then went onto describe the preliminary X-ray analysis of the crystal and the results of the complete data collection. A data set to 1.63 Å resolution was obtained and the crystals belonged to the space group *I*422.

The electron density is the Fourier transform of the structure factors, which contain both the amplitudes and the phase angles of the diffracted rays (for a mathematical description see Appendix D). To determine the electron density it is necessary to measure the intensity of the reflections, which are proportional to the square of the amplitude of the structure factors. However, the phase angle for each diffracted ray can not be measured directly. The lack of phase information is commonly called the “phase problem”. Different methods have been developed to solve the phase problem, including multiple isomorphous replacement (MIR), multiple anomalous scattering or multiple anomalous dispersion (MAD), and molecular replacement (MR). The processed data in Chapter 3 was used in molecular replacement to find the phases so that complete structure factors could be obtained.

Molecular replacement is a method which relies on the existence of a previously solved protein structure homologous to the unknown structure from which the diffraction data is derived. The crystal structure of the rTxln-1-trypsin complex (Chapter 5) was solved prior to the structure of the free rTxln-1, therefore this molecule could be used as the probe molecule in MR studies. The MR method encompasses calculations to determine the orientation and position of the molecule in the unit cell. In general, these calculations are carried out in two steps consisting of one rotational search and one translational search. If successful, a preliminary model will be obtained by correctly orienting and positioning the homologous structure. This solution can then be further optimized by rigid body refinement. Finally, the
model goes through cycles of refinement and model building to reduce bias introduced by the starting model.

In obtaining a solution to the phase problem two factors can be critical; the search model should not deviate significantly from the unknown structure, and the dataset needs to be highly complete to at least ~3 Å resolution. For a detailed description of the methods see Appendix D and E.

4.3 Phase determination of free rTxln-1
Several attempts were made to determine the structure of free rTxln-1 using the molecular replacement program EPMR (Kissinger et al., 1999) (see Appendix E.3). The intensities measured were converted to structure factors using TRUNCATE (French and Wilson, 1978). CAD (CCP4, 1994) sorted the data and put it into the correct asymmetric unit and MTZ2VARIABLES (Dodson, 1992) produced a file for EPMR and for refinement calculations. The initial studies were based on the hypothesis that there were two molecules of rTxln-1 in the asymmetric unit. These results gave a plausible packing arrangement, but the refinement stalled with an $R_{\text{free}}$ value of 0.48. Since this result appeared to not be correct the program Phaser (McCoy et al., 2005) (see Appendix E.4 for Phaser and maximum-likelihood method) within CCP4 (CCP4, 1994) was next tried for molecular replacement studies. In Phaser, data between 22.41 Å and 2.5 Å resolution was used for all calculations and the search model was the rTxnl-1 that was determined when bound to trypsin (Chapter 5).

In keeping with the EPMR studies, a search for two molecules of rTxln-1 in the asymmetric unit was attempted but subsequently a search for three molecules in the asymmetric unit was also undertaken. For selection of solutions Phaser uses the Z-score and the Log-Likelihood Gain (LLG) score. The Z-score is the number of standard deviations above the mean value of the maximum-likelihood function. After a fast translation function, this should generally be $> 5$ as an indication that the fitting of a molecule is correct. The LLG score is the difference between the log-likelihood score of a molecular replacement trial and the log-likelihood calculations for a Wilson distribution, and it indicates how much better the data can be predicted from the model than with a random distribution of the same atoms (Storoni et al., 2004).
The search for three molecules in the asymmetric unit gave four peaks in the fast rotation function with values greater than 75% of the top peak. The top solution showed an LLG score of 22 and a Z-score of 5. All four peaks were passed to the fast translation function and the top solution had an LLG score of 114 and a Z-score of 14. The crystal packing was checked and no clashes were observed. This solution was then refined with Phaser which improved the LLG to 131 (Table 4-1).

<table>
<thead>
<tr>
<th>Table 4-1</th>
<th>Molecular replacement solution: orientation of a single molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler angles (°)</td>
<td>Fractional coordinates</td>
</tr>
<tr>
<td>54.9</td>
<td>83.4</td>
</tr>
</tbody>
</table>

In the next stage, the first molecule was fixed, and a search was conducted to find the second molecule. Three peaks were obtained in the fast rotation function. The top solution had a LLG score of 152 and a Z-score of 5. These three peaks were then passed to the fast translation function with the top solution showing an LLG score of 338 and a Z-score of 19. The packing of this solution showed no clashes and refinement improved the LLG to 390 (Table 4-2).

<table>
<thead>
<tr>
<th>Table 4-2</th>
<th>Molecular replacement solution: orientations of two molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler angles (°)</td>
<td>Fractional coordinates</td>
</tr>
<tr>
<td>34.9</td>
<td>96.8</td>
</tr>
<tr>
<td>350.0</td>
<td>115.6</td>
</tr>
</tbody>
</table>

The fast rotation function was run again to search for a third molecule. Four peaks were found. The top solution showed a LLG score of 411 and a Z-score of 6. These four peaks were passed to the fast translation function with the top solution showing an increased LLG score of 764 and a Z-score of 23, with no clashes in the packing. The final refinement of this solution gave a LLG of 1311 (Table 4-3).

<table>
<thead>
<tr>
<th>Table 4-3</th>
<th>Molecular replacement solution: orientations of three molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euler angles (°)</td>
<td>Fractional coordinates</td>
</tr>
<tr>
<td>34.4</td>
<td>97.0</td>
</tr>
<tr>
<td>277.1</td>
<td>151.1</td>
</tr>
<tr>
<td>9.2</td>
<td>64.1</td>
</tr>
</tbody>
</table>

Examination of the crystal packing indicated reasonable contacts between the three molecules and no overlap of symmetry related molecules (Figure 4-1).
4.4 Initial model building of the rTxln-1 crystal structure

The results from Phaser were used in the software suite ARP/wARP (Morris et al., 2003) (see Appendix E.5) to build the initial model of rTxln-1. This set of programs fit the sequence of rTxln-1 into the electron density. For selection of solutions the value of the connectivity index should increase if the tracing is successful. The connectivity index indicates the progress of the number of consecutive residues in the sequence traced in the electron density. A value below 0.6 is considered not very promising, a value around 0.8 indicates good progress, while a value above 0.95 indicates an essentially complete tracing. Fragments of five residues or shorter are converted back to free atoms. The sequence coverage is reported which is defined as the ratio between the number of docked residues and the total number of traced residues. A value higher than 0.8 is judged as being good sequence coverage. The first building cycle showed 159 residues in five chains, with the longest chain comprising 54 residues, and a connectivity index of 0.94. After the last building cycle 169 residues in three chains were obtained. These had a connectivity index of 0.94 and a sequence coverage of 99%.
4.5 Refinement of model of free rTxln-1
The model determined by ARP/wARP was refined with REFMAC5 (Murshudov et al., 1997) (see Appendix E.7) by maximum likelihood. Restrained refinement was undertaken with no prior phase information. Data used in refinement ranged from 22.41–1.63 Å, with 5.1% of reflections used for the $R_{\text{free}}$ test set (1259 reflections). All atoms were refined with individual isotropic B-factors. Arp_water (Lamzin and Wilson, 1993) was used to place the water molecules. Model building between the cycles of refinement was undertaken with the program Coot (Emsley and Cowtan, 2004) (see Appendix E.6).

4.6 Refinement statistics and quality of free rTxln-1 structure
There are three rTxln-1 molecules in the asymmetric unit. The electron density map is, in general, well defined and the majority of the amino acid residues could be fitted into the electron density (Figure 4-2). In the final model, residues 1–58 for molecule A, residues 2–59 for molecule B, and residues 1–58 for molecule C, one sulfate ion, two PEG molecules, and 240 water molecules could be fitted. The final $R_{\text{factor}}$ and $R_{\text{free}}$ are 0.191 and 0.218, respectively. A difference between the two values of only 0.027 is an indication that the model has not been over refined. The final model has excellent geometry with an rmsd from ideal bond lengths of 0.01 Å, and an rmsd from ideal bond angles of 1.24° (Table 4-4).

![Figure 4-2](image)

Figure 4-2 Stereoview of the $2F_o-F_c$ electron density map contoured at 1σ for residues 16-20 and residue 40 of molecule A of rTxln-1. The electron density map is displayed as a light grey mesh. Carbon, nitrogen and oxygen are colored green, blue and red respectively.
Table 4-4 Refinement statistics for rTxln-1

<table>
<thead>
<tr>
<th>Non-hydrogen atoms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecule A</td>
<td>462</td>
</tr>
<tr>
<td>Molecule B</td>
<td>457</td>
</tr>
<tr>
<td>Molecule C</td>
<td>462</td>
</tr>
<tr>
<td>Total number of protein atoms</td>
<td>1381</td>
</tr>
<tr>
<td>Solvent (H₂O)</td>
<td>240</td>
</tr>
<tr>
<td>Solvent (SO₄)</td>
<td>5</td>
</tr>
<tr>
<td>Solvent (PEG)</td>
<td>12</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>22.41 – 1.63</td>
</tr>
<tr>
<td>( R_{\text{factor}} )</td>
<td>0.191</td>
</tr>
<tr>
<td>( R_{\text{free}} )</td>
<td>0.218</td>
</tr>
<tr>
<td>rmsd from ideal geometry</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.010</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.24</td>
</tr>
</tbody>
</table>

\( R_{\text{factor}} = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|}, \) 5.1% of the data were excluded from the refinement to calculate \( R_{\text{free}} \).

Thermal parameters (B-values) reflect to a first approximation positional variation of atoms over time. The B-value is a consequence of the dynamic disorder in the crystal caused by the temperature-dependent vibration of the atoms in the structure. High B-values can also serve as markers for problem areas in models undergoing crystallographic refinement. A high B-value indicates that the atom moves extensively, which gives low registration of the intensity and weak electron density, resulting in difficulties to build the model. This can lead to an incorrect structure description in these areas. The highest values are often found in flexible surface loops and flexible N- and C-termini. In this structure, the average B-values for all main-chain atoms are 17.10 Å², 17.72 Å² and 16.21 Å² for molecules A, B, and C, respectively, while for side-chain atoms these values are 20.00 Å², 20.52 Å² and 18.87 Å² (Table 4-5). An analysis of the B-values as a function of position in the polypeptide of the three rTxln-1 molecules is shown in Figure 4-3. It shows that the N- and C-terminal tails of molecule A and B have the largest B-values. The next highest B-values observed are for residues 27-29 in all the three molecules. These residues form a solvent exposed loop. No obviously aberrant B-values were observed for the three rTxln-1 molecules.
Table 4-5  Mean B-values (Å²) for the three rTxln-1 molecules in the asymmetric unit

| Molecule A     | Main-chain | 17.1 |
|               | Side-chain | 20.0 |
|               | all atoms (462 atoms) | 18.5 |
| Molecule B    | main-chain | 17.7 |
|               | side-chain | 20.5 |
|               | all atoms (457 atoms) | 19.1 |
| Molecule C    | main-chain | 16.2 |
|               | side-chain | 18.9 |
|               | all atoms (462 atoms) | 17.5 |
| H₂O (240 atoms) |       | 30.6 |
| SO₄ (5 atoms) |       | 26.0 |
| PEG chain G (6 atoms) |   | 43.4 |
| PEG chain H (6 atoms) |   | 33.8 |
| Total         | main-chain | 17.0 |
|               | side-chain and waters | 22.8 |
|               | all atoms (1638 atoms) | 20.4 |

Figure 4-3  B-values (Å²) for rTxln-1 molecules A, B and C. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values.
A Ramachandran plot informs as to which regions of the polypeptide adopt secondary structure and illustrates the various regions of dihedral space (phi and psi backbone angles) that are the most energetically favored (Ramachandran and Sässiekharan, 1968; R. A. Laskowski et al., 1993). It can therefore detect unusual backbone conformations in the protein. The Ramachandran plot of the free rTxln-1 crystal structure showed 89.4% of residues in most favored regions, 8.5% of residues in additionally allowed regions, and 2.1% of residues in generously allowed regions (Table 4-6). There are no residues in disallowed regions. The amino acid side-chains that have additionally allowed backbone dihedral angles are Arg17 in molecules A and B, Arg19 in molecule C, Asn43 in molecule A and C, and Glu41, Asn45 and Asn46 in all three molecules (Figure 4-4). The consistent occurrence of Glu41, Asn45 and Asn46 adopting high energy dihedral angles in all three of the molecules indicate that these conformations are nonetheless favorable for the structures and do not occur as the result of crystal packing.

Table 4-6 Ramachandran plot statistics for the three free rTxln-1 molecules

<table>
<thead>
<tr>
<th>Region</th>
<th>Residues (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favored regions</td>
<td>89.4</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
<td>8.5</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>2.1</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure 4-4 Ramachandran plot of the crystal structure of free rTxln-1. A (α-helix), B (β-sheet) and L (left handed α-helix) are the most favored regions. a, b and c are additionally allowed regions, and ~a, ~b, ~l, ~p are generously allowed regions. Glycine residues are identified as triangles. Residues 17, 18, 19, 41, 43, 45 and 46 of rTxln-1 are labeled. Amino acids with high backbone energy conformations are Arg17 in molecules A and B, Val18 in molecule B, Arg19 in molecule C, Asn43 in molecule A and C, and Glu41, Asn45 and Asn46 in molecules A, B and C.
4.7 Crystal structure of free rTxln-1

The three rTxln-1 molecules in the asymmetric unit all have a shape of an ellipsoid and measure 35 Å x 17 Å x 15 Å. Their accessible surface areas are 4041 Å$^2$, 4047 Å$^2$ and 3926 Å$^2$ for molecules A, B and C, respectively, as calculated by CCP4 AREAIMOL (B. Lee and Richards, 1971). All three rTxln-1 molecules consist of; one $\beta$-sheet, which is twisted through 180°, and together with the loop that connects the $\beta$-strands form a $\beta$-hairpin motif in the central part of the protein. Three disulfide bonds stabilize the overall structure, Cys7-Cys57, Cys16-Cys40 and Cys32-Cys53. Cys7-Cys57 joins the N- and C-termini and Cys16-Cys40 links the canonical loop (Pro15-Phe20) with the secondary binding loop (Ile36-Glu41) (see Chapter 5). Cys32-Cys53 stabilizes the structure by connecting the second $\beta$-strand with the C-terminal $\alpha$-helix. The wiring diagram of the secondary structure is shown in Figure 4-5, and the stereodiagram of the overall fold is shown in Figure 4-6. The overall surface is mainly negatively charged except for the canonical loop region which contains Arg17-Val18-Arg19. The side-chains of Arg17 and Arg19 point out from the molecule giving the canonical loop a positively charged surface with a high degree of flexibility due to the length of these side-chains. The side-chain of Val18 points in the opposite direction to the side-chains of Arg17 and Arg19 (Figure 4-7).

Figure 4-5 Wiring diagram of the secondary structure of rTxln-1.

H1 = N-terminal $3_{10}$-helix, H2 = C-terminal $\alpha$-helix, A = $\beta$-strands that form the antiparallel $\beta$-sheet, yellow circles = cysteine residues forming disulfide bridges (orange solid line defines the connection), $\beta$ = $\beta$-turn, $\gamma$ = $\gamma$-turn and $\gamma$ = $\gamma$-turn and $\beta$-hairpin. The three $\beta$-turns are formed by Asn26-Glu29, Pro27-Lys30 and Asn43-Asn46. The $\gamma$-turn is formed by Ala44-Asn46. The $\beta$-hairpin is formed by Phe20-Asn26. Picture generated by PDBsum (R. A. Laskowski et al., 2005).
Figure 4-6  Stereo view of the overall fold of molecule A of rTxln-1. Helices are colored red, β-strands are colored yellow, the loops are colored green and the disulfide bonds are shown as orange sticks. The same numbering is applied as for the wiring diagram (Figure 4-5).

Figure 4-7  Connolly surface with electrostatic surface potentials overlaid for molecule A of rTxln-1. Regions of positive potential are colored blue, negative potential are colored red and neutral regions are white. The side-chains of Arg17 and Arg19 irradiate out from the core structure making the top portion of the molecule positively charged. The side-chain of Val18 points away from the positively charged surface and finds stabilizing hydrophobic contacts with the side-chain of Phe20.

Least-squares fitting of topologically equivalent Cα atoms (and all atoms including side-chains) of residues 2-58 in rTxln-1 using the program Superpose (E. Krissinel and Henrick, 2004), gave an rmsd of 0.81 Å (1.63 Å), 0.88 Å (2.15 Å) and 1.21 Å (2.34 Å), for molecule A and B, A and C and B and C, respectively. On visual inspection of the superposition, the majority of the polypeptide backbone overlays well except for the canonical loop region and the N- and C-termini (Figure 4-8). Both termini of molecule B differ from the position of the termini of molecules A and C. The N- and C-termini of rTxln-1 are regions with fewer internal contacts and
consequently less restrictions in their conformations, which is reflected in their high B-values.

The canonical loop region adopts a different conformation in molecule C compared to molecules A and B. This difference is highlighted by the rmsd values after least-squares fitting of Pro15-Phe20, which are 0.50 Å (1.88 Å), 1.69 Å (4.73 Å) and 1.43 Å (4.93 Å) for A and B, A and C, and B and C, respectively. The largest separation after superimposition is between Val18Ca in molecule A and the equivalent atom in molecule C (Figure 4-8). This distance is 5.8 Å. The change is due to rotations in the backbone dihedral angles around Arg17 and Arg19. The overall result is that the side chain of Val18 is flipped through ~180° in molecule C relative to its position in the A or B molecules. The (Φ, Ψ) angles of Arg17 have changed from (-126.4°, 33.1°) in molecule A or (-125.9°, 36.4°) in molecule B to (-54.2°, 133.9°) in molecule C. Arg17 in molecule C has a slightly more favorable dihedral angle conformation compared to Arg17 in molecule A and B (Figure 4-4). Thus Arg17 changes from a β-sheet conformation in molecule A and B to an α-helical region in molecule C. The (Φ, Ψ) angles of Arg19 have changed from (-83.1°, 102.9°) in molecule A or (-133.6°, 110.9°) in molecule B to (-140.9°, -18.0°) in molecule C (Figure 4-4). Thus Arg19 changes from α-helical in molecule A and B to β-sheet in molecule C. Therefore, the three backbone conformations all appear to be energetically favored but may ultimately be influenced by neighboring contacts within the protein or the crystal.
Figure 4-8  Stereo view of superimposition of the three free rTxln-1 molecules within the asymmetric unit.
Molecule A is green, molecule B is blue and molecule C is purple. Disulfide bonds are shown as orange sticks. The canonical loop of molecule C has a different conformation compared to molecule A and B. The distance between Val18Cα of molecule A and C is highlighted by a black line and is in Å units. In general, the main chains of all thee copies of rTxln-1 overlay well and the chirality of the three disulfide bonds is the same in all three molecules.

A B-value analysis reveals that the canonical loop (Pro15-Phe20) and secondary loop of all three molecules (Ile36-Glu41) have relatively low overall B-factors for the main-chain atoms. The canonical loop in molecule C is firmly held in place, with about equally rigidity as molecule B and more rigidly than molecule A (Figure 4-3; Figure 4-9). The interactions that stabilize molecule C are; van der Waals contacts between the side-chain of Val18 and Thr13, and Ile36, and the backbone nitrogen of Arg19 hydrogen bonds to the carbonyl oxygen of Tyr37 (3.05 Å). The χ₁ angle of Thr13 is rotated ~180° in molecule C compared to molecule A and B. Instead of Thr13Oγ₁ forming a hydrogen bond with Ile36O as in molecule A and B, in molecule C it forms a hydrogen bond with Gly14O (3.07 Å). Thr13 is locked in position in all three structures as reflected by it having the smallest overall B-values of any residue (Figure 4-3). There is one water molecule stabilizing the location of the canonical loop in molecule C. One of the Arg19Nη atoms in molecule C forms a hydrogen bond to this water molecule (2.98 Å), which in turn form hydrogen bonds to Arg17O (2.85 Å) and Gly38O (2.81 Å) in molecule C (Figure 4-10).
**Figure 4-9** Visualization of the main-chain B-values for the three rTxln-1 molecules. High B-values are in yellow, thick tubes. Lower B-values are in blue, thin tubes.

**Figure 4-10** Stereoview of interactions stabilizing the conformation of the canonical loop in molecule C of rTxln-1. Carbon is purple, nitrogen dark blue and oxygen red. The water molecule is displayed as a red sphere. Distances are shown as dashed lines with their lengths in Å units.
Most globular proteins contain a core of hydrophobic residues that are inaccessible to solvent in the folded state. In rTxln-1, several residues with aromatic side-chains are partially buried in the structure. Tyr25 is the most buried with a solvent accessible area of only 3.6 Å² (M. E. Chen et al., 2007). Phe47, Phe35, Tyr37, Phe23 and Tyr24 are all partially buried with solvent accessible surface areas of 10.2 Å², 20.7 Å², 35.3 Å², 36.7 Å² and 41.9 Å², respectively. However, Phe6 and Phe20, have solvent accessible surface areas of 66.2 Å² and 105.5 Å², respectively. Likewise the hydrophobic amino acids Leu9, Ala11, Val18, Leu33, Ile36, Ala44, Ile48, Ala58 and Ala59 are all on the surface of rTxln-1. Thus >1/5 of all the surface amino acids are hydrophobic. In comparison, the structure of aprotinin is much less hydrophobic with Leu33 a glutamine, Ala44 an arginine and Ile48 a lysine in aprotinin.

Some of the core residues in rTxln-1 appear to have a stabilizing effect on the structures of the canonical and secondary loops. Phe35 and Tyr37 are located in the centre of the rTxln-1 structure with some of the atoms in the side-chain within hydrophobic contact. Tyr37, in turn, stabilizes the side-chain of the Phe20 residue (~4.50 Å) which is in contact with Val18 in molecule A and B (3.48 Å and 3.79 Å, respectively) and stabilizes the conformation of the canonical loop in molecule A and B. Moreover, the aromatic side-chain of Tyr37 fits under the backbone of the secondary loop and the hydroxyl group of Tyr37 forms a hydrogen bond with Glu41O (2.70 Å, 2.66 Å and 2.62 Å in molecule A, B and C, respectively), contributing to the position of the secondary loop in all three structures.
Figure 4-11 Superimposition of the three free rTxIn-1 molecules. Carbon atoms from molecule A are green, molecule B are blue and molecule C are pink. The side-chains are displayed as sticks. Oxygen atoms are red and nitrogen atoms are blue. Disulfide bridges are displayed as yellow sticks.
There are three disulfide bonds that stabilize rTxln-1; Cys16-Cys40, Cys7-Cys57, and Cys32-Cys53. The structure of a disulfide bond can be described by its $\chi_3$ dihedral angle between the C$\beta$ – S$\gamma$ – S$\gamma'$ – C$\beta'$ atoms. A right-handed disulfide bond has a $\chi_3$=95.0±10.3° and a left-handed disulfide bond has a $\chi_3$=-86.4±8.5°. The three disulfide bonds in rTxln-1 molecules A, B and C overlay perfectly (Figure 4-8). Cys7-Cys57 ($\chi_3$=-82.3°) and Cys32-Cys53 ($\chi_3$=-85.8°) are left-handed, while Cys16-Cys40 is right-handed ($\chi_3$=95.5°) (average $\chi_3$ for molecule A, B and C). Cys16 (from the canonical loop) and Cys40 form a solvent exposed disulfide bond and its location suggests that this bond can play a role in serine protease inhibition. The N-terminal $\beta$-hairpin loop in rTxln-1 consists of three residues Pro27, Asp28 and Glu29. The kink introduced by Pro27 in rTxln-1 makes the loop form a negatively charged protrusion on the surface due to the presence of Asp28 and Glu29 (Figure 4-12).

Glycine is often present in proteins because of its unique flexibility and proline because it can place a "kink" in polypeptide chains (Kuntz, 1972; Branden and Tooze, 1991). rTxln-1 contains four glycine residues (14, 38, 39 and 42). All are found in loop regions, close to the canonical loop and secondary loop or part of the secondary loop. The five proline residues (4, 10, 15, 21 and 27) produce significant bends in the structure, especially Pro15 which starts the canonical loop and Pro21 which directly proceeds the canonical loop. The $\beta$-hairpin loop in rTxln-1 consists of three residues Pro27, Asp28 and Glu29. The kink introduced by Pro27 in rTxln-1 makes the loop form a negatively charged protrusion on the surface due to the presence of Asp28 and Glu29 (Figure 4-12).
Figure 4-12 Two views, rotated by 90°, of the Connolly surface with electrostatic potentials overlaid for molecule A of rTxln-1. Regions of positive potential are colored blue, negative potential are colored red and neutral regions are white. The β-hairpin loop Pro27-Asp28-Glu29 protrudes from the surface and the side-chains of Asp28 and Glu29 extend out to the solvent. This region forms a negatively charged protrusion on the surface.

The region of polypeptide chain from Cys16 to Thr49 which crosses the β-sheet is stabilized by the hydrogen bond formed between the amide nitrogen atom of Phe47 and the carbonyl oxygen atoms of Phe23 in the β-sheet. Further hydrogen bonds stabilizing this part of the structure are formed between the side-chain Oᵦ₁ and Nᵦ₂ atoms of Asn45, and the backbone oxygen of Leu9, and amide nitrogen and oxygen atoms of Tyr25 (Figure 4-11).

In general, polar residues in the core are energetically unfavorable except when they are able to form intra-molecular hydrogen bonds. Four polar residues are buried to a different extent in the structure of rTxln-1, these are; Ser22, Asn43, Asn45 and Asn46. Stabilizing hydrogen bonds are formed between the side-chain hydroxyl group of Ser22 and the backbone carbonyl oxygens of Phe20 and Pro21. The side-chain Nᵦ₂ atom of Asn43 forms stabilizing hydrogen bonds with the backbone carbonyl oxygen of Pro10, and its Oᵦ₁ atom form hydrogen bonds with the backbone nitrogen atom of Asn45 and one internal water molecule. The backbone dihedral angles of Asn45 and Asn46 in all three structures are in the additionally allowed region of the Ramachandran plot, however, their low B-values confirm that they have in rigid conformations (Figure 4-3; Figure 4-4). The side-chain of Asn45 is completely buried and its side-chain Nᵦ₂ atom forms hydrogen bonds with the backbone carbonyl oxygens of Leu9 and Tyr25. The side-chain of Asn46 is not
hydrogen bonded to any protein atoms, its closest potential hydrogen bond partners are Ser22Oγ and Tyr37OH which are situated ~4 Å away. However, the side-chain Nδ2 atom does form a hydrogen bond with a buried water molecule.

There are three conserved water molecules buried in the core of all three rTxl-n-1 molecules. These molecules have B-values between 10.8 Å² and 13.7 Å², significantly lower than those water molecules found on the surface of rTxl-n-1 where the average B-value is 30.6 Å², and lower than the average B-value for protein atoms (18.4 Å²). Two water molecules are centrally located in the structure of rTxl-n-1 and are hydrogen bonded to each other. One is hydrogen bonded to Asp12N, Asn45O and the side-chain Asn43Oδ1, while the other is hydrogen bonded to Asp12O, Asn43N, Asn43O and Asn46Nδ2. The third of the water molecules forms hydrogen bonds to Cys16N, Thr13O and Cys40O (Figure 4-14). This water molecule therefore plays a critical role in stabilizing and orienting the canonical and secondary loop regions.

Figure 4-13 Three conserved buried water molecules. Superimposition of molecule A (green), B (blue) and C (purple) of rTxl-n-1 and the three conserved and buried water molecules.
Figure 4-14  Stereo diagram of the hydrogen bond formed between an internal water molecule and Thr13O, Cys16N and Cys40O in molecule A of rTxIn-1. Carbons are shown as green sticks, nitrogen in dark blue, oxygen in red and the disulfide bridge in orange. The water molecule is displayed as a red sphere. (a) The 2F_o-F_c electron density is displayed as light grey mesh and contoured at 1σ. (b) Hydrogen bonds are marked with dashed black lines and their lengths are in Å units.

4.8 Crystal contacts

The protein backbone conformation as well as the side-chain conformations are influenced by the surrounding environment, and the densely packed environment of proteins in crystal structures can be dissimilar to their environment in the native state (Janin and Rodier, 1995; Carugo and Argos, 1997). Packing contacts in protein crystals can be artifacts due to the way the crystals grow and can be considered nonspecific protein-protein interactions referred to as “crystal contacts”. Regions of a protein in crystal contact are more rigid and have lower B factors than these atoms exposed to solvent (Kossiakoff et al., 1992; Jacobson et al., 2002). These contacts make use of the same forces that govern specific recognition in protein-protein complexes and oligomeric proteins. Thus, crystal contacts provide examples of nonspecific protein-protein interactions which can be compared to interactions that are biologically relevant. All possible crystal contacts from surrounding molecules in the rTxIn-1 crystal were investigated to see if the important features of rTxIn-1 were
affected by these artificial contacts and to find out if there existed an oligomeric formation.

Analysis of the crystal packing reveals that there are crystal contacts between all three rTxln-1 molecules in the asymmetric unit and that these three molecules also have crystal contacts with symmetry related molecules. Molecule C has the most number of surrounding atoms within a distance of 5 Å (Figure 4-15). These adjacent atoms are distributed over the canonical and secondary loops, the N- and C-terminal helices, the β-hairpin and the groove on the opposite side of the β-hairpin of molecule C. Molecule A and B have about the same number of surrounding atoms within a distance of 5 Å (Figure 4-15). These are distributed at about the same place as in molecule C. However, the side-chains of Arg17 and Arg19 in molecule B are surrounded with more atoms from adjacent molecules and more rigid than equivalent side-chains in molecule A. This is reflected in the B-values of the side-chains in the canonical and secondary loops. The side-chains of Arg17 and Arg19 are highly mobile in molecule A (30.9 Å² and 24.6 Å², respectively) but are relatively stable in molecule B (19.1 Å² and 18.6 Å²) and C (20.2 Å² and 19.4 Å²) (Figure 4-3).
Figure 4-15 Crystal contacts for molecule A, B and C of rTxln-1. The Connolly surface of molecule A is green, molecule B is blue and molecule C is purple. Atoms within 5 Å are shown as stick models. Carbon atoms from molecule A are green, from molecule B are blue and from molecule C are purple. Carbons atoms from PEG molecules are yellow or salmon. Sulfur atoms from sulfate ions are yellow. Nitrogen is blue and oxygen is red.

In the crystal, molecules A and C face one another with their β-sheets forming the interface, while molecule B is located with its canonical loop region and secondary loop region nearest to molecule A and the amino and carboxy terminal regions nearest to molecule C. Molecules A and C are related by a pseudo-2-fold rotation (Figure 4-16).

Figure 4-16 Stereoview of the packing of the three rTxln-1 molecules in the asymmetric unit. Molecule A is colored green, molecule B is colored blue and molecule C is colored purple. The β-sheets and molecules A and C face each other.

Mass spectrometric analysis of the free rTxln-1 indicated that a portion of these molecules form dimers in solution (see Chapter 2). Apart from NMR spectroscopy, there is no other method to visualize how this dimer might look in solution. However, calculations that determine the amount of surface area that is buried upon interaction between two neighboring molecules in a crystal can sometimes be a good
indication as to whether or not such an interaction might also be important in solution. In comparing the surface areas buried, molecules A and C have the largest buried surface of 1286 Å² while for molecules A and B this value is 499 Å² and for molecules B and C it is only 397 Å² (AREAICL (B. Lee and Richards, 1971)). Thus the most likely mimic for the dimer in solution is the A and C pairing.

Closer analysis of the interactions between molecules A and C shows that this pairing is stabilized by six inter-molecular hydrogen bonds involving (Thr13, Arg17, Tyr24, Glu29 and Ile36 from molecule A and Thr13, Glu41, Tyr24, Lys31 and Ala11 from molecule C). Thr13N(A) forms a hydrogen bond with Thr13Oγ(C) (2.94 Å) and Thr13Oγ(A) forms a hydrogen bond with Thr13N(C) (2.85 Å). The canonical and secondary loops are also involved in monomer interactions; Arg17 Ne(A) forms an ionic bond with the carboxylate group of Glu41(C) (2.86 Å). Ile36N(A) and Ala11O(C) form a hydrogen bond (2.82 Å). A further ionic bond is observed between Glu29Oε1(A) and Lys31Nζ (2.91 Å). Three water molecules bridge between residues in molecule A and C in the interface. Ala11O(A) forms a hydrogen bond to a water molecule (2.68 Å), which in turn forms a hydrogen bond to Ile36N(C) (2.98 Å) and Ile36O(C) (2.80 Å). Glu34N(A) form a hydrogen bond to the second water molecule (2.97 Å), which hydrogen bonds to Asp28Oδ2(C) (2.80 Å). Tyr24OH(A) hydrogen bonds to the third water molecule (2.74 Å), which bonds to Glu34O(C) (2.89 Å) and Asn26Nδ2(C) (2.97 Å). This overall interaction has stabilized the β-hairpin loop in molecule A and C. This is reflected in the higher B-values for this region in molecule B, as compared to both molecule A and C (Figure 4-3; Figure 4-16). Thus, not only is there a very large contact surface, there are fourteen hydrogen bonds that add to the stabilization of this pairing. It is therefore highly conceivable that this arrangement of molecules is the dimer that is observed in the mass spectrometer.
Figure 4-17 Interactions between molecule A and C in the free rTln-1 structure. Molecule A is green and molecule C is purple. Residues making contact are shown as sticks and labeled. Red is oxygen and blue is nitrogen. Water molecules are red spheres. The two molecules form a contact surface of 1286 Å² which is stabilized by six inter-molecular hydrogen bonds and three bridging water molecules. Hydrogen bonds are black dashed lines. (a) Distances are in Å units. (b) Stereo view of the interactions.
The AB interface is mostly stabilized by ionic interactions between Arg17 and Arg19 from molecule A and Glu41 from molecule B. Hydrophobic interactions are also observed between Phe20 in A and the Cys16-Cys40 disulfide bridge in molecule B. This disulfide bridge in molecule B, together with the same disulfide bridge in molecule C from another asymmetric unit form a hydrophobic crevice which the side-chain of Phe20 from A fits into. The BC interface is formed by hydrogen bonds between Asp2 and Asp5 in B and Lys1 and Asp5 in molecule C. No water molecules are visible in any of the small interfaces that include molecule B.

There are also stabilizing interactions that hold adjoining asymmetric units together. The largest contact surface is between molecule A and molecule C from another asymmetric unit related by the symmetry operation (y, 1-x, z). Here the buried surface area is 844 Å². The next largest inter asymmetric unit buried surface is 703 Å² which is between molecule B and the same C molecule.

The canonical loop in molecule C is situated between two neighboring A molecules where it is stabilized by inter-molecular bonds. One A molecule (A1) is from the same asymmetric unit as molecule C, the other (A2) is from an adjacent asymmetric unit related by (1-y, x, z). The side-chain of Val18(C) points toward a neutral surface on one side of molecule A1, while the side-chains of Arg17(C) and Arg19(C) face the opposite direction and point towards a negatively charged groove on molecule A2.
Hydrogen bonds between the canonical loop of molecule C and molecule A2 are observed between the following, Arg17N of molecule C and Gly39O of molecule A2 (2.81 Å), Arg17N\eta_1 atom of molecule C and Glu41O of molecule A2 (2.90 Å) and Arg17N\eta_2 and Ans43O of molecule A1 (2.99 Å), Arg19N\eta_1 atom of molecule C and Asn46O\delta_1 of molecule A2 (2.79 Å), Arg19N\eta_1 of molecule C and a water molecule (2.86 Å), which also forms a hydrogen bond to Ser22O\gamma (2.87 Å) of molecule A2.

There are two water molecules forming bridges between the canonical loop in the molecule C and the two A molecules. Val18N of molecule C form a hydrogen bond with a water molecule (2.94 Å), which in turn forms a hydrogen bond to Asp12O\delta_2 (2.79 Å) of the molecule A1. This water molecule forms a hydrogen bond with a negatively charged sulphate ion (2.64 Å), which is located ~5 Å from the positively charged guanidino group of Arg17 in molecule C.

There is extra electron density which makes it possible to model an alternate conformation of the guanidino group of Arg17 in molecule C where its Ne atom and one of its N\zeta atoms make two hydrogen bonds to the sulphate ion (2.95 Å and 2.97 Å, respectively). There is one PEG molecule located in the space created by molecules A1-C-A2. It is situated ~4 Å from the Cys16-Cys40 disulfide bridge in the molecule C and molecule A1.
Figure 4-19  Stereoview of the canonical loop in molecule C located between two A molecules. Arg17, Val18 and Arg19 of molecule C have purple carbon atoms. Molecule A within the same asymmetric unit is green. The adjoining molecule A is blue. For all structures, oxygen is red and nitrogen is blue. Cys16-Cys40 disulfide bond is shown in orange in all three molecules. A sulfate ion and PEG molecule are represented as balls and sticks. (a) Hydrogen bonds are drawn with black dashed lines and their lengths in Å units. (b) Stereo view of the location of the canonical loop in molecule C relative to neighboring A molecules. (c) The $2F_o-F_c$ electron density contoured at $1\sigma$ is displayed as light grey mesh.
A more extensive view of the molecular packing within the crystal shows that an alternating pattern of A and C molecules occurs such that a circle of eight molecules is apparent. The buried surfaces between A and C alternate between 1286 Å² and 844 Å². Four B molecules are attached to the AC pair to form a dodecamer of rTxln-1 molecules. However, the buried surface areas for the AB interface and BC interface are only 499 Å² and 397 Å², respectively. Four sulphate ions and eight PEG molecules are located in the circle, but these are on opposite faces. This situation is somewhat analogous to what is observed in the crystal packing of aprotinin, where two rings of five molecules are formed and five sulphate ions are situated in between the two pentamers to make a decamer (Hamiaux et al., 2000).

![Figure 4-20](image)

**Figure 4-20** Expanded view of the crystal packing of the free rTxln-1 molecules. A molecules are colored green, B molecules are colored blue and C molecules are colored purple. (a) Four A and four C molecules form a tight circle with four sulphate ions located internally, the associated B molecule packing is also shown. (b) Eight PEG molecules are also located internally but on the opposite face. (c) Sixteen PEG molecules located between two dodecamers of rTxln-1.

Solution studies of aprotinin showed that decamer formation is not simply a crystal packing artifact, and that aprotinin becomes a mixture of two particles, a monomer and decamer, in acidic or neutral pH, high salt concentration, high protein concentration, and the fraction of decamers increased with increasing salt concentration (Hamiaux et al., 2000). Three types of buried surface area are observed in the aprotinin decamer. One aprotinin molecule has two buried surfaces of 650 Å² with its two neighbors within a pentamer involving four to six hydrogen bonds. Two other buried surfaces with subunits belonging to the facing pentamer; one surface of 750 Å² that was mainly hydrophobic (only two hydrogen bonds) and a more extensive surface of 1200 Å², with at least eleven hydrogen bonds (Hamiaux et al., 2000).
Possible interpretations of the crystal packing analysis:

- Molecule A and B have fewer crystal contacts than molecule C, but any of these conformations could plausibly exist in solution (Section 4.10, Figure 4-29).
- The AC dimer pair could be the species observed in the mass spectrometer.
- The view of the expanded crystal packing is analogous to what occurs in aprotinin and suggests dodecamers of rTxln-1 could occur.

### 4.9 Comparison between textilinin-1 and aprotinin

To compare the structures of rTxln-1 and aprotinin (PDB entry 1BPI), least-squares fitting and visual analysis were undertaken. Least-square fitting of topologically equivalent Ca atoms of aprotinin and residues 3-58 of molecules A, B and C of free rTxln-1, resulted in rmsd values of 0.91 Å, 1.10 Å, and 0.95 Å, respectively. Thus, the overall folds of rTxln-1 and aprotinin are very similar. On visual inspection, the polypeptide chains overlay well and the chirality of the dihedral angles around the sulfur-sulfur disulfide bonds are conserved. Least-squares fitting of main-chain atoms in the canonical loop region (residues 15-20 rTxln-1 numbering) of aprotinin and rTxln-1 resulted in rmsd values of 0.60 Å, 0.42 Å and 1.31 Å for molecules A, B and C, respectively. Thus, residues 15-20 in molecule A and B of rTxln-1 have a typical canonical loop structure, whereas this loop in molecule C is substantially different. The largest difference is in the location of the Ca atoms of Val18 in molecule C of rTxln-1 and Ala16 in aprotinin. These two atoms are 4.31 Å apart after superimposition. This change is due to rotations in the backbone dihedral angles around Arg17 and Arg19 in molecule C of rTxln-1. The side-chain of Thr11 in aprotinin has the same conformation as molecule A and B in rTxln-1, while it has rotated ~180° in molecule C of rTxln-1. This is to accommodate the side-chain of Val18 in molecule C (Section 4.7).

Another region where the three dimensional structure of the two proteins deviate significantly is between residues 40 and 45 (rTxln-1 numbering), with the maximal difference at Gly42 in Txln-1 (Ala40 in aprotinin) where the positions of the Ca atoms of Gly42 of rTxln-1 in molecule A and Ala40 of aprotinin deviate by 3.60 Å. For Glu41 of rTxln-1 (Arg39 of aprotinin), the dihedral angle Φ stays the same, however, Ψ changes by ~180°. Average values for Glu41 in molecule A, B and C are
Φ=57.2 and Ψ=−137.6°, compared to Φ=58.4° and Ψ=43.8° in Arg39 of aprotinin. These angles correspond to a generously allowed region in the Ramachandran plot for Glu41 of rTxl-n-1 and a most favored region for Arg39 in aprotinin. In rTxl-n-1 Glu41O forms a hydrogen bond with the side-chain hydroxyl of Tyr37 (2.70 Å, 2.66 Å and 2.62 Å in molecule A, B and C, respectively). The side-chain of Tyr37 in rTxl-n-1 is located in the centre of the structure and its role is likely to be of major importance in stabilizing the core of the structure. However, Arg39O and Tyr35OH in aprotinin do not form such a hydrogen bond. Thus the higher energy dihedral angle observed for Glu41 in rTxl-n-1 is offset by the formation of a hydrogen bond to Tyr37. The following residue in the sequence is Gly42 in rTxl-n-1 and Ala40 in aprotinin. In the rTxl-n-1 structure, the Φ, Ψ dihedral angles for Gly42 in molecule A, B and C are (77.1°, -152.3°), (69.1°, -149.1°) and (86.9°, -148.7°), respectively. This is a generously allowed region in the Ramachandran plot that can be occupied by a glycine residue, however, it is a high energy area for all of the other amino acid residues including alanine and therefore not in a conformational space that could easily be explored by aprotinin. The Φ, Ψ dihedral angles for Ala40 in aprotinin are -53.9° and 146.9°, values that are in the β-sheet region.

In rTxl-n-1 the side-chain of Asn43Oδ1 forms a hydrogen bond with an internal water molecule (section 4.7), while the side-chain of Lys41 in aprotinin extends out into the solvent. Thus the side-chain of Asn43 plays a role in stabilization of the rTxl-n-1 fold, whereas the Lys41 in aprotinin appears to have no such function. One underlying reason for the difference in structure in this region is the nearby presence of Tyr10 in aprotinin (Asp12 in rTxl-n). In aprotinin the side-chain of this residue is situated on the surface making van der Waals contacts with both the side-chain and backbone of Arg39 and side-chain of Lys41. Asp12 in rTxl-n-1 has its side-chain extending out into the solvent. If the Arg39O in aprotinin was flipped and in the same position as Glu41O appears in rTxl-n-1 and Ala40 in aprotinin is in the same position as Gly42 in rTxl-n-1, the tyrosine side-chain would make unfavourable contacts (< 1 Å) with both backbone atoms and the side chain of Ala40. Therefore, these changes in sequence between rTxl-n-1 and aprotinin gain both a formation of a hydrogen bond between Glu41O and Tyr37OH in rTxl-n-1 and rigidify the structure by forming hydrogen bonds to the side-chain of Asn43 in rTxl-n-1.
Figure 4-21  Stereoview of the superimposition of the three rTxln-1 molecules and aprotinin. Molecule A of the rTxln-1 is colored green, molecule B is colored blue, molecule C is colored purple and aprotinin (1BPI) is colored yellow. Disulfide bonds are shown as orange sticks. The largest deviations in structure are in the canonical loop of molecule C of rTxln-1 and in the region between residues 40-45 (rTxln-1 numbering). Distances represented by dashed lines are in Å.

Figure 4-22  Stereoview of the region between residues 40-45 (rTxln-1 numbering) after superimposition of molecule A of rTxln-1 and aprotinin. rTxln-1 is colored green and aprotinin (1BPI) is colored yellow. Distances are marked as black dashed lines and are in Å units.
The sequence identity between rTxln-1 and aprotinin is 47 % with the major differences ascribed to residues that appear on the surfaces of the two proteins. To analyze these differences Connolly surfaces with electrostatic surface potentials mapped were constructed. While the overall surfaces vary in electrostatic potential, with a dominating negative potential on the surface of the rTxln-1 molecule and positive potential on aprotinin, the surfaces of the canonical loop in both molecules are similar. In rTxln-1 the helices and the β-hairpin loop have a negatively charged surface, while in aprotinin this helix is less negatively charged. The β-hairpin loop in aprotinin has a neutral surface but with one positively charged region where the side-chain of Lys26 points out to the solvent.

An interesting observation with the aprotinin structures is that there are four highly conserved water molecules buried in the core of the structure (Figure 4-24) (Deisenhofer and Steigemann, 1975). Three of these water molecules in aprotinin are also found in the same location in rTxln-1. The fourth water molecule has been replaced by the side-chain of Asn43 in rTxln-1. The equivalent residue is Lys41 in aprotinin where its side-chain extends in solvent (Figure 4-22).
Figure 4-23 The Connolly surface and electrostatic potentials for molecule A of rTxln-1 and aprotinin. Regions of positive potential are colored blue, negative potential are colored red and neutral regions are white. Two views (rotated by 180°) of each molecule are shown. The major part of the canonical loop has a positively charged surface in both rTxln-1 and aprotinin (1BPI).
Figure 4-24 Internal water molecules in rTxln-1 and aprotinin. Molecule A of rTxln-1 is colored green and aprotinin (1BPI) is colored blue. Disulfide bonds are displayed as orange sticks. Water molecules are displayed as spheres. (a) There are three internal water molecules in rTxln-1, displayed as red spheres. (b) There are four internal water molecules in aprotinin, displayed as red spheres. (c) Stereo view of rTxln-1 and aprotinin superimposed.
Several aprotinin structures were analyzed in order to get an overview of the polypeptide structure. In particular, I was interested to see if the canonical loop structure observed in molecule C of rTxln-1 is also found in any other aprotinin structures. Superimposition of 41 structures of aprotinin (free aprotinin, aprotinin in complexes, and mutants of aprotinin) (1AAL, 1G6X, 1QLQ, 1NAG, 1TPA, 2PTC, 2TGP, 2TPI, 3TPI, 4TPI, 1CBW, 1MTN, 1BTH, 2KAI, 1C07, 3TGI, 3TGI, 1D0D, 1FAK, 1K6U, 7PTI, 2HEX, 1BHC, 1B0C, 1BPI, 1BZ5, 4PTI, 5PTI, 6PTI, 7PTI, 8PTI, 9PTI, 1PIT (NMR), 1LD5 (NMR), 1LD6 (NMR), 1JV8 (NMR), 1JV9 (NMR), 1BTI, 1FAN, 1BPT and 1BRB) showed that only two of these exhibited major differences in their structure. These were a crystal structure of a Tyr35Gly mutant (8PTI) (Housset et al., 1991), and the NMR structure of an aprotinin mutant with residues Thr11, Pro13, Arg17, Ile18, Ile19, Val34, Gly37, Arg39 substituted by alanine, Lys15 was mutated to arginine and Met52 was mutated to leucine (1LD6) (Cierpicki and Otlewski, 2002). These alterations, except for the Met52Leu mutation, are mutations of residues influencing the canonical loop and the secondary loop (Figure 4-25). While none of these structures showed a loop with an identical conformation as to that observed in molecule C of rTxln-1, the two 8PTI and 1LD6 showed equally large structural variation as observed in molecule C of rTxln-1.

![Amino acid sequence alignment of Txln-1, aprotinin and two aprotinin mutants (8PTI and 1LD6). The cysteine residues are highlighted in orange and the mutations in aprotinin in yellow.](image)

In both rTxln-1 and aprotinin, the side-chain of Tyr37 in rTxln-1 and the topologically equivalent Tyr35 in aprotinin is situated in the centre of the structures, with the hydroxyl group positioned inside a semi circle lined by the backbone atoms of the residues Gly39-Gly42 (rTxln-1 numbering) (Figure 4-26). The B-values of Tyr37 main-chain and side-chain of molecule A, B and C of rTxln-1 and are among the lowest for all the protein atoms (11.0 Å² and 12.4 Å² for molecule A, 12.4 Å² and 12.3 Å² for molecule B, and 12.3 Å² and 11.9 Å² for molecule C, average B value for
protein atoms is 18.37 Å²), which confirm a stable position. In the Tyr35Gly aprotinin mutant, rearrangements of the canonical loop and secondary loop region have occurred, especially the secondary loop. This change in conformation was reported (Housset et al., 1991) to be one of the largest alterations in aprotinin arising from a single amino acid substitution. The secondary loop has collapsed and the backbone has moved 3.54 Å relative to aprotinin (PDB entry 1BPI), measured at the backbone nitrogen atom of residue Arg39 (aprotinin numbering, which corresponds to Glu41 in rTxln-1). Therefore, this indicates that Tyr37 of rTxln-1 has a central role in the structure of rTxln-1. Since the secondary loop and the canonical loop are linked by a disulfide bond (Cys14-Cys38 in aprotinin), the movement of the secondary loop also moved the disulfide bond and the canonical loop (Figure 4-27). The side-chains of the canonical loop were still pointing towards the solvent, and the aprotinin Thr35Gly mutant was reported to still be an inhibitor of trypsin (Housset et al., 1991). This furthermore suggests that rTxln-1 with the canonical loop in the conformation found in molecule C could still be a weak inhibitor, since the side-chains in the canonical loop are exposed and can still interact with trypsin or potentially other serine protease inhibitors.

The aprotinin mutant with eight alanine substitutions and Lys15Arg and Met52Leu mutations showed a 8.2 Å shift of the Tyr35 position relative to aprotinin (PDB entry 1BPI) measured between Ca atoms of Tyr35 (aprotinin numbering). The Tyr35 side chain has rotated out towards the solvent, and the secondary loop is on the opposite side of the canonical loop when compared to aprotinin (Figure 4-27). It was suggested that the loss of hydrophobic interactions between the aromatic side chain of Tyr35 and the side chains of Ile18, Ile19 and Val34, which had been mutated to alanine residues in this protein, were a probable reason for the alteration in the structure (Cierpicki and Otlewski, 2002). These three hydrophobic amino acids superimpose with three hydrophobic residues of rTxln-1 (Phe20, Pro21, and Ile36) and the side-chains of Phe20 and Tyr37 are close enough to have a stabilizing hydrophobic effect on each other (~4.5 Å). Both aprotinin mutants lost the interaction between the side chains of Tyr35 and Asn44 (aprotinin numbering) (Housset et al., 1991; Cierpicki and Otlewski, 2002). Asn44 in aprotinin is located at the same position as Asn46 in rTxln-1 and suggests that also Asn46 has a stabilizing effect on Tyr37 in rTxln-1. Although the canonical loop in the multiply mutated
aprotinin is less structured than aprotinin, and mainly consisting of alanine residues Ala11-Gly12-Ala13-Cys14-Arg15-Ala16-Ala17-Ala18-Ala19, the side chain of Arg15 is exposed to solvent and the cysteine bridge still located on the same side of the first loop as aprotinin. The mutant was reported to have inhibitory activity against trypsin and chymotrypsin, suggesting that the canonical conformation of the binding loop can be regained to enable interactions with these enzymes (Cierpicki and Otlewski, 2002). This again suggests that rTxln-1 with the canonical loop in the conformation found in molecule C can still interact with serine proteases.

*Figure 4-26* Stereo view of the position of the side-chain of Tyr37 in molecule A of rTxln-1 and Tyr35 in aprotinin. rTxln-1 is colored green and aprotinin (1BPI) is colored blue. The tyrosine side-chains are shown as sticks and labeled accordingly. Disulfide bonds are shown as orange sticks. The tyrosine side-chain is partially buried in the core of the protein.
Figure 4-27 Structural comparison of rTxln-1 and two aprotinin mutants. Molecule A, B and C in rTxln-1 are green. Disulfide bonds are shown as orange sticks. (a) Superposition of molecule A, B and C of rTxln-1 and aprotinin mutant 8PTI (pink). The secondary loop has collapsed. The movement of the Cys14-Cys38 disulfide bridge in aprotinin has also moved the canonical loop. (b) Superposition of molecule A, B and C of rTxln-1 and 10 NMR models of aprotinin mutant 1LD6 (yellow). The secondary loop is on the opposite side of the canonical loop compared to rTxln-1. (c) Superposition of molecule A of rTxln-1, 8PTI (yellow) and model 1 of 1ld6 (purple).
4.10 Comparison between textilin-1 and proteins with highest amino acid sequence identity

A search for proteins with similar fold and sequence to rTxln-1 in the PDB was performed using the program BLAST (Schäffer et al., 2001). This was done to identify homologs, compare their structures to rTxln-1 and analyze if rTxln-1 could have other biological functions unrelated to serine protease inhibition. Several molecules with aprotinin-like fold isolated from snake venoms are Ca$^{2+}$ channel blockers e.g calcicludine or K$^+$ channel blockers e.g $\alpha$-dendrotoxin, dendrotoxin K and dendrotoxin I. rTxln-1 is the first crystal structure determined from a molecule isolated from an Australian snake and justifies such a comparison even though its trypsin and plasmin inhibition already is well established. The BLAST search resulted in the top four hits listed in Table 4-7.

<table>
<thead>
<tr>
<th>PDB entry</th>
<th>Score (bits)</th>
<th>Amino acid sequence identity (%)</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1JC6</td>
<td>65.5</td>
<td>34/59=57</td>
<td>NMR structure of Banded krait, <em>(Bungarus fasciatus)</em>, fraction IX, a chymotrypsin inhibitor</td>
</tr>
<tr>
<td>1ZR0</td>
<td>58.2</td>
<td>27/56=48</td>
<td>Crystal structure of Kunitz domain I of tissue factor pathway inhibitor-2</td>
</tr>
<tr>
<td>1DTK</td>
<td>58.2</td>
<td>29/52=55</td>
<td>NMR structure of dendrotoxin K from Black mamba <em>(Dendroaspis polylepis polylepis)</em></td>
</tr>
<tr>
<td>1LD5</td>
<td>56.6</td>
<td>28/57=49</td>
<td>NMR structure of aprotinin mutant Lys15Arg, Ala16Val and Met52Leu</td>
</tr>
</tbody>
</table>

The four proteins in Table 4-7 have different targets. The protease inhibitor from Banded krait was reported to inhibit chymotrypsin and not trypsin (C. Chen et al., 2001a). Kunitz domain I of tissue factor pathway inhibitor-2 inhibits trypsin, plasmin and factor VIIa/tissue factor (Schmidt et al., 2005). Dendrotoxin K blocks K$^+$ channels but has negligible trypsin and chymotrypsin inhibitor activities (Dufton, 1985; Berndt et al., 1993; Smith et al., 1997). Dendrotoxin K has positively charged residues in the N-terminus and the $\beta$-hairpin loop. It is believed these could interact with K$^+$ channels (Smith et al., 1997).
The three disulfide bonds in rTxln-1 (Cys7-Cys57, Cys16-Cys40 and Cys32-Cys53) are conserved in all four proteins (Figure 4-28). Superimposition of the structures of these proteins shows that the disulfide bonds overlay well (Figure 4-29). The largest difference between the backbone conformation between rTxln-1 and these four proteins is seen between rTxln-1 and the inhibitor from Banded krait. The major difference between the backbone conformation of rTxln-1 and the inhibitor from Banded krait is from residue Glu8 to Phe20 (rTxln-1 numbering) (Figure 4-29 (a)). There are only four residues that are conserved (Leu8, Thr13, Gly14 and Cys16, rTxln-1 numbering) in this region which is a contributing factor to the difference. The structures of rTxln-1 and the Kunitz domain 1 of tissue factor pathway inhibitor-2 do not show significant differences in the backbone conformation, except when compared to the canonical loop in molecule C in rTxln-1 (as discussed in the subsections 4.7 and 4.8) (Figure 4-29 (b)). This protein is the only one in Table 4-7 that has a glutamate at position 41 (rTxln-1 numbering) and the sequence from Tyr37 to Phe47 (rTxln-1 numbering) is completely conserved suggesting its importance for plasmin binding and selectivity. A major difference between the backbone structure between rTxln-1 and dendrotoxin K is in the region which forms the β-hairpin loop (Figure 4-29 (c)). The change from Pro27 in rTxln-1 to Trp25 in dendrotoxin K contributes to the different directions of the backbone of the β-hairpin loop. The loop residue Asp28 in rTxln-1 is Lys26 in dendrotoxin K and Glu29 in rTxln-1 is Ala27 in dendrotoxin K. These differences result in this loop being negatively charged in rTxln-1 and neutral to positively charged in dendrotoxin K. The Trp25 and Lys26 in dendrotoxin K have been reported to be responsible for recognition of K⁺ channels (F. C. Wang et al., 1999). Thus it is unlikely that rTxln-1 could have a similar function.

Figure 4-28  Sequence alignment of rTxln-1 and the proteins mentioned in Table 4-7.
The backbone of the canonical loop region of the Kunitz domain 1 of tissue factor pathway inhibitor-2 and dendrotoxin K (Berndt et al., 1993) have a similar conformation to molecule A and B of rTxln-1 (Figure 4-29). The backbone of the canonical loop region in the NMR structure of inhibitor from Banded krait (C. Chen et al., 2001a) varies in position with two NMR models having similar structures as in molecule A and B of rTxln-1, and eight NMR models having similar structures to molecule C of rTxln-1 (Figure 4-29). In the inhibitor from Banded krait, Asn17 is flipped through ~180° in a similar way as occurs when molecule A and B of rtxln-1 are compared to molecule C.

The BLAST hit number five, the Lys15Arg, Ala16Val and Met52Leu triple aprotinin mutant has a reduced association constant with trypsin, chymotrypsin and plasmin compared to wild type aprotinin. The Ala16Val substitution was reported to destabilize the canonical loop conformation in the triple aprotinin mutant. However, a glycine, alanine, serine, leucine, arginine or tryptophan residue at this position did not destabilization the canonical loop conformation (Grzesiak et al., 2000; Cierpicki and Otlewski, 2002). This indicates that the valine at this position may be unique in contributing to the different conformations of the canonical loop in molecules A, B and C.
Figure 4-29 Structural comparison of rTxln-1 and four other proteins with the aprotinin-like fold. Molecule A, B and C of rTxln-1 are colored green. Disulfide bonds are shown as orange sticks. (a) rTxln-1 and ten NMR models of inhibitor from Banded krait (1JC6) (yellow) superimposed. (b) rTxln-1 and two crystal structures of Kunitz domain 1 of tissue factor pathway inhibitor-2 (1RZ0) (blue) superimposed. (c) rTxln-1 and twenty NMR models of dendrotoxin K (1DTK) (pink) superimposed. (d) rTxln-1 and ten NMR models of the triple mutant of aprotinin (1LD5) (purple) are superimposed. Structural differences are discussed in the text.
Apart from the cysteine residues there are six other conserved residues in the five proteins (Table 4-7; Figure 4-28). These are Gly14, Tyr25, Phe35, Tyr37, Gly39, Asn45 and Phe47 (rTxln-1 numbering) (Figure 4-30). All these residues overlay well upon superimposition of the structures. The side-chain of Asn45 bridges between the N-terminal 3\textsubscript{10}-helix and the second $\beta$-strand, which suggests that its conservation is important for stabilization. The conservation of Gly14 and Gly39 indicate that these residues are important in orienting the canonical and secondary loops. The Phe35 and Tyr37 play important roles in stabilizing the hydrophobic core of the fold. The general roles of the other conserved amino acids may be more obscure.

It is also evident that all the aromatic residues in rTxln-1 (Phe6, Phe23, Tyr24, Tyr25, Phe35, Tyr37 and Phe47), except Phe20, correspond to residues with aromatic side-chains in the other three proteins. The conservation of the residues with aromatic side-chains again signifies their importance in stabilization of the hydrophobic core. Phe20 in rTxln-1 corresponds to a leucine residue in the Kunitz domain 1 of tissue factor pathway inhibitor-2 and an isoleucine residue in the inhibitor from Banded krait and dendrotoxin K. Leucine and isoleucine are aliphatic therefore suggesting that a hydrophobic side-chain is needed in this position.
Figure 4-30 Location of six conserved residues. Superposition of rTxln-1 molecule A (green), inhibitor from Banded krait (1JC6) (yellow), Kunitz domain 1 of tissue factor pathway inhibitor-2 (1ZR0) (blue), dedrotoxin K (1DTK) (pink) and aprotinin triple mutant (1LD5) (purple). The conserved residues Gly14, Tyr25, Phe35, Tyr37, Gly39, Asn45 and Phe47 are labeled and their side-chains are shown as sticks. Disulfide bonds are orange sticks.

Aprotinin has four buried waters, with three of these also conserved in rTxln-1. In both rTxln-1 and dendrotoxin K, one internal water molecule is replaced by the side-chain atoms of Asn43 (rTxln-1 numbering). The water molecule, located between the canonical loop and the secondary loop in rTxln-1 and aprotinin, is replaced by the hydroxyl group of the side-chain of Ser36 (Gly38 in rTxln-1 and Gly36 in aprotinin) in dendrotoxin K (Berndt et al., 1993). This indicates that it is important to maintain a hydrogen bond between the secondary and canonical loops in this region. The replacement of a serine for a glycine and water molecule could subtly alter the structural relationships between canonical and secondary loops and therefore may be important for inhibitor interactions, specificity and selectivity.
Figure 4-31  Superimposition of rTxln-1 and dendrotoxin K. Molecule A of rTxln-1 is colored green and dendrotoxin K is yellow. Waters are shown as spheres. One internal water molecule in rTxln-1 is replaced by a serine side-chain in dendrotoxin K. Molecule A of rTxln-1 is green. Dendrotoxin K is yellow and its Ser36 side-chain is displayed as a yellow stick with an oxygen colored red. Internal water molecules are represented as green spheres.

Figure 4-32  Replacement of one internal water molecule, (in aprotinin), by an asparagine side-chain in rTxln-1 and dendrotoxin K. Waters are shown as spheres. rTxln-1 is green, aprotinin blue and dendrotoxin K yellow. (a) Overall view. (b) Zoomed in on the asparagine residues.
4.11 Comparison between the six amino acid sequences of textilinin-1 to 6

The sequences of five other textilinin molecules have been determined (Filippovich et al., 2002). They all possess the six conserved cysteine residues, which form three disulfide bridges in rTxln-1, and the six conserved residues that make up the aromatic hydrophobic core (Phe23, Tyr24 (His in Txln-5 and -6), Tyr25, Phe35, Tyr37 and Phe47) in rTxln-1. Thus it is likely these molecules have an aprotinin like fold. Interestingly, the residues Tyr37-Gly38-Gly39 in the secondary loop are also absolutely conserved.

One region where rTxln-1 and aprotinin differ in backbone structure is between Gly42 and Asn46. In all of the textilinin sequences this region is also conserved. One reason for the difference in the structure is that the side-chain of Tyr10 in aprotinin makes van der Waals contacts with both the side-chain and backbone of Arg39 and side-chain of Lys41, while the corresponding side-chain Asp12 in rTxln-1 extends out into the solvent. In the sequences of Txln-1, -2, -4, -5 and -6, position 12 is an asparagine while in Txln-3 it is a glutamate. The conservation of residues 42-46 and an asparagine or a glutamate at position 12 suggests that the other five textilinin molecules will adopt a similar structure to rTxln-1 in this region. Gly42 in rTxln-1 gives the polypeptide chain of rTxln-1 flexibility to form a stabilizing hydrogen bond to the hydroxyl group of Tyr37, Asn43 replaces one of the four internal water molecules typically observed in aprotinin, Asn45 bridges the N-terminal 3_10-helix (residue 9) with the β-hairpin and Asn46 is likely to contribute with stability of the hydroxyl group of Tyr37. The conservation of these residues indicates that they have a similar role in all textilinin molecules.

The greatest variation of residue type occurs in the canonical loop region and the β-hairpin loop region. Arg17 in the canonical loop of rTxln-1 is a determinant for its plasmin inhibition activity. Txln-1 and 2 have an arginine residue and Txln-4 has a lysine residue at position 17, followed by a small amino acid (valine residue in textilinin-1 and 2, and a glycine residue in textilinin-4). This suggests that Txln-4 has a specificity for serine proteases that binds residues with a positively charged side-chain in their specificity pocket similar to trypsin and plasmin. Residues 12-18 of Txln-3 align with identical residues in the canonical loop of the inhibitor from
Banded krait, which has been reported to bind to chymotrypsin with the side-chain of Asn17 protruding into the specificity pocket (C. Chen et al., 2001a). This indicates Txln-3 may be a chymotrypsin inhibitor even though chymotrypsin typically binds a hydrophobic side-chain in its specificity pocket. Both Txln-5 and -6 have a residue with a negatively charged side-chain in the P1 position (Glu17 in Txln-5 and Asp17 in Txln-6) followed by the sequence Asp18-Phe19-Thr20-Gly20-Ala21. This indicates that these inhibitors may have similar targets to each other but have different specificities when compared to Txln-1.

Phe20 in rTxln-1 is isoleucine in Txln-3, valine in Txln-4 and threonine in Txln-5 and Txln-6. Phe20 in rTxln-1 is at the extremity of the canonical and may only play a minor role in binding to proteases. The large variation in sequence suggests it is a residue responsible for fine tuning the selectivity and specificity of these inhibitors.

The sequence in the β-hairpin loop in the rTxln-1 (27-29) all the six Txln molecules varies considerably. Position 27 is generally occupied by a small hydrophobic residue, typically proline for Txln-1, -2, -3 and 6. The amino acid at position 28 is always charged, with three Txln molecules having aspartate and three having arginine. Position 29 tends to be negatively charged with three Txln molecules having a glutamate and one aspartate. In contrast aprotinin has the sequence Ala25-Lys26-Ala27 for the β-hairpin loop.

Figure 4-33  Amino acid sequence alignment of the six textilinin molecules; Txln-1 to 6. 24 amino acid pro-peptide and 59 amino acid mature peptide, the cysteine residues and conserved regions are shadowed and the disulfide bonds in Txln-1 are shown. Canonical, secondary and β-hairpin loops are marked with orange bars.
4.12 Summary

The structure of rTxln-1 has been determined by molecular replacement to a resolution of 1.63 Å. Unusually, three molecules of rTxln-1 are observed in the asymmetric unit of the crystals. All of these molecules have an aprotinin-like fold stabilized by three disulfide bonds and three internal water molecules. For two of the rTxln-1 molecules, the canonical loop adopts a classical shape, extending out into the solvent. However, for the third molecule this loop is inverted, with the side-chain of Val18 forming non-covalent interactions with Thr13 and Ile36. While it is a rare observation that this loop can become inverted, it is not without precedence, as a similar structure has been observed in the NMR models of the inhibitor from Banded krait (C. Chen et al., 2001a). In addition, an Ala16Val mutant of aprotinin was also shown to have a destabilized canonical loop (Grzesiak et al., 2000; Cierpicki and Otlewski, 2002). If the canonical loop is more flexible in rTxln-1 compared to aprotinin, this could begin to explain the faster off-rate this molecule has for plasmin than what does aprotinin (Chapters 1 and 6).
Chapter 5
Crystal structure of the rTxln-1-trypsin complex

5.1 Introduction

rTxln-1 is a nanomolar inhibitor of both human plasmin ($K_i=14\pm3$ nM and $K_i^* = 0.49\pm0.02$ nM) and bovine trypsin ($K_i^* = 0.76\pm0.2$ nM) (Flight et al., unpublished). Kinetic analysis revealed a competitive, reversible mechanism for the binding of rTxln-1 to trypsin, prior to inhibitor modification. Previously, a slow inactivation of rTxln-1 was observed, presumably due to enzymatic cleavage (Willmott et al., 1995). Regulation of serine proteases by inhibition has evolved to include three types of natural inhibitors that have been categorized based on their mode of binding to their target serine protease. These inhibitors are canonical, non-canonical and serpin inhibitors (Krowarsch et al., 2003). The canonical inhibitors have an exposed loop, which interacts in a substrate-like manner with their target enzyme (Bode and Huber, 1992; Krowarsch et al., 2003). Aprotinin is a canonical inhibitor of serine proteases and rTxln-1 is predicted to be a canonical inhibitor of trypsin and plasmin. Both molecules have the same fold and disulfide bond connectivities (Chapter 4). However, aprotinin is greater than $1.3 \times 10^4$ times more efficient than rTxln-1 in inhibiting trypsin and 15 times more efficient in the inhibition of plasmin (Flight et al., unpublished). The canonical loop of aprotinin fits with precise structural complementary to the active site of trypsin (pdb file 2PTC (Marquart et al., 1983)). The fit is tight with strong interactions and a lifetime, of the order of several months. Based on these data it is suggested that rTxln-1 binds in a similar but not identical fashion to trypsin as aprotinin. Some preliminary mutagenesis studies were carried out to identify the binding residues of rTxln-1 (Filippovich et al., 2002). This work showed that plasmin inhibition was completely lost with the Arg17Ala mutation suggesting that the side-chain of this residue binds in the specificity pocket of plasmin or trypsin, and is a major contributor to the affinity of rTxln-1 towards these proteases (Filippovich et al., 2002).

rTxln-1 and aprotinin only share 45% sequence identity, suggesting that there should be differences in the mode of binding of these inhibitors. The slow inactivation of rTxln-1 is an observation in common with the serpin inhibitors, but is in direct contrast to most other serine proteases inhibitors, which remain active after being
cleaved by their target enzyme (M. J. Laskowski and Kato, 1980; M. J. Laskowski and Qasim, 2000).

Trypsin has a strong preference for cleaving protein substrates after the positively charged residues arginine or lysine (Craik et al., 1985). The important catalytic triad consists of an aspartate, histidine and serine. In trypsin, Ser195 (chymotrypsin numbering system which is used throughout this chapter) acts as a nucleophile in the hydrolysis, His57 is believed to act as a general base and Asp102 stabilizes the positive charge developed during the catalytic reaction (Warshel et al., 1989; Perona and Craik, 1995). A high-affinity calcium binding site is also a structural feature of trypsin (Bode and Schwager, 1975).

A nomenclature to describe the interaction of a substrate with a protease applies numbers to the amino acid residues of the polypeptide substrate on the N-terminal side of the scissile bond as P3, P2, P1, and those residues on the C-terminal side as P1', P2', P3' (Schechter and Berger, 1967). The subsites on the protease that accommodate the substrate residues are numbered S3, S2, S1, S1', S2', S3' accordingly. In trypsin, the substrate specificity is primarily determined by the negatively charged Asp189 side-chain, which lies at the bottom of the S1 specificity pocket and interacts with the positively charged side-chain of the P1 residue of the substrate (Graf et al., 1987). In the interaction between a canonical inhibitor and protease, the side-chain of the P1 residue of the canonical loop is inserted into the S1 specificity pocket and the carbonyl oxygen of the P1 residue extends towards the oxyanion hole formed by the main chain amides of Gly193 and Ser195 (Ruehlmann et al., 1973; M. J. Laskowski and Kato, 1980; Bode and Huber, 1992). Another highly conserved feature of canonical inhibitor-serine protease complexes is a short antiparallel β-sheet interaction that occurs between the P1-P3 residues of the canonical inhibitor and the three residues on the surface of the protease (214-216) (Ruehlmann et al., 1973; Otlewski et al., 2005).

Trypsin is produced in the pancreas as the inactive precursor, trypsinogen. It moves to the small intestine where it becomes activated and degrades other proteins. Pancreatitis is inflammation of the pancreas. When the pancreas becomes inflamed, digestive enzymes attack the tissue that produces them. One is trypsin, which causes
tissue damage and bleeding, and can cause the pancreatic blood cells and blood vessels to swell. Pancreatitis most often begins at an acute stage and may become chronic. One risk group is patients who undergo endoscopic procedures (mainly endoscopic retrograde cholangiopancreatography), while other contributing factors include alcohol or gallstones. Hereditary and/or sporadic pancreatitis may be due to enhanced trypsinogen autoactivation and/or reduce trypsin autolysis (J.-M. Chen et al., 2001b) (Whitcomb et al., 1996). Trials using aprotinin to therapeutically treat pancreatitis have had mixed success. One study using a high dose of aprotinin reduced mortality (Trapnell et al., 1974), however, subsequent studies did not confirm this result (MRC, 1977; Imrie et al., 1978; MRC, 1980). Knowledge of the crystal structure of the rTxln-1-trypsin complex will help to further establish the structural features necessary for the development of trypsin specific inhibitors that could be developed to treat diseases such as pancreatitis.

A crystal structure of the rTxln-1-trypsin complex allows the determination of the binding mode of rTxln-1 to bovine trypsin and helps to understand the specificity and potency of binding. It also permits comparison with the structure of free rTxln-1 to determine the structural changes upon binding and comparison with aprotinin binding to trypsin. This chapter describes the procedures to determine the structure of the rTxln-1-trypsin complex and then goes on to analyse this structure by comparison with the structures described above.

5.2 Phase determination, model building and refinement of rTxln-1-trypsin

The program EPMR (Kissinger et al., 1999) which carries out a six-dimensional search for molecular replacement solutions by randomly changing the three positional coordinates and the three rotational parameters and then applying an evolutionary search algorithm was chosen to solve the phase problem. The search model for the study was bovine trypsin with pdb code 3PTN.

The calculation of correlation coefficient for the molecular replacement is based on structure factors. Therefore, the intensities measured (Chapter 3) were converted to structure factors using TRUNCATE (French and Wilson, 1978). This data was then sorted and put into the correct asymmetric unit using CAD (CCP4, 1994).
MTZ2VARIOUS (Dodson, 1992) converted the file to import into EPMR and CNS (Brünger et al., 1998). Based on solvent content arguments it appeared most likely that there was only one complex in per asymmetric unit. Therefore, the search model was a single molecule of trypsin. The space group for the crystal was assumed to be $P3_{1}21$ (Chapter 3) with all the data between 15–4 Å resolution used in these calculations. The evolutionary search procedure was set to run ten times unless the correlation coefficient value for any solution was greater than 0.30. In that case the program would terminate early. Each evolutionary search was carried out over 50 generations using a population size of 300. The correlation coefficient between $F_o$ and $F_c$ was used to evaluate the fitness of the solutions. In this search, the program stopped after the seventh run by finding a solution with a correlation coefficient of 0.60 and a crystallographic $R_{\text{factor}}$ of 0.40.

5.3 Refinement of the rTxln-1-trypsin complex crystal structure

The electron density was visualized in the graphics program O (Jones and Kjeldgaard, 1998) and revealed the presence of a rTxln-1 molecule bound in the active site of the trypsin molecule. The rTxln-1 was then docked into the vacant electron density. Examination of the packing within the unit cell indicated reasonable contacts between the complexes and no overlap of symmetry related molecules (Figure 5-1). The structure factors used for refinement in CNS were divided into two sets, with 90% of the data used for refinement ($R_{\text{factor}}$) and 10% used for cross-validation ($R_{\text{free}}$). To encourage adjustment of the model to the data, the two molecules in the asymmetric unit were initially treated as two single rigid bodies. Next, simulated annealing was carried out, using the data from 50.0-2.5 Å resolution. This reduced the $R_{\text{factor}}$ to 0.32 and $R_{\text{free}}$ to 0.35. Several cycles of model building and simulated annealing were then performed and water molecules were added reducing the $R_{\text{factor}}$ to 0.23.
Figure 5-1  The crystal packing of the rTxln-1-trypsin complex. rTxln-1 is yellow and trypsin is purple. The unit cell is shown in green with the a and c axes marked in red.

This crystal structure was used as the starting point for further refinement when a second data set, collected to 1.63 Å was obtained. This refinement was carried out in REFMAC5 (Murshudov et al., 1997), with modeling using Coot (Emsley and Cowtan, 2004). Data used in the refinement ranged from 37.42–1.64 Å, with 5% of reflections used for the $R_{free}$ test set (2413 reflections). All atoms were refined with individual isotropic B-factors. Arp_water (Lamzin and Wilson, 1993) was used to place all of the water molecules and a Ca$^{2+}$ ion.

5.4 Refinement statistics and quality of rTxln-1-trypsin complex crystal structure

The final model includes residues 3-58 of rTxln-1 and residues 16-245 of bovine trypsin. The electron density map is, in general, well defined for both the rTxln-1 molecule and trypsin molecule and an unambiguous trace of the backbone atoms could be made. One Ca$^{2+}$ ion with 60% occupancy and 279 water molecules were also fitted to the electron density. The final $R_{factor}$ and $R_{free}$ were 0.21 and 0.24, respectively. The difference of only 3% suggests that the model was not overrefined.
or biased. The overall geometry of the model is excellent with an rmsd from ideal bond lengths of 0.019 Å and an rmsd from ideal bond angles of 1.67° (Table 5-1).

Table 5-1  Refinement statistics for the rTxln-1-trypsin complex

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-hydrogen atoms</td>
<td></td>
</tr>
<tr>
<td>rTxln-1</td>
<td>445</td>
</tr>
<tr>
<td>Trypsin</td>
<td>1629</td>
</tr>
<tr>
<td>Total number of protein atoms</td>
<td>2074</td>
</tr>
<tr>
<td>Solvent (H2O)</td>
<td>279</td>
</tr>
<tr>
<td>Ca2+</td>
<td>1</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>37.42 – 1.64</td>
</tr>
<tr>
<td>( R_{\text{factor}} )</td>
<td>0.2127</td>
</tr>
<tr>
<td>( R_{\text{free}} )</td>
<td>0.2455</td>
</tr>
<tr>
<td>rmsd from ideal geometry</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.019</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.67</td>
</tr>
</tbody>
</table>

\( R_{\text{factor}} = \frac{\sum||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum|F_{\text{obs}}|} \), 5% of the data were excluded from the refinement were used to calculate \( R_{\text{free}} \).

The average B-value for the main-chain (side-chains) atoms is 29.8 Å² (33.7 Å²) for rTxln-1 and 22.4 Å² (24.4 Å²) for trypsin (Table 5-2). Thus, rTxln-1 is on average marginally more mobile than trypsin. Residues 26–31 and the N-terminal and C-terminal residues in the rTxln-1 and the C-terminal residues and the solvent exposed loop consisting of residue 125-155 in trypsin have the highest B-values (Figure 5-2; Figure 5-3; Figure 5-4). The rigid positions of Pro15, Cys16 and Arg17 in the canonical loop are demonstrated by a local minimum B-value for both backbone and side-chain atoms (Figure 5-2). In trypsin, B-values for residues in the catalytic triad His57, Asp102 and Ser195 are lower than the average B-values for the other atoms in trypsin (Figure 5-3). This B-value analysis which shows gradual changes in values throughout the polypeptide chains supports that the structure has been correctly refined. Moreover, high B-values are generally associated with mobile loop regions, N- and C-termini or exposed side-chains while low B-values correspond to the core amino acids residues.
Table 5-2  Mean B-values (Å²) for the rTxln-1-trypsin complex

<table>
<thead>
<tr>
<th></th>
<th>rTxln-1</th>
<th>Trypsin</th>
<th>H₂O (279 atoms)</th>
<th>Ca²⁺ (1 atom)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>main-chain</td>
<td>29.8</td>
<td>22.4</td>
<td>35.1</td>
<td>23.0</td>
<td>23.8</td>
</tr>
<tr>
<td>side-chain</td>
<td>33.7</td>
<td>24.4</td>
<td>23.0</td>
<td>28.5</td>
<td>28.5</td>
</tr>
<tr>
<td>all atoms (445 atoms)</td>
<td>31.7</td>
<td>23.2</td>
<td>23.0</td>
<td>26.2</td>
<td>26.2</td>
</tr>
</tbody>
</table>

Mean B-values were calculated using the CCP4 program BAVERAGE (CCP4, 1994)

Figure 5-2  B-values (Å²) for rTxln-1 in the complex. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values.
Figure 5-3  B-values (Å$^2$) for trypsin in the complex. Glycine residues have no side-chain B-value.
Figure 5-4  Visualization of the B-values for the rTxln-1-trypsin complex. High main-chain high B-values are represented by red, thick tubes. Low B-values are represented by blue, thin tubes. Side-chains are shown as sticks and are color coded.

The Ramachandran plot shows 87.6% of residues in the most favored regions, 12.0% of residues in additionally allowed regions, and 0.4% in generously allowed regions (Table 5-3). There are no residues in disallowed regions. The only residues of rTxln-1 outside the most favored regions in the plot are Glu41, Asn43, Asn45 and Asn46. These residues are in additionally allowed regions except Glu41 which is in a generously allowed region, as previously identified in free rTxln-1 (Chapter 4). Arg17, Val18 and Arg19 which are part of the canonical loop all have favorable dihedral angles. There are no trypsin residues in disallowed regions. His57 and Ser195 of the catalytic triad have backbone dihedral angles in the most favored regions, while Asp102 is in a generously allowed region (Figure 5-5).

| Table 5-3  Ramachandran plot statistics for the rTxln-1-trypsin complex |
|---------------------------------|----------|
| Residues (%)                   |          |
| Most favored regions           | 87.6     |
| Additionally allowed regions    | 12.0     |
| Generously allowed regions      | 0.4      |
| Disallowed regions              | 0.0      |
Figure 5-5 Ramachandran plot of the rTxln-1-trypsin complex. A (α-helix), B (β-sheet) and L (left handed α-helix) are the most favoured regions. a, b and c are additionally allowed regions, and ~a, ~b, ~l, ~p are generously allowed regions. Glycine residues are identified as triangles. Residues of rTxln-1 are denoted by an I after their residue number. Arg17I, Val18I, Arg19I, Glu41I, Asn43I, Asn45I and Asn46I of rTxln-1 and the catalytic triad His57, Asp102 and Ser195 of trypsin are marked with white circles. This plot was generated by Procheck (R. A. Laskowski et al., 1993).

5.5 Crystal structure of the rTxln-1-trypsin complex

The rTxln-1-trypsin complex measures 56 Å in length, with the trypsin and rTxln-1 having dimensions of 32 x 45 x 36 Å and 35 x 17 x 15 Å, respectively. The complex has the shape of a mushroom, with trypsin acting as the head and rTxln-1 as the stalk. The trypsin model consists of one 223 amino acid polypeptide chain that comprises two open-ended β-barrels, one in the N-terminal domain with 7 strands, and one in the C-terminal domain with 6 strands (Figure 5-6). The active site is made up of loop regions from each domain and is situated in a crevice between the two domains, with the catalytic residues His57, Asp102, and Ser195 bridging the two barrels. The calcium binding loop consists of residues Glu70-Glu80 which connects two antiparallel β-strands. This ion is bound by six ligands, Glu70Oε1, Asn72O, Val75O, Glu80ε2 and two water molecules positioned at the edges of an almost regular octahedron. Trypsin and the rTxln-1 have the same secondary structures as observed in their free structures.
Figure 5-6 Stereodiagram of the overall fold of the rTxln-1-trypsin complex. The rTxln-1 is colored yellow (β-strands), red (α-helices) and green (loops). Trypsin is colored purple (β-strands), blue (α-helices), pink (loops) and green (Ca²⁺ as a sphere). Shown is how rTxln-1 docks into the crevice created by the two β-barrels in trypsin.

5.6 The interface between rTxln-1 and trypsin

The total buried surface area in the complex is 1354 Å², 588 Å² from rTxln-1 and 768 Å² from trypsin (CCP4 AREAIMOL, (B. Lee and Richards, 1971)). Only a minor portion of both molecules are in contact in the complex with 8 residues out of 59 of the inhibitor, and 15 residues out of 225 of trypsin forming interactions. However, the fit is highly complementary and the interface is tightly packed due to a complex network of hydrogen bonds, an ionic bond, hydrophobic interactions and numerous van der Waals contacts. The majority of interactions are between residues 15-19 (P3-P2') and trypsin (Figure 5-7). This region binds in a canonical-like manner to trypsin (Bode and Huber, 1992). Residues 36-40 (secondary binding loop) also form interactions with trypsin.
Figure 5-7 Interactions between rTxln-1 and trypsin. Distances between atoms are marked with dashed lines and are in Å units. rTxln-1 residues are labeled in red and are identified by an I. Trypsin residues are labeled in blue. An ionic bond is formed between Arg17(I) and Asp189 which is at the bottom of the specificity pocket of trypsin. Ser195 of trypsin makes a closer than van der Waals interaction (2.68 Å) with the carbonyl carbon of Arg17(I). (Figure generated using ChemSketch (ACD/Labs, 2003).

One common structural feature in canonical inhibitor serine protease complexes is an antiparallel β-sheet interaction formed between the P1-P3 residues, Pro15-Cys16-Arg17 in rTxln-1, and three residues in trypsin (Gly216-Trp215-Ser214). In this complex a hydrogen bond between Pro15O(I) and Gly216O of trypsin (3.01 Å) is observed but Arg17N(I) and Ser214O are just outside hydrogen bonding distance (3.31 Å). A kink in the rTxln-1 polypeptide at Pro15(I) puts its amide nitrogen 4.48 Å from Gly216O of trypsin. However, Cys16O(I) and the side-chain nitrogen of Gln192 do form a hydrogen bond (2.90 Å). Pro15(I) is further stabilized by hydrophobic contacts to Trp215 (Figure 5-7).
The S1 specificity pocket in trypsin is lined by backbone atoms from residues 190-195, 214-220 and 226-227, the disulfide bridge Cys191-Cys220, and the side-chains of Tyr228 and Ser190. Asp189 is situated at the bottom of this pocket. The P1 residue of rTxln-1, Arg17(I), protrudes into this pocket with the guanidine group forming an ionic bond with the carboxylate of Asp189 (2.67 Å and 2.87 Å) and a hydrogen bond to Ser190O (3.15 Å) and Ser190Oγ (2.91 Å). While Arg17Nη2(I) hydrogen bonds to Gly219O (2.85 Å). There are two buried water molecules present in the pocket. One bridges between Arg17Nη1(I) (3.23 Å) and Val227O (3.04 Å). The other water molecule bridges between Arg17Nε(I) (2.86 Å), Gly216O (2.87 Å) and Gln192Ne2 (2.91 Å). The carbonyl oxygen of Arg17(I) extends towards the oxyanion hole formed by the backbone atoms of Gly193N (2.75 Å) and Ser195N (2.85 Å) (Figure 5-7).

The S1' subsite in trypsin is lined by the side-chains of Phe41, and His57, and the Cys42-Cys58 bridge. The side-chain of the P1' residue, Val18(I), points inside this pocket and is stabilized by hydrophobic interactions with the side-chains of these residues. The S2' subsite is a half open channel bordered by the side-chains of Tyr39 and Tyr151. The side-chain of the P2' residue, Arg19(I), fills this site but is also exposed to solvent and forms a hydrogen bond to a water molecule (2.80 Å) which bridges to His40O (3.06 Å) of trypsin. Arg19N(I) also forms a hydrogen bond to Phe41O (3.16 Å) (Figure 5-7).

The disulfide bond between Cys16(I) and Cys40(I) connects the canonical loop to the secondary loop. It interacts with the surface of trypsin by forming a hydrophobic contact with side-chain of Leu99 (~ 3.7 Å). Both the Gly38(I)Cα and the side-chain of Ile36(I) are in hydrophobic contact with the side-chain of Gln192 (3.8 Å and 3.4 Å, respectively) (Figure 5-7).

The reaction carried out by trypsin is to cleave the P1-P1' bond. It is believed that initially, Ser195 transfers a proton from its side chain hydroxyl group to the imidazole ring of His57 in trypsin. This activates the hydroxyl oxygen of Ser195 to attack the carbonyl carbon atom of the susceptible peptide bond in the substrate, leading to the formation of a covalent bond between the carbonyl carbon atom in the substrate and Ser195Oγ (Barrett et al., 1998; Radisky et al., 2006). In the rTxln-1-
trypsin complex Ser195Oγ is 2.68 Å away from the P1 carbonyl carbon atom. This distance is therefore shorter than a van der Waals bond (>3.4-4 Å), but longer than a covalent bond (~1.43 Å). The bond angle between Ser195Cβ-Ser195Oγ-Arg17C is 103.3°. This angle is close to that for a sp³ bent hybridization configuration of an oxygen atom in a water molecule (the H-O-H angle in water is 105°). His57Neε2 in trypsin is 2.62 Å away from Ser195Oγ, which is the shortest distance to any of its surrounding neighbors, with a bond angle of 97.4° for His57Cε₁-His57Neε₂-Ser195Oγ. The His57Neε₂ is therefore ideally placed to deprotonate Ser 195Oγ activating the oxygen for nucleophilic attack on the carbonyl carbon of the P1 residue. However, the normal proteolytic reaction is arrested somewhere along the reaction coordinate between the Michaelis complex and the tetrahedral intermediate (Figure 5-8).

![Figure 5-8](image)

Figure 5-8 Stereo view of rTxln-1 bound in the active site of trypsin. Residues Arg17 and Val18 of rTxln-1 are yellow, the catalytic triad of trypsin (His57, Asp102 and Ser195) is purple. Oxygen and nitrogen atoms are red and blue respectively. The hydrogen atom (grey) of Ser 195Oγ has been modeled into the active site to show it bonding to His57Neε₂. Distances are in Å units.

The electron density around the scissile peptide bond is well defined. It is clear that in the crystal structure no tetrahedral intermediate formation has taken place, and that the P1-P1' bond is still intact. The shape of the electron density suggests that there is some distortion towards the tetrahedral intermediate. Unfortunately, it is not possible to determine the extent of movement towards the intermediate in detail at a resolution of 1.64 Å (Figure 5-9).
5.7 Crystal contacts

Mass spectrometry (Chapter 2) indicated that not only a 1:1 complex of trypsin and rTxln-1 could form upon incubation ($M_W = 30$ kDa) but also a species of 37 kDa was present, suggesting that one molecule of trypsin could combine with two molecules of rTxln-1. An analysis of the crystal packing contacts could reveal how the ternary
complex could form. Therefore, all possible crystal contacts were investigated to see if two rTxln-1 molecules docked to trypsin and if the important features of rTxln-1 or trypsin were affected by these contacts.

Analysis of the crystal packing revealed that rTxln-1 and trypsin are forming crystal contacts. In forming the complex, one rTxln-1 molecule binds in the active site of trypsin but a second rTxln-1 molecule within the crystal is observed to bind tightly to trypsin with a buried surface area of 852 Å² (Figure 5-11(a)). Contact is through interactions between the α-helix of rTxln-1 and the C-terminal α-helix of trypsin. In particular the side-chains of Ile48(I) and Thr49(I) are shielded from the solvent by Cys128 and Cys232 of trypsin. A hydrogen bond is formed between one side-chain oxygen atom of Glu51(I) and the side-chain atom Asn233Nδ2 of trypsin (3.11 Å). This oxygen atom also forms a hydrogen bond to a water molecule (2.62 Å), which bridges to Ser178O (2.61 Å) and Lys230Nζ (2.76 Å) of trypsin, and to another water molecule (2.39 Å) bridging to Asn233Oδ1 (2.76 Å) of trypsin. The second side-chain oxygen atom of Glu51(I) forms a hydrogen bond to Ser178Oγ (2.74 Å). The side-chain of Glu52(I) locks neatly into the C-terminal α-helix of trypsin, with its side-chain interacting with the backbone nitrogen atoms of Val235, Ser236 and Trp237 and with three water molecules (Figure 5-11(b)). This is the only surface where a second molecule of rTxln-1 interacts with trypsin and is therefore the only candidate representative to show how a second molecule of rTxln-1 could bind to trypsin based on crystal packing.
Figure 5-11 Crystal contacts between a symmetry related rTxln-1 molecule and trypsin. rTxln-1 is yellow and trypsin is purple. (a) The buried surface area between the C-terminal α-helix of rTxln-1 and the C-terminal α-helix of trypsin is 852 Å². (b) Residues and waters that form interactions at the interface. Side-chains are represented as sticks with carbon in yellow for rTxln-1 and purple for trypsin, nitrogen atoms are in blue and oxygen atoms are in red. Water molecules are shown as red spheres.

Figure 5-12 Two of the largest interfaces in the crystal packing of rTxln-1-trypsin complex. The rTxln-1 is yellow and trypsin purple. Two large contact areas one 842 Å² and the other 787 Å² are formed between trypsin molecules.
Two large buried surface areas are formed between the central trypsin molecule and two neighboring trypsin molecules at 842 Å$^2$ and 787 Å$^2$ (Figure 5-12). Neither of these contacts is in the active site or affects the conformation of the canonical loop or the secondary loop of rTxln-1. Thus none of the crystal contacts appears to influence how rTxln-1 inhibits trypsin in this structure.

5.8 Comparison between rTxln-1 in the trypsin complex and free rTxln-1

To study whether the interaction between rTxln-1 and trypsin induces conformational changes, the structures of the free and complexed rTxln-1 to trypsin are compared. The B-value graphs for all atoms can also be compared as a measure of differences in conformational dynamics between the rTxln-1 structures.

Visual analysis of a superposition of rTxln-1 bound to trypsin and molecule A, B and C of free rTxln-1, reveals that the overall main-chain overlays well, the disulfide bonds all have the same chirality and the location of three internal water molecules is completely conserved (Figure 5-13). Least-square fitting of $C\alpha$ atoms (and all atoms including side-chains) for residues 3-58 of rTxln-1 bound to trypsin compared to molecules A, B and C of free rTxln-1, resulted in rmsd values of 0.64 Å (1.96 Å), 0.41 Å (1.72 Å), and 0.88 Å (2.63 Å), respectively. The difference in the canonical loop structure explains the high rmsd when molecule C of rTxln-1 is compared with trypsin bound rTxln-1. The rmsds between residues in the canonical loop (residues 15-20), comparing trypsin bound rTxln-1 with molecule A, B and C of free rTxln-1 are 0.49 Å (2.71 Å), 0.24 Å (2.39 Å), and 1.46 Å (5.43 Å), respectively, showing that molecule B and to lesser extent molecule A in the free rTxln-1 structure mimic the structure when bound to trypsin. Comparing only trypsin bound rTxln-1 to molecule A and B of free rTxln-1, the dihedral angles for Arg17, Val18 and Arg19 are $\Phi=-108.3^\circ$ and $\Psi=32.6^\circ$, $\Phi=-73.24^\circ$ and $\Psi=159.4^\circ$ and $\Phi=-103.8^\circ$ and $\Psi=110.4^\circ$, respectively, in trypsin bound rTxln-1, $\Phi=-126.35^\circ$ and $\Psi=33.1^\circ$, $\Phi=-96.0^\circ$ and $\Psi=101.6^\circ$ and $\Phi=-83.1^\circ$ and $\Psi=102.9^\circ$, respectively, in molecule A of free rTxln-1, and $\Phi=-125.9^\circ$ and $\Psi=36.4^\circ$, $\Phi=-110.5^\circ$ and $\Psi=176.7^\circ$ and $\Phi=-103.8^\circ$ and $\Psi=110.4^\circ$, respectively, in molecule B of free rTxln-1. Thus, the dihedral angles for Arg17, Val18 and Arg19 vary less between trypsin bound rTxln-1 and molecule B of free rTxln-1 (the total variation is 113.0$^\circ$) than trypsin bound rTxln-1 and molecule A of
free rTxln-1 (the total variation is 127.1°). Thus overall there is only a small difference between the structure of the backbone of the canonical loop when bound to trypsin and in molecule A and B of free rTxln-1. However, most of the side-chains in trypsin bound rTxln-1 and molecule A and B of free rTxln-1 have different dihedral angles. The side-chain of Arg17 has to rotate its dihedral angle $\chi_4$ 120° in molecule A of free rTxln-1 and its dihedral angle $\chi_3$ 120° in molecule B of free rTxln-1 to obtain the elongated conformation observed in the trypsin bound rTxln-1. The side-chain of Val18 in molecule A of free rTxln-1 has the same $\chi_1$ as observed in the rTxln-1-trypsin complex, but in molecule B of free rTxln-1 this value differ by 100°. The dihedral angles $\chi_1$, $\chi_3$ and $\chi_4$ of Arg19 in molecule A of free rTxln-1 have to rotate by 210°, 70° and 160°, while the dihedral angles $\chi_3$ and $\chi_4$ of Arg19 in molecule B of free rTxln-1 have to rotate by 160° and 180° to have a similar conformation as Arg19 in trypsin bound rTxln-1 (Figure 5-14). Thus, the side-chains of Arg17 and Arg19 in molecule A and B of free rTxln-1 and the side-chain of Val18 in molecule B of free rTxln-1 have to adjust their dihedral angles to dock into trypsin.

**Figure 5-13** Stereo view of the superposition of free rTxln-1 and trypsin bound rTxln-1. Trypsin bound rTxln-1 is yellow, molecule A of free rTxln-1 is green, molecule B of free rTxln-1 is blue and molecule C of free rTxln-1 is purple. Disulfide bonds are displayed as orange sticks. The canonical loop of trypsin bound rTxln-1 has a similar conformation as molecule A and B of free rTxln-1 but is different compared to molecule C of free rTxln-1. The distance between Val18Ca of trypsin bound rTxln-1 and free rTxln-1 molecule C is shown as a dashed black line and is in Å units.
B-values reflect to a first approximation positional variation of atoms over time and are therefore provide insight into the dynamics of the protein (Ringe and Petsko, 1986; Parthasarathy and Murthy, 2000). However, B-values are also influenced by crystal order, diffraction resolution, atom occupancy and restraints used for the refinement. The crystal morphology and size varied between the crystals of rTxl-1-trypsin and free rTxl-1, and most likely the overall order. It is evident from Figure 5-15 that the overall patterns of the main-chains and side-chains B-values are similar for the structures of trypsin bound rTxl-1 and free structures. However, the main-chain B-values for trypsin bound rTxl-1 are systematically larger than those for the free rTxl-1 structures. The overall patterns of the side-chain B-values are similar. The most striking change upon forming the complex is the reduction in B-factor value of the side-chain of Arg17. The main-chain B-values for residues 18-22 are in a more obvious peak in the graph in the rTxl-1-trypsin complex compared to the corresponding graph for molecule B and C of the free structure, and is a dip in the free structure of molecule A, which is a bit surprising as this indicates that this part of the canonical loop has not found a stabilizing position upon forming the complex with trypsin (Figure 5-15; Figure 5-16). The side-chain of Glu52 is in a dip which is due to the crystal contacts described in Section 5.7 (Figure 5-15).
Figure 5-15  B-values (Å²) for rTxln-1 bound to trypsin and the three free rTxln-1 molecules.
The orange blocks highlight the canonical loop and secondary loop region. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values.
Figure 5-16 Visualization of the B-values for rTxln-1 bound to trypsin. The main-chain of the rTxln-1 molecule is displayed with high B-values in red, thick tubes, low B-values in blue, thin tubes and side-chains as sticks. The Connolly surface of the trypsin molecule is colored grey. (a) rTxln-1 docked to trypsin. (b) Zoom in of the active site of trypsin. The side-chain of Arg17 is locked into the specificity pocket of trypsin. Arg19 fits loosely in the active site.

5.9 Comparison of the rTxln-1 trypsin complexes with an aprotinin trypsin complex

To analyze differences in binding modes between rTxln-1 and aprotinin with trypsin (pdb file 2PTC determined to 1.90 Å resolution (Marquart et al., 1983)), the structures of the complexes are compared.

Visual analysis of a superposition of rTxln-1 bound to trypsin and aprotinin bound to aprotinin bound to trypsin, shows that the main-chain atoms overlay well and the disulfide bonds have the same chirality. Least-squares fitting of topologically equivalent Cα atoms between rTxln-1 and aprotinin bound to (residues 3-59 in rTxln-1 and 1-57 in aprotinin) gives an rmsd of 0.83 Å. The canonical and secondary loops do not deviate significantly in their backbone structures. However, a large difference between the two structures is in the region between residues 40-45 (rTxln-1 numbering). This variation is discussed in Chapter 4, since it is also present when the free molecules are compared (Figure 5-17).
Figure 5-17 Superposition of rTxlN-1 and aprotinin bound to trypsin.

rTxlN-1 and aprotinin bound to trypsin (2PTC) are colored yellow and green respectively. Disulfide bonds are shown as orange sticks. The distance between Gly42 of rTxlN-1 and Ala40 (red) aprotinin is 3.07 Å.

On visual inspection of a superposition of the trypsin molecules, the main-and side-chain atoms overlay well, with the residues of the catalytic triad perfectly superposed. The least-squares fitting of topologically equivalent Cα (and all atoms including side-chains) of the two trypsin molecules gave an rmsd of 0.53 Å (0.64 Å) (residues 16-245). Interestingly, however, the superposition shows that the rTxlN-1 and aprotinin have different docking angles in their complexes with trypsin. The difference is ~24.5° (Figure 5-18). In this alignment (trypsin alone), the only region where the rTxlN-1 and aprotinin structures overlay is the canonical loop.
Figure 5-18  Superimposition of trypsin in the rTxln-1-trypsin complex and the aprotinin-trypsin complex (2PTC).

rTxln-1 is colored yellow, aprotinin green and trypsin purple. Only the trypsin molecule from the rTxln-1-trypsin complex is shown, the trypsin molecule from aprotinin complex was omitted for clarity. rTxln-1 deviates 24.5° from the binding direction of aprotinin to trypsin.

Least-squares fit of the main-chain atom of the P3-P3' residues of rTxln-1 and aprotinin gives an rmsd of 0.44 Å, which verifies that the backbone conformation of this region is conserved in the two proteins. The main-chain atoms of the P3-P1 residues, overlay best with an rmsd of 0.17 Å while the P1'-P3' residues overlay with an rmsd of 0.36 Å. This implies that the P3-P1 region of the canonical loop is bound in a more consistent manner to trypsin than the P1'-P3' (Figure 5-19). This is mainly due to the short antiparallel β-sheet formed between residues 214-217 of trypsin and the P3-P1 residues of rTxln-1 or aprotinin and the conserved location of the specificity pocket and the oxyanion hole in trypsin.
Figure 5-19 Residues P3-P4' in rTxlN-1 and aprotinin.
Superimposition of the trypsin in the rTxlN-1-trypsin complex and in the aprotinin-trypsin complex (2PTC). rTxlN-1 is colored yellow, aprotinin green. Oxygen is red, nitrogen blue and sulfur is orange. The Connolly surface of the trypsin molecule from the rTxlN-1-trypsin complex is transparent and shown in purple. (a) Residues P3-P4' are shown as sticks. The disulfide bridge linking the canonical and secondary loops is shown as stick. Residues in rTxlN-1 are labeled in yellow and residues in aprotinin are labeled in white and black. (b) Stereo view.

The differences in sequence of the canonical loop are Arg17 in rTxlN-1 vs Lys15 in aprotinin, Val18 in rTxlN-1 vs Ala16 in aprotinin and Phe20 in rTxlN-1 vs Ile18 in aprotinin. The two Arg17Nη atoms in rTxlN-1 form an ionic bond with the carboxylate of Asp189, while Lys15 in aprotinin only uses the 15Nζ atom to form an ionic bond. However, the side-chain of Lys15 in aprotinin does not form a hydrogen bond to Gly219O in trypsin or a water bridge such as the one between Arg17Nε in rTxlN-1 and Gly216O and Gln192Nε2. For Arg17 in rTxlN-1 Φ= -108.3° and Ψ=32.6°, compared with Lys15 in aprotinin where Φ= -116.9° and Ψ=39.1°. The next residue, Val18 in rTxlN-1 is more bulky than Ala16 in aprotinin. For Val18 in rTxlN-1 Φ= -73.2° and Ψ=159.4° compared with Ala16 in aprotinin where Φ= -87.5° and
Thus, the dihedral angles ($\Phi$ and $\Psi$) for the P1 and P1' residues are similar in rTxln-1 and aprotinin.

The dihedral angles for Arg19 in rTxln-1 and Arg17 in aprotinin are $\Phi=-103.8^\circ$ and $\Psi=110.4^\circ$, and $\Phi=-112.4^\circ$ and $\Psi=79.3^\circ$, respectively. Thus $\Phi$ varies by $8.6^\circ$ and $\Psi$ by $31^\circ$. Therefore, the canonical loop structure in the two proteins starts to deviate here on the C-terminal side of the scissile bond. For Phe20 in rTxln-1 its dihedral angles are $\Phi=-122.6^\circ$ and $\Psi=101.3^\circ$ compared to Ile18 in aprotinin where $\Phi=-105.5^\circ$ and $\Psi=121.7^\circ$. Thus for this residue, both dihedral angles vary by $\sim 20^\circ$. Pro21 in rTxln-1 has $\Phi=-149.0^\circ$ and $\Psi=128.0^\circ$. Compared to $\Phi=-89.5^\circ$ and $\Psi=122.4^\circ$ for Ile19 in aprotinin. Thus the $\Phi$ angle varies by $59.5^\circ$. This difference can be described to conformational restriction on main-chain due to the presence of proline in rTxln-1 (Figure 5-19; Figure 5-20).

**Figure 5-20** The P1’-P4’ residues in rTxln-1 and aprotinin. The Connolly surface of the trypsin molecule from the rTxln-1-trypsin complex is shown in purple. Residues in rTxln-1 are shown as stick with carbon atoms in yellow. Residues in aprotinin (2PTC) are shown as sticks with carbon atoms in green. Oxygen is red and nitrogen blue. (a) Stereo view. (b) Residues in rTxln-1 are labeled in yellow. Residues in aprotinin are labeled in white and black. The dihedral angles vary $\sim 20^\circ$ between Phe20 of rTxln-1 and Ile18 of aprotinin. The following residue in the sequence of rTxln-1 is Pro21, while it is Ile19 in aprotinin.
On the N-terminal side of the scissile bond all residues of rTxln-1 and aprotinin, have similar dihedral angles (Figure 5-17; Figure 5-19; Figure 5-21). For Cys16 in rTxln-1 and Cys14 in aprotinin the dihedral angles are $\Phi=-70.5^\circ$ and $\Psi=155.5^\circ$, and $\Phi=-65.7^\circ$ and $\Psi=149.8^\circ$, respectively. For Pro15 in rTxln-1 and Pro13 in aprotinin the dihedral angles are $\Phi=-83.9^\circ$ and $\Psi=-27.4^\circ$ and $\Phi=-77.5^\circ$ and $\Psi=-29.4^\circ$. However, the position of the Gly14C\(\alpha\) in rTxln-1 and Gly12C\(\alpha\) in aprotinin differs with 0.96 Å (Figure 5-21).

**Figure 5-21** The P4-P1 residues in rTxln-1 and aprotinin. The Connolly surface of the trypsin molecule from the rTxln-1-trypsin complex is shown in purple. Residues in rTxln-1 are shown as stick with carbon atoms in yellow. Residues in aprotinin (2PTC) are shown as sticks with carbon atoms in green. Oxygen is red, nitrogen is blue and sulfur is orange. (a) Stereo view. (b) Residues in rTxln-1 are labeled in yellow. Residues in aprotinin are labeled in white. The distance between Gly14C\(\alpha\) in rTxln-1 and Gly12C\(\alpha\) in aprotinin is 0.96 Å.
In comparing the docking of aprotinin and rTxln-1 to trypsin all of the side-chains of Val18, Phe20 and Pro21 in rTxln-1 would come too close to the surface of trypsin if it was in the same orientation as aprotinin. The side-chain of Val18 in rTxln-1 would be 2.18 Å from the catalytic His57 and 2.61 Å from Ser195, while the side-chain of Phe20 would be 2.42 Å from His57O and Pro21 would be only 1.04 Å from Tyr39Oη of trypsin. Therefore, rTxln-1 can not dock to the active site of trypsin in the same way as aprotinin. It can be seen in Figure 5-22 that the hydrophobic residues Val18 and Phe20 on the surface and Phe35 and Tyr37 which are buried in rTxln-1 move in unison with the residues of the inhibitor while the P3-P1' residues in the canonical loop remain fixed.

![Figure 5-22](image-url)

**Figure 5-22** rTxln-1 and aprotinin bound to trypsin. rTxln-1 is colored yellow, aprotinin (2PTC) green. The Connolly surface of the trypsin molecule from the rTxln-1-trypsin complex is shown in purple. Residues in rTxln-1 are labeled in yellow and residues in aprotinin are labeled in white. The side-chains of Val18, Phe20, Tyr37 and Phe35 in rTxln-1, and Ala16, Ile18, Tyr35 and Phe33 in aprotinin are shown as sticks. Oxygen is red.

The secondary loop of rTxln-1 and aprotinin are 2.07 Å apart as measured by the distance between rTxln-1’s Gly39Ca and the aprotinin’s Gly37Ca (Figure 5-23). The major difference in sequences in this loop is a change from Glu41 in rTxln-1 to Arg39 in aprotinin. While the glutamate residue can form water bridges to trypsin, the difference in loop position allows the arginine residue of aprotinin to form a direct hydrogen bond with the carbonyl oxygen atom of trypsin’s Asn97 from trypsin.

The different location of the second loop reflects upon the size of the area of rTxln-1 and aprotinin which contacts trypsin. The buried surface area of aprotinin in contact with trypsin is 658 Å² while for rTxn-1 this value is 588 Å².
In comparing the binding of rTxln-1 and aprotinin to trypsin the interaction between the P3-P1' residues and the enzyme is highly conserved. This is in spite of the change from lysine to arginine at the P1 position. The major difference in binding occurs at the P1'-P3', where there are substantial changes in sequence Ala16-Arg17-Ile18 in aprotinin vs Val18-Arg19-Phe20 in rTxln-1. These changes to the sequence appear to affect the way which the two molecules dock into the active site of trypsin. The $K_i$ value for aprotinin is $1.3 \times 10^4$ times tighter than for rTxln-1. An obvious difference is the change from arginine to lysine at the P1 site which could contribute significantly to the difference in binding. It is also observed that the buried surface area of rTxln-1 is $77 \, \text{Å}^2$ less, suggests that this molecule would have a reduced affinity as compared to aprotinin.

**Figure 5-23** Variation in the location of the secondary loop.
Superimposition of the trypsin structure in the rTxln-1-trypsin complex and in the aprotinin-trypsin complex (2PTC). rTxln-1 is colored yellow, aprotinin green and trypsin purple. The Connolly surface of the trypsin molecule from the rTxln-1-trypsin complex is shown in purple. Residues of rTxln-1 are labeled in yellow and residues of aprotinin are labeled in green. Distance is a dashed black line and are in Å units. Glu41 in rTxln-1 and Arg39 in aprotinin are shown as sticks with nitrogen atoms in blue and oxygen atoms given in red. Disulfide bonds are orange. The canonical loops overlay well, while the shorter loop in rTxln-1 has moved as compared to aprotinin. Glu41 of rTxln-1 can make water mediated contact to trypsin and Arg39 of aprotinin makes direct contact with trypsin.
5.10 Summary
The structure of the rTxIn-1-trypsin complex has been determined by molecular replacement to a resolution of 1.64 Å. rTxIn-1 docks to trypsin through a canonical loop and a secondary loop in a similar mode to that which occurs when aprotinin binds to trypsin. The side-chain of Arg17 protrudes into the specificity pocket of trypsin forming an ionic bond with Asp 189. The distance between the carbonyl carbon of Arg17 and the catalytic Ser195Oy of trypsin is sub van der Waals. At this resolution, it is not possible to tell if the carbonyl carbon of Arg17 is pyramidal or distorted. However, it is clear that the complex is arrested somewhere along the reaction coordinate between the Michaelis complex and the formation of the tetrahedral intermediate. The secondary loop of aprotinin docks closer to the surface of trypsin than rTxIn-1. This is reflected in the greater surface area of aprotinin that is in contact with trypsin than rTxIn-1. If rTxIn-1 would dock as aprotinin to trypsin, the side-chains of Val18, Phe20 and Pro21 would clash with the surface of trypsin.
Chapter 6
Crystal structure of the rTxln-1 recombinant human microplasmin complex

The primary function of plasmin is to remove intravascularly formed thrombin by the degradation of fibrin. rTxln-1 is a serine protease inhibitor that works as an antifibrinolytic agent by blocking the activity of plasmin and/or other proteases involved in fibrinolysis. rTxln-1 has been shown to inhibit plasmin with a $K_i$ of 14 nM, Table 1-2 in Chapter 1 (Filippovich et al., 2002; Flight et al., 2007). It is therefore highly likely that the physiological target for rTxln-1 is plasmin.

In surgery, the plasmin inhibitor aprotinin is administered to control excessive bleeding. It has previously been shown that rTxln-1 is a less potent and more rapidly reversible inhibitor of plasmin in comparison to aprotinin and therefore less likely to cause the adverse risk of thrombosis (Masci et al., 2000; Filippovich et al., 2002).

Plasmin is centrally involved in a number of physiological processes other than hemostasis. The phenotype of homozygous plasminogen-deficient mice shows not only serious thrombosis (Bugge et al., 1995; Ploplis et al., 1995), but also impaired wound healing (Romer et al., 1996) and reduced fertility (Ploplis et al., 1995). It is believed that plasmin functions either directly or indirectly in processes such as macrophage invasion in inflammation, breaking of the follicular wall for ovulation (Reich et al., 1985), angiogenesis (Gross et al., 1983) and keratinocyte accumulation after wound healing (Morioka et al., 1987). Plasmin can also degrade extracellular matrix molecules such as fibronectin, laminin, thrombospondin and type IV collagen. (Edmonds-Alt et al., 1980; Ott et al., 1982; Coligan and Slayter, 1984), activates metalloproteinases (Stricklin et al., 1977; He et al., 1989; Shah, 1997) and collagenase (Gross et al., 1983), protolyses mediators of the complement system (Pillemzer et al., 1953; Ratnoff and Naff, 1967), modify coagulation factors (Omar and Mannj, 1987) and activate growth factors (Sato and Rifkin, 1989; Khalil et al., 1996).
The crystal structures of microplasmin in complex with staphylokinase (Parry et al., 1998) or with streptokinase (X. Wang et al., 1998) have helped to understand why plasmin is such a diverse molecule. For example, the purpose (Parry et al., 1998) of the α domain of streptokinase (X. Wang et al., 1998; Loy et al., 2001; Wakeham et al., 2002) is to create a new subsite for improved substrate presentation in plasmin. The suggested function of the γ domain of streptokinase (X. Wang et al., 1998) or the N-terminal flexible peptide of streptokinase (Terzyan et al., 2004) is to induce conformational changes in microplasminogen for microplasmin activation.

Knowledge of the crystal structure of the rTxlN-1-microplasmin complex will help to establish the structural features necessary for the development of a plasmin selective inhibitor. It is suggested that rTxlN-1 binds in a similar fashion to microplasmin as it is bound to trypsin but there will be differences since the substrate binding sites for plasmin and trypsin are different. Mutagenesis studies have been carried out to identify the binding residues of rTxlN-1 (Filippovich et al., 2002). Plasmin inhibition was completely lost with the Arg17Ala mutation. This further suggested that side-chain of Arg17 binds in the specificity pocket of plasmin (Filippovich et al., 2002). An analysis of a crystal structure of the rTxlN-1-microplasmin complex allows the determination of the contacts between rTxlN-1 and microplasmin. Comparison with a structure of the free rTxlN-1 (Chapter 4) will show whether the structure of rTxlN-1 is altered upon binding. rTxlN-1 is speculated to induce a non-reactive conformation of the active site residues of microplasmin not seen in the other structures of microplasminogen/microplasmin. Knowledge of the rTxlN-1 inhibited microplasmin structure is useful to compare with structures of both microplasminogen (Peisach et al., 1999; Terzyan et al., 2004), staphylokinase and streptokinase inhibited microplasmin. As a result, a selective plasmin inhibitor might be able to control inflammation, cancer metastasis and angiogenesis (Ossowski and Reich, 1983; Ossowski, 1992; Chapman, 1997).

This chapter describes the procedures used to solve the structure of the rTxlN-1-microplasmin complex, a description of the structure and comparison of this structure with free rTxlN-1, rTxlN-1-trypsin complex, microplasmin-staphylokinase, microplasmin-streptokinase and microplasminogen structures.
The identity of the amino acids in the peptide substrate is as described in Chapter 5.

6.1 Phase determination of rTxln-1-microplasmin complex

The molecular replacement program EPMR (Kissinger et al., 1999) was used to solve the phase problem. The coordinates of microplasmin from the complex with streptokinase (pdb file 1BML) were used as a search model. Based on solvent content and Matthews coefficient arguments (Chapter 3) it was anticipated that there are two complexes per asymmetric unit.

The calculation of correlation coefficient is based on structure factor values. Therefore TRUNCATE converted the intensities to structure factors (French and Wilson, 1978), CAD sorted the data and put it into the correct asymmetric unit for CCP4 (CCP4, 1994) and MTZ2VARIOUS (Dodson, 1992) produced a reflection file for CNS (Brünger et al., 1998).

A search for the location of the two complexes in space group P2₁ using EPMR was performed with data in the resolution range 15-4 Å. The evolutionary search procedure was run ten times and carried out over 50 generations using a population size of 300. In the search for a single molecule of plasmin a correlation coefficient of 0.24 and a crystallographic $R_{\text{factor}}$ of 0.56 were first obtained. The contribution of this first solution was added to the calculation to search for the second molecule. This search was repeated twenty times, again using data from 15-4 Å. The highest correlation coefficient was 0.46 and the lowest crystallographic $R_{\text{factor}}$ was 0.47.

6.2 Refinement of the rTxln-1-microplasmin crystal structure

The two plasmin molecules were refined by simulated annealing in CNS and subsequently using REFMAC5 in maximum likelihood mode. The structure factors were divided into two sets, 90% of the data was used for refinement ($R_{\text{factor}}$) and 10% for cross-validation ($R_{\text{free}}$). Simulated annealing reduced the $R_{\text{factor}}$ and $R_{\text{free}}$ to 0.33 and 0.41, respectively. The electron density was visualized in the graphics program O (Jones and Kjeldgaard, 1998). Inspection of the initial electron density maps revealed the presence of a rTxln-1 molecule bound in the active site of each
microplasmin molecule. The coordinates of rTxln-1 were then manually docked into
the electron density. Examination of the packing within the unit cell indicated
reasonable contacts between the complexes and no overlap of symmetry related
molecules (Figure 6-1). Simulated annealing refinement of the two rTxln-1-
microplasmin complexes reduced the $R_{\text{factor}}$ to 0.29 and $R_{\text{free}}$ to 0.36.

Figure 6-1  The crystal
packing of the rTxln-1-
microplasmin complex.
Microplasmin is blue and
rTxln-1 is yellow. The unit
cell is overlayed.

Subsequent refinement was carried out in REFMAC5 (Murshudov et al., 1997) using
restrained refinement with no prior phase information and the weighting term
between X-ray and geometric residuals set to 0.01. Two different groups of non-
crystallographic symmetry (NCS) restraints were defined; group 1 consisted of
residues 4–56 of rTxln-1, and group 2 consisted of residues 550-559 and 562–785 of
human microplasmin. Tight non-crystallographic symmetry restraints were set for
main-chain atoms and medium restraints for side-chains atoms. Data used in this
refinement ranged from 40.03–2.77 Å, with 5% of reflections used for the $R_{\text{free}}$ test
set (800 reflections). All atoms were refined with individual isotropic B-factors. The
program Arp_water (Lamzin and Wilson, 1993) was used to place water molecules
and the last cycles of model building were carried out with Coot (Emsley and
Cowtan, 2004).
6.3 Refinement statistics and quality of rTxln-1-microplasmin crystal structure

The electron density map of the structure is, in general, well defined for both the rTxln-1 molecules and microplasmin molecules. An unambiguous trace of the backbone of the rTxln-1 and microplasmin molecules could be made. There are two rTxln-1-microplasmin complexes in the asymmetric unit. The final model includes residues 547-559 and 562-791 of human microplasmin and residues 3-58 of rTxln-1 in complex A and residues 545-559 and 562-791 of human microplasmin and residues 3-59 of rTxln-1 in complex B. Due to insufficient electron density for the side chains, the following residues were mutated to alanine in the final model; 552, 554, 557, 582, 583, 585 and 750 in human microplasmin, 3, 5, 8, 30 and 31 in rTxln-1 in complex A, and 546, 554, 582, 583, 585, 615, 629 and 750 in human microplasmin, and 3 and 31 in rTxln-1 in complex B. 43 water molecules and two sulfate ions could also be fitted to the electron density. Refinement with NCS restraints brought the $R_{\text{factor}}$ and $R_{\text{free}}$ down to 0.24 and 0.29, respectively. The difference of only ~5% suggests that the model is not overrefined or biased. The overall geometry of the model is excellent with an rmsd from ideal bond lengths of 0.006 Å and an rmsd from ideal bond angles of 0.98° (Table 6-1).

<table>
<thead>
<tr>
<th>Table 6-1 Refinement statistics of the rTxln-1-microplasmin complexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-hydrogen atoms</td>
</tr>
<tr>
<td>rTxln-1 complex A</td>
</tr>
<tr>
<td>Microplasmin complex A</td>
</tr>
<tr>
<td>rTxln-1 complex B</td>
</tr>
<tr>
<td>Microplasmin complex B</td>
</tr>
<tr>
<td>Total number of protein atoms</td>
</tr>
<tr>
<td>Solvent (H2O)</td>
</tr>
<tr>
<td>Solvent (SO4)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>$R_{\text{factor}}$</td>
</tr>
<tr>
<td>$R_{\text{free}}$</td>
</tr>
<tr>
<td>rmsd from ideal geometry</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
</tbody>
</table>

$R_{\text{factor}} = \frac{\sum|\text{F}_{\text{obs}}| - |\text{F}_{\text{calc}}|}{\sum|\text{F}_{\text{obs}}|}$, 5% of the data were excluded from the refinement were used to calculate $R_{\text{free}}$.

The refinement of the two rTxln-1-microplasmin structures with NCS gave similar B-values for the two rTxln-1 molecules and for the two microplasmin molecules. B-values are strongly correlated to completeness of data and/or restraints (Kleywegt,
1996, 2001). Therefore, a direct comparison of temperature factors between NCS-related molecules is not always possible, since the NCS-related molecules in the crystal sometimes may have different average temperature factors. However, the correlation coefficient of the temperature factors of the NCS-related Cα atoms should be high (Kleywegt, 1996).

The average B-value for main-chain (side-chain) atoms for rTxln-1 in complex A and B are 22.9 Å² (22.7 Å²) and 22.9 Å² (23.2 Å²), respectively, and for microplasmin in complex A and B are 19.6 Å² (19.8 Å²) and 19.7 Å² (19.7 Å²), respectively (Figure 6-2). The same residues in the rTxln-1 structures described in Chapter 4 and 5 for free rTxln-1 and rTxln-1 in complex with trypsin have higher B-values than average in the rTxln-1 of complex A and B. These are the residues 26–31 and the N-terminal and C-terminal residues. In microplasmin, the short solvent exposed chain as well as the N-terminus of molecule B and two solvent exposed loops consisting of residues 583-627 and 625–628 have higher B-values than the rest of the molecule (Figure 6-3; Figure 6-4; Figure 6-2). Overall, the rTxln-1 molecules are marginally more mobile than the microplasmin. The lowest B-values for both backbone and side-chain atoms in rTxln-1 in the complex with microplasmin are residues 15-20 (Figure 6-3). This shows that Arg17 has both a stable main-chain and side-chain which is clearly visible as a dip in the graph over the B-values for the side-chains (Figure 6-3). Both the main-chain and side-chain B-values for the catalytic triad His603, Asp646 and Ser741 of microplasmin are also low (Figure 6-4). No obvious aberrant B-values were noticed for the atoms in the three rTxln-1 molecules in the asymmetric unit. Therefore, this B-value analysis supports that the structure is correct.
Table 6-2 Mean B-values (Å²) for the rTxln-1-microplasmin complexes in the asymmetric unit

<table>
<thead>
<tr>
<th></th>
<th>rTxln-1 in complex A</th>
<th>Micrplasmin in complex A</th>
<th>rTxln-1 in complex B</th>
<th>Micrplasmin in complex B</th>
<th>H₂O (43 atoms)</th>
<th>SO₄ (10 atoms)</th>
<th>Total</th>
<th>main-chains</th>
<th>side-chains</th>
<th>all atoms (4611 atoms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>main-chain</td>
<td>22.9</td>
<td>19.6</td>
<td>22.9</td>
<td>19.7</td>
<td>7.0</td>
<td>30.4</td>
<td>20.3</td>
<td>20.2</td>
<td>20.2</td>
<td></td>
</tr>
<tr>
<td>side-chain</td>
<td>22.7</td>
<td>19.8</td>
<td>23.2</td>
<td>19.7</td>
<td>30.4</td>
<td>20.2</td>
<td>20.2</td>
<td>20.2</td>
<td>20.2</td>
<td></td>
</tr>
<tr>
<td>all atoms (423 atoms)</td>
<td>22.8</td>
<td>all atoms (1838 atoms)</td>
<td>23.0</td>
<td>all atoms (1852 atoms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all atoms (4611 atoms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-2 Visualization of the main-chain B-values in the rTxln-1-microplasmin. High B-values are represented by red thick tubes and lower B-values are represented by blue thin tubes.
Figure 6-3  B-values (Å²) for molecule A (blue) and B (pink) of rTxln-1 in complex with microplasmin. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values.

Figure 6-4  B-values (Å²) for molecule A (blue) and B (pink) of microplasmin mol A and B from respective complexes. Glycine residues have no side chain B-value.
The two complexes within an asymmetric unit refined using NCS would be expected to have a similar set of backbone dihedral angles (Kleywegt, 1996). This, for example, is observed in Figure 6-5, which shows that Arg17 has similar backbone dihedral angles in both molecules. The Ramachandran plot of the complex shows 83.2% of residues in most favored regions, 15.7% of residues in additionally allowed regions, 0.6% in generously allowed regions, and 0.4% in disallowed regions (Table 6-3). Amino acid residues Glu41, Asn45 and Asn46 in rTxl-1 have unusual backbone dihedral angles which are also observed in free rTxl-1 (Chapter 4) and rTxl-1 in complex with trypsin (Chapter 5). Cys57 and Ala 58, in complex B have backbone dihedral angles that are in additionally and generously allowed regions, respectively. Arg17, Val18 and Arg19 in the canonical loop of rTxl-1 are located in approximately the same position in the Ramachandran plot as in the rTxl-1-trypsin complex. In the rTxl-1-microplasmin complex, Arg17 and Arg19 of rTxl-1 have dihedral angles less commonly observed in proteins, while Val18 has dihedral angles found in β-sheet structures of proteins. One residue, Phe715 of microplasmin in complex A and B, is found in a region classified as “generously allowed”, and one N-terminal residue, Asp547 of microplasmin in complex B, is found in a region classified as “disallowed”, in the Ramachandran plot, these residues do not form interactions between rTxl-1 and microplasmin. The His603 and Ser741 of the catalytic triad in microplasmin have backbone dihedral angles most favorable in proteins, while Asp646 has dihedral angles less commonly observed in proteins, similar is observed in the catalytic triad of trypsin in the rTxl-1-trypsin complex (Figure 6-5).

<table>
<thead>
<tr>
<th>Table 6-3 Ramachandran plot statistics for the two rTxl-1-microplasmin complexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in Ramachandran plot (%)</td>
</tr>
<tr>
<td>Most favored regions</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
</tr>
<tr>
<td>Generously allowed regions</td>
</tr>
<tr>
<td>Disallowed regions</td>
</tr>
</tbody>
</table>
Figure 6-5 Ramachandran plot of the rTxln-1-microplasmin complex. A (α-helix), B (β-sheet) and L (left handed α-helix) are the most favoured regions. a, b and c are additionally allowed regions, and ~a, ~b, ~l, ~p are generously allowed regions. Glycine residues are identified as triangles. Residues Arg17, Val18, Arg19, Glu41, Asn45 and Asn46 of rTxln-1 in complex A and B, and Cys57 and Ala58 in complex B are marked with white circles and discussed in the text. The catalytic triad His603, Asp646 and Ser741, and Phe715 of microplasmin in complex A and B, and Asp547 in complex B are marked with white circles and discussed in the text. This plot was generated by Procheck (R. A. Laskowski et al., 1993).

6.4 Crystal structure of the rTxln-1 microplasmin complex

The rTxln-1-microplasmin complex measures 57 Å in length with the microplasmin being 31 x 42 x 39 Å, and the rTxln-1 is 35 x 16 x 15 Å. The complex has a mushroom shape similar to the rTxln-1-trypsin complex, with the microplasmin acting as the head and the rTxln-1 acting as the stalk. The two microplasmin structures in the asymmetric unit consist of two polypeptide chains of 13 and 230 residues (complex A), and 15 and 230 residues (complex B). The short and long chains in each complex are linked by two disulfide bonds (Cys548-Cys666 and Cys558-Cys566). The tertiary structure of microplasmin is similar to trypsin with two open-ended β-barrels, one in the N-terminal domain with 7 strands, and one in the C-terminal domain with 6 strands (Figure 6-6). The active site is situated between the two domains, with the Ca atoms of the catalytic residues His603, Asp646, and Ser741 located in the same positions in the structure as in trypsin. The secondary and
tertiary structure of rTxln-1 bound to microplasmin is similar to that seen in free rTxln-1 and rTxln-1 bound to trypsin.

**Figure 6-6** Stereodiagram of the overall fold of complex A of rTxln-1-microplasmin. The rTxln-1 is colored yellow (β-strands), red (α-helices) and green (loops). Microplasmin is colored blue (β-strands), purple (helices) and pink (loops). The image shows how rTxln-1 docks into the crevice created by the two β-barrels in microplasmin.

After least-squares fitting of topologically equivalent Ca (and all atoms including side-chains) using CCP4 Superpose (E. Krissinel and Henrick, 2004), the two rTxln-1 molecules have an rmsd of 0.21 Å (0.75 Å) for residues 3-58, while the rmsd of the two microplasmin molecules is 0.09 Å (0.68 Å) for residues 545-559 and 562-791. Therefore the polypeptide traces of the rTxln-1 and microplasmin structures are essentially identical within experimental error. However, least-square fitting of the two complexes as a whole shows an rmsd of 0.76 Å (1.01 Å), is significantly larger. Inspection of the superposition shows that there is a difference in the relative docking angle of 7.8° for the two rTxln-1 molecules (Figure 6-7). rTxln-1 interacts with microplasmin through its canonical loop and secondary loop. The canonical loops of the two rTxln-1 molecules superpose well, but there is a difference in the location of the remainder of the structure (Figure 6-10). It can be speculated that the difference in position between the two rTxln-1 molecules is due to crystal packing contacts. However, the different docking angles could also give information about the flexible nature of the interaction between rTxln-1 and microplasmin.
6.5 The interface between rTxln-1 and microplasmin

The total buried surface area in complex A is $1383 \, \text{Å}^2$, $621 \, \text{Å}^2$ from rTxln-1 and $762 \, \text{Å}^2$ from microplasmin and in complex B $1384 \, \text{Å}^2$, $626 \, \text{Å}^2$ from rTxln-1 and $758 \, \text{Å}^2$ from microplasmin (CCP4 AREAIMOL, (B. Lee and Richards, 1971)). The fit between rTxln-1 and microplasmin is highly complementary and tightly focused. Only a small portion of both rTxln-1 and microplasmin are in contact in the complex, with 9 out of 59 residues of rTxln-1 and 15 out of 249 residues of microplasmin forming interactions. The contacts are characterized by a complex network of hydrogen bonds, van der Waals contacts, hydrophobic interactions and ionic bonds.

rTxln-1 is docked into the active site of microplasmin through its canonical loop (residues 15-20) and secondary binding loop (36-40). Residues 15-20 (P3-P3’) are all in direct contact with microplasmin and bind in a canonical-like manner to microplasmin (Bode and Huber, 1992).
Figure 6-8 Interactions between the canonical loop and secondary loop of rTxln-1 and the active site of microplasmin.
Distances between atoms are marked with dotted lines and their lengths are in Å. Values for complex A and B are given. rTxln-1 residues are labeled in red and identified by an I. Microplasmin residues are labeled in blue. Residues 15-20 and 39-40 of rTxln-1 are in direct contact with microplasmin through hydrogen bonds, hydrophobic interactions, van der Waals interactions, cysteine-aromatic ring interaction and an ionic bond. There is a short antiparallel β-sheet formed between Pro15-Arg17 of rTxln-1 and Ser760-Gly762 of microplasmin. An ionic bond is made between the guanidino group of Arg17 of rTxln-1 and the carboxylate group of microplasmin’s Asp735 which is at the bottom of the specificity pocket. Ser741 in microplasmin makes a closer than a van der Waals interaction (2.85 and 3.02 Å in complex A and B, respectively) with the carbonyl carbon of Arg17 in rTxln-1. Figure generated using ChemSketch (ACD/Labs, 2003).
A short antiparallel β-sheet is formed between the S3-S1 residues (Gly762-Trp761-Ser760) and Pro15-Cys16-Arg17, the P3-P1 residues of rTxln-1. In this complex two hydrogen bonds are formed, one is between Pro15O(I) and Gly762N of microplasmin (2.99 Å and 2.77 Å in complex A and B, respectively) and the other is between Arg17N(I) and Ser760O of microplasmin (3.01 Å in both complexes). Pro15(I) makes a kink which positions its amide nitrogen too far away from the microplasmin surface to form any hydrogen bonds with microplasmin. Pro15(I) is however stabilized by hydrophobic contacts to Trp761. Another stabilizing hydrogen bond is between Cys16O(I) and the side-chain nitrogen of Gln738 (2.66 Å and 2.71 Å) (Figure 6-8).

The specificity pocket in microplasmin consists mainly of backbone atoms for residues 760-765 and 736-741, the Cys735-Cys765 disulfide bridge, the side-chains of Tyr774, Ser736 and Asp735 which is situated at the bottom of this pocket. The side-chain of residue Arg17(I) is inserted into the specificity pocket (S1). An ionic bond is formed between the guanidino group of Arg17(I) and the carboxylate of the Asp735 (2.49 Å and 2.76 Å). Hydrogen bonds are also formed between Arg17Nη(I) and Gly764O (2.98 Å and 2.81 Å) and between the second Arg17Nη(I) and Ser736O (2.92 Å and 2.81 Å). In complex B Arg17Nη(I) forms a hydrogen bond with Ser736Oγ (2.68 Å). The carbonyl oxygen atom of Arg17(I) is positioned in the oxyanion hole of microplasmin, fixed via a bifurcated hydrogen bond to Gly739N (2.83 Å and 2.61 Å) and Ser741N (2.85 Å and 3.02 Å) (Figure 6-8). There is a continuous electron density between the Ser741Oγ and the carbonyl carbon Arg17(I). A sub-van der Waals distance (3.00 Å and 2.88 Å) is observed, but due to the limited resolution, the exact geometry around the carbonyl carbon atom cannot be fully described (Figure 6-8).

The S1' subsite is partly open, delimited by the hydrophobic side-chain of Phe587 and the Cys588-Cys604 disulfide bridge. The side-chain of the P1' residue Val18(I) has different conformations in the two complexes in the asymmetric unit (Figure 6-9; Figure 6-11). This result shows that the S1' subsite in microplasmin cannot be completely filled by a single valine side-chain. The S2' subsite is also partly open. The side-chain of Arg19(I) binds in this subsite in an extended conformation pointing towards solvent. The Arg19(I) side-chain is not firmly packed against any
side-chains of microplasmin. However, this subsite has three backbone carbonyl oxygens and the side-chains of both Glu623 and Glu686 point towards the guanidino group of Arg19(I). These negatively charged or polarized groups are all within 5 Å of this side-chain (Figure 6-11). The main-chain of Arg19(I) is stabilized by a hydrogen bond between the amide nitrogen of Arg19(I) and Phe587O (2.84 Å and 3.07 Å). The S3' subsite consists only of the side-chain of Phe587 and the hydrophobic part of the side-chain of Lys607 with the aromatic group of Phe20(I) stacked in between (Figure 6-8). The side-chain of Phe20(I) also stabilizes the side-chain of Val18(I) (Figure 6-11).

An unusual arrangement of the side-chains of the catalytic triad, Ser741-His603-Asp646, is observed in the rTxln-1-microplasmin complex. The side-chain of the catalytic His603 points out of the active site cleft, which creates a more open S1' subsite compared to if the side-chain is in its normal position. This space is occupied by a water molecule hydrogen bonded to Ser741Oγ and one of the side-chain Oδ atoms of Asp646 (2.94 Å and 2.47 Å in complex A, and 3.02 Å and 2.57 Å in complex B). It is not possible to unambiguously assign the position of the Nδ1 and Ne2 atom of the imidazolium ring of His603. However, there is a possible interaction between the side-chain of His603 and the side-chain of Asp646, if the Nδ1 atom is oriented toward the carboxylate of Asp646 (3.36 Å and 3.06 Å in complex A, and 3.23 Å and 2.81 Å in complex B). In either conformation, the imidazolium ring of His603 forms a stabilizing interaction with the Cys16(I)-Cys40(I) disulfide bridge. In this orientation, the His603Ne2 atom is within ~3.5 Å of a water molecule (Figure 6-9).
The canonical loop residues 16-19 and Cys40 of rTxln-1 as yellow sticks. The catalytic triad, His603, Asp646 and Ser741 of microplasmin, is shown as blue sticks. Nitrogen atoms are in dark blue, oxygen atoms in red and disulfide bridge in orange. The water molecules are displayed as red spheres. (a) The $2F_o-F_c$ electron density is displayed as light grey mesh and contoured at $1\sigma$. (b) Both rotamers of Val18 are shown. Polar contacts are marked with dotted black lines and their lengths in Å.

The secondary binding loop of rTxln-1 is stabilized by a hydrogen bond between Gly39O(I) and the side-chain of Glu606 together with the earlier described cysteine-aromatic ring interaction (Figure 6-8). In the Txln-1 structure, the $\phi, \psi$ dihedral angles for Gly39(I) are 122.73° and -22.57° in complex A and 123.27° and -20.97° in complex B, which are generously allowed values for other amino acid residues including alanine and therefore not in a conformational space that could easily be explored by any residue other than glycine. One of the contacts with the shortest
distance in the interface between rTxl-n-1 and microplasmin is between the Gly39O(I) and the carboxylate of Glu606 (2.59 Å and 2.71 Å in complex A, and 2.69 Å and 3.00 Å in complex B) (Figure 6-8). Attempts to model the side-chain of Glu606 in any other position resulted in a negative electron density indicating the wrong location. Positive electron density only appeared at the position close to the Gly39O(I). This suggests that the Glu606 carboxylate group is protonated, which is unusual, since this side-chain is generally assumed to be deprotonated at a pH >3. A water molecule not visible in the electron density or the imidazolium ring of His603 could be the proton donor. The catalytic Ser741Oγ or the water molecule situated outside the active site cleft may have protonated His603 which was then transfered to the carboxylate of Glu606. There is a hydrogen bond between the Glu606 carboxylate and the His603 imidazolium ring. If the His603 has the Nδ1 atom closest to the Glu606 carboxylate, the distance becomes shorter (2.90 Å and 4.23 Å) than if the His603 imidazolium ring Ne2 atom mediated the suggested ionic bond (3.85 Å and 5.16 Å) (Figure 6-10). The distance between Gly39O(I) and His603Ne2 is 4.81 Å in complex A, and 3.91 Å in complex B, alternatively, 4.13 Å in complex A, and 3.10 Å in complex B, if His603Ne1 is placed closest to Gly39O(I). It can be speculated that the Gly39O(I) is protonated instead of the side-chain of Glu606. However, the pKa for protonation of carbonyl oxygen atoms is estimated to be ~0 (Eriksson et al., 1995; Shan and Herschlag, 1996) and it is therefore more likely that the side-chain of Glu606 with a pKa of 4.07 (Dawson et al., 1986) is protonated.
Figure 6-10 Stereo view of the interaction between His603 and Glu606 of microplasmin in complex A.
The carbon atoms of the secondary loop of rTxln-1, consisting of residues 38, 39 and the Cys16-Cys40 disulfide bridge are shown as yellow sticks. His603, Glu606 and Asp646 of microplasmin are shown as blue sticks. Nitrogen atoms are in dark blue, oxygen atoms in red and the disulfide bridge in orange. The water molecule is displayed as a red sphere. (a) The $2F_o-F_c$ electron density is contoured at $1\sigma$ and displayed as a light grey mesh. (b) & (c) show the two different possible orientations of the imidazolium ring of His603 in microplasmin and the different possible contacts marked with dotted black lines. All lengths are in Å.
Figure 6-11 The interface between rTxln-1 and microplasmin in complex A. rTxln-1 residues are shown as sticks, carbon is yellow, nitrogen is blue, oxygen is red, disulfide bridge is orange. A transparent Connolly surface is overlayed. The water molecule that has replaced the side chain of the catalytic His603 in the active site of microplasmin is represented as a green sphere. The Connolly surface of microplasmin is shown with the electrostatic surface mapped. Residues of rTxln-1 and the water molecule are labeled in black and residues of microplasmin are labeled in white except Gln738 which is labeled in grey. The side chain of Arg17 of rTxln-1 protrudes into the specificity pocket and its carbonyl oxygen atom is positioned in the oxyanion hole. Val18 of rTxln-1 is shown with its two different rotamers from complex A and B. The hydrophobic side chain of Phe20 is positioned inside the active site lined up against the side chain of Lys607 of microplasmin. Arg19 reaches out towards solvent. The imidazolium ring of His603 in microplasmin is lined up close to the Cys16-Cys40 bridge in rTxln-1. Gly39O in rTxln-1 makes a hydrogen bond to Glu606 in microplasmin.

An explanation for the difference in location of the secondary loop of rTxln-1 in the active site of microplasmin could be that complex A and complex B show two different steps in the reaction between rTxln-1 and microplasmin. It could be that the different steps are favoured as the result of differences in crystal packing. The larger average B-values for the rTxln-1 molecules compared to the microplasmin molecules support this idea. The side-chains of Asp646 and Ser741 in the catalytic triad overlay almost identically. However, there is a difference in the positions of the side-chains
of the catalytic His603 (maximum 0.6 Å between Cε1 atoms) and the side-chain of Glu606 (maximum 1.4 Å between Oε atoms) (Figure 6-12). A difference of ~1 Å is observed in the location of the Gly39O(I) and the disulfide bridge Cys16(I)-Cys40(I) in rTxln-1 in complex A and complex B (Figure 6-13). It is also obvious that the rTxln-1s have different docking angles to microplasmin (Figure 6-7). Thus, the secondary loop of rTxln-1 has different position in complex A and B of microplasmin. However, it is not possible to deduce that the different positions of the secondary loop are due to the different positions of the side-chains of His603 and Glu606 as a step in the inhibition mechanism. Therefore, it is not clear if the structures of the two rTxln-1-microplasmin complexes show two different steps in the reaction between rTxln-1 and microplasmin.

Possible interpretations are;

- Two modes of binding exist which leads to the observation that there are two molecules per asymmetric unit
- In one of the complexes, the secondary loop has not completely locked in to its optional bending conformation
- The observations are solely due to crystal packing effects
Figure 6-12  Stereo view of the superimposition between complex A and B. The canonical and secondary loop of rTxln-1 from complex A and B colored yellow and green, respectively. Stick representation of rTxln-1’s Cys16-Cys40 disulfide bond and Gly39O. His603 and Glu606 of microplasmin colored light blue and magenta in complex A and B, respectively. Nitrogen atoms are given in dark blue, oxygen atoms in red and disulfide bridge in orange. (a) and (b) 2Fo-Fc electron density map contoured at 1σ for the disulfide bond Cys16-Cys40 and Gly39 of rTxln-1 and His603 and Glu606 of microplasmin, in complex A and B, respectively. (c) Superposition of rTxln-1 and residues His603 and Glu606, from complex A and B. Displaying the differences in location.

Figure 6-13  Comparison of the binding modes of the two molecules of rTxln1 in the asymmetric unit of the rTxln-1-microplasmin complex. rTxln-1 from complex A is colored yellow, rTxln-1 from complex B is colored green. The Connolly surface of microplasmin from complex A is colored blue. Microplasmin from complex B is not shown. The canonical loop of rTxln-1 in complex A and B overlay well, while the secondary loop has moved by up to 1.04 Å (measuring between Cys16Ca in molecule A and B of rTxln-1).
6.6 Crystal contacts

The protein backbone conformation as well as the side-chain conformations are influenced by the surrounding environment, and the dense packed environment of proteins in crystal structures is dissimilar from the environment in the native state of proteins (Janin and Rodier, 1995; Carugo and Argos, 1997). Packing contacts in protein crystals are artifacts from the way the crystals grow and are considered nonspecific protein-protein interactions called “crystal contacts”. Regions of a protein in crystal contact are more rigid and the B factor of atoms having crystal contacts is lower than solvent exposed atoms (Kossiakoff et al., 1992; Jacobson et al., 2002). These contacts make use of the same forces that govern specific recognition in protein-protein complexes and oligomeric proteins. Thus, crystal contacts provide examples of nonspecific protein-protein interaction which can be compared to biologically relevant ones. All possible crystal contacts from surrounding molecules in the rTxlIn-1-microplasmin crystal were investigated to see if the important features of rTxlIn-1 or microplasmin were affected by these artificial contacts and to find out if there are other binding sites than the canonical and secondary loop of rTxlIn-1 and the active site cleft of microplasmin.

Analysis of the crystal packing reveals that there are different crystal contacts for rTxlIn-1-microplasmin complex A and rTxlIn-1-microplasmin complex B (Figure 6-14). The rTxlIn-1 molecules are held in the crystalline array both by crystal packing contacts from symmetry related molecules and through their attachments to the microplasmin molecules. Interfaces formed between the rTxlIn-1 molecules and adjacent molecules are shown in Figure 6-15 and Figure 6-16.
Figure 6-14 Crystal contacts for complex A and B of rTxlIn-1-microplasmin
Connolly surface of complex A (a) and B (b), rTxlIn-1 is yellow and microplasmin is blue. Atom within 5 Å are shown as spheres. Symmetry related rTxlIn-1 atoms are green, symmetry related microplasmin atoms are salmon. Yellow and blue spheres are atoms from the complex within the same asymmetric unit.

There is a small buried surface area of 221 Å² formed between the rTxlIn-1-microplasmin complex B and the rTxlIn-1 molecule from complex A in the asymmetric unit (Figure 6-15; Figure 6-16). In this interface is one crystal contact that could be argued to affect the true native interface between rTxlIn-1 and microplasmin. The side-chain of the P2' residue Arg19 of rTxlIn-1 from complex B, which is situated in the partly open S2' subsite of microplasmin, does not point towards the surface of microplasmin but extends out to make use of an ionic contact with the side chain Asp28 of rTxlIn-1 from complex A (3.03 Å and 3.99 Å; Figure 6-17). The conformation and location of the side-chain of Arg19 of rTxlIn-1 is similar in both complexes. Therefore, it appears that the interaction between rTxlIn-1 and microplasmin is unaffected by crystal contacts.
Figure 6-15  Crystal contacts between rTxln-1 in complex A and surrounding molecules. Molecules making contacts with rTxln-1 in complex A are shown. rTxln-1 in complex A and complex B are colored yellow. Microplasmin in complex A is colored blue. Symmetry related rTxln-1 and microplasmin molecules are colored green and purple, respectively. The interface area between rTxln-1 in complex A and surrounding molecules are marked.

Figure 6-16  Crystal contacts between rTxln-1 in complex B and surrounding molecules. Molecules making contacts with rTxln-1 in complex B are shown. rTxln-1 in complex A and B are colored yellow. Microplasmin in complex B is colored blue. Symmetry related rTxln-1 and microplasmin molecules are colored green and purple, respectively. (a) Molecules in contact with rTxln-1 in complex B are shown. (b) Same molecules as (a) in a different orientation. The interface surface areas are marked.
Figure 6-17  Crystal contact between Arg19 of rTxln-1 and Asp28 of rTxln-1 from complex B. rTxln-1 is yellow and microplasmin blue. Side-chains are represented as sticks, nitrogen atoms are blue and oxygen atoms are red. (a) Crystal contact close to the interface of complex B. (b) Crystal contact between the side-chain of Arg19 of rTxln-1 from complex B and Asp28 of rTxln-1 from complex B.

The rTxln-1 molecules from complex A and complex B form buried surfaces with different crystallographic symmetry related microplasmin molecules (258 Å² and 431 Å² in complex A and B, respectively) (Figure 6-15; Figure 6-16) which include a key contact between Glu41 of rTxln-1 and Arg 644 (3.43 Å and 4.41 Å in complex A and B, respectively; Figure 6-18). The side-chain of Glu41 of rTxln-1 does not make a direct contact to microplasmin and therefore does not have a contribution to the conformation of the canonical loop of rTxln-1 to microplasmin, thus the crystal contact of Glu41 of rTxln-1 does not affect the structure of the canonical loop bound to microplasmin.
Figure 6-18 Crystal contacts between Glu41 and Arg644. rTxln-1 is yellow and microplasmin blue. Side-chains are represented as sticks, nitrogen atoms are blue and oxygen atoms are red. Crystal contacts between the side-chain of Glu41 of rTxln-1 and Arg644 of a symmetry related microplasmin (a) and (b) showing complex A, (c) and (d) showing complex B.

The largest of the buried surface areas with crystal contacts is 684 Å² (Figure 6-15; Figure 6-16) and made between the C-terminal α-helix of rTxln-1 from complex A and B. Two negatively charged residues (Glu51 and Glu52) and two polar residues (Thr56 and Ser55) form the interactions in the interface (Figure 6-19).
Figure 6-19 Largest crystal contact between two rTxln-1 molecules. rTxln-1 from complex A is yellow, rTxln-1 from complex B is green and microplasmin blue. Side-chains are represented as sticks, nitrogen atoms are blue and oxygen atoms are red. (a) The buried surface area between the C-terminal $\alpha$-helices of rTxln-1 from complex A and B is 684 Å$^2$. (b) Residues Glu51, Glu52, Ser55 and Thr56 form contacts in this interface.

If the different docking angles of rTxln-1 in the two complexes are not due to two different steps in the reaction between rTxln-1 and microplasmin, then one explanation might be that rTxln-1 and microplasmin have crystal contacts locking their bodies relative each other in different positions in complex A and B. After superposition of the microplasmin molecules, rTxln-1 from complex A clashes with the crystal contacts if it was docked as rTxln-1 from complex B to microplasmin.
(Figure 6-20). However, rTxln-1 from complex B does not clash with any crystal contacts if it was docked as rTxln-1 from complex A to microplasmin. The crystal contacts around the α-helix may keep rTxln-1 from complex B in its position (Figure 6-21). Therefore, the different crystal contacts in complex A and B may account for the two observed docking angles of rTxln-1 to microplasmin.

**Figure 6-20** Difference in crystal contacts between rTxln-1 from complex A and B  
Complex A, rTxln-1 is yellow and microplasmin is blue. Connolly surface of complex A is transparent. rTxln-1 from complex B is magenta. Atoms within 5 Å are shown as spheres. Green is symmetry related rTxln-1 atoms, salmon is symmetry related microplasmin atoms. Yellow and blue spheres are atoms from the complex within the same asymmetric unit. (a) Complex A, (b) rTxln-1 from complex A clash with atoms at distance ≤ 5 Å if it would be docked as rTxln-1 from complex B.
Figure 6-21 Difference in crystal contacts between rTxl-1 from complex A and B
Complex B, rTxl-1 is yellow and microplasmin is blue. Connolly surface of complex B is transparent. rTxl-1 from complex A is magenta. Atoms within 5 Å are shown as spheres. Green is symmetry related rTxl-1 atoms, salmon is symmetry related microplasmin atoms. Yellow and blue spheres are atoms from the complex within the same asymmetric unit. (a) Complex B. (b) rTxl-1 from complex B does not clash with atoms at distance ≤ 5 Å if it would be docked as rTxl-1 from complex A.

6.7 Comparison between rTxl-1-microplasmin and existing microplasminogen and microplasmin structures

The structure of the entire plasminogen or plasmin molecule has not been determined (Chapter 1). However, the structures of a number of the domains and complexes have been determined separately. These are the structures of kringle 1 (Wu et al., 1994), kringle 2 (Marti et al., 1999), kringle 4 (Stec et al., 1997), kringle 5 (Chang et al., 1998), microplasminogen with the nonessential mutations Met585Gln, Val673Met and Met788Leu (pdb file 1QRZ) (Peisach et al., 1999), the inactive microplasminogen Ser741Ala mutant (pdb file 1DDJ) (X. Wang et al., 2000), the Lys698Met mutant (pdb file 1RJX) (Terzyan et al., 2004), the microplasminogen Arg561Ala and Ser741Ala mutant in complex with deletion mutant of streptokinase α domain (pdb file 1L4D and 1L4Z) (Wakeham et al., 2002), two microplasmin molecules in complex with staphylokinase (pdb file 1BUI) (Parry et al., 1998) and microplasmin Ser741Ala mutant in complex with streptokinase (pdb file 1BML) (X. Wang et al., 1998).
Superposition of the microplasminogen and microplasmin structures illustrates large regional differences, but overall small structural differences. Human plasminogen has in its catalytic domain (microplasmin) a loop (558-566) where, upon activation by the physiological activators tissue-type plasminogen activator and urokinase-type plasminogen activator, a peptide bond located between Arg561 and Val562 which is proteolytically cleaved resulting in conformational changes and formation of a functional active site. Urokinase-type plasminogen activator bound to sepharose was used to convert microplasminogen to active microplasmin. Due to the limited resolution, residue 560-562 could not be modeled in either complex A or B. However, the distance between Pro559Ca and Val562Ca is ~15 Å and there is no extra density around Val562. Therefore, confirming that the peptide bond between Arg561 and Val562 is cleaved. The cleavage produced two polypeptide chains, one short at the N-terminus (13 and 15 residues in complex A and B, respectively) and one longer (230 residues in both complexes), connected by two disulfide bonds (Cys548-Cys666 and Cys558-Cys566). The Val562Ca has moved, in the microplasmin complex structures (rTxln-1-microplasmin, 1BUI and 1BML), an average distance of ~12 Å, from a solvent exposed position in the structures of microplasminogen (1DDJ, 1QRZ, 1L4D, 1L4Z and 1RJX) to a position in the core of microplasmin. The microplasmin α-amino group of Val562 forms an essential ionic bond with the side-chains of Asp740 in microplasmin, 2.91 Å and 2.63 Å in the rTxln-1-microplasmin complex A and B, respectively (Figure 6-22). The same ionic bond formation is found in the complex between microplasmin and staphylokinase and in the streptokinase-microplasmin complex.
Figure 6-22  Superposition of microplasmin and microplasminogen. Microplasmin from complex A of rTxln-1-microplasmin is blue and microplaminogen (1QRZ) is brown. Val562 moves ~12 Å from a solvent exposed position in microplaminogen to a position in the core of microplasmin.

The Gln738Ca and Gly739Ca atoms have moved ~6 Å in microplasmin (rTxln-1-microplasmin, 1BUI and 1BML) compared to microplaminogen (1DDJ and 1QRZ). This movement allowed the creation of the oxyanion hole. In the rTxln-1-microplasmin structure Gly739N has moved 6.87 Å compared to the microplaminogen structures (1DDJ and 1QRZ), and Gly739N and Ser741N make hydrogen bonds to Arg17O in rTxln-1. The Gln738 side-chain of microplasmin is directed away from the oxyanion hole in the rTxln-1-microplasmin structure, which is also the case in the complex between microplasmin and staphylokinase. The Gln738 side-chain of microplasmin in the streptokinase-microplasmin structure is occupying about the same position as the Arg17O in rTxln-1 (Figure 6-23).
Figure 6-23 Formation of the oxyanion hole in microplasmin. Microplasmin and rTxln-1 of complex A are in blue and yellow, respectively. Nitrogen atoms are blue and oxygen atom is red. Microplasmin (1BML) is green. Microplasminogen (1QRZ) is brown. Distances are in Å.

The loop region made up of residues 737-741 is disulfide linked by Cys737-Cys765 to the specificity (S1) entrance residues 760-765. The Cα atoms of the Cys737-Cys765 disulfide bridge have moved ~3 Å and ~5 Å, respectively, in the rTxln-1-microplasmin structure compared to the microplasminogen structures (1DDJ and 1QRZ). This has also been observed in microplasmin complex structures (1BUI and 1BML). This movement expands the entrance size of the S1 specificity pocket relative to microplasminogen. One of the most important structural changes is the location of the side-chain of Trp761. In microplasminogen, the side-chain of Trp761 covers the S1 entrance preventing binding of the side-chain of a P1 residue (1DDJ and 1QRZ). In the rTxln-1-microplasmin complex, the side-chain of Trp761 is folded away from the entrance of the specificity pocket (Figure 6-24), as in the other microplasmin complex structures (1BUI and 1BML). This allows the side-chain of Arg17 in rTxln-1 to protrude into the specificity pocket. The Asp735Cα of microplasmin has moved an averaged distance of 7.5 Å, from its solvent exposed position in microplasminogen (1DDJ and 1QRZ) to the bottom of the specificity pocket of microplasmin in all the microplasmin complex structures (rTxln-1-microplasmin, 1BUI and 1BML) and determines the specificity for a positively charged P1 residue. The position of the residues 737-741, the disulfide bridge Cys737-Cys765, and side-chain of Trp731 are the same in the microplasminogen structures 1RJX, 1L4D and 1L4Z, but somewhat different to that in either the
microplasminogen structures 1DDJ and 1QRZ and the microplasmin structures of the rTxln-1-microplasmin complex, 1BUI and 1BML.

Contrary to the physiological activators tPA and uPA that protelytically activate plasminogen, the two plasminogen activators from bacteria, staphylokinase and streptokinase, form tight-binding stoichiometric complex with human plasminogen or plasmin. These complexes act as plasminogen activators, a property neither plasminogen nor plasmin have by themselves. The structure of the microplasmin complex with staphylokinase was reported to show that staphylokinase does not affect the active-site geometry, but instead provide additional docking sites for enhanced presentation of the Arg561-Val562 bond in other plasminogen molecules (Parry et al., 1998). Superimposition of the microplasmin from the complex of rTxln-1 and staphylokinase reveals that a molecule such as rTxln-1 can fit into the active site cleft of microplasmin when staphylokinase has bound to microplasmin (Figure 6-24).

**Figure 6-24** Formation of the specificity entrance of microplasmin. Microplasmin of rTxln-1-microplasmin complex A is blue. Microplasminogen (1QRZ) is brown. The side-chain of Trp761 is represented as a stick model. Nitrogen atoms are blue and oxygen atoms are red. Disulfide bonds are orange.
Therefore it may suggest that staphylok inase could contribute to an amplified specificity for a number of molecules similar in size and shape to rTxln-1 and possibly occlude larger ones by limiting the size of the active site cleft of microplasmin.

Streptokinase consists of an α, β and γ domain and is a much larger molecule then staphylok inase. The N-terminal α domain of streptokinase was suggested to provide substrate recognition when bound to microplasmin (X. Wang et al., 1998). Moreover, it was suggested that the C-terminal γ domain of streptokinase induced an active site formation in microplasminogen (X. Wang et al., 1998), but was later proposed to be the function of the N-terminus of streptokinase (Terzyan et al., 2004). It was suggested that binding of streptokinase to microplasminogen made it possible to form a functional active site because of the formation of the essential ionic bond with Asp740 of microplasminogen. In activation of microplasminogen by streptokinase, this critical ionic bond was proposed to be formed between the Lys698 side-chain and Asp740 side-chain of microplasminogen (X. Wang et al., 1998), but it was later suggested that the N-terminus of streptokinase assumes the role in critical ionic bond formation with Asp740 of microplasminogen (Terzyan et al., 2004), which induces a catalytic active site without cleaving the activation bond between Arg561 and Val562 of microplasminogen. Superimposition of the microplasmin from the complex of rTxln-1 and streptokinase shows that a molecule with the same

Figure 6-25 Complex of staphylokinase-microplasmin and rTxln-1-microplasmin. (a) Structure of microplasmin and staphylokinase (1BUI), magenta and green, respectively. (b) RTxln-1 fits into the active site of microplasmin when staphylokinase is docked to microplasmin. Rtxln-1 and microplasmin complex A are blue and yellow, respectively.
size and shape as rTxln-1 could attach to the active site cleft of microplasmin when staphylokinase has bound to microplasmin (Figure 6-26). The $\alpha$ and $\beta$ domains of streptokinase were closer to rTxln-1 and the active site of microplasmin, than the $\gamma$ domain of streptokinase. By superimposing all the three structures of microplasmin onto each other shows that the staphylokinase and the $\alpha$ domain of streptokinase overlay well. It may suggests that both the $\alpha$ domain of streptokinase and staphylokinase could increase specificity of molecules similar to rTxln-1, however, the $\beta$ and $\gamma$ domains in streptokinase may even contribute to a more stringent specificity for a range of molecules similar in size and shape to rTxln-1, than staphylokinase, which consist of only one domain.

![Figure 6-26](image)

**Figure 6-26** Complex of streptokinase-microplasmin and rTxln-1-microplasmin. (a) Structure of streptokinase-microplasmin (1BML), dark blue and green, respectively. (b) rTxln-1 fits into the active site of microplasmin when streptokinase is bound to microplasmin. Rtxln-1 and microplasmin from complex A are light blue and yellow, respectively.

The backbone atoms of the catalytic triad, His603, Asp646 and Ser741, superposed at about the same place in the microplasminogen and microplasmin structures, however, the conformations of the side-chains varies. Overall, the $\chi_1$ angle of the Ser741 residues varies by $80^\circ$ in the structures of microplasminogen and microplasmin. The side-chain of His603 points out of the active site in the microplasminogen structures in two directions that differ by $\sim70^\circ$ (1QRZ, 1RJX, three of four molecules in the asymmetric unit of the 1DDJ structure in one group and 1LD4, 1L4Z, one of the 1DDJ molecules in the other group) (Figure 6-27). However, the side-chain of His603 points into the formed active site in the structures of microplasmin in complex with staphylokinase or streptokinase. In this position the imidazolium ring of His603 can make interactions with Ser741 and Asp646 and
hydrolyse substrates. The side-chain of the catalytic His603 in the rTxln-1-microplasmin complex has an orientation not observed in the other microplasmin structures. As mentioned earlier, the His603 side-chain of microplasmin from the rTxln-1-microplasmin complex is replaced by a water molecule in the active site, so it can not participate in a catalytic activity with Ser741 and Asp646, but has found stabilizing bonds with the Cys16-Cys40 disulfide bridge of rTxln-1, the solvent and possibly also an ionic bond with the side-chain of Glu606. The side-chain of His603, in the microplasminogen structures 1DDJ and 1QRZ also has a water molecule in the position of a catalytically functional His603 imidazolium ring of microplasmin, and is oriented away from Asp646 while forming an ionic bond with the side-chain of Glu606 (Peisach et al., 1999; X. Wang et al., 2000) (Figure 6-28; Figure 6-29). All microplasminogen structures have the His603 side-chain and the Glu606 side-chain facing each other at an ionic bond distance, except the 1L4D structure. Therefore, it is possible to assume that the side-chain of His603 points out of the active site in microplasminogen or when microplasmin is not active, and a water molecule can form the necessary bonds in the active site, while the side-chain of His603 is forming an ionic bond with the side-chain of Glu606.

**Figure 6-27** Stereo view of the different positions of the side-chain of the catalytic histidine. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. The side-chain of His603 of microplasmin structure 1BML is green, His603 of microplasminogen structure 1L4Z is purple and His603 of microplasminogen structure 1DDJ is pink.
Figure 6-28  Stereo view of a water molecule in the position of catalytically functional His603. rTxln-l and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. The side-chain of His603 of microplasminogen structure 1DDJ is pink. The water molecule belonging to the rTxln-l-microplasmin structure is red and the water molecule belonging to the microplasminogen structure 1DDJ is pink.

Figure 6-29  Stereo view of catalytic histidine forming an ionic bond with Glu606. rTxln-l and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. His603 and Glu606 of microplasminogen structure 1L4Z are pink, and His603 and Glu606 of microplasminogen structure 1DDJ are purple.

All of the Asp646 side-chains in the microplasminogen structures point out of the active site, except one, the side-chain of Asp646 from one of four molecules in the asymmetric unit of the 1DDJ structure which points inwards just as the side-chain of Asp646 of microplasmin (Figure 6-30). Visual inspection shows that the microplasminogen structures 1L4Z and 1QRZ have a water molecule in the position of Asp646 in microplasmin (Figure 6-31). Therefore it is possible to assume that a water molecule can replace the side-chain of Asp646 in a non-functional active site of microplasminogen.
Figure 6-30 Stereo view of different positions of Asp646. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. His603 and Asp646 of microplasminogen structure 1L4Z are purple, and His603 and Asp646 of microplasminogen structure 1QRZ are brown.

Figure 6-31 Stereo view of a water molecule in the position of the catalytically functional Asp646. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. His603 and Asp646 of microplasminogen structure 1L4Z are purple, and His603 and Asp646 of microplasminogen structure 1QRZ are brown. The water molecule belonging to the microplasminogen structure 1L4Z is purple and the water molecule belonging to the microplasminogen structure 1QRZ is colored brown.

The complex of microplasmin with staphylokinase is a ternary complex where two microplasmin molecules and one staphylokinase have combined. One microplasmin has a covalent bound chloromethylketone inhibitor (deoxo-methylarginine) bound to the catalytic His603 and Ser741 of microplasmin, and the inhibitor’s arginine side-chain is projecting into the specificity pocket. Superimposing the microplasmin with the chloromethylketone inhibitor onto the rTxln-1-microplasmin structure shows that the inhibitors overlays well (Figure 6-32). The other microplasmin molecule in this microplasmin-staphylokinase-microplasmin complex is docking in a substrate-like manner onto the first microplasmin molecule. In this structure it is proposed that an ionic bond between Lys556 of the microplasmin (that is substrate-like) and Glu606
assists in proper substrate orientation (Figure 6-33). Therefore it is possible to suggest that the side-chain of Glu606 of microplasmin has more than one function. It stabilizes the His603 side-chain, as in the microplasminogen and rTxln-1-microplasmin structures, and it binds to a molecule docked into the active site of microplasmin, as in the rTxln-1-microplasmin complex and the microplasmin-staphylokinase-microplasmin complex.

**Figure 6-32** Stereo view of inhibitor in the active site of microplasmin. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. Microplasmin structure 1BUI is green and the chloromethylketon inhibitor is purple. Superimposing shows that the inhibitors overlay well.

**Figure 6-33** Stereo view of interaction between Glu606 and bound molecule. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. Microplasmin structure 1BUI is green and purple. Glu606 of microplasmin forms an interaction with bound molecule.
6.8 Comparison between rTxln-1 in the microplasmin complex, in the trypsin complex and free rTxln-1

To study conformational differences between the rTxln-1 structures in the microplasmin complex, trypsin complex and the free structure, least-squares fitting and visual analyses of superposition were undertaken. The B-value patterns were examined as a measure of differences in conformational dynamics between the rTxln-1 structures.

Visual analysis of superposition of the structures shows that the overall main-chains overlay well, except for the free structure of molecule C, and all the disulfide bonds have the same chirality. Least-squares fitting of topologically equivalent Ca (and all atoms including side-chains) in residues 3-58 of all rTxln-1 structures resulted in rmsd values given in Table 6-4. The largest rmsd values appear when comparing the free rTxln-1 molecule C to all the other structures of rTxln-1. The high rmsd value is explained by the large structural difference of the canonical loop of molecule C (Figure 6-34).

| Table 6-4 Table of rmsd values between the rTxln-1 structures, Ca (all atoms) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
|                  | rTxln-1 free A   | rTxln-1 free B   | rTxln-1 free C   | rTxln-1 trypsin cplx | rTxln-1 µplm cplx A | rTxln-1 µplm cplx B |
| rTxln-1 free A   | 0                | 0                | 0                | 0                | 0                | 0                |
| rTxln-1 free B   | 0.66 (1.19)      | 1.09 (1.76)      | 0.88 (1.95)      | 0.43 (1.25)      | 0.43 (1.01)      | 0.43 (1.01)      |
| rTxln-1 free C   | 0.87 (1.66)      | 0.41 (1.25)      | 0.43 (1.95)      | 0.38 (1.86)      | 0.38 (1.86)      | 0.38 (1.86)      |
| rTxln-1 trypsin cplx | 0.64 (1.48) | 0.41 (1.25) | 0.87 (1.95) | 0.40 (1.25) | 0.40 (1.01) | 0.40 (1.01) |
| rTxln-1 µplm cplx A | 0.43 (1.22) | 0.43 (1.01) | 0.87 (1.86) | 0.40 (1.01) | 0.40 (1.01) | 0.40 (1.01) |
| rTxln-1 µplm cplx B | 0.54 (1.27) | 0.41 (1.11) | 0.93 (1.86) | 0.38 (1.06) | 0.38 (1.06) | 0.38 (1.06) |

185
Figure 6-34  Superimposition of all rTxln-1 structures.  
Microplasmin complexes are in yellow. rTxln-1 from the trypsin complex (pink) and all three structures of free rTxln-1 (molecule A green, molecule B blue and molecule C purple). Disulfide bonds are displayed as orange sticks. rTxln-1 from microplasmin-rTxln complex is yellow, and from trypsin-rTxln complex is in pink. Free rTxln-1 molecules are green, blue and purple. It is a 5.2 Å distance between Val16Cα in rTxln-1 in the rTxln-1-microplasmin complex and Val16Cα in molecule C of free rTxln-1.

B-values reflect to a first approximation positional variation of atoms over time and are therefore likely to provide insight into the dynamics of the protein (Ringe and Petsko, 1986; Parthasarathy and Murthy, 2000). However, B-values are also influenced by crystal order, diffraction resolution, atom occupancy and restraints used for the refinement. The crystal morphology and size varied between the crystals of the rTxln-1-microplasmin, rTxln-1-trypsin and free rTxln-1, as well as the overall order, as reflected by the resolution limits of 2.8 Å, 1.63 Å and 1.63 Å, for each of the crystals. Furthermore, because of the lower resolution of data the rTxln-1-microplasmin complex was also refined with NCS restraints. However, as illustrated by Figure 6-35, it can be seen that there is more motion in the same parts among the rTxln-1 structures.
**Figure 6-35** Visualization of the B-values in rTxln-1.
The highest main-chain B-values are in red thick tubes and lowest main-chain B-values are in blue thin tubes. Side-chains are shown as sticks and color coded according to B-values. (a) The three free rTxln-1 molecules. (b) rTxln-1 bound to trypsin. (c) The two rTxln-1 molecules bound to microplasmin.

**Figure 6-36** B-values (Å\(^2\)) for all rTxln-1 structures. Orange lines mark the canonical loop and the secondary loop of rTxln-1. Residues 14, 38, 39 and 42 are glycine and therefore have no side-chain B-values.
The most striking change in the B-values profiles for rTxlN-1 upon complexation with microplasmin or trypsin is the reduction in B-factor for the side-chain of Arg17. Thus, confirming the stabilization of this residue upon forming the complex with serine protease. There is a notable difference in the B-values profile for rTxlN-1 in the complex with microplasmin or trypsin between residues 18-22. The results suggest that the P' residues are bound more rigidly to the plasmin than to trypsin in these complexes. Also, the main-chain B-values for residues 38-40 of the secondary loop are in a dip in the rTxlN-1-microplasmin complex and in a peak in the rTxlN-1-trypsin complex. Thus the bond between Gly39O of rTxlN-1 and the side-chain of Glu606 of microplasmin makes the secondary loop more stable in the rTxlN-1-microplasmin complex than in the rTxlN-1-trypsin complex.

Analysis of the water structure in the rTxlN-1-microplasmin complex showed that, only one water molecule between the canonical and secondary loops is visible in complex B. It is highly likely that if higher resolution data were available the water molecules in the other highly conserved sites would also be visible.

6.9 Comparison between the rTxlN-1-microplasmin complex, the rTxlN-1-trypsin complex and an aprotinin-trypsin complex

Microplasmin has a similar overall structure as trypsin. However, the active site cleft is more open in microplasmin than in trypsin. The differences are confined to external loops at four places around the active site, one in the loop region between residues 37-39 in trypsin where two residues are omitted in microplasmin, and the second is due to the removal of six residues deleted in a loop region of microplasmin, the so-called 94-shunt in microplasminogen (Parry et al., 1998), between residues 94-101 of trypsin, and a third where the residues 145-152 of trypsin have a different conformation to that of microplasmin. The formation of the 94-shunt of microplasmin may permit entry of larger molecules such as cross-linked fiber clots (Peisach et al., 1999). Nevertheless, there is one region where microplasmin narrows the active site cleft more than in trypsin, which is a loop region with five residues insertion between residues 605-613 of microplasmin (Figure 6-37; Figure 6-38).
Figure 6-37 Amino acid sequence alignment of microplasmin and trypsin. Orange bars highlight places in the sequences which correspond to difference in external loops close to the active site of microplasmin and trypsin. These loops are marked out in Figure 6-38.

**Microplasmin**
AAPSFDGKPQVEPKCCPGVGGCVAHPSWQDPRFCMHPFCG

**Trypsin**
IVGGYTCNANTPVQVSLNS-Q-YHFCGG

TLISPENWLTAHCLEKSPRSYKVLGAHQEVEYNLEPHQIEVSRLFLEPTRF-----KD
SLINSQWVSAAHYC--S--GQVRLGDNVNVNENEQFISASKSVHSYSNTLND

IALLKLSPAVTDKVPACPLPSNYYAVDRTECFITGWGETQTF--FGA--GLKEAQLPVIEN
IMLKLKSAASLSRVSAISLPTS--ASAGTQCLISGWNTSSGTSYDPVLKCLKAPILSD

KVCNRYEFLNGRVQSTELCAGHLAGGTQOSGAGPVLCEFADKYLQCVTSGLGCCARPNI
SSCKS--AYPQQITSNMFCAHLEGKDSCQGSGPPVVCST---KLQIVSWSGCQAQKNK

PGVYTVRVSFVTWIEGVMRNN
PGVYTKVCNYVWIKQTIASN

Figure 6-38 Superposition of microplasmin and trypsin. The active site cleft is more open in microplasmin than in trypsin. Microplasmin is blue and trypsin purple. Differences in external loop close to the active site are marked out.

The interfaces between rTxln-1 and microplasmin and rTxln-1 and trypsin cover areas which are about the same in size, only 2% larger in the rTxln-1-microplasmin complexes (1384 Å² in the rTxln-1-microplasmin complexes and 1356 Å² in the rTxln-1-trypsin complex). Comparing the respective contact areas in the rTxln-1-
microplasmin complex (762:621 Å² in complex A and 758:626 Å² in complex B) and rTxln-1-trypsin complex (768:588 Å²) reveals that more surface of rTxln-1 is covered by microplasmin. This is mostly due to the five residues insertion between residues 605-613 of microplasmin. The side-chain of His603 together with the side-chains of Glu606 and Lys607 of microplasmin create a “wall” (Figure 6-39; Figure 6-40), which covers the canonical loop of rTxln-1 and makes contact with the secondary loop of rTxln-1, in the rTxln-1-microplasmin complex. The contact areas of microplasmin and trypsin are about the same. However, differences are that the S1’ subsite is more open in microplasmin compared to the S1’ subsite in the rTxln-1-trypsin complex, due to the side-chain of microplasmin’s His603 is pointing away from the active site (Figure 6-41). Another noticeable difference is that there are more contacts to the side-chain of Arg19 of rTxln-1 in the rTxln-1-trypsin complex.

**Figure 6-39** Differences between microplasmin and trypsin. rTxln-1 and microplasmin from complex A are light blue and yellow, respectively. Trypsin is purple. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. The side-chains of Glu606 and Lys607 of microplasmin are situated in a loop, which is not present in trypsin.
Figure 6-40 Interface between rTxln-1 and microplasmin or trypsin.
The Connolly surface of microplasmin and trypsin is shown with the electrostatic surface mapped. RTxln-1 is yellow sticks. Nitrogen atoms are blue and oxygen atoms are red. Residues of rTxln-1 are labeled in white and residues of microplasmin and trypsin are labeled in black. (a) The interface of the rTxln-1-microplasmin complex A. A “wall” is formed by the side-chains of His603, Glu606 and Lys607 of microplasmin, which cover the canonical loop of rTxln-1 and makes contacts to the secondary loop of rTxln-1. (b) The interface of the rTxln-1-trypsin complex. The side-chain of Arg19 of rTxln-1 is fitted inside a half open tunnel of trypsin.

Figure 6-41 Interface between rTxln-1 and microplasmin or trypsin.
The Connolly surface of microplasmin and trypsin is shown with the electrostatic surface mapped. RTxln-1 is yellow sticks. Nitrogen atoms are blue and oxygen atoms are red. Residues of rTxln-1 are labeled in white and residues of microplasmin and trypsin are labeled in black. (a) The interface of the rTxln-1-microplasmin complex A. The side-chain of His603 is pointing out of the active site of microplasmin. The P1’ residue Val18 of rTxln-1 has different rotamers in the rTxln-1-microplasmin complex A and B. (b) The interface of the rTxln-1-trypsin complex. The side-chain of His57 is positioned inside the active site. This creates a more open S1’ subsite in microplasmin than trypsin.
Visual analysis of the superposition of microplasmin from complex A and B and trypsin reveals that the docking angles between all three rTxln-1 molecules vary considerably. The largest angle is between the rTxln-1 bound to trypsin and rTxln-1 bound to microplasmin in complex B, 18.3°, while the angle is 11.4° between the rTxln-1 bound to trypsin and rTxln-1 bound to microplasmin in complex A (Figure 6-42). The canonical loops overlays well, but the distances between the secondary loops are 1.21 Å and 2.24 Å, measured between Cys40Ca of rTxln-1 from the trypsin complex to the microplasmin complex A and B, respectively (Figure 6-43). Therefore, it is possible to speculate that the extra contact surface of rTxln-1 to the “wall” of microplasmin has influenced the docking and created a change in docking angle of 11.4° and 18.3° and a corresponding change in position of 1.21 Å and 2.24 Å of Cys40Ca.

Figure 6-42 The docking angles between the bound rTxln-1 molecules when superimposing trypsin onto microplasmin. rTxln-1 bound to microplasmin is yellow and purple bound to trypsin. Microplasmin is light blue. (a) Docking angle between rTxln-1 from complex A and rTxln-1 from trypsin is 11.4°. (b) Docking angle between rTxln-1 from complex B and rTxln-1 from trypsin is 18.3°.
Figure 6-43 The difference in position of the secondary loop of rTxln-1 when superimposing trypsin onto microplasmin.

rTxln-1 bound to microplasmin is yellow and purple bound to trypsin. Cys16-Cys40 disulfide bond of rTxln-1 is orange. The Connolly surface of microplasmin is light blue. (a) Distance between Cys40α of rTxln-1 from complex A and rTxln-1 from trypsin is 1.21 Å. (b) Distance between Cys40α of rTxln-1 from complex B and rTxln-1 from trypsin is 2.24 Å.

The catalytic triads are located in the same spatial arrangement at the same positions and align well in all complexes and the side-chain of rTxln-1 Arg17 is pointing into the S1 specificity pocket in all complexes. However, the side-chain of the catalytic His603 has rotated out of the active site in the microplasmin complex, while it is positioned inside the active site in the trypsin complex (Figure 6-44). Therefore it seems possible to speculate that the reaction between rTxln-1 and microplasmin and rTxln-1 and trypsin are different.

Figure 6-44 The positions of the catalytic triads and rTxln-1 when microplasmin and trypsin are superimposed.

The rTxln-1 molecules bound to microplasmin are yellow and rTxln-1 bound to trypsin is purple. Microplasmin is blue. The catalytic triad of trypsin is green. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. Residues are labeled and labels are colored according to their residue color. The catalytic triads overlay well when microplasmin and trypsin are superimposed. The rTxln-1 side-chain Arg17 point in the same direction in the two structures.
In most crystal structures of hydrolases, the side chain of the catalytic histidine residue is rotated towards the catalytic active site, often with the His-Asp/Glu pair hydrogen bonded (Jing et al., 1998). However, in some instances the side chain of the histidine has rotated out of its normal location in the active site. Comparing two structures of *Rhodococcus sp.* cocaine esterase, one with an inhibitor and the other with a product bound in the active site, revealed that both the catalytic serine and histidine side-chains had rotated ~45° away from their catalytically functional positions when the product was bound (Larsen et al., 2001). In the structures of *Streptomyces griseus* protease A the histidine was rotated by ~90° from a catalytically functional position in a structure with a tetrapeptidealdehyde in the active site but not with a tetrapeptide (Brayer et al., 1979; James et al., 1980). An additional example of a mobile histidine side-chain was revealed when comparing two structures of *Torpedo californica* acetylcholinesterase complexed with two products consecutively formed after each other when reacting with an organophosphorous inhibitor. The structures of *Torpedo californica* acetylcholinesterase showed reversible movement of the histidine, where the conformational changes involved forming a bond to two different glutamatic acid residues (Millard et al., 1999). In the structure of tonin, the catalytic histidine is positioned out of the active site and has a role as a ligand for a Zn ion (Fujinaga and James, 1987). In structures of factor D (Narayana et al., 1994; Cole et al., 1997; Cole et al., 1998; Jing et al., 1998), the catalytic hisitidine rotates 180° out of the active site and into a stabilizing pocket, after an obstructing loop has moved. It can be concluded that the different positions of the catalytic histidine side-chain are dependent on the environments surrounding the histidine side-chain.

It is possible to theoretically model the microplasmin side-chain of His603 without having it clashing with any other side-chains to a catalytic functional position inside the active site, however, the short distance between His603Cδ₂ of microplasmin and one of the Val18Cγ atoms of rTxln-1 (~1.8 Å) make it unfavorable (Figure 6-47). Therefore, it can be speculated that the rTxln-1 side-chain of Val18 forces the side-chain of His603 to rotate out.

In bovine trypsin the side-chain of His57 can be theoretically modeled out of the active site, but not at the same location as the one in the rTxln-1-microplasmin
complex because it come too close to the side-chains of Tyr94 and Leu99 of bovine trypsin. The histidine has been seen to rotate \(\sim 90^\circ\) away from the catalytically functional position when the residue at the same location as Tyr94 in bovine trypsin is phenylalanine in a *Rattus norvegicus* trypsin mutant (Sprang et al., 1987) (Figure 6-45) or tryptophan in porcine pancreatic elastase (Sawyer et al., 1978; Meyer et al., 1985; Radhakrishnan et al., 1987; Meyer et al., 1988) (Figure 6-46). Modeling the side-chain of His57 of bovine trypsin in the rTxln-1-trypsin complex in any other position away from its catalytically functional position and without it clashing with rTxln-1 reveals that there is no residue such as Glu606 of microplasmin at the same position as in microplasmin to stabilized trypsin’s His47 side-chain (Figure 6-47). Therefore, it may be less likely that the side-chain of His57 would rotate out of the active site.

![Figure 6-45](image)

**Figure 6-45** Comparison of bovine trypsin and Asn102 *Rattus norvegicus* trypsin mutant. The Asn102 *Rattus norvegicus* trypsin mutant (1TRM) is yellow. His57, Phe94, Asn102 and Ser195 in the *Rattus norvegicus* trypsin mutant are shown as sticks with carbon atoms colored purple. Oxygen atoms are red and nitrogen blue. (a) His57 in the *Rattus norvegicus* trypsin mutant is partitioned between to position related with \(\sim 90^\circ\). (b) Bovine trypsin from the rTxln-1-trypsin complex is superimposed onto the *Rattus norvegicus* trypsin mutant. His57, Tyr94, Asp102 and Ser195 in bovine trypsin from the rTxln-1-trypsin complex are shown as sticks with carbon atoms colored cyan.
Figure 6-46  Comparison of bovine trypsin and porcine pancreatic elastase.
Porcine pancreatic elastase is green (1JIM). (a) His57, Phe94, Asp102 and Ser195 in porcine pancreatic elastase are shown as sticks. Carbon atoms are colored orange in elastase (3EST) with the His57 in its normal position. Carbon atoms are colored salmon (1INC) and yellow (1JIM) with the side-chain of His57 ~90° from its normal position. Oxygen atoms are red and nitrogen blue. (b) Bovine trypsin from the rTxln-1-trypsin complex is superimposed onto porcine pancreatic elastase. His57, Tyr94, Asp102 and Ser195 in bovine trypsin from the rTxln-1-trypsin complex are shown as sticks with carbon atoms colored cyan.

Figure 6-47  Stereo view of the environment around the catalytic histidine.
rTxln-1 and microplasmin from complex A are yellow and light blue, respectively. Trypsin is purple. Side-chains are represented as sticks. Nitrogen atoms are blue and oxygen atoms are red. Residues are labeled and labels are colored according to their residue color. The side-chain of microplasmin’s His603 would be very close to the side-chain of rTxln-1’s Val18 if it is positioned at the location of the side-chain of trypsin’s His57. The side-chain of trypsin’s His57 would be too close to the side-chains of trypsin’s Tyr94 and Leu99 if it is positioned at the location of the side-chain of microplasmin’s His603.
A structure of the aprotinin-microplasmin complex has not been determined to date. However, by superimposing trypsin from the aprotinin-trypsin complex (pdb file 2PTC) onto microplasmin reveals that the secondary loop of aprotinin docks even closer to the surface of microplasmin than the secondary loop of rTxln-1 (Figure 6-48). An additionally favorable contact can form between Lys46Nζ of aprotinin and Lys607N of microplasmin (2.73 Å). Furthermore, the side-chain of Arg20 of aprotinin have to adjust not to clash with the side-chain of Lys607 of microplasmin, having the potential to form a favorable “sandwiched” interactions among the side-chains of Arg20 of aprotinin, Lys607 of microplasmin and Ile18 of aprotinin. Another contribution to a possible closer contact of the secondary loop of aprotinin to the surface of microplasmin is the residue Arg39 of aprotinin, which has a longer side-chain than the corresponding side-chain of Glu41 of rTxln-1. This theoretical approach may explain why aprotinin binds harder to microplasmin than rTxln-1 and also why aprotinin is less reversible than rTxln-1. However, another important difference between rTxln-1 and aprotinin is the change from a valine in rTxln-1 to alanine in aprotinin at the P1' position. Alanine is smaller than valine and is likely to not expel the His603 side-chain of microplasmin, which could greatly affect the off rate.

![Figure 6-48](image-url) The location of aprotinin when superposing the trypsin molecule from the aprotinin-trypsin complex onto microplasmin. rTxln-1 from the microplasmin complex is yellow, rTxln-1 from the trypsin complex is purple and aprotinin from a complex with trypsin is green. Microplasmin is light blue. (a) Visualizing the different docking angles and aprotinin docks closer to the microplasmin surface than rTxln-1. (b) The Connolly surface of microplasmin is light blue. Aprotinin is approaching close to the surface of microplasmin, the canonical loops, secondary loops and disulfide bridge overlay well.
6.10 Discussion of rTxln-1 dimer

Mass spectrometry (Chapter 2) of both rTxln-1 alone and of the rTxln-1 trypsin complex indicated that rTxln-1 might form a dimer. A possible dimer formation of rTxln-1 was seen in the crystal packing (Chapter 4). Mass spectroscopy of the rTxln-1-trypsin complex showed the presence of a species of 37 kDa, suggesting that one molecule of trypsin could combine with two molecules of rTxln-1 (Chapter 2). The structural proofs of two rTxln-1 molecules docking to microplasmin or trypsin within this study are discussed briefly in the following.

In the crystal lattice of rTxln-1-microplasmin, two rTxln-1 molecules have an interface of $342 \text{ Å}^2$ between their C-terminal $\alpha$-helices. One of the rTxln-1 molecules is docked to the active site of microplasmin and the other rTxln-1 molecule has also an interface with the microplasmin molecule. This latter interface is $215 \text{ Å}^2$ and $129 \text{ Å}^2$, in complex A and B respectively (Figure 6-49). The interface of $342 \text{ Å}^2$ between the two rTxln-1 molecules are likely to be too small for a dimer formation, therefore, this is not likely to occur (Figure 6-49; Figure 6-50; Figure 6-51).

![Figure 6-49](image_url)

**Figure 6-49** Two rTxln-1 molecules and one microplasmin molecule. The rTxln-1-microplasmin complex is yellow and light blue. The symmetry related rTxln-1 is green. Interfaces are marked. (a) The rTxln-1-microplasmin complex A and symmetry related rTxln-1, interfaces are $324 \text{ Å}^2$ and $215 \text{ Å}^2$. (b) The rTxln-1-microplasmin complex B and symmetry related rTxln-1, interfaces are $324 \text{ Å}^2$ and $129 \text{ Å}^2$. 

---

198
Figure 6-50  Residues in interface between complex A and symmetry related rTxln-1. The rTxln-1 and microplasmin from complex A are yellow and light blue. The symmetry related rTxln-1 is green. Residues in the interface are marked.

Figure 6-51  Residues in interface between complex B and symmetry related rTxln-1. The rTxln-1 and microplasmin from complex B are yellow and light blue. The symmetry related rTxln-1 is green. Residues in the interface are marked.
Another possible dimer formation of rTxln-1 with an interface area of 643 Å² was mentioned in Chapter 4. Superposition of the structure of free rTxln-1 molecule A or molecule C onto the rTxln-1 molecule in the trypsin complex and rTxln-1 in the microplasmin complex A and B, reveals that this dimer formation of rTxln-1 may not function as an efficient inhibitor of plasmin or trypsin because the canonical loop is hindered to some extent to enter the active site, less so with microplasmin than with trypsin (Figure 6-52). Therefore, this is not an ideal dimer formation of rTxln-1 as efficient inhibitor of plasmin or trypsin.

![Figure 6-52](image)

**Figure 6-52** Possible dimer binding to microplasmin. The rTxln-1 and microplasmin from complex B are yellow and light blue. The dimer formation of free rTxln-1 molecule A and C is purple. (a) The rTxln-1-microplasmin complex A. The free rTxln-1 molecule A is superposed onto rTxln-1 from complex A and showing the position of the free rTxln-1 molecule C. (b) The rTxln-1-microplasmin complex B. The free rTxln-1 molecule A is superposed onto rTxln-1 from complex B and showing the position of the free rTxln-1 molecule C. The free rTxln-1 molecule C clashes with microplasmin.

### 6.11 Summary

The structure of the rTxln-1-microplasmin complex has been determined by molecular replacement to a resolution of 2.78 Å. rTxln-1 docks to microplasmin through its canonical loop and secondary loops in a similar manner as it docks to trypsin. A stable Michaelis complex is formed with a buried surface of 1384 Å². The Arg17 from rTxnl-1 protrudes into the specificity pocket of microplasmin making an ionic bond with aspartate at the bottom of this pocket, while the distance
between the carbonyl carbon of Arg17 and the catalytic Ser741Oγ in microplasmin are within a sub van der Waals distance (2.94 Å). His603 in microplasmin adopts an altered non-catalytically competent position such that its side-chain has swiveled around its χ1 bond and out of its classical catalytic triad location. To compensate, a water molecule is observed bridging, the serine and aspartate side-chains. The reason for this movement appears to be due to the close approach of Val18 in rTxln-1 forcing His603 to move. The side-chain orientation of His603 has not previously been observed in other plasmin structures. In addition, the His603 side-chain is situated close to the Cys16-Cys40 disulphide bridge in rTxln-1.

An interesting observation is that the two complexes in the asymmetric unit differ in docking angles by, ~8° suggesting that there is flexibility at the interface. This is either due to rTxln-1 not being completely locked into its optimal conformation in one of the complexes or as a result of crystal packing effects. This variability may be related to rTxln-1’s fast off rate for plasmin.
Chapter 7

Crystal structure of native ACII-4 from Australian King Brown snake (*Pseudechis australis*) venom

Snake venoms are a rich source of molecules that have a PLA₂ fold. As well as their phospholytic activities, the PLA₂s from venom possess a wide variety of pharmacological activities such as presynaptic and/or postsynaptic neurotoxicity, myotoxicity, initiation and/or inhibition of platelet aggregation, haemolytic, anticoagulant, convulsant, hypotensive, cardiotoxic and oedema-inducing effects (Kini and Evans, 1989; Arni and Ward, 1996).

PLA₂ isolated by our group from the Australian King Brown snake, ACII-4, has low phospholipase activity and strong anticoagulant activity. It consists of a single polypeptide chain of 118 amino acid residues and has a molecular weight of 14 kDa. The phospholipase activity has been shown to be dependent on the presence of Ca²⁺ (Masci, 2000). However, the phospholipase and anticoagulant activity are not dependent on each other and should therefore relate to different regions of the molecule. ACII-4 has also been reported to have both presynaptic neurotoxic and myotoxic activities on isolated nerve-muscle preparations (Rowan *et al.*, 1989). Nitrophenylsulfonylation of tryptophanyl residues at positions 31 and 69 in ACII-4 caused loss of all activities (Takasaki *et al.*, 1990a). The region containing the anticoagulant activity in PLA₂s from the genus *Naja* (group I) was correlated to the region around a trypsin residue which corresponds to Phe61 in ACII-4 (Kini and Evans, 1987; Kini, 2006). It has been suggested that the ACII-4 molecule is functional as a dimer (Masci, 2000). A search for similar sequences to ACII-4 in the Protein Data Bank using the ‘Basic Local Alignment Search Tool’ (BLAST) (Schäffer *et al.*, 2001) shows the highest identity, 65%, with a presynaptic neurotoxic PLA₂, Notexin, from Mainland Tiger snake (*Notechis scutatus scutatus*). All fourteen cysteine residues in these molecules are in conserved locations in these two proteins.

The overall architecture of molecules with a PLA₂ fold consists in general of three major and two minor α-helices, and a double-stranded antiparallel β-sheet (referred to as the β-wing), a Ca²⁺ binding loop and the presence of several disulfide bonds (Arni and Ward, 1996). For PLA₂ to be catalytically active a Ca²⁺ ion needs to be
bound to the Ca$^{2+}$ binding loop (Dijkstra et al., 1983; Steiner et al., 2001). One of the ligands coordinating the Ca$^{2+}$ ion is generally an aspartate at position 41, contributing its two side-chain oxygen atoms to the coordination.

Crystal structures have been solved of PLA$_2$s from the venom of Chinese cobra, Naja naja atra, Indian cobra, naja naja sagittifera (Jabeen et al., 2005), Guangxi cobra, Naja naja kaouthia (Gua et al., 2002), Eastern Cottonmouth snake, Agkistrodon piscivorus piscivorus, (Maraganore and Heinrickson, 1986; van den Bergh et al., 1988, 1989), Bothrops pirajai (Rigden et al., 2003) and a pit viper Bothrops asper (Arni et al., 1995). An individual PLA$_2$ can exhibit one or more pharmacological activities. For example, a PLA$_2$ from Russell's viper, Vipera russelli russelli, has anticoagulant properties while PLA$_2$ from Australian Tiger snake, Notechis scutatus scutatus, has presynaptic neurotoxic properties (Carredano et al., 1998), and the PLA$_2$ from King cobra, Ophiophagus hannah has cardiotoxic and myotoxic activity (Zhang et al., 2002).

There are three phospholipases in the venoms of the Australian King Brown snake, Pseudechis australis. These have been identified as PA11, PA3 and PA15 (Takasaki et al., 1989; Takasaki et al., 1990b). ACII-4 corresponds to PA11. The other two PLA$_2$s exhibit different activities. One has very strong phospholipase activity and the second has strong haemolytic activity. From a sequence alignment of PA11, PA3 and PA15 and other phospholipases it is not clear what region corresponds to the specific activity. However, structural analysis and comparison with other PLA$_2$ sequences and structures should help to better define the source of the individual activities, thereby leading to novel lead drug candidates.

This chapter describes the steps used to solve the structure of the ACII-4, and compares the structure of this molecule with other molecules with a PLA$_2$ fold.

### 7.1 Introduction to hemihedral crystal twinning

During the growth of a crystal, an intergrowth of two crystals can sometimes occur in a symmetrical configuration. This so called twinning can be classified as merohedral or nonmerohedral (Friedel, 1928). Nonmerohedrally twinned crystals share only some sets of the same crystal lattice points in three dimensions, while for
merohedrally twinned crystals diffraction patterns overlay exactly. The most common type of merohedral twinning is hemihedral twinning, where only two different orientations of the different crystals are present (Parsons, 2003). Diffraction patterns of hemihedrally twinned crystals look the same as diffraction patterns of crystals without twinning. It is not possible to distinguish a diffraction pattern of a hemihedrally twinned crystal from a crystal without twinning without further analysis. In a diffraction pattern from a hemihedrally twinned crystal, each observed diffraction intensity is a weighted sum of two twin-related reflections (Stanley, 1955, 1972; Dauter, 2003). The weights are determined by the fractional volume of each of the two crystal orientations within the twinned crystal, represented by the “twinning fraction” $\alpha$, corresponding to the fractional volume of the crystal that the smaller of the two domains occupies. The observed intensity $I_{\text{obs}}$ is equal to

$$I_{\text{obs}}(hkl) = (1-\alpha) I(hk'l') + \alpha I(h'k'l')$$

(7.1)

where $(hkl)$ and $(h'k'l')$ are twin-related Miller indices. The twinning fraction takes values between 0.0 and 0.5. The specimen is untwinned if $\alpha = 0$. When $\alpha$ is nonzero but less than 0.5 it is referred to as “partial twinning” and the special case of $\alpha = 0.5$ as mimetic or “perfect twinning”. In the case of partial twinning, the structure can still be solved by statistically estimating the twinning fraction followed by detwinning the data and refining the structure against the detwinned dataset.

As outlined and described below, the diffraction pattern of the ACII-4 crystal was produced from a hemihedrally twinned crystal. It was therefore possible to determine the twinning fraction and to detwin the processed intensities. Assuming that the ACII-4 molecule had a structure similar to other PLA$_2$ structures molecular replacement could be applied as the technique to solve the phase problem, since several structures of PLA$_2$s are deposited in the PDB.

### 7.2 Phase determination of free ACII-4

Extensive attempts at solving the phase problem were carried out. Initially, AMoRe (Navaza, 1994) (see Appendix E.2), was used to perform fast rotation, translation functions and rigid body refinements. Based on the analysis of the X-ray diffraction pattern the space group $P3_121$ was assigned to this crystal form. Solvent content
arguments indicated that there is only one molecule in the asymmetric unit. The rotation function was run searching for Patterson correlation within a sphere centered on the origin. This allowed the Patterson correlation function to be expressed in terms of spherical harmonics, and the calculation to exploit fast Fourier transform techniques (see Appendix D and E for AMoRe and Patterson calculations). Searching for one molecule and subsequently checking for two molecules in the translation step was carried out by employing the Crowther and Blow translation function. This function tests each orientation solution in turn and searches for the best resultant translational. There was no improvement when searching for two molecules. After carrying out rigid-body refinement, the solution with one molecule gave a correlation coefficient of 0.49 and an $R_{\text{factor}}$ of 0.48. Refinement by utilizing the CNS suite of programs (Brünger et al., 1998) and model building with the Program O (Jones and Kjeldgaard, 1998) could not refine the structure to better than an $R_{\text{factor}}$ of 0.38. The data was reprocessed in space group P3$_1$ and AMoRe was run in space groups P3$_1$, P3$_2$ and P3, but no correct solution was found. The molecular maximum-likelihood program Phaser (McCoy et al., 2005) (see Appendix E.4 for Phaser and maximum-likelihood method), available through the CCP4 software suite, was also tried to find the initial set of phases, using the same space groups. A model was found in P3$_1$, but this would not refine any better than to an $R_{\text{factor}}$ of 0.38 using CNS. ARP/wARP (Morris et al., 2003) was used to auto build a completely new model based on phases from the previous molecular-replacement solution. However, only dummy atoms could be placed and a new model could not be built. The conclusion was that the data had either been processed in the wrong space group or the data was twinned.

The CCP4 program SFCHECK was used to test for merohedral twinning and to detwin the data. SFCHECK first tested the reflection file which had been processed in space group P3$_1$ for merohedral twinning. Initially the perfect twinning test was performed. Each observed intensity is a sum of two true crystallographic intensities. The crystallographic intensities obey ordinary Wilson statistics (Wilson, 1949), while the sum of two such values does not. Expected properties of intensity distributions for perfect hemihedral twinning have been given by (Stanley, 1972). For acentric data the formulae for the perfect twinning test are
\[
\frac{\langle I_{\text{obs}}^2 \rangle}{\langle I_{\text{obs}} \rangle^2} = 2 \quad \text{for untwinned data} \quad (7.2a)
\]
\[
\frac{\langle I_{\text{obs}}^2 \rangle}{\langle I_{\text{obs}} \rangle^2} = 1.5 \quad \text{for perfectly twinned data} \quad (7.2b)
\]

The calculation showed a value of 1.67 for the dataset collected from the ACII-4 crystal indicating that the crystal is neither untwinned nor perfectly twinned. Next a partial twin test parameter H (Yeates, 1997) was computed for the pairs of acentric reflections

\[
H = \frac{| I_{\text{obs}}(h) - I_{\text{obs}}(h') |}{(I_{\text{obs}}(h) + I_{\text{obs}}(h'))} \quad (7.3)
\]

where \((h)\) and \((h')\) represent \((hkl)\) and \((h'k'l')\), respectively. To calculate H requires that the twin operator is already known. The twin operator is the symmetry operation that relates the two orientations in the twinned crystal. SFCHECK tested all possible twin operators and reported \((-h-k,+k,-l)\) as the correct one. If Wilson statistics applies to the two independent intensities it is possible to derive the expected distribution for H as a function of \(\alpha\) (the twinning fraction). The expected cumulative distribution for H, S(H), is linear in H for acentric reflections and given by

\[
S(H) = H / (1 - \alpha) \quad (7.4)
\]

The cumulative distribution of H for the observed intensity data was calculated in Figure 7-1 and the plot indicates a twinning fraction of about 0.43.

![Figure 7-1 Cumulative distribution of H. Cumulative distribution of H as a function of H and twinning fraction \(\alpha\).](image-url)
The curves plotted with + are the linear curves based on Wilson statistics. The curve plotted with o is based on the observed intensity data and indicating a twinning fraction $\alpha \approx 0.43$. The partial twinning test was followed by calculating the twinning fraction $\alpha$ according to (Yeates, 1997)

$$\alpha = \frac{1}{2} - <H>$$

(7.5)

which gave a twinning fraction $\alpha = 0.438$. This is in agreement with the graph produced by SFCHECK.

The processed reflection file, from 26.79–1.56 Å, was detwinned by utilizing the determined twinning fraction and the following equations

$$I_{\text{obs}}(h) = (1 - \alpha) I_o(h) + \alpha I_o(h')$$

(7.6a)

$$I_{\text{obs}}(h') = \alpha I_o(h) + (1 - \alpha) I_o(h')$$

(7.6b)

thus

$$I_o(h) = \frac{(1 - \alpha) I_{\text{obs}}(h) - \alpha I_{\text{obs}}(h')}{1 - 2\alpha}$$

(7.7a)

$$I_o(h') = \frac{(1 - \alpha) I_{\text{obs}}(h') - \alpha I_{\text{obs}}(h)}{1 - 2\alpha}$$

(7.7b)

where $I_{\text{obs}}$ is the observed intensity, $I_o$ is the true, detwinned intensity, $h$ and $h'$ denote twin-related indices. For perfect twinning, $\alpha = 0.5$, the Equations 7.6 and 7.7 give no solutions as the denominators are zero.

The resultant intensities were converted into a mtz file and used in MrBump that directs CCP4 supported programs to automatically discover and prepare a search model in addition to carrying out cycles of refinement (Keegan and Winn, 2007) (see Appendix E.8). MrBump was given the amino acid sequence of the ACII-4 molecule, and the detwinned experimental structure factors. The run was entered to search for homologous structures, and created a set of suitable search models from the template structures by using the default multiple sequence alignment program, Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh et al., 2005), with the E-value for the Fasta search set to 0.02. The sequence was compared to the sequences in the SwissProt database using MAFFT to find similarities. The obtained data for these sequences are reported as percent identical residues, Z-score and E-value. If two proteins have over 45% identical residues in their optimal alignment, they...
probably have very similar structures and a similar function. The Z-score is a measure of how a score compares to the mean and standard deviation of the alignments of randomized sequences. If the alignment has a score S, the Z-score is given by

\[
Z\text{-score} = \frac{(S-\text{mean})}{\text{standard deviation}}
\]  

(7.9)

A Z-score of 0 means that the similarity is no better then the average of the random permutations of the sequence and might have happened by chance. A higher Z-score increases the probability that the observed alignment has not happened by chance. A Z-score ≥5 is considered significant. The E-value of an alignment is the expected number of sequences that give the same score or better. If the E-value ≤0.02 the sequences are probably homologous (Lesk, 2002).

The Fasta search found several good alignments with the three best scores:

<table>
<thead>
<tr>
<th>PDB</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AE7</td>
<td>5.3e-37</td>
</tr>
<tr>
<td>2NOT Mol A</td>
<td>3.2e-36</td>
</tr>
<tr>
<td>2NOT Mol B</td>
<td>3.2e-36</td>
</tr>
</tbody>
</table>

The models were prepared with Molrep (Vagin and Teplyakov, 1997) and Chainsaw (Stein, 2006). Molrep contains a model preparation function which aligns the template sequence with the target sequence and prunes the non-conserved side-chains accordingly. Chainsaw removes un-aligned residues from the model and prunes non-conserved residues back to the C\(\gamma\) atom, which preserves more atoms than a polyalanine model. After the models had been prepared with Molrep and Chainsaw, the log file showed the three best scores as the following:

<table>
<thead>
<tr>
<th>Prepared search models</th>
<th>Molrep Score</th>
<th>Chainsaw Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AE7</td>
<td>0.346</td>
<td>0.66</td>
</tr>
<tr>
<td>2NOT Mol A</td>
<td>0.345</td>
<td>0.65</td>
</tr>
<tr>
<td>2NOT Mol B</td>
<td>0.353</td>
<td>0.65</td>
</tr>
</tbody>
</table>
MrBump was set to use PHASER for the molecular replacement and thereafter 30 cycles of restrained refinements in REFMAC5 (Murshudov et al., 1997). The prepared search model 1AE7 was used in PHASER. The log file of MrBump showed two molecules in the asymmetric unit (Figure 7-2); the results of the final translation function in PHASER had a Z-value of 18, and the refinement results from REFMAC5 had an $R_{\text{factor}}$ of 0.38. Examination of the crystal packing indicated reasonable contacts between the two molecules and no overlap of symmetry related molecules (Figure 7-2).

**Figure 7-2** The two ACII-4 molecules in the asymmetric unit. The two ACII-4 molecules in the asymmetric unit are colored blue and purple. The monoclinic unit cell is shown in purple with the a and b axis marked in black.

### 7.3 Refinement of model of PLA$_2$ crystal structure

The model and map from REFMAC5 was visualized with Coot (Emsley and Cowtan, 2004). The side-chains that differed were mutated into the amino acid sequence of ACII-4 and fitted into the electron density. The model was fitted into the 2Fo-Fc map, using a sigma level of 1 for both the 2Fo-Fc and Fo-Fc maps. The model building was refined with cycles of restrained refinement with REFMAC5 with no prior phase information, and with the weighting term between X-ray and geometric residuals set to 0.4. Data used in refinement ranged from 22.46-1.56Å, with 5% reflections used for $R_{\text{free}}$ test set (1142 reflections). All atoms were refined with
individual isotropic B-factors. The first rounds were cycled with the ARP_water feature in REFMAC5 to analyze the solvent model.

7.4 Refinement statistics and quality of ACII-4 crystal structure

There are two ACII-4 molecules in the asymmetric unit. The electron density map is, in general, well defined and the majority of the amino acid residues for the two ACII-4 molecules in the asymmetric unit could be fitted into the electron density. The final model included residues 1–118 in both molecules A and B in the asymmetric unit. 309 water molecules, two sulfate ions, four PEG molecules and two Ca$^{2+}$ ions could also be fitted into the electron density. It was not necessary to truncate any of the side-chains to alanine. The final $R_{\text{factor}}$ and $R_{\text{free}}$ are 0.20 and 0.24, respectively. A difference between the two values of only 0.04 is an indication that the model has not been over refined. The final model has excellent geometry with an rmsd from ideal bond lengths of 0.014 Å, and an rmsd from ideal bond angles of 1.60° (Table 7-3).

<table>
<thead>
<tr>
<th>Table 7-3</th>
<th>Refinement statistics of ACII-4 crystal structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-hydrogen atoms</strong></td>
<td></td>
</tr>
<tr>
<td>Molecule A</td>
<td>899</td>
</tr>
<tr>
<td>Molecule B</td>
<td>899</td>
</tr>
<tr>
<td>Total number of protein atoms</td>
<td>1798</td>
</tr>
<tr>
<td>Solvent (H$_2$O)</td>
<td>309</td>
</tr>
<tr>
<td>Solvent (Ca$^{2+}$)</td>
<td>2</td>
</tr>
<tr>
<td>Solvent (SO$_4$)</td>
<td>10</td>
</tr>
<tr>
<td>Solvent (PEG)</td>
<td>28</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>22.46 – 1.56</td>
</tr>
<tr>
<td>$R_{\text{factor}}$</td>
<td>0.20</td>
</tr>
<tr>
<td>$R_{\text{free}}$</td>
<td>0.24</td>
</tr>
<tr>
<td>rmsd from ideal geometry</td>
<td></td>
</tr>
<tr>
<td>rmsd bond lengths (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>rmsd bond angles (°)</td>
<td>1.520</td>
</tr>
</tbody>
</table>

$R_{\text{factor}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}|/\sum |F_{\text{obs}}|$, 5% of the data were excluded from the refinement were used to calculate $R_{\text{free}}$.

The average B-values for all main-chain atoms are 13.1 Å$^2$ and 14.0 Å$^2$ for molecule A and B, respectively, while for side-chains atoms these values are 14.7 Å$^2$ and 15.9 Å$^2$ (Table 7-4). This shows that the two molecules display a similar level of structural order. An analysis of the B-values as a function of position in the
polypeptide of the two ACII-4 molecules shows that there are four areas with higher values than the average. They consist of residues 10–24 (part of solvent exposed loop), 29–36 (part of Ca\(^{2+}\) binding loop), 70–84 (part of β-wing), and the C-terminal residues (Figure 7-4). Lys81 and Pro82 were the only two residues with weaker electron density and difficult to model. These two residues are both situated in a solvent exposed loop region and Lys81 has, except for the C-terminal Lys118, the highest B-values in the structure.

**Table 7-4** Mean B-values (Å\(^2\)) for the two ACII-4 molecules.

<table>
<thead>
<tr>
<th></th>
<th>Molecule A</th>
<th></th>
<th>Molecule B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>main-chain</td>
<td>13.1</td>
<td>main-chain</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>side-chain</td>
<td>14.7</td>
<td>side-chain</td>
<td>15.6</td>
</tr>
<tr>
<td>all atoms (899 atoms)</td>
<td>13.9</td>
<td></td>
<td>all atoms (899 atoms)</td>
<td>14.7</td>
</tr>
<tr>
<td>H(_2)O (309 atoms)</td>
<td>22.1</td>
<td></td>
<td>H(_2)O (309 atoms)</td>
<td>22.1</td>
</tr>
<tr>
<td>Ca(^{2+}) (2 atoms)</td>
<td>13.8</td>
<td></td>
<td>Ca(^{2+}) (2 atoms)</td>
<td>13.8</td>
</tr>
<tr>
<td>SO(_4) (10 atoms)</td>
<td>28.7</td>
<td></td>
<td>SO(_4) (10 atoms)</td>
<td>28.7</td>
</tr>
<tr>
<td>PEG (28 atoms)</td>
<td>25.3</td>
<td></td>
<td>PEG (28 atoms)</td>
<td>25.3</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>13.6</td>
<td></td>
<td>17.3</td>
</tr>
<tr>
<td>main-chains</td>
<td>13.6</td>
<td></td>
<td>main-chains</td>
<td>17.3</td>
</tr>
<tr>
<td>side-chains and waters</td>
<td>17.3</td>
<td></td>
<td>side-chains and waters</td>
<td>17.3</td>
</tr>
<tr>
<td>all atoms (2147 atoms)</td>
<td>15.6</td>
<td></td>
<td>all atoms (2147 atoms)</td>
<td>15.6</td>
</tr>
</tbody>
</table>

**Figure 7-3** Visualization of the B-values for the two molecules of ACII-4 in the asymmetric unit. Main-chain high B-values are in yellow, thick tubes and lower B-values are in blue, thin tubes. Side-chains are shown as sticks and color coded. Molecule A and B are labeled.
Figure 7-4 B-values (Å²) for molecule A and B of ACII-4. Glycine residues have no side-chain B-value.

The Ramachandran plot calculated over the two molecules showed 89.9% of residues in most favored regions, 9.1% residues in additional allowed regions, and 1.0% in generously allowed regions. There were no residues in disallowed regions (Table 7-5). The main-chains that have unusual dihedral angles are Lys81 and Pro82 (Figure 7-5). This was the only place in the structure where the electron density was weak. The cyclic side-chain of proline residues limit the range of Phi angles to values of around -60°, making it the most conformationally restricted residue.

<table>
<thead>
<tr>
<th>Residues (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favored regions</td>
</tr>
<tr>
<td>Additionally allowed regions</td>
</tr>
<tr>
<td>Generously allowed regions</td>
</tr>
<tr>
<td>Disallowed regions</td>
</tr>
</tbody>
</table>
Figure 7-5  Ramachandran plot of the crystal structure of ACII-4. A (α-helix), B (β-sheet) and L (left handed α-helix) are the most favored regions. a, b and c are additionally allowed regions, and ~a, ~b, ~l, ~p are generously allowed regions. Glycine residues are identified as triangles. Amino acids with high backbone energy conformations are Lys81 and Pro82 in both molecule A and B.

Pro82 is situated in a solvent exposed loop region and has created a bend in the loop. The Cys78-Cys89 disulfide bond holds this loop close to the core of ACII-4. This has set restraints on the side-chain of Lys81 which was modeled in a cis-conformation. The side-chain of this residue is aligned with the core of the protein. A water mediated bond is also formed between Lys81 and Lys85.

Lys81 is part of the tryptic fragment for which a mass has been determined (ms). However, no ms/ms data have been obtained to unambiguously assign the sequence (personal communication, Geoff Birrell). Otherwise, the undetermined regions in the sequence (Gly59-Lys63, Pro82-Lys85, Cys98-Lys106 and Lys115-Lys118) have well defined electron density except for small parts of the side-chains of Lys85, Arg116 and Lys118 in molecule A and Lys118 in molecule B. The tryptic digest and
mass spectrometry results together with the crystal structure confirm that ACII-4 is the same molecule as the previously isolated PA11 (Takasaki et al., 1990b).

### 7.5 Crystal structure of free ACII-4

ACII-4 has an L shape with a height of a 34 Å (measured from residue Gly59Cα to Lys105Cα) and a width of 40 Å (measured from Gly73Cα to Lys114Cα). The solvent accessible surface of each ACII-4 is 6719 Å² (average of the two molecules in the asymmetric unit) AREAIMOL (B. Lee and Richards, 1971). Both ACII-4 molecules consist of; three major α-helices (residues 2-12, 40-57 and 83-101), one minor α-helix (113-116), two minor 3_10-helices (residues 19-22 and 106-108) and one double-stranded antiparallel β-sheet (residues 69-72 and 75-78). Seven disulfide bonds stabilize the structure (Cys11-Cys71, Cys27-Cys117, Cys29-Cys45, Cys44-Cys98, Cys51-Cys91, Cys60-Cys84 and Cys78-Cys89). Three disulfide bonds; Cys44-Cys98, Cys51-Cys91 and Cys60-Cys84, rigidify the two core helices (40-57 and 83-101). Cys78-Cys89 secure the β-sheet, referred to as the β-wing, to one of the core α-helices (83-101) and Cys11-Cys71 secure the β-sheet to N-terminal α-helix. Two disulfide bridges (Cys 27-Cys117 and Cys29-Cys45) secure the correct relative orientation of the Ca²⁺ binding loop (residues 28-32). The Ca²⁺ ion is bound by seven ligands: the carbonyl oxygens of Tyr28O, Gly30O, Gly32O, the side-chain oxygens of Asp49, and two water molecules. The geometry of the Ca²⁺ coordination is best described as a distorted pentagonal bipyramid. The wiring diagram of the secondary structure is shown in Figure 7-6, and the stereo-diagram of the overall fold is shown in Figure 7-7.
Figure 7-6 Wiring diagram of the secondary structure of ACII-4
H1 = N-terminal α-helix, residues 2-12 H2= 3_{10}-helix, residues 19-22, H3= α-helix 40-57, residues, H4= α-helix, residues 83-101, H5= 3_{10}-helix, residues 106-108, H6= C-terminal α-helix, residues 113-116, A = antiparallel β-strands forming a β-sheet, yellow circles = cysteine residues forming disulfide bridges (orange solid line defines the connection), β = β-turn, γ = γ-turn and ——— = β-hairpin. Picture generated by PDBsum (R. A. Laskowski et al., 2005).

Figure 7-7 Stereo view of the overall fold of molecule A of ACII-4 represented as a cartoon drawing.
The strands are colored yellow, helices red, loops green and the Ca^{2+} is illustrated as a blue colored sphere.

Least-squares fitting of topologically equivalent Ca atoms (and all atoms including side-chains) of residues 1–118 in ACII-4 using the program Superpose (E. Krissinel and Henrick, 2004), resulted in an rmsd of 0.41 Å (0.98 Å). All seven disulfide bonds in both molecules have the same chirality (Figure 7-8). Five of the disulfide bonds have left handed chirality Cys27-Cys117 (\(\chi_3 -87.0^\circ\)), Cys44-Cys98 (\(\chi_3 -74.6^\circ\)),...
Cys51-Cys91 (\( \chi^3 -81.1^\circ \)), Cys60-Cys84 (\( \chi^3 -84.5^\circ \)) and Cys78-Cys89 (\( \chi^3 -97.5^\circ \)) and two disulfide bonds have right handed chirality Cys11-Cys71 (82.9\(^\circ \)) and 29-45 (94.2\(^\circ \)). A closer look at the superposition shows that the only region with different conformations in the two molecules is the segment of residue Trp31-Ser43, which is part of the Ca\(^{2+}\) loop. This difference is discussed in the next section.

![Figure 7-8](image)

**Figure 7-8** Superposition of the two ACII-4 molecules. Molecule A (blue) and B (green) have the same overall fold. All disulfide bonds (orange sticks) have the same chirality. The two disulfide bridges that stabilize the \( \beta \)-wing in ACII-4 are labeled.

The \( \beta \)-wing in ACII-4 is built up by a \( \beta \)-hairpin motif created by the two antiparallel \( \beta \)-strands, (residues 69-72 and 75-78), and a short loop (residues Gly73-Asn74) that connects the \( \beta \)-strands. This region is further stabilized by two disulfide bonds (Cys11-Cys71 and Cys78-Cys89) connecting the \( \beta \)-wing to the \( \alpha \)-helical core of the molecule. The \( \beta \)-wing extends away from the core of ACII-4, and the ‘wingtip’ (Thr72-Gly73-Asn74-Val75) is poorly anchored (Figure 7-8).

### 7.6 The calcium ion in the structure of ACII-4

The crystal structure of ACII-4 contains one calcium ion coordinated to seven ligands. No Ca\(^{2+}\) ions were added during the purification of ACII-4 or in the crystallization conditions. Thus the Ca\(^{2+}\) ion must have been present with the ACII-4
and carried through from the venom. The Ca\(^{2+}\) ion has an average B-value of 13.8 Å\(^2\) which is similar to the average B-value of 15.6 Å\(^2\) for all atoms (Table 7-4), indicating that the Ca\(^{2+}\) ion is firmly bound. This Ca\(^{2+}\) ion is bound to the protein by the seven ligands consisting of the carbonyl oxygen atoms of Tyr28 (2.27 Å), Gly30 (2.36 Å), Gly32 (2.26 Å), the side-chain oxygens of Asp49 (2.58 Å and 2.62 Å) and two water molecules (2.45 Å and 2.35 Å). Five of the seven Ca\(^{2+}\) ligands in ACII-4 lay in one plane (Gly30O, Gly32O, side-chain oxygens of Asp49 and one water molecule), with Tyr28O below the plane and one water molecule above the plane. The ligands form a distorted pentagonal bipyramid (Figure 7-9). Two disulphide bridges (Cys 27-Cys117 and Cys29-Cys45) secure the correct relative orientation of the Ca\(^{2+}\) binding loop (Figure 7-10).

![Figure 7-9](image1.png)

**Figure 7-9** Stereo view of the seven ligands that coordinate to the Ca\(^{2+}\) ion form a distorted pentagonal bipyramid. ACII-4 is colored green with the Ca\(^{2+}\) as a yellow sphere. Two water molecules are represented as red spheres. The side-chain of Asp49 is shown as a stick model. The carbonyl oxygen of Gly28, Gly30 and Gly32 are shown as stick models. Carbon atoms are colored green and oxygen red.

![Figure 7-10](image2.png)

**Figure 7-10** Stereo view of the two disulfide bonds securing the correct relative orientation of the Ca\(^{2+}\) binding loop. Molecule A of ACII-4 is colored green. The Cys27-Cys117 and Cys29-Cys45 disulfide bridges are shown as sticks. Sulfur atoms are orange.
In bovine pancreatic PLA₂ (Dijkstra et al., 1981b; Yu et al., 1993; Arni and Ward, 1996; Steiner et al., 2001) and porcine pancreatic PLA₂ (Dijkstra et al., 1983; van den Bergh et al., 1989; Finzel et al., 1991) a calcium ion is also an essential cofactor. A superimposition is shown in Figure 7-11.

(a) Superposition of a PLA₂ from porcine pancreas (4P2P) (colored pink with Ca\(^{2+}\) as a purple sphere) and a PLA₂ from bovine pancreas (1BP2) (colored steel with Ca\(^{2+}\) ion as a green sphere). (b) Superposition of a PLA₂ from porcine pancreas (4P2P) and molecule A of ACII-4 (colored green with Ca\(^{2+}\) as a yellow sphere).

Superimposition of ACII-4 with the PLA₂s from bovine pancreas (Dijkstra et al., 1981b) and Indian cobra, *Naja naja sagittifera* (Jabeen et al., 2005) showed that the seven ligands and the Ca\(^{2+}\) ion were in nearly identical positions in all three PLA₂s. However, the Ca\(^{2+}\) binding loop has different conformations. A reason for this is that the indole group of Trp31 in ACII-4 is lined against the surface of ACII-4. Similarly the side-chain of the Leu31 in the bovine pancreatic PLA₂ rotates towards the surface of the molecule. However, the side-chain of Arg31 in the PLA₂ from Indian cobra does not interact with the surface of the molecule but instead is rotated out towards the solvent (Figure 7-12). A PEG molecule packed close to the Trp31 side-chain (~3.5 Å) in the ACII-4 structures is likely to influence the orientation of the indole group.
The Ca\(^{2+}\) binding loop has different conformations in the two ACII-4 molecules in the asymmetric unit. The largest separation, 2.66 Å, is between Gly32Cα in molecule A and the equivalent atom in molecule B (Figure 7-13). The reason for the difference in the conformation of the Ca\(^{2+}\) binding loop between the two ACII-4 molecules are the three glycine residues (Gly30, Gly32 and Gly33) which contribute to a flexibility of this loop yet still allow three main chain oxygen atoms (Tyr28O, Gly30O and Gly32O) to participate in coordinating the calcium ion.

**Figure 7-12** The Ca\(^{2+}\) binding loop. The different position of the side-chain of residue 31 in the Ca\(^{2+}\) binding loop between ACII-4, PLA\(_2\) from bovine pancreas and Indian cobra. Molecule A (green) and molecule B (blue) of ACII-4. PLA\(_2\) from bovine pancreas (1BP2) is purple and PLA\(_2\) from Indian cobra is orange (1YXH). The side-chains of residue 31 are shown as sticks. Oxygen is red and nitrogen blue.

**Figure 7-13** The Ca\(^{2+}\) binding loop in ACII-4. The largest distance between the Ca\(^{2+}\) binding loop in molecule A (green) and molecule B (blue) of ACII-4 is 2.66 Å (between Gly32Cα A and Gly32Cα B).

Several crystal structures of PLA\(_2\)s have been determined with Ca\(^{2+}\) ion bound. These include PLA\(_2\)s from Indian cobra, *Naja naja naja*, Chinese cobra, *Naja naja atra* and Indian cobra, *Naja naja sagittifera* (Gua et al., 2002; Jabeen et al., 2005). The sequence of the Ca\(^{2+}\) binding loop is the same in these four PLA\(_2\)s. It has also been shown that PLA\(_2\) from porcine pancreas can bind a second Ca\(^{2+}\) ion outside the Ca\(^{2+}\) binding loop (Slotboom et al., 1978; Dijkstra et al., 1983; Finzel et al., 1991) (Figure 7-14(a)). The crystal structure of a PLA\(_2\) from the snake venom of the Chinese cobra, *Naja naja atra*, contains an additional Ca\(^{2+}\) with a functional role in phospholipase activity (Scott et al., 1990; White et al., 1990) (Figure 7-14(b)). A human nonpancreatic extracellular PLA\(_2\) also contains a second Ca\(^{2+}\) with a role in the phospholipase activity (Scott et al., 1991) (Figure 7-14(c)).
Figure 7-14  Ca$^{2+}$ ions bound to PLA$_2$.
(a) PLA$_2$ from porcine pancreas (4P2P) colored red with two Ca$^{2+}$ ions as blue spheres. (b) PLA$_2$ from Chinese cobra (1POA) colored brown with two Ca$^{2+}$ ions as light blue spheres. (c) Human PLA$_2$ (1POE) colored purple with two Ca$^{2+}$ ions as green spheres.

However, a myotoxin protein from pit viper species Bothrops asper (Arni et al., 1995) and a PLA$_2$ from the Eastern Cottonmouth snake, Agkistrodon piscivoris piscivorus, has a lysine residue at position 49 with the side-chain ε-amino group filling the site normally occupied by the Ca$^{2+}$ ion situated in the Ca$^{2+}$ binding loop (Holland et al., 1990). Non-catalytic PLA$_2$-like proteins from snake venom are characterized by the substitution of an aspartate at position 49 to a lysine (Figure 7-15).

Figure 7-15  PLA$_2$ with Lys49.
The Ca$^{2+}$ is replaced by the ε-amino group of the side-chain of Lys49. (a) Superposition of a myotoxin protein from Pit viper (1CLP) (light blue) and a PLA$_2$ from Eastern Cottonmouth snake (1PPA) (light yellow). Lys49 is shown as a stick model. Nitrogen atoms are colored blue. (b) Molecule A of ACII-4 is superimposed onto the myotoxin protein from Pit viper (1CLP). ACII-4 is colored green with the Ca$^{2+}$ ion as a yellow sphere and Asp49 as a stick model. Nitrogen atoms are colored red.
7.7 Active site

Polar groups of either monomeric or micellar lipid substrates can easily diffuse into the active site of PLA₂s (Roberts, 1996). Two polyethylene glycol molecules were found in molecule A and B of ACII-4, at close to identical positions, occupying the site for a phospholipid substrate (Figure 7-16).

![Figure 7-16 The polyethylene glycol molecules in ACII-4. Molecule A and B of ACII-4 superposed onto each other. ACII-4 molecule A is colored green with its polyethylene glycol molecules as green spheres. ACII-4 molecule B is colored blue with its polyethylene glycol molecules as blue spheres.](image)

The polyethylene glycol molecule fits (situated furthest) inside a pocket lined by Leu2, Phe5, Ile9, Phe99, and the hydrophobic part of the side-chains of Tyr22, Tyr52 (Figure 7-17). The hydrophobic side-chains of residue Leu2, Ile3, Leu20 and Ala23 guard the entrance of the hydrophobic channel and the side-chain of Trp31 is shielding the Ca²⁺ ion (Figure 7-18). Earlier studies on PA11 showed that chemical modification of tryptophan residues at positions 31 and 69 caused loss of all activities (Takasaki *et al.*, 1990a). This indicated an important role for either or both of these two residues. As seen in the crystal structure Trp31 with its large bulky side-chain is located close to the active site and it can be speculated that it can easily move out into the solvent like a gate with the possibility to open and close.
Figure 7-17 The hydrophobic channel leads into the active site in ACII-4. Molecule A of ACII-4 is green and the Ca$^{2+}$ ion is a yellow sphere. The polyethylene glycol molecule located inside the active site is represented as spheres, with carbon atoms in white and oxygen atoms in red.

Figure 7-18 Hydrophobic entrance to the active site of ACII-4. Residues Leu2, Ile3, Phe5, Ile9, Ser19, Leu20, Ala23 and Trp31 guard the entrance of the hydrophobic channel and the side-chain of Trp31 shields the Ca$^{2+}$ ion. The Connolly surface with the electrostatic surface potentials in molecule A of ACII-4 is displayed. Regions of positive charge are colored blue, negative charges are colored red and neutral regions are white. A polyethylene glycol molecule (PEG1) located deep inside the active site is represented as a stick with green carbon and red oxygen atoms.

One end of the polyethylene glycol molecule is positioned close to the catalytic diad (His48 and Asp92) and a water molecule (Figure 7-19; Figure 7-20; Figure 7-21). His48N$\delta_1$ forms a hydrogen bond to this water molecule (2.94 Å and 2.99 Å) and the His48N$\varepsilon_2$ atom forms a hydrogen bond to the Asp92O$\delta_2$ (2.93 Å and 2.87 Å) while
Asp92 is stabilized by Tyr52 and Tyr67. Tyr52O\(_{\beta}\) forms a hydrogen bond to Asp92O\(_{\beta1}\) (2.73 Å and 2.66 Å) and Tyr67O\(_{\beta}\) forms a hydrogen bond to Asp92O\(_{\beta2}\) (2.68 Å and 2.69 Å) (Figure 7-22).

**Figure 7-19** The 2Fo-Fc electron density map contoured at 1\(\sigma\) for His48, Asp49 and Asp92, two water molecules, one polyethylene glycol molecule and one Ca\(^{2+}\) ion for ACII-4. The electron density map is displayed as a light grey mesh. Carbon, nitrogen and oxygen are colored green, red and blue respectively. The polyethylene glycol molecule is displayed as a balls and stick model with purple carbon and red oxygen atoms. Water molecules are represented as red spheres and the Ca\(^{2+}\) ion as a yellow sphere.

**Figure 7-20** The polyethylene glycol molecules and the catalytic diad. Molecule A of ACII-4 is colored green with Ca\(^{2+}\) as a yellow sphere. The polyethylene glycol molecules (PEG1 and PEG2) are represented as balls and stick with purple carbon atoms and red oxygen atoms. The catalytic diad; Asp92 and His48 are represented as sticks with green carbon atoms, red oxygen atoms and blue nitrogen atoms. Water molecules are shown as red spheres. (a) With PEG molecules. (b) Without PEG molecules.
Figure 7-21 The catalytic diad and the catalytic water molecule. Molecule A of ACII-4 colored green with Ca$^{2+}$ as a yellow sphere. The catalytic diad; Asp92 and His48 are represented as sticks with green carbon atoms, red oxygen atoms and blue nitrogen atoms. Water molecules are shown as red spheres.

Figure 7-22 Stereo view of the two tyrosine residues stabilizing Asp92. Molecule A of ACII-4 is colored green. His48, Tyr52, Tyr67 and Asp92 are shown as sticks. Carbon atoms are colored green, oxygen red and nitrogen blue. Distances are in Å.

Superimposing a PLA2 from porcine pancrease with a tetrahedral mimic bound in the active site onto ACII-4 confirmed that the polyethylene glycol molecules are overlapping with the tetrahedral mimic. The polyethylene glycol molecule furthest inside the hydrophobic channel is situated close to the phosphate group of the tetrahedral mimic which is coordinated to the calcium ion. The polyethylene glycol molecule outside the active site touches the alkyl chain of the tetrahedral mimic which extends out from the hydrophobic channel (Figure 7-23).
Figure 7-23  Comparison of active sites of ACII-4 and PLA₂ from porcine pancreas. PLA₂ from porcine pancreas (1FXF) and molecule A of ACII-4 are superimposed. Carbon atoms are colored green for ACII-4 and pink for the PLA₂ from porcine pancreas. The tetrahedral mimic (salmon) and the polyethylene glycol (blue) in ACII-4 are shown as balls and sticks. Residues His48 and Asp49 are shown as stick models. Oxygen is colored red, nitrogen blue, phosphate orange and fluorine white. The Ca²⁺ ion is shown as a purple sphere. The catalytic water is shown as a red sphere. (a) The polyethylene glycol molecules and tetrahedral mimic in the active site. (b) Zoomed in on the active site. (c) Stereo view.

Superimposition of a PLA₂ from Common Indian krait, *Bungarus caeruleus*, (Singh *et al.*, 2005a) with a fatty acid bound in the active site onto ACII-4 showed that the
side-chains of residues Asp48, His49 and Asp92 (ACII-4 numbering) aligned well. Two carbon atoms of polyethylene glycol in the active site superposed onto two carbon atoms of the fatty acid in the active site of the PLA2 from Common Indian krait (Figure 7-24).

Figure 7-24 Comparison of active sites of ACII-4 and PLA2 from Common Indian krait venom.
PLA2 from Common Indian krait venom (1TC8) and molecule A of ACII-4 (green) are superimposed. The fatty acid (salmon) in the PLA2 from Common Indian krait and the polyethylene glycol (blue) in ACII-4 are shown as balls and sticks. The catalytic diad is shown with green carbon atoms for ACII-4 and yellow carbon atoms for the PLA2 from Common Indian krait. Oxygen is colored red and nitrogen blue. The Ca++ ion is shown as a purple sphere. (a) The catalytic diad, catalytic water and polyethylene glycol molecules/fatty acid shown in the active site. (b) Stereo view.

A change in the conformation of the side-chain of Trp19 PLA2 from Common Indian krait venom, which is located at the entry of the substrate-binding site, is observed when comparing a structure of the free PLA2 (1FE5) and complex (1TC8) (Singh et al., 2005b). The side-chain of Trp19 in the free structure of the PLA2 from Common Indian krait venom would have clashed with the fatty acid if it was not rotated. In the complex, the side-chain of Trp19 is more rotated inwards towards the core of the protein. The corresponding amino acid in ACII-4 is Ser19 and its side-chain does not have the ability to obstruct the entrance to the active site. In a crystal structure of a
PLA₂ from the venom of Russell’s viper, *Daboia russelli pulchella*, showed a dimer where a peptide inhibitor (FLSYK) had entered only one of the monomers. The orientation of the Trp31 side-chain was suitable for the binding of the inhibitor in that monomer and it was assumed that in a micellar environment the two monomers would separate and each bind to the phospholipid interface and be catalytically effective (Chandra *et al.*, 2002). Superposing the PLA₂ from Russell’s viper onto ACII-4 showed that the polyethylene glycol molecules in ACII-4 and the peptide inhibitor were located in the catalytic active site of the PLA₂s (Figure 7-25). The side-chain of Trp31 in the inhibited monomer from Russell’s viper had the same rotation as the side-chain of Trp31 in ACII-4. However, the side-chain of Trp31 in the uninhibited monomer from Russell’s viper clashed with the polyethylene glycol molecules (Figure 7-26). I speculate that rotation of the side-chain of Trp31 could regulate the binding of molecules in the active site of the PLA₂.

Figure 7-25 Comparison of bound molecules in the active site. (a) The PLA₂ from Russell’s viper venom (1JQ9) and molecule A of ACII-4 (green) are superimposed. Only ACII-4 with its Ca²⁺ ion (purple sphere) and the polyethylene glycol molecules (light blue carbon atoms) and the inhibitor (yellow carbon atoms) in the active site of the PLA₂ from Russell’s viper venom are shown. (c) Stereo view of the active site. Oxygen is colored blue and oxygen is colored red.
The PLA₂ from Russell’s viper venom (1JQ9) (purple) and molecule A of ACII-4 (green) are superimposed. (a) The location of the inhibitor and polyethylene glycol molecules in the active site. (b) Trp31 in PLA₂ from Russell’s viper is shown as sticks with carbon atoms colored purple and Trp31 in ACII-4 is shown as a stick with carbon atoms colored green. The polyethylene glycol molecules are shown as spheres with white carbon atoms. Oxygen atoms are colored red. The side-chain of Trp31 in the uninhibited monomer of PLA₂ from Russell’s viper venom clashes with the polyethylene glycol molecules in ACII-4. The side-chain of Trp31 in the uninhibited monomer of the PLA₂ from Russell’s viper venom is in a position not observed in ACII-4.

The activity of secreted PLA₂s towards lipid micelles, membranes, and vesicles is several times higher than that on molecular dispersed substrates (Yuan et al., 1990).

It has been established that the PLA₂s bind to the substrate interface via an interfacial binding surface which is different from the site where the phospholipase activity occurs (Verger and de Haas, 1976; Dennis, 1983). The area referred to as the interfacial binding surface have been proposed to consists of a ring of positively charged side-chains surrounding a hydrophobic channel that leads to the centrally located active site (Scott et al., 1990; Scott et al., 1994). Twelve lysine residues and two arginine residues (Lys14, Arg17, Lys57, Lys58, Lys63, Lys70, Lys81, Lys85, Lys105, Lys106, Lys114, Lys115, Arg116 and Lys118) form a ring of positively charged side-chains which surrounds the active site (Figure 7-27; Figure 7-28).

![Figure 7-27 The amino acid sequence of ACII-4](image)

The positively charged (blue) residues and hydrophobic residues (orange) labeled in Figure 7-28 are highlighted.
7.8 Comparison of ACII-4 with other PLA$_2$ molecules

Division of PLA$_2$s into several sub-classes of PLA$_2$s have been carried out based on phospholytic activity, Ca$^{2+}$ dependency and structural features that include the number of disulfide bonds and the number of loop extensions (Dennis, 1994; Arni and Ward, 1996; Dennis, 1997; Six and Dennis, 2000). PLA$_2$s from snake venoms can be classified into two main groups, I and II (Davidson and Dennis, 1990; Dennis, 1994; Six and Dennis, 2000). Group I includes PLA$_2$s from venoms of elapid snakes, sea snakes and from mammalian pancreas (Heinrikson et al., 1977; Davidson and Dennis, 1990; Dennis, 2000). The PLA$_2$s in group I are cross-linked by seven disulfide bridges which stabilize the globular fold. The group is characterized by a disulfide bridge between residues 11 and 77 which secure the $\beta$-wing to the N-terminal major $\alpha$-helix.

ACII-4 belongs to IA. This subgroup is characterized by a surface loop connecting the second major $\alpha$-helix to the $\beta$-wing. This surface loop is between the second major $\alpha$-helix and the first $\beta$-strand (residues Lys58-Trp69 in ACII-4). The loop is located near the entrance of the hydrophobic channel in the active site and is referred to as the “elapid loop” (Figure 7-29; Figure 7-30). In subgroup IB where the PLA$_2$’s from mammalian pancreas is found, this loop is defined by the presence of a few
extra residues to the elapid loop and is referred to as the “pancreatic loop” (Fremont et al., 1993; Carredano et al., 1998; Alape-Girón et al., 1999; S. Xu et al., 2003).

**Figure 7-29** Sequence alignment of PLA₂ to show the elapid loop. Alignment carried out with BLAST and adjusted by manual tracing the superposition of the structures. The elapid loop region is marked with an orange bar underlining residues 56-62 which is also marked in Figure 7-30.

**Figure 7-30** Comparison of group IA and group II PLA₂. The elapid loop and disulfide bridge Cys11-Cys71 in group IA is missing in group II. Molecule A of ACII-4 (green), notechis II-5 from Australian Tiger snake venom (2NOT) (blue) and a PLA₂ from Russell’s viper venom (1VIP) (pink) are superimposed. ACII-4 and notechis II-5 belong to the subgroup IA of group I and the PLA₂ from Russell’s viper belong to group II. All PLA₂ in group I has a Cys11-Cys77 disulfide bridge. Subgroup IA has an elapid loop. Labeled and marked out in orange. Group II does not have an elapid loop and not a disulfide bridge between residues 11 and 77. The disulfide bridge is labeled. The C-terminus of PLA₂s from group II is a few residues longer than PLA₂s from group I.
A search for proteins with similar fold and sequence to ACII-4 in the PDB was performed using the program BLAST (Schäffer et al., 2001). The six top hits are shown in Table 7-6. Their overall structure are similar (Figure 7-32).

**Table 7-6** Hits from BLAST search versus the PDB

<table>
<thead>
<tr>
<th>PDB entry</th>
<th>Amino acid sequence identity (%)</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AE7</td>
<td>65</td>
<td>Notexin from Australian Tiger snake, <em>Notechis scutatus scutatus</em> venom</td>
</tr>
<tr>
<td>2NOT</td>
<td>64</td>
<td>Notechis II-5 from Australian Tiger snake, <em>Notechis scutatus scutatus</em> venom</td>
</tr>
<tr>
<td>5P2P</td>
<td>59</td>
<td>Porcine pancreatic PLA₂ mutant, residues 62-66 deleted, Trp3Phe and Leu31Trp</td>
</tr>
<tr>
<td>1GP7</td>
<td>56</td>
<td>PLA₂ from King cobra, <em>Ophiophagus Hanna</em> venom</td>
</tr>
<tr>
<td>3P2P</td>
<td>58</td>
<td>Porcine pancreatic PLA₂ mutant, residues 62-66 deleted</td>
</tr>
<tr>
<td>1YXH</td>
<td>58</td>
<td>PLA₂ from Indian cobra, <em>Naja naja sagittifera</em> venom</td>
</tr>
</tbody>
</table>

**Figure 7-31** Sequence alignment of PLA₂ proteins homologous to ACII-4. Alignment carried out with BLAST and adjusted by manual tracing the superposition of the structures. The elapid/pancreatic loop region is marked with an orange bar underlining residues 56-67 (ACII-4 numbering). The green bar underlines a loop region between the second β-strand and the third major α-helix consisting of residues 79-83 (ACII-4 numbering).
Figure 7-32  Comparison of the structures of ACII-4 and the six top hits from a BLAST search. Superposition of molecule A of ACII-4 (green), notexin from Australian Tiger snake (1AE7) (white), notechis II-5 from Australian Tiger snake venom (1NOT) (blue), porcine pancreatic PLA₂ mutant (5P2P) (orange), PLA₂ from King cobra venom (1GP7) (yellow), porcine pancreatic PLA₂ mutant, with residues 62-66 deleted (3P2P) (steel blue) and PLA₂ from Indian cobra venom (1YXH) (cerise).

The two PLA₂ from the Australian Tiger snake, Notechis scuatus scutatus, venom, notexin (1AE7) (Westerlund et al., 1992) and notechis II-5 (2NOT) (Carredano et al., 1998) and the PLA₂ from Indian cobra, Naja naja sagittifera, venom (1YXH) (Jabeen et al., 2005) belong to subgroup IA. Notexin is presynaptic neurotoxic and myotoxic and notechis II-5 is presynaptic neurotoxic (Halpert and Eaker, 1975). Notexin is three times as lethal as its relative notechis II-5 (Halpert and Eaker, 1975).

The structure of the PLA₂ from Indian cobra is the first structure of a PLA₂ from group I with strong anticoagulant activity (Jabeen et al., 2005). ACII-4, notexin, notechis II-5 and the PLA₂ from Indian cobra all have an elapid loop. However, the PLA₂ from Indian cobra has one more residue inserted in this region which makes its loop more protruding and exposed (Figure 7-31).

Interestingly, the structure that differs the most from ACII-4 is notexin which has the highest amino acid identity. The largest variation between the structures of ACII-4 and notexin are situated in two areas (Figure 7-31; Figure 7-35). One is the loop region after the second major α-helix (residues Gly56-Tyr67 in ACII-4). The other is the loop region after the second β-strand (residues Asn79-Gly83). Cys60α in ACII-4 and Cys61α in notexin differ 1.93 Å in position. This difference misalign the sequences and instead of Phe61-Pro62 in ACII-4 and Phe67-Pro68 in notexin superimposing onto each other, the hydrophobic side-chain of Phe67 in notexin is in
the same position as the side-chain of Pro62 in ACII-4. The difference in the position of the Pro62 in ACII-4 and Pro68 in notexin steer the loops in different directions and the loops does not align again until Tyr67 in ACII-4 and Tyr73 in notexin. One reason for the difference in structure in this region could be due to binding of a sulfate ion in the notexin structure. Instead of the side-chain of Lys69 pointing out to the solvent as the side-chain of Lys69 in the notechis II-5 structure, the side-chain of Lys69 in notexin points inwards the molecule and form a bond to the sulfate ion (Figure 7-33). This indicates that the loop between the second major α-helix and the first β-strand in this group of PLA$_2$ molecules can have different conformations.
Figure 7-33 Differences between ACII-4 and notexin. Molecule A of ACII-4 (green) and notexin (blue). Sulfate ion is shown as balls. Oxygen is colored red, nitrogen blue and sulfur is yellow. (a) Residues mentioned in the text are labeled. Distance is in Å. (b) Stereo view.

A single amino acid insertion in the region after the second β-strand and a change from Pro82 in ACII-4 to Lys88 in notexin have given notexin a loop structure in this region which is more protruding compared to ACII-4. However, the Cys60-Cys84 disulfide bridge (ACII-4 numbering) anchors the loop to the core of the structure (Figure 7-34).
Figure 7-34 Differences between ACII-4 and notexin.
Molecule A of ACII-4 (green) and notexin (blue). Oxygen is colored red, nitrogen blue and sulfur is yellow. (a) Residues mentioned in the text are labeled. Distance is in Å. (b) Stereo view.

The structures of ACII-4 and notechis II-5 follow the same path after the second major α-helix. Notechin II-5, like notexin, has an insertion of one residue in this region and a lysine residue replacing Pro82 in ACII-4. Thus, after the second β-strand the structure of notechin II-5 and notexin follow the same trail.

Figure 7-35 Structural differences between ACII-4 and the BLAST top hits belonging to group IA.
Molecule A of ACII-4, notexin from Australian Tiger snake venom (1AE7) (white), notechis II-5 from Australian Tiger snake venom (1NOT) (blue) and PLA2 from Indian cobra venom (1YXH) (cerise) are superimposed. The regions with the largest variations between the structures are marked with orange bars. These regions are the elapid/pancreatic loop region residues 56-67 in ACII-4 and a loop region between residues 79-83 in ACII-4.
The pancreatic loop of the mutants of the PLA2 from porcine pancreas (5P2P and 3P2P) (Kuipers et al., 1989; Thunnissen et al., 1990a) have been shortened to the same size as the elapid loop of the PLA2 from the Indian cobra venom. Thus the elapid loop in the two mutants of the PLA2 from porcine pancrease and the PLA2 from the Indian cobra differ with one amino acid to the elapid loop in ACII-4. The PLA2 from King cobra, Ophiophagus Hannah, venom (1GP7) (Zhang et al., 2002) is the first crystal structure of a PLA2 from venom with a pancreatic loop and differ with six residues to the elapid loop in ACII-4, notexin and notechis II-5 (Figure 7-31; Figure 7-36). The elapid loop in the two mutants of the PLA2 from porcine pancrease, the PLA2 from Indian cobra and the pancreatic loop in the PLA2 from King cobra are therefore more protruding and exposed compared to ACII-4 (Figure 7-32). The loop region after the second β-strand in the β-wing in the two mutants of the PLA2 from porcine pancreas, the PLA2 from Indian cobra and the PLA2 from King cobra venom follow the same path which differs from ACII-4 (Figure 7-32; Figure 7-37).

Figure 7-36  The elapid loop in PLA2 belonging to group IA and the pancreatic loop in PLA2 belonging to group IB.
Molecule A of ACII-4 (green), the PLA2 from Australian Tiger snake venom referred to as notechis II-5 (2NOT) (blue) and the PLA2 from King cobra venom (1GP7) (yellow) are superimposed.
Figure 7-37 Structural differences between ACII-4 and the BLAST top hits belonging to PLA$_2$s from group II. Molecule A of ACII-4, porcine pancreatic PLA$_2$ mutant (5P2P) (orange), PLA$_2$ from King cobra, Ophiophagus Hanu venom (1GP7) (yellow) and porcine pancreatic PLA$_2$ mutant (3P2P) (blue steel) are superimposed. The regions with the largest variations between the structures are marked with orange bars. These regions are the elapid/pancreatic loop region residues 56-67 in ACII-4 and a loop region between residues 79-83 in ACII-4.

7.9 Crystal contacts and electrostatic surface

The crystal packing is very tight with only 43% solvent content (Chapter 3). Analysis revealed that there are about the same number of surrounding atoms within a distance of 5 Å in both molecule A and B of ACII-4. These contacts are distributed over similar areas. The crystal contacts are around the N-terminal α-helix, the catalytic site of the phospholipase activity, the loop region Thr36-Glu41 prior to the N-terminal side of the second major α-helix, the C-terminal side of the second major α-helix, the first β-strand of the β-wing, the minor 3$_{10}$-helix consisting of residues 106-108 and the C-terminal minor α-helix (Figure 7-38). It is likely that these regions that form crystal contacts are also important for biological activity. The most clear cut example of this is that the PEG molecules in the active site mimic the phospholipase substrate (Section 7.7) (Volwerk et al., 1974; Scott et al., 1990; Thunnissen et al., 1990a).
Figure 7-38 Contacts between the surrounding atoms and molecule A of ACII-4. Connolly surface of molecule A of ACII-4 is green. Atom within 5 Å are shown. Carbon atoms are green from molecule A and blue from molecule B. Carbons atoms from polyethylene glycol molecules are white, yellow and salmon. Sulfur atoms from sulfate ions are yellow. Nitrogen is blue and oxygen is red. Figures are rotated 90° around the vertical axis.

The crystal contacts are similar in both molecule A and B of ACII-4 (Figure 7-45; Figure 7-46). The region corresponding to residues 53-70 in ACII-4 (Figure 7-39) contain positively charged residues and is the most likely source of the anticoagulant activity (Kini and Evans, 1987; Carredano et al., 1998; Jabeen et al., 2005; Prijatelj et al., 2006). The crystal contacts that are made by residues 53-70 in ACII-4 are marked out and are discussed below (Figure 7-41; Figure 7-42; Figure 7-43; Figure 7-44).

![Figure 7-39](image)

The amino acid sequence of ACII-4.

In both molecule A and B of ACII-4 a polyethylene glycol molecule forms a semicircle around the Lys63Nζ at a distance of ~3 Å and this polyethylene glycol molecule is also in hydrophobic contact to Trp31 (~3.5 Å) in both molecules. This molecule is also close to Phe61 (~4 Å) from an adjacent molecule (Figure 7-41; Figure 7-42). Leu64 and Phe61 from both molecules form a hydrophobic cluster and sandwich the polyethylene glycol molecules, together with the side-chain of Trp31 which is in hydrophobic contact (~4 Å) to the side-chain of Lys57, in both molecule A and B (Figure 7-40; Figure 7-41; Figure 7-42).
Figure 7-40  Region 61-64 and Trp31 of ACII-4 participating in a hydrophobic interface. Carbon atoms in molecule A of ACII-4 are colored green and molecule B is colored blue. Oxygen atoms are colored red and nitrogen blue. Trp31, Lys57, Lys63, Leu64 and Phe61 are shown as sticks. Polyethylene glycol is shown as a ball and stick model with carbon atoms yellow (PEG2) or salmon (PEG2’) are forming a semicircle around the Lys63Nʐ.

Trp69 and Lys70 are situated at the start of a β-strand of the β-wing, form crystal contacts with the N-terminal α-helix in both molecule A and B of ACII-4 (Figure 7-41). A sulfate ion (Figure 7-41 and) is close to the N-terminal α-helix and bound to Gln4Nɛ2 (2.96 Å and 2.94 Å), Asn7Nδ2 (2.94 Å and 2.89 Å) and Trp69Nɛ1 (2.84 Å and 2.92 Å). The second sulfate ion is bound to Lys115Nʐ(A) (3.01 Å) and Arg17Nη2(A) (2.41 Å) and Arg17Nɛ(A) (2.80 Å). Arg17 and Lys115 are part of the region with positively charged residues referred to as the interfacial binding surface (Scott et al., 1990; Scott et al., 1994; Mounier et al., 2000) (Section 7.10). Gln54Oɛ1 in both molecule A and B is forming hydrogen bonds to Gln54Nɛ2 (3.13 Å in both molecule A and B) and Lys58Nʐ from the adjacent molecule (2.93 Å in both molecule A and B) (Figure 7-41). The β-wing also forms crystal contacts. Asn74 is situated in the short loop connecting the two antiparallel β-strands, the wing-tip, and its side-chain extends out from the molecule. In both molecule A and B of ACII-4, the Asn74Nδ2 atom forms a hydrogen bond to Gly15O in the adjacent molecule (2.98 Å and 2.96 Å in molecule A and B, respectively) (Figure 7-44; Figure 7-49; Figure 7-50).
Connolly surface of molecule A of ACII-4 is shown with electrostatic surface potentials overlaid. Regions of positive potential are colored blue, negative potential are colored red and neutral regions are white. Atom within 5 Å are shown. Carbon atoms are green from molecule A and blue from molecule B. Carbons atoms from polyethylene glycol molecules and are white (PEG1), yellow (PEG2) and pink (PEG2'). Sulfur atoms from sulfate ions are yellow. Nitrogen is blue and oxygen is red. Atoms labeled in black or white are part of molecule A. Atoms labeled in light blue are the residues 53-70 making crystal contacts to molecule A.

Molecule A of ACII-4 is rotated by 90° around the vertical axis compared to Figure 7-41. The same labeling system as in Figure 7-41 is used.
Figure 7-43 Crystal contacts for ACII-4. Molecule A of ACII-4 is rotated by 90° around the vertical axis compared to Figure 7-42. The same labeling system as in Figure 7-41 is used.

Figure 7-44 Crystal contacts for ACII-4. Molecule A of ACII-4 is rotated by 90° around the vertical axis compared to Figure 7-43. The same labeling system as in Figure 7-41 is used.
Figure 7-45  Four views of molecule A of ACII-4.
Crystal contacts between the surrounding atoms and molecule A of ACII-4. Connolly surface of molecule A of ACII-4 is shown with electrostatic surface potentials overlaid. Atoms within 5 Å are shown. Carbon atoms are green from molecule A and blue from molecule B. Carbons atoms from polyethylene glycol molecules are white, yellow and pink. Sulfur atoms from sulfate ions are yellow. Nitrogen is blue and oxygen is red. Figures are rotated 90° around the vertical axis.

Figure 7-46  Four views of molecule B of ACII-4.
Same description of molecule and atoms as in Figure 7-45.

7.10  Dimerization of ACII-4

Previous studies showed that ACII-4 could exist as a dimer (Masci, 2000). Two possible dimers are shown, dimer AB and dimer AB', in Figure 7-47 and Figure 7-48. The buried surface of the AB dimer (1141 Å²) is larger then the AB' dimer (686 Å²) (AREA1MOL (B. Lee and Richards, 1971) which would suggest this is the most likely dimer in solution. Further more, the AB dimers have both active sites exposed to solvent, where as the AB' dimer has the active site located close to the buried surface. The electrostatic areas are differentely distributed in the two dimers AB and AB' of ACII-4 (Figure 7-47; Figure 7-48). The AB interface is more hydrophilic than the AB' interface. One of the sulfate ions is important for mediating the major intermolecular contacts in the AB interface. Except for the contacts mediated by the sulfate ion and the contacts formed by the Asn74 residue situated in the wing-tip, two more hydrogen bonds are formed between Asn7Oδ₁ and Cys71N (2.90 Å and 2.75 Å in molecule A and B, respectively), and Asn7Nδ2 and Trp69O (3.27 Å and 3.07 Å in
molecule A and B, respectively) (Figure 7-49; Figure 7-50). The interactions in the 
AB’ interface are the hydrophobic contacts formed by the polyethylene glycol 
molecules, residues 61-64 and Trp31 (Figure 7-51).

Figure 7-47  AB dimer formation.
(a) Showing the two ACII-4 molecules in the asymmetric unit. The Ca\(^{2+}\) ions are shown as 
yellow spheres. A sulphate ion is located in between the two molecules represented as 
spheres and sulfur atoms are yellow and oxygen red. The two polyethylene glycol molecules 
are visible close to the active site as spheres with carbon atoms colored pink and oxygen 
atoms colored red. (b) Connolly surface with the electrostatic surface potentials mapped out. 
Positive charge is colored blue, negative charge is colored red and neutral is white. The 
polyethylene glycol molecules located close to the active site are represented as green 
spheres.

Figure 7-48  AB’ dimer formation.
(a) Showing molecule A from the asymmetric unit and a symmetry related molecule B. The 
Ca\(^{2+}\) ions are shown as yellow spheres. (b) The two polyethylene glycol molecules are 
visible close to the active site as spheres with carbon atoms colored pink and oxygen atoms 
colored red. (c) Connolly surface with the electrostatic surface potentials mapped out. 
Positive charge is colored blue, negative charge is colored red and neutral is white. The 
polyethylene glycol molecules located close to the active site are represented as green 
spheres.
Figure 7-49  Interactions in the AB interface
Molecule A (green) and molecule B (blue) of ACII-4 pack together with a sulfate ion shown as balls and stick mediating the major intermolecular contacts. The residues Asn7, Gly15, Trp69, Cys71 and Asn74 form contacts in the interface. Oxygen atoms are colored red and nitrogen blue. Distances are in Å.
Figure 7-50  Stereo view of the interactions in the AB interface.  
Description of the picture is given in Figure 7-49

Figure 7-51  Stereo view of the interactions in the AB’ interface.  
Molecule A (green) and molecule B (blue) of ACII-4 pack together with two polyethylene glycol molecules shown as balls and stick in the D interface. The carbon atoms in PEG2 are colored yellow and in PEG2’ salmon. Trp31, Lys57, Lys63, Leu64 and Phe61 are shown as sticks. Oxygen atoms are colored red and nitrogen blue.
Oligomerization of PLA₂ has been stated to affect the pharmacological activities (Soares et al., 2000) and affect the phospholytic activity (Rigden et al., 2003). Superposing the crystal structure of the PLA₂ from Common Indian Krait, Bungarus caeruleus, venom with a fatty acid bound to the active site (Singh et al., 2005a) onto each monomer of ACII-4 in the different dimer formations showed that the fatty acid fits into both dimer constellations (Figure 7-52).

![Figure 7-52](image.png)

**Figure 7-52** Superposing the PLA₂ from Common Indian krait (1TC8) onto each monomer of the two ACII-4 dimer formations shows the fitting of the fatty acid. Molecule A (green) and molecule B (blue) of ACII-4. The fatty acid represented as spheres with purple carbon atoms and red oxygen atoms. (a) The fatty acid can bind in the AB dimer formation. (b) The fatty acid can bind in the AB’ dimer formation.

Superposing the crystal structure of the PLA₂ from Russell’s viper, Daboia russelli pulchella, venom with a peptide inhibitor (FLSYK) in the active site (Chandra et al., 2002) on each of the monomers in the two ACII-4 dimer formations showed that the inhibitor clashes with the facing monomer in the AB dimer formation but fits in the AB’ dimer formation. This indicates that the active site is more restricted in the AB dimer than the AB’ dimer.

![Figure 7-53](image.png)

**Figure 7-53** Superposing the PLA₂ from Russell’s viper (1JQ9) onto each monomer of the two ACII-4 dimer formations shows the fitting of the protein inhibitor. The two ACII-4 molecule in the asymmetric unit are green and blue. The protein inhibitor represented as spheres with yellow carbon atoms and red oxygen atoms. (a) The protein inhibitor clashed with the facing monomer in the AB dimer formation. (b) The protein inhibitor can bind in the AB’ dimer formation.
7.11 Pharmacological activities

It has previously been shown that ACII-4 has Ca$^{2+}$ dependent anticoagulant activity and phospholipase activity (Masci, 2000). However, the individual activities are not dependent on each other (Masci, 2000). The observation suggests that the different activities are located in separate regions of the molecule.

A comparison of sequences of PLA$_2$s from the genus *Naja* (group I) with different grades of anticoagulant potency has shown that the region corresponding to residues 53-70 in ACII-4 is strongly anticoagulant when positively charged but weakly or non-anticoagulant when negatively charged (Kini and Evans, 1987). A comparison of the sequences of a PLA$_2$ from Black-necked Spitting cobra, *Naja nigricollis*, venom (group I) and a PLA$_2$ from Russell’s viper, *Vipera russelli russelli*, venom (group II) suggested their strong anticoagulant activity is due to a positively charged ridge of lysine residues situated in this region (marked with blue in the amino acid sequence of the PLA$_2$ from Black-necked Spitting cobra venom and Russell’s viper venom; Figure 7-54), together with a negatively charged residue (marked with red in the amino acid sequence of the PLA$_2$ from Black-necked Spitting cobra venom and Russell’s viper venom; Figure 7-54) (Kini and Evans, 1987; Carredano *et al.*, 1998). ACII-4 contains four positively charged residues in this region (Lys57, Lys58, Lys63 and Lys70) and a negatively charged glutamate at position Glu53.

The strong anticoagulant property from Indian cobra, *Naja naja sagittifera*, venom (group I) (Section 7.8) is suggested to be in the region from Glu57 to Glu71 the PLA$_2$ (marked with red in the amino acid sequence of the PLA$_2$ from Indian cobra venom; Figure 7-54). This sequence includes three charged residues (marked with blue in the amino acid sequence of the PLA$_2$ from Indian cobra venom; Figure 7-54) (Figure 7-55) (Jabeen *et al.*, 2005).

Studies on an anticoagulant PLA$_2$ from the venom of *Agkistrodon halys* Pallas (group II) suggested that the residues Glu53 and Trp70 in this PLA$_2$ (Figure 7-54), might play an important role in the anticoagulant activity (Zhao *et al.*, 2000) (marked with red in the amino acid sequence of the PLA$_2$ from *Agkistrodon halys* Pallas venom; Figure 7-54) (Figure 7-55). These residues also correspond to a negatively charged and a hydrophobic residue, Glu53 and Leu64, in ACII-4.
Figure 7-54 Sequence alignment of PLA2 with anticoagulant activity.
The alignment was carried out using BLAST and adjusted by manually tracing the superposition of the structures (1VIP, 1YXH, 1JIA and ACII-4). Reference for the amino acid sequences for *Naja nigricollis* not found at the PDB is (Kini and Evans, 1987; Carredano *et al.*, 1998). Residues mentioned in the text are marked.

Superimposition of the structures of the PLA2 from Russell’s viper (Carredano *et al.*, 1998), Indian cobra (Jabeen *et al.*, 2005) and *Agkistrodon halys* Pallas (Zhao *et al.*, 2000) venom onto molecule A of ACII-4 and comparing the region between residue 53-70 (ACII-4 numbering) shows the conserved residues at equivalent positions in the structures. Tyr67 (ACII-4 numbering) is conserved and its side-chain has the same conformation in all four structures. In ACII-4, and in the three other structures, the side-chain of Tyr67 is buried in the core and confirms that this residue is not significantly important for interface interactions but does provide structural support and aids in the catalytic activity by stabilizing the carboxylate group of the catalytic Asp92 (ACII-4 numbering). All four PLA2s have an aromatic side-chain, either a tryptophan or a tyrosine, at position 69 (ACII-4 numbering). This aromatic side-chain creates a hydrophobic environment for the side-chain of Tyr67 and Tyr67 which in turn stabilizes the catalytic Asp92 (ACII-4 numbering).

The Cys60–Cys84 disulfide bridge and the Pro62 are also conserved in all four structures (ACII-4 numbering). The conservation of these residues indicates that they have important structural roles. The Cys60–Cys84 disulfide bridge stabilizes the loop region between residues 53-70 and the Pro62 produces a kink in the loop. All four PLA2s have a hydrophobic residue, either an alanine or a valine residue, at position...
55 (ACII-4 numbering). This residue is situated in between the two major α-helices which run antiparallel and is not exposed to solvent.

Of the positively charged residues in the region between residue 53-70 (ACII-4 numbering) only one (Lys63, ACII-4 numbering) is conserved in all four structures. Importantly, its side-chain is exposed and free for molecular interaction. In addition, all four PLA₂s have a hydrophobic residue, either a leucine or a tryptophan, at position 64 (ACII-4 numbering) and a polar residue, either a serine or a threonine, at position 68 (ACII-4 numbering) exposed and free for molecular interaction. Three of the PLA₂s (ACII-4, PLA₂ from Russell’s viper and PLA₂ from Agkistrodon halys Pallas venom) have a glutamate at position 53 (ACII-4 numbering). Suggesting this residue may also be important for the anticoagulant activity. The PLA₂ from Russell’s viper and Agkistrodon halys Pallas venom do not have a residue which aligns with Lys57 in ACII-4 while the PLA₂ from Indian cobra venom has a negatively charged residue in this location. Lys58 in ACII-4 is a polar or aliphatic residue in the other three PLA₂s. Lys70 in ACII-4 has changed into a polar or negatively charged residue, in the other three PLA₂s (PLA₂ from Russell’s viper, Indian cobra and Agkistrodon halys Pallas venom). This also suggests that out of the positively charged residues in the region between residue 50-73 in ACII-4, only Lys63 in ACII-4 may have the same role in anticoagulant activity as the PLA₂ from Russell’s viper, Indian cobra and Agkistrodon halys Pallas venom.

However, Lys57 and Lys58 are conserved in ACII-4, notexin and notechis II-5. ACII-4, notexin and notechis II-5 have presynaptic neurotoxic and myotoxic activities (Rowan et al., 1989; Westerlund et al., 1992; Carredano et al., 1998). This indicates that Lys57 and Lys58 may be significant in these pharmacological activities.
Figure 7-55 Comparison of residues with possible anticoagulant effect.

Residues 53-76 (ACII-4 numbering) are shown as sticks models. ACII-4 colored green with side-chains yellow. (a) PLA$_2$ from Russell’s viper (1VIP) and molecule A of ACII-4 are superimposed. PLA$_2$ from Russell’s viper colored purple with side-chains blue. (b) The PLA$_2$ from Indian cobra (1YXH) and ACII-4 are superimposed. PLA$_2$ from Indian cobra colored pink with side-chains blue. (c) The PLA$_2$ from Agkistrodon halys Pallas (1JIA) and ACII-4 are superimposed. PLA$_2$ from Agkistrodon halys Pallas colored purple with side-chains light blue. Oxygen atoms are colored red and nitrogen atoms are blue.

The anticoagulant activity of PLA$_2$s arises from different mechanisms. It has been suggested that enzymatic activity is essential for the anticoagulant properties in some cases e.g. by hydrolyzing procoagulant phospholipids (Verheij et al., 1980a; Evans and Kini, 1997; Andrião-Escarsoa et al., 2000). Other investigations have proposed that some PLA$_2$s compete with clotting proteins for the lipid surface by interacting with phospholipids and thus inhibit coagulation (Prigent-Dachary et al., 1980). On the other hand, a study on three anticoagulant PLA$_2$s from Black-necked spitting cobra, _Naja nigricollis_, venom (group I) showed that the extrinsic tenase (tissue factor-factor VIIa) complex was inhibited by all three PLA$_2$s but the prothrombinase complex was inhibited by only one of the PLA$_2$s (Stefansson et al., 1989; Kini and Evans, 1995; Kini, 2006). Another study showed that secreted PLA$_2$s from human (group II) bind to coagulation factor Xa thus inhibiting formation of the prothrombinase complex and inhibiting coagulation (Mounier et al., 2000). Two regions of the secreted PLA$_2$s from human (group II) have been proposed to exert the
anticoagulant effect. One corresponds to residues 51-75 in ACII-4 (marked with blue in the amino acid sequence of the secreted PLA2 from human; Figure 7-56) (Mounier et al., 1998). The other region reported to bind to factor Xa and inhibit prothrombinase activity are made up by residues in the putative interfacial binding surface; Arg7, Lys10, Lys16, Lys38, Lys74, Lys87, Arg92, Lys110, Lys115, Lys116, Lys124 and Lys127 in the secreted PLA2 from human (marked with red in the amino acid sequence of the secreted PLA2 from human; Figure 7-56) (Snitko et al., 1997; Mounier et al., 2000).

Superimposing the structure of the secreted PLA2 from human (1POE) and ACII-4 showed that there are three conserved positively charged residues in the region that corresponds to residues 51-75 in ACII-4. These are Lys57, Lys58 and Lys63 in ACII-4 (Figure 7-56; Figure 7-57). Three positively charged residues in the secreted PLA2 from human have changed in ACII-4. Glu53, Gln54 and Ser68 have become Lys53, Arg54 and Lys68 in human PLA2. On the other hand Lys70 in ACII-4 is Ser69 in the secreted PLA2 from human. In the putative interfacial binding surface, five positively charged residues are conserved; Arg17, Lys85, Lys106, Lys114 and Lys118 in ACII-4 (Figure 7-57). A substantial number of residues have change from a positive charge; one from positive to negative (Glu107 in ACII-4), four have changed from positive to polar (Asn7, Gln10, Ser80 and Ser68 in ACII-4), one has changed from positive to neutral (Val38 in ACII-4) and one is an insertion in the secreted PLA2 from human (Lys102 in the secreted PLA2 from human is an insertion) (Figure 7-57). The lack of many conserved residues and difference in the charge distribution between ACII-4 and the secreted PLA2 from human, in the region corresponding to residues 51-75 in ACII-4 and the putative interfacial binding

![Figure 7-56](attachment:image.png) Sequence alignment of PLA2 with anticoagulant activity. Alignment carried out with BLAST and adjusted by manual tracing the superposition of the structures (1POE and ACII-4).
surface, indicate that these areas are likely to have different specificities in these two PLA2s. Overall, this supposes that the anticoagulant activity of ACII-4 arises from the same mechanism as the closely related PLA2 from the Black-necked Spitting cobra, and target the extrinsic coagulation cascade, e.g. the extrinsic tenase (tissue factor-factor VIIa)) complex (Stefansson et al., 1989; Kini and Evans, 1995).
Figure 7-57  Comparison of residues with possible anticoagulant effect.  
The human secreted PLA₂ (1POE) and molecule A of ACII-4 are superimposed. ACII-4 colored green with side-chains yellow. Oxygen atoms are colored red and nitrogen atoms are blue. (a) Residues 51-75 (ACII-4 numbering) are shown as sticks. Human secreted PLA₂ colored purple with side-chains blue. (b) The side-chains in the interfacial binding surface are shown as sticks. ACII-4 colored green with side-chains yellow. Human secreted PLA₂ colored purple with side-chains pink.

A study on an anticoagulant PLA₂ from the venom of *Vipera ammodytes ammodytes* (group II) showed that four positively charged residues Arg72, Lys74, His76 and Arg77 (marked with blue in the amino acid sequence of the PLA₂ *Vipera ammodytes ammodytes* venom; Figure 7-58) bind directly to human coagulation factor Xa and are involved in inhibition of prothrombinase complex formation and are therefore critical for the anticoagulant effect (Prijatelj *et al.*, 2006). However, only His76 in this PLA₂ aligns with a positively charged residue in ACII-4 (Lys70). Basic residues located at the C-terminus, Arg118, Lys128 and Lys132 in the PLA₂ from the *Vipera ammodytes ammodytes* venom have also been proposed to bind to factor Xa (Prijatelj *et al.*, 2006). Arg118 is the only one of these residues that can be aligned with ACII-4 (marked with blue in the amino acid sequence of the PLA₂ *Vipera ammodytes ammodytes* venom; Figure 7-58) (correspond to Tyr109 in ACII-4 from the sequence...
alignment) the other (Lys128 and Lys132) are part of the extra residues in the C-terminus which exist in the PLA₂ₘ from group II but not group I. This further suggests that the anticoagulant activity of ACII-4 does not arise from interaction with factor Xa but has a similar mechanism to the closely related PLA₂ from the Black-necked Spitting cobra, thereby targeting the extrinsic coagulation cascade (Stefansson et al., 1989; Kini and Evans, 1995).

The crystal structure of the anticoagulant PLA₂ from the venom of Agkistrodon halys Pallas (group II) showed that the molecule contains a positively charged face at the C-terminus. It is suggested that this region might be involved in hemolytic activity (Zhao et al., 2000). There are four residues contributing to a positively charge area in the C-terminus in ACII-4 (Lys114, Lys115, Arg116 and Lys118), however, only Lys118 aligns with a positively charged residue from the PLA₂ from the Vipera ammodytes ammodytes venom (Figure 7-58).

Lys114, Lys115 and Arg116 (ACII-4 numbering) are conserved in ACII-4, notexin and notechis II-5, thus it is possible that this region is important for presynaptic neurotoxic and myotoxic activity.

<table>
<thead>
<tr>
<th>1 5 10 15 20 25 30 35 40 45</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACII-4</td>
</tr>
<tr>
<td>NLIQFQGMICANKGSRPSL-DYADYGCGWGGSSGTPVDEKLRCCQVHD</td>
</tr>
</tbody>
</table>

Vipera ammodytes ammodytes
SLLEFGMILGETG--KNPLTSYSFYGCYGVGGKGTKPSTKDATDRCCFDVHD

1JIA Agkistrodon halys Pallas
HLLQFRKMIKKMTG--KEFVVSAYAFIICYGSGSRGKPKDATDRCCFDVHD

<table>
<thead>
<tr>
<th>50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 124</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCYEQAGKGG-CFPKLTILYSKCT--GNVPTCNS--KGCKSFGVCCDKAAAACDKFAKA--FVRKRKNI--DTKTKKCRK---</td>
</tr>
<tr>
<td>CCYG---NLIPDC3PSTKDGIFNGA--IVCG--KGSCHNRCEDDAAACICFRKNLKTYNYTAYHYP-D-FLCKKESEKC</td>
</tr>
<tr>
<td>CCYEKVT--T-GCDPKWDDYTSWKNG--TIVCGG--DDEPKKEVCECDKAAAICFRDNLKTYKRYMAYP-D-ILCSKSEKC</td>
</tr>
</tbody>
</table>

Figure 7-58 Sequence alignment of PLA₂ with anticoagulant activity.
The amino acid sequences for Vipera ammodytes ammodytes is taken from (Prijatelj et al., 2006).

A study on three PLA₂ from Small-eye snake, Micropechis ikaheka, venom (group I) showed an increasing number of positive charges on the surface of the C-terminal face and higher levels of oligomerization, correlated with higher anticoagulant potency (Lok et al., 2005) (Figure 7-59). The C-terminal face of ACII-4 has several areas with positive charge (Figure 7-59 (a)). Comparing the charge distribution over the C-terminal face of ACII-4 and the PLA₂ from Small-eye snake shows that ACII-
4 is more similar to the C-terminal face of the PLA₂ with the strongest anticoagulant strength than the PLA₂ with the lowest anticoagulant strength (Figure 7-59).

![Figure 7-59](image)

**Figure 7-59** Comparison of the electrostatic surface potentials.
The C-terminal face of ACII-4 and two PLA₂ from Small-eye snake venom. Connolly surface with the electrostatic surface potentials mapped out. Positive charge is colored blue, negative charge is colored read and neutral is white. (a) Molecule A of ACII-4. (b) The PLA₂ from Small-eye snake venom with the strongest anticoagulant strength. (c) The PLA₂ from Small-eye snake venom with the lowest anticoagulant strength.

### 7.12 Summary

The structure of ACII-4 has been determined by molecular replacement to a resolution of 1.56 Å after detwinning the reflection data. There are two ACII-4 molecules in the asymmetric unit. ACII-4 has a characteristic PLA₂ fold stabilized by seven disulfide bonds stabilizing the fold. One Ca²⁺ ion is bound by seven ligands, which are the carbonyl oxygens of Tyr28O, Gly30O, Gly32O, the side-chain oxygens of Asp49, and two water molecules, positioned at the edges of a distorted pentagonal bipyramid. A PEG molecule is situated in the active site and appears to mimic the binding of fatty acid substrate. The PEG molecule is close to a water molecule and the catalytic His48 and Asp92. The carboxylate group of Asp92 is stabilized by the hydroxyl group of Tyr52 and Tyr67.

Two possible dimer formations of ACII-4 can be deduced from the crystal lattice. One is a hydrophilic interface with a buried surface of 1141 Å² and the other is a hydrophobic interface with a buried surface of 686 Å².
Chapter 8
Conclusions and Future Directions

This study has focused on structural investigations of two molecules isolated from the venoms of Australian snakes. While the venom itself does provide a rich source of a range of different compounds, the amounts that are produced for each component is relatively minute, thus making it impossible to concentrate this sample to the level needed for crystallization trials. If restricted to only working with protein from the native source it is unlikely that the project would have been a success. For this work, I was in the fortunate position to have recombinant protein available. Tens of milligrams of rTxln-1 were required to initially crystallize the free molecule, though lesser amounts were needed in crystallization trials of the complexes. For rTxln-1, three crystal structures have been determined and the results of the structural studies from this work are presented in Chapters 4, 5 and 6.

Nowadays, the availability of robots to pipette solutions, prepare crystallization plates and monitor the results of experiments has meant that much smaller samples of purified protein are required to carry out the crystallization process. It is therefore conceivable that, in the future, there could be a return to structure determination of proteins from their native sources. As an example of the efficiency of the robotics approach, the “Mosquito” can dispense volumes as low as 100 nL for crystallization trials. If three hundred crystallization trials were attempted on a sample concentrated to 10 mg/mL only ~0.3 mg of protein would be required. Thus it is conceivable that the structures of a large majority of venom proteins could now be determined in this way.

My study on the crystallography of ACII-4 is an example where a protein from a native source has been successfully crystallized. In this instance, a relatively small sample of protein was provided, <5 mg in total, but I was able to produce crystals that diffracted to better than 1.6 Å resolution. With the use of the new robotics systems that are now available it is clear that I would have been able to determine the structure with a sample of much less than 1 mg of total protein.
In a recent proteomics study of 18 Australian snake venoms it has been shown that there are of the order of 200 or more unique protein/peptide components per snake (Birrell et al., 2007). With this number in mind, and given that this study is the first of its kind, it is apparent that the field of “structural venomics” of Australian snakes is wide-open. Of particular interest would be a complete analysis of the structure and function of all of the different Kunitz type inhibitors from all of the venomous snakes in Australia. In addition to the studies with Txtln-1, our group has recently identified two new Kunitz-type molecules that are potent inhibitors of plasma kallikrein but are only weak inhibitors for plasmin. Furthermore, using a genomics approach, five other Textilinins and more than twenty other Kunitz-type inhibitors have been identified from Australian snake venoms (Filippovich et al., 2002). It will therefore be interesting to investigate the structure, specificity and inhibitory activities of each of these.

As with the Kunitz-type inhibitors, there are also a large number of PLA$_2$s present in snake venoms. It is estimated that there are at least 15 different isoforms present in each individual snake (Braganca and Sambray, 1967; Takasaki et al., 1990b). Once the biological activities of all of these have been fully characterized it is certain that very clearly defined locations on these proteins will be ascribed to specific biological functions. My analysis has shown that some links between structure and function can now be made, but they are by no means definitive (Chapter 8; Figure 8-1). Further characterization of PLA$_2$s form the venoms of Australian snakes is required if we are to precisely identify the regions on these proteins that are responsible for their activities that include anticoagulant, hemolytic and myotoxic.

Another approach to deconvoluting the individual activities of each of the PLA$_2$s is by site-directed mutagenesis. Such studies require that recombinant forms of the PLA$_2$s become available. While the cloning and expression of these molecules has a high guarantee of success, the production of fully folded/ fully active protein is not assured. There is however reason to be optimistic that active recombinant versions of PLA$_2$s from Australian snake venoms can be produced using a systematic approach. For example, it has recently been shown that a recombinant version of a PLA$_2$, OS$_2$, from the Australian Taipan snake, *Oxyuranus scutellatus scutellatus* is active and correctly folded (Rouault et al., 2006).
Figure 8-1 Two orthogonal views of the surface of the putative ACII-4 dimer. Molecule A of ACII-4 is colored green and molecule B is colored cyan, with the predicted anticoagulant sites in red. The entrance to the active site is identified by the presence a balls and stick model for a polyethylene glycol molecule, which was in the crystallization buffer. The C-terminus region is highlighted in yellow. This region is predicted to be important for presynaptic, neurotoxic and myotoxic activity.

8.1 How can we modify the potency and selectivity of textilinin-1

Aprotinin (Trasylol®) has been in clinical use for over fifteen years as an anti-bleeding agent during surgery. In the last year and half, a number of studies have raised concerns that the use of this drug can lead to an increased risk of death, serious kidney damage or stroke. Thus it is timely to consider if snake venom components that are analogues of this molecule could be developed as safer alternatives. Now that the structure of rTxln-1 has been determined as the free inhibitor and in complex with microplasmin, the way is open for rational structure-based design of new inhibitors. There are several approaches that could be undertaken to develop new
plasmin inhibitors. For example, the structure of rTxln-1 inhibited microplasmin could be used in virtual screening or docking simulations with other known protease inhibitors to see if there is potential for complementarity of fit. If such molecules are identified, they can be screened against the enzyme for inhibitory activity. Another approach is to use the structure of rTxln-1 as bound to microplasmin as a template for the synthesis of peptides or peptide mimetics. Such molecules would be expected to form similar interactions as those observed for rTxln-1. While both of these approaches are valid, the question that arises is, will they be able to lead to the discovery of molecules that are highly potent, and highly specific for a given target protease? A better, strategy may be to start with rTxln-1 or a similar Kunitz-type inhibitor and, in a rational way, make changes to the amino acid sequence that will modify its inhibitory activity. Clearly, alterations to residues that make direct contact to plasmin are likely to have the greatest impact, though residues that are distal to the canonical or secondary loops could also be modified to provide more subtle variations.

The P1 residue, which protrudes into the specificity pocket of plasmin and other proteases is the most logical place to start mutagenesis studies. In Txtn-1 this residue is arginine, and in aprotinin it is lysine. An interesting comparison would therefore be to mutate this residue to lysine in Txtn-1 and investigate its inhibitory properties. This molecule would still be expected to inhibit proteases with trypsin like activity but would it bind more or less tightly to plasmin? In rTxln-1, the P1’ position is occupied by valine, forcing the histidine in plasmin to rotate out of its position in the catalytic triad. Would a change to a shorter hydrophobic residue such as glycine or an alanine alter the inhibition mechanism, leaving the histidine in its catalytically active orientation? In light of the fact that valine can have two different side-chain conformations when rTxln-1 is in complex with microplasmin, a bulkier side-chain such as leucine or isoleucine may also be accommodated in this site. Modeling studies suggest that the binding affinity to plasmin would be either unchanged or slightly enhanced, but how would this change affect the binding to other proteases? It could at least be assumed that proteases that have a less flexible active site than plasmin would become precluded from forming interactions with such an inhibitor. Threonine has a similar shape to valine, but provides an hydroxyl group which could form an hydrogen bond with Phe587O in the case of plasmin.
However, the energy gained from the formation of the hydrogen bond may well be offset by the loss of the energy due to a reduction in the hydrophobic effect. Residues in the P2’ and P3’ sites and those in the secondary binding loop that interact with plasmin are also prime candidates to consider for mutagenesis experiments.

Two crystal structures of human plasma kallikrein (pdb file 2ANW and 2ANY) are available, therefore it is possible to make some suggestions as to how to design an inhibitor that might be selective for either of these two proteases. By comparison with microplasmin, the S2’ and S3’ sites in plasma kallikrein are compact, precluding a large aromatic side-chain such as phenylalanine, tyrosine, tryptophan or histidine from binding. Thus the presence of a phenylalanine at P3’ in microplasmin could be an important arbiter in selectivity. Within the secondary binding loop, an Ile36Tyr mutation could allow the formation of a hydrogen bond to Gln738O or the side-chain of Glu687 in microplasmin, whereas such a large side-chain would make it difficult for rTxln-1 to interact with plasma kallikrein.

In terms of X-ray crystallographic studies, a structure of aprotinin bound to plasmin would be the next most interesting structure to determine. It would show us if aprotinin also inhibits by the same mechanism forcing the histidine side-chain out of its catalytically active site. My prediction is that this will not be the case and that the catalytic triad will remain intact.

It was postulated that rTxln-1 could dimerize in solution. If this were to happen its inhibitory activity might be decreased (Chapter 4). Further analysis of a dimerization of rTxln-1 should be undertaken. Studies could include, dynamic light scattering, circular dichroism, and NMR spectroscopy. Investigations at different pH values and salt concentrations should also be undertaken.

### 8.2 Further investigations with ACII-4 and other Australian snake phospholipases

The crystal structure of ACII-4 is a starting point to build a hypothesis around the different activities of three phospholipases isolated from the Australian King Brown snake. Strikingly, these molecules, although having very similar sequences but exhibit different activities. The first has a strong anticoagulant activity and low
phospholipase activity, the second has very strong phospholipase activity and the third has strong haemolytic activity. Crystal structures, along with biological assay data for each of these is therefore required if we are to rationally explain each of the different activities.

The precise target or targets for ACII-4 in human plasma are unknown. It is suggested that binding studies \textit{(i.e.} surface plasmon resonance) with factor Va, Xa, IXa, XIa and XIIa be undertaken to determine which of these are the binding partners for ACII-4. Once these have been established X-ray crystallographic studies of the complexes should be carried out to better define their binding sites.
Appendix A
Common and Latin names of Australian venomous snakes

Table A-1 Family Elapidae

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acanthophis</td>
<td><em>A. antarcticus</em></td>
<td>Common Death Adder</td>
</tr>
<tr>
<td>Austrelaps</td>
<td><em>A. superbus</em></td>
<td>Lowland Copperhead</td>
</tr>
<tr>
<td></td>
<td><em>A. ramsayi</em></td>
<td>Highland Copperhead</td>
</tr>
<tr>
<td>Cryptophis</td>
<td><em>C. nigrescens</em></td>
<td>Smalled-eyed Snake</td>
</tr>
<tr>
<td>Demansia</td>
<td><em>D. vestigiata</em></td>
<td>Black Whip Snake</td>
</tr>
<tr>
<td>Hoplocephalus</td>
<td><em>H. stephensi</em></td>
<td>Stephen’s Banded Snake</td>
</tr>
<tr>
<td>Notechis</td>
<td><em>N. scutatus</em></td>
<td>Mainland Tiger Snake</td>
</tr>
<tr>
<td></td>
<td><em>N. ater niger</em></td>
<td>Island Tiger Snake</td>
</tr>
<tr>
<td></td>
<td><em>N. ater serventi</em></td>
<td>Island Tiger Snake</td>
</tr>
<tr>
<td>Oxyuranus</td>
<td><em>O. scutellatus</em></td>
<td>Taipan</td>
</tr>
<tr>
<td></td>
<td><em>O. microlepidotus</em></td>
<td>Fierce Snake</td>
</tr>
<tr>
<td>Pseudechis</td>
<td><em>P. porphyriacus</em></td>
<td>Red-bellied Black Snake</td>
</tr>
<tr>
<td></td>
<td><em>P. australis</em></td>
<td>King Brown Snake</td>
</tr>
<tr>
<td></td>
<td><em>P. guttatus</em></td>
<td>Blue-bellied Black Snake</td>
</tr>
<tr>
<td></td>
<td><em>P. colletti</em></td>
<td>Collett’s Snake</td>
</tr>
<tr>
<td>Pseudonaja</td>
<td><em>P. textilis</em></td>
<td>Common Brown Snake</td>
</tr>
<tr>
<td></td>
<td><em>P. nuchalis</em></td>
<td>Western Brown Snake</td>
</tr>
<tr>
<td></td>
<td><em>P. affinis</em></td>
<td>Dugite</td>
</tr>
<tr>
<td></td>
<td><em>P. inframacula</em></td>
<td>Peninsula Brown Snake</td>
</tr>
<tr>
<td>Tropidechis</td>
<td><em>P. carinatus</em></td>
<td>Rough Scaled Snake</td>
</tr>
</tbody>
</table>
Appendix B
Laboratory methods

B.1 Analytical ultracentrifugation

Analytical ultracentrifugation can be used to determine the molecular weight and oligomeric state of proteins. The sample is monitored through either of two optical detection systems, interference or absorption optics. This allows the operator to observe the evolution of the sample concentration versus the axis of rotation profile as a result of the applied centrifugal field. These observations are electronically digitized and stored for further mathematical analysis. Two kinds of experiments can be performed: sedimentation velocity or sedimentation equilibrium experiments. Sedimentation velocity is a hydrodynamic technique and is sensitive to the mass and shape of the macromolecular species. In sedimentation velocity, a moving boundary is formed on application of a strong centrifugal field. A series of scans (i.e. measurements of sample concentration, c, as a function of radial distance, r) are recorded at regular intervals to determine the rate of movement of the boundary as a function of time. Sedimentation velocity experiments aim to interpret the entire time-course of sedimentation, and report on the shape and molar mass of the dissolved macromolecules, as well as their size-distribution. In contrast, sedimentation equilibrium is carried out at a lower speed for a longer time period until the solute is in equilibrium with opposing forces operating. Thus sedimentation velocity and equilibrium measurements provide complementary information, and it is often useful to apply both techniques to a given problem. Computer programs such as Sedfit (Schuck, 1998) then calculate the frictional coefficient to determine the principal shape of a molecule.

In equilibrium sedimentation, the equilibrium (Tinoco et al., 1995; Hensley, 1996) is given by

\[ c_r = c_A \exp \left( M_A (1 - \nu' \rho) \omega^2 \left( r^2 - r_m^2 \right) / 2RT \right) \]

where \( c_r \) and \( c_A \) are the concentrations of macromolecule at a radial position \( r \) and the meniscus, respectively. \( M_A \) is the molecular mass of A. \( \nu' \) is the partial specific volume of A, \( \omega \) is the angular velocity, \( \rho \) is the solvent density, and \( r \) and \( r_m \) are radial positions at an arbitrary position and at the meniscus, respectively. \( R \) is the gas constant and \( T \) is the absolute temperature. Parameter dimensions are; \( M_A \) in
kg mol\(^{-1}\), \(v'\) in m\(^3\) kg\(^{-1}\), \(\rho\) in kg m\(^{-3}\), \(\omega\) in rad s\(^{-1}\), \(r\) in m, \(R = 8.314\) kg m\(^2\) s\(^{-2}\) K\(^{-1}\) mol\(^{-1}\), T in K and c arbitrary. By choosing the speed of rotation \(\omega\), it is possible to sediment molecules of any molecular weight.

The following equation is used to determine the molecular weight from the sedimentation equilibrium experiment, which is the logarithm of Equation B.1.

\[
\ln(c_r/c_A) = M_A (1 - v'\rho) \omega^2 (r^2 - r_m^2) / 2RT \quad (B.2)
\]

The equation shows that \(\ln(c_r/c_A)\) as a function of \(r^2\) is a straight line with a slope \(\lambda\).

\[
\lambda = M_A (1 - v'\rho) \omega^2 / 2RT \quad (B.3)
\]

and the absolute molecular weight becomes

\[
M_A = \lambda 2RT / ((1 - v'\rho) \omega^2) \quad (B.4)
\]

Hence, by measuring absorbance versus radial distance and plotting \(\ln\) Abs as a function of the square of radial distance the molecular weight can be obtained if the plot is linear.
Appendix C
Crystallization of protein

C.1 Introduction
The first step in structure determination is to obtain crystals of the protein of interest. The quality of the crystal sets a limit on the maximal resolution to which data can be measured. The amount of data collected determines the accuracy of the final model.

C.2 The process of crystal growth
There are two phases to the crystallization of molecules, nucleation and growth. In the first phase, molecules must overcome an energy barrier to form a nucleus, a microcrystal. In the second phase, growth can be accomplished if the solid state is made more attractive to individual molecules than the free solution state.

For a protein to start crystallization it is necessary that it reaches a supersaturated state in solution so that the growing of a three-dimensional periodic arrangement of molecules is initiated. Formation of a crystal nucleus and the subsequent growth is shown in the Figure C-1.

The solubility curve divides the concentration space into two areas, the undersaturated and the supersaturated. On this curve each point corresponds to concentrations at which the solution is in equilibrium with the precipitating agent. The curve represents a situation either at the end of the crystal growth process from supersaturated solution or to a situation when crystal dissolution occurs in an undersaturated solution. In the area under the solubility curve, the solution is undersaturated and crystallization will never occur. Above the solubility curve is the supersaturation zone. For a given concentration of precipitating agent in this zone, the protein concentration is higher than that at the equilibrium. The supersaturated area is subdivided into three zones, depending on the kinetics to reach equilibrium and the level of supersaturation.
They are:

- Precipitation zone: Excess of protein molecules immediately separates from the solution to form amorphous aggregates.
- Nucleation zone: Excess of protein molecules aggregate in a crystalline form.
- Metastable zone: Spontaneous nucleation does not occur unless it is induced by vibration or a seed is introduced that will promote heterogeneous nucleation.

Figure C-1 Phase diagram for protein solubility.

A favorable way to grow well-ordered crystals is to start with the formation of one or more nuclei in the nucleation zone just above the metastable zone. When the crystals grow the solution moves to the metastable zone and the creation of more should stop. The existing nuclei continue to grow at a decreasing rate until equilibrium is attained.

C.3 Crystallization methods

C.3.1 Vapor diffusion

The vapor diffusion technique is by far the most used crystallization method for protein (Bergfors, 1999). A small drop of the order of 0.5 – 5 µL of the protein is mixed with about the same volume of the crystallization reagent. The word reagent, or agent, is a collective name for buffer, salt and precipitant. The drop is placed on a
cover slip and placed over a reservoir of reagent. Vapor phase equilibration can then occur.

The drop contains a lower concentration of reagents than the reservoir. To achieve equilibrium, water leaves the drop and migrates to the reservoir. As water leaves, the sample undergoes an increase in supersaturation. Both the sample and the reagent increases in concentration as water leaves the drop for the reservoir. Equilibrium is reached when the reagent concentration in the drop is approximately the same as that in the reservoir. A schematic description of the process is shown in Figure C-1.

The protein will start to concentrate from an undersaturated state, point A, and then reaches a supersaturated state, point B. The first crystals take form and the concentration of protein starts to decrease. The crystal then grows until the concentration of the protein in the drop reaches the solubility curve, point C.

C.3.2 Crystallization vessels
In the hanging drop technique the precipitant solution is placed in a container. A drop consisting of the protein and the same volume of the precipitant solution is placed on a siliconized glass cover slip. The slip is placed on the container with the drop facing the precipitant solution. The cover slip is sealed to the container using grease. The set-up is shown schematically in Figure C-2.

The sitting drop technique is similar to the hanging drop except that the sample sits on a shelf, Figure C-2.

![Figure C-2](image)

**Figure C-2** Protein crystallization vessels. Schematic representations of three frequently used apparatus; hanging drop and sitting drop.
C.4 Preparation for crystallization, screening, optimizing

There are many parameters that influence the crystallization process. Deciding upon which conditions may lead to success is difficult to predict. Protein crystallization is considered a trial-and-error experiment in which a wide range of conditions are initially screened to yield promising lead conditions.

To find suitable conditions for crystallization, screening and optimization experiments must be undertaken. Examples of initial grid screening and optimization protocols have been published (Weber, 1990; Shaw and Khimasia, 1994) or an alternative to a first attempt of screening is to use spare matrix screens from Hampton Research (Jancarik and Kim, 1991). By far the most important factor in crystallization is the purity of the protein sample. Screening is performed by varying one or several of the parameters; protein concentration, precipitate concentration, pH, temperature, method, protein. To refine the possible conditions that give crystals, screening around conditions that give microcrystals is suggested.
Appendix D
Molecular replacement and model refinement

D.1 Introduction
This appendix describes the phase problem in X-ray crystallography, molecular replacement model refinement and reliability indices.

D.2 Structure factors and the phase problem
Diffracted rays are recorded as spots on an imaging plate. The diffracted X-rays are the Fourier transforms of the molecular structure being analyzed. In the mathematical theory of crystallography the diffracted X-rays are called structure factors and are denoted by $F$. They can be derived (Drenth, 1999) by a three-dimensional Fourier transform of the electron density distribution of the molecule and as a function of the Miller indices.

\[ F(h k l) = V \int \int \int \rho(x, y, z) \exp[2\pi i(hx + ky + lz)] dx \, dy \, dz \quad (D.1) \]

Here $F(h k l)$ is the structure factor, a complex number, with an absolute value equal to the amplitude of the diffracted wave and the argument corresponding to the phase of the wave. The Miller indices are denoted $h$, $k$ and $l$. The integration is performed over all electrons in the unit cell, $\rho(x, y, z)$ describing the electron density distribution of the molecule, and $x$, $y$ and $z$ being the fractional coordinates of the unit cell. $V$ is the volume of the unit cell. The Equation D.1 illustrates that each structure factor contains contributions from the electron density of the entire unit cell.

In protein X-ray crystallography the primary goal is not to calculate the diffraction pattern but to be able to calculate the electron density $\rho$ at every position $x, y, z$ in the unit cell and hence make it possible to display the whole electron density map. This can be accomplished by calculating the inverse Fourier transform. $F(h k l)$ is the transform of $\rho(x, y, z)$, but there is an inverse Fourier transform of $F(h k l)$ that gives $\rho(x, y, z)$. According to the Laue conditions diffraction occurs only in discrete directions and the inverse Fourier integration can therefore be replaced by a summation.
It should now be possible to calculate the electron density \( \rho(x y z) \) at all positions \((x y z)\) in the unit cell if we know \( F(h k l) \). Unfortunately, the measured intensities, \( I \), carry no information about the phase angles as \( I \) is proportional to \(|F|^2\). Rewriting \( F \) in the form \( F = |F| \exp[i\alpha] \) and introducing in Equation D.2 gives

\[
\rho(x y z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F(h k l) \exp \left[ -2\pi i (hx + ky + lz) \right] \tag{D.3}
\]

This equation demonstrates the phase angle problem in crystallography. The absolute values \(|F(h k l)|\) can be derived from the recorded intensities \( I(h k l) \) but the phase angles \( \alpha(h k l) \) cannot. There are methods available to solve the phase problem.

### D.3 Solving the phase problem

The X-ray diffraction pattern from a crystal contains no information about the relative phase angles of the reflections. As seen from Equation D.3 the electron density, \( \rho(x y z) \), can therefore not be calculated due to the lack of phases. There exist today three principal techniques (Drenth, 1999) for solving the phase problem in protein X-ray crystallography; (i) the isomorphous replacement method which requires the attachment of heavy atoms to the protein, (ii) the multiple wavelength anomalous diffraction method, which depends on the presence of sufficiently strong anomalous scattering atoms in the protein structure itself and (iii) molecular replacement for which a similar structure to that of the unknown has to be available.

In this work the molecular replacement method has been used for structure determination. A useful tool in this type of work is the Patterson function, which is described in the below.
D.4 The Patterson function

The Patterson function (Patterson, 1934) is a Fourier summation of the square of the structure factor amplitudes and with no phases. The value of the three-dimensional Patterson function at a position $u,v,w$ is calculated as

$$P(u,v,w) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F(hk l)|^2 \cos[2\pi(hu + kv + lw)]$$  \hspace{1cm} (D.4)

where $u,v$ and $w$ are relative coordinates in the unit cell.

It can be shown that this is equivalent to

$$P(u,v,w) = \iint_{\text{unit cell}} \rho(x,y,z) \rho(x+u, y+v, z+w) \, dV$$  \hspace{1cm} (D.5)

This expression means that the Patterson function is the product of the electron density in two points in space related by the vector $u,v,w$ summed over the whole unit cell. The Patterson function calculated from the amplitudes of a diffraction experiment will thus have large values for points $u,v,w$ that represent vectors between positions with high electron density. This is the reason for choosing the coordinates $u,v,w$, which are coordinates in vector space, rather than $x,y,z$, which are used in real space.

D.5 Molecular replacement

In the molecular replacement method a preliminary set of protein phases are derived from a known structure, if there is sufficient structural similarity between the unknown (observed) and the known structures. The known structure serves as a first model that can subsequently be refined. This procedure is based on the observation that proteins, homologous in their amino acid sequence, have a very similar folding of their polypeptide chain. The structure factors $F(hk l)$ of the model (amplitude and phase), can be calculated using Equation D.1. The calculated phases and the observed $|F|$ values can then be applied to Equation D.3 to calculate a first electron density of the observed structure, which later can be refined. Before the phases of the model can be used for the observed structure the model has to be put in an orientation in agreement with the observed structure. The above is the essence of the macromolecular replacement technique (Rossmann and Blow, 1962). The known structure is referred to with several names; search model, probe model or just model.
The process of orienting the model in proper orientation and precise location within the unit cell is performed using the Patterson function (Drenth, 1999; Liljas, 2003). This function can be calculated from experimental data and for the model using Equations D.4 and D.5. The transformation of the model is performed in two steps. First a rotation and then a translation.

It is important to distinguish between different types of vectors in the crystal. Vectors between atoms within a molecule are called self-vectors, and vectors between different molecules in the crystal are called cross-vectors. All the self-vectors will be shorter than, or equal to, the longest dimension of the molecule, but the majority of them will be relatively short. The cross-vectors can have any length. Most of the self-vectors will be shorter than most of the cross-vectors, as long as the molecule does not have a very peculiar shape.

The set of self-vectors of a molecule will be the same independent of the packing of the molecule in the crystal. When molecules have different orientations in the crystal, the self-vectors will differ only by a rotation around the origin. When a molecule is present in different crystals, or in multiple copies in one crystal, the relative orientation of the molecules can therefore be found by comparing the self-vectors. One of the molecules is fixed and the other rotated in order to find a rotation that gives the best overlap of the Patterson functions.

In the rotation function R below, the best overlap is found by locating the maximum value of the product of a stationary Patterson function $P_{\text{obs}}(u)$ and a rotated Patterson function $P_{\text{model}}(u_r)$, integrated over a suitable volume U. Here $u_r$ denotes the coordinates of the rotated molecule and the Patterson function will be rotated in the same way.

$$R = \int_{U} P_{\text{obs}}(u) P_{\text{model}}(u_r) \, du$$  \hspace{1cm} (D.6)

The volume U is chosen so that it contains most of the self-vectors, but few of the cross-vectors. The R-function is calculated as a function of the orientation of the rotated molecule.
Once the orientation of the model is known, the model is put in the native unit cell for the translation search (Crowther and Blow, 1967) to determine the absolute position of the unknown molecule. The cross-vectors between the model and a symmetry-related molecule of the space group are calculated as a function of its position in the cell. A cross Patterson function $P_{12}$ (the calculated cross-vectors between symmetry-related molecules 1 and 2) is calculated as a function of the position of the first molecule. When the relative position is correct, the observed Patterson function, $P_{obs}$, will contain the same cross-vectors as the calculated Patterson function. Calculating the product of the two Patterson functions as a function of an intermolecular vector $t$ gives the translation function (Giacovazzo et al., 2002).

$$T(t) = \int \frac{P_{12}(u,t) P_{obs}(u)}{V} \, du$$

(D.7)

This function can be expected to show large values for the $t$ corresponding to the correct translation, giving the position of the molecule.

**D.6 Model refinement**

The experimental phases will normally have significant errors (Liljas, 2003). The ‘observed’ electron density, the $F_{obs}$ map, calculated from initial calculated phases, $\alpha_{calc}$, and the amplitudes of the observed structure factors using Equation D.3 might make the map (at least locally) difficult to interpret due to the errors. There might even be other reasons for errors in the form of small inaccuracies or even mistakes in model building. It is therefore necessary to refine the model coordinates to correct these errors. There are various methods that can be applied to correct the errors. These methods try to minimize the difference between experimental data for the crystal, for example the amplitude of reflections $F_{obs}$, and the corresponding quantities calculated from the current model $F_{calc}$. Most refinement programs also try to minimize deviations from standard geometry and to avoid collisions between the atoms. The new refined coordinates are used to calculate improved phases and electron density maps. The new maps make it possible to perform manual rebuilding of regions of the model, where the refinement program did not succeed to correct the model. The refinement is an iterative process that gradually improves the model.
The validation step in the structure determination is where parameters such as rmsd values for bond lengths and angles, Ramachandran plot and correspondence between calculated and observed amplitudes are measured and compared with other models to decide the quality of the model.

Manual refinement is performed using the $F_{\text{obs}} - F_{\text{calc}}$ electron density maps. Negative densities in the $F_{\text{obs}} - F_{\text{calc}}$ tend to reveal mistakes in the model, such as misplaced atoms or extraneous atoms that should not belong to the model, while positive densities indicate the omissions in the model. The model is manually adjusted and can be further improved by refining coordinates, B-factors (isotropic and anisotropic), and adding solvent or other relevant molecules, see below. It is also common to use $2F_{\text{obs}} - F_{\text{calc}}$ density maps for model building to reduce the bias towards the model (McRee, 1993). Improved models lead to better values of $a_{\text{calc}}$ and subsequently a better observed electron density map.

*Refinement of atomic parameters*

The atomic model of a molecule obtained from building in the electron density map will have small errors. The model can be improved by automatic refinement methods. There is a major problem as in general the observed data-to-parameter ratio is low. A ratio above 10 is generally required for accurate and independent determination of each parameter (Giacovazzo et al., 2002). This is often the case for small molecules, but for macromolecules the number of refined parameters (number of atoms in the model x 4 (x,y,z and B-factors)) might be of the same order as the number of observations (the number of reflections). The errors in the measurements will then lead to incorrect models.

*Rigid body refinement*

Rigid body refinement is a useful method often applied in early stages to improve positioning and orientation. A rigid group can be the entire molecule or domains of a molecule. Each rigid group is moved as a whole to refine its position and orientation in the asymmetric unit.
**Restrained refinement**

The problem with the unfavorable ratio of observations to parameters is overcome by adding information about the model that can be seen as extra “observations”. These observations are derived from careful studies of small molecules as well as other molecules, where the stereochemistry of the polypeptide has been well defined. The known bond lengths and angles are added to the refinement as restraints of the corresponding properties in the model. The restraints are mostly sufficient to guide the refinement to better models. Besides bond lengths and angles the restraints can include; planarity of groups (e.g. aromatic rings), non-bonded contacts (avoiding collisions of atoms) and chirality.

**Simulated annealing**

A considerable problem with the refinement is that the minimization might stop at a local minimum. Simulated annealing allows the structure to find new parts of the conformational space and escape entrapment in a local energy minimum. This is done by using molecular dynamics. The rearrangement of the atoms is made by allowing the coordinates and velocities to change according to Newton’s Law of motion. The movements of the atoms in the molecule are analyzed as a function of the potential energy of the system. By computationally heating the molecule to a very high temperature, typically 3000 to 4000K, the atom can pass energy barriers. The potential energy of the molecule can be combined with a pseudo-energy based on the similarity of $F_{\text{obs}}$ and $F_{\text{calc}}$ using a least squares criteria of these entities. When the system is “annealed” by cooling it down to normal temperatures, the model can be refined even if it has relatively large errors from the beginning. The procedure can increase the radius of convergence.

**D.7 Reliability indices**

In the course of obtaining a final model there are several causes of errors that have to be controlled. This is commonly done by analysis of indices that can unveil if the obtained results are acceptable. It starts with the measurement of diffraction intensities and their errors which need to be evaluated in a reliable manner. In molecular replacement the quality of the initial $F_{\text{obs}}$ map is given by the initial phases from the model. The improvement of the subsequent calculations of the electron
density maps may be unduly biased towards the model at hand rather than the true model due to the use of the $a_{\text{calc}}$ in the calculations. Manual intervention when fitting the model to electron density may also introduce errors due to subjectivity. Improvement in the model by adding more parameters may result in over-fitting (Kleywegt and Jones, 1997). Objective measures of the quality during the development process of a model are necessary. Commonly a number of reliability indices are used for this purpose.

The quality of X-ray diffraction data is examined by several parameters, such as completeness, intensity to noise ratio and an index called $R_{\text{sym}}$. The latter is a measure of the agreement among the independent measures of symmetry related reflections in a crystal data set.

$$R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |I_i - <I>|}{\sum_{hkl} \sum_i |I_i|}$$  \hspace{1cm} (D.8)

where $<I>$ is the mean of all the symmetry related reflections with indices $(hkl)$ and $I_i$ is the intensity of each symmetry related reflection. Symmetry related reflections should have identical intensities. If they do not, it suggests some type of measurement error and may also be caused by morphology and absorption.

A variation of $R_{\text{sym}}$ is the $R_{\text{merge}}$. In this case it is a measure of agreement among multiple measurements of the same reflections (not symmetry related), with the different measurements being in different frames of data or different data sets.

Completeness is the number of unique reflections relative to the theoretical number of unique reflections. The intensity to noise ratio is given by $<I>/<\sigma(I)>$, where $<$ is the mean value. This ratio is sometimes written as $I/\sigma$. The ratio shows the quality of the recorded measurements. These statistics can be calculated for the entire data set or for subsets such as resolution shells. Using resolutions shells makes it possible to set an upper resolution limit beyond which data is not useful. By choosing the values of the criteria a balance can be achieved between maximizing the useful data set and still keeping a high quality of the overall data. Choosing the following figures
for the highest resolution shell is considered good practice; completeness at least 50%, I/σ above 2 and \( R_{\text{sym}} \) and \( R_{\text{merge}} \) below 30%.

The agreement between the magnitudes of observed and calculated structure factors are commonly used to obtain a measure of the quality of the model structure. The measure is quantified by a residual index called \( R \)-factor

\[
R = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \tag{D.9}
\]

with summations over (hkl). It has been shown that the \( R \)-factor, or \( R_{\text{factor}} \), can reach surprisingly low values in the refinement of protein structural models that appear later to be incorrect, one reason being that the number of model parameters used is too high. To avoid this situation Brünger (1992, 1993), suggested the introduction of a free \( R \)-factor, which is unbiased by the refinement process. In this method 5-10% of the reflections are randomly set aside as a test set and not used in the refinement. The \( R \)-factor for the test set is called the free \( R \)-factor or \( R_{\text{free}} \) and is calculated using Equation D.9. If the refinement improves the model this is reflected in the concurrent decrease in \( R_{\text{factor}} \) (working set) and \( R_{\text{free}} \) (test set). If errors in the model occur, \( R_{\text{free}} \) will increase or remain high (Kleywegt and Brünger, 1996). In general an \( R \)-value \( \sim 20\% \) and an \( R_{\text{free}} \) not more than an additional 5% higher is recommended.

In addition to \( R \)-factors a linear correlation coefficient between observed and calculated structure factor amplitudes are used for a quality check of the structure. The correlation coefficient is given by

\[
C = \frac{\sum (|F_{\text{obs}}| - <|F_{\text{obs}}|>)(|F_{\text{calc}}| - <|F_{\text{calc}}|>)}{(\sum (|F_{\text{obs}}| - <|F_{\text{obs}}|>)^2 \times \sum (|F_{\text{calc}}| - <|F_{\text{calc}}|>)^2)^{1/2}} \tag{D.10}
\]

The correlation coefficient has a value -1 \( \leq C \leq 1 \). The closer \(|C|\) is to 1 the greater is the correlation between observed and calculated structure factors.
Appendix E

Computer programs

E.1 Introduction

This appendix describes computer programs that have been used in the thesis. The major part of the programs belong to the Collaborate Computational Project, No.4 (CCP4) program suite (CCP4, 1994).

E.2 AMoRe

AMoRe (Navaza, 1994) consists of three core programs, (i) orientation of the model, (ii) translation of the model, (iii) rigid-body refinement of the model (Navaza and Vernoslova, 1995; Navaza and Soludjian, 1997).

Calculating the rotation function is computationally very time-consuming. One approach to facilitating calculation times is to use the Fast Rotation Function developed by (Crowther, 1972). In this method the Patterson function is calculated in polar coordinates. A spherical volume in the Patterson function is assumed. The new treatment with spherical harmonics has the advantage that Fast Fourier Transforms (FFT) can be utilized thereby improving calculation times. However, it has a limitation in that the maximal radius of integration is dependent on the resolution of the diffraction. In AMoRe the rotation function is calculated using a variation of the Fast Rotation Function that does not have the size/resolution limitations of the original procedure.

A common problem in molecular replacement is that the correct solution to the rotation function cannot be found in the background noise. To overcome this problem all rotation function solutions are tested in translation function calculations. The parameters for orientation are then improved by rigid body refinement. The program then calculates the correlation coefficient between the calculated model data and the observed data.

The model is placed in an orthogonal P1 unit cell, about four times the size of the model molecule. The combination of integration sphere and unit cell size when
computing the Patterson function is chosen so that it contains only self-vectors and no cross-vectors between atoms. The next steps are the translation and the rigid body refinement. There are several recommendations on how to choose the size of the artificial P1 unit cell and the parameters for the computations (*e.g.* CCP4 AMoRe Manual, (Navaza, 1994)).

### E.3 EPMR

EPMR an Evolutionary Program for Molecular Replacement (Kissinger *et al.*, 1999) applies an evolutionary search algorithm to find crystallographic molecular replacement solutions. The program optimizes a target function for a search model with respect to the three rotational and the three positional parameters. The evolutionary nature of the program makes it significantly faster than the brute force approach. The target function is the correlation coefficient between the observed and calculated structure factors.

The program (Kissinger *et al.*, 2001) uses the following steps:

- Generates a set of random solutions.
- The correlation coefficient is calculated for each orientation.
- A fraction of the highest scoring orientations are kept. These are used to regenerate a complete set of new trial orientations. This is done by applying random alterations to the orientation angles and translations for each remaining solution.
- The correlation coefficients for the new population are calculated. A new population is again regenerated from the top scoring solutions, and this procedure is repeated for a number of cycles.
- A traditional conjugate gradient minimization is carried out. This is a local, rigid-body refinement of the search molecule.

A correct solution on every run due to the stochastic nature of the optimization process is not always possible. The success rate is normally high if you have a good model. It is recommended to try 10 times in the case of a difficult problem. For a single molecule in the asymmetric unit a correlation coefficient of 0.5 or more is expected for a correct solution with a very good search model. For poor search...
models the efficiency can be low. If other molecular replacement methods have failed to give a solution, it is a good idea to try an extensive run with EPMR.

E.4 PHASER

The maximum likelihood method Fisher (1912) was introduced 1992 by Bricogne (1992) as a tool for solving molecular replacement optimization.

Earlier, the least squares refinement was the predominant method, but the conditions for least squares refinement are not always fulfilled (de La Fortell and Bricogne, 1994; Pannu and Read, 1996; Tickle et al., 1998). The maximum likelihood is a more general method and includes the least squares as a special case, where the errors in the parameters are simple Gaussians, rather than more complex functions.

The maximum likelihood method is based on Bayesian statistics. A good introduction to the subject is given in (McCoy, 2004). The relationship between the probabilities of the model and the data according to Bayes’ theorem is

\[
P(\text{model}; \text{data}) = \frac{P(\text{model})}{P(\text{data})} \times P(\text{data}; \text{model})
\]  

(E.1)

where \( P \) is the probability and the semicolon “;” stands for “given the known values of”. The data is the measured data and the model is the search model to be optimized. Applying the theorem in crystallography (Drenth, 1999) the likelihood function \( L_{\text{total}} \) for the joint probability for all reflections can be written as

\[
L_{\text{total}} = \prod_{h} P [ | F_h (\text{obs}) | ; F_h (\text{calc}) ]
\]  

(E.2)

where \( h = (h,k,l) \). \( L_{\text{total}} \) is a product of the many structure factors and it is easier to compute and maximize a sum. \( L_{\text{total}} \) is therefore replaced by

\[
- \log L_{\text{total}} = - \sum_{h} \log (P [ | F_h (\text{obs}) | ; F_h (\text{calc}) ])
\]  

(E.3)

which has its minimum at the same place as \( L_{\text{total}} \) has its maximum.
The aim of the maximum likelihood is to minimize the target function $-\log L_{\text{total}}$ with respect to the parameters of the model structure.

Phaser is built around the log-likelihood function. The molecular replacement process is a two step process aimed at maximizing the log-likelihood target function. Firstly, the search model is rotated to find the orientation. Secondly, the oriented search model is translated to find its position.

The program starts with a correction of anisotropy as the likelihood target assumes that data are isotropic. Next a likelihood-enhanced fast rotation function (Storoni et al., 2004) is performed. This rotation function is an effective approximation to the full likelihood target but is several orders of magnitude faster. Translation is calculated using a likelihood-enhanced fast translation function (McCoy et al., 2005). Both the fast translation and the fast rotation functions are based on series approximations to the full likelihood targets and FFT. The full likelihood target, which is more computationally intensive, can be used for final optimization. The program checks if the solutions pack into the unit cell without overlap.

The selection of good candidates in the different runs is done using Z-score and log-likelihood gain (LLG) score. The Z-score is the number of standard deviations above the mean value and should generally be greater than five. The LLG score is the difference between the log-likelihood scores for a molecular replacement trial and the log-likelihood calculated for a Wilson distribution. It measures how much better the data can be predicted from the model than with a random distribution of the same atoms. The LLG makes it possible to compare different models against the same data set, but the LLG values for different data sets should not be compared with each other.

**E.5 ARP/wARP**

The Automated Refinement Procedure, ARP/wARP, (Lamzin and Wilson, 1993) (Lamzin and Wilson, 1997) (Morris et al., 2003) is a package of programs that are used for automated model building and refinement of protein structures. When the phase information is available for the measured structure factor amplitudes, a three-dimensional electron density map can be computed. The task is then to derive an
atomic model of the molecule. The solution process is a combination of electron
density interpretation using the concept of hybrid model, pattern recognition in the
electron density map and maximum likelihood model parameter refinement. If the
measured structure factor data has a resolution of at least 2.6 Å (ARP/wARP 6.1) and
the initial phase estimates are of reasonable quality, ARP/wARP is capable of
deriving a fairly complete and refined model.

The main chain of a protein molecule can to a high degree of accuracy be determined
by the positions of the Cα atoms (Esnouf, 1997). If a correct main chain is at hand,
the side-chain placements can be performed as the side-chains are well defined by
the Cα atoms. Therefore, the first step usually in the interpretation of a protein
electronic density map is to find the main chain of the protein. In the ARP/w/ARP
program suite the warpNtrace module is designed to perform such a task.

The basic idea of ARP/wARP is that the derived model consists only of what is
found in the electron density map. ARP/wARP starts to reproduce the given electron
density distribution by the placement of dummy atoms, usually oxygen, in regions of
significant electron density. In this concept atoms can be added or removed, which is
not possible in traditional refinement. The introduced dummy atoms are referred to
as “free atoms” (Isaacs and Agarwal, 1985).

The model is expanded from random atoms by stepwise addition of atoms in
significant electro density and at bonding distances from existing atoms. The
geometrical criteria remain the same and all the atoms in the model are of the same
type. The density threshold is gradually reduced and new atoms are positioned in the
low density areas of the map.

Atoms placed by ARP/wARP are normally within 0.5 Å of the corresponding
position in the final structure (Lamzin and Wilson, 1997) but depend on the quality
of the data. It is the task of the model-building algorithms to identify their
connectivity. When the bonds and atom types are determined the stereochemical
restraints can be established and used in restrained refinement. The set of built
protein fragments with the remaining free atoms are referred to as a hybrid model.
By iterating the interpretation of the map, the model building and refinement cycle results in an improvement of the phase estimates and give the final protein model.

The wARP part of the name of the program suite ARP/wARP refers to a concept (Perrakis et al., 1997) that improves crystallographic phases by weighted averaging of multiple refined dummy atom models as an additional step after density modification methods such as solvent flattening and averaging. The initial phase set is used to generate a number of dummy atom models which are improved by maximum likelihood refinement and the iterative model updating. The extra models are built from intermediates that gave the first model or created by random positional shifts of 0.5 Å. A weighted combination of the individual phases of the models improves the quality of the maps compared to maps for the individual models.

The default refinement procedure is the REFMAC program using maximum likelihood optimization.

The progress of the refinement is reflected in the value of a connectivity index that should increase if the tracing is successful (see ARP/wARP Manual). A value below 0.6 is not promising. A value around 0.8 indicates good progress. A value above 0.95 indicates an essentially complete tracing. The R-factor from ARP/REFMAC is not a good indicator as it varies with the ARP/REFMAC refinement cycles. Sequence coverage is defined as the ratio between the number of docked residues and the total number of traced residues. A value higher than 0.8 is considered as good convergence.

E.6 COOT

Coot is a new program package (Emsley and Cowtan, 2004) that is a stand alone portion of the CCP4’s Molecular Graphics project. Coot, abbreviation for Crystallographic Object Oriented Toolkit, is for crystallographic model building and manipulation. Coot is a completely independent system that has some features that resemble program O (Jones and Kjeldgaard, 1998).
A molecular graphics program is a useful tool when it comes to functions of the type; side-chain placement, loop, ligand and fragment fitting, structure comparison, analysis and validation.

The program functions of Coot are primarily built around two major libraries; mmdb (E. B. Krissinel *et al.*, 2004) which is a library for the handling of macromolecular coordinates, and Clipper (Cowtan, 2003) which is a library for crystallographic objects and their computation.

The generation of the electron density maps is based on the Clipper libraries. After reading a file of structure factors including phase information a map can be generated. The Coot program can read mtz, mmCIF and phs files, the output from refinement programs, and display electron density maps. To start model building a pdb-file is input and there are functions available to analyze the structure and to manipulate, extend and build a molecule. There is also a possibility to run REFMAC (Section E.7) from Coot. The user can decide to use REFMAC to refine the current coordinates after having done some interactive model building.

The rigid body refinement makes use of the Clipper library that provides functions for map gradients. After the selection of a coordinate set the map gradients at atomic centers are averaged. The next computation applies a small shift that is some fraction of the average gradient. The transformations continue until the average shift length is less than 0.0005 Å.

Coot can auto-fit rotamers using the backbone independent library of Dunbrack & Cohen (1997). The library is built from a large sample set of 850 chains. Coot picks a set of most likely rotamers, makes rigid body refinement, including backbone, and accepts the rotamer that best fits the density.

A useful feature is the possibility to regularize and refine the coordinates of the model. This can be applied to fragments which are chosen by clicking on two atoms defining a zone of interest. Coot reads from REFMAC the mmCIF dictionaries which contain ideal values and estimated standard deviations for bond lengths, angles, torsion planes and chiral centers. The target function for the optimization of the
coordinates is a sum of squared differences of the parameters. The optimization is performed using Polak-Ribiere variant of the BFGS (Brogden, Fletcher, Goldfarb, Shanno) conjugate-gradient multi-variable function minimizer to optimize the model coordinates. In another function, fitting to the map, which is called ‘Refinement’ in Coot, the map gradients are just multiplied by a scaling factor determined by the user and added to the geometric terms in the target function.

Further, Coot has functions for finding ligands and to add terminal residues. Skeletonization and $C_\alpha$ building is also available.

**E.7 REFMAC**

The REFMAC program package (Murshudov et al., 1997; Murshudov et al., 1999) is a macromolecular program that is part of the CCP4 program suite. The program can perform rigid-body, restrained or unrestrained refinement against X-ray data. The optimization is carried out by adjusting the coordinate parameters to minimize either a maximum likelihood (Section E.4) or a least squares (Glusker et al., 1994) crystallographic residual. If the user does not input experimental phases, REFMAC can run to generate phases. Anisotropic refinement is performed using TLS parameterization (Winn et al., 2001).

**E.8 MrBUMP**

MrBUMP is a new automated procedure for macromolecular structure solution by molecular replacement (Keegan and Winn, 2007). The aim of MrBUMP is to start from native structure factors and a target sequence and to deliver a positional and partly refined model suitable for further model rebuilding, model completion and/or refinement.

MrBUMP takes advantage of the increasing number of models of macromolecular structures deposited in the Protein Data Bank (Berman et al., 2000) that continues to grow fast, supporting further structure determination through the use of molecular replacement. On the highest level the process consists of three stages; (i) discovery of search model templates, (ii) construction of search models from these templates, (iii) molecular replacement itself.
In the first stage, the target sequence is used to search for related proteins in the PDB, using a simple pairwise alignment as implemented in the FASTA package (Pearson and Lipman, 1988). The user can manually add models from local PDB files if such exists. Optionally the FASTA search can be complemented by search in the SSM service (E. Krissinel and Henrick, 2004), SCOP database (Murzin et al., 1995; Lo Conte et al., 2002) and the PQS database (Henrick and Thornton, 1998).

The set of sequences of the template chains and domains are aligned using a multiple alignment program. From this multiple alignment, pairwise alignments between the target and each of the templates are extracted and are used in the model building step. The multiple alignment is also used to score the template models.

A subset of the top ranked template models is passed to the search model preparation stage of MrBUMP, which generates an actual MR search model from the template structure. The most sophisticated methods used for this are CHAINSAW (Stein, 2007) and MOLREP (Vagin and Teplyakov, 1997).

The list of search models is passed to the MR stage, which uses MOLREP or Phaser (McCoy et al., 2005) or both. If the program finds a solution, then the positioned model is passed to REFMAC for 30 cycles of refinement. A criterion based on the behaviour of the free $R$-factor is used to inform the user about a possible successful structure solution.
References


Read, R. J., James, M. N. G. (1986). Introduction to the protein inhibitors: X-ray crystallography. In G. Salvesen & Barrett (Eds.), *Proteinase Inhibitors* (pp. 301 - 336): Elsevier Science Publisher BV.


