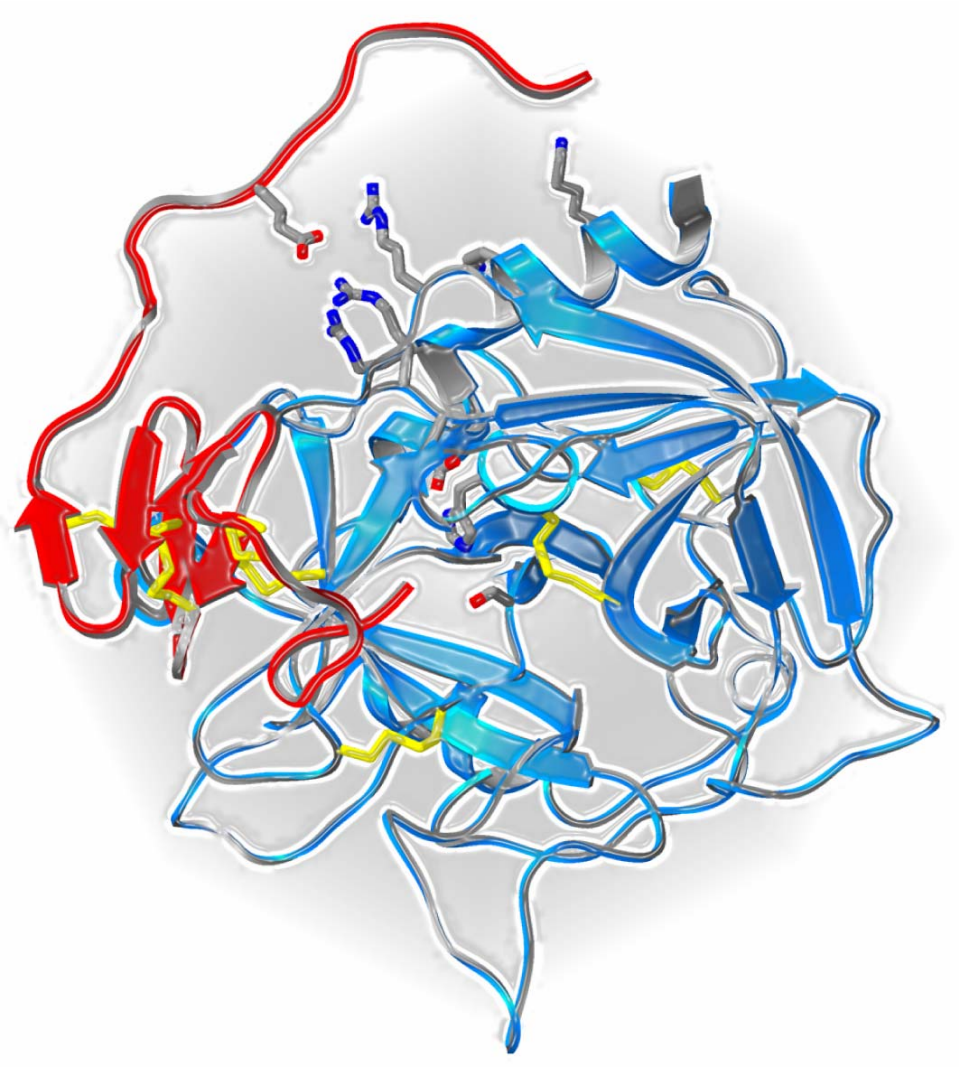
Structural and Kinetic Characterization of the Leech Derived Inhibitor Haemadin in Complex with Human α-Thrombin

Structural Analysis of the Tsetse Thrombin Inhibitor in Complex with Bovine α-Thrombin



John Lee Richardson

Max-Planck-Institut für Biochemie

Abteilung Strukturforschung

D-82152 Martinsried

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

Structural and Kinetic Characterization of the Leech Derived Inhibitor Haemadin in Complex with Human α-Thrombin

Structural Analysis of the Tsetse Thrombin Inhibitor in Complex with Bovine α-Thrombin

John Lee Richardson

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Dr. A. Bacher
Prüfer der Dissertation:	
	1. apl.-Prof. Dr. h.c. R. Huber
	2. Univ.-Prof. Dr. W. Hiller

Die Dissertation wurde am 23.01.02 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 26.04.02 angenommen.

TABLE OF CONTENTS

1. PREFACE

Haemadin is indeed a unique inhibitor of thrombin, an inhibitor that was thought for a long time to be similar to another inhibitor from another leech. That could have been the end of the story except that thrombin is an essential enzyme in the blood coagulation cascade, which makes it an attractive target for inhibitors as antithrombotic drugs against cardiovascular diseases. Coupled with the fact that until now only four X-ray structures of thrombin in complex with inhibitors from blood sucking animals have been solved, structural solution was undertaken with surprising results. Upon determining the X-ray crystal structure of the human α-thrombin-haemadin complex, a puzzle arose as to whether this inhibitor, with so much similarity to another inhibitor, was binding in the same manner as the other inhibitor or not. With help from kinetic experiments on mutant thrombins the puzzle was resolved, and much information about the nature of the inhibitor, which turned out to be unique in many ways, was also provided.

Mystery also surrounded an inhibitor from the tsetse fly called tsetse thrombin inhibitor, attempts to reproduce the potency of the native inhibitor artificially had mysteriously failed. In addition, perfectly twinned crystals, along with a confusing X-ray crystal structure in complex with bovine α-thrombin added more hurdles to discovering concrete facts about the inhibitor. Again with a little help from thrombin mutants and enzyme kinetics as well as other experiments, enough data about the inhibitor was acquired to model its binding mode from the structural data.

2. SUMMARY

Thrombin is a central enzyme in the coagulation cascade where it regulates blood coagulation. However thrombin also plays important roles in haemostasis and inflammation regulation, making it an attractive target for antithrombotic therapies. Thrombin has the fold of a typical chymotrypsin like serine proteinase, however some insertion loops and areas of positive charge on thrombin's surface discriminate thrombin to certain functions. With such a central regulatory role in coagulation and haemostasis it is no wonder that several bloodsucking animals have evolved proteinacious inhibitors towards thrombin to prevent coagulation whilst feeding. However despite 22 types of natural proteinacious anticoagulants being known the exact molecular mode of inhibition of them is only known for four types to date.

The thrombin inhibitor haemadin, a 57 amino acid peptide from the Indian land living leech *Haemadipsa sylvestris* was, due to sequence homology with another leech inhibitor hirudin from the European leech *Hirudo medicinalis* assumed to have the same mode of action as hirudin. However the 3.1 Å X-ray crystal structure of the human α-thrombin-haemadin complex was found to contain evidence that the inhibitor was significantly different. Despite haemadin having the same active site binding region with the N-terminus, forming a parallel β-strand with residues Ser 214 to Gly 216, and fold as hirudin, the location of binding of the C-terminal tail of the inhibitor was exosite II and not exosite I. This constituted one of the most dramatic structural and functional rearrangements observed in inhibitors from organisms so closely related phylogenetically.

Further slow and tight binding kinetic analysis of haemadin with thrombin mutants confirmed this proposal, and the inhibitor was thoroughly characterised kinetically with thrombin and 10 thrombin mutants. Data analysis allowed both the ionic and non-ionic binding contributions to be ascertained, with ionic interactions contributing -17 kJ per mole to the Gibbs free energy of binding, this being the equivalent of up to six salt bridges. These salt bridges make up 20% of the total binding energy at zero ionic strength, and have been attributed to the C-terminal tail alone.

In addition, the contributions of the N-terminal and C-terminal regions of haemadin were ascertained by using derivatives of both haemadin and thrombin. Limited proteolysis using formic acid produced haemadin cleaved between residues 40 and 41 removing the majority of the C-terminal tail. This truncated haemadin displayed a 20,000-fold reduced affinity for thrombin, and was no longer a tight binding inhibitor. A form of thrombin in which the active site serine was blocked by diisopropyl fluorophosphate bound to haemadin, but with a 72,000-fold reduced affinity, indicating that the N-terminus is more important than the C-terminus for strong binding.

The tsetse thrombin inhibitor is a potent and specific low molecular mass (3,530 Da) anticoagulant 32 amino acid peptide from the salivary gland extracts of *Glossina morsitans morsitans*, the tsetse fly, a carrier of sleeping sickness. Lacking homology to any previously described family of molecules, the tsetse thrombin inhibitor appeared to represent a unique class of naturally occurring proteinase inhibitors. This was confirmed when the chemically synthesised peptide was solved in complex with bovine thrombin to 2.9 Å. Perfect hemihedral twinning hindered structural solution until detwinning using the method of Redinbo and Yeates was successfully carried out on the data. The X-ray structure revealed a completely new mode of binding to the active site, which was subsequently confirmed by kinetic experiments using thrombin and a thrombin mutant. Binding to exosite I of thrombin was also confirmed by the structure and kinetic experiments. Gel filtration experiments revealed that the X-ray crystal structure contained artefacts, and the binding of this synthetic tsetse thrombin inhibitor was modelled as to its possible structure as a 1:1 complex with thrombin.

3. NOMENCLATURE

3.1. Abbreviations

Å	1×10^{-10} m
AMC	aminomethyl coumarin
APP	amyloid precursor protein
ATU	anti thrombin units
BPTI	basic pancreatic trypsin inhibitor
Da	dalton
DIPF	diisopropyl fluorophosphate
DIP-Thrombin	diisopropylphosphoryl-thrombin
DNA	deoxyribose nucleic acid
ddH_2O	doubly distilled water
Chromazyme-TH	tosyl-Gly-Pro-Arg- para-Nitroanilide
EDTA	ethylene diamine tetraaceticacid
EGF	epidermal growth factor
FPLC	fast performance liquid chromatography
Fmoc	fluorene methoxy carboxyl
fw	fresh water
HPLC	high performance liquid chromatography
rHPLC	reverse phase high performance liquid chromatography
HM2 (1-41)	hirudin variant 2 from *Hirudinaria manillensis* residues 1 to 41
haemadin (1-40)	haemadin residues 1 to 40
hirudin (1-43)	hirudin from *Hirudo medicinalis* residues 1 to 43
IL X	interleukin X
IPTG	isopropyl thiogalactoside
kDa	kilo Dalton
MAD	multiple wavelength anomalous diffraction
MBP	maltose binding protein
MES	morpholine ethane sulphonate
MIR	multiple isomorphous replacement
Mw	molecular weight

mar	marine
NMR	nuclear magnetic resonance
NO	nitrous oxide
n.k.	not known
PAF	platelet activating factor
PAGE	Polyacrylamide-gelelectrophoresis
PARs	proteinase-activated receptors
PEG	poly(ethylene glycol)
pNA	para-Nitroanilide
RCSB	Research Collaboratory for Structural Bioinformatics
rTTI	recombinant tsetse inhibitor
S-2238	D-Phe-pipecolyl-Arg-para-Nitroanilide
sTTI	synthesised tsetse thrombin inhibitor
TAFI	thrombin-activatable fibrinolysis inhibitor
TFA	trifluoro acetic acid
TME456	thrombomodulin EGF domains 4, 5 and 6
TTI	tsetse thrombin inhibitor
Tos-GPR-AMC	tosyl-Gly-Pro-Arg-aminomethyl coumarin
Tris	tris(hydroxymethyl)aminomethane
× g	times gravity
v/v	volume per volume (100 ml)
w/v	weight per volume (100 ml)

3.2. Codes for amino acids

A	Ala	Alanine	M	Met	Methionine
B	Asx	Asn or Asp	N	Asn	Asparagine
C	Cys	Cysteine	P	Pro	Proline
D	Asp	Aspartic acid	Q	Gln	Glutamine
E	Glu	Glutamic acid	R	Arg	Arginine
F	Phe	Phenylalanine	S	Ser	Serine
G	Gly	Glycine	T	Thr	Threonine
H	His	Histidine	V	Val	Valine
I	Ile	Isoleucine	W	Trp	Tryptophan
K	Lys	Lysine	Y	Tyr	Tyrosine
L	Leu	Leucine	Z	Glx	Gln or Glu

4. INTRODUCTION

4.1 The serine proteinase thrombin

Thrombin is 33.7 kDa protein that is present at a low concentration in human blood, and like the pancreatic serine proteinases it is synthesised as a zymogen, which in thrombin's case is called prothrombin, which is the form most present in blood at a concentration of 100 μg/ml. Prothrombin is a 72 kDa glycosylated protein consisting of 579 residues (human) and is synthesised in the liver, where it contains both a signal peptide and a pro peptide, which are both cleaved off before the prothrombin reaches the blood. In the blood prothrombin consists of a gla domain (which contain ten copies of the unusual amino acid gamma carboxy glutamate) and two kringle domains as well as a catalytic domain. Two cleavages, first at an arginine (271) threonine bond releasing a 32 kDa fragment, and then at an arginine (320) leucine bond yield active thrombin. Activation is caused by a rearrangement of the two chains of thrombin (alpha chain: 36 residues and beta chain: 259 residues, held together by a disulphide bond), and this is caused by the formation of an ion pair between the amino group of leucine 15 and aspartate 194 (chymotrypsinogen numbering) (Bode *et al*., 1992) (Bode and Huber, 1976). Thrombin is evolutionary related to the family of serine proteinases by fold and the presence of the catalytic triad of residues His 57, Asp 102 and Ser 195 which are present in all serine proteinases. Thrombin plays a central role in thrombosis and haemostasis (Fenton, 1988) it possesses both enzymatic and hormonal functions as well as having both pro and anti coagulant functions. After fibrinogen, prothrombin is the second most abundant soluble protein in the blood. As thrombin its major task is to cleave fibrinogen which it cleaves at four arginine-glycine bonds to release 4 fibrinopeptides, two being 18 residues each, and two being 20 residues each to leave a fibrin monomer. This fibrin monomer is much less soluble than fibrinogen, and the monomers spontaneously associate to form a soft fibrin clot. Fibrin is subsequently strengthened by crosslinking through specific lysine and glutamine chains via factor XIII, a transglutaminase generated by thrombin, which cross-links the fibrils, thus stabilising the clot (Dahlbäck, 2000; Bailey *et al*., 1951).

4.2. The coagulation cascade

4.2.1. Introduction

Thrombin, as well as catalysing the conversion of fibrinogen to fibrin to form a clot, also regulates the coagulation cascade by providing stimulatory and inhibitory feedback.

Essentially the blood coagulation system is a cascade of proteinases which cleave zymogens to form active enzymes (see figure 1). This in turn acts an amplification system that culminates in the formation of more thrombin. The centrality of thrombin to this system makes it an extremely powerful enzyme in the coagulation cascade, especially as a procoagulant and therefore makes it the most suitable enzyme to be targeted for both anticoagulant drug therapy (Fenton *et al*., 1998) and by various blood feeding animals (Markwart, 1994).

After initiation of the coagulation cascade and thrombin generation is achieved, the cascade is maintained by feedback effects of thrombin. Thrombin also actives an array of other factors in the coagulation cascade (Davie *et al*., 1991). This stimulatory feedback is achieved by the activation of factors V, VIII and XI (Narayanan, 1999). Factor V, when activated, associates with activated factor X on a membrane surface, and cleaves prothrombin to thrombin. Factor VIII, when activated, associates with activated factor IX and activates factor X. Factor XI, when activated, activates factor IX.

4.2.2. The coagulation cascade and blood clotting

In a tissue damaging event if platelets fail to close the wound, a blood clot must be formed, however there are two semi-independent ways this is achieved, the intrinsic pathway and the extrinsic pathway (See figure 1). The intrinsic pathway is so named as it only involves blood components, whereas the extrinsic pathway is triggered by extravascular tissue damage. These separate pathways both culminate in the activation of prothrombin to thrombin.

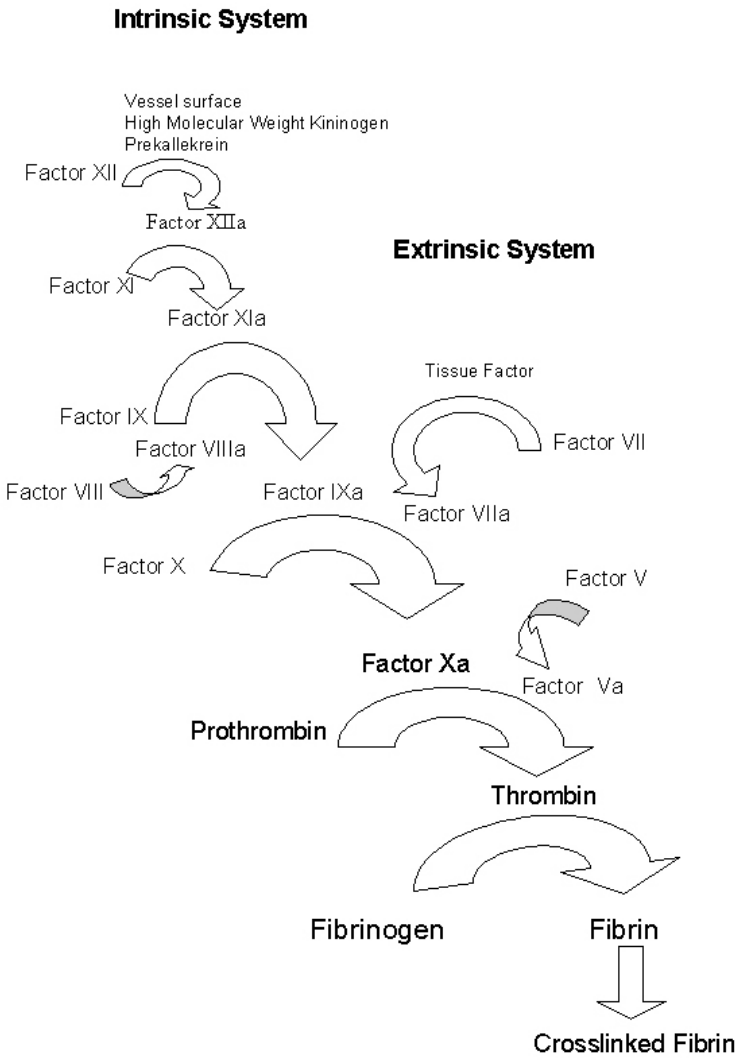


Figure 1: The Coagulation cascade. Glycoprotein components of the intrinsic pathway include factors XII, XI, IX, VIII, X and V, prothrombin and fibrinogen. Glycoprotein components of the extrinsic pathway, initiated by the action of tissue factor include factors VII, X and V, prothrombin and fibrinogen.

4.2.2.1. The extrinsic pathway

The extrinsic pathway is much quicker than the intrinsic pathway and takes about 15 seconds to start a clotting event. The damaging of tissue releases tissue thromboplastin (factor III), which along with calcium ions and factor VIII plus tissue phospholipids and enzymes from the damaged tissue, activate Factor X. Factor Xa is where both the pathways converge.

4.2.2.2. The intrinsic pathway

The intrinsic pathway is slower than the extrinsic pathway and takes between 2 to 6 minutes to get started. The intrinsic pathway is a genuine cascade which starts with factor XII associated with platelet phospholipids activating factor XI, which along with Ca^{2+} activates factor IX, then factor IX combines with factor VIII, Ca^{2+} and platelet phospholipids to activate factor X.
The endothelial release of factor XII causes platelet aggregation as well as changes in platelet structure. The platelets release lipids, such as prostacyclins, which initiate the clotting cascade, and this leads to factor X activation, exactly as in the extrinsic pathway.

4.2.2.3. The common pathway

Both pathways converge at the activation of Factor X. Factor Xa along with factor Va, on the phospholipid surface of the cell membrane of the damaged tissue, along with lipids released from the platelets all form the prothrombin activation complex. Also required are several other factors; factor IV (which is Ca^{2+} ions), factor V, and factor VII.

4.3 Thrombin's other roles in haemostasis

4.3.1 Thrombin outside the coagulation cascade

As seen above, the coagulation cascade is upregulated by thrombin, and a very large amount of fibrin is produced in this way from the initial stimulus. However thrombin fulfils many other roles in haemostasis. In a complementary role to fibrin clot formation, thrombin also prevents the lysis of the clot (Broze, 1996). Clot lysis is performed by the plasma carboxypeptidase enzyme "thrombin activatable fibrinolysis inhibitor" (TAFI), which is an inactive proenzyme in the blood, and is activated by thrombin once bound to the endothelial receptor thrombomodulin. TAFI inhibits fibrinolysis by cleaving the carboxy-terminal lysine residues on the fibrin polymers (Tilburg *et al*., 2000). These residues play an important role in assembling the fibrinolytic system components and their removal inhibits fibrinolysis.

Conversely thrombin also regulates its own production by it participation in an inhibitory feedback loops. This is activated by the binding of thrombin to thrombomodulin, a vascular endothelial cell protein. A consequence of this is the activation of protein C, which in turn causes the inactivation of coagulation factors V and VIII, which results in a down-regulation of thrombin generation. This leads to activation of protein C (causing inactivation of coagulation factors V and VIII and thus down-regulation of thrombin generation), and inhibition of thrombin's ability to form fibrin and activate factor XIII, platelets and coagulation feedback stimulatory proteins (Esmon 2000). This protein C mediated form of thrombin inhibition is extremely potent, and necessitates its confinement to the injury site. This is achieved by the expression of thrombomodulin only on damaged endothelial cell walls, where it is part of the complex with protein C, thus ensuring its localisation at the site of injury.

In addition to its above mentioned role in inhibiting fibrinolysis, thrombin participates indirectly in the stimulation of this process. Thrombin acts as a neutrophil chemoattractant, which play a role in fibrin clot degradation (Sonne, 1988) and additionally thrombin causes the endothelial cells to release plasminogen activators which lead to plasmin generation. Plasmin then initiates fibrinolysis and inactivates prothrombin (Fenton *et al*., 1998).

This role that thrombin plays in the enhancement and inhibition of fibrinolysis seems anomalous, but it exemplifies the high degree of regulation which is exercised in maintaining haemostasis. Haemostasis has to be extremely well controlled to prevent the both excessive haemorrhage events and excessive blood coagulation, both of which would prove fatal and as mentioned previously thrombin plays a central part in this regulation.

4.3.2. Thrombin as a platelet activator

As well as being involved in coagulation, thrombin also plays a vital role in primary haemostasis, as a platelet activator. Thrombin-induced platelet activation is a critical factor in haemostasis (Hung *et al*., 1992). Platelet activation by thrombin, as with any platelet agonist, results in intra-platelet events such as the activation of phospholipase C, inhibition of adenylate cyclase, and mobilisation of calcium, culminating in platelet aggregation (Vu *et al*., 1991; Hayes *et al*., 1994). The initiating factor is thrombin's cleavage of a receptor on the surface of the platelet the proteinase-activated G-protein-coupled receptor (PARs), which subsequently activate various secondary messenger systems, of which the inositol triphosphate/calcium system is probably the most important. This thrombin interaction with the platelet receptor has been extensively studied. The receptor, a member of the seven-transmembrane domain family, has been cloned and expressed (Hung *et al*., 1992). These interactions are interesting because the reaction kinetics suggest that it is not a simple ligand-receptor interaction, but an enzyme-substrate interaction where thrombin enzymatically cleaves the receptor/substrate (Hayes *et al*., 1994). The receptors amino terminal extracellular domain contains a cleavage site for thrombin, and has a structure similar to the anticoagulant hirudin and is consequently able to bind thrombin (Liu *et al*., 1994). After thrombin cleavage of this site between Arg 41 and Ser 42 a new amino terminal peptide is exposed, which is a ligand for the receptor itself and is termed a "tethered ligand" (Liu *et al*., 1994).

Thrombin participation in platelet activation is not an isolated event, but is closely tied in with thrombin's other roles in coagulation, particularly fibrin formation. With platelet aggregation and coagulation both being necessary during a bleeding event, it

is important that the molecule central to the control of one system is also important in the regulation of the other.

4.3.3. Thrombin's role in inflammation

Thrombin acts as a powerful chemoattractant for neutrophils (Esmon, 2000) and monocytes (Becker *et al*., 1998), which induces them to follow a chemical gradient of thrombin concentration which is where the trauma has occurred. Once there, the neutrophils and monocytes perform phagocytic roles on any bacteria present.

Thrombin also stimulates the production of the cell-anchoring protein P-selectin from Weibel-Paladi bodies in the endothelial cells, leading to its expression on their membranes (Esmon, 2000). P-selectin plays an important role in leucocyte "rolling", where leucocytes loosely bound to the vessel wall begin to slow down, and roll along the endothelium, eventually stopping at their required destination. Additionally thrombin causes endothelial cells to produce platelet activating factor (PAF) which is a powerful neutrophil activator especially to P-selectin bound neutrophils.

Thrombin has an additional role in the induction of other pro-inflammatory and pro-coagulant substances from endothelial cells, these include von Willebrand factor, growth factors and cytokines, and also induces changes in endothelial cells including their shape and increases their permeability (Coughlin, 1999).

Thrombin's action on monocytes induces the production of proinflammatory cytokines such as interleukin 6 (IL6) and IL8, and also acts on endothelial cells to produce other inflammatory cytokines (Cate, 2000). The production of these cytokines, initiates an inflammation response, and has been shown to induce thrombin production by mononuclear cells, and this also leads a coagulation response (Cate, 2000).

Once again thrombin acts as a link between the vital processes of coagulation and inflammation which are both necessary at the site of the trauma.

4.3.4. Thrombin's other functions

Thrombin affects many cell types and it has been suggested recently that thrombin contributes to nervous development, and has a role in the pathophysiology of Alzheimer's disease (Turgeon and Houenou, 1997). Alzheimer's disease pathology consists of vascular and cerebral plaques which are largely composed of amyloid beta protein (A beta) protein. Thrombin causes the precursor of this protein; amyloid precursor protein (APP) to be secreted from endothelial cells and therefore is implicated in Alzheimer's disease pathology (Ciallella *et al*., 1999).

Cell proliferation in a number of cell types is caused by thrombin these include smooth muscle cells and macrophages. Thrombin in this role is implicated in the disease process of atherosclerosis and the subsequent formation of the atherosclerotic plaque which consists of smooth and muscle cells macrophages (Becker *et al* 1998).

4.3.5. Conclusions

As can be seen from it above functions thrombin has an extremely wide range of biological roles, and is vital in the regulation of many processes particularly as a pro-coagulant and pro-inflammatory mediator.

4.4. Structure of thrombin

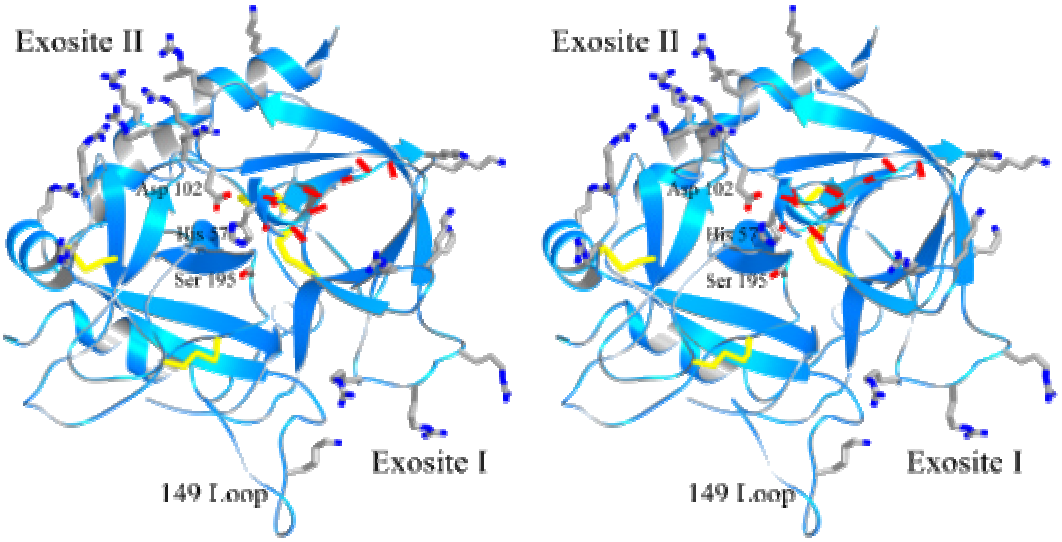


Figure 2 Ribbon representation of thrombin in the 'standard orientation'. Standard orientation is where the active site cleft, with the catalytic residues at its base, runs from left to right with a substrate molecule binding from the carboxy terminal to the amino terminal (Bode *et al.*, 1992). Labelled in the diagram are the regions of thrombin described in chapter 4.4.2., with the exception of the 60 loop which is portrayed with its carbonyl oxygens as red bars. The side chains of the three catalytic residues are shown specifically, as well as the side chains of residues which make up exosites I and II.

4.4.1. Introduction

Thrombin is a trypsin like serine proteinase of approximate dimensions $45 \times 45 \times 50$ $Å^3$, with a deep canyon containing the active site (Stubbs *et al.*, 1993). Figure 2 shows the major interaction sites of thrombin, with thrombin in the 'standard orientation'**.**

The thrombin molecule (human) consists of the disulphide linked 36 residue A chain and a 259 residue B chain. The A chain forms an integral structural part of the thrombin molecule, but is far removed from the active site cleft (Bode *et al.*, 1989; Bode and Huber, 1992).

As with other members of the serine proteinase family thrombin is divided into two six stranded barrel domains. The catalytic triad consisting of Ser 195, His 57 and Asp 102 (chymotrypsin numbering) is located at the junction of the two barrels.

Thrombin contains several insertions compared to trypsin which surround the active site making it deeper and less accessible to macromolecular substrates, this being the main cause of thrombin's specificity.

4.4.2. Features of thrombin

4.4.2.1 The 60 insertion Loop

The active site is closed on the top face by the highly conserved Tyr60A to Trp60D insertion loop. This loop is rigid in practically all recorded thrombin structures and deviates by less than 2 Å in all the determined structures (the only exception being the (E192Q) thrombin-BPTI structure (van de Locht *et al.*, 1997)). The small degree of movement correlates with the binding of various molecules and this loop effectively forms a hydrophobic lid to the active site. This loop as mentioned previously, restricts the approach of most protein substrates, and allows thrombin to be more discriminatory in its selection of substrates. This hypothesis has been realised with thrombin mutants lacking this loop (LeBonniec *et al.*, 1993).

4.4.2.2. The 149 insertion Loop

At the bottom of the molecule, a five residue insertion near Trp 148 is opposite to the 60 loop. Despite being well defined in many thrombin structures, it is often found in different conformations, contrary to the rigidity of the 60 loop, and is able to move about the nine residues from Glu 146 to Gly 149E. Although devoid of an absolute structure this loop play important roles in the binding of hirudin in PPACK inhibited thrombin (Banner, 2000) and deletion of the residues Glu 146 to Trp 148 drastically alters thrombin's ability to form interactions with serpins and reduces clotting activity (LeBonniec *et al.* 1992).

4.4.2.3. Active site

The active site catalytic triad, His 57, Asp 102 and Ser 195 displays a geometry virtually indistinguishable from other members of the trypsin like serine proteinase family. The active site takes shape only upon cleavage of the A and B chains of thrombin where the free amino group of Ile 16 forms a salt bridge with Asp 194, explaining the inactivity of prothrombin towards substrates. Similar to trypsin thrombin cleaves after basic residues at the P1 position, but prefers arginine, and has a narrower substrate specificity. The S1-S3 pocket are occupied in an antiparallel fashion by Arg 3, Pro 2 and D-Phe 1 by the covalently binding PPACK substrate (See figure 3) (Bode *et al.*, 1992). The S1 pocket is slightly less polar than trypsin, and is well suited to binding arginine residues which can form a salt bridge with Asp 189 at the base of the pocket.

The S2 pocket is a hydrophobic pocket formed by His 57, Leu 99, Tyr 60A and Trp 60D, the last two residues forming the "lid" of the pocket. It is well suited to binding medium sized non polar residues which are frequently present in the P2 position of thrombin substrates.

The S2 pocket is close to another hydrophobic pocket containing the residues Arg 97, Glu 97A, Asn 98, Leu 99, Ile 174 and Trp 215. This pocket is particularly well suited for the binding of aromatic groups, for example D-Phe 1 of PPACK and Tyr 3 of hirudin, as well as many small molecular inhibitors. This site is referred to as the "aryl binding site" or S4 pocket, to discriminate it from the S2 pocket and together these form a single common hydrophobic surface previously identified as the "apolar binding site" (Berliner and Shen, 1977). The P3 to S3 interaction for substrates is similar to that of other serine proteinases, with an anti parallel beta sheet being formed between Gly 216 and the P3 residue. The S1' site of thrombin is of reduced size due to the interdiction of Lys 60F into it and restrict this pocket to small residues such as Gly, Ala or Ser which explains their frequent occurrence in the P1' position in thrombin substrates.

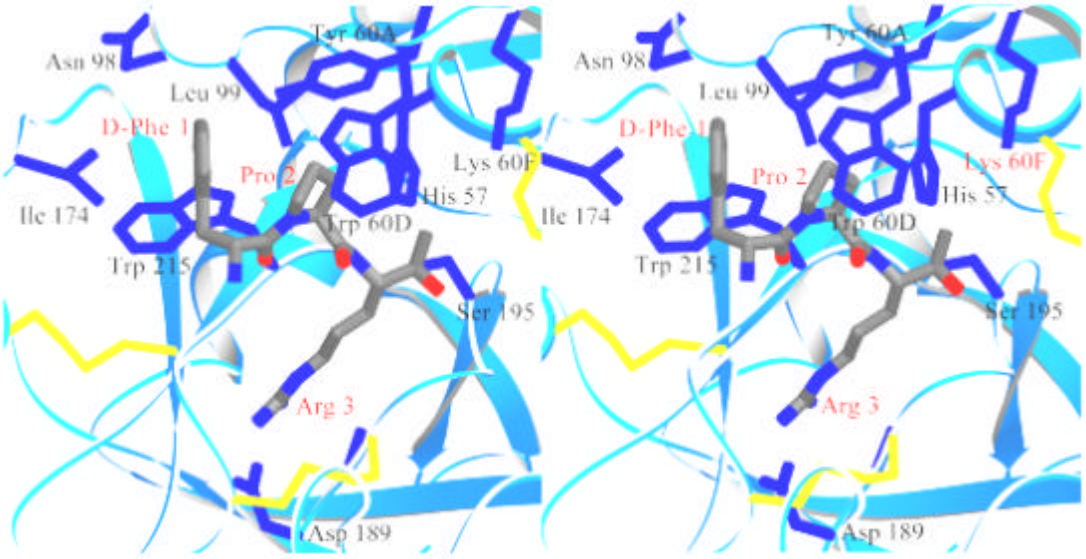


Figure 3: Stereo ribbon view of the active site of thrombin. Thrombin is shown in blue, PPACK in grey. Shown are the side chains of thrombin (Dark blue) ascribed to certain subsites in the active site. For clarity the covalent bonds between His 57 and Ser 195 are not shown, along with the majority of the side chains of thrombin, as well as all water molecules.

4.4.2.4. Fibrinogen-binding site (exosite I)

This site consists of residues Arg 35, Lys 36, Arg 73, Arg 75, Arg 77A, Lys 81, Lys 109, Lys 110 and Lys 149E, and is centred around the Lys70 to Glu 80 loop, which is stabilised by a number of buried salt bridges similar to the calcium binding loop of trypsin. This site is highly positively charged and the charge is only partially compensated by the residues Glu 77 and Glu 80. This exosite is mainly involved in fibrinogen and thrombin receptor binding and cleavage (Hogg and Jackson, 1989), and is also used by thrombomodulin (Suzuki *et al.*, 1990), heparin cofactor II (Myles *et al.*, 1998), as well as hirudin (Rydel *et al.*, 1990) and other anticoagulant proteins for binding. Proteins that bind to this site show little or no conventional structure homology except that they possess several negatively charged aspartate or glutamate residues.

4.4.2.5. Heparin-binding site (exosite II)

At the other end of the molecule is the heparin binding site, which has been confirmed by mutagenesis data (Sheehan and Sadler, 1994) to consist of residues Arg 93, Arg 233, Lys 236 and Lys 240 and may include Arg 97, Arg 101, Arg 126, Arg 165, Arg 173 and Arg 175. This positively charged region is primary used by heparin (and other glycosamino glycans) to bind to thrombin as well as some thrombin inhibitors. In the case of heparin this region is used to form a bridge when binding the natural blood born inhibitor serpin antithrombin III (Olson and Björk, 1992). In prothrombin the heparin binding site has the F2 fragment (the second Kringle domain) bound to it which prevents thrombin being inhibited in the early stages of coagulation (Martin *et al.*, 1997). This exosite also plays some role in the binding of thrombomodulin, whose chondroitin sulphate moiety provides a second low affinity binding site on thrombin (Koyama *et al.*, 1991).

4.5. Natural inhibitors of thrombin from blood sucking animals

4.5.1. Introduction

Haematophagous animals have a vested interest in preventing their host animals blood from clotting both before and after feeding. For this reason blood sucking animals possess an arsenal of anticoagulants of which some are directed against thrombin. Other strategies these animals have developed include, inhibition of prothrombin activation, inhibition of factor Xa, inhibition of platelet aggregation, fibrinolytic activity, depolymerisation of thrombin, glycoprotein IIa/IIIb antagonism, plasminogen activation and even release of NO to cause vasodilation and inhibit platelet aggregation (Markwart, 1994; Weichsel *et al.*, 1998). As can bee seen from table 1 these proteins number 22, discounting the many hirudin variants. The species concerned are mainly flies, mosquitoes, bugs, ticks, leeches and a snake.

The vast majority of these animals can be bred or procured in large amounts in order to identify, purify and characterise the minute amounts of the agents involved. It is this factor that seems to be the limiting factor that prevents an even greater number of these agents being identified.

Table 1: Known natural thrombin inhibitors with their properties and dates of discovery

Species	Common name	Inhibitor	MW	Ki	Reference
Ixodes ricinus	Tick	Ixin	n.k.	250 ATU/mg	Hoffmann *et al*., 1991
Triatoma infestans	Assasin bug	Triatomin	n.k.	100 ATU/mg	Markwardt & Schultz, 1960
Eutriatoma maculata	Assasin bug	Maculatin	100,000-200,000 KDa	n.k.	Hellmann & Hawkins, 1966
Rhodnius prolixus	Assasin bug	Prolixin Rhodniin	n.k. 12.4 kDa	100 ATU/mg 200fM	Hellmann & Hawkins, 1965 Friedrich *et al*., 1993
Tabanus bovinus	Horse fly	Tabanin	n.k.	200 ATU/mg	Markwardt & Leberecht, 1959
Hirudo medicinalus	Leech	Hirudins (×11)	6.97 kDa	21fM	Markwardt *et al*., 1967
Haemadipsa sylvestris	Leech	Haemadin	6.26 kDa	100 fM	Strube *et al*., 1993
Hirudo Nipponia	leech	Leech granulin Hirujin(A+B)	6 kDa 5.35 kDa	n.k. n.k.	Hong *et al*., 1999 Hong
Theromyzon tessulatum	Leech	Theromin	14.49 kDa	12 fM	Salzet *et al*., 2000
Hirudinaria manellensis	Buffalo leech	Bufrudin Hirudin P6 Hirudin P18 HM1 HM2	6.58 kDa 7.42 kDa 7.20 kDa 6.73 kDa 6.80 kDa	14,000 ATU/mg n.k. n.k. n.k. 780 fM	Electricwala *et al*., 1991 Steiner *et al*., 1990 Scacheri *et al*., 1993
Ornithodorus savignyi	Tick	Savignin	12.43 kDa	1.39 pM	Nienaber *et al*., 1999
Ornithodoros moubata	Soft tick	Ornithodrin	12.7 kDa	1 pM	van deLocht *et al*., 1996
Boophilus microplus	Cattle tick	Boophilin	13.9 kDa	2 nM	Peireira, 1999
Glossina morsitans	Tsetse fly	TTI	3.5 kDa	5 fM	Cappello *et al*., 1996
Anopheles albimanus	Mosquito	Anophelin	6.34 kDa	5.87 pM	Valenzuela *et al*., 1999
Dipetalogaster maximus	Bug	Dipetalogastin	11.8 kDa	125 fM 9,300 ATU/mg	Lange *et al*., 1999
Triatoma pallidipennis	Triatomine bug	Triabin	16 kDa	3 pM	Noeske-Jungblut *et al*., 1995
Bothrops jararaca	Snake	Bothrojaracin	27 kDa	600 pM	Zingali *et al*., 1993

4.5.2. Structural details of natural thrombin inhibitors

Despite the large number of natural thrombin inhibitors yet discovered and the huge number of X-ray structures of thrombin in complex with small molecule inhibitors, to date only four classes of inhibitor crystal structures exist of these agents in complex with thrombin. In addition a number of NMR structures exist of hirudin alone. Of the structures solved the structures of hirudin, rhodniin**,** ornithodorin and triabin in complex with either human or bovine thrombin are shown below in figure 4. As can be seen all these inhibitors show a unique mode of binding to both the active site of thrombin and exosite I, except triabin which binds only to exosite I.

Hirudin consists of a globular N-terminal domain and an extended C-terminal peptidic tail. The first six N-terminal amino acids bind to the active site in a non-canonical fashion, whereas the C-terminal tail binds to exosite I of thrombin mainly via electrostatic interactions. The last five C-terminal residues form a helical loop that interacts with the proteinase mainly via hydrophobic contacts.

Rhodniin's two Kazal like domains bind to different sites of thrombin. The amino-terminal domain binds in a substrate like manner to the active site the carboxy terminal domain with its distorted "active site" loop cannot form the canonical conformation and binds to exosite I via extensive electrostatic interactions. Ornithodorin too consists of two domains, each domain resembling the Kunitz domain of basic pancreatic trypsin inhibitor (BPTI), with a distorted fold. None of the "reactive site loops" touches thrombin, instead the N-terminal portion binds to thrombin in a non-canonical fashion resembling hirudin, and the C-terminal domain binds at exosite I mainly through electrostatic forces.

Triabin is a compact one-domain molecule consisting essentially of an eight-stranded ß-barrel. Triabin interacts with thrombin exclusively via exosite I and surprisingly, most of the interface interactions are hydrophobic with the exception of one salt bridge. This exosite I blocking activity is to prevent fibrinogen binding and subsequently being cleaved thus preventing blood coagulation.

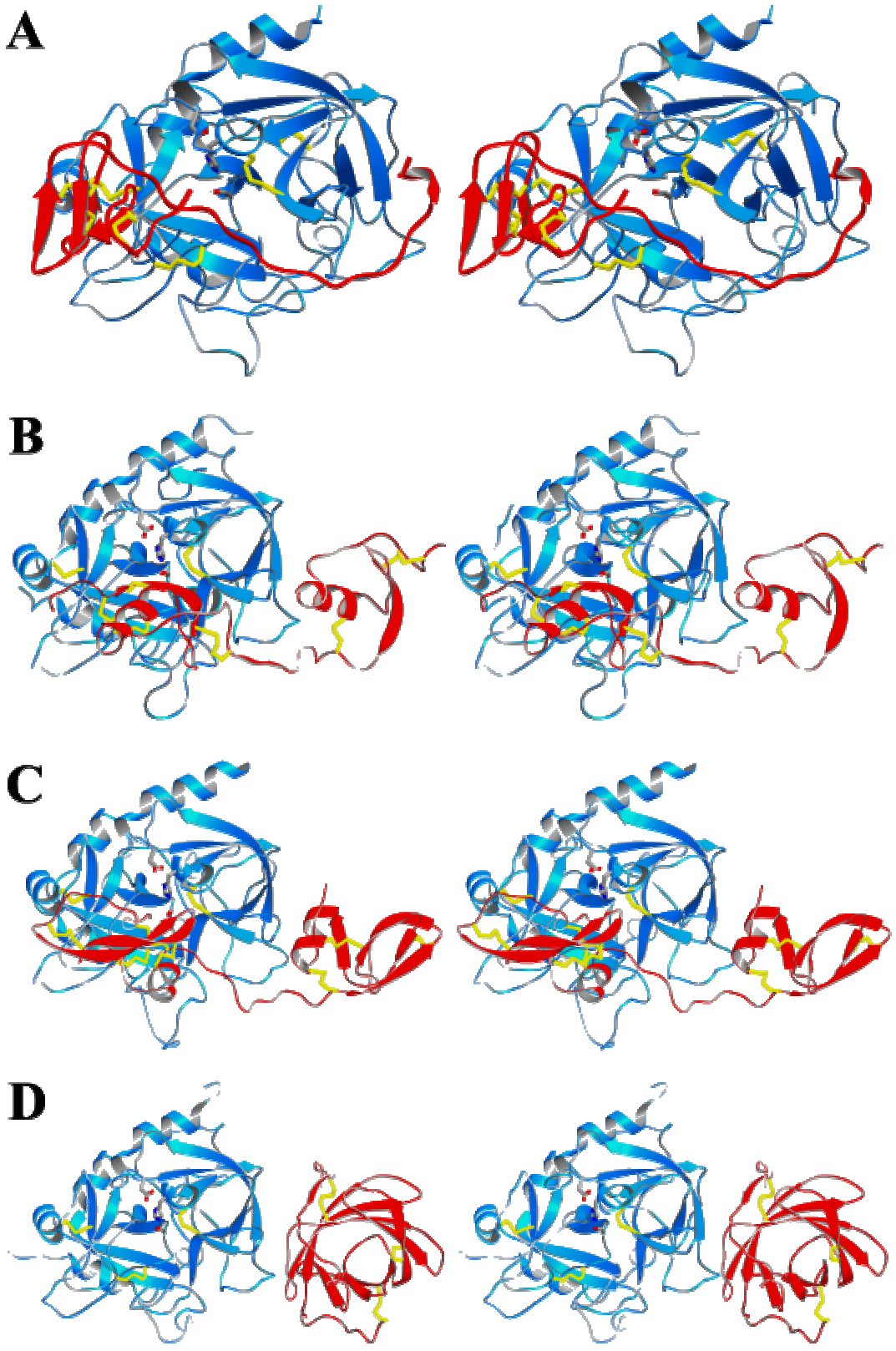


Figure 4: Stereo ribbon diagrams of solved X-ray structures of thrombin in complex with natural peptidic inhibitors. In each diagram thrombin is shown in blue and the inhibitor in red. Shown explicitly are the side chains of the catalytic triad. The complexes in order of appearance are the (A) hirudin/human thrombin, (B) rhodniin/bovine thrombin, (C) ornithodorin/human thrombin and (D) triabin/bovine thrombin.

Although these are the only structures of natural thrombin inhibitors, biochemical investigations or primary structure similarity has given us an insight into the structure of other thrombin inhibitors, and this information is printed overleaf in table 2.

Table 2: Structural details of known natural peptidic thrombin inhibitors to date

Species	Designation	Inhibitor	Structure	Binding mode	Reference
Ixodes ricinus	Tick	Ixin	×	Unknown	
Triatoma infestans	Assasin bug	Triatomin	×	Unknown	
Eutriatoma maculata	Assasin bug	Maculatin	×	Unknown	
Rhodnius prolixus	Assasin bug	Prolixin Rhodniin	× 1TBQ 3.1 Å 1TBR 2.6 Å	Unknown Kazal, active site plus exosite I	van de Locht *et al*., 1995
Tabanus bovinus	Horse fly	Tabanin	×	Unknown	
Hirudo medicinalus	Leech	Hirudins (×11)	3-4HTC 2.3 Å 1HRT 2.8 Å NMR 4-6HIR 1HIC	Active site plus exosite I	Rydel *et al*., 1990, 1991 Vitali *et al*., 1992 Folkers *et al*., 1989 Szyperski *et al*., 1992
Haemadipsa sylvestris	Leech	Haemadin	1E0F 3.1 Å	Active site plus exosite II	Richardson *et al*., 2000
Hirudo Nipponia	Leech	Leech Granulin Hirujins A+B	× ×	Unknown Hirudin like	
Theromyzon tessulatum	Leech	Theromin	×	Unknown dimer	
Hirudinaria manellensis	Buffalo leech	Bufrudin Hirudin P6 Hirudin P18 HM1 HM2	× × × × ×	Hirudin like Hirudin like Hirudin like Hirudin like Hirudin like	
Ornithodorus savignyi	Tick	Savignin	×	Ornithodorin like	
Ornithodoros moubata	Soft tick	Ornithodrin	1TOC 3.1 Å	Kunitz type active site plus exosite I	van de Locht *et al*., 1996
Boophilus microplus	Cattle tick	Boophilin	×	Kunitz type	
Glossina morsitans	Tsetse fly	TTI	2.9 Å	Active site plus exosite I	
Anopheles albimanus	Mosquito	Anophelin	×	Unknown peptide	
Dipetalogaster maximus	Bug	Dipetalogastin	×	Rhodniin like	
Triatoma pallidipennis	Triatomine bug	Triabin	1AVG 2.6 Å	Exosite I only	Fuentes-Prior *et al*., 1997
Bothrops jararaca	Snake	Bothrojaracin	×	Exosites I and II	

4.5.3. Introduction to haemadin

The thrombin inhibitor haemadin was first isolated at BASF Ludwigshafen from the heads of decapitated *Haemadipsa sylvestris* leeches (Strube *et al.*, 1993). *Haemadipsa sylvestris* is a land-living leech largely found in India and the subcontinent, and attacks mainly cattle and occasionally humans, but is not considered as a major pest. BASF's interest in the leech was to develop a thrombin inhibitor useful for the treatment of coronary thrombosis. After homogenising upwards of 600 leech heads, the extracted crude homogenate (after clearing) was run over a thrombin affinity column. Treatment of the column with acid removed the specific thrombin binding proteins. The only thrombin binding protein from the column to be identified was haemadin, and it was subsequently sequenced revealing the first 45 amino acids. rHPLC analysis of the protein, however, revealed there may be as many as ten isoforms. The gene coding for haemadin was found in the leech and by the use of recombinant gene technology an effective expression system as a maltose binding protein conjugate was developed in *E. coli.* The gene revealed that the protein was synthesised as a 77 amino acid precursor that was post translationally processed to 57 amino acids. Upon production of large amounts of the protein, a kinetic analysis of the inhibitor was undertaken. The kinetic analysis classified haemadin as a very specific thrombin slow tight-binding inhibitor, and gave the K_i as $99 \pm 26 \times 10^{-15}$ M. At the time due to its sequence similarity to hirudin (See figure 5), it was assumed to have the same mode of action as hirudin.

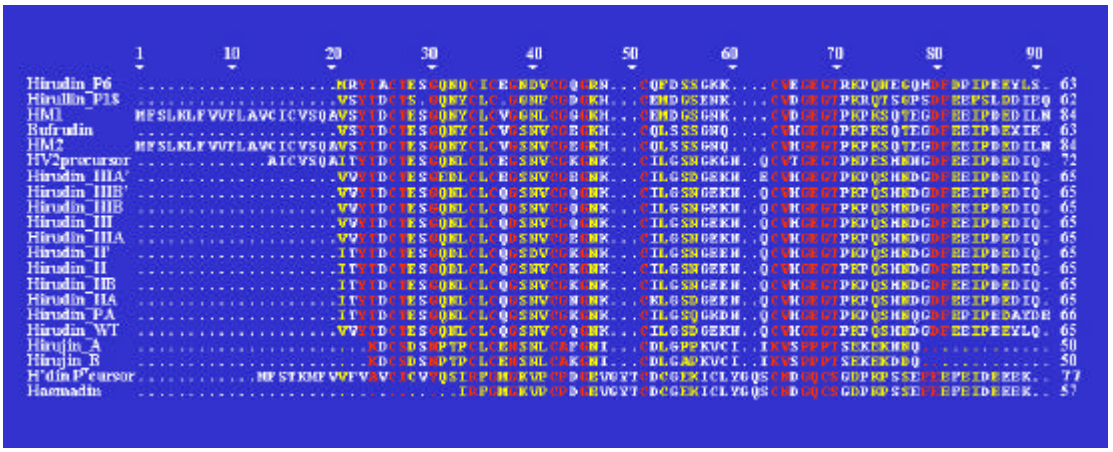


Figure 5: Alignment of the amino acid sequences of haemadin and of the other 20 known hirudin variants. Residues with particularly close homologies are boxed in yellow, identities in red. The numbering system is absolute with a full stop (.) indicating a gap has been introduced for the maximal homology.

4.5.4. Introduction to tsetse thrombin inhibitor

The tsetse thrombin inhibitor (TTI) was first isolated from *Glossina morsitans morsitans* salivary gland extracts at the Yale University school of Medicine (Cappello *et al.*, 1998). *Glossina morsitans morsitans* is commonly know as the tsetse fly, and is an important vector of African trypanosomiasis, a potentially fatal disease known trivially as sleeping sickness. The inhibitor was shown to be a low molecular weight peptide (MW = 3,530 Da) by laser desorption mass spectrometry, after purification using a combination of size-exclusion chromatography and reverse-phase, high-performance liquid chromatography. Amino terminal sequencing of the purified protein revealed it to be a 32 amino acid peptide, with no homology to any previously identified serine proteinase inhibitor or naturally occurring anticoagulant. Discovery of the gene in *Glossina morsitans morsitans* revealed the inhibitor was produced as a 53 amino acid precursor which was post translationally processed. The tsetse thrombin inhibitor (TTI) is a stoichiometric inhibitor of thrombin, with an equilibrium dissociation inhibitory constant (K_i) of 9.2×10^{-15} M. In addition, it is also a potent inhibitor of thrombin-induced platelet aggregation. Attempts to reproduce the efficacy of the inhibitor by solid phase chemical synthesis (sTTI) and recombinant expression in *E. coli* (rTTI) produced peptides with a significantly lower K_i of about

2.5×10^{-9} M, with the same primary structure as the native inhibitor. This apparent anomaly between the native and man made forms to date has not been resolved.

4.6. Twinning

Perfect crystals consist of molecules adopting exactly the same conformations along with exactly related positions and orientations. However in most crystals this is an elusive phenomenon and occurs in the case of only a few small molecule crystals. These strict crystallinity rules may be broken in many ways. Conformational differences between the molecules in space or time causes the diffraction to decrease with increasing scattering angle and the crystal resolution is worsened.

Additionally, series of minor crystal growth defects may cause a crystal, which despite being highly ordered locally, not to be highly ordered globally. This phenomenon is called mosaicity and it causes broadening or smearing of the diffraction pattern. This phenomenon is present in most crystals to some degree or another, and despite being disadvantageous their effects are predictable.

However another type of crystal growth disorder is twinning which is much less frequent and can often remain undetected. Usually twinning is thought of as being a crystal growth disorder where many crystals grow together in non specific clusters such as those in figure 6, however these sorts of macroscopic growth defects are usually excluded from the strict definition of crystal twinning given below. Occasionally concave crystal morphology indicates crystal twinning, but often microscopic examination of twinned crystal is uninformative as is seen from the twinned crystals shown in figure 6.

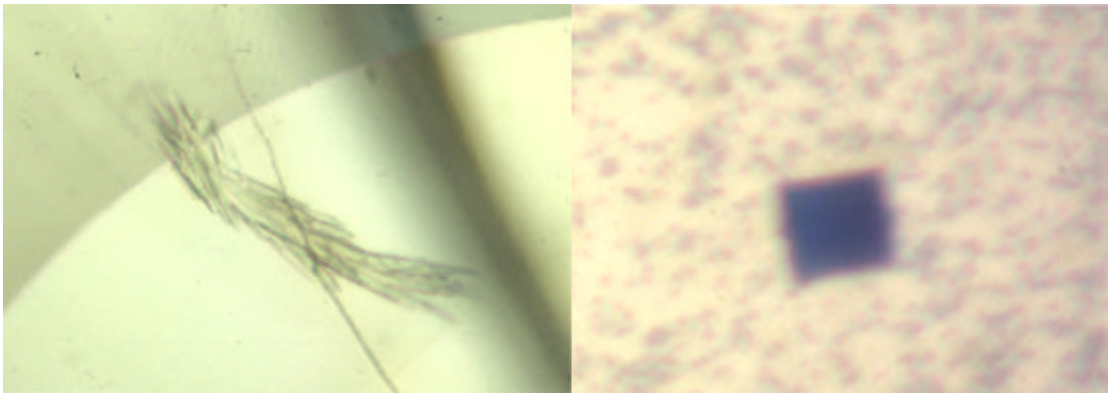


Figure 6: Photographs of crystals taken under a microscope. The crystals on the left of the bovine thrombin-bothrojaracin complex exhibit clear macroscopic twinning, whereas the crystal on the right of the bovine thrombin-sTTI complex appear to be untwinned macroscopically. The bovine thrombin-sTTI complex crystal shown have approximate dimensions of $0.1 \times 0.1 \times 0.1$ mm^3 and is stained blue with bromophenol blue.

4.6.1. Twinned crystals

During the growth of a crystal, nuclei may associate in different orientations that achieve good molecular packing but violate the symmetry of the crystal. This results in a crystal which consists of two or more crystalline domains, with orientations which differ in a definite way, and this is what crystallographers refer to as twinning. In small-molecular crystallography twinning is relatively common, however there are many more small molecular crystal structures and amongst these twinning has a long history (Catti *et al.*, 1976), so much so that small molecule crystallography programs often have strategies to deal with twinned crystals.

In protein crystallography twinning is less common and has only been reported in a few cases. In the RCS crystal database there are 29 entries (out of 13,621 (1/11/01)) which match the search criterion twin, hemihedral and merohedral. These correspond to 18 different proteins which are summarised below in table 3.

Table 3: Twinned X-ray crystal structures in the RCS data bank

Protein	Space group	Resolution	α = 0.5	RCS PDB No.	Reference
Nitrate reductase	R3	2.5 - 3.0	✕	1CNE 1CNF 2CND	Lu *et al*., 1995
Arginase	P3	1.7, 2.3	✓	1D3V 1HQ5	Cox *et al*., 1999
Rubredoxin oxygen oxidoreductase	P21212	2.5	✕	1E5D	Frazao *et al*., 2000
Autoinducer 2 production protein	P42	2.4	✕	1JOE	Unpublished
Rat mastcell proteinase II	P31	1.9	✕	3RP2	Remington *et al*., 1988
IGG2A-κ antibody 26-10 light chain	P21	2.5	✕	1IGJ	Jeffrey *et al*., 1993
Bacteriorhodopsin	R32 P63	1.55 - 2.3	✓ (1)	1BRX 1C3W 1C8R 1C8S 1F4Z 1F50	Luecke *et al*., 1998, 1999, 1999, 2000
Glycerol-3-phosphate cytidylyltransferase	P21	1.8	✕	1LOZ	Booth *et al*., 1997
Deacetoxycephalo-sporin C synthase	R3	1.3, 1.5	✕	1DCS 1RXF	Valegard *et al*., 1998
Canavalin	R3	1.7	✕	1DGW	Unpublished
Phycoerythrin	R3	2.25	✓	1EYX	Contreras-Martel et al., 2001
HSLU Adenosine-5‘ -diphosphate	P63	2.3, 2.8	✕	1G41 1IM2	Trame *et al*., 2001
Antartic fish haemaglobin	P21	2.2	✕	1HBH	Ito *et al*., 1995
Lipase B	P622	2.6	✕	1LBS	Uppenberg *et al*., 1995
Mersacidin	P32	1.06	✕	1QOW	Schneider *et al*., 2000
GTPase activating protein	I41	2.66	✕	1YRG	Hillig *et al*., 1999
Plastocyanin	P32	1.5	✕	2PLT	Redinbo & Yeates, 1993
D-glyceraldehyde-3-Phosphate dehydrogenase	C2	3.5	✓	3GPD	Moras *et al*., 1975 Mercer *et al*., 1976
Lambda Phage Display Platform Protein Gpd	P1 21	1.1	✕ (MAD)	1C5E	Forrer *et al*., 2000
Human α-lactalbumin	P21212	1.8	✕	1A4V	Chandra *et al*., 1998
Arcelin-1	P21212	1.9	✕	1AVB	Mourey *et al*., 1998

4.6.2. Twinning types

There are two basic twinning types; non-merohedral (or epitaxial), and merohedral, these are shown below in figure 7.

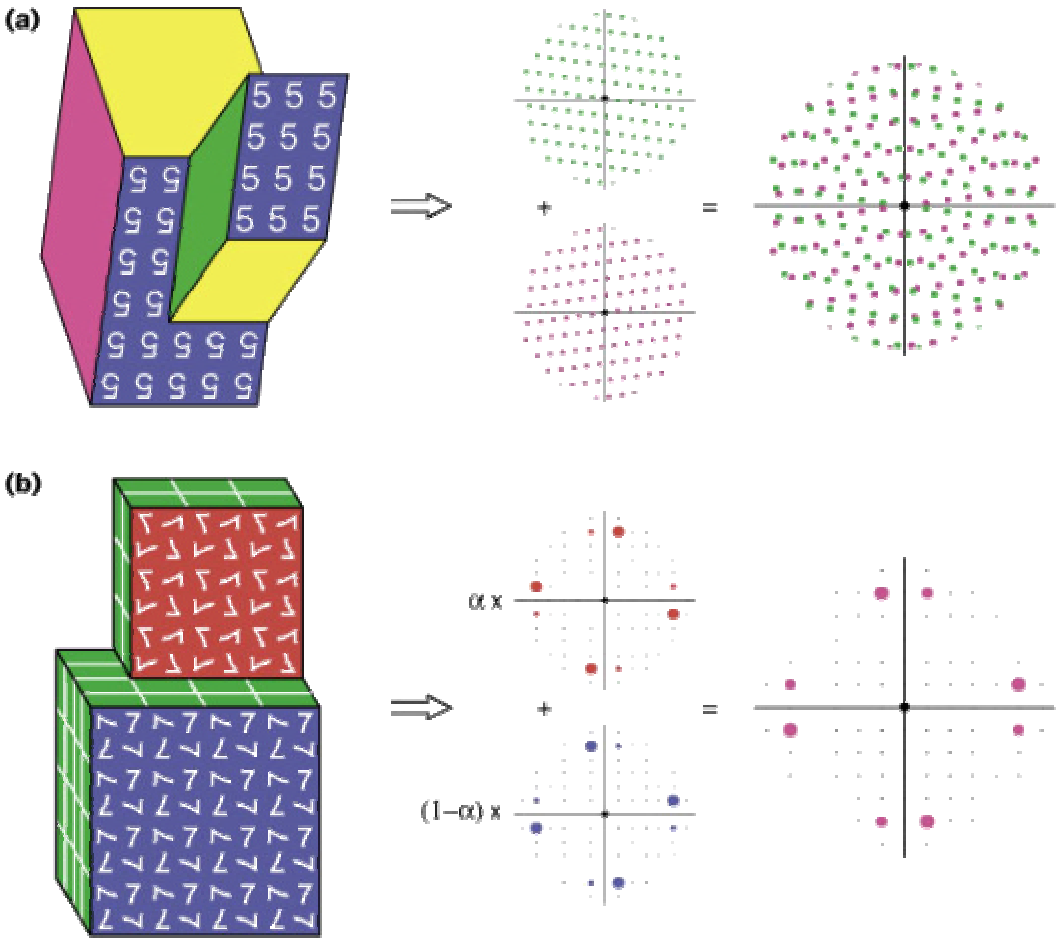


Figure 7: Schematic representation of non-merohedral and merohedral twinning. (a) Non-merohedral twinning, the separate domain lattices align in fewer than three dimensions, in this case at the meeting surface, thus the diffraction pattern contains interpenetrating lattices. (b) Merohedral twinning, the domain lattices superimpose completely in three dimensions with the resulting diffraction pattern being a single lattice. Shown in detail is one pair of twin-related reflections (and their symmetry mates in P4). The observed intensities have domain contributions according to their volume. When the twin fraction (α) = 0.5, the twin-related observations are equal and the diffraction pattern acquires erroneously high symmetry. In the diagram, the domains are represented as spatially distinct, however, merohedral twins may grow with completely interpenetrating domains. This figure is used with permission from structure.

4.6.2.1. Non-merohedral twinning

In non-merohedral or epitaxial twinning the lattices of the distinct domains may match in two dimensions (e.g. at the surface where the two domains meet), but may be incompatible in three dimensions. With the two lattices not entirely matching, the diffraction from this type of crystal contains two (or more) interpenetrating lattices. For this reason non-merohedral twinning is easy to detect, in the case where one domains is the major component, processing of the data may only integrate the reflections from the major lattice component (Sparks, 1997). A small amount of the reflections may be too close to the minor lattice spots, but a virtually complete data set may be collected. With its ease of detection and elimination at an early stage of structure solution non-merohedral will not be considered further here.

4.6.2.2. Merohedral twinning

In merohedral twinning the lattices of the distinct domains are completely superimposable in three dimensions. This is possible only in crystal systems where the lattice symmetry exceeds the underlying symmetry of the crystal. In all macromolecular cases detected so far only two distinct domain orientations (related by a 180° rotation) are involved, hence this is referred to as hemihedral twinning.
Merohedrally twinned crystals diffraction patterns seem perfectly normal, because the separate domain lattices are superimposable. However every measured diffraction intensity actually contains contributions from the two twin-related reflections, and depends on the ratio of the two domain volumes. Where the twinning ratio (α) is not equal to 0.5 the merohedral twinning is referred to as partial merohedral twinning. When α is exactly equal to 0.5 the twinning is referred to as perfect merohedral twinning. Depending on which form of merohedral twinning is present, two different strategies for detwinning of the data are possible.

4.6.2.3. Partial merohedral twinning

In the case where the two domains differ substantially in volume ($\alpha < 0.45$) the secondary domain will make a minor contribution to the diffraction intensities. This effect can easily remain undetected and will prevent MIR and MAD structure determination both of which require very accurate data. However structure solution by molecular replacement will usually proceed without hitch, but problems will occur later during refinement. After detection of such twinning if α is much less than 0.5, the intensities can be recovered from the observed data. For each pair of reflections, two measurements are obtained that are linear combinations of the two unknown crystallographic intensities. The mixing of twin-related reflections that is caused by twinning can be reversed mathematically (Grainger, 1969). The downside of this treatment is that detwinning increases the measurement errors by a factor of $1/(1-2\alpha)$, and another downside is that the accurate value of α must be known. The best estimation of α is by comparing twin-related observations, there are many different approaches to this (Britton, 1972; Murray-Rust, 1973; Fisher and Sweet; 1980; Rees, 1982; Yeates, 1988). So a low well determined value of α leads to the determination of the true crystallographic intensities, and normal structure determination may continue.

4.6.2.4. Perfect merohedral twinning

When α is close or equal to 0.5 the two twin-related observations contain equal contributions from the two reflections. That is the twin-related observations are the same, and this leads to two problems. Firstly, for every pair of unknown crystallographic intensities, only one independent observation, the average is known, so the separate desired intensities cannot be extracted. Secondly, with the twin-related reflections being equal, an erroneously high symmetry is imposed on the diffraction pattern and the resultant data set. If the intensity statistics are not checked, it is possible that the correct crystal symmetry will not be determined (Wilson, 1949; Stanley, 1972).

With perfect twinning the crystal structure may only be determined without direct knowledge of the true crystallographic intensities.

Molecular replacement has been performed on perfectly hemihedrally twinned structures, and has been checked by solution of the same crystals with a lower α (Redinbo and Yeates, 1993). The solution of a perfectly hemihedrally twinned crystal structure by MAD and MIR methods has yet to be performed but has been theoretically considered in the case of MIR (Yeates and Rees, 1987).

4.6.2.4.1. Treatment of hemihedrally twinned structures

As has been mentioned previously it is possible to use replacement methods in order to solve a perfectly hemihedrally twinned structure.

In general there are two method for refining a structure with a hemihedral twin ratio (α) equals 0.5. In one method, which has been implemented in small molecule crystallography the quantities to be minimised would be the differences between the observed structure factor amplitudes and values obtained by twinning the structure factors calculated from the model $[F^2_{calc}(h_1) + F^2_{calc}(h_2)]^{0.5}$ (Wei, 1969; van Koningsveld, 1983). Another method developed by Redinbo and Yeates (Redinbo and Yeates, 1993) obtains estimates of the diffraction intensities from each of the separate twin domains. For a hemihedrally twinned crystal with a twin ratio (α) of 0.5:

$$I_{obs}(h_1) = I_{obs}(h_2) = I(h_1) + I(h_2) \qquad (1)$$

Where $I(h_1)$ and $I(h_2)$ are the expected intensities contributed by each of the twin domains. $I_{obs}(h_1)$ can be measured, but this provides only a single equation in terms of the two unknowns $I(h_1)$ and $I(h_2)$. Further equations can be obtained by approximating the untwinned diffraction intensities with values calculated from the positioned molecular replacement model $I_{calc}(h_1)$ and $I_{calc}(h_2)$:

$$I(h_1) \approx I_{calc}(h_1) \qquad (2)$$

$$I(h_2) \approx I_{calc}(h_2) \qquad (3)$$

In the case where a model closely resembles the actual structure substitution into equation (1) gives:

$$I(h_1) = I_{obs}(h_1) - I_{calc}(h_2) \quad (4)$$

$$I(h_2) = I_{obs}(h_1) - I_{calc}(h_1) \quad (5)$$

This gives two estimates for the untwinned intensities for each twin domain, (2) and (4) for domain 1 and (3) and (5) for domain 2. Least squares minimisation reduces these pairs of equations to the arithmetic mean of the two estimates,

$$I_{detwin}(h_1) = [I_{obs}(h_1) + I_{calc}(h_1) - I_{calc}(h_2)]/2 \quad (6)$$

$$I_{detwin}(h_2) = [I_{obs}(h_1) + I_{calc}(h_2) - I_{calc}(h_1)]/2 \quad (7)$$

Using (6) and (7), the untwinned intensities from each of the twin domains can be estimated. These artificially 'detwinned' data can be used to refine the model and to generate electron density maps. As the model of molecule 1 improves through refinement and model building $I_{calc}(h_1)$ and $I_{calc}(h_2)$ improve, which also improves the estimate of $I_{detwin}(h_1)$ and facilitates further refinement. This pattern of cyclic model improvement and detwinning is continued until structural refinement is complete. Two different *R* factors can be calculated when refining in this way:
The first is the *R* factor calculated using the data against which the structure is refined, which for molecule 1 is:

$$R_{untwin} = \sum_i | F_{calc}(h_1) - F_{detwin}(h_1)_i | \; / \sum_i | F_{detwin}(h_1)_i | \quad (8)$$

The second R factor is calculated against the original data and is given by:

$$R_{twin} = \sum_i | [F^2_{calc}(h_1) + F^2_{calc}(h_2)]^{0.5} - F_{obs}(h_1)_i | \; / \sum_i | F_{obs}(h_1)_i | \quad (9)$$

R_{untwin} is simply the usual crystallographic *R* factor with the F_{obs} replaced by F_{detwin}. The other R_{twin} replaces the usual F_{calc} by the F_{calc} of the twinned crystal. R_{twin} is expected to be lower than R_{untwin} by a factor of $2^{0.5}/2$ as the model errors are averaged over two twin related reflections, it is however unbiased as it directly compares the model to the observed data. R_{twin} can be calculated for reflections (h_1) for which (h_2) has not been measured and which are not included in refinement.

4.7. Aim of the work

The framework of this work was to determine the crystal structures and the modes of binding of natural bloodsucking animals inhibitors of thrombin. To date only four of such structures have been determined, and the generation of further structures should help us achieve a greater understanding of how these creatures inhibit thrombin in order to feed from mammals. As seen in the introduction, thrombin is a many faceted enzyme, which plays central roles in the blood coagulation system and haemostasis. With its central role in coagulation, the inhibition of thrombin has been undertaken from a medical standpoint to prevent the onset of coronary thrombosis (or myocardial infarction or heart attack), from which there are more deaths than any other medical disorder in the western world. The use of thrombin inhibitors to prevent and treat this condition is a very active area of attention for pharmaceutical companies as seen from figure 8. The generation of unique thrombin inhibitor structures should be helpful for the creation of better thrombin inhibitors, with which to fight these diseases, replacing reliance on the traditional treatments of heparin and warfarin.

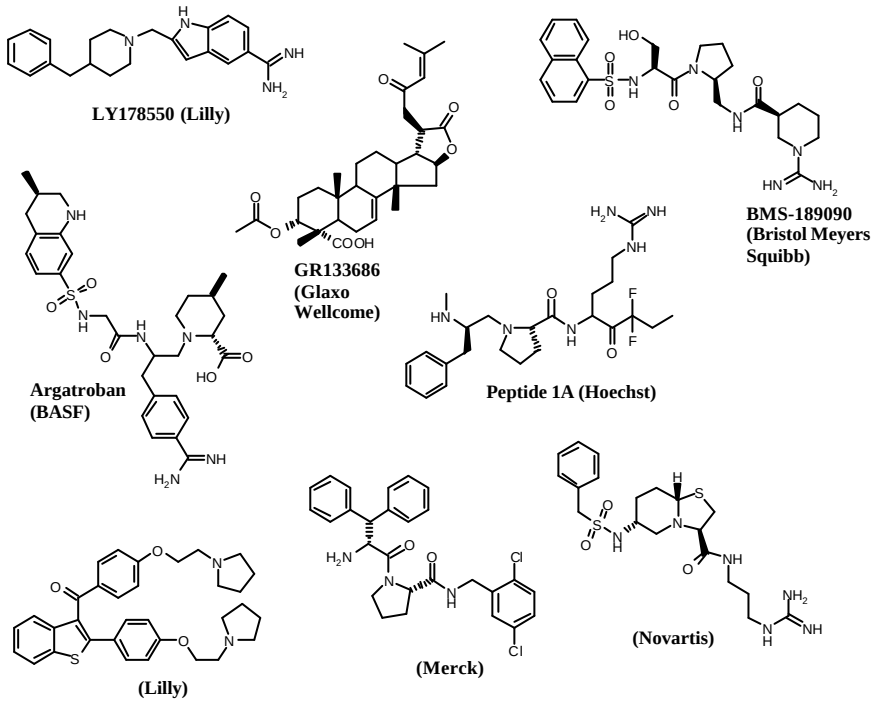


Figure 8: The chemical structures of small molecule inhibitors as worked on by major pharmaceutical companies. Listed are the compounds generic names (if appropriate) along with the company where they were developed. Of all the compounds to date only Agartroban has received approval for treatment of deep vein thrombosis of the lower legs.

5. MATERIALS AND METHODS

5.1. Materials and equipment used

5.1.1. Chemicals

Ammonium peroxodisulphate (BioRad, München)
Bacto-Trypton, Bacto-Agar and yeast extract (Difco Laboratories USA)
Bradford protein assay (BioRad, München)
Chromazyme-TH, Tosyl-GPR-AMC (Sigma, Deisenhofen)
Coomassie brilliant blue R-250 (Serva)
IPTG (Biomol, Hamburg)
Protogel (30% w/v Acrylamide / 0.8% (w/v) Bis-acrylamide) (National Diagnostics, Manville, USA)

In addition other chemicals of p.a. quality from the firms of Fluka (Neu-Ulm), Merck (Darmstadt), Roth (Karlsruhe), Serva (Heidelberg) or Sigma (Deisenhofen).

5.1.2. Proteins

Factor Xa (Boehringer Mannheim/Roche, Penzberg)
Oxyranus scutelatus venom (LATOXAN, Rosans, France)

5.1.3. Bacteria

E. coli DH5α (Stratagene GmbH, Heidelburg) (Hanahan, 1983)
Genotype: F^-, *deo*R, *end*A1, *gyr*A96, *hsd*R17($r_k^- m_k^+$), *rec*A1, *rel*A1, *sup*E44, *thi*-1, Δ (*lac*ZYA-*arg*F), ϕ80*lac*ZΔM15$F^-\lambda^-$

All bacteriological phyla were obtained from chemically competent cells. The temperature of storage was -80°C.

5.1.4. Plasmids

The plasmid for the haemadin-maltose binding protein conjugate was constructed from haemadin cDNA and the bacterial expression vector pMAL-p2 (New England Biolabs) at the Department of Biotechnology BASF Ludwigshafen, and its construction is detailed in Strube *et al.* 1993.

5.1.5. Equipment

Äkta FPLC system with gradient programmer GP-250 and two P500 pumps. (Amersham Pharmacia, Freiburg)
Autoclave GVA 570 (Fritz Gössner GmbH & Co., Hamburg)
Diode array photometer DU 7500 (Beckman, München)
Eagle Eye II (Stratagene, Heidelberg)
Goniometer head (Huber Diffractionstecknik, Rimsting)
Image plate (MarResearch, Hamburg)
Leica MZ 12 microscope (Leica, Germany)
PHM 82 standard pH meter (Radiometer Copenhagen, Denmark)
Refridgerated centrifuge J2-21 with JA10 and JA20 rotors (Beckmann, München)
Rotating anode RU2000 (RIGAKU)
SMART system (Amersham Pharmacia, Freiburg)
UVIKON 943 double beam spectrophotometer (BIOTEK-KONTRON Instruments, Milan)

For working with diffraction data, workstations from the firms of DIGITAL alpha stations (OSF), SILICON GRAPHICS work-stations and computer servers (IRIX) or IBM compatible personal computers were used.

5.1.6. Consumables

1 mm thick silanised quartz capillaries (W. Müller, Berlin)
24 well CrysChem polystyrene plastic plates (Charles Supper, Natick, USA)
Centricon 30 tubes (Amicon, Lexicon, Massachusetts, USA)
Millex-GV 0.22 µm sterile filters (Millipore, Molsheim, France)
Sealing wax (Harvard, Berlin)

5.1.7. Column materials

Affi-Gel 10 (Biorad, Hercules, California USA)
DEAE-5PW column (TosoHaas, Germany)
Heparin Sepharose (Amersham Pharmacia, Freiburg)
Q-Sepharose Fast Flow (Amersham Pharmacia, Freiburg)
ResourceQ (Pharmacia Biotech, Uppsala, Sweden)
Reverse phase source 5RPC ST column (Amersham Pharmacia Biotech)

5.2. Microbiological methods

All work with bacteria was conducted under sterile conditions, meticulous sterilisation of working equipment and solutions was achieved via autoclaving, non-autoclavable solutions were sterile filtered through Millex-GV 0.22 µm sterile filters. Standard microbiological procedures were carried out in accordance to Sambrook *et al.* (Sambrook *et al.* 1989).

5.2.1. Bacterial culture

Autoclaved LB (Luria-Bertani) medium was used for the expression of recombinant haemadin.
Antibiotic solution: Ampicillin 500 fold solution cold filtered with Millex-GV 0.22 µm sterile filters 50 mg/ml in ddH_2O.

RB-Medium:	In 1 litre ddH_2O:
	10 g NaCl
	10 g Bacto-Tryptone
	5 g yeast extract
	5 g glucose
	NaOH added until pH 8.0 achieved

This is essentially the recipe for LB medium with added glucose.

5.2.2. Culture of bacterial plates

LB medium with 15 g/l agar was autoclaved. After cooling the solution to about 60°C, ampicillin was added to it to a final concentration of 100mg/l and the solution was poured into sterile Petri dishes. The Petri dishes were stored overnight in a sterile environment and were stored inverted and sealed at 4°C. Bacterial cultures were added with a sterile spreader and were incubated at 37°C for 12 hours.

5.2.3. Transformation of foreign DNA into competent *E. coli* DH5a cells

Forty microlitres of competent cells were kept on ice for 30 minutes along with 0.01 µg of plasmid DNA. The cells were then heated at 42°C for 90 seconds (heat shock). The entire cell solution was then transferred into a sterile reagent glass containing 100 ml LB medium. After shaking for 1 hour at 37°C, 100 µl of the solution was plated onto LB ampicillin plates and kept at 37°C overnight. Viable colonies were picked using sterile pipette tips and added to 100 ml LB medium with 100 mg/l ampicillin added and were shaken at 37°C overnight. The resulting cell cultures were kept frozen along with 30% glycerol at -80°C for further use.

5.2.4. MBP-haemadin expression in fermented *E. coli*

One millilitre of cell culture was added to 100 ml LB medium and grown at 37°C overnight. Four millilitres of this culture were added to baffled 2 l Erlenmeyer flasks containing 400 ml RB medium with 100 mg/l ampicillin. The flasks were shaken at 37°C until the optical density of the cells reached 0.6 at an wavelength of 600 nm against a blank of RB medium. At this point induction of expression was achieved using 1 mM IPTG and an additional 100 mg/l ampicillin was added. Optimal expression of MBP-haemadin was achieved after 4 hours.

5.2.5. Periplasmic extraction

Preparation of the periplasmic extract was achieved using the cold osmotic shock method (Neu and Heppel, 1965). This method utilises two solutions A + B the contents of which are disclosed below.

Solution A: 30 mM Tris-HCl
20% (w/v) sucrose
1 mM EDTA
pH 8.0

Solution B: 5 mM $MgSO_4$
4°C

The above grown RB medium cells were centrifuged at 4,000 × g for 20 min at 4°C, and the supernatant was discarded. The cell pellet was resuspended in solution A (400 ml per litre grown RB medium) with mild stirring for 10 min.

The resulting suspension was centrifuged at 8,000 × g for 20 min at 4°C, the supernatant was removed, and the pellet was resuspended on ice with solution B (400 ml per litre grown RB medium) with mild stirring for 10 min.

The resulting suspension was centrifuged at 8,000 × g for 20 min at 4°C, the supernatant (the cold osmotic shock fluid) contained the periplasmically expressed protein.

5.3. Protein biochemical methods

5.3.1. Bovine and human a-thrombin preparation

Bovine and human prothrombin were purified from bovine blood and human plasma by a modification of the procedure of Mann (Mann, 1976). Purified snake venom from the snake *Oxyranus scutelatus* bound to a matrix, as described below, was used to activate the prothrombin to thrombin.

One hundred milligrams of lyophilised snake venom in 10 ml of 300 mM NaCl, 20 mM MOPS, pH 6.5 was added to 25 ml of Affi-Gel 10. The resulting mixture was rotated at 4°C for 5 h until the extinction at 280 nM no longer decreased. The gel was drained and washed with ten volumes of the buffer that was used to bind the snake venom. The washed gel was suspended for an hour in a solution of 1 M ethanolamine pH 7.5 at 4°C to block any remaining active groups. The gel was washed again with ten volumes of snake venom buffer and the resulting gel was added to prothrombin solutions along with $CaCl_2$ to a final concentration of 5 mM, where the formation of thrombin was assay using activity tests, as well as using SDS-PAGE (four hours usually sufficing). After dilution tenfold and the addition of EDTA to a final concentration of 10 mM, the resulting solution was added to Heparin Sepharose column pre equilibrated in 10 mM MES pH 6.0. After column washing with ten volumes of 10 mM MES pH 6.0 the protein was eluted using a linear gradient of 0-1 M NaCl with 10 mM MES pH 6.0, with most of the thrombin eluting at 0.5 M NaCl. Fractions corresponding to thrombin were pooled assayed and frozen at -20°C in aliquots.

5.3.2. Purification of MBP-haemadin

Fifty millilitres of Q-Sepharose per litre of cold osmotic shock fluid was packed in a column and equilibrated with a solution of 20 mM Tris-HCl pH 8.0. The cold osmotic shock fluid had 1 M Tris-HCl pH 8.0 added to it to achieve a final concentration of 20 mM. The resulting solution was pumped onto the column overnight at 4°C. The column was then washed with 20 mM Tris-HCl pH 8.0 until

the A_{280} of the column flow through was the same value when the column was equilibrated. The proteins were eluted with a 800 ml linear gradient per 50 ml bed volume of from 0 to 0.65 M NaCl in 20 mM Tris-HCl pH 8.0. The majority of the MBH-haemadin was eluted from the column between 0.41 and 0.45 M NaCl concentration and was detected using UV-light at a wavelength of 280 nm.

5.3.3. Thrombin activity assay

Chromogenic amidolytic assays of thrombin activity were performed using 150 µM Tos-Gly-Pro-Arg-pNA as a substrate and a standard thrombin concentration of 6 nM. Reactions were performed at 25°C in 100 mM Tris-HCl buffer (pH 7.8), containing 100 mM NaCl and 0.1% PEG6000. Reaction progress was monitored at 405 nm on a UVIKON 943 double beam spectrophotometer. The molar amount of inhibitor was calculated assuming a 1:1 ratio of thrombin to inhibitor binding at concentrations in the nM range, being equal to the fractional loss of initial thrombin activity measured against a 6 nM thrombin standard, multiplied by 6 nM.

5.3.4. Cleavage of the fusion protein and production of haemadin

The fractions containing MBP-haemadin were diluted until a concentration of 0.1 M NaCl was achieved and additional 1 M Tris-HCl pH 8.0 was added to a final concentration of 50 mM, as well as 1 M $CaCl_2$ to a final concentration of 1 mM.
Factor Xa was added by weight at a ratio of 1/600th of the weight of the fusion protein (calculated using the method of Bradford (Bradford *et al.* 1976)). The formation of haemadin was followed using the thrombin activity assay. When the calculated amount of haemadin was perceived as not increasing (usually after 2h) the reaction was stopped using 2 mM EDTA. Haemadin was purified on a DEAE-5PW column (TosoHaas, Germany) using a linear gradient of 0-500 mM NaCl, in 20 mM Tris-HCl, pH 8.0. The pure protein eluting at around 0.41 M NaCl. The amount of protein was calculated by both the Bradford method and the thrombin activity assay and was lyophilised.

5.3.5. Formation and purification of the human α-thrombin haemadin complex

Lyophilised haemadin was added to human α-thrombin (4.1 mg/ml, 10 mM Mes pH 6.0, 500 mM NaCl) in slight excess of a 1:1 molar ratio. After dialysis against 20 mM Tris-HCl, pH 8.0, the complex was purified on a DEAE-5PW column using a linear gradient of 0-250 mM NaCl, in 20 mM Tris-HCl, pH 8.0. Fractions corresponding to the equimolar complex were exhaustively dialysed against 10 mM Tris-HCl pH 8.0, 20 mM NaCl and 0.02% (w/v) NaN_3, and concentrated in Centricon 30 tubes) to a concentration of approximately 5 mg/ml.

5.3.6. Polyacrylamide-gel electrophoresis (PAGE)

Discontinuous polyacrylamide-gel electrophoresis was carried out according to a modification of Laemmli (Laemmli, 1970). This method can be used with or without SDS, and is most frequently used with SDS (and called SDS-PAGE) taking advantage of the fact that proteins can bind up to 1.4 g SDS to 1 g protein thereby denaturing them. The resulting denatured proteins are similarly charged so the resulting separation is due to the size of the protein. In this way proteins with molecular mass differences of around 2% can be resolved using SDS-PAGE. Without SDS the separation is achieved according to the charge to size ratio of the protein, and is normally used without denaturation, which is a useful method to detect complex formation. Separation is achieved by applying a voltage across a polyacrylamide matrix which contains certain sized pores depending on the percentage of acylamide used to make the gel. Complex formation in solution was analysed using non-denaturing polyacrylamide gel electrophoresis (PAGE). After preincubation for two minutes in 20 mM Tris-HCl (pH 6.9), the protein solutions were subjected to electrophoresis in 10% polyacrylamide gels containing 365 mM Tris-HCl (pH 8.8). The migration buffer consisted of 100 mM glycine and 25 mM Tris-HCl (pH 8.8).

5.3.7. Staining of gels

Finished gels were stained for 2 hours with shaking in the following solution :

2.5 g Coomassie brilliant blue R-250
250 ml ethanol
80 ml glacial acetic acid
670 ml ddH_2O
After the gel was a deep dark blue the colour was removed by shaking the gel in:
250 ml ethanol
80 ml acetic acid
670 ml ddH_2O
leaving only the protein bands stained in the gel.

5.3.8. Mass spectroscopy

Mass spectroscopy was carried out on a triple-quadrapole Ionspray-Mass spectrometer API 365 (SciEx, Thornhill, Ontario, Canada). The m/z scales were calibrated using ammonium salts of polypropylene glycol oligomers after multiple loadings (Covey *et al.*, 1988; Mann *et al.*, 1989). Spectra were run by Miss Sylvia Koerner at the Max-Planck-Institut für Biochemie in Martinsried in the Protein Analytical Department.

5.3.9. Amino terminal sequencing

Automated sequencing using the principles of Edman (Edman and Henschen, 1975) was carried out with 10 µg of haemadin on a ProBlott PVDF membrane in a ProSpin Cartridge (Applied biosystems, Foster City, USA). Systematic phenyl isothiocyanate derivatisation, and release and detection of the subsequently formed Phenylthiohydantoin derivative by HPLC against known standards, was used to elucidate the primary sequence. This work was performed on a Applied Biosystems Sequencer 470A operated by Dr. Mann at the Max-Planck-Institut für Biochemie in Martinsried.

5.3.10. Peptide synthesis

The synthesis of synthetic peptides, was achieved by the solid phase method (1 mmol scale) employing the classical protocols for N-α-Fmoc strategy (Walsh *et al.*, 1981). After cleavage from the resin, the products were precipitated with methyl tert-butyl ether/hexane, and the crude peptides were purified by preparative RP-HPLC eluted using a linear gradient from 0 to 100% of 0.1% aqueous TFA/0.08% TFA in CH_3CN, with the products being obtained upon lyophilization. The resultant proteins were judged pure from mass spectral analysis. This work was carried out in the Biological Chemistry department of Professor Moroder at the Max-Planck-Institut für Biochemie in Martinsried.

5.3.11. Sequence analysis

Alignment and analysis of protein primary sequences was accomplished using the programs BESTFIT and PILEUP of the GCG package (Wisconsin Package, Genetics Computer Group, Madison, USA).

5.4. Enzyme kinetics

In order to study an enzymatic reaction an assay is necessary. This assay is a measurement of a chemical reaction, which usually involves measuring the formation of the product. In the case of thrombin this assay is the cleavage of a small tripeptide nitroanilide substrate whose amino acids are the same, or are similar to the specificity of the active site. Once the para-Nitroaniline moiety is cleaved from the tripeptide the lone pair on the nitrogen contribute to a conjugated p system which absorbs at 405 nm, and this absorption is proportional to the concentration of para-Nitroaniline in solution. In this way the amount of product formed against time (the rate) can be measured. For thrombin the best substrates (with the lowest K_m values and highest K_{cat} values) are S-2238 which is D-Phe-pipecolyl-Arg-pNA and Chromazyme-TH which is tosyl-Gly-Pro-Arg-pNA. For more sensitive measurements, such as progress curves with picomolar amounts of enzyme, fluorescent AMC derivatives are used instead of para-Nitroanilides.

5.4.1. Slow tight-binding inhibitors

In some cases, as in the case of haemadin, the inhibitor has an extremely low K_i which causes significant inhibition at low concentrations which are comparable to the concentration of its target enzyme. *In vitro* study of this inhibitor causes complications as the inhibitor concentration is so low that it is altered as a result of binding to the enzyme. This is what is meant by tight-binding inhibition. With an effect partly attributable to their low concentrations, tight-binding inhibitors often show slow-binding characteristics. In contrast to most inhibitors that act almost immediately (or on the msec time scale), slow-binding inhibitors can take up to several seconds, minutes or hours for their effect to become fully apparent. It should be noted that each of these characteristics can exist independently, and indeed there are tight-binding inhibitors that act quickly, and slow-binding inhibitors that are not particularly strong. However the incidence between slow-binding and tight-binding is fairly common.

5.4.2. Kinetic theory

The inhibition of an enzyme in the presence of substrate can be illustrated by the following scheme where E = Enzyme, S = Substrate (S-2238 or Tos-GPR-AMC), I = Inhibitor and P = Products:

$$\begin{array}{ccccc} & & K_m & K_{cat} & \\ E & + S & \leftrightarrow ES & \rightarrow & E + P \\ + & & & & \\ I & & & & \\ k_{off} \uparrow\downarrow & k_{on} & & & \\ EI & & & & \end{array}$$

The dissociation constant of the inhibitor (K_i) is given by k_{off} / k_{on}, but K_i can also be expressed the following way:

$$K_i = [E]\,[I] / [EI]$$

So essentially the K_i of a given inhibitor is the concentration where 50% inhibition is achieved.

In the case of tight binding inhibition, where binding of the inhibitor causes a significant depletion in the amount of free inhibitor, the variation of the steady state velocity (v_s) with the total inhibitor concentration (I_t) is given by the following equation:

$$v_s = (v_o / 2E_t)[[(K_{i'} + I_t - E_t)^2 + 4K_{i'}E_t]^{0.5} - (K_{i'} + I_t - E_t)] \qquad [1]$$

Where v_o is the velocity in the absence of inhibitor, E_t is the total enzyme concentration and $K_{i'}$ is the apparent inhibition constant (Williams and Morrison, 1979).

In the case of competitive inhibition, the dissociation constant (K_i) of a tight binding inhibitor can be calculated from the apparent inhibition constant ($K_{i'}$) via the following equation:

$$K_{i'} = K_i (1 + S / K_m) \quad [2]$$

Where S is the competing substrate concentration and K_m is the Michaelis constant of the substrate for the enzyme (Williams and Morrison, 1979).

In order to determine the k_{on} and k_{off} for an inhibitor progress curves of the reaction must be followed. In the case of slow binding inhibition, the amount of product formed (P) in a certain time (t) is described by the following equation:

$$P = v_s t + (v_o - v_s) (1 - \exp(-k_{app} t))/k_{app} + d \quad [3]$$

Where v_s is the steady state velocity, v_o is the initial velocity, k_{app} is the apparent rate constant (unrelated to $K_{i'}$) and d is a time displacement term inserted to account for the fact that at $t = 0$ the amount of product may not have been known accurately.

From the k_{app} the constants k_{on} and k_{off} can be calculated from the equation below (Morrison, 1982):

$$k_{app} = [k_{on} / (1 + S / K_m)] I + k_{off} \quad [4]$$

From the K_i the standard Gibbs free energy of formation of the complex (ΔG°_b) (also referred to as the binding energy) is given by the following equation where R = the gas constant and T = absolute temperature:

$$\Delta G^\circ_b = RT \ln K_i \quad [5]$$

The ΔG°_b can be divided into an ionic component (ΔG°_{ion}) and a nonionic (ΔG°_{nio}) component:

$$\Delta G^\circ_b = \Delta G^\circ_{ion} + \Delta G^\circ_{nio} \quad [6]$$

From the Debye-Hückel theory the value of ΔG°_{ion} will be dependent on the ionic strength (I) according to the equation:

$$\Delta G^{\circ}_{ion} = \Delta G^{\circ}_{ion0} \exp(-C_1 \nu I) \qquad [7]$$

Where ΔG°_{ion0} is the ionic interaction energy at an ionic strength of zero and C_1 is a constant that is related to those terms in the Debye-Hückel screening parameter that are independent of ionic strengths, and the distance between the two charges (Adamson, 1979).

However Debye-Hückel theory is only applicable at low ionic strength, and several semiempirical terms must be introduced to account for deviations, one such treatment has been shown to be applicable to higher ionic strengths in the case of proteins (Meyer *et al.*, 1984; Tollin *et al.*, 1984; Przysiecki *et al.*, 1985):

$$\Delta G^{\circ}_{ion} = \Delta G^{\circ}_{ion0} \ (exp(-C_1 \nu I)) / (1 + C_1 \nu I) \qquad [8]$$

Substitution of equations [7] and [8] into equation [6] gives the following expression for dependence of binding energy on ionic strength:

$$\Delta G^{\circ}_{b} = \Delta G^{\circ}_{nio} + \Delta G^{\circ}_{ion} \ (\exp(-C_1 \nu I)) / (1 + C_1 \nu I) \qquad [9]$$

In the case of ionic interactions, the k_{on} will also be dependent on ionic strength, and the following equation can be derived from equations [7] and [8] to describe this dependence:

$$\ln k_{on} = \ln k_{on8} \ (-\Delta G^{\circ}_{ion0} / RT) \ (\exp(-C_1 \nu I)) / (1 + C_1 \nu I) \qquad [10]$$

Where k_{on8} is the association rate constant at infinite ionic strength.

A simple way to evaluate ionic forces is to treat the solvent as being a continuum having a dielectric constant e (Laidler, 1987):

$$\Delta G^{\circ}_{ion} = (L \ z_A \, z_B \, e^2) / (4\pi \ e_0 \ e \ d_{AB}) \qquad [11]$$

Where L is the Avogadro constant, z_A and z_B are the charges of the ions, e is the elementary charge, e_0 is the vacuum permittivity, e is the effective dielectric constant of the solvent and d_{AB} is the distance between the charge centres.

5.4.3. Amidolytic assays of thrombin

Consideration of ionic strength effects for the Debye-Hückel theory requires that low ionic strengths be used. However a modification of the Debye-Hückel theory, which has previously been shown to be applicable at higher ionic strengths (Meyer *et al.*, 1984; Tollin *et al.*, 1984; Przysiecki *et al.*, 1985), and has been applied in the case of thrombin (Stone *et al.*, 1989) has proved to fit the data more accurately than the Debye-Hückel theory. The dissociative and associative rate constants for thrombin and four thrombin mutants were determined over a range of ionic strengths from 0.0695 to 0.5195 (ten points), the use of these higher ionic strengths was to circumvent losses of protein which frequently occur at lower ionic strengths. The subsequent graphs were fitted using nonlinear regression to equation [9] for the K_i values and to equation [10] for the k_{on} values, with the points weighted to the inverse square of their standard errors.

Slow and tight binding assays were performed at 25°C as previously described (Strube *et al.*, 1993). Tight binding assays were performed with a final concentration of 500 pM human-α-thrombin (or mutant), being preincubated for upwards of 30 min with inhibitor (0.4-2.4 × [E_0]), in 50 mM Tris-HCl (pH 8.3), = 50 mM NaCl and 0.1% PEG6000. A plot of velocity against concentration of haemadin (seven points) was fitted using nonlinear regression to equation [1] to obtain the $K_{i'}$. This procedure was repeated using varying substrate concentrations (100-500 μM S-2238), and the resulting $K_{i'}$ values (5 points) were plotted against substrate concentration using weighted linear regression to equation [2], with the points weighted to the inverse square of their standard errors in order to obtain the K_i.

Slow binding assays were conducted using 25 μM Tos-GPR-AMC as a substrate in 50 mM Tris-HCl (pH 8.3), = 50 mM NaCl and 0.1% PEG6000 with varying inhibitor concentrations (15-30 × [E_0]). The reaction was started by the addition of 10 pM of human α-thrombin or mutant thrombin (final concentration), and a plot of product formed against time (33 points) was fitted using nonlinear regression to equation [3] in order to obtain the k_{app}. A plot of k_{app} against inhibitor concentration (four points) was fitted to equation [4] to obtain the k_{on}. The k_{off} was calculated from the K_i value and the k_{on} value by using the equation $K_i = k_{off} / k_{on}$.

Ionic strength was calculated assuming a pK_a of 8.1 for Tris at 25°C (Ellis and Morrison, 1982), with the ionic strength being checked by measuring the conductance of the solutions. In all cases the ionic strength measured was within 5% of the theoretical value.

The kinetic constants for each of the mutant thrombin's towards the substrates S-2238 and Tos-GPR-AMC were determined using a concentration of 5×10^{-10} M of thrombin, with the concentrations of both substrates (S-2238 and Tos-GPR-AMC) being varied from between 5 and 25 μM. The initial velocities obtained (five points) were fitted to the Michaelis-Menten equation using linear regression (Segel, 1975).

5.4.4. Haemadin cleavage and kinetics

Haemadin was cleaved between residues Asp 40I and Pro 41I using the formic acid method to produce a truncated form; haemadin (1-40) which has been isolated to homogeneity as described below. A solution of haemadin (175 μg ml^{-1}) had formic acid added until it was a 70% (v/v) formic acid solution and was heated at 40 °C for 48 h. The resultant solution was diluted with twenty volumes of water and applied to a reverse phase source 5RPC ST column (Amersham Pharmacia Biotech), where it was eluted using a linear gradient from 10 to 90 % acetonitrile/water 0.1% TFA on an HPLC system. The resultant protein was judged pure from protein sequencing data and from mass spectral analysis (*m/z* 4198.7, M_{calc} 4200.1). The concentration of the inhibitor as determined for usage in the experiments described below was determined by UV absorption using a UVIKON 943 double beam spectrophotometer. The absorption coefficient for the inhibitor ($e_{280\ nm}$ = 0.696 mg^{-1} cm^2) was determined using molecular absorption coefficients of 1280 M^1 cm^{-1} and 120 M^1 cm^{-1} for tyrosine and cysteine residues respectively (Gill and von Hippel, 1989). In a similar way the concentration of intact haemadin was determined for use in the experiments described using the same method, with the absorption coefficient ($e_{280\ nm}$ = 0.468 mg^{-1} cm^2) being calculated in the same manner.

K_i values for human α-thrombin-haemadin (1-40) complex were determined as follows: Solutions of 5×10^{-10} M of thrombin in 50 mM Tris-HCl (pH 8.3), 50 or 500 mM NaCl and 0.1% PEG6000 at 25°C with varying concentrations of haemadin (1-

40) were taken, and the assays were started by the addition of substrate. Initial rate velocities were recorded at four different substrate concentrations (20 to 125 $\times 10^{-6}$ M) and used to construct Dixon plots. The relationship of (initial velocities)$^{-1}$ *versus* inhibitor concentration (1 to 4 $\times 10^{-8}$ M (four points)) were analysed using linear regression, with the K_i values being determined by standard procedures using the equation which describes competitive inhibition (Segel, 1975). Associative rate constant (k_{on}) values for the thrombin-haemadin (1-40) complex were determined using pseudo-first order kinetics: A 500 µl volume of a solution of 1×10^{-8} M of haemadin (1-40) in 50 or 500 mM Tris-HCl (pH 8.3), 50 mM NaCl and 0.1% PEG6000 was taken, and the assay was started by adding thrombin to a final concentration of 1×10^{-9} M. Every 5 min a 25 µl aliquot was taken and added to 225 µl l of 217 µM S-2238 in the same buffer as above, where the initial velocity was measured. Alongside this reaction a reaction without inhibitor was performed as a control. The initial velocity is proportional to the amount of uninhibited thrombin present in the assay, and obeys the equation $V_i = V_0 \exp(-k_{app}t)$ (Myles *et al.*, 1998). Nonlinear regression on velocities taken against time was fitted to this equation to give a value for k_{app}, and the k_{on} value was determined by dividing k_{app} by the inhibitor concentration.

5.4.5. DIP-thrombin preparation and kinetics

DIP-thrombin was prepared by a modification of the method of Stone *et al.* (Stone *et al.*, 1987). A solution of human α-thrombin (12 µM) in 55 mM NaCl, 0.5 mM MES pH 6.0 was incubated at room temperature with 1 mM diisopropyl fluorophosphate (DIPF) for 30 min. After this time, an aliquot was taken and tested for thrombin activity using the chromagenic substrate S-2238. If the thrombin activity was greater than 0.01% of the activity of the original solution, more DIPF was added until the concentration was 1 mM greater than previously, and after 30 min the solution was reassayed for thrombin activity. This procedure was repeated until the thrombin activity was 0.01% of the original activity or below. The thrombin solution was then exhaustively dialysed against 50 mM Tris-HCl (pH 8.3), 50 mM NaCl and 0.1%

PEG6000, and its concentration was determined by measuring its absorption at 280 nm according to the method of Fenton *et al.* (Fenton *et al.*, 1977).

The inclusion of DIP-thrombin into assays of steady state rates against inhibitor concentration (conducted as described previously in section 5.4.3., using 500 µM S-2238) caused an increase in the K_i, which was linear when plotted against the concentration of DIP-thrombin added. The variation of the value of the dissociation constant obeyed equation [2] as would be expected for a competitive ligand. The K_i, derived from the steady state experiments without DIP-thrombin, divided by the slope of this plot (from 0 to 50 nM DIP-thrombin (6 points)), gave a value of the K_i for the DIP-thrombin-haemadin complex.

5.4.6. Kinetics of haemadin peptides and sTTI

K_i values for human α-thrombin-peptide (or sTTI) complexes were determined as follows: Solutions of 6.1×10^{-9} M of thrombin (or thrombin K217E for TTI) in 50 mM Tris-HCl (pH 8.3), 100 mM NaCl and 0.1% PEG6000 at 25°C with varying concentrations of each peptide were taken, and the assays were started by the addition of substrate. Initial rate velocities were recorded at a substrate concentration of 150×10^{-6} M and used to construct Dixon plots. The relationship of (initial velocities)$^{-1}$ *versus* inhibitor concentration (0 to 500×10^{-6} M (seven points)) were analysed using linear regression, with the K_i values being determined by standard procedures using the equation which describes competitive inhibition (Segel, 1975).

5.5. Biochemistry of sTTI with thrombin

5.5.1. Gel filtration experiments on the bovine α-thrombin-sTTI complex

Gel filtration experiments on the bovine α-thrombin-sTTI complex were carried out on a Smart system equipped with a Superdex 75 pc 3.2/30 column. The runs were carried out using a concentration of 5 mg/ml protein in PBS buffer (0.01M phosphate, 0.138 M NaCl, 0.0027 M KCl; pH 7.4) the same buffer being subsequently used for the elution at 80 µl/min with a run time of 60 min. Detection of the eluted proteins was at 280 nm and bovine serum albumin was used to calibrate the equipment prior to use.

5.5.2. Competition experiments between sTTI and triabin

Solutions of 6.1×10^{-9} M of thrombin in 50 mM Tris-HCl (pH 8.3), 100 mM NaCl and 0.1% PEG6000 at 25°C containing 6.1×10^{-7} M sTTI were treated with 6.1×10^{-9} M triabin, and the rates of the reaction monitored at certain time intervals up to one hour.

5.6. X-ray analytical methods

5.6.1. Crystallisation

In order to test a large range of conditions the principle of sparse matrix sampling was used (Carter and Carter, 1979; Jancarik and Kim, 1991). Crystallisation was achieved using the vapour diffusion method with hanging or sitting drops (McPherson, 1982) in airtight 24 well CrysChem polystyrene plastic plates. Normally 3 µl droplets consisting of 2 µl of protein solution and 1 µl of precipitant were equilibrated against 500 µl of precipitant solution at 4ºC or 22ºC, and the wells were sealed using clear adhesive tape. Wells were checked regularly using a Leica MZ 12 microscope and the condition of the drops noted.

5.6.2. Data collection

In order to measure diffraction data crystals were mounted in 1 mm thick silanised quartz capillaries. After placement of the crystal, in solution, in the capillary, the excess solution was removed with exceptionally thin pieces of filter paper. The capillary was at its ends sealed with wax, incorporating a little of the well solution at one end to prevent the crystal drying out during measurement, which was conducted at 4ºC. For cryo-data collection the crystals were captured in nylon loops, transferred into a cryo-buffer (usually the well buffer of the drop the crystals grew in plus ethylene glycol, glycerol, glucose, polyethylene glycol, MPD or sucrose) recaptured in the same loop, and then frozen immediately in liquid nitrogen. At all times when the crystals were being mounted or measured, they were present in a stream of frozen nitrogen gas at 77 K.

Data sets were collected as a series of rotation pictures with a turning angle of 1° over a period of 20 minutes. In this way the crystal was subjected to graphite monochromatised Cu-K_α X-rays (λ = 1.5418 Å) generated from a Rigaku rotating anode and diffraction signals were recorded on an electronic image plate.

5.6.3. Data analysis

The crystal orientation, space group and elementary cell parameters as well as the integration of the reflection intensities were evaluated using the using the MOSFLM package (Leslie, 1994). Data reduction was handled by programs supported by the CCP4 program suite (Collaborative Computational Project Number 4, 1994).

5.6.4. Structure determination

The structure was solved using Patterson search techniques. Rotational and translational searches to determine the orientation and position of the thrombin components were performed with the program AMoRe (Navaza, 1994). For these searches some of the residues from the A-chain of thrombin were occluded as well as most of the residues forming the autolysis loop (Glu146 to Glu149E).

5.6.5. Model building and refinement

Atomic models of omitted parts of thrombin, as well as the segments of the inhibitors in contact with thrombin, were built using the program MAIN (Turk, 1992) into Fourier maps calculated after appropriate positioning of the thrombin molecules in the crystal cell. The complexes were crystallographically refined using the conjugate gradient method with molecular dynamics using the program X-PLOR (Brünger, 1991), with the crystallographic R-factor as a target function. Target geometric parameters (bond angles and lengths) were the parameters of Engh and Huber (Engh and Huber, 1991). In the case of the thrombin sTTI complex detwinning using the built structure was performed between refinement steps, with the new fob file used for the subsequent refinement. Following refinement of coordinates individual temperature factors (*B* factors) were refined. This process was cyclically repeated a few times until no more inhibitor sequence could be fitted into the density. Water molecules were added at stereochemically reasonable positions where the ($2F_{obs}$ - F_{calc})

density and (F_{obs} - F_{calc}) difference density maps were above 1σ and 2σ levels, respectively.

5.6.6. Analysis of the atomic model

Statistical analysis of the bond lengths and angles as well as the temperature factors was carried out using the program CNS version 1.1 (Brünger *et al.*, 1998). Stereochemical analysis of the atom model was accomplished via Ramachandran plots using the program PROCHECK Version 2.0 (Laskowski *et al.*, 1993).

5.6.7. Graphical representation

Graphical representations of the molecular models were achieved using the programs SETOR (Evans, 1990), for representation of the molecular surface the program GRASP (Nicholls *et al.*, 1993) was used. In addition, for formatting, croping and labelling, the programs GHOSTSCRIPT (Aladin Enterprises, USA), X-VIEW (J. Bradley, Pennsylvania, USA), PHOTOSHOP (Adobe Systems Inc., USA) and CorelDRAW (Corel Corperation Limited, 1998) were used.

6. RESULTS

6.1. The human α-thrombin-haemadin complex

6.1.1. Crystallisation

Crystals of the human α-thrombin-haemadin complex of approximate dimensions 0.3 × 0.15 × 0.15 mm^3 were grown by vapour-diffusion using the sitting drop method. Droplets of 3 µl consisting of 2 µl of protein solution and 1 µl of precipitant (100 mM Na citrate pH 5.6, 14% (w/v) PEG4000, 12.5% (v/v) isopropanol), were equilibrated against 500 µl of precipitant solution at 22ºC. Crystals appeared after 2 days and achieved maximum size after 1 week. Occasionally precipitation seeding was required to produce diffraction quality crystals. This consisted of leaving the drop in the open air until precipitation occurred around the edge of the drop. Subsequent sealing of the drop caused the precipitate to disappear again (probably due to reuptake of isopropanol), but would leave behind minute microcrystals to act as seeds. These seeds would often result in a multiplicity of crystals, but with carefully controlled conditions about one in fifty wells would contain a single crystal. Figure 9 below shows a crystal used for measurement, at two stages of growth.

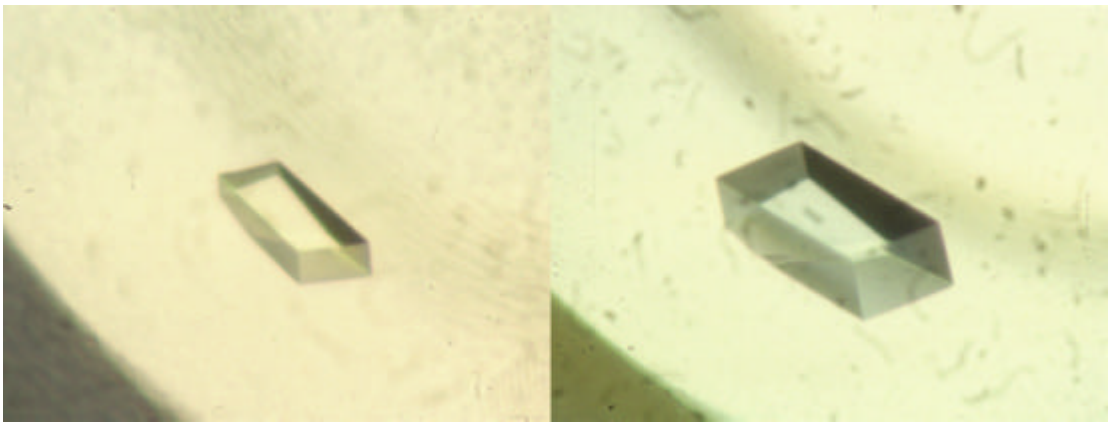


Figure 9: Photographs of a crystal of the human α-thrombin-haemadin complex taken under a microscope. The pictures are of a crystal used for measurement at different stages of growth with final dimensions of $0.3 \times 0.15 \times 0.15$ mm^3. Notice how the single crystal is stuck to the bottom of the crystallisation well.

6.1.2. Data collection

Data was collected form two crystals of the human α-thrombin-haemadin complex at 4°C. The crystals initially diffracted to 2.8 Å resolution but radiation damage caused rapid decay to 3.1 Å resolution. Cryo-cooling could not attenuate this damage, as cryo-cooling irreversibly damaged the crystals with the eight cryoprotectants tested so far. A picture of the diffraction pattern is shown below in figure 10, and the data processing statistics are shown in table 4.

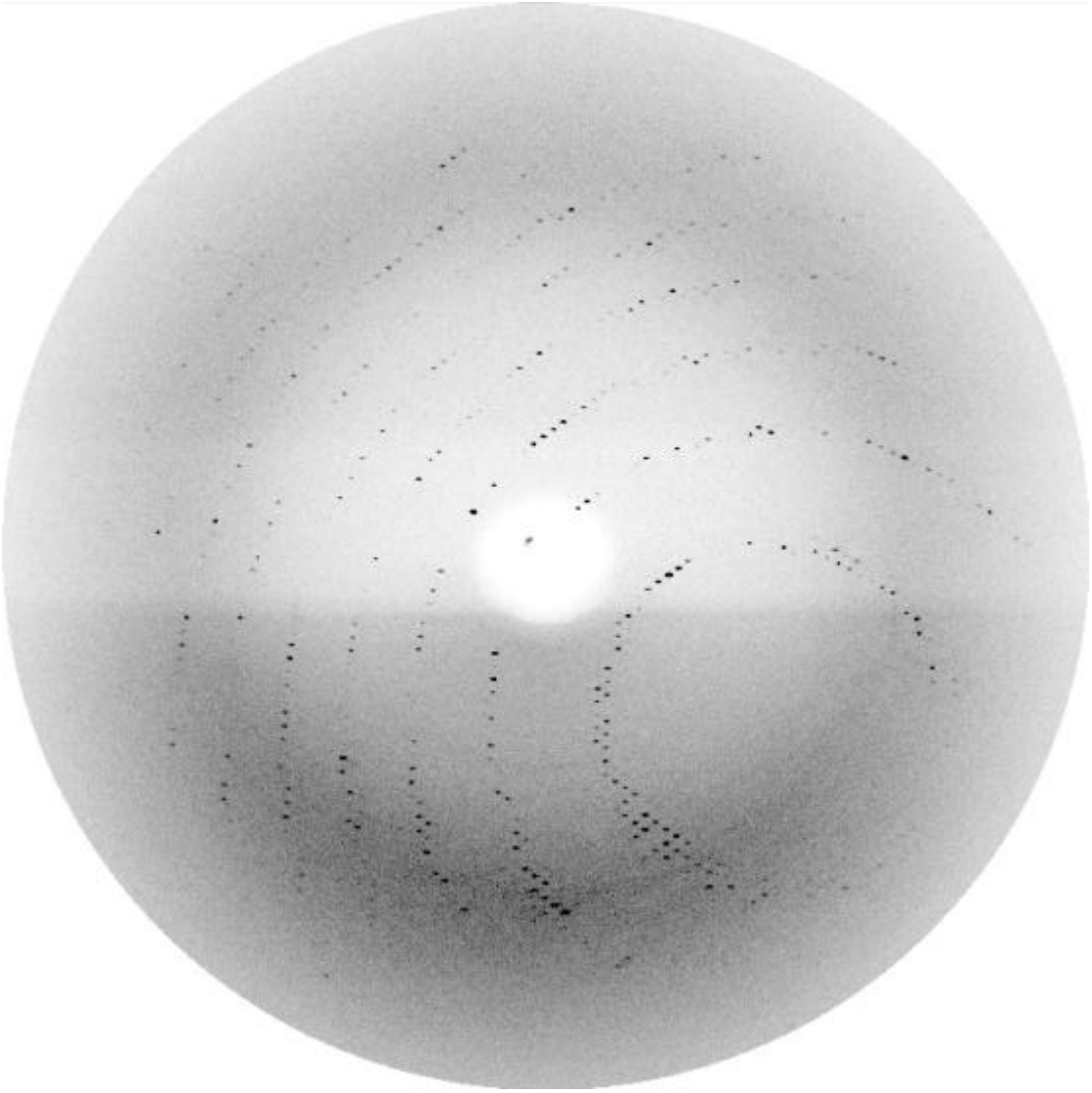


Figure 10: Image plate image of the diffraction pattern from a crystal of the human α-thrombin-haemadin. The crystal was rotated through 1° for twenty minutes, with the outer edge of the diffraction pattern extending to 2.8 Å.

Table 4: Data processing statistics of the human α-thrombin-haemadin complex.

Crystal system	Monoclinic
Space group	P2
Cell constants	*a* = 121.67 Å
	b = 50.57 Å
	c = 129.74 Å
	α = 90°
	β = 114.76°
	γ = 90°
Cell volume	724886
V_m (Å^3/Da)	3.03 (59% solvent)
Number of crystals	2
Limiting resolution, Å	3.1
R_{merge}, %[1]	10.9
Reflections collected	126,754
Unique reflections	23,938
Completeness (∞ to 3.1 Å), %	81.2

[1] $R_{merge} = [S_h S_i|I(h,i) - \langle I(h)\rangle| / S_h S_i I(h,i)] \times 100$, where I(h,i) is the intensity value of the i-th measurement of h and <I(h)> is the corresponding mean value of h for all i measurements of h. The summation is over all measurements.

6.1.3. Matthews parameter

The Matthews parameter is a measure of volume per molecular mass that this volume occupies. It acts as a measure of the packing density for the crystal, having units of Å^3/Da. From this the contents of the asymmetric unit can be estimated, as proteins have relatively uniform densities, and their crystals usually contain from 27 to 65% solvent (which converts to Matthews parameters from between 1.7 to 3.5 Å^3/Da). Normally the solvent content of the crystal is estimated to be 50%, and the number of

molecules that this is closest to, is taken to be the number of molecules in the asymmetric unit. In the case of the human α-thrombin-haemadin complex the solvent content was 59% H_2O for three complexes per unit cell, with a Matthews parameter of 3.03 $Å^3$/Da.

6.1.4. Patterson search

The structure of the human α-thrombin-haemadin complex was solved using Patterson search techniques. As a search model the coordinates of human α-thrombin from the thrombin-hirudin complex (Rydel *et al.*, 1990), from which residues Thr1H to Asp1A and Asp14L to Arg15 of the A-chain, and most of the residues forming the autolysis loop (Glu146 to Glu149E) were excluded was used. Rotational and translational searches to determine the orientation and position of the thrombin components were performed with the program AMoRe (Navaza, 1994) using diffraction data from 10.0 to 3.5 Å. As expected from the Matthews parameter for a trimer, three rotation solutions were found. After translational search and rigid body refinement, these solutions resulted in a correlation value of 0.67 and an R-factor of 0.34 (next highest peak: c = 0.46, R = 0.44).

6.1.5. Model building and refinement

The human α-thrombin-haemadin complex was built using the program MAIN (Turk, 1992). Into an interpretable 3.1 Å ($2F_{obs}$-F_{calc}) Fourier map calculated after appropriate positioning of the thrombin molecules in the crystal cell, some of the omitted parts of thrombin and the segments of haemadin in contact with thrombin were built. The complexes were then crystallographically refined with X-PLOR (Brünger, 1991) using the target parameters of Engh and Huber (Engh and Huber, 1991), and consisted of normal crystallographic refinement, B-factor refinement, as well as simulated annealing at 2500 K. This process was repeated cyclically a few times until most of the inhibitor sequence could be fitted into the density. Sixty-eight water molecules were added at stereochemically reasonable positions where the ($2F_{obs}$-F_{calc}) density and

(F_{obs}-F_{calc}) difference density maps were above 1σ and 2σ levels, respectively. The *R* factor of the final model was 0.208 (with a free *R* factor of 0.255). The refinement statistics are given in table 5. The atomic coordinates of the complex were deposited in the Protein Data Bank with the entry code 1E0F.

Table 5 : Refinement statistics for the human α-thrombin-haemadin complex

Resolution range, Å	10.0-3.1
Reflections used for refinement	22,278
R-value, %[1]	20.8
R_{free}, %[2]	25.5
Non hydrogen protein atoms	8,381
Solvent molecules	68
Standard deviations:	
Bond lengths, Å	0.006
Main chain bond angles, °	1.52
Side chain bond angles, °	23.55
Improper angles, °	1.17
No. of residues excluding glycine and proline	836
(Φ,Ψ) angle distribution in	
Most favoured region	648 (77.2%)
Additionally allowed region	168 (20.1%)
generously allowed region	13 (1.6 %)
Disallowed region	10 (1.1%)
Average *B*-factor	30.4 ± 16.7
Min *B*-factor	8.00
Max *B*-factor	79.66
Occupancy	96.7%

[1] *R*-value = $\{S(|F_o|-|F_c|) / S|F_o|\} \times 100$, where $|F_o|$ is the observed and $|F_c|$ the calculated structure factor amplitude of reflection *hkl*.

[2] R_{free} was calculated randomly omitting 5% of the observed reflections from refinement.

6.1.6. Quality of the model

The refined model consist of 3 human α-thrombin-haemadin complexes containing 8,381 non hydrogen atoms and 68 water molecules. Below (figure 11) is the Ramachandran plot of the final model.

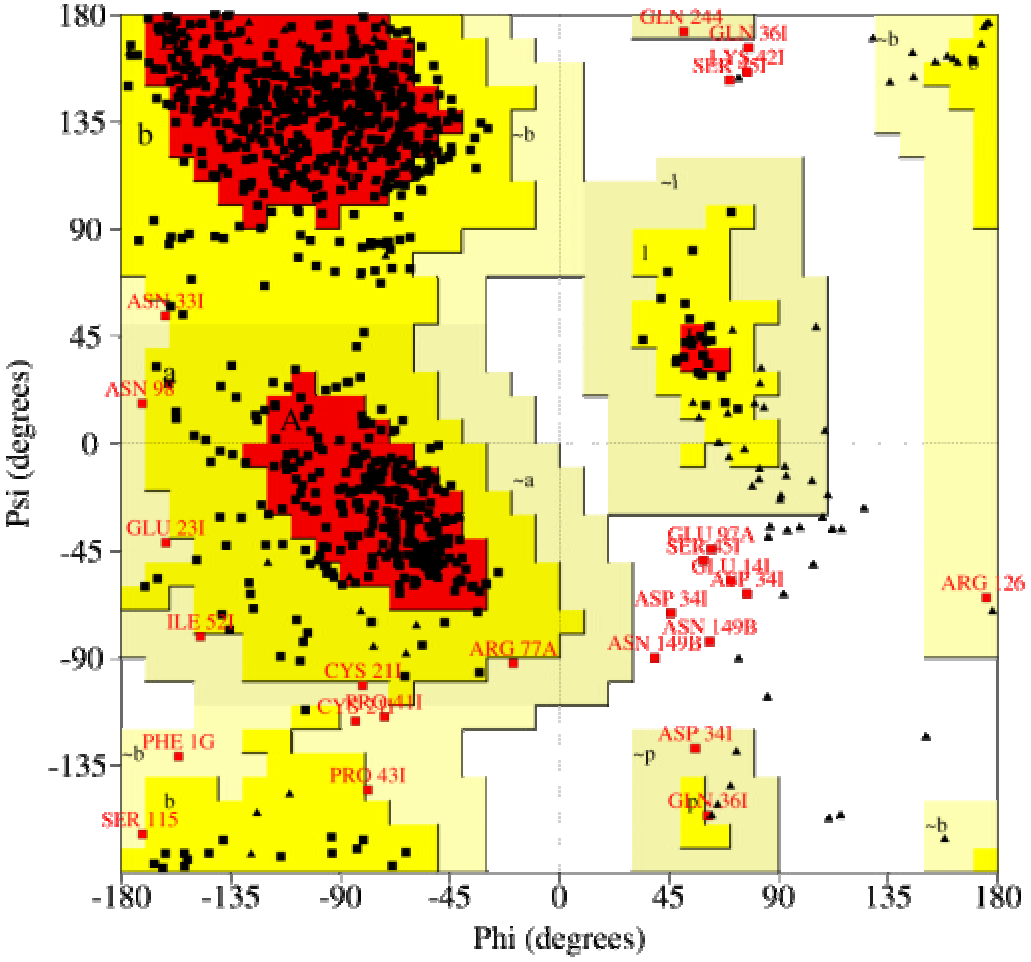


Figure 11: Ramachandran plot of the Phi and Psi angles of the atomic model of the human α-thrombin-haemadin complex.

6.1.7. Molecular packing and inter molecular contacts

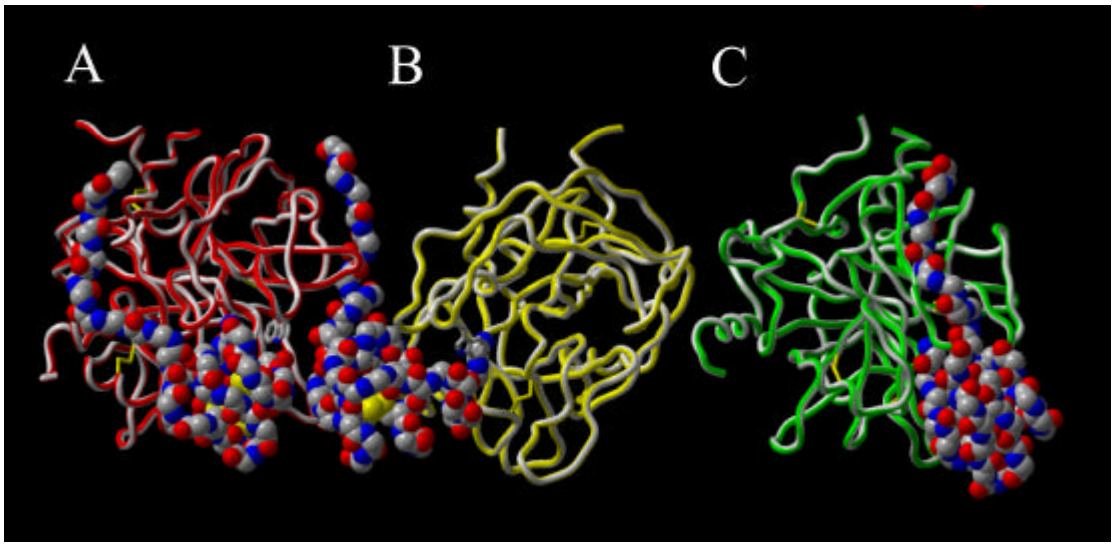


Figure 12: Crystal structure of the human α-thrombin-haemadin complex. Structure of the crystallographic trimer present in the asymmetric unit. Monomers are labelled A, B and C. Thrombin molecules are shown as red, yellow and green ribbons; the Cα traces of the three inhibitors are presented as colour-coded van der Waals spheres (red, oxygen; blue, nitrogen; grey, carbon).

The human α-thrombin-haemadin complex crystallises as a trimer in the monoclinic space group P2, with the crystallographically independent complexes (denoted A, B and C) aligned almost parallel to the a-c plane (See figure 12). In all three complexes, the main body of the inhibitor binds to the target proteinase adopting extremely similar conformations, while its C-terminal tail shows slightly different structures (See figure 13). This remarkably acidic peptide of haemadin sandwiches between the heparin-binding exosite of the cognate thrombin molecule and the fibrinogen-recognition exosite of a neighbouring molecule, and is thus essential for formation of the crystallographic A-B-C trimer. As seen in Figure 12, the C-terminal tail of haemadin in complex B fills the A-B interface by simultaneously binding to its cognate thrombin moiety (B) and to the neighbouring thrombin molecule (complex A). Similar contacts are made between the haemadin peptide of the boundary complex A and the symmetry-related molecule C', while the tail of haemadin molecule C contacts both its "own" thrombin (C) and the symmetry-related molecule A' (not depicted in figure 12). Strong thrombin-thrombin contacts involving both charged and hydrophobic residues of the adjacent antiparallel stretches Ser 11 to Ile 14K (light chains, chymotrypsinogen numbering) (Bode *et al.*, 1992) and Ser 20 to

Gly 25 (heavy chains) in complexes B and C further stabilise the asymmetric unit trimer. Figure 14 shows how six of the human alpha thrombin complex join together to form a hexamer which is not the biologically relevant complex. In the crystal these hexamers form a two dimensional layer rather like hexagons do on a parquet floor. Additionally the hexameric rings stack directly on top of each other, and due to the fact they contain a cavity in the centre they form giant channels that extend through the crystal.

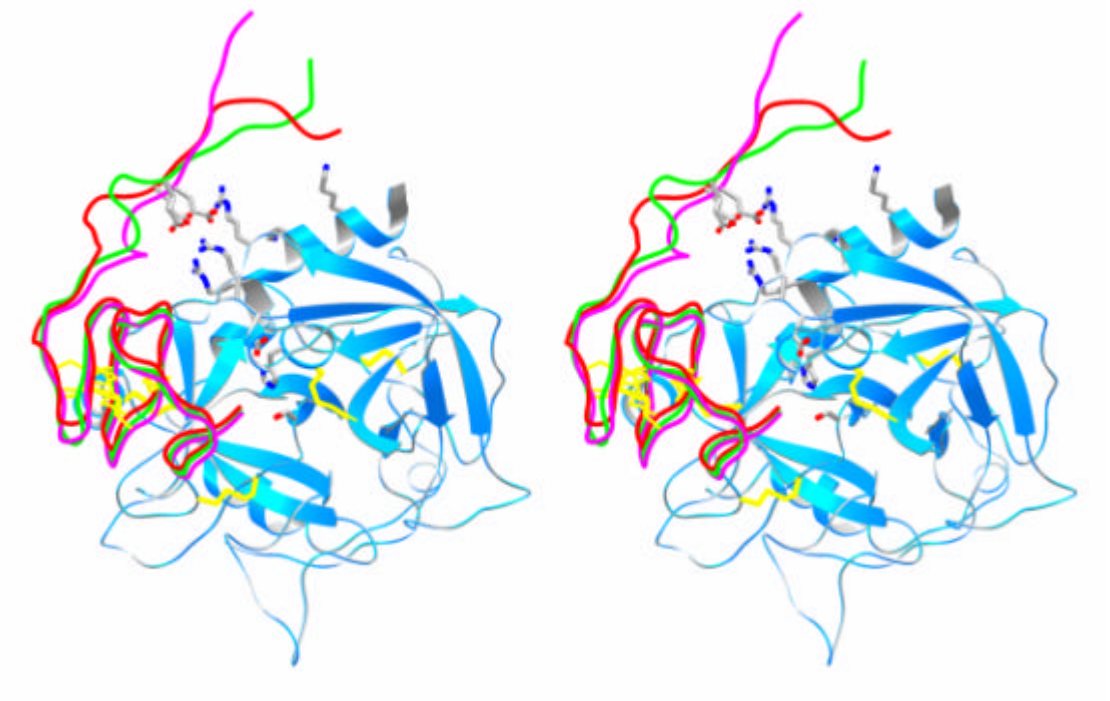


Figure 13: Stereo view of a ribbon representation of an overlay with respect to thrombin of the three haemadin molecules present in the asymmetric unit. Thrombin is in the 'standard orientation' (see figure 2), and is shown in blue. The three inhibitors are shown in purple green and red. Shown specifically are the side chains of the catalytic triad as well as some of the basic residues of exosite II.

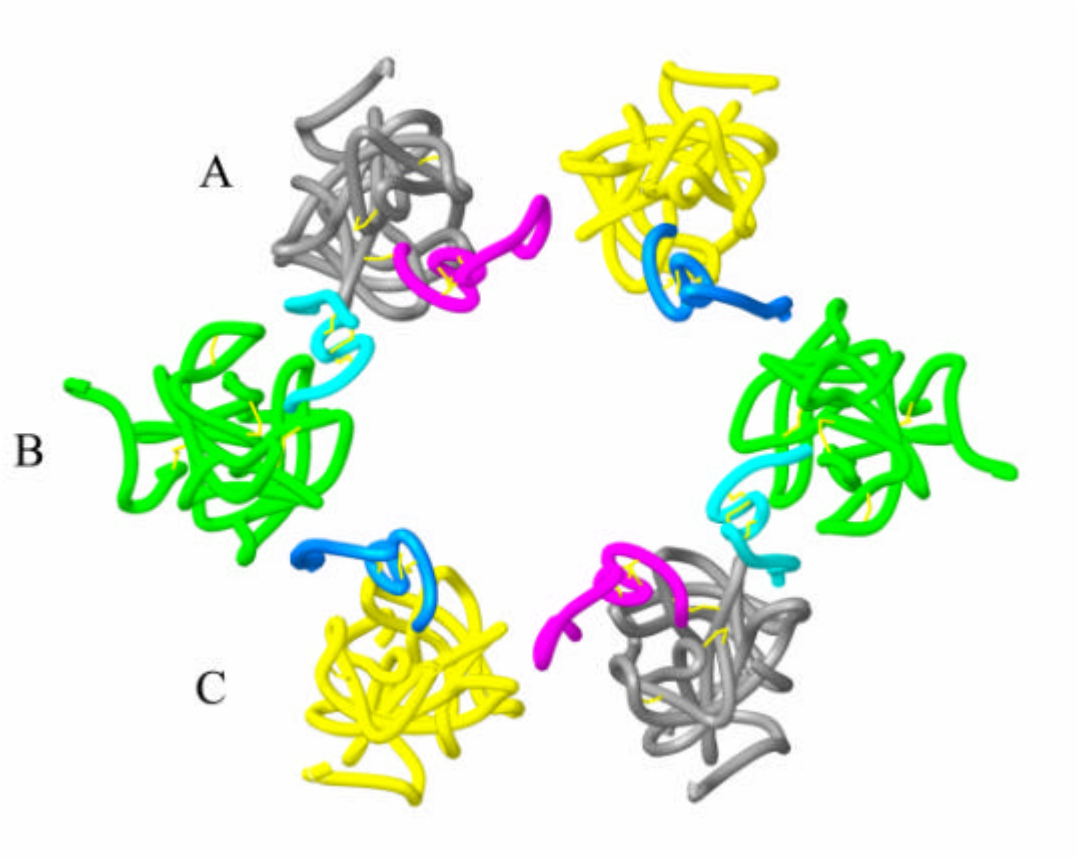


Figure 14: The hexameric rosette of the human α-thrombin-haemadin complex. Thrombins are shown in yellow, grey and green, with haemadin shown in purple, blue and cyan. The letters A, B and C denote the complexes of the asymmetric unit, similar coloured complexes being the symmetry complexes. This hexameric arrangement is not the physiological relevant species.

6.1.8. Structure description

6.1.8.1. Overall structure

However, the remarkable polymeric arrangement of the thrombin-haemadin complexes as described in the section molecular packing and inter molecular contacts does not seem to be relevant in diluted protein solutions and at more physiological conditions. Evidence discussed in section 6.3. below entitles us to assume that the solution structure of the thrombin-haemadin complex corresponds to the monomeric 1:1 complex shown in more detail in figure 15.

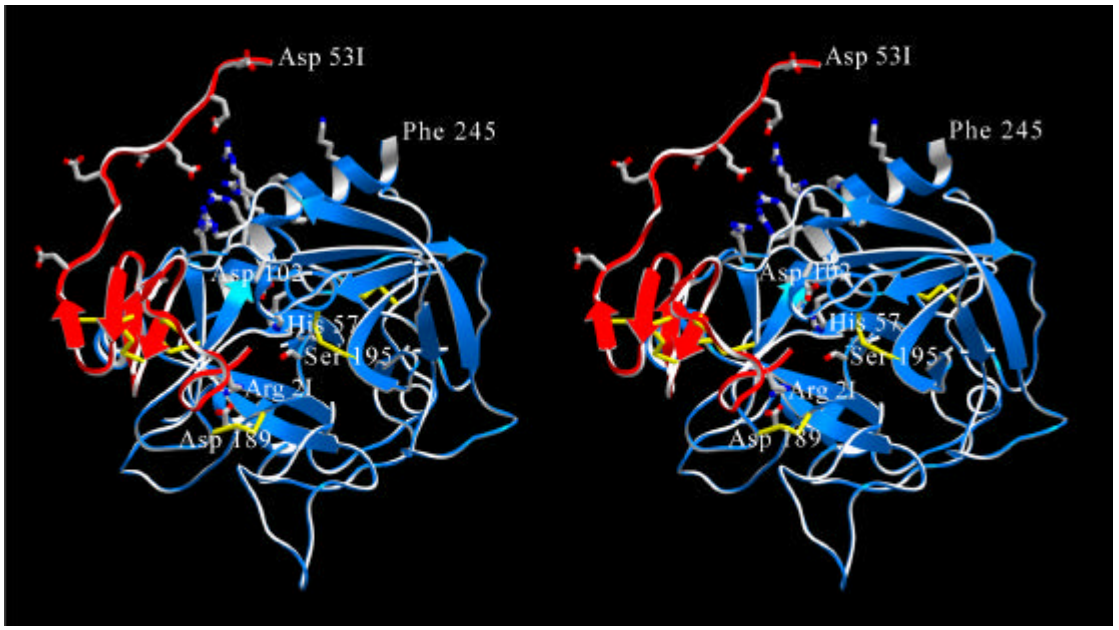


Figure 15: Stereo diagram of complex molecule A. The proteinase is shown in its 'standard orientation' (see figure 2). Side chains of the catalytic triad residues are explicitly shown, as well as the side chains of the interacting residues Asp189 (thrombin) and Arg 2I (haemadin) (colour coded as in figure 12). Also shown (unlabelled) are the side chains of the basic residues of the heparin-binding site (thrombin), as well as the side chains of the acidic residues of haemadin's C-terminal tail.

6.1.8.2. Structure of thrombin

The structure of haemadin-bound thrombin is similar to that of thrombin in the human α-thrombin-hirudin structure (Grütter *et al.*, 1990; Rydel *et al.* 1990); this also applies to the active site catalytic triad, which is almost identical in both structures. Large deviations in the main chain are essentially limited to the so called "autolysis loop" or 149-loop (residues Thr 147 to Lys 149E). This loop possesses a well-reported conformational flexibility (Bode *et al.*, 1992) and has missing or very weak electron density for up to four residues in the three thrombin-haemadin complexes. Electron density is also weak or missing in the light chain termini (residue Thr1H of complex C as well as residues Gly14M and Arg15 of complex A) and in the C-terminus of the heavy chain (residues Glu 247 of all complexes, Gly 246 of complexes A and B, and Phe 245 of complex B). Weak density accounting for the first *N*-acetyl-glucosamine moiety attached to the amide group of Asn 60G is present in all three complexes.

6.1.8.3. Structure of haemadin

Haemadin residues Pro 9I to Ser 38I (the suffix "I" identifies haemadin residues) are folded into a compact ellipsoidal core of approximate dimensions $15 \times 16 \times 18$ $Å^3$. This core contains five short β-strands (β1 to β5), which are arranged in two antiparallel distorted sheets formed by strands β1-β4-β5 and β2-β3 facing each other (for a schematic drawing of haemadin's primary and secondary structures, see figure 16). This β-sandwich is stabilised by six enclosed cysteines arranged in a [1-2, 3-5, 4-6] disulphide-pairing (See figure 17), resulting in a disulphide-rich hydrophobic core that is largely inaccessible to bulk solvent. The close proximity of the disulphide bonds [3-5] and [4-6] organises haemadin into four distinct loops. The largest of these loops (loop A, residues Pro11I to Thr18I) contains β-strand β1 (Gly 13I to Val 15I), which hydrogen-bonds residues Gln 30I to Cys 32I (β4). The shorter loops B (residues Gly 22I to Ile 25I), C (Leu 27I to Ser 31I) and D (Asn 33I to Gln 36I) are essentially β-hairpin turns connecting consecutive β-strands β2-β3, β3-β4 and β4-β5, respectively.

The conformations adopted by both the N- (Ile 1I to Val 8I) and especially the C-terminal segments (Gly 39I to Lys 57I) seem to be largely determined by contacts to the cognate thrombin and/or to neighbouring molecules in the crystal (see below). In particular, the acidic C-terminal tail protrudes away from the main body of the inhibitor (See figures 12, 15 and 17). This extended stretch lacks secondary structure and adopts slightly different conformations in each of the three crystallographically independent complexes (See figure 13). Moreover, the last four C-terminal residues Glu 54I to Lys 57I, are not defined in the electron density, suggesting enhanced conformational flexibility.

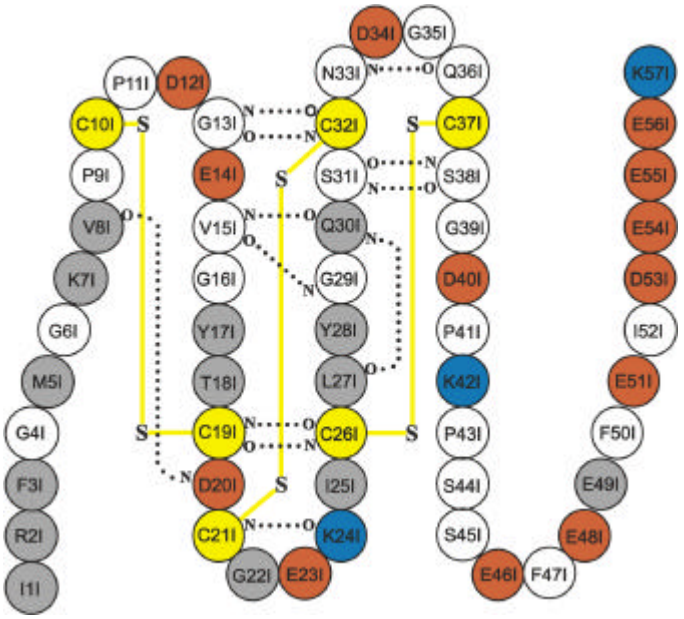


Figure 16: Primary and secondary structure of haemadin. Acidic, basic and cysteine residues are represented by red, blue and yellow circles, respectively. Main chain to main chain hydrogen bonds formed in at least two of the three independent molecules forming the asymmetric unit are indicated with dotted lines. Disulphide bridges are shown with yellow lines. Residues involved in interactions with thrombin in the crystal structure are represented by grey circles.

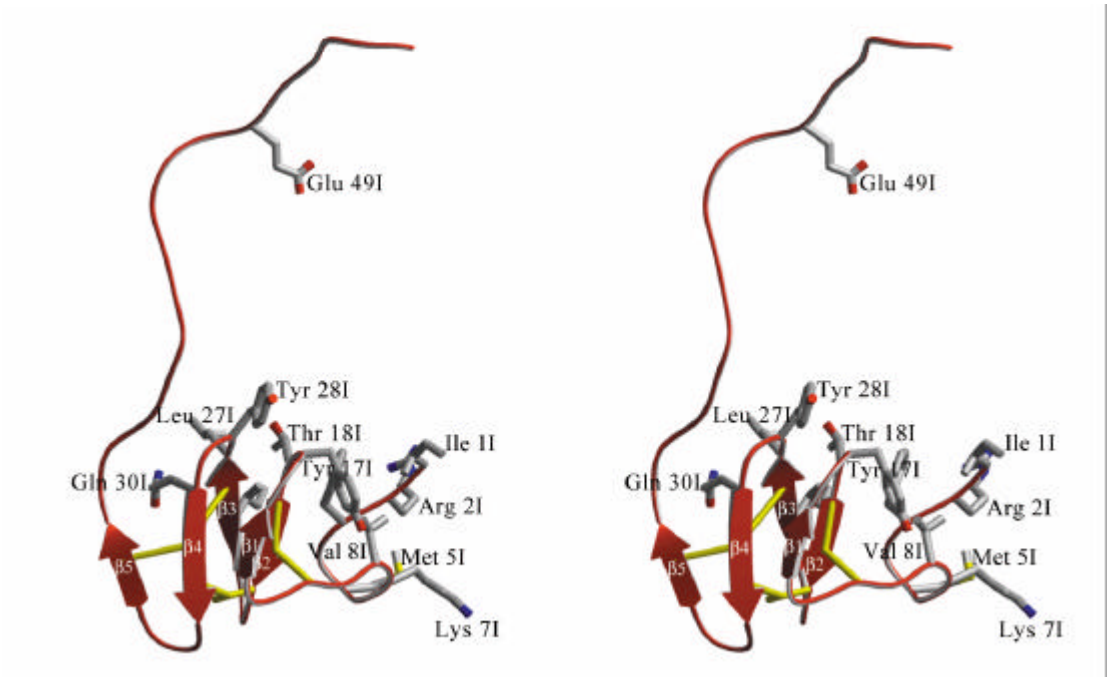


Figure 17: Ribbon diagram of haemadin's structure, highlighting elements of secondary structure and the disulphide-rich core. Side chains of residues involved in major interactions with thrombin are explicitly shown (colour-coded as in figure 12).

6.1.8.4. Binding of the inhibitor

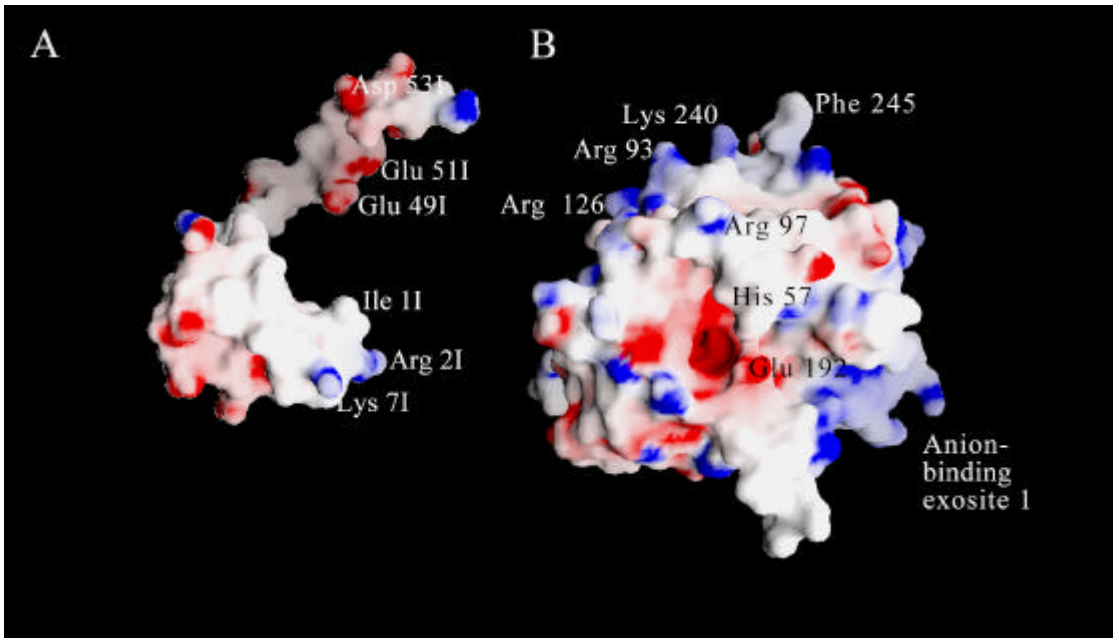


Figure 18: Space-filling models of human α-thrombin and haemadin, showing the surface potential of the two molecules. Positive charges are displayed in blue, negative charges in red, with darkest blue and red colours corresponding to electrostatic potentials beyond -10 kT/e and +10 kT/e, respectively. The thrombin component (B) is shown in the 'standard orientation' (see figure 2), while haemadin (A) is rotated along the *y* axis to present the thrombin binding surface to the viewer. Some of the residues of the interacting interfaces are labelled.

Analysis of the interacting surfaces of haemadin and thrombin reveals a striking complementarily of their electrostatic surface potentials (See figure 18), which is most certainly a major driving force for complex formation. Solvent accessible molecular surfaces of 780 $Å^2$ and 200-230 $Å^2$ are buried at the active-site and exosite II regions, respectively. Beyond the association of thrombin's highly basic heparin-binding exosite and the acidic C-terminal tail of haemadin, also the negatively charged active site region is compensated by the N-terminal alkaline side chains of Arg 2I and Lys 7I. In the complex, the first three N-terminal residues of haemadin are inserted into the active site cleft of thrombin, forming a parallel β-sheet with thrombin residues Ser 214 to Gly 216. Asides from several intermolecular main chain to main chain hydrogen bonds that stabilise this β-sheet, the side chains of these three residues make numerous contacts with thrombin's active site (See figures 19 and 20).

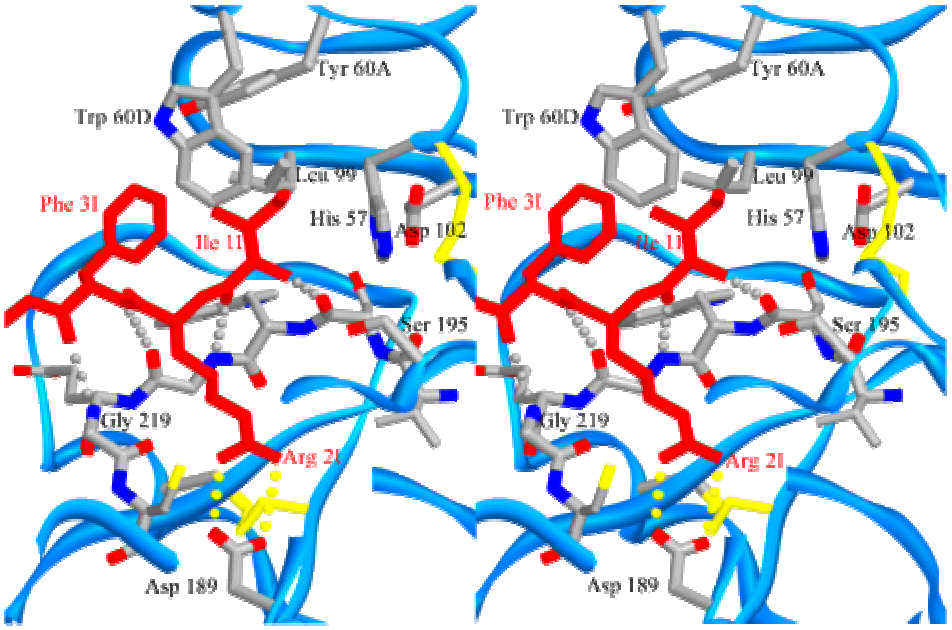


Figure 19: Close-up stereoview of the active site cleft of haemadin-bound human α-thrombin. Side chains of selected residues are depicted colour-coded. Intermolecular hydrogen bonds are shown as white dotted lines for the parallel interaction of haemadin's N-terminal loop with thrombin residues Ser 214 to Gly 216 or yellow dotted lines for contacts made inside the S1 pocket. Some residues of thrombin and the inhibitors along with all water molecules are omitted for the sake of simplic ity.

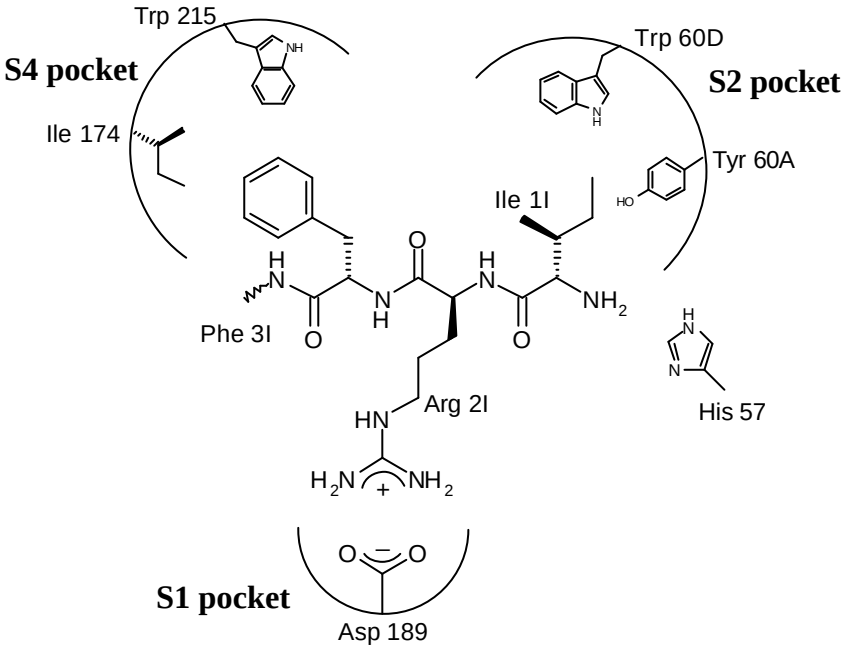


Figure 20: Schematic diagram of the interactions between the first three residues of haemadin and the active site of human α-thrombin. Haemadin residues are represented as their chemical line drawing formulas. Thrombin's non-primed part of the active site is schematically represented showing explicitly the side chains of the residues with which haemadin makes major contacts.

Close to the active site of thrombin is Ile 1I, with its isobutyl group occupying thrombin's S2 pocket, while its amino group hydrogen-bonds the carbonyl of Ser 214. The protruding side chain of Arg 2I occupies the S1 specificity pocket, where it forms a hydrogen-bonded ion pair with the carboxylate of Asp 189, but enters this pocket from a position that in canonically binding inhibitors is occupied by P3 residues. In addition, Arg 2I Nη1 donates a hydrogen bond to the carbonyl of Gly 219. The phenyl moiety of Phe 3I protrudes up into the S4 pocket of thrombin; this aromatic side chain and the aliphatic Ile 1I also participate in intramolecular van der Waals contacts with Val 8I as well as with some residues belonging to the compact core of haemadin (e.g., Thr 18I, Ile 25I). These interactions stabilise the N-terminal peptide in its characteristic loop shape whilst bound to thrombin (See figures 15 and 17).

Emerging from the active site, the carbonyl of Gly 4I accepts a hydrogen bond from the side chain of thrombin's Arg 221A, while the side chain of Met 5I is interspersed between the thrombin carboxylates of Glu 146 and Glu 192. The prominent Glu 192 residue is in turn sandwiched between the side chains of Met 5I and Lys 7I, experiencing the positive electrostatic potential of the latter (See figure 18). However, the distance of about 4.1 Å between Lys 7I Nζ and Glu 192 Oε1 and their unfavourable relative orientation precludes formation of a direct hydrogen-bonded salt-bridge. Val 8I makes hydrophobic interactions with the indole moiety of Trp 60D; this insertion (60-)loop acts as a centre of hydrophobicity and places residues Phe 3I and Tyr 17I within the realms of van der Waals contacts.

Solvent-exposed side chains of haemadin's core, in particular those of loops A and C, participate in several major interactions with surface loops of thrombin that surround the 'west' and 'north' sides of the active site ('Standard orientation', see figure 2). At the end of loop A, Tyr 17I edges thrombin's 60- and 96-loops, making close van der Waals contacts with residue Pro 60C and the indole moiety of Trp 96, while the Thr 18I amide hydrogen-bonds Arg 97 O. The guanidinium group of the latter donates a hydrogen bond to Gly 16I's carbonyl, thus mimicking formation of a short parallel β-sheet. The exposed side chain of Ile 25I (loop B) makes strong van der Waals contacts with thrombin's Glu 217, as well as a weaker interaction with Ile 174. Emerging from loop C, Tyr 28I is tilted against the aliphatic parts of Arg 97 and Glu 97A. Further, the side chain of Leu 27I is in close van der Waals distance to both the

backbone and side chains of Arg 173 to Ile 174. The guanidinium group of Arg 173 donates hydrogen bonds to both the carbonyl of Leu 25I and Gln 30I Oε1.

The extended C-terminal peptide of haemadin covers the heparin-binding exosite of thrombin. This surface patch of the proteinase is dominated by unbalanced positive charges (e.g., Arg 93, Arg 97, Arg 101, Arg 233, Lys 240; see figure 18), whereas the C-terminal domain of haemadin harbours a large number of negatively charged residues (eight out of 19 residues are aspartates or glutamates). In spite of this overall electrostatic complementarity, the only strong intermolecular salt-bridges observed in the crystal are those linking the carboxylate of Glu 49I to the guanidinium groups of Arg 93 and Arg 101 (See figure 15).

6.2. Binding studies of haemadin in solution

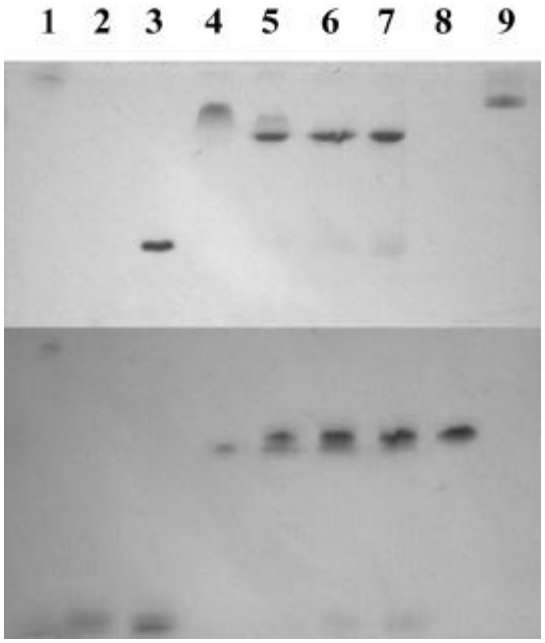


Figure 21: Binding studies in solution. (Top) Formation of the ternary complex haemadin-thrombin-triabin, as followed using non-denaturant polyacrylamide gel electrophoresis. Lane 1, human α-thrombin (5 μg); lane 2, haemadin (10 μg); lane 3, triabin (20 μg); lane 4, thrombin-haemadin complex (10 μg); lanes 5-7, 10 μg of thrombin-haemadin complex incubated with increasing amounts of triabin (1:1, 1:1.5 and 1:2 equivalents); lane 9, human α-thrombin-triabin complex (10 μg). (Bottom) The same conditions as the top gel, except that haemadin has been replaced by an equimolar amount of hirudin, lane 8 has the same conditions as lane 9, and lanes 2 and 3 have the same conditions as lane 3.

As a result of ambiguity in the X-ray crystal structure of the human α-thrombin-haemadin structure "band shift" studies were conducted to deduce the exact nature of binding of the complex. These studies were conducted on native non-reducing PAGE gels with native proteins. To achieve this a known exclusively exosite I binding protein, triabin, (See figure 4) (Fuentes-Prior *et al.*, 1997) was added to both complexes of thrombin and haemadin and thrombin and hirudin to see the results, which are shown in figure 21. As a control measure, to eliminate the possibility of new bands being other complexes, thrombin, haemadin and triabin were run alone. A further control lane used the thrombin triabin complex. Yet further experiments were conducted using meizothrombin (Martin *et al.*, 1997), whose exosite II is blocked by the second kringle domain (See figure 22) and the thrombomodulin fragment TME456 (Fuentes-Prior *et al.*, 2000) and the results of these are shown in figure 23.

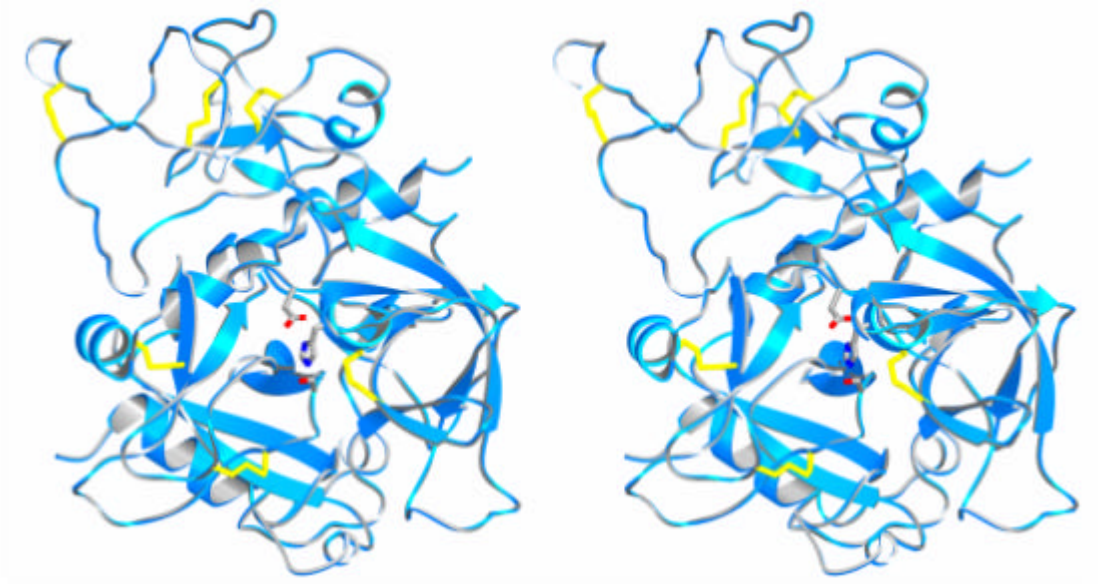


Figure 22: Stereo ribbon representation of the bovine meizothrombin showing the occluding of exosite II by the second Kringle domain. Meizothrombin is shown in 'standard orientation' (see figure 2). Shown unlabelled are the side chains of the catalytic triad.

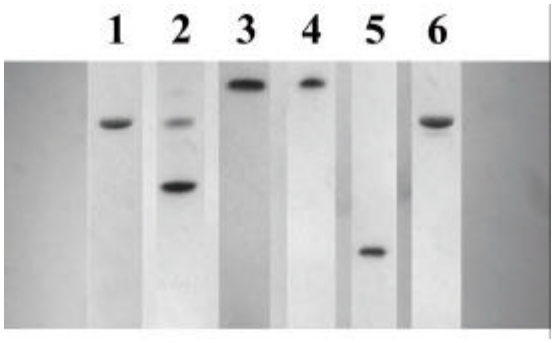


Figure 23: Haemadin binds to thrombomodulin-bound thrombin, but not to meizothrombin. Lane 1, 10 μg of human thrombin-TME456; lane 2: 10 μg of the complex incubated with 1 μg haemadin; lane 3: 10 μg meizothrombin; lane 4, 10 μg meizothrombin incubated with 1.5 μg haemadin; lane 5, 20 μg triabin; lane 6, human thrombin-triabin complex (10 μg).

6.3. Enzyme kinetics of the human a -thrombin-haemadin complex

6.3.1. Kinetic constants of haemadin with native and mutant thrombins

The kinetic constants of haemadin with human α-thrombin, and with ten other thrombin mutants, where a basic residue has been mutated to a glutamic acid (see figure 24), at an ionic strength of 0.0695 are shown in table 6. For the K_i measurements a typical tight binding assay, a plot of velocity against concentration of haemadin, is shown in figure 25, with the graph derived from it a plot of K_i, against substrate concentration being show in the same figure. For the k_{on} measurements a typical graph of a progress curve, and the graph derived from several progress curves are shown in figure 26. With the K_i and k_{on} being known for each mutant, this enabled the k_{off} for all the thrombins to be determined. These figures are shown in table 6 along with their standard deviations.

Additionally the K_i of the thrombin-TME456 complex was determined by the same method and found to be $296.3 \pm 94.2 \times 10^{-15}$ M, or 1.21 times the figure for the native human α-thrombin-haemadin complex.

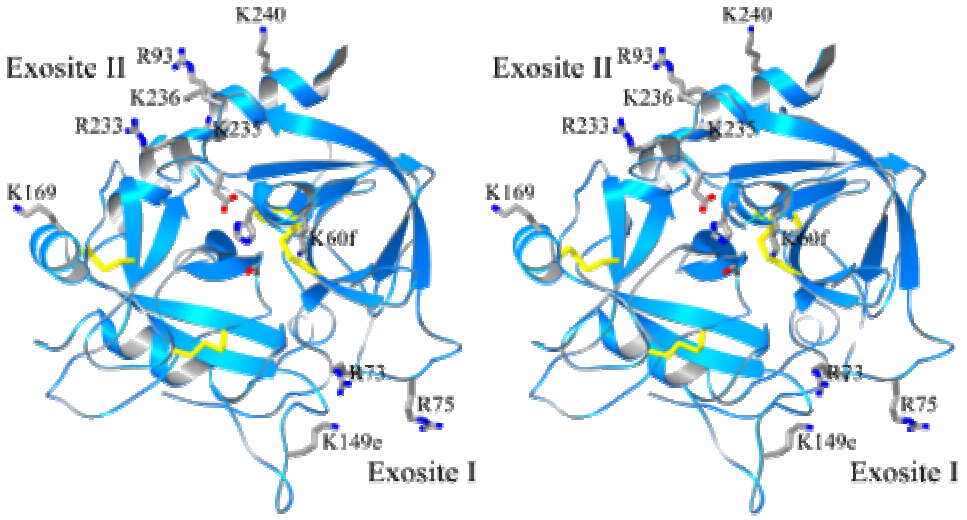


Figure 24: Stereo ribbon representation of human α-thrombin showing the location of mutated residues used in kinetic experiments. Thrombin is shown in the 'standard orientation' (see figure 2). Shown explicitly are the side chains of arginines or lysines which have been mutated singularly to glutamic acid which have been used in the present investigation. The side chains of the catalytic triad (Asp 102, His 57 and Ser 195) are shown unlabelled, and also indicated in the diagram are exosites I and II of thrombin.

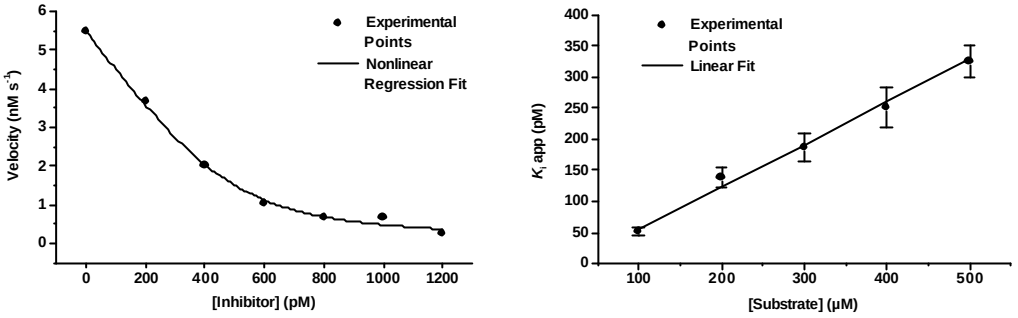


Figure 25: Graphical representation of experiments to determine K_i values. (Left) Graph of a tight binding assays of the human α-thrombin-haemadin complex. Plotted are rate of substrate hydrolysis against inhibitor concentration in this case for native thrombin with haemadin at an ionic strength of 0.6852 with 100 μM S-2238 as the substrate. The seven points were fitted using nonlinear regression to equation [1] to obtain the $K_{i'}$. (Right) A graph of $K_{i'}$ against substrate concentration for native thrombin with haemadin at an ionic strength of 0.6852. After repeating the tight binding experiment shown on the left with four other substrate concentrations (200-500 μM S-2238), the resulting $K_{i'}$ values (5 points) were fitted using weighted linear regression to equation [2], with the points weighted to the inverse square of their standard errors in order to obtain the K_i.

Table 6: Kinetic constants of thrombin mutants at 0.05 M Tris pH 8.3, 0.05M NaCl, 0.1% PEG 6000, 25 °C

Thrombin Mutant	K_i (10^{-13} M)	Fold Increase to native	k_{on} (10^{7} M^{-1} s^{-1})	k_{off} (10^{-5} s^{-1})
Native	2.44 ± 0.14	1.00	9.32 ± 1.65	2.27 ± 0.42
K60fE	2.48 ± 0.57	1.02	10.5 ± 0.23	2.62 ± 0.60
R73E	3.37 ± 0.52	1.38	7.20 ± 0.39	2.42 ± 0.40
R75E	2.51 ± 0.73	1.03	9.46 ± 0.38	2.37 ± 0.70
R93E	24.1 ± 0.12	9.89	1.46 ± 0.38	3.51 ± 0.92
K149eE	2.34 ± 0.35	0.96	11.7 ± 0.57	2.73 ± 0.43
K169E	2.29 ± 0.63	0.94	9.78 ± 0.84	2.24 ± 0.65
R233E	7.33 ± 0.45	3.01	4.44 ± 0.49	3.25 ± 0.41
K235E	2.45 ± 0.41	1.01	9.15 ± 0.92	2.24 ± 0.43
K236E	4.22 ± 0.30	1.73	6.55 ± 0.30	2.77 ± 0.24
K240E	12.9 ± 0.87	5.30	2.70 ± 0.24	3.48 ± 0.39

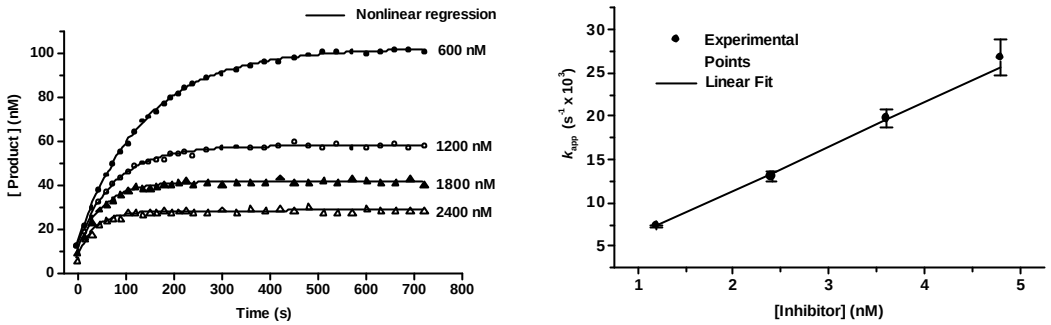


Figure 26: Graphical representation of experiments to determine k_{on} values. (Left) Graph of a slow tight-binding progress curve of human α-thrombin with haemadin. Plotted are the amount of substrate (Tos-GPR-AMC) cleaved against time, shown is the thrombin mutant K236E at an ionic strength of 0.3457. The four lines correspond to different concentrations of haemadin present. (15-30 × $[E_0]$). The 33 points from each individual curve were fitted using nonlinear regression to equation [3] to obtain the k_{app}. (Right) A plot of k_{app} against inhibitor concentration for the four curves shown on the left. The four points were fitted using weighted linear regression to equation [4], with the points weighted to the inverse square of their standard errors to obtain the k_{on}.

6.3.2. Effect of ionic strength on the K_i of haemadin with thrombin and thrombin mutants

The investigation of ionic interactions between the C-terminal tail of haemadin required that mutants of the heparin binding site (exosite II) be considered which, as shown in table 6, have a significant effect on the K_i compared to native thrombin or other thrombin mutants. The K_i values of six mutants (K60fE, R73E, R75E, K149eE, R169E and K235E) were measured and their similarity in binding energy to native thrombin effectively ruled them out as being involved in thrombin-haemadin interactions.

Of the exosite II mutants, which were implicated in haemadin binding as shown from their higher K_i values (compared to native thrombin) in table 6 (R93E, R233E, K236E and K240E), along with native thrombin, the effect of ionic strength on the K_i is shown in figure 27. As can be seen from the graph the ΔG°_b decreases as the ionic strength increases. The data from these curves was fitted to equation [9] using weighted nonlinear regression, the results of these analyses being shown in table 7.

This analysis allowed the binding energy (ΔG°_{b}) to be divided into a non-ionic component (ΔG°_{nio}) and an ionic component (ΔG°_{ion0}).

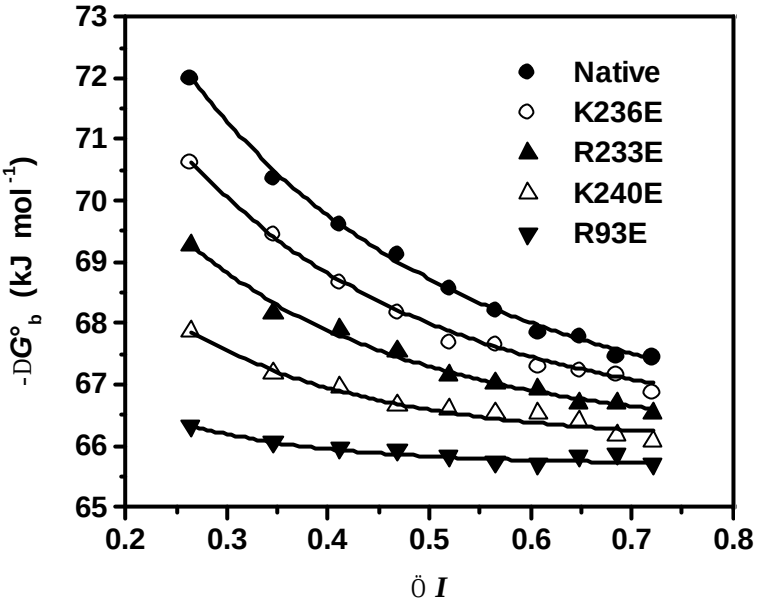


Figure 27: Effect of ionic strength on the binding energies of complexes of haemadin with thrombin and thrombin mutants. The binding energies of thrombin or thrombin mutants in complex with haemadin are plotted against the square root of ionic strength. Points are experimental derived from the K_i using equation [5] and the lines drawn represent the best fit to equation [9] with the points weighted to the inverse square of their standard errors.

Table 7: Parameters for the effect of ionic strength on the binding energy of the thrombin-haemadin complexes

Thrombin Mutant	C_1	$-\Delta G^{\circ}_{ion0}$ (kJ mol^{-1})	$-\Delta G^{\circ}_{nio}$ (kJ mol^{-1})
Native	2.26 ± 0.24	16.9 ± 0.86	66.2 ± 0.32
K236E	2.50 ± 0.35	14.3 ± 1.20	66.2 ± 0.29
R233E	2.72 ± 0.52	11.2 ± 1.51	66.1 ± 0.27
K240E	3.52 ± 0.70	8.91 ± 1.82	66.0 ± 0.15
R93E	4.67 ± 1.60	4.84 ± 2.51	65.7 ± 0.07

As table 7 shows, of the five thrombins analysed the ΔG°_{nio} values are very similar, and have a weighted mean value of -65.8 ± 0.10 kJ mol^{-1}, which suggests that the mutations made have only affected the ionic interactions between the proteins. However the ΔG°_{ion0} values of native thrombin and the mutants were different in each case, which suggests that the binding energies for these particular amino acids are different with respect to their binding to the tail of haemadin. Examination of the C_1 values shows a general increase as the binding energies of the mutants decreases.

6.3.3. Effect of ionic strength on the k_{on} of haemadin with thrombin and thrombin mutants

In addition these mutants were used to evaluate the effect of ionic strength on the associative rate constant k_{on}, this data being shown in figure 28, which shows that the association rate constant k_{on} decreased as the ionic strength was increased. The data from these curves were fitted by weighted nonlinear regression according to equation [10], and the results of these analyses are shown in table 8. The results of these analyses gave the values of $k_{on\infty}$ for thrombin and the thrombin mutants, this value representing the k_{on} at an infinite ionic strength where only non-ionic forces play a role in the association between the molecules. These figures as shown in table 8 are remarkably close, having a weighted mean value of 1.08 ± 0.08 × 10^7 M^{-1} s^{-1}. The analysis of these data allowed another set of figures for ΔG°_{ion0} to be obtained. The figures from both analyses were generally in agreement in the case of native thrombin, but in the case of the mutant thrombins the figures were an average of 2.56 kJ mol^{-1} lower than the ones derived from figure 27. The C_1 values for the complexes from table 8 were generally quite similar having a weighted mean value of 2.14 ± 0.30.

In general, when the data from figure 27 were fitted to the Debye-Hückel equation [7], and the data from figure 28 were fitted to the Debye-Hückel version of equation [10] $\{\ln k_{on} = \ln k_{on\infty} (-\Delta G^{\circ}_{ion0} / RT) (\exp(-C_1\sqrt{I}))\}$, the sum of the squares and the standard errors were slightly higher than when equations [8] and [10] were used. In all cases the C-values were about 60% higher and the ΔG°_{ion0} values were on average about 15% lower, whilst the ΔG°_{nio} and the $k_{on\infty}$ values were approximately the same.

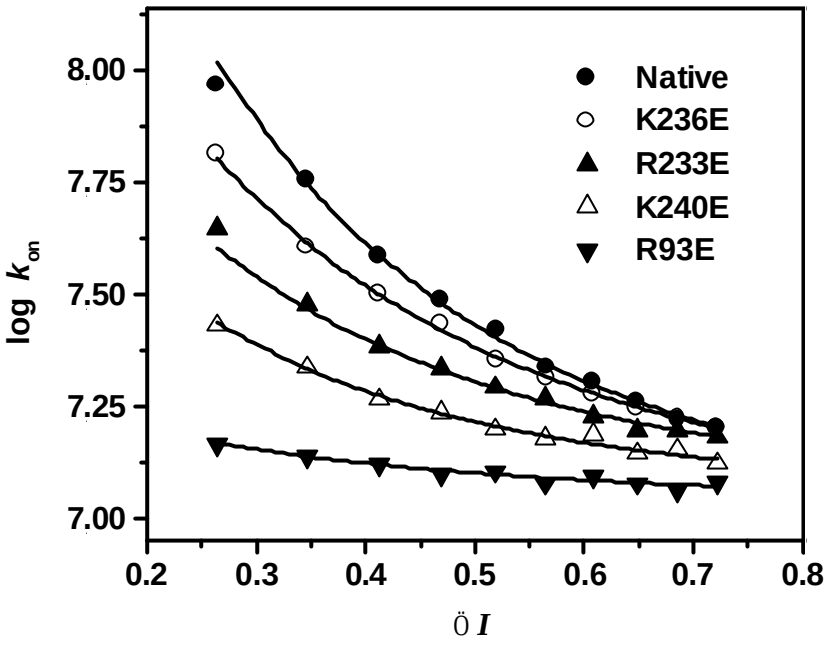


Figure 28: Effect of ionic strength on the associative rate constant k_{on} between haemadin with thrombin or thrombin mutants. Values of the associative rate constant are plotted against the square root of ionic strength. Points are experimental and the lines drawn represent the best fit to equation [10] with the points weighted to the inverse square of their standard errors.

Table 8: Parameters for the effect of ionic strength on the association rate constant between thombins and haemadin

Thrombin Mutant	C_1	$-\Delta G^{\circ}_{ion0}$ (kJ mol^{-1})	$k_{on\infty}$ (10^7 M^{-1} s^{-1})	k_{off} (10^{-5} s^{-1})
Native	2.37 ± 0.37	17.7 ± 2.61	1.00 ± 0.12	2.51 ± 0.22
K236E	1.99 ± 0.20	11.8 ± 0.73	1.00 ± 0.08	2.89 ± 0.38
R233E	2.16 ± 0.40	8.64 ± 1.21	1.14 ± 0.10	3.29 ± 0.47
K240E	2.44 ± 0.41	6.75 ± 0.98	1.14 ± 0.06	3.54 ± 0.33
R93E	1.90 ± 1.52	1.85 ± 0.78	1.09 ± 0.11	3.64 ± 0.48

6.3.4. Production and kinetics of haemadin (1-40)

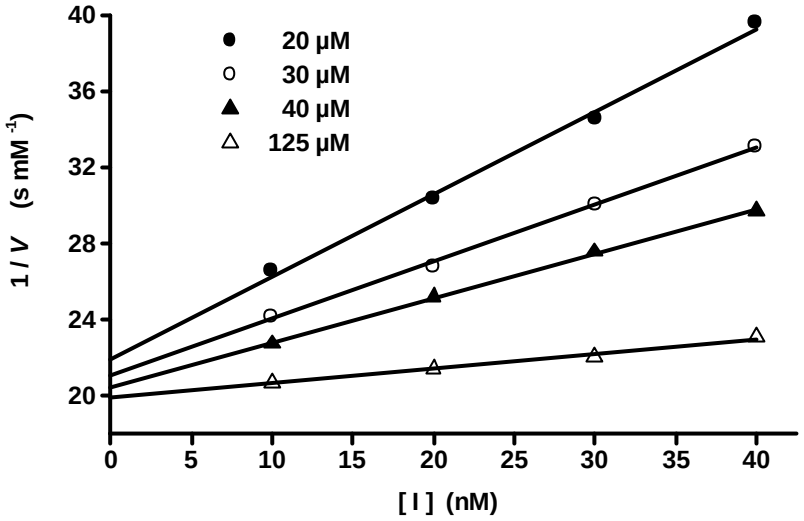


Figure 29: Dixon plot of human α-thrombin inhibition by haemadin (1-40) with four chromagenic substrate concentrations. Points are experimental and the lines drawn represent the best fit by linear regression. In all cases the linear regression coefficients were >99%.

The removal of the C-terminal tail of haemadin required a specific cleavage strategy in order to produce a pure product for kinetic analysis. The amino acid sequence of haemadin contains 1 arginine and 4 lysine residues, of which the N-terminal head region of haemadin contains both a lysine and arginine. Therefore the use of trypsin like enzymes to cleave off the C-terminal tail region of haemadin would have also cleaved off the N-terminal head of the inhibitor. Likewise the presence of hydrophobic amino acids, proline, glutamic acid or aspartic acid residues, either in the N-terminal region, or in haemadin's loops made the use of other enzymes such as chymotrypsin, proline endopeptidase, endoproteinase Glu-C and endoproteinase Asp-N also unattractive. However the presence of an aspartic acid-proline bond, allowed the selective hydrolysis of haemadin with 70% formic acid. After purification, the product was characterised by N-terminal sequencing, as well as by mass spectroscopy, with the purity being confirmed by quantitative N-terminal analysis with Ile 1I being greater than 99%.

The resulting product haemadin (1-40) no longer exhibited tight binding inhibition, but was still a competitive inhibitor, and its K_i was determined from a plot of $1/V$

against inhibitor concentration (Dixon plot) as shown in figure 29. As is seen in table 9, haemadin (1-40) possesses an inhibition constant 20,000 times higher than intact haemadin ($4.93 \pm 0.14 \times 10^{-9}$ M compared to $2.44 \pm 0.14 \times 10^{-13}$ M), indicating that loss of the acidic tail severely impairs haemadin (1-40) binding to human α-thrombin. A similar situation was seen for the k_{on}, which was less for haemadin (1-40) ($2.76 \pm 0.22 \times 10^{5}$ M^{-1} s^{-1}) than for intact haemadin ($9.32 \pm 1.65 \times 10^{7}$ M^{-1} s^{-1}).

6.3.5. Kinetics of haemadin with DIP-thrombin

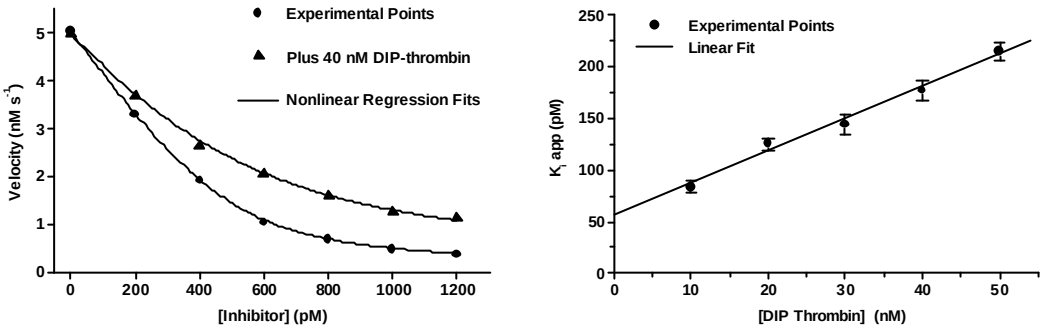


Figure 30: Graphical representation of experiments to determine K_i value for the DIP-thrombin-haemadin complex. (Left) A comparison of two tight binding curves, one with 500 pM native thrombin at an ionic strength of 0.2636 using 500 μM S-2238, and the other at the same conditions plus 40 nM DIP-thrombin. Plots of velocity against inhibitor concentration (7 points) were fitted using nonlinear regression to equation [1]. (Right) A plot of $K_{i'}$ against amount of DIP-thrombin for the 40 nm DIP-thrombin curve shown on the left and another 4 concentrations of DIP-thrombin. Weighted linear regression (5 points) gave the slope of the graph along with the standard error. The $K_{i'}$ derived from the steady state experiments without DIP-thrombin, divided by the slope of this plot, gave a value of the K_i for the DIP-thrombin-haemadin complex at this ionic strength.

With haemadin binding at sites distinct from the active site (Richardson *et al.*, 2000), thrombin with a modified active site would be expected to compete with α-thrombin for the formation of a complex with haemadin. Indeed the inclusion of DIP-thrombin in steady state measurements caused an increase in the $K_{i'}$, which was linear when plotted against the concentration of DIP-thrombin added (figure 30). The alteration of thrombin's active site using DIPF caused a 72,000 increase in the K_i of the inhibitor (from $1.76 \pm 0.11 \times 10^{-8}$ M to $2.44 \pm 0.14 \times 10^{-13}$ M) indicating that haemadin binds to DIP-thrombin less strongly than to α-thrombin.

The kinetic constants of the human α-thrombin-haemadin (1-40) complex and the DIP-human α-thrombin-haemadin complex at two different ionic strengths are shown in table 9 for comparison. In particular it is noteworthy that haemadin (1-40) shows little change in either its K_i (from 4.93 ± 0.14 M to 5.67 ± 0.16 M) or its k_{on} (from 2.76 ± 0.22 M^1 s^{-1} to 2.43 ± 0.23 M^1 s^{-1}) at low (0.0695) or high (0.5195) ionic strengths.

Table 9: Values of kinetic constants at extremes of ionic strength for thrombin and haemadin derivatives

Complex / Measurement	Value at Ionic strength of 0.0695	0.5195
Haemadin (1-40)-thrombin / K_i (10^{-9} M)	4.93 ± 0.14	5.67 ± 0.16
Haemadin (1-40)-thrombin / k_{on} (10^5 M^{-1} s^{-1})	2.76 ± 0.22	2.43 ± 0.23
Haemadin-DIP-thrombin / K_i (10^{-8} M)	1.76 ± 0.11	12.3 ± 0.74

6.3.6. Kinetics of synthetic haemadin peptides with thrombin

The synthetic amino terminal peptides H-IRFGMGKV-NH_2 and H-IRFGMGKVP-NH_2 based on the first 8 and 9 amino acids of haemadin, had K_i values in the micro-molar range and were determined as 3.27 ± 0.30 × 10^{-6} M and 3.01 ± 0.14 × 10^{-6} M respectively.

6.4. The bovine α-thrombin-tsetse inhibitor complex

6.4.1. Crystallisation

Crystals of the bovine α-thrombin-sTTI complex of approximate dimensions $0.2 \times 0.2 \times 0.2$ mm^3 were grown by vapour-diffusion using the sitting drop method. Droplets of 3 µl consisting of 1.5 µl of protein solution (bovine α-thrombin-sTTI complex, 9 mg/ml) and 1.5 µl of precipitant (2 M $(NH_4)_2SO_4$, 5% (v/v) isopropanol), were equilibrated against 500 µl of precipitant solution at 22ºC. Crystals appeared after 3 days and achieved maximum size after 1 week. Problems with macroscopic twinning were encountered (See figure 31) in some drops, but in others apparently no twinning was evident.

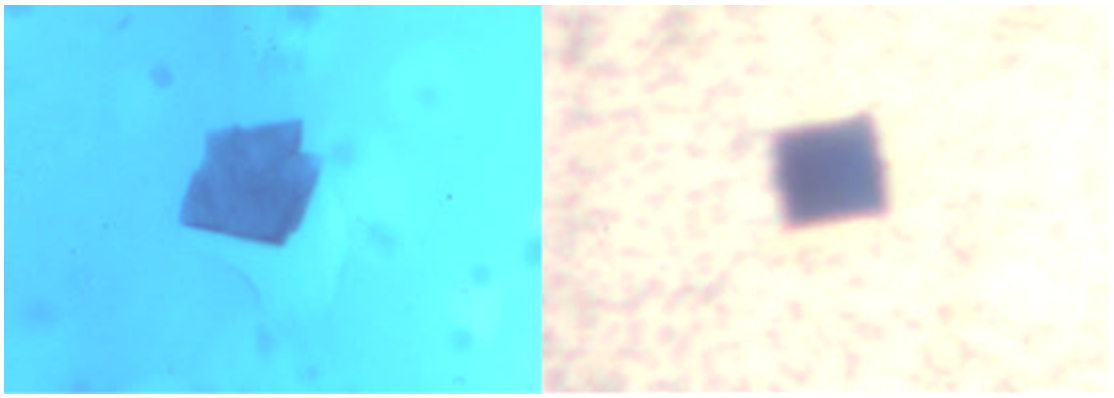


Figure 31: Photographs of crystals of the bovine α-thrombin-sTTI complex taken under a microscope. The crystals on the left exhibit clear macroscopic twinning, whereas the crystals on the right appear to be macroscopically untwinned. The crystals shown have approximate dimensions of $0.1 \times 0.1 \times 0.1$ mm^3 and are blue due to staining with bromophenol blue.

6.4.2. Data collection

Data was collected form one crystal of the bovine α-thrombin-haemadin complex at 77 K, with the crystals diffracting to 2.9 Å resolution. Cryo-cooling was achieved by the addition of 20% glycerol to the well solution. A picture of the diffraction pattern is shown below in figure 32, and the data processing statistics are shown in table 10.

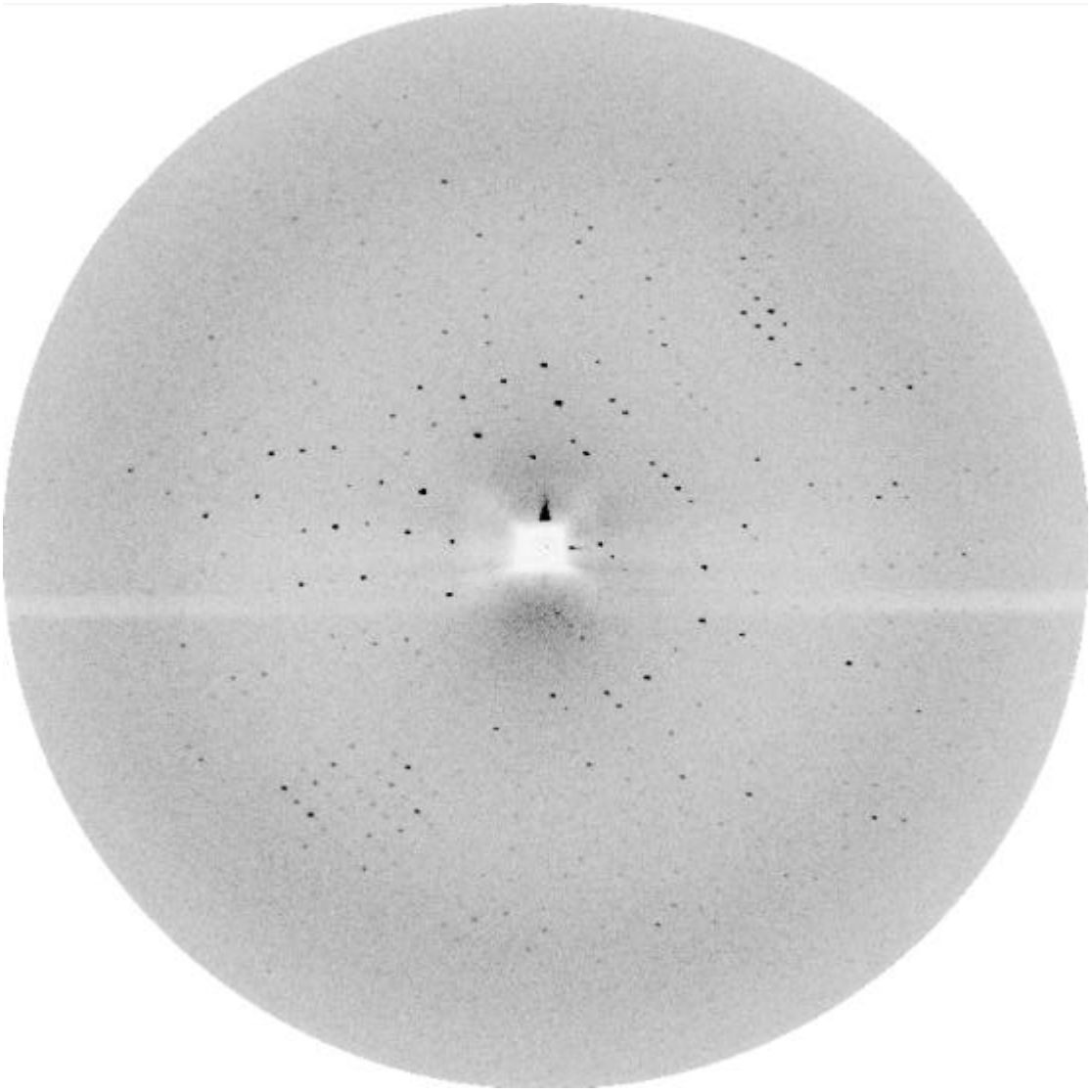


Figure 32: Image plate image of the diffraction pattern from a crystal of the bovine α-thrombin-sTTI complex. The crystal was rotated through 1° for twenty minutes, with the outer edge of the diffraction pattern extending to 2.9 Å.

Table 10: Data processing statistics of the bovine α-thrombin-sTTI complex

Crystal system	Orthorhombic
Space group	P2(1)2(1)2
Cell constants	*a* = 87.03 Å
	b = 87.43 Å
	c = 99.34 Å
	α = 90°
	β = 90°
	γ = 90°
Cell volume	755881
V_m (Å^3/Da)	2.27 (46% solvent)
Number of crystals	1
Limiting resolution, Å	2.9
R_{merge}, %[1]	12.5
Reflections collected	207,862
Unique reflections	17,407
Completeness (∞ to 2.9 Å), %	99.4

[1] $R_{merge} = [S_h S_i|I(h,i) - \langle I(h)\rangle| / S_h S_i I(h,i)] \times 100$, where I(h,i) is the intensity value of the i-th measurement of h and <I(h)> is the corresponding mean value of h for all i measurements of h. The summation is over all measurements.

6.4.3. Matthews parameter

In the case of the bovine α-thrombin-sTTI complex the solvent content of the crystals was 46% for two thrombin molecules per unit cell, with a Matthews parameter of 2.27 Å^3/Da.

6.4.4. Patterson search

The structure of the bovine α-thrombin-sTTI complex was solved using Patterson search techniques. As a search model the coordinates of bovine α-thrombin from the thrombin-triabin complex (Fuentes-Prior *et al.*, 1997), from which residues Phe 1U to Gly 1D and Glu 14L to Arg 15 of the A-chain, and most of the residues forming the autolysis loop (Arg 145 to Glu 149E) were excluded, was used. Rotational and translational searches to determine the orientation and position of the thrombin components were performed with the program AMoRe (Navaza, 1994) using diffraction data from 43.8 to 3.7 Å. As expected from the Matthews parameter for a dimer, two rotation solutions were found. After translational search and rigid body refinement, these solutions resulted in a correlation value of 0.50 and an R-factor of 0.37 (next highest peak: c = 0.33, R = 0.42).

6.4.5. Detection of twinning

Due to perfect hemihedral twinning the original data set was originally interpreted as a tetragonal system with unit cell dimensions of 87.188, 87.188 and 99.355. After molecular replacement, the 2.9 Å ($2F_{obs}$-F_{calc}) Fourier map revealed good density for thrombin and some of the density of the inhibitor. Attempts to build either the missing parts of thrombin or the inhibitor, led upon refinement, to an increase in the free *R* value. Additionally the Ramachandran plot progressively worsened upon further refinement. Suspecting that something was amiss the data was submitted to the crystal twinning server at UCLA (http://www.doe-mbi.ucla.edu/Services/Twinning/). The server looks for an intensity distribution that does not follow Wilson statistics. This involves comparing $\langle I^2 \rangle$ / $\langle I \rangle^2$, where the expected values are 1.5 for (acentric) twinned data and 2.0 for (acentric) untwinned data. The test must be performed on normalised data or in thin shells. Overleaf (figure 33) is a graphical representation of the results.

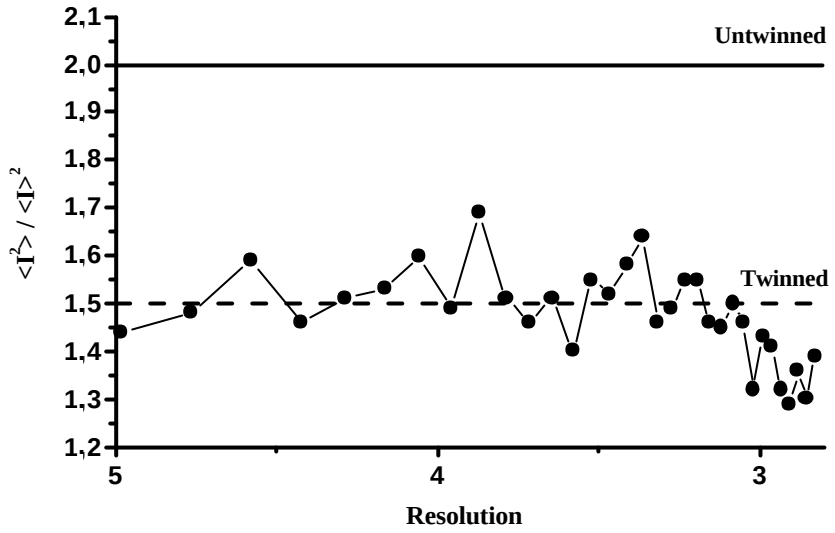


Figure 33: A test for perfect merohedral (hemihedral) twinning for acentric data from the bovine α-thrombin-sTTI complex. The ratios are computed in thin resolution bins of 500-600 reflections.

6.4.6. Detwinning

Detwinning was accomplished using the XPLOR script of Redinbo and Yeates (Chandra *et al.*, 1999). This program essentially rotates the model to the expected orientation of the molecule in the other unit cell. Once this is accomplished the F_{calc} figures for the rotated model are scaled to the original data, and subtracted from it where there are data.

6.4.7. Model building and refinement

The bovine α-thrombin-sTTI complex was built using the program MAIN (Turk, 1992). Into an interpretable 2.9 Å ($2F_{obs}$-F_{calc}) Fourier map calculated after appropriate positioning of the thrombin molecules in the crystal cell, some of the omitted parts of thrombin and the segments of sTTI in contact with thrombin were built. After detwinning using the method of Redinbo and Yeates (Redinbo and Yeates 1993), the complexes were then crystallographically refined with X-PLOR (Brünger, 1991)

using the target parameters of Engh and Huber (Engh and Huber, 1991). This usually consisted of normal crystallographic refinement, B-factor refinement, as well as simulated annealing at 2500 K. This process was repeated cyclically a few times until most of the inhibitor sequence could be fitted into the density. Twenty-three water molecules were added at stereochemically reasonable positions where the ($2F_{obs}$-F_{calc}) density and (F_{obs}-F_{calc}) difference density maps were above 1σ and 2σ levels, respectively. The *R* factor of the final model was 0.265 (with a free *R* factor of 0.262). The refinement statistics are given in table 11.

Table 11: Refinement statistics for the bovine α-thrombin-sTTI complex

Resolution range, Å	43.8-2.9
Reflections used for refinement	15,242
R-value (R_{untwin}), %[1]	26.2
R_{free}, %[2]	26.5
Non hydrogen protein atoms	4,772
Solvent molecules	23
Standard deviations:	
Bond lengths, Å	0.008
Main chain bond angles, °	1.286
Side chain bond angles, °	22.88
Improper angles, °	1.16
No. of residues excluding glycine and proline	482
(Φ,Ψ) angle distribution in	
Most favoured region	368 (75.6%)
Additionally allowed region	101 (21.0%)
generously allowed region	8 (1.7 %)
Disallowed region	8 (1.7%)
Average *B*-factor	34.99 ± 7.60
Min *B*-factor	7.60
Max *B*-factor	73.55
Occupancy	98.0%

[1] $R_{untwin} = \sum_l | F_{calc}(h_1) - F_{detwin}(h_1)_l | / \sum_l | F_{detwin}(h_1)_l | \times 100$, see section 4.6.2.4.1.

[2] R_{free} was calculated randomly omitting 5% of the detwinned reflections from refinement.

6.4.8. Quality of the model

The refined model consist of 2 molecules of bovine α-thrombin and 1 sTTI inhibitor molecule (consisting 19 amino acids) making a total of 4,772 nonhydrogen atoms, with an additional 23 water molecules present in the structure. Below (figure 34) is the Ramachandran plot of the final model.

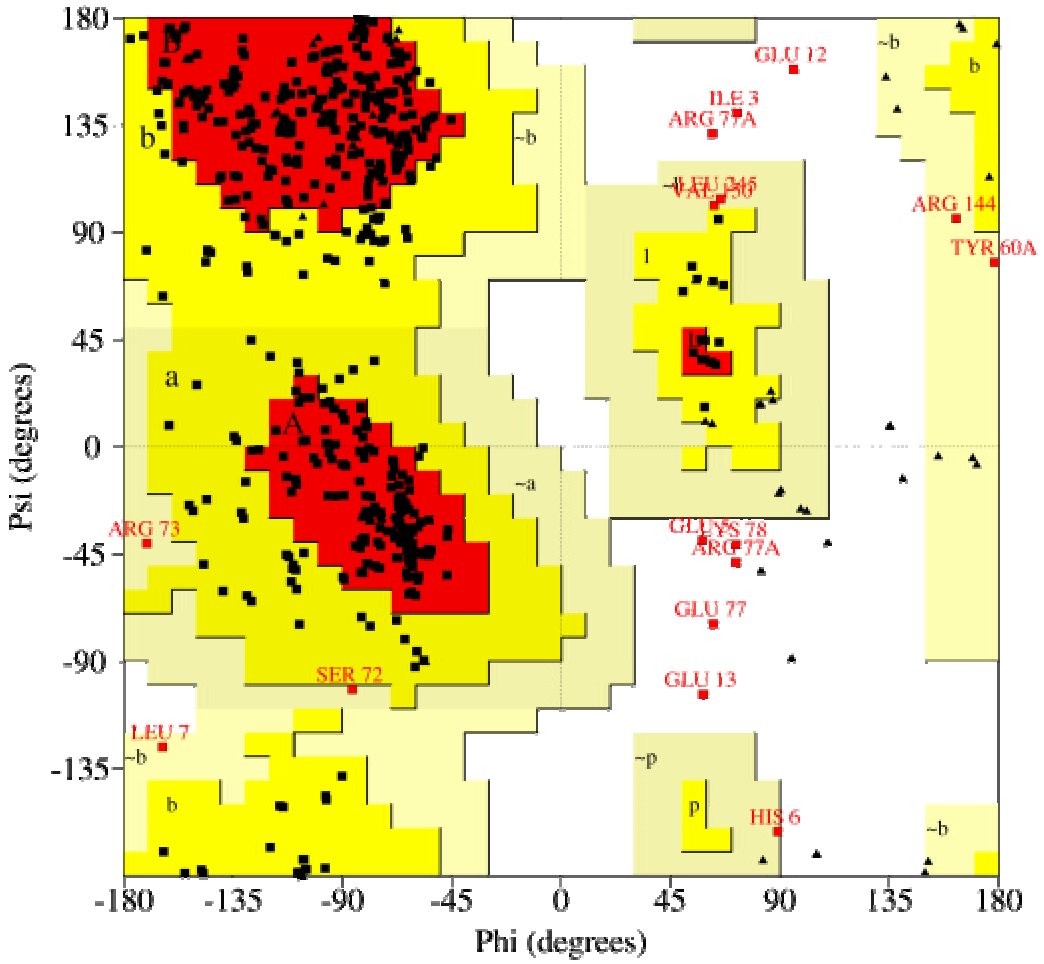


Figure 34: Ramachandran plot of the Phi and Psi angles of the atomic model of the bovine α-thrombin-sTTI complex.

6.4.9. Molecular packing and inter molecular contacts

The bovine α-thrombin-sTTI complex crystallises as a dimer in the orthorhombic space group P2(1)2(1)2, with the crystallographically independent complex consisting

of two thrombin molecules. The thrombin molecules are placed active site to active site with interactions between the 60 loops via hydrophobic contacts with Pro 60C to Pro 60C. Additionally there are contacts between the 90 loop and 60 loops with hydrophobic contacts between Trp 60D and Trp 96A as well as a salt bridge between Lys 97 and Asp 60E. In Addition the 170 loops and 220 loops form minor hydrophobic contacts between each other. The 180 loop of molecule B pushes the 140 loop of molecule A aside somewhat. Between the asymmetric thrombins there are further interactions, with molecule A interacting with its 170 loop with the 180 loop of its symmetry thrombin, and it also interacting with its 200 loop with the 30 loop of its symmetry thrombin. For molecule B the 37 loop clashes with the 200 loop of its symmetry molecule.

Concerning the inhibitor, it runs from the active site of one thrombin molecule to the exosite I of the other thrombin molecule, and is thus essential for formation of the crystallographic A-B dimer.

As seen in figure 35, the peptide chain of sTTI fills the A-B interface by simultaneously binding to one thrombin and to the other thrombin molecule.

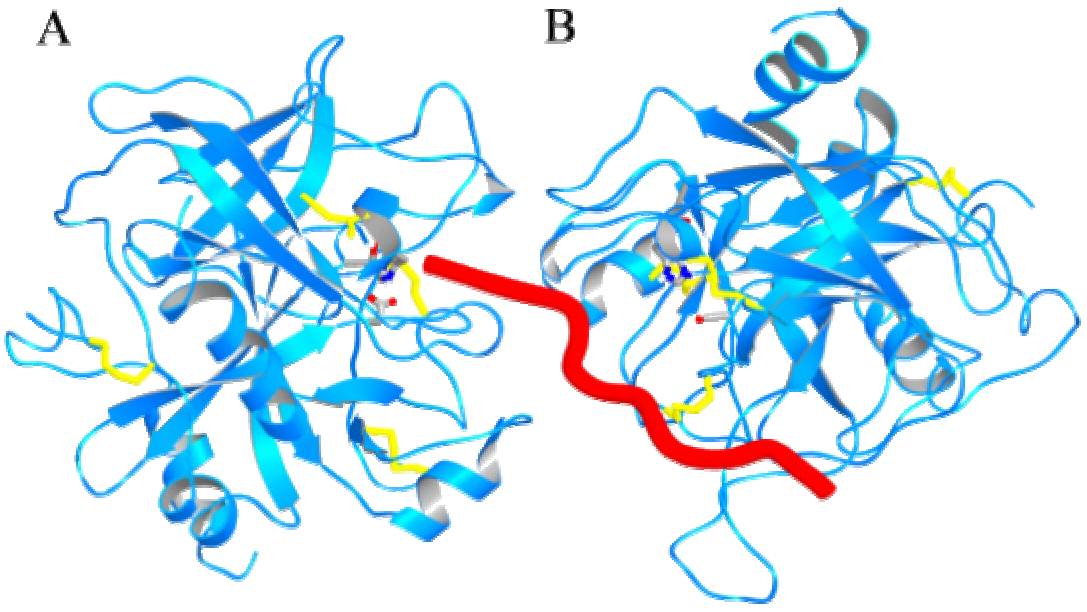


Figure 35: Ribbon representation of the bovine α-thrombin sTTI dimer. Thrombin is depicted in blue and sTTI in red. The side chains of the catalytic triads of the two thrombins are shown explicitly. The two thrombins are labelled explicitly as they are referred to in the text.

6.4.10 Structure description

6.4.10.1. Structure of thrombin

The structure of sTTI-bound thrombin is similar to that of thrombin in the bovine α-thrombin-triabin structure (Fuentes-Prior *et al.*, 1990), the active site catalytic triad, being almost identical in both structures. Large deviations in the main chain are essentially limited to the so called 'autolysis loop' or 149-loop (residues Thr 147 to Lys 149E) which in molecule B is displaced considerably. This loop possesses a well-reported conformational flexibility (Bode *et al.*, 1992) and has missing or very weak electron density in the two thrombins of the asymmetric unit. Electron density is also weak or missing in the light chain termini (residues Thr 1U to Glu 1C of complex A, as well as residue Arg15 of complex A, and residues Thr 1U to Glu 1C of complex B. In addition slight perturbances due to crystal contacts are seen in the 70 loop of molecules A and B as well as the 180 loop of molecule B.

6.4.10.2. Structure of sTTI

sTTI being a small peptide has no secondary structure, and displays no secondary structure in complex with thrombin. The inhibitor has missing density for the residues Gly 1I to Ser 16I, as well for Gly 20I to Thr 22I.

6.4.10.3. Binding of the inhibitor

sTTI binds to the active site of thrombin with its carboxy terminus within hydrogen bonding distance of Ser 195, the position of the side chain of Ile 32I occupying a similar position to the Ile 1I of haemadin, except in the case of haemadin the residue is at the N-terminus. Many of the interactions between sTTI and thrombin are of a hydrophobic nature. This residue makes many hydrophobic interactions with residues His 57, Tyr 60A and Leu 99 (See figure 36). The Arg 31I of sTTI unlike the Arg 2I of haemadin does not extend into the S1 pocket, instead it binds to Glu 217 forming a

salt bridge at a near optimal distance of 3.1 Å. Additionally Arg 2I makes two hydrogen bonds to Gly 216 in a manner similar to substrate binding. Continuing to residue Leu 3I, this residue occupies a position quite similar to Tyr 3I of hirudin and mostly interacts with residue Tyr 60A via van der Waals forces. The residue Ile 27I forms hydrophobic interactions with the side chain carbon atoms of Arg 173.

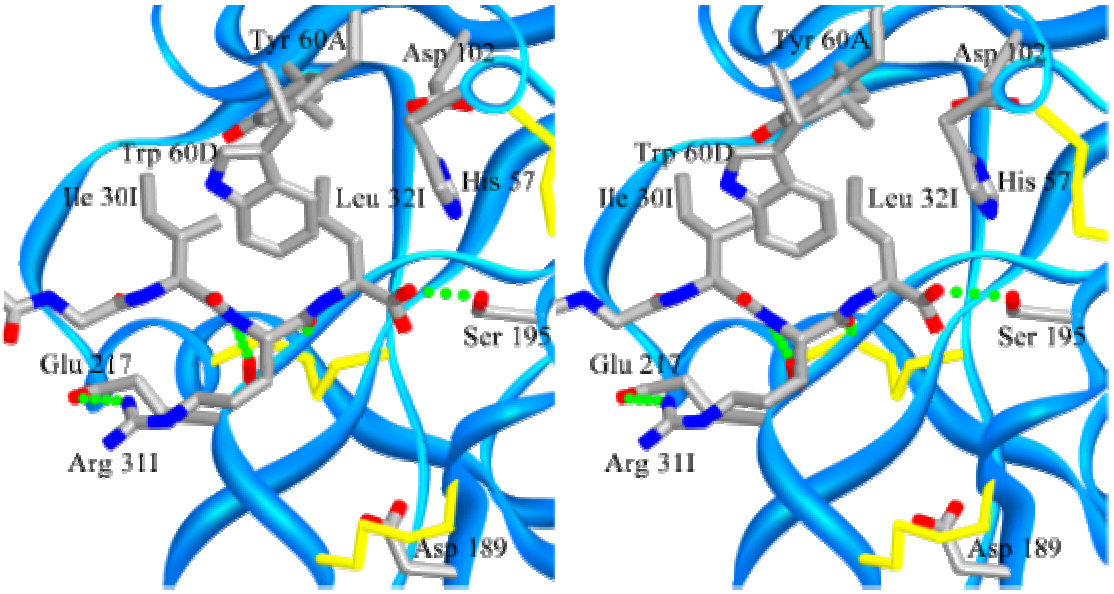


Figure 36: Close-up Stereo representation of sTTI binding to the active site of bovine α-thrombin. Side chains of selected residues are depicted element colour-coded. Major interactions including the hydrogen bonds and ionic interactions discussed in the text are shown as green dotted lines. The three side chains if the catalytic triad are shown explicitly, also shown is the side chain of residue Asp 189 of the S1 pocket which is included to show its proximity to Arg 31I. Some residues of thrombin and the inhibitor along with all water molecules are omitted for the sake of simplicity.

Coming away from the active site Glu 26I forms a salt bridge with Lys 60F which is bent away from its usual position close to the active site, whilst its side chain carbon atoms interact in a hydrophobic manner with Trp 60D (See figure 37). The residue Leu 24I make hydrophobic interactions with the alkyl chain of Glu192, in a similar fashion to Thr 21I which also makes a similar interaction with Glu 25. After Pro 23I there are three amino acids, a threonine and two glycines that are not visible in the electron density. The interactions continue with Ile 19I which forms a hydrophobic interaction with Phe 34, in a manner similar to Phe 56 of hirudin, additionally its backbone carbonyl atom makes a hydrogen bond to the side chain NH of Gln 151. Also in a manner similar to hirudin after Ile 19I there are two glutamate residues (Glu

17I and Glu 18I), however these residues have no opportunity to bind, as the 200 loop of a symmetry thrombin molecule binding to thrombin B, prevents them from binding.

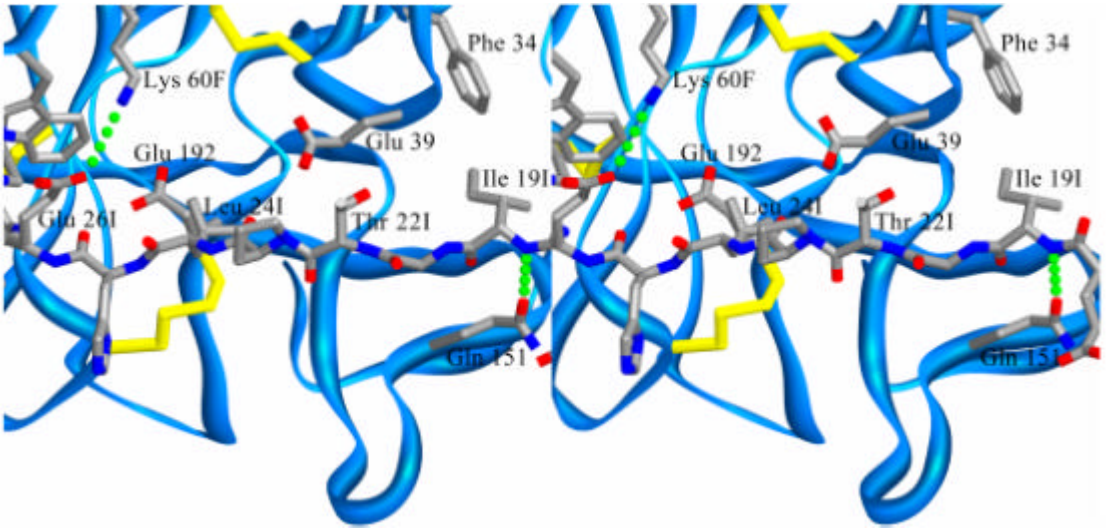


Figure 37: Stereo close-up view of sTTI binding to bovine α-thrombin. Side chains of selected residues are depicted element colour-coded. Major interactions including the hydrogen bonds and ionic interactions discussed in the text are shown as green dotted lines. The peptide chain of the inhibitor runs left to right as C-terminal to N-terminal. Some residues of thrombin and the inhibitor along with all water molecules are omitted for the sake of simplicity.

6.5. Biochemistry of the bovine a-thrombin-sTTI complex

6.5.1. Kinetics of sTTI with bovine a-thrombin

The K_i of the bovine α-thrombin-sTTI complex was $1.98 \pm 0.04 \times 10^{-8}$ M whereas the K_i of the bovine α-thrombin E217K-sTTI complex was $8.36 \pm 0.12 \times 10^{-7}$ M. These two K_i values convert by equation [5] to ΔG°_b values of -43.94 kJ mol^{-1} and -34.69 kJ mol^{-1} respectively.

6.5.2. Gel filtration of the bovine a-thrombin-sTTI complex

The calculated molecular weight from the gel filtration experiment was 40.1 KDa which correspond to the molecular weight of the monomeric complex being 37.1 KDa.

6.5.3. Displacement of sTTI by triabin

After one hour in the presence of triabin, the rate of thrombin cleavage of the substrate S-2238 was 48% compared to the rate without triabin present. After 10 min the rate was 5% of the rate without triabin.

7. DISCUSSION

7.1. The human α-thrombin-haemadin complex

7.1.1. The true nature of haemadin binding

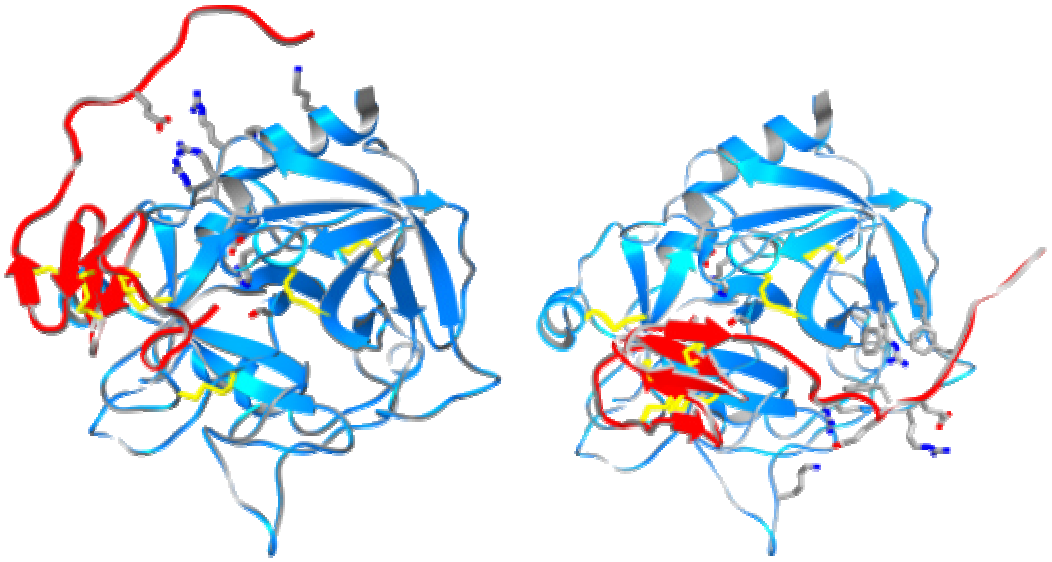


Figure 38: Ribbon representation of the two possible binding modes of haemadin to human α-thrombin. Thrombin is shown in 'standard orientation' (see figure 2), shown explicitly are the side chains of the catalytic triad, as well as the side chains of haemadin shown in the crystal structure to bind to residues of an exosite. Also shown are some of the basic residues of exosites I and II. On the left is monomer A (See figure 12) of the crystal structure which shows the C-terminal tail over exosite II. On the right is a model of the complex with the C-terminal tail positioned as in the X-ray structure, binding to exosite I. Note how the C-terminal tail is too short to allow it to be simply transposed from one exosite to another, without the body of the inhibitor being moved significantly.

After the crystal structure of the human α-thrombin-haemadin complex was built there were major concerns as to the mode of binding of the inhibitor. This was largely due to the location of the C-terminal tail which in the crystal structure is bound to exosite I of a neighbouring thrombin in a mode similar to hirudin (See figure 14). Therefore it was conceivable that the inhibitor could be adopting a hirudin like binding mode as is shown modelled in figure 38. The first evidence that haemadin bound to exosite II was that all three inhibitors in the asymmetric unit adopted similar conformations (See figure 13) with the exception of their C-terminal tails. This

seemingly innocuous information was important as figure 38 shows, that in order to bind to exosite I, not only the tail has to move but the whole inhibitor. Seeing as the inhibitor occupies the same position in all three of the asymmetric inhibitors seemed to suggest that the inhibitor did indeed bind to exosite II. Further evidence from "band shift" studies (See figure 22) conducted with the exosite I binding inhibitor triabin (Fuentes-Prior *et al.*, 1997), showed that a triple complex was formed between thrombin, haemadin and triabin, whereas in the case of hirudin no such triple complex was formed. The presence of exosite II binding was also hinted at by the non-binding of haemadin to meizothrombin whose exosite II is obstructed by the second kringle domain (See figure 22). However these results although ruling out exosite I binding, do not conclusively prove binding to exosite II. The kinetics experiments with thrombin exosite I and II mutants (See section 6.3.1.) unequivocally prove that the complex is as shown on the left in figure 38, which exposes the crystal structure as having artefacts.

7.1.2. Comparison of haemadin with hirudin

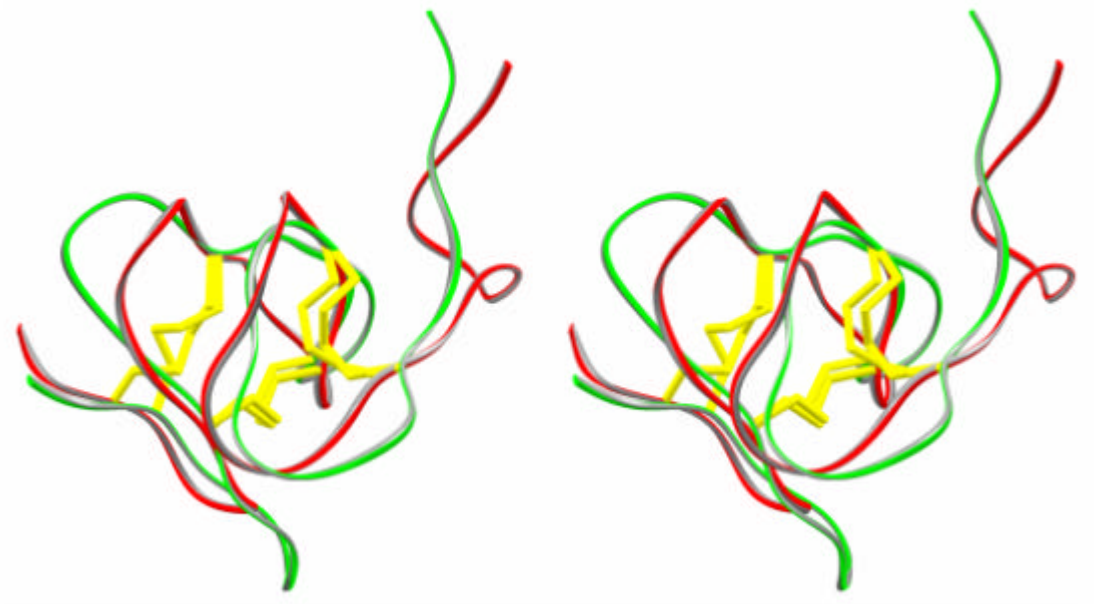


Figure 39: Stereoview of the main chain of haemadin (red, residues Ile 1I to Ser 38I) and hirudin (green, residues Ile 1H to Val 40H) after optimal least-squares fit. Only the side chains of the first three residues of both molecules are shown explicitly. Note the different location of the N-terminal segments, indicating divergent arrangements of the compact domains relative to thrombin (compare figure 41).

The disulphide-rich cores of haemadin (residues Pro 9I to Ser 38I) and *Hirudo medicinalis* hirudin (Thr 4H to Val 40H; the suffix 'H' denotes hirudin residues) can be overlaid with a root-mean-square deviation of 1.15 Å for 22 pairs of equivalent Cα atoms. As shown in figure 39, all three disulphide bonds are spatially similar, but the four loops described earlier for haemadin are somewhat offset in the two structures. Some of the differences can be accounted for by loop size discrepancies, but in the case of loop C, which is of identical size, the displacement is due to Gly 23H following the disulphide bond [4-6] (Cys 22H to Cys 39H) in hirudin. A structure-based sequence alignment of haemadin with four hirudin variants is presented in figure 40; it highlights that the overall conservation of the three-dimensional structure is only marginally matched at the sequence level. Additionally the active site regions of haemadin and hirudin, as shown in figure 41, are extremely similar when overlaid with respect to thrombin, the majority of this similarity lies in their parallel binding mode to Ser 214 to Gly 216 and the similar nature of their side chains.

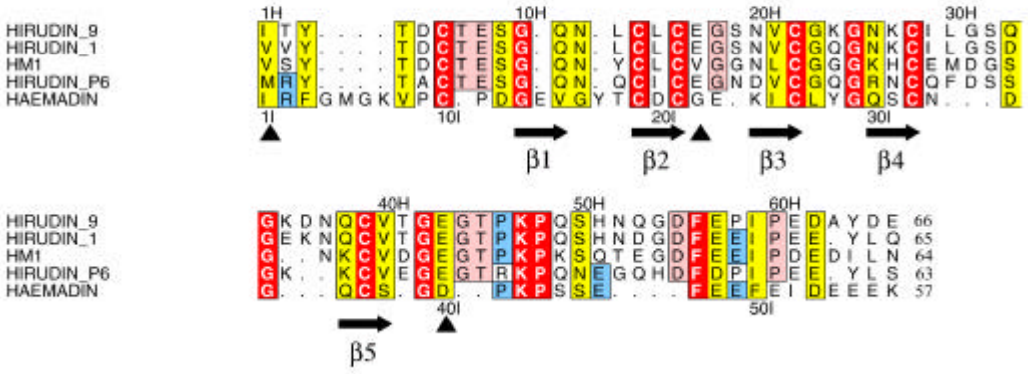


Figure 40: Structure-based alignment of the amino acid sequences of haemadin and of four representative hirudin variants. Nomenclature follows the work of Steiner and co-workers Steiner *et al.*, 1992). Residues with particularly close homologies are boxed in yellow, identities in red. Residues conserved in hirudin but not haemadin are shadowed pink; those common to haemadin and some hirudin variants are shadowed blue. Numbers refer to the sequences of hirudin (above) and haemadin (below the alignment). The secondary structure of haemadin is also given. The intron-exon boundaries (full arrows) are those determined for *Hirudinaria manillensis* (Scacheri *et al.*, 1993). The aligned sequences were formatted using the program ALSCRIPT.

The considerable similarities of the C-terminal tails manifest themselves in the binding of the C-terminal peptides of haemadin to the fibrinogen-recognition exosites of neighbouring thrombin molecules in the current crystal structure (See figures 12

and 42). The main chains of residues Glu 46I to Glu 51I and Asp 55H to Pro 60H can be superimposed, with Cα atoms deviating less than 1.3 Å. This similarity extends to the conformation of several side chains and thus to the contacts made with thrombin (See figure 42).

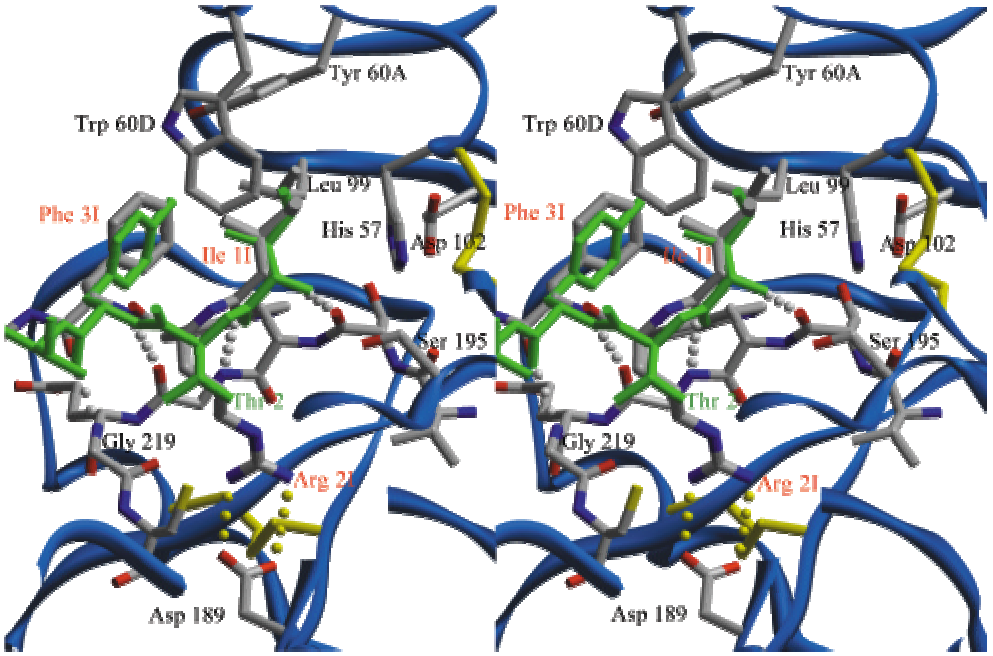


Figure 41: Close-up stereoview of the active site cleft of haemadin-bound human α-thrombin compared with the three N-terminal amino acid residues of hirudin (green). Side chains of selected residues are depicted colour-coded. Intermolecular hydrogen bonds are shown as white dotted lines for the parallel interaction of haemadin's N-terminal loop with thrombin residues Ser 214 to Gly 216 or yellow dotted lines for contacts made inside the S1 pocket. Some residues of thrombin and the inhibitors along with all water molecules are omitted for the sake of simplicity.

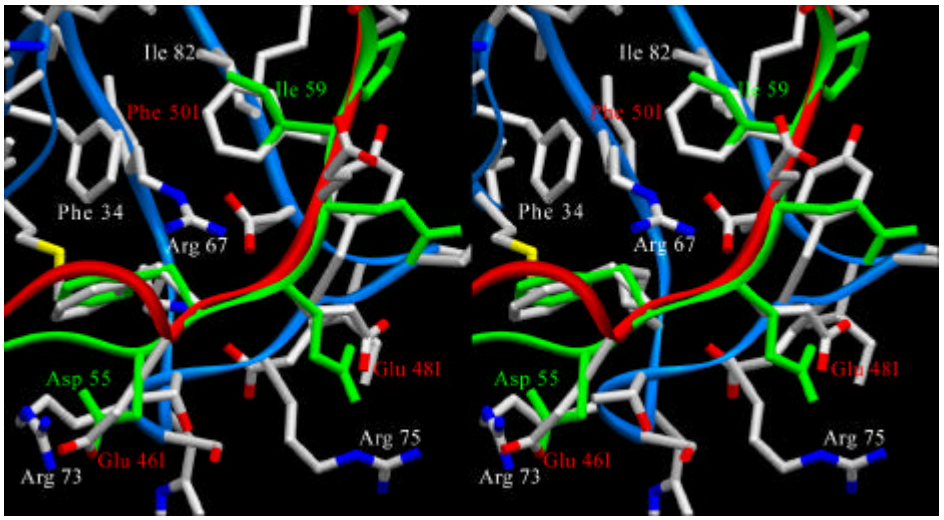


Figure 42: Close-up stereoview comparing the interactions of the C-terminal tails of haemadin (red) and hirudin (green) with the fibrinogen-recognition exosite of a neighbouring thrombin molecule (blue). Side chains of interacting thrombin / inhibitor residues are explicitly labelled. Notice the close agreement between the phenyl moieties of Phe 47I and Phe 56H; also the side chains pairs Phe 50I and Ile 59H, and Glu 48I and Glu 57H occupy similar positions.

7.1.3. Comparison of haemadin with other natural thrombin inhibitors

Serine proteinase substrates bind to the active-site cleft of their cognate proteinase by building an antiparallel β-strand with residues Ser 214 to Gly 216 (chymotrypsinogen numbering). Although this "canonical" mode of binding has been encountered in a natural thrombin inhibitor, rhodniin (van de Locht *et al.*, 1995), evolutionary pressure appears to have favoured a different type of thrombin-inhibitor association, in which the inhibitor forms instead a parallel β-strand with enzyme residues Ser 214 to Gly 216. This parallel retro-binding alignment has first been observed in the leech-derived inhibitor hirudin (Grütter *et al.*, 1990; Rydel *et al.*, 1990), and subsequently also in the double-headed Kunitz-type inhibitor ornithodorin (van de Locht *et al.*, 1996) and in the small molecule nazumamide A (Nienaber and Amparo, 1996). The parallel mode of binding has been exploited by developing potent thrombin inhibitors, termed hirunorms (De Simone *et al.*, 1998). In spite of a definite sequence variability within hirudin variants isolated from different leeches (Scacheri *et al.*, 1993; Steiner *et al.*, 1992) (See figure 40), the hirudin structure and thus its mode of inhibition appeared to be essentially conserved (Folkers *et al.*, 1989; Grütter *et al.*, 1990;

Haruyama and Wuthrich, 1989; Priestle *et al.*, 1993; Rydel *et al.*, 1990; Vitali *et al.*, 1992).

7.1.4. Evolutionary origin of haemadin

As shown in figure 43 the phylogenetic relationship between haemadin and hirudin is the furthest of any hirudin variant. Although *Hirudo medicinalis* is a water dwelling leech and *Haemadipsa sylvestris* is a land living leech, the *Haemadipsa* branched off before the *Hirudo* in their evolutionary histories (See figure 44) as classified by cocoonology, parental care, habitat preference, vectorology and dietry preferences.
Although essentially no conclusions can be drawn about molecular evolution from life history evolution, only three leech species to date have been found to possess hirudin like thrombin inhibitors, the *Hirudo*, *Hirudinaria* and *Haemadipsa* species. It has been assumed that the *Hirudinaria* species inhibitors are similar in action to hirudin (De Filippis *et al.*, 1999). However little is known about the hirujins, but their primary structure is the closest to haemadin, even though it is a *Hirudo* species. This seems to suggest that haemadin evolved into hirudin, unless, of course, the hirujins have back evolved. The evolutionary relationships are not entirely clear, especially as the hirujins possess a seven amino acid sequence prior to the first disulphide bond. The length of this sequence is important as we believe it plays an important role in distinguishing haemadin from hirudin (see section 7.1.6.). In hirudin this sequence is five amino acids long and in haemadin it is nine amino acids. Further difficulties lie in that the first 5 amino acids of hirujin are not known, and the N-terminal amino acid is believed to be derivatised, which is somewhat different to the situation in haemadin.

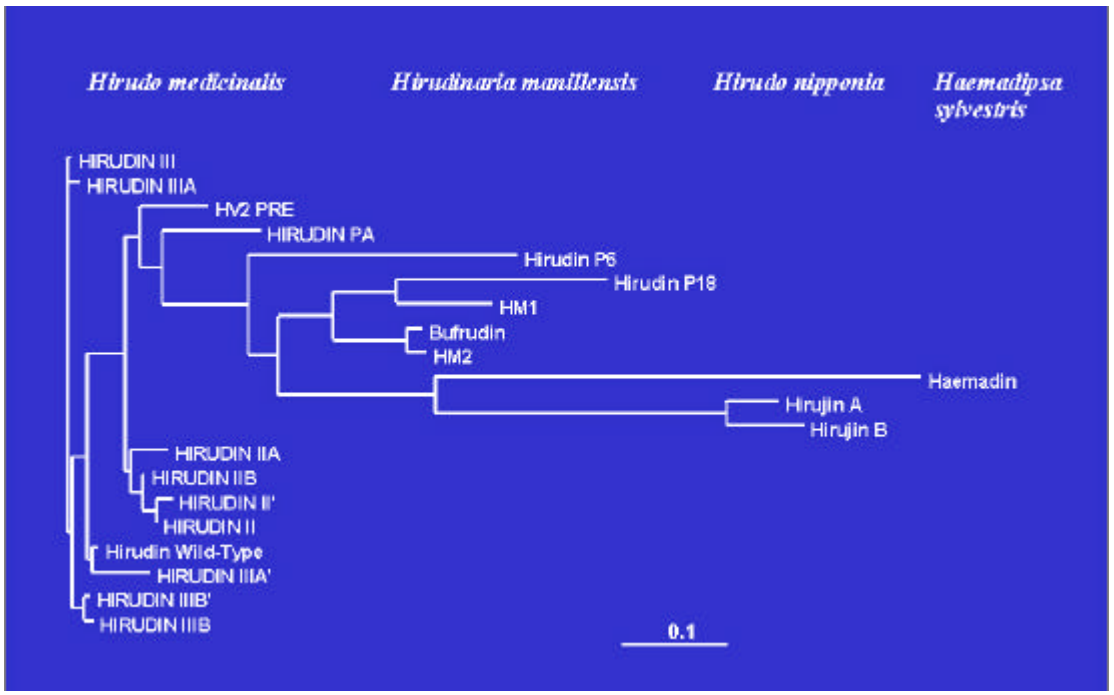


Figure 43: Phylogenetic relationship between haemadin and the 20 known hirudin variants. The phylogenetic relationships were determined by ClustalW from 21 aligned sequences and the tree was displayed with TreeView software (version 1.5.2). The scale "0.1" displays 0.1 amino acid substitutions per residue.

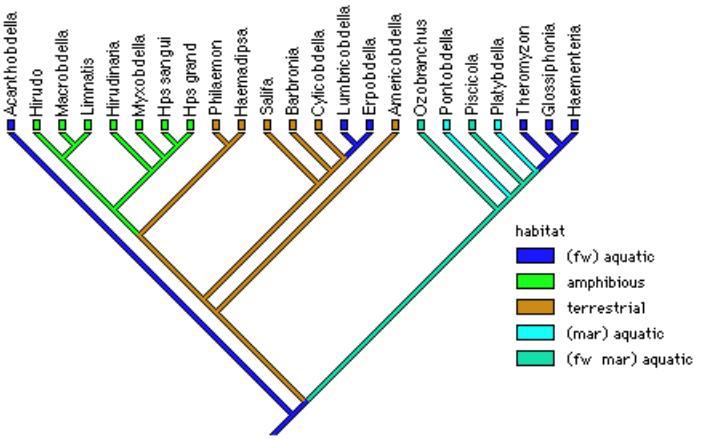


Figure 44: Evolutionary history of leech species with habitats. (fw = freshwater, mar = marine). The depiction is the most parsimonious optimisations of life-history characteristics of leeches and the resulting cladogram is that of Siddall and Burreson (Siddall and Burreson, 1995).

7.1.5. Active site binding of haemadin compared to hirudin

Biochemical studies conducted with another leech-derived inhibitor, haemadin, pointed to some differences in the detailed thrombin-inhibitor contacts, within an overall conserved inhibition mechanism. The current crystal structure shows that the N-terminal loop of haemadin forms indeed a non-canonical, hirudin-like parallel β-sheet with thrombin residues Ser 214 to Gly 216. Interestingly, the N-terminal amino acid residues of haemadin seem to be better adapted to the active-site features of thrombin than those of *Hirudo medicinalis*' hirudin (See figure 41). In particular Arg 2I is strongly preferred over a valine due to its favourable interaction with Asp 189 at the bottom of the S1 specificity pocket. Experimental data confirms the preference for a basic arginine side chain, as the recombinant hirudin variant Val 2H→Arg possesses a nine-fold higher affinity to thrombin compared with the wild-type form (Betz *et al.*, 1992). The following Phe 3I seems to be more appropriate than the conserved Tyr 3H of hirudin to occupy the hydrophobic S4 pocket. Once again mutational analysis are consistent with this proposal, as the Tyr3Phe hirudin mutant possesses a six-fold lower K_i value than wild-type hirudin (Lazar *et al.*, 1991). Amino acid residues at positions two and three are therefore optimised in haemadin for non-canonical binding to the active site of thrombin.

7.1.6. Divergence of haemadin binding from hirudin

Additional interactions made by haemadin's unique four-residue insertion Gly 4I-Met-Gly-Lys 7I, in particular by residues Met 5I and Lys 7I, contribute to strengthen the binding of the inhibitor to thrombin. Unexpectedly, the presence of this insertion and the divergence of the main chains en route to the different exosites excludes any further similarities between both thrombin-inhibitor complexes. Most strikingly, the shorter C-terminal tail of haemadin would not be able to reach exosite I of its cognate thrombin molecule (as shown in figure 38). This essential binding site for thrombin substrates and inhibitors is occupied by hirudin's C-terminal tail (Grütter *et al.*, 1990; Rydel *et al*. 1990); the same is also true for the isolated C-terminal peptide, hirugen (Skrzypczak-Jankun *et al.*, 1991). The interaction of the corresponding haemadin tail

with exosite II clearly differentiates haemadin from other members of the hirudin family. This constitutes one of the most dramatic structural and functional rearrangements observed in inhibitors from organisms so closely related phylogenetically.

7.1.7. The unique binding features of haemadin

It is noteworthy that the disulphide-rich core of haemadin is more than just a rigid connector between the extended N- and C-terminal peptides, since it is involved in major contacts with thrombin's 60-, 96-, 174- and 222-loops. The presence of a unique triplet of glutamate residues Glu 54I to Glu 56I further points to the adaptation of haemadin to the more basic exosite II of thrombin. In spite of the overall complementarity of electrostatic potentials, however, only Glu 49I is involved in direct ion pairs with thrombin residues Arg 93 and Arg 101 (See figure 15). Both residues form, together with Arg 97, a positive surface patch that mediates physiologically important interactions with negatively charged glycosaminoglycans such as heparin (Ye *et al.*, 1994; Sheehan and Sadler, 1994). It is tempting to speculate that optimal adaptation to thrombin's active site features along with several insertions-deletions within a rather conserved disulphide-stabilised core (See figure 39) have transformed hirudin into an inhibitor exhibiting a markedly different mechanism, haemadin.

7.1.8. Inferences from binding studies in solution

Even though the shorter C-terminal tail of haemadin would not allow simultaneous binding to the active site and anion-binding exosite I, the involvement of this acidic peptide in 'hirudin-like' contacts with a neighbouring thrombin molecule (Figs 12 and 42) prompted us to analyse the character of the thrombin-haemadin interactions in solution. We have recently shown that the thrombin inhibitor triabin binds solely to the fibrinogen-recognition exosite of the proteinase, leaving its active site unoccupied (Fuentes-Prior *et al.*, 1997). Therefore, haemadin and triabin would not compete with each other for thrombin binding. Band shift studies (as shown in section 6.2.) showed

that this is indeed the case, and that a ternary haemadin-thrombin-triabin complex could be isolated.

A similar line of reasoning applies to the active thrombomodulin fragment comprising the three C-terminal EGF-like repeats (TME456). Exosite I is blocked by domains TME5 and TME6, but no contacts are made with either the active site region or with exosite II (Fuentes-Prior *et al.*, 2000). Again, observation of a ternary haemadin-thrombin-TME456 complex is in accord with the occupation of exosite II by haemadin. More relevantly for future therapeutic applications, the K_i value determined for the inhibition of the thrombin-TME456 complex by recombinant haemadin is in the same order of magnitude as the figure determined for free human α-thrombin. This observation is in line with the absence of major allosteric rearrangements in the active site of thrombomodulin-bound thrombin (Fuentes-Prior *et al.*, 2000). The slight decrease in inhibitory potency might result from minor clashes between the C-terminal residues of the inhibitor and thrombomodulin residues located in close proximity to the "upper" part of exosite I.

The contribution of positively charged residues of thrombin exosite II to haemadin binding seems to closely correspond to the interactions observed in the thrombin-haemadin crystal structure. In particular, the replacement of three basic amino acids in exosite II (Arg 93, Lys 240, and Lys 233) with glutamic acid led to ten, five and threefold reductions in the inhibition constants, whereas exosite I mutants show virtually no change in inhibition constants.

Our preliminary results also indicate that haemadin does not bind to the important intermediate activation form meizothrombin. This observation is explained by the occupation of exosite II by the second kringle domain of meizothrombin (Martin *et al.*, 1997), which covers a surface on the catalytic domain about six-fold larger than the interface between haemadin's tail and thrombin. Superposition of the (desF1) meizothrombin structure on that of α-thrombin-haemadin reveals that residues of the inhibitor C-terminal to Pro 43I would not be able to contact exosite II, besides severe clashes involving exposed side chains of the inhibitor (e.g., Tyr 28I) and kringle 2 of meizothrombin. In contrast, exosite I remains free in meizothrombin, and hirudin binds to this intermediate form as strongly as it does to α-thrombin (Fischer *et al.*, 1998).

7.1.9. Haemadin and the search for antithrombotics

The search for heparin substitutes in light of the drawbacks of heparinotherapy is an area of intensive research. A recent study reports the synthesis of novel heparin mimetics (Petitou *et al.*, 1999). Peptides and peptide mimetics that target proteinase regions distant from the active site (exosites) may also play an important role as antithrombotic drugs. This opens promising avenues, not only for thrombin, but it has also recently been shown in the case of coagulation factor VIIa (Dennis *et al.*, 2000).
The combined crystallographic and biochemical evidence conclusively establishes that haemadin, binds to the heparin-binding exosite but does not interact with the fibrinogen-recognition exosite. This feature allows haemadin to inhibit not only free α-thrombin, but also the thrombomodulin-bound proteinase. From the viewpoint of the development of novel antithrombotics, haemadin offers therefore a unique advantage over current concepts based on hirudin and hirudin-like molecules. Membrane-bound active intermediates formed during the conversion of prothrombin to α-thrombin seem to play major physiological roles (Côté *et al.*, 1997; Doyle and Mann, 1990). More relevantly, meizothrombin possesses a six-fold higher protein C cofactor activity than α-thrombin, pointing to a probable anticoagulant role *in vivo* (Côté *et al.*, 1997). Haemadin binding only to forms possessing a freely accessible exosite II would selectively target circulating α-thrombin, without interfering with the anticoagulant and possibly also antifibrinolytic activities of meizothrombin. Finally, the ability of the C-terminal peptide of haemadin to bind a second thrombin molecule could become relevant at high thrombin concentrations, as might be found in the clot.

7.2. Kinetics of the human α-thrombin-haemadin complex

7.2.1. Introduction

The aim of this work was to determine the component of the binding energy of the human α-thrombin-haemadin complex that is due to ionic interactions, and evaluate its ionic strength dependence. In particular, how much of these ionic interactions are from either the acid rich tail of haemadin, or from the active site binding N-terminal peptide of haemadin was also investigated. Also considered, was how modifications of thrombin or haemadin effected the rate of association between them and how these effects themselves were affected by ionic strength.

7.2.2. Comparison of kinetic profile with hirudin

It is useful when referring to haemadin to compare it with hirudin, a 65 amino acid protein from the medicinal leech *Hirudo medicinalis* (Rydel *et al.*, 1990), due to their similarities, particularly in their folds, and in their non-canonical binding modes of their N-termini to the active site of thrombin. However these two leech derived peptides differ drastically when it comes to the binding location of their acidic C-terminal peptides, with hirudin binding to exosite I, and haemadin to the almost diametrically opposite exosite II (Richardson *et al.*, 2000).

When comparing haemadin to hirudin, it is important to note that there are two forms of hirudin; the native form contains a sulphated tyrosine residue (Y63) in its C-terminal tail and has a K_i of 2×10^{-15} M, which is about ten-fold lower than its recombinant unsulphated counterpart. The weighted mean of the ΔG°_{nio} (-65.8 ± 0.10 kJ mol^{-1}) for the human α-thrombin-haemadin complexes derived from the figures in table 7 is comparable to the figure determined for human α-thrombin-hirudin complexes (-63.7 ± 0.1 kJ mol^{-1}) (Stone *et al.*, 1989). However, when the K_i values at infinite ionic strength are calculated from the figures in table 7, and from the corresponding figures for hirudin contained within (Stone *et al.*, 1989) using equation [5], the respective K_i values for the human α-thrombin-haemadin complex and the human α-thrombin-recombinant hirudin-complex are 2.97×10^{-12} M and 1.87×10^{-11}

M. This shows that the K_i of recombinant hirudin is 6.3 times higher than haemadin, indicating haemadin is the better binder at infinite ionic strength. The situation at blood ionic strength (around 154 mM) is slightly different with both the human α-thrombin native and recombinant hirudin complexes having higher ΔG°_b values of -79.9 kJ and -74.4 kJ respectively, compared to -69.9 kJ for the human α-thrombin-haemadin complex. Using equation [5] these figures convert to K_i values of 3.53×10^{-14}, 2.91×10^{-13} and 5.65×10^{-13} M respectively, with the K_i of recombinant hirudin (which is the form that has been approved for use in the clinic (Salzet, 2001)) being about half that of haemadin at this ionic strength. The comparison of k_{on} values for these three proteins with human α-thrombin at this physiological ionic strength shows a slightly lower k_{on} value for haemadin (4.28×10^7 M^{-1} s^{-1} compared to 6.39×10^7 M^{-1} s^{-1} for recombinant hirudin), with the figure for native hirudin being 1.16×10^8 M^{-1} s^{-1}. These k_{on} values for thrombin inhibitors have been shown to be more important for therapeutic usage than the K_i value for *in vivo* thrombin inhibition, due to the concentration of thrombin in the thrombus (Elg, *et al.*, 1997; Stone and Tapparelli, 1995). The weighted mean $k_{on\infty}$ for the human α-thrombin-haemadin complexes from table 8, compared to the equivalent figure for human α-thrombin-hirudin complexes (Stone *et al.*, 1989) is $1.08 \pm 0.08 \times 10^7$ M^{-1} s^{-1} against $1.32 \pm 0.25 \times 10^6$ M^{-1} s^{-1}, or 8 times higher. This indicates that the high k_{on} of the human α-thrombin-haemadin complex is less dependant on ionic forces than is the case for the human α-thrombin-hirudin complex.

7.2.3. Interpretation of ionic effects

Apparent from tables 2 and 3 is that replacement of residues involved in this binding does not lead to an equal energy loss upon haemadin binding for each mutant. For example, there is a 13.95 ± 1.89 kJ energy loss on replacement of R93 with a glutamate which leads to virtual abolition of the ΔG°_{ion0} for the interaction of this mutant thrombin with haemadin. However, other mutants examined in this study, showed less of an energy loss (see tables 2 and 3), which suggests that the ionic interactions are specific in nature when it comes to complex formation. These results are in line with those from hirudin mutants, where ionic residues mutated to neutral

residues caused different changes in the ΔG°_{b} values in their complexes with human α-thrombin (Braun *et al.*, 1988).

The k_{off} figures for all the human α-thrombin-haemadin complexes tested were similar having a weighted mean of $2.73 \pm 0.40 \times 10^{-5}$ s^{-1}. Analysis of the k_{off} values at differing ionic strengths revealed essentially no dependence on ionic strength, and the weighted mean figures increased slightly as the K_i of the human α-thrombin-haemadin complex increased (see table 8). For the remaining human α-thrombin-haemadin complexes the k_{off} values were very similar, with a weighted mean of $2.45 \pm 0.49 \times 10^{-5}$ s^{-1}, which is close to the values for both the human α-thrombin native and recombinant hirudin complexes which are 0.98×10^{-5} s^{-1} and 3.17×10^{-5} s^{-1} respectively (Braun *et al.*, 1988). This shows that the major kinetic difference between the complexes lies in the association rate constants (k_{on}), which for both complexes have different profiles when plotted against ionic strength.

7.2.4. Kinetics of haemadin (1-40)

Removal of the tail of haemadin using formic acid causes a drastic 20,000 fold decrease in the K_i of the inhibitor, this figure corresponding to a 24.58 kJ decrease in the binding energy. This decrease is much larger than the binding energy of all the ionic binding contributions in the human α-thrombin-haemadin complex (ΔG°_{ion}), which at the same ionic strength (0.0695) is equal to -5.83 ± 0.28 kJ mol^{-1}, this figure being calculated from the figures in table 7 using equation [9]. This five-fold disparity indicates that the C-terminal tail of haemadin contributes more than just ionic interactions to the energy of binding. This additional loss of energy may be attributable to the large reduction in the k_{on} of the complex, which decreases from $9.32 \pm 1.65 \times 10^{7}$ M^{-1} s^{-1} for the human α-thrombin-haemadin complex, to $2.76 \pm 0.22 \times 10^{5}$ M^{-1} s^{-1} (338-fold decrease) for the human α-thrombin-haemadin (1-40) complex at an ionic strength of 0.0695. The corresponding change in k_{off} values from $4.54 \pm 0.84 \times 10^{-5}$ s^{-1} to $1.36 \pm 0.12 \times 10^{-3}$ s^{-1} is much less (60-fold increase), indicating that the C-terminal tail of haemadin has greater implications for the k_{on} than the k_{off}.

The K_i of haemadin (1-40) is lower when compared to other hirudin variants which have been cleaved enzymatically to remove the majority of the tail region, leaving only a few amino acid "overhang" from the main body of the inhibitor. In this respect the human α-thrombin-haemadin (1-40) complex having a K_i of $4.93 \pm 0.14 \times 10^{-9}$ M is about two magnitudes stronger than the equivalent peptide complexes from HM2 (1-41) ($K_i = 400 \pm 50 \times 10^{-9}$ M (Vindigni *et al.*, 1994)) and hirudin (1-43) ($K_i = 299 \pm 12 \times 10^{-9}$ M (Chang *et al.*, 1990)). Of further interest is that haemadin (1-40) binding is virtually unaffected by ionic strength (see table 9), with a modest 1.15 fold increase upon increasing the ionic strength from 0.0695 to 0.5195. This indicates that the majority, if not all of the ionic interactions relevant for complex formation come from the tail region of haemadin. Further proof for this hypothesis comes from experiments using DIP-thrombin with intact haemadin, which unlike haemadin (1-40) has a dependency on ionic strength. The difference in binding energies for the DIP-thrombin-haemadin complex at ionic strengths of 0.0695 and 0.5195 calculated using equation [1] is 4.82 kJ mol^{-1}, which is similar to the difference in the binding energies of the human α-thrombin-haemadin complex at these two ionic strengths, being 4.56 kJ mol^{-1}. Yet further proof comes from the fact that the k_{on} values for haemadin (1-40) are virtually unchanged at ionic strengths of 0.0695 and 0.5195, with the figure being 12% lower at the higher ionic strength (see table 9).

7.2.5. The contribution of haemadin's C-terminal peptide

The fact that haemadin's C-terminal tail is responsible for most if not all of the -16.96 ± 0.96 kJ of ionic interactions at an ionic strength of zero (this figure being the weighted mean of the values from tables 2 and 3) allows us to estimate how many ionic interactions take place between the tail region of haemadin and exosite II of thrombin. Using a figure of -4.0 kJ per participating ionic contribution, as was calculated for the interaction between the tail of hirudin and exosite I of thrombin (Stone et al., 1989) (The crystal structure of the human α-thrombin-hirudin complex (1HTC, RCSB protein data bank) contains almost no direct salt bridges, but contains several indirect ones), this converts to approximately four residues playing a role in interactions with exosite II. This figure is close to our observations where the four

mutants present in tables 2 and 3 (Arg 93, Lys 236, Arg 233 and Lys 240), as well as Arg 101 (as seen from the crystal structure) are all implicated in ionic binding of haemadin to thrombin. A more accurate assessment comes from the fact that by Debye-Hückel theory the constant C_1 should be equal to one third of the separation distance between the charges (Stone et al., 1989). Using the weighted mean of the C_1 values for the native human α-thrombin-haemadin complex (2.29 ± 0.27) to obtain this distance, and substituting it into equation [11] as the value for d_{AB}, with 80 (the value for water) being the value for *e*, gives an average energy of an ionic interaction as -2.53 kJ mol^{-1}. If the ΔG°_{ion0} of the human α-thrombin-haemadin complex (-16.96 ± 0.96 kJ) is divided by this number, it yields an even higher number, 6.7, as the number of ionic interaction in the complex. Interestingly this number is close to the value derived from reported data from the human α-thrombin-recombinant hirudin complex (Stone et al., 1989), where the result is 6.25. The fact that the C_1 value for the human α-thrombin-haemadin complex is approximately twice the value for the human α-thrombin-recombinant hirudin complex (Stone et al., 1989) shows that the C-terminal tail of haemadin isn't as tightly bound by ionic forces. However, when the distance between the charges is calculated for the human α-thrombin-hirudin complex using equation [11], the distance is 4.3 Å (Stone et al., 1989), which is 66% of the figure for the human α-thrombin-haemadin complex. These distances appear to be correct when the X-ray structures are inspected, with the C-terminal tail of hirudin being closer to thrombin than the C-terminal tail of haemadin (1E0F, 1HTC, RCSB protein data bank entries).

Knowing that most of the ionic interactions in the human α-thrombin-haemadin complex come from the C-terminal tail, suggests that the large ΔG°_{b} decrease, with virtual abolition of all ionic interactions upon mutation of R93 to a glutamate, is caused by a complete disruption of the tail segment of haemadin from exosite II of thrombin. As can be seen from figure 13, residue R93 is located close to residue R101, which has been shown to from a salt bridge with residue E49I of haemadin in the X-ray crystal structure (Richardson *et al.*, 2000). It is conceivable that when R93 is mutated into a glutamate, this residue makes a salt bridge with residue R101. This process would eliminate two basic exosite II residues from interacting with the acidic groups on haemadin's tail, which would then not be anchored at its fulcrum, and would disrupt any interactions further on down the C-terminal tail.

7.2.6. The contribution of haemadin's N-terminal peptide

Concerning the amino terminal peptide of haemadin, which inserts into the active site, blocking the active site with diisopropyl fluorophosphate causes a roughly 72,000 fold reduced affinity for thrombin, this figure being higher than the 20,000 fold reduced affinity haemadin (1-40) has for thrombin. This indicates that the amino terminal peptide of haemadin is more important for binding than the tail region. The synthetic amino terminal peptides H-IRFGMGKV-NH_2 and H-IRFGMGKVP-NH_2 based on the first 8 and 9 amino acids of haemadin respectively, are reasonably strong inhibitors of thrombin having K_i values in the micromolar range. Using equation [5], their K_i values convert to ΔG°_b values of -31.31 kJ and -31.52 kJ respectively, and it is interesting that these values closely match the loss of free energy due to thrombin being blocked by diisopropyl fluorophosphate, which for the human α-thrombin-haemadin complex is 27.74 kJ. The similarity in these binding energies also suggests that the binding of the entire N-terminal peptide is abolished by the presence of the diisopropyl phosphate group in the active site, and this is not surprising since as shown from the X-ray crystal structure, the N-terminal peptide's configuration involves intramolecular contacts whilst it is in the active site (Richardson *et al.*, 2000). The loss of binding energy of the DIP-thrombin-haemadin complex is also much greater than the figure calculated for the DIP-thrombin-hirudin complex, which loses 18 kJ mol^{-1} binding energy due to the presence of the diisopropyl group in the active site (Stone *et al.*, 1987). This indicates that haemadin binds more strongly to the active site than hirudin, which would also be expected, especially as there is a salt bridge between its Arg 2I and Asp189 of thrombin's S1 pocket, which is the amino acid expected for this specificity pocket.

7.2.7. Conclusions

In conclusion the contribution of ionic interactions in the human α-thrombin-haemadin complex is much less than for the human α-thrombin-hirudin complex, with it relying much less on ionic interactions in forming a tight complex and in its rate of association with thrombin. However at least five residues have been implicated in binding the C-terminal tail of haemadin by thrombin, and overall the same number of residues are implicated as being involved in ionic bonds for both haemadin and hirudin with thrombin, the major difference being that the estimated binding distances are 50 to 100% greater in the case of haemadin. So, as is shown in figure 15, the location of haemadin's C-terminal tail is approximately correct in terms of its location to residues it is known to interact with, and in terms of its distance from them. As has also been shown, most of the ionic interactions in the human α-thrombin-haemadin complex comes from its C-terminal acidic tail, but it seems that unlike hirudin these ionic residues are not as vital for its high rate of association with thrombin, which both inhibitors achieve through their C-terminal tails. Thus the tail of haemadin can be seen as an orienting structure but using less long range ionic forces to orient it than hirudin. This suggests that other factors, such as some form of coopravity in binding, play a role in the rate of association, as has been seen for ionic residues in the human α-thrombin-hirudin complex (Myles *et al.*, 2001). In this way the haemadin's C-terminal tail may differ from hirudin's which is in a stable conformer when it binds to thrombin (Ni *et al.*, 1992), whereas haemadin's may actively seek an orientation, and this seeking behaviour may also apply to its N-terminal peptide.

7.3. The bovine a-thrombin-sTTI complex

7.3.1. The nature of sTTI binding to thrombin

When the tsetse thrombin inhibitor (TTI) was first isolated from *Glossina morsitans morsitans* salivary gland extracts, the inhibitor was characterised to its primary structure by Edman degradation (Cappello *et al.*, 1997). Subsequent production of the peptide by solid phase methodology or recombinant expression in *E. coli* resulted in a peptide of the same primary sequence, but of low efficacy against thrombin. This change in efficacy was probably not due to lack of a post translational modification, as, for the native inhibitor, this would have been detected by an increased molecular weight, compared to the calculated molecular weight, in the mass spectrum. Alternatively the presence of a *d*-amino acid could also explain this discrepancy, however with the cDNA sequence being identical to the primary sequence this can be discounted. This seems to leave the possibility of a folding problem as the major reason for lower activity. This is not very likely as secondary structure is usually not a feature of such short peptides. One explanation may be, that the presence of four proline residues in the inhibitor, may cause certain structural changes due to the cis and trans conformations around the planar peptide bonds in these residues. To date there is no evidence to support this conclusion.

As to the structure of the inhibitor binding to thrombin, the structure must be taken tentatively as the dimer structure represented in figure 35 is now known to be non-physiological due to the results from gel filtration experiments (See section 6.5.2.) which suggested the form in solution was, in fact, a monomer.

The presence of the unexpected binding mode of Arg 31I shown in figure 36 which instead of binding in the S1 specificity pocket binds instead to Glu 217 has indeed been confirmed as being correct (see section 6.5.1.) with a 9.25 kJ mol^{-1} difference in binding energies. This leads us to believe that the active site binding portion of the inhibitor is essentially correct. As to the binding site of the rest of the peptide, experiments with triabin detailed that triabin eventually displace sTTI from thrombin suggesting that exosite I is important for sTTI binding to thrombin.

7.3.2. Structure based model of sTTI binding to bovine a -thrombin

With the dimer shown in figure 35 having its active sites placed closed together it is possible to model the sTTI peptide binding to a single bovine thrombin molecule with what has been built. For the modelling most of the peptide remained in its built positions, but the positions of three residues had to be changed considerably. The residues in question were Ile 27I, Pro 28I and Gly 29I, with Gly 29I being the most displaced residue. In order to make the peptide bend round sharply as shown in figure 38 the residues had to be included in a ß turn. Subsequent minimisation in the program XPLOR without using an X-ray term caused the modelled structure to be energy minimised with force fields applied. The resulting modelled structure is shown in figure 45.

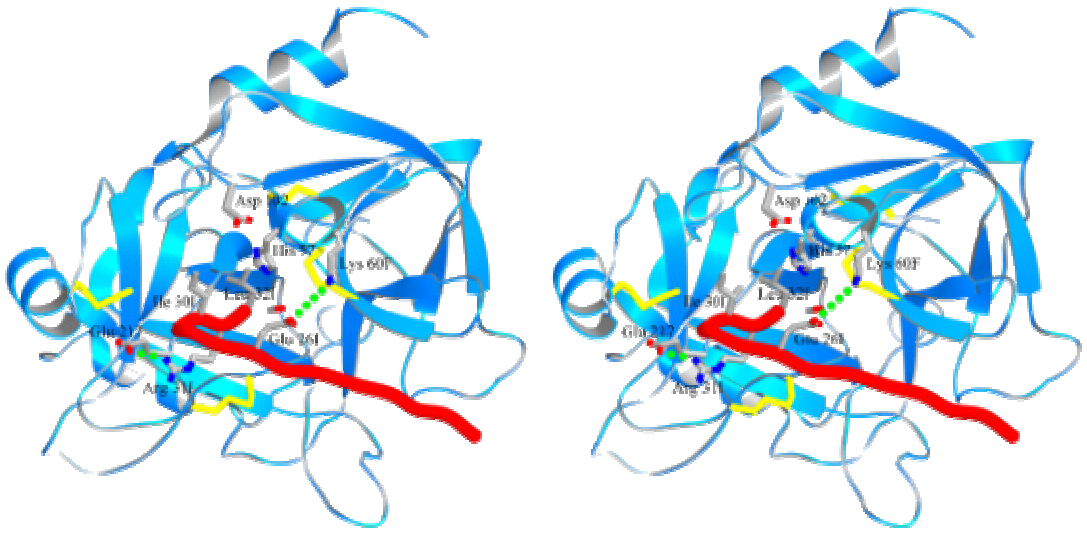


Figure 45: Stereo ribbon representation of the 1:1 model of sTTI binding to bovine α-thrombin. Shown explicitly are the side chains (element coloured) of the catalytic triad of thrombin, along with the first three C-terminal residues of sTTI and other important interacting residues. Shown as green dotted lines are the two major ionic interactions of the inhibitor. The majority of the model was built using structural information, the remaining bridging region between the active site and the "tail" of sTTI was built in MAIN and energy minimised using the program XPLOR.

8. BIBLIOGRAPHY

Adamson, A. W. (1979) *Textbook of Physical Chemistry*, pp 461-478, Academic Press, New York.

Bailey, K.F., Bettelheim, R., Lorand, L. and Middlebrook, W.R. (1951) Action of thrombin in the clotting of fibrinogen. *Nature 167*, 233-234.

Banner, D. (2001) Personal communication. XVIIIth Congress of the International Society on Thrombosis and Haemostasis (ISTH), July 6-12 2001, Paris, France.

Becker, R.C., Bovill, E.G., Seghatchian, M.J. and Samama, M.M. (1998) Pathobiology of thrombin in acute coronary syndromes. *American Heart Journal 136*, S19-31.

Berliner, L.J. and Shen, Y.Y. (1977) Physical evidence for an apolar binding site near the catalytic center of human alpha-thrombin. *Biochemistry*. *16*, 4622-4626.

Betz, A., Hofsteenge, J. and Stone, S.R. (1992) Interaction of the N-terminal region of hirudin with the active-site cleft of thrombin. *Biochemistry 31*, 4557-4562.

Bode, W. and Huber, R. (1976) Induction of the bovine trypsinogen-trypsin transition by peptides sequentially similar to the N-terminus of trypsin. *FEBS Letters. 68*, 231-236.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge J. (1989) The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO Journal. 8*, 3467-3475.

Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.*, *204*, 433-451.

Bode, W., Turk, D. and Karshikov A. (1992) The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Science. 1,* 426-471.

Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature 385*, 787-793.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry 72*, 248-254.

Braun, P. J., Dennis, S., Hofsteenge, J., and Stone, S. R. (1988) Use of site-directed mutagenesis to investigate the basis for the specificity of hirudin. *Biochemistry*. *27,* 6517-6522.

Britton, D. (1972) Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. *Acta Crystallogr. A 28*, 296-297.

Broze, G.J. (1996) Thrombin-dependent inhibition of fibrinolysis. *Current Opinions in Haematology 3*, 390-394.

Brünger, A.T. (1991) Crystallographic phasing and refinement of macromolecules. *Curr. Opin. Struct. Biol. 1*, 1016-1022.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren GL. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D 54*, 905-921.

Cappello, M., Bergum, P.W., Vlasuk, G.P., Furmidge, B.A., Pritchard, D.I. and Aksoy, S. (1996) Isolation and characterization of the tsetse thrombin inhibitor: a potent antithrombotic peptide from the saliva of *Glossina morsitans morsitans*. *American Journal of Tropical Medicine & Hygiene 54*, 475-480.

Cappello, M., Li, S., Chen, X.O., Li, C.B., Harrison, L., Narashimhan, S., Beard, C.B. and Aksoy, S. (1998) Tsetse thrombin inhibitor - bloodmeal-induced expression of an anticoagulant in salivary glands and gut tissue of *Glossina morsitans morsitans*. *Proc. Natl. Acad. Sci. U S A 95*, 14290-14295.

Carter, C.W., Jr. and Carter, C.W. (1979) Protein crystallization using incomplete factorial experiments. *Journal of Biological Chemistry*. *254*, 12219-12223.

Cate, H. (2000) Pathophysiology of disseminated intravascular coagulation in sepsis Critical Care Medicine. *28*, S9-11.

Catti, M. and Ferraris, G. (1976) Twinning by merohedry and X-ray crystal structure determination. *Acta Crystallogr. A 32*, 163-165.

CCP4 (1994) The CCP4 Suite; Collaborative Computational Project Number 4. *Acta Crystallogr. D 50*, 760-763.

Chandra, N., Brew, K. and Acharya, K. R. (1998) Structural evidence for the presence of a secondary calcium binding site in human alpha-lactalbumin. *Biochemistry 37*, 4767-4772.

Chandra, N., Acharya, K.R. and Moody, P.C.E. (1999) Analysis and characterization of data from twinned crystals. *Acta Crystallogr. D 55*, 1750-1758.

Ciallella, J.R., Figeuiredo, H., Smith-Swintosky, V. and McGillis, J.P. (1999) Thrombin induces surface and intracellular secretion of amyloid precursor protein from human endothelial cells. *Thrombosis and Haemostasis 81*, 630-637.

Contreras-Martel, C., Martinez-Oyanedel, J., Bunster, M., Legrand, P., Piras, C., Vernede, X. and Fontecilla-Camps, J. C. (2001) Crystallization and 2.2 A Resolution Structure of R-Phycoerythrin from Gracilaria Chilensis: A Case of Perfect Hemihedral Twinning. *Acta Crystallogr. D57*, 52-60.

Coughlin, S.R. (1999) How the proteinase thrombin talks to cells. *Proc. Natl. Acad. Sci. U S A 96*, 11023-11027.

Côté, H.C., Bajzar, L., Stevens, W.K., Samis, J.A., Morser, J., MacGillivray, R.T. and Nesheim, M.E. (1997) Functional characterization of recombinant human meizothrombin and meizothrombin (desF1). Thrombomodulin-dependent activation of protein C and thrombin-activatable fibrinolysis inhibitor (TAFI), platelet aggregation, antithrombin-III inhibition. *Journal of Biological Chemistry 272*, 6194-6200.

Covey, T.R., Bonner, R.F., Shushan, B.I. and Henion, J. (1988) The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Communications in Mass Spectrometry*. *2*, 249-256.

Cox, J. D., Kim, N. N., Traish, A. M. and Christianson, D. W. (1999) Arginase-Boronic Acid Complex Highlights a Physiological Role in Erectile Function. *Nat. Struct. Biol. 6*, 1043-1047.

Dahlbäck, B. (2000) Blood coagulation. *Lancet 355*, 1627-1632.

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

De Filippis, V., Russo, I., Vindigni, A., Di Cera, E., Salmaso, S. and Fontana A. (1999) Incorporation of noncoded amino acids into the N-terminal domain 1-47 of hirudin yields a highly potent and selective thrombin inhibitor. *Protein Science. 8*, 2213-2217.

Dennis M.S., Eigenbrot, C., Skelton, N.J., Ultsch, M.H., Santell, L., Dwyer, M.A., O'Connell, M.P., and Lazarus, R.A. (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature 404*, 465-470.

De Simone, G., Lombardi, A., Galdiero, S., Nastri, F., Della Morte, R., Staiano, N., Pedone, C., Bolognesi, M. and Pavone, V. (1998) Hirunorms are true hirudin mimetics. The crystal structure of human α-thrombin-hirunorm V complex. *Protein Science 7*, 243-253.

Doyle, M.F. and Mann, K.G. (1990) Multiple active forms of thrombin. IV. Relative activities of meizothrombins. *Journal of Biological Chemistry 265*, 10693-10701.

Edman, P. and Henschen, A. (1975) Sequence determination. Protein sequence determination. Springer Press, Heidelberg *2*, 232-279.

Electricwala, A., Sawyer, R.T., Jones, C.P. and Atkinson, T. (1991) Isolation of thrombin inhibitor from the leech Hirudinaria manillensis. *Blood Coagulation & Fibrinolysis. 2*, 83-89.

Elg, M., Gustafsson, D., and Deinum, J. (1997) . The importance of enzyme inhibition kinetics for the effect of thrombin inhibitors in a rat model of arterial thrombosis. *Thrombosis & Haemostasis. 78*, 1286-1292.

Ellis, K. J., and Morrison, J. F. (1982) Buffers of constant ionic strength for studying pH-dependent processes. *Methods Enzymol. 87*, 405-426.

Engh, R. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A47*, 392-400.

Esmon, C. (2000) The protein C pathway. *Critical Care Medicine 28*, S44-48.

Evans, S.V. (1990) SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics 11*, 134-138.

Fenton, J. W., II, Fasco, M. J., Stackrow, A. B., Aronson, D. L. Young A. M., and Finlayson, J. S. (1977) Human thrombins. Production, evaluation, and properties of alpha-thrombin. *Journal of Biological Chemistry 252*, 3587-3598.

Fenton, J.W.D. (1988) Regulation of thrombin generation and functions. *Semin. Thromb. Hemos. 14*, 234-240.

Fenton, J.W., Ofosu, F.A., Brezniak, D.V. and Hassouna, H.I. (1998) Thrombin and antithrombotics. *Seminars in Thrombosis and Hemostasis 24*, 87-91.

Fischer, B.E., Schlokat, U., Himmelspach, M. and Dorner, F. (1998) Binding of hirudin to meizothrombin. *Protein Eng. 11*, 715-721.

Fisher, R.G. and Sweet, R.M. (1980) Treatment of diffraction data from protein crystals twinned by merohedry. *Acta Crystallogr. A 36*, 755-760.

Folkers, P.J., Clore, G.M., Driscoll, P.C., Dodt, J., Kohler, S. and Gronenborn, A.M. (1989) Solution structure of recombinant hirudin and the Lys-47----Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry*. *28*, 2601-2617.

Frazao C., Sieker, L., Coelho, R., Morais, J., Pacheco, I., Chen, L., LeGall, J., Dauter, Z., Wilson, K. and Carrondo, M.A. (1999) Crystallization and preliminary diffraction data analysis of both single and pseudo-merohedrally twinned crystals of rubredoxin oxygen oxidoreductase from Desulfovibrio gigas. *Acta Crystallogr. D 55*, 1465-1467.

Frazao, C., Silva, G., Gomes, C. M., Matias, P., Coelho, R., Sieker, L., Macedo, S., Liu, M. Y., Oliveira, S., Teixeira, M., Xavier, A. V., Rodrigues-Pousada, C., Carrondo, M. A. and Le Gall, J. (2000) Structure of a Dioxygen Reduction Enzyme from Desulfovibrio Gigas. *Nat. Struct. Biol. 7*, 1041-1045.

Friedrich, T., Kroger, B., Bialojan, S., Lemaire, H.G., Hoffken, H.W., Reuschenbach, P., Otte, M. and Dodt, J. (1993) A Kazal-type inhibitor with thrombin specificity from Rhodnius prolixus. *Journal of Biological Chemistry*. *268*, 16216-16222.

Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W.D., Huber, R. and Bode W. (1997) Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. *Proc. Natl. Acad. Sci. U S A 94*, 11845-11850.

Fuentes-Prior, P., Iwanaga, Y., Huber, R., Pagila, R,. Rumennik, G., Seto, M., Morser, J., Light, D.R. and Bode W. (2000) Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature. 404*, 518-525.

Gill, S. G., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*. *182*, 319-326.

Grainger, C.T. (1969) The treatment of overlapped data. *Acta Crystallogr. A 25*, 427-434.

Grütter, M,G., Priestle, J.P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J. and Stone, S.R. (1990) Crystal structure of the thrombin-hirudin complex: a novel mode of serine proteinase inhibition. *EMBO Journal. 9*, 2361-2365.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*. *166*, 557-580.

Haruyama, H. and Wüthrich, K. (1989) Conformation of recombinant desulfatohirudin in aqueous solution determined by nuclear magnetic resonance. *Biochemistry 28*, 4301-4312.

Hayes, K.L., Leong, L., Henriksen, R.A., Bouchard, B.A., Ouellette, L., Church, W.R. and Tracy, P.B. (1994) α-thrombin - induced human platelet activation results solely from formation of a specific enzyme - substrate complex. *Journal of Biological Chemistry 269*, 28606-28612.

Hellmann, K. and Hawkins, R.I. (1965) Prolixins-S and prolixin-G; two anticoagulants from Rhodnius prolixus Stål. *Nature 207*, 265-267.

Hellmann, K. and Hawkins, R.I. (1966) An antithrombin (maculatin) and a plasminogen activator extractable from the blood-sucking hemipteran, Eutriatoma maculatus. *British Journal of Haematology 12*, 376-384.

Hillig, R. C., Renault, L., Vetter, I. R., Drell, T., Wittinghofer, A. and Becker, J. (1999) The Crystal Structure of RNA1P: A New Fold for a Gtpase-Activating Protein. *Mol. Cell 3*, 781-791.

Hoffmann, A., Walsmann, P., Riesener, G., Paintz, M. and Markwardt, F. (1991) Isolation and characterization of a thrombin inhibitor from the tick ixodes ricinus. *Pharmazie 46*, 209-212.

Hogg, P.J. and Jackson C.M. (1989) Fibrin monomer protects thrombin from inactivation by heparin-antithrombin III: implications for heparin efficacy. *Proc. Natl. Acad. Sci. U S A 86*, 3619-3623.

Hong, S.J. and Kang, K.W. (1999) Purification of granulin-like polypeptide from the blood-sucking leech, *Hirudo nipponia*. *Protein Expression & Purification 16*, 340-346.

Hung, D.T., Vu, T.H., Wheaton, V.I., Ishii, K. and Coughlin, S.R. (1992) Cloned platelet thrombin receptor Is necessary for thrombin - induced platelet activation. *Journal of Clinical Investigation 89*, 1350-1353.

Ito, N., Komiyama, N. H. and Fermi, G. (1995) Structure of deoxyhaemoglobin of the antarctic fish Pagothenia bernacchii with an analysis of the structural basis of the root effect by comparison of the liganded and unliganded haemoglobin structures. *J. Mol. Biol. 250*, 648-658.

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

Jeffrey, P. D., Strong, R. K., Sieker, L. C., Chang, C. Y., Campbell, R. L., Petsko, G. A., Haber, E., Margolies and M. N., Sheriff, S. (1993) 26-10 Fab-digoxin complex: affinity and specificity due to surface complementarity. *Proc. Natl. Acad. Sci. U S A 90*, 10310-10314.

Koyama, T., Parkinson, J.F., Aoki, N., Bang, N.U., Müller-Berghaus G. and Preissner, K.T. (1991) Relationship between post-translational glycosylation and anticoagulant function of secretable recombinant mutants of human thrombomodulin. *British Journal of Haematology 78*, 515-522.

Laidler, K. J. (1987) *Chemical Kinetics*, pp 191-194, Harper and Row, New York.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature. 227*, 680-685.

Lange, U., Keilholz, W., Schaub, G.A., Landmann, H., Markwardt, F. and Nowak, G. (1999) Biochemical characterization of a thrombin inhibitor from the bloodsucking bug Dipetalogaster maximus. *Haemostasis 29*, 204-211.

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

Lazar, J.B., Winant, R.C. and Johnson, P.H. (1991) Hirudin: amino-terminal residues play a major role in the interaction with thrombin. *Journal of Biological Chemistry*. *266*, 685-688.

Leslie, A.G.W. (1994) Mosflm User Guide, Mosflm version 5.23. MRC Laboratory of Molecular Biology, Cambridge UK.

Lu, G., Lindqvist, Y., Schneider, G., Dwivedi, U. and Campbell, W. (1995) Structural studies on corn nitrate reductase: refined structure of the cytochrome b reductase fragment at 2.5 Å, its ADP complex and an active-site mutant and modeling of the cytochrome b domain. *J. Mol. Biol. 248*, 931-948.

Luecke, H., Richter, H. T. and Lanyi, J. K. (1998) Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science 280*, 1934.

Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. and Lanyi, J. K. (1999) Structure of Bacteriorhodopsin at 1.55 Angstrom Resolution. *J.Mol.Biol. 291*, 899.

Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. and Lanyi, J. K. (1999) Structural Changes in Bacteriorhodopsin During Ion Transport at 2 Å Resolution. *Science 286*, 255-261.

LeBonniec, B.F., Guinto, E.R., Esmon, C.T. (1992) Interaction of thrombin des-ETW with antithrombin III, the Kunitz inhibitors, thrombomodulin and protein C: structural link between the autolysis loop and the Tyr-Pro-Pro-Trp insertion of thrombin. *Journal of Biological Chemistry 267*, 19341-19348.

LeBonniec, B.F., Guinto, E.R., MacGillivray, R.T.A., Stone, S.R. and Esmon, C.T. (1993) The role of thrombin's Tyr-Pro-Pro-Trp motif in the interaction of fibrinogen, thrombomodulin, Protein C, antithrombin II, and the Kunitz inhibitors. *Journal of Biological Chemistry 268,* 19055-19061.

Luecke, H., Schobert, B., Cartailler, J. P., Richter, H. T., Rosengarth, A., Needleman, R. and Lanyi, J. K. (2000) Coupling Photoisomerization of Retinal to Directional Transport in Bacteriorhodopsin. *J.Mol.Biol. 300*, 1237-1255.

Liu, L., Freedman, J., Hornstein, A., Fenton, J.W. and Ofosu, F.A. (1994) Thrombin binding to platelets and their activation in plasma. *British Journal of Haematology 88,* 592-600.

Mann, K.G. (1976) Prothrombin. *Method. Enzymol. 45*, 123-516.

Mann, Y. , Meng, M. and Fenn, C.K. (1989) Interpreting mass spectra of multiply charged ions. *Anal. Chem*. *61,* 1702-1708.

Markwardt, F. and Leberecht, E. (1959) Untersuchung über den blutgerinnung hemmenden Wirkstoff der Tabaniden. *Naturwissenschaften 46*, 17.

Markwardt, F. and Schultz, E (1960) Über einen Hemmstoff des Gerinnungs fermentes Thrombin aus blutsaugenden Raubwanzel (Reduviiden). *Naturwissenschaften 47,* 43.

Markwardt, F., Schafer, G., Topfer, H. and Walsmann, P. (1967) [Isolation of hirudin from medicinal leech]. *Pharmazie 22*, 239-41.

Markwardt, F. (1994) Coagulation inhibitors from blood-sucking animals. *Pharmazie 49,* 313-316.

Martin, P.D., Malkowski, M.G., Box, J., Esmon, C.T. and Edwards, B.F. (1997) New insights into the regulation of the blood clotting cascade derived from the X-ray crystal structure of bovine meizothrombin des F1 in complex with PPACK. *Structure 5*, 1681-1693.

Matthews, B. W. (1968) The Solvent Content of Protein Crystals. *J. Mol. Biol. 33*, 491-497.

McPherson, A. (1989) Preparation and analysis of protein crystals. (McPherson, A.,ed.), Krieger Publishing Co., Malabar, Florida.

Mercer, W. D., Winn, S. I. and Watson, H. C. (1976) Twinning in crystals of human skeletal muscle D-glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol. 10*, 277-283.

Meyer, T. E., Watkins, J. A., Przysiecki, C. T., Tollin, G., and Cusanovich, M. A. (1984) Electron-transfer reactions of photoreduced flavin analogues with c-type cytochromes: quantitation of steric and electrostatic factors. *Biochemistry 23*, 4761-4767.

Moras, D., Olsen, K. W., Sabesan, M. N., Buehner, M., Ford, G. C. and Rossmann, M. G. (1975) Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *Journal of Biological Chemistry 250*, 9137-9162.

Morrison, J. F. (1982) The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. *Trends Biochem. Sci. 7*, 102-105.

Mourey, L., Pedelacq, J. D., Birck, C., Fabre, C., Rouge, P. and Samama, J. P. (1998) Crystal structure of the arcelin-1 dimer from Phaseolus vulgaris at 1.9-A resolution. *Journal of Biological Chemistry 273*, 12914-12922.

Murray-Rust, P. (1973) Crystal structure of $[Co(NH_3)_6]_4Cu_5Cl_{17}$. *Acta Crystallogr. B 29*, 2559-2566.

Myles, T., Church, F. C., Whinna, H. C., Monard, D., and Stone, S.R. (1998) Role of thrombin anion-binding exosite-I in the formation of thrombin-serpin complexes. *Journal of Biological Chemistry 273,* 31203-31208.

Myles, T., Le Bonniec, B. F., Betz, A., and Stone, S. R. (2001) Electrostatic steering and ionic tethering in the formation of thrombin-hirudin complexes: the role of the thrombin anion-binding exosite-I. *Biochemistry 40*, 4972-4979.

Narayanan, S. (1999) Multifunctional roles of thrombin. *Annals of Clinical & Laboratory Science. 29,* 275-280.

Navaza, J. (1994) An automated package for molecular replacement. *Acta Crystallogr. A50,* 157-163.

Neu, H.C. and Heppel, L.A. (1965) The release of enzymes from Escherichia coli by osmotic shock and during the formation of spheroplasts. *Journal of Biological Chemistry 240*, 3685-3692.

Ni, F., Ripoll, D. R., and Purisima, E. O. (1992) Conformational stability of a thrombin-binding peptide derived from the hirudin C-terminus. *Biochemistry 31*, 2545-2554.

Nicholls, A., Bharadwaj, R. and Honig, B. (1993) GRASP - graphical representation and analysis of surface properties. *Biophys. J. 64*, A166.

Nienaber, V.L. and Amparo, E.C. (1996) A noncleavable retro-binding peptide that spans the substrate binding cleft of serine proteinases. Atomic structure of nazumamide A: human thrombin. *J. Am. Chem. Soc. 118*, 6807-6810.

Nienaber, J., Gaspar, A.R. and Neitz, A.W. (1999) Savignin, a potent thrombin inhibitor isolated from the salivary glands of the tick Ornithodoros savignyi (Acari: Argasidae). *Experimental Parasitology. 93*, 82-91.

Noeske-Jungblut, C., Haendler, B., Donner, P., Alagon, A., Possani, L. and Schleuning, W.D. (1995) Triabin, a highly potent exosite inhibitor of thrombin. *Journal of Biological Chemistry*. *270*, 28629-28634.

Olson, S.T. and Björk, I. (1992) Regulation of thrombin by antithrombin and heparin cofactor II. In Berliner LJ ed. *Thrombin: Structure and Function*. New York, Plenum, pp 159-217.

Pereira, Barbosa. José. Pedro. (1999) PhD Thesis, Institute of Biomedical Sciences University of Portugal, pp 90-98.

Petitou M., Hérault J.-P., Bernat, A., Driguez, P.-A., Duchaussoy, P., Lormeau, J.-C. and Herbert, J.-M. (1999) Synthesis of thrombin-inhibiting heparin mimetics without side effects. *Nature 398*, 417-422.

Priestle, J.P., Rahuel, J., Rink, H., Tones, M. and Grütter, M.G. (1993) Changes in interactions in complexes of hirudin derivatives and human α-thrombin due to different crystal forms. *Protein Science 2*, 1630-1642.

Przysiecki, C. T., Cheddar, G., Meyer, T. E., Tollin, G., and Cusanovich, M. A. (1985) Kinetics of reduction of high redox potential ferredoxins by the semiquinones of Clostridium pasteurianum flavodoxin and exogenous flavin mononucleotide. Electrostatic and redox potential effects. *Biochemistry 24*, 5647-5652.

Redinbo, M.R. and Yeates, T.O. (1993) Structure determination of plastocyanin from a specimen with a hemihedral twinning fraction of one-half. *Acta Crystallogr. D 49*, 375-380.

Rees, D.C. (1982) A general theory of X-ray intensity statistics for twins by merohedry. *Acta Crystallogr. A 38*, 201-207.

Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W. and Neurath, H. (1988) The structure of rat mast cell proteinase II at 1.9-Å resolution. *Biochemistry 27*, 8097-8105.

Richardson, J. L., Kröger, B., Hoeffken, W., Sadler, J. E., Pereira, P., Huber, R., Bode, W., and Fuentes-Prior, P. (2000) Crystal structure of the human alpha-thrombin-haemadin complex: an exosite II-binding inhibitor. *EMBO Journal 19*, 5650-5660.

Rydel, T.J., Ravichandran, K.G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C. and Fenton, J.W. 2nd. (1990) The structure of a complex of recombinant hirudin and human alpha-thrombin. *Science. 249*, 277-280.

Rydel, T.J., Tulinsky, A., Bode, W. and Huber, R. (1991) Refined structure of the hirudin-thrombin complex. *Journal of Molecular Biology*. *221*, 583-601.

Salzet, M., Chopin, V., Baert, J., Matias, I. and Malecha, J. (2000) Theromin, a novel leech thrombin inhibitor. *Journal of Biological Chemistry*. *275*, 30774-30780.

Salzet, M. Anticoagulants and inhibitors of platelet aggregation derived from leeches. (2001) *FEBS Letters 492*, 187-192.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Scacheri, E., Nitti, G., Valsasina, B., Orsini, G., Visco, C., Ferrera, M., Sawyer, R.T. and Sarmientos, P. (1993) Novel hirudin variants from the leech *Hirudinaria manillensis*. Amino acid sequence, cDNA cloning and genomic organization. *European Journal of Biochemistry*. *214*, 295-304.

Schneider, T. R., Km-Drcher, J., Pohl, E., Lubini, P. and Sheldrick, G. M. (2000) Ab Initio Crystal Structure Determination of the Lantibiotic Mersacidin. *Acta Crystallogr. D56*, 705-713.

Segel I. H. (1975) *Enzyme Kinetics*, J. Wiley & Sons, New York.

Sheehan, J.P. and Sadler, J.E. (1994) Molecular mapping of the heparin-binding exosite of thrombin. *Proc. Nat. Acad. Sci. U S A 91*, 5518-5522.

Siddall, M.E. and Burreson, E.M. (1995) Phylogeny of the euhirudinea - independent evolution of blood feeding by leeches. *Canadian Journal of Zoology*. *73*, 1048-1064.

Skrzypczak-Jankun, E., Carperos, V.E., Ravichandran, K.G., Tulinsky, A., Westbrook, M. and Maraganore, J.M. (1991) Structure of the hirugen and hirulog 1 complexes of α-thrombin. *J. Mol. Biol. 221*, 1379-1393.

Sonne, O. (1988) The specific binding of thrombin to human polymorphonuclear leucocytes. *Scandinavian Journal of Clinical & Laboratory Investigation. 48*, 831-838.

Sparks, R.A. (1997) Twinning. Programs for Indexing, Structure Refinement and Determining the Relationship Between the Twin Components. 54. *American Crystallographic Association Abstracts*, St. Louis, Missouri, USA.

Stanley, E. (1972) The identification of twins from intensity statistics. *J. Appl. Crystallogr. 5*, 191-194.

Steiner, V., Knecht, R., Grütter, M., Raschdorf, F., Gassmann, E. and Maschler, R. (1991) Isolation and purification of novel hirudins from the leech Hirudinaria manillensis by high-performance liquid chromatography. *Journal of Chromatography*. *530*, 273-282.

Steiner, V., Knecht, R., Bornsen, K.O., Gassmann, E., Stone, S.R., Raschdorf, F., Schlaeppi, J.M. and Maschler, R. (1992) Primary structure and function of novel *O*-glycosylated hirudins from the leech *Hirudinaria manillensis*. *Biochemistry 31*, 2294-2298.

Stone, S. R., Braun, P.J., and Hofsteenge, J. (1987) Identification of regions of alpha-thrombin involved in its interaction with hirudin. *Biochemistry 26*, 4617-4624.

Stone, S. R. ,Dennis, S., and Hofsteenge, J (1989) Quantitative evaluation of the contribution of ionic interactions to the formation of the thrombin-hirudin complex. *Biochemistry 28*, 6857-6863.

Stone, S. R., and Tapparelli, C. (1995) Thrombin inhibitors as antithrombotic agents: the importance of rapid inhibition. *Journal of Enzyme Inhibition 9*, 3-15.

Strube, K.H., Kroger, B., Bialojan, S., Otte, M. and Dodt, J. (1993) Isolation, sequence analysis, and cloning of haemadin. An anticoagulant peptide from the Indian leech. *Journal of Biological Chemistry*. *268*, 8590-8595.

Stubbs, M.T. and Bode, W. (1993) A player of many parts: the spotlight falls on thrombin's structure. *Thrombosis Research. 69*, 1-58.

Suzuki, K., Nishioka, J. and Hayashi T. (1990) Localization of thrombomodulin-binding site within human thrombin. *Journal of Biological Chemistry*. *265*, 13263-13267.

Szyperski, T., Guntert, P., Stone, S.R. and Wuthrich, K. (1992) Nuclear magnetic resonance solution structure of hirudin(1-51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *Journal of Molecular Biology*. *228*, 1193-1205.

Tollin, G., Cheddar, G., Watkins, J. A., Meyer, T. E., and Cusanovich, M. A. (1984) Electron transfer between flavodoxin semiquinone and c-type cytochromes: correlations between electrostatically corrected rate constants, redox potentials, and surface topologies. *Biochemistry 23*, 6345-6349.

Trame, C. B. and Mckay, D. B. (2001) Structure of Haemophilus Influenzae Hslu Protein in Crystals with One-Dimensional Disorder Twinning. *Acta Crystallogr. D57*, 1079-1090.

Turgeon, V.L. and Houenou, L.J. (1997) The role of thrombin-like (serine) proteinases in the development, plasticity and pathology of the nervous system. *Brain Research - Brain Research Reviews. 25*, 85-95.

Turk, D. (1992) Weiterentwicklung eines Programms für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD Thesis, Technische Universität, München.

Uppenberg, J., Ohrner, N., Norin, M., Hult, K., Kleywegt, G. J., Patkar, S., Waagen, V., Anthonsen, T. and Jones, T. A. (1995) Crystallographic and molecular-modeling studies of lipase B from Candida antarctica reveal a stereospecificity pocket for secondary alcohols. *Biochemistry 34*, 16838-16851.

Valegard, K., van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J. and Andersson, I. (1998) Structure of a cephalosporin synthase. *Nature 394*, 805-809.

Valenzuela, J.G., Francischetti, I.M. and Ribeiro, J.M. (1999) Purification, cloning, and synthesis of a novel salivary anti-thrombin from the mosquito Anopheles albimanus. *Biochemistry*. *38*, 11209-11215.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W. and Bode W. (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO Journal. 14*, 5149-5157.

van de Locht, A,. Stubbs, M.T., Bode, W., Friedrich ,T., Bollschweiler, C., Hoffken, W. and Huber, R. (1996) The ornithodorin-thrombin crystal structure, a key to the TAP enigma?. *EMBO Journal. 15*, 6011-6017.

van de Locht, A., Bode, W. Huber, R., LeBonniec, B.F., Stone, S.R., Esmon, C.T. and Stubbs, M.T. (1997) The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *EMBO Journal. 16*, 2977-2984.

van Koningsveld, H. (1983) Twinning by pseudomerohedry in ammonium tetrachlorozincate(II), $(NH_4)_2[ZnCl_4]$. A reinvestigation of the crystal structure at room temperature. *Acta Crystallogr. C 39*, 15-19.

van Tilburg, N.H., Rosendaal, F.R. and Bertina, R.M. (2000) Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. *Blood. 95*, 2855-2859.

Vindigni, A., De Filippis, V., Zanotti, G., Visco, C., Orsini, G., and Fontana, A. (1994) Probing the structure of hirudin from Hirudinaria manillensis by limited proteolysis. Isolation, characterization and thrombin-inhibitory properties of N-terminal fragments. *European Journal of Biochemistry 226*, 323-333.

Vitali, J., Martin, P.D., Malkowski, M.G., Robertson, W.D., Lazar, J.B., Winant, R.C., Johnson, P.H. and Edwards, B.F. (1992) The structure of a complex of bovine alpha-thrombin and recombinant hirudin at 2.8-Å resolution. *Journal of Biological Chemistry*. *267*, 17670-17678.

Walsh, K.A., Ericsson, L.H. Parmelee, D.C. and Titani K. (1981) Advances in protein sequencing. *Annual Review of Biochemistry*. *50*, 261-284.

Wei, C. H. (1969) Structural Analyses of Tetracobalt Dodecacarbonyl and Tetrarhodium Dodecacarbonyl. Crystallographic Treatments of a Disordered Structure and a Twinned Composite. *Inorg. Chem*. *8*, 2384-2397.

Weichsel A. Andersen JF. Champagne DE. Walker FA. Montfort WR. (1998) Crystal structures of a nitric oxide transport protein from a blood sucking insect. *Nature Structural Biology 5*, 304-309.

Williams, J. W., Morrison, J. F. (1979) The kinetics of reversible tight-binding inhibition. *Methods in Enzymology 63,* 437-467.

Wilson, A.J.C. (1949) The probability distribution of X-ray intensities. *Acta Crystallogr. 2,* 318-321.

Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E., Steven, A. C., Pluckthun, A. and Wlodawer, A. (2000) Novel Fold and Capsid-Binding Properties of the Lambda Phage Display Platform Protein Gpd. *Nat. Struct. Biol. 7,* 230-237.

Ye, J., Rezaie, A.R. and Esmon, C.T. (1994) Glycosaminoglycan contributions to both protein C activation and thrombin inhibition involve a common arginine-rich site in thrombin that includes residues arginine 93, 97, and 101. *Journal of Biological Chemistry 269,* 17965-17970.

Yeates, T.O. and Rees D.C. (1987) An isomorphous replacement method for phasing twinned structures. *Acta Crystallogr. A 43,* 30-36.

Yeates, T.O. (1988) Simple statistics for intensity data from twinned specimens. *Acta Crystallogr. A 44,* 142-144.

Zingali, R.B., Jandrot-Perrus, M., Guillin, M.C. and Bon, C. (1993) Bothrojaracin, a new thrombin inhibitor isolated from Bothrops jararaca venom: characterization and mechanism of thrombin inhibition. *Biochemistry*. *32*, 10794-10802.

9. APPENDIX

9.1. Table index

9.2. Figure index

9.3. Publications

Parts of this works have appeared as the following publications:

Richardson, J.L., Kroger, B., Hoeffken, W., Sadler, J. E., Pereira, P., Huber, R., Bode, W., and Fuentes-Prior, P. (2000) "Crystal structure of the human alpha-thrombin-haemadin complex: an exosite II-binding inhibitor." *EMBO Journal. 19(21)*, 5650-5660.

Richardson, J.L., Fuentes-Prior, P., Sadler, E.J., Huber, R. and Bode, W. "Characterisation of the residues in the human α-thrombin-haemadin complex: an exosite II-binding inhibitor", in press.

9.4. Congresses attended

Richardson, J.L., Pereira-Barbosa, J.P., Kröger, B., Huber, R., Bode, W. and Fuentes-Prior, P. *Crystal structure of human a-thrombin complex; a new class of thrombin inhibition ?* Poster presentation at the VIth International Symposium on Proteinase Inhibitors and Biological Control, June 9-13 1999, Brdo by Ljubljana, Slovenia.

Richardson, J.L., Kröger, B., Pereira-Barbosa, J.P., Huber, R., Bode, W. and Fuentes-Prior, P. *X-ray crystal structure of the human a-thrombin-haemadin complex*. Seminar presentation at the XVIIIth Winter School on Proteinases and their Inhibitors Recent Developments, March 8-11 2000, in Tiers, Italy.

Ploom, T., Thony, B., Yim, J., Lee, S., Nar, H., Leimbacher, W., Richardson, J., Huber, R., and Auerbach, G. *Crystallographic investigations on the mechanism of 6-pyruvoyl tetrahydropterin synthase.* Seminar presentation at the EU Network Mid Term Review Meeting, July 7 2000, Martinsried, Germany.

Richardson, J.L., Sadler, J.E, Huber, R., Bode, W. and Fuentes-Prior, P. *Characterisation of residues involved in the human alpha-thrombin-haemadin complex: an exosite II-binding inhibitor.* Seminar presentation at the XIX Winter School on Proteinases and their Inhibitors Recent Developments, February 28-March 4 2001, in Tiers, Italy.

Richardson, J.L., Sadler, J.E, Huber, R., Bode, W. and Fuentes-Prior, P. *Characterisation of residues involved in the human alpha-thrombin-haemadin complex: an exosite II-binding inhibitor.* Seminar presentation at VIIth International Symposium on Proteinase Inhibitors and Biological Control, June 16-20 2001, in Portorož, Slovenia.

Richardson, J.L., Sadler, J.E, Huber, R., Bode, W. and Fuentes-Prior, P. *Characterisation of residues involved in the human alpha-thrombin-haemadin complex: an exosite II-binding inhibitor.* Presidential prize winning poster presentation at the XVIIIth Congress of the International Society on Thrombosis and Haemostasis (ISTH), July 6-12 2001, in Paris, France.

Richardson, J.L., Sadler, J.E, Cappello, M., Huber, R., Bode, W. and Fuentes-Prior, P. *Characterisation of residues involved in the human a-thrombin-haemadin complex: an exosite II-binding inhibitor, and Structure of bovine thrombin in complex with tsetse thrombin inhibitor.* Poster presentation at the 2nd General Meeting of the International Proteolysis Society (IPS) associated with the International Conference on Proteinase Inhibitors (ICPI), October 31-November 4 2001, Freising near Munich, Germany.

9.5. Acknowledgements

I am indebted to Prof. Dr. Robert Huber for giving me the chance to begin this work, and for his friendly guidance and his scientific advice. I would similarly like to extend my thanks to Prof. Dr. Wolfram Bode for introducing me to the area of the coagulation cascade.

I would also like to extend my thanks to Dr. Pablo Fuentes-Prior for his support and advice over the past three years.

I would like to thank all the members of the Max-Planck Institute in particular Rainer Friedrich, Tobias Ullrich and Iris Fritze for their friendship during my PhD studies as well as all the members of lab K 303, past and present including Dr. Matthias Bochtler, Dr. Michael Groll, Dr. Nedilko Budisa, Dr. Ravichandran Ravishankar, Arne Ramsperger, Thomas Steiner, Jay Hun Bay, Marina Rubini, Irena Bonin and Dr. Lars Linden. Not forgetting the members of my office K 201, both past and present Dr. Hans Brandstetter, Dr. Carlos Fernandez-Catalan, Dr. David Reverter, Otto Kyrieleis, Li Chi Chang and Stephan Gerhardt.

Also to the secretaries Renate Rüller and Gina Beckmann for their supreme help with problems of a non-scientific nature.

I would like to thank my parents for their support and encouragement throughout my PhD. I will always be indebted to them for giving me a secure childhood, which enabled me to go to University and tackle the many problems of life. I would also like to express my thanks to my brother Andrew and sister Claire whom I love dearly.

I would also like to thank Anja Dietel whose patience and understanding during the last three years were a great help to me, and whom I love dearly.

I would also like to give a special thanks to both Prof. Phillip Parsons of the University of Sussex and Dr. Leslie Johnson of the University of Southampton, for cultivating my interest in science.