

Chapter 3: Homology modeling of savignin, a thrombin inhibitor from the tick *Ornithodoros savignyi**

*Part of the work presented in this chapter has been accepted for publication in *Insect Biochemistry and Molecular Biology* (Mans, Louw and Neitz, 2002a)

3.1.1 Introduction: The mechanism of serine protease activity

Serine proteases are a ubiquitous family that has been found in vertebrates, invertebrates, plants and prokaryotes. Serine proteases function as digestive enzymes (trypsin, chymotrypsin and elastase), clotting agents (blood coagulation cascade) or play a role in invertebrate immunity. Phylogenetic studies indicate that most serine proteases are homologous except for subtilisin from *B. subtilis* that attained its similar mechanism of action through convergent evolution (Perona and Craik, 1997). The central mechanism of action of serine proteases consists of a nucleophilic attack on the carboxyl group of the scissile peptide bond by a reactive serine (Fig. 3.1).

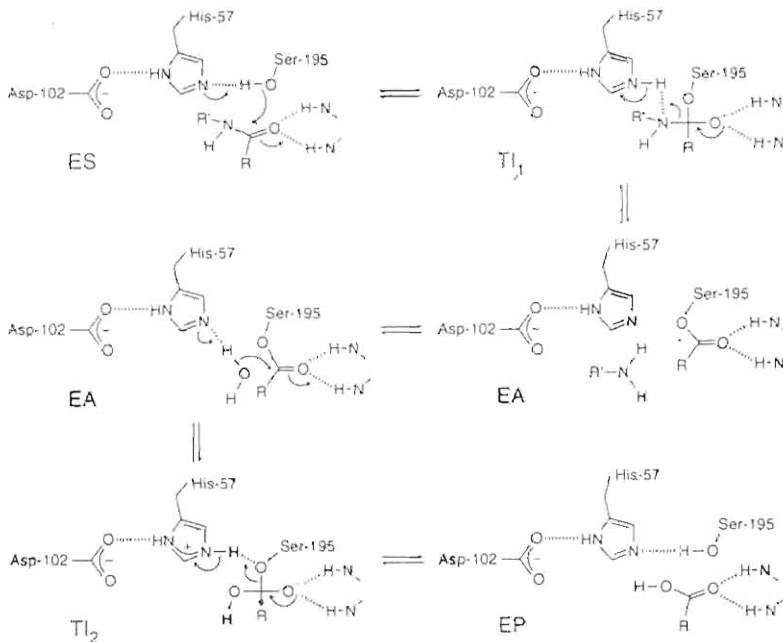


Fig. 3.1: Mechanism of action of serine proteases. The serine proteases form a noncovalent enzyme-substrate complex. Attack by the hydroxyl group of serine gives a tetrahedral intermediate that collapses to give the acylenzyme and the released amine. The acylenzyme hydrolyzes to form the enzyme-product complex via another tetrahedral intermediate followed by release of the carboxyl group. Adapted from Fersht (1999).

The active-site serine is highly reactive due to the specific conformation formed with histidine and aspartic acid (catalytic triad) that facilitates removal of the proton from serine's hydroxyl group, thus increasing the nucleophilic character of the oxygen, making serine more reactive (Fig. 3.2).

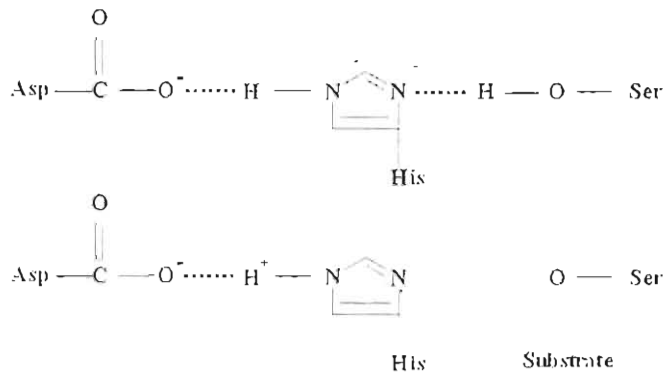


Fig. 3.2: The catalytic triad of the serine proteases. Due to the charge relay system of the catalytic triad, aspartic acid and histidine facilitates the removal of a proton from serine's hydroxyl group, allowing nucleophilic attack by serine. Adapted from Fersht (1999).

3.1.2 Specificity of serine proteases

Serine proteases are specific with regard to the sequences that are hydrolyzed. Trypsin and chymotrypsin for example hydrolyzes at the C-terminal side of basic (R, K) or aromatic residues (F, Y, W), respectively. Specificity is determined by a series of subsites in the binding pocket that recognize the scissile residue as well as the adjacent amino acids (Fig. 3.3). Loops close to the reactive site, can enhance steric control to increase specificity. Protease inhibitors that can bind tightly into these restricted conformations are not hydrolyzed because the leaving amino group is constrained and does not diffuse away from the active site of the enzyme (Fersht, 1999).

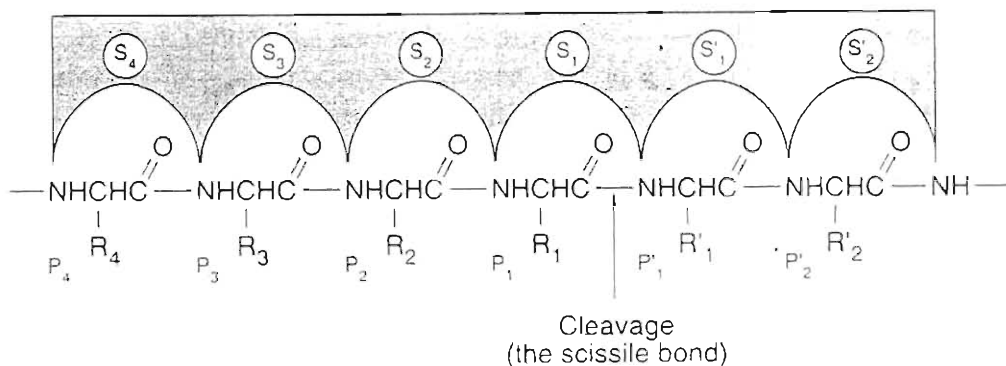


Fig. 3.3: Schechter-Berger notation for binding sites in the binding pocket (Schechter and Berger, 1967). The active site involved in the catalytic active site is named S_1 , while the scissile residue is referred to as P_1 . Upstream of this residue, the residues are named P'_2 - P'_n and downstream to the scissile residue, P_1 - P_n . Adapted from Fersht (1999).

3.1.3 Serine protease zymogens and inhibitors

Most serine proteases occur as inactive zymogens that are activated by proteolysis. Activity of the zymogens is very low due to conformational restraints on the binding site. This is an important regulatory mechanism by which levels of active proteases are kept in check, while active serine proteases are regulated by specific inhibitors of which there are at least 16 different families (Bode and Huber, 1992). The PI-1, PI-2, STI-Kunitz, Bowman-Birk and squash seed inhibitors are all from plants. The most extensively characterized inhibitors are the serpins, the Kazal and BPTI (basic/bovine pancreatic trypsin inhibitor) families. The serpins are large glycoproteins (>40kDa) that interact transiently via an exposed binding loop with the active site, until hydrolysis of this loop and release (Bode and Huber, 1992). Most small inhibitors react with their enzymes via an exposed binding loop (reactive site) with a characteristic canonical conformation. Most of these inhibitors have a compact conformation with a hydrophobic core stabilized by disulphide bonds. While the protein folds of these inhibitors might be conserved, the reactive site residue (P_1) is normally hypervariable, so that replacement with another residue normally leads to a change in inhibition specificity. This is in contrast to other proteins, where the active site residues are normally very conserved (Laskowski and Kato, 1980).

3.1.4 The BPTI /Kunitz family of serine protease inhibitors

The BPTI/Kunitz family of serine protease inhibitors is small (50-65 residues), with 6 cysteines arranged in a characteristic disulphide bond pattern (Laskowski and Kato, 1980) (Fig. 3.4A). The basic structure consists of an N-terminal 3_{10} -helix around the first cysteine, a central double stranded anti-parallel β -sheet linked by a hairpin loop and a C-terminal three turn α -helix (Fig. 3.4B). The binding loop exhibits a characteristic conformation from P_3 to P'_3 and is stabilized by a cysteine at P_2 that is disulphide-connected to the hydrophobic core. The binding-site loop associates with the catalytic residues of the cognate enzyme in a similar manner as the productively bound substrate, with the P_1 carbonyl carbon fixed in contact with the reactive serine. The scissile peptide bond remains intact, with a slight out-of-plane deformation of the carbonyl oxygen. The P_3 - P_3' sites also interact with their cognate enzymes, while secondary contacts can also occur (Bode and Huber, 1992).

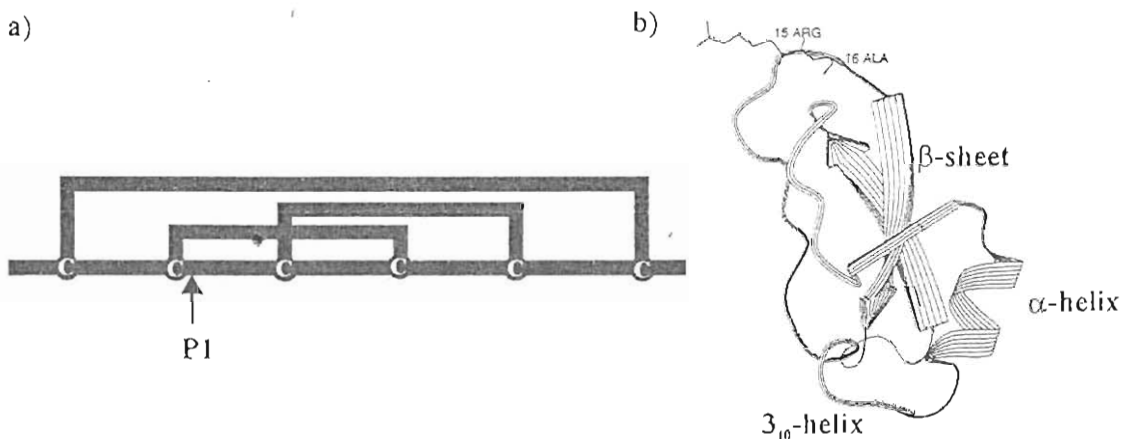


Fig. 3.4: BPTI -Kunitz inhibitor structure. (a) The characteristic disulphide bond pattern of BPTI-Kunitz inhibitors. Cysteines are indicated with dots and the P_1 site with an arrow. (b) The structure of BPTI with its reactive arginine indicated. Note the N-terminal 3_{10} -helix, central β -sheet and C-terminal α -helix. Adapted from Bode and Huber (1992).

3.1.5 Tick-derived serine protease inhibitors of fXa and thrombin

Factor Xa (TAP and fXaI) and thrombin (ornithodorin and savignin) inhibitors from soft ticks are part of the BPTI/Kunitz family. The mechanism of inhibition is distinct from that found for the canonical BPTI inhibitors. Both tick inhibitors insert their N-terminal residues into the active site of their enzymes, in a manner reminiscent to that of hirudin, a

thrombin inhibitor from the leech *Hirudo medicinalis* (van de Locht *et al.* 1996; Wei *et al.* 1998). Hirudin inserts its first three N-terminal amino acids into the active-cleft of thrombin, forming a parallel β -sheet structure with thrombin segment Ser214-Gly219. This is in contrast to the anti-parallel binding of the canonical BPTI inhibitors. The catalytic Ser195 of thrombin is not blocked. The extended carboxy-terminal tail of hirudin (I48-I65) runs along a groove extending from the active-site cleft of thrombin to the positively charged fibrinogen secondary recognition exosite, where it interacts electrostatically (Grütter *et al.* 1990, Rydel *et al.* 1990).

3.1.6 The TAP-fXa complex

TAP is the first inhibitor of fXa purified from soft ticks. It consists of 60 amino acids, with a molecular mass of 6850 Da (Waxman *et al.* 1990). It was recombinantly expressed in yeast and rTAP exhibited all the characteristics of the wild type inhibitor (Neeper *et al.* 1990). TAP has limited homology to the Kunitz-type inhibitors, but determination of its disulphide bond pattern showed that it shared the characteristic disulphide bond pattern of the prototype BPTI-fold (Sardana *et al.* 1991). Determination of the solution NMR structure of TAP also indicated that the β -sheet and α -helical secondary structure elements were similar to that of the BPTI-fold. Due to insertions and deletions in the primary structure of TAP the loops before and after the β -sheets differ extensively in conformation from the prototype BPTI-fold (Antuch *et al.* 1994). It was shown that TAP is a tight-binding competitive inhibitor of fXa that binds to fXa via a two-step mechanism that involves a secondary binding-site (Jordan *et al.* 1990; Jordan *et al.* 1992). Site-directed mutagenesis indicated two areas of the primary structure involved in fXa interaction, the primary recognition site being the first four N-terminal amino acid residues as well as a secondary site between residues 40-54 (Dunwiddie *et al.* 1992). The mechanism of TAP interaction is thus completely different from that of canonical BPTI-inhibitors. Determination of the crystal structure of the TAP-fXa complex (Fig. 3.5) confirmed these biochemical studies and showed that the three N-terminal residues bind inside the P1, P2 and aryl binding pocket of the active-site while Asp47-Tyr49 and Asp55-Ile60 interact with a secondary binding site close to the active-site (Wei *et al.* 1998). To account for the two-step kinetic mechanism observed for TAP, it has been

suggested that an initial slow-binding step occurs at the secondary binding site, which induces a conformation change in the N-terminal residues with concomitant binding into the active site. This could in part explain the slow-tight binding kinetics observed for these inhibitors. Recently, the structure for a TAP-BPTI complex has been determined and it was shown that the conformation of the N-terminal residues probably differ quite extensively for the free and fXa bound TAP, which supports the hypothesis that the N-terminal undergoes a conformational change upon binding (St Charles *et al.* 2000).

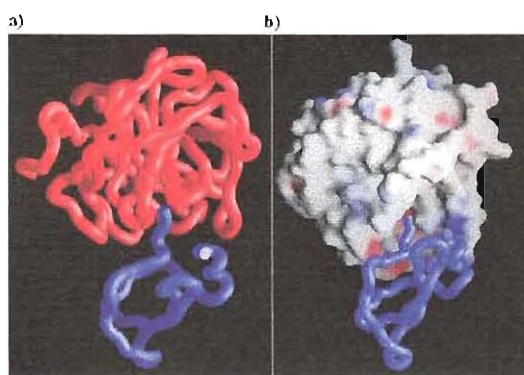


Fig. 3.5: The fXa-TAP complex. (a) Factor Xa is indicated in red and TAP in blue. (b) TAP is indicated in blue and fXa as a surface model. Blue and red surfaces correspond to positive and negative electrostatic potentials, respectively. The coordinates were obtained from the protein databank (PDB: 1K1G).

fXaI is orthologous to TAP and has been purified from *O. savignyi*. It is a slow-tight binding inhibitor ($K_i \sim 0.83$ nM) with a molecular mass of 7183 Da, consisting of 60 amino acids and it shows 46% identity and 87% similarity to TAP (Fig. 3.6) (Gaspar *et al.* 1996; Joubert *et al.* 1998).



Fig. 3.6: Alignment of fXaI and TAP. Shown is the conserved BPTI-fold disulphide bond pattern and similarity according to the PAM250 matrix (boxed in black).

3.1.7 The ornithodorin-thrombin complex

Thrombin is characterized by its high specificity. This is in part due to insertion loops (loop 60 and 149) present around the active site, which restrict access to the active site, so that typical BPTI-like inhibitors cannot bind (Stubbs and Bode, 1993). Thrombin also

contains a basic fibrinogen recognition exosite, which is important for substrate recognition. Ornithodorin, a thrombin inhibitor from the tick *O. moubata* has been co-crystallized with thrombin (van de Locht, 1996). It consists of a N-terminal BPTI-like domain (1'-53') and a C-terminal BPTI-like domain (60'-119') connected by 7 amino acid residues. The N-terminal domain is involved in interaction with the thrombin active site, via its N-terminal residues as well as secondary interaction with the insertion 60 loop. Its C-terminal domain interacts with the basic fibrinogen recognition site via the C-terminal α -helix and possibly the overall negative electrostatic potential of this domain (Fig. 3.7). No mechanism of conformational rearrangement has as yet been proposed for ornithodorin.

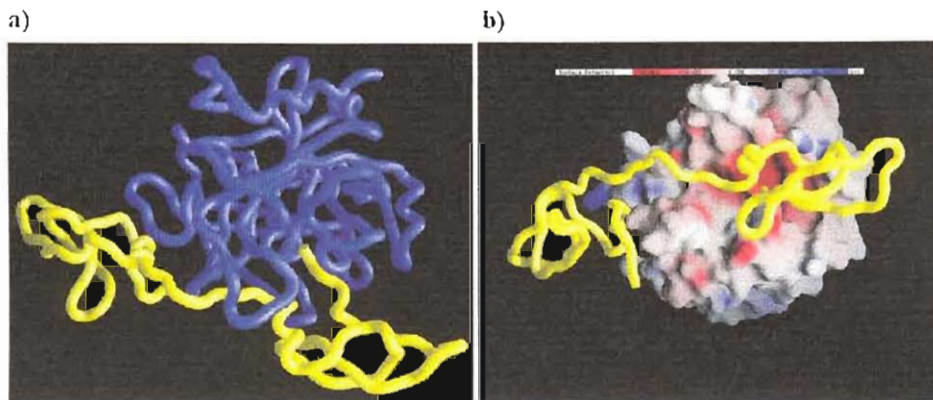


Fig. 3.7: The ornithodorin-thrombin complex. (a) Insertion of the N-terminal residues of ornithodorin (yellow) into the active-site of thrombin (blue) are shown, as well as the proximity of the insertion loops of thrombin to its active-site. (b) Interaction of the C-terminal helix with the basic fibrinogen recognition site is clear. The blue and red surfaces indicate basic and acidic electrostatic potentials, respectively. Coordinates are (PDB:1TOC).

Savignin, an orthologous inhibitor from the tick *O. savignyi* has been kinetically characterized and was shown to be a potent slow-tight binding inhibitor ($K_i \sim 5 \text{ pM}$), although no specific mechanism could be assigned to it. A much lower affinity ($K_i \sim 22 \text{ nM}$) for γ -thrombin, which lacks the fibrinogen recognition exosite, indicated that this site is important for high affinity binding. Changes in the ionic strength had no effect on the K_i of savignin, which indicates that electrostatic interaction is not the main type of interaction (Nienaber, Gaspar and Neitz, 1999). This chapter deals with the further characterization of savignin on the molecular level.

3.2 Materials and methods

3.2.1 Cloning, sequencing and sequence analysis of savignin

The experimental procedures described in Chapter 2 have essentially been followed to clone and sequence savignin. To obtain the coding gene and 3' untranslated region (3' UTR), a degenerate primer (ThrombA: YTN AAY GTI MGI TGY AAY AA) was designed using the first seven amino acids (LNVRCNN) obtained previously (Nienaber, Gaspar and Neitz, 1999). To obtain the 5' UTR and signal peptide sequence a gene specific primer (ThrombC1: CTC GAG TTC CAT TGA AAC GCC ACA) complementary to the coding sequence of the last six amino acids of savignin (CGVSME) was designed. 3'RACE and 5'RACE was performed as described and a 500bp product obtained for 3'RACE and a 450bp product for 5'RACE. These were cloned into the pGEM T-Easy Vector for sequencing.

3.2.2 Molecular modeling of savignin

The deduced protein sequence of savignin, was submitted to the SWISS-MODEL Automated Comparative Protein Modeling Server (Peitsch 1995; Peitsch 1996; Guex and Peitsch, 1997) for modeling. Procheck provided Ramachandran plot parameters (Laskowski *et al.* 1996) and ProFit V1.8 the root mean square deviation (RMSD) (<http://www.biochem.ucl.ac.uk/~martin/swreg.html>) values for carbon α backbone fits of ornithodorin and savignin. WHATIF (Vriend, 1990) was used to validate the model obtained and LIGPLOT (Wallace, Laskowski and Thornton, 1995) to calculate which residues interact between inhibitors (savignin and ornithodorin) and thrombin. The structure of the ornithodorin-thrombin complex (PDB ID: 1TOC) was obtained from the RCSB Protein Databank (Berman *et al.* 2000; <http://www.rcsb.org/pdb/>). All worm figures and surface models were constructed with the Graphical Representation and Analysis of Surface Properties (GRASP) program (Nicholls, Sharp and Honig, 1991). Molecular distances were measured using Rasmol v3.7. The nomenclature used was adapted from van de Locht *et al.* (1996) so that savignin and ornithodorin residues are indicated by primes.

3.3 Results

3.3.1 RACE and sequencing of savignin

3'RACE under optimized conditions with the degenerate primer designed from the N-terminal sequence of savignin gave a product of approximately 500 bp which include the ORF and 3'-UTR. 5' RACE resulted in a product of approximately 450 bp, which include the ORF and 5'-UTR (Fig. 3.8). The cDNA sequences obtained revealed all primers used during PCR and the ORF, 5' and 3' UTR's. The cDNA also contains the stop codon (TAG) and an unusual poly-adenylation signal AATACA.



Fig. 3.8: RACE of savignin. The ORF and 3'UTR of savignin were obtained from single-stranded cDNA with 3'RACE using a 5' N-terminal degenerate primer and 3' poly-T anchor primer (lane 2). The 5' UTR and signal peptide sequence were obtained from double-stranded cDNA using 5'RACE with a 3' gene specific primer and a 5' adapter primer (lane 3). Lane 1 is 100 bp ladder with the 500 bp marker showing twice the intensity of the other markers.

3.3.2 Analysis of the recombinant amino acid sequence of savignin

The translated amino acid sequence gave a protein of 134 amino acids, while the mature chain consisted of 118 amino acids. The first 11 amino acids of the ORF corresponded to that obtained with Edman degradation (Fig. 3.9; Table 1). Analysis of the immature protein using SignalP (Nielsen *et al.* 1997), correctly predicted the presence of the signal peptide (16 amino acids) and the correct cleavage site.

```

aactcactatagggctcgagcggccgccggggcggtgctttaccagccagaagatgctcttttacgtcgtaataact -75
      M L F Y V V I T - 8
-----
ctcgtcgtggaacggtttctggattgaacggttcgatgcaacaaccgcatactgccaactgcgaaaatggtgcaaag -150
L V A G T V S G L N V R C N N P H T A N C E N G A K - 34
-----
cttgagagctattttagggaggggaaacgtgcgtagggtcaccagcatgtcctggagaaggatagccactaaggag -225
L E S Y F R E G E T C V G S P A C P G E G Y A T K E - 60
-----
gactgtcagaaggcctgtttccctggcggggagaccacagcactaatgtcgacagctcatgctttggtaaccgcc -300
D C Q K A C F P G G G D H S T N V D S S C F G Q P P - 86
-----
acttctgcgagactggagcggaggtaacctactacgattctggtagcagaacgtgtaaggactacaacatggctgt -375
T S C E T G A E V T Y Y D S G S R T C K V L Q H G C -112
-----
ccatcgagtgaaaacgcattcgattcagagattgagtgccaagtcgcttggcggttcaatggaatagcagggctgt -450
P S S E N A F D S E I E C Q V A C G V S M E -134
-----
aggaagacacagcgtgaagtcggcatctgaaccgaaccaatctaatacatgacacagaaatacagccctgtagtaaaa -525
-----
aagtcacaaaaaaaaaaaaaaaaagagtggttggtaatgatagc -569
    
```

Fig. 3.9: cDNA sequence and deduced protein sequence of savignin. A cDNA sequence of 569 base pairs was obtained. The 5' adapter, 3' gene specific and 3' anchor primer are shown in bold. The stop codon (TAG), and unusual poly-adenylation signal (AATACA) and the poly-A tail are boxed. The N-terminal sequence previously obtained with N-terminal Edman degradation is underlined while the N-terminal sequence used for degenerate primer design is shown in bold. The signal sequence is underlined with a dashed-line.

3.3.4 Comparison of the recombinant sequence data with data from native savignin

The predicted amino acid composition of recombinant savignin corresponds well with that obtained for the native inhibitor (Fig. 3.10).

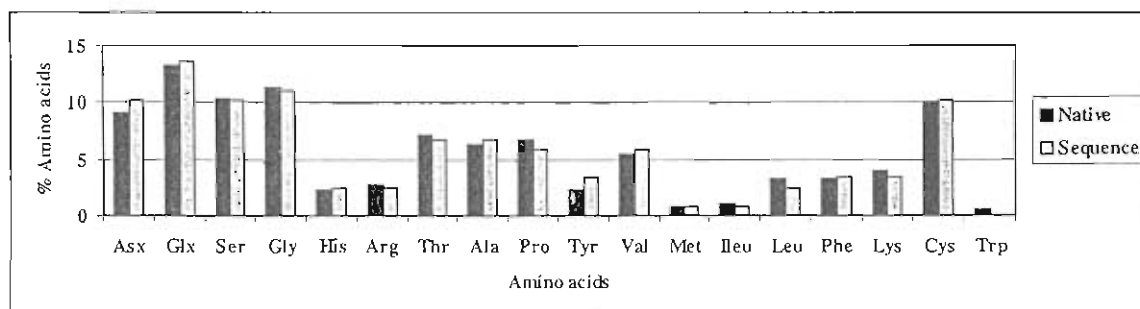


Fig. 3.10: Amino acid composition of native and recombinant savignin. Comparison of the amino acid analysis values obtained for the native protein (adapted from Nienaber, Gaspar and Neitz, 1999) and the deduced amino acid composition obtained for recombinant savignin.

Furthermore, the molecular mass of savignin as determined by electrospray mass spectrometry (Nienaber, Gaspar and Neitz, 1999), correlated with the calculated mass obtained for the recombinant inhibitor (Table 3.1). The number of basically charged residues also corresponded with the number of charged species obtained during ES-MS (results not shown). The predicted iso-electric point is midway between the empirically determined points of the two iso-forms described previously. Taken together these results confirmed the correctness of the cDNA sequence as well as the deduced amino acid sequence of savignin.

Table 3.1: Experimental parameters for native savignin compared with the values calculated from the recombinant sequence. The N-terminal sequence, molecular mass and iso-electric point of native savignin (Nienaber, Gaspar and Neitz, 1999), compared with the deduced amino acid sequence and molecular mass and iso-electric point calculated from the deduced amino acid sequence. Iso-electric points were calculated using compute pI/Mr at the ExPasy server (Bellqvist *et al.* 1993)

	Native protein	Recombinant sequence
N-terminal sequence	LNVRXNNPHTA	LNVRCNNPHTA
Molecular mass	12430.4 Da	12435.4 Da
Iso-electric point	4.2, 5.0	4.55

3.3.4 Sequence alignment of savignin with ornithodorin

The DNA sequence of savignin had 83% identity with the DNA fragment (Genbank accession code: A23191) that codes for the thrombin inhibitor ornithodorin (Fig. 3.11), from the related soft tick, *O. moubata* (van de Locht *et al.* 1996).

```

cttaatgtgcggtgcaacaacccgcatactgccaactgcgaaaatgtgcaaagcttgagagctattttagggag -75
ttgaatgtgttgtccaataacccgcatacggccgattgcaaaaatgatgcacaggttgacagtattttagggag -75

ggggaaacgtgcgtaggggtcaccagcatgtcctggagaaggatacgccactaaggaggactgtcagaaggcctgt -150
ggggaaacgtgcctaatgtcccagcatgcacgagcgaaggatacgccctccagacgaatgtcctcaggcctgc -150

ttccctgccggggagaccacagcactaaatgtcgacagctcatgctttggtcaaccgccacttcctgcgagact -225
ttgttgccggggagaccacagcactgaaatgcacagctcatgctttggtgaccgccacttcctgcgcggaa -225

ggagcggaggtaacctactacgattctgtagcagaacgtgtaaggtactacaa -279
ggcacggacactaacctactacgattctgatagcaaaacatgtaaggtactagca -279
  
```

Fig. 3.11: cDNA sequence alignment of savignin (top rows) with ornithodorin (bottom rows). The cDNA fragment of ornithodorin obtained for the cloning gene (which include the first 275 bp) corresponds with residues 103-379 of savignin (Fig. 3.9). Identity (83%) is boxed in black. Genbank accession codes are savignin (AAL37210) and ornithodorin (A23191).

BLAST analysis of the deduced protein sequence of savignin indicated significant similarity (E-value: $1e^{-30}$) to ornithodorin (Genbank accession code: P56409). Alignment using the Dayhoff PAM250 matrix gave an identity of 63% and similarity of 89% (Fig. 3.12).



			
Savignin:	1'-LNVRCNNPHTANCE NGAKLESYFREG ETCVGSPAC P EGYATKED CC KACFPGGGDHSTN	60'	Domain
Ornithodorin:	1'-LNVLCNNPHTAD EN DAOVDR YFREG ITCLMSPAC T SEGYAS Q HE CC QACFVGGEDHSE	60'	1
			
Savignin:	61'-VDSSCF G OPPTSC ET GA EV TY YD SGSR T CKV L QHG CP SS EN AF DS E IE CVAC CV SM E -	118'	Domain
Ornithodorin:	61'-MHSS CL GD PP TSC A EGTD I TY YD SD SK TCKV L AAS CP SG EN TF ES EV EC VAC CA PI ES -	119'	2

Fig. 3.12: Protein sequence alignment of savignin with ornithodorin. Identity (63%) is boxed in black while similar residues (89%) using the PAM 250 matrix (DENQH, SAT, KR, FY and LIVM) are shaded in gray. The N-terminal BPTI-like domain is from residue 1-53 and the C-terminal BPTI-like domain from residue 61-118. The distinct disulphide bond pattern of the BPTI fold is indicated for each domain. Genbank accession codes are savignin (AAL37210) and ornithodorin (P56409).

3.3.5 Homology modeling of savignin

Superposition of the α -carbon backbone structure of savignin onto that of ornithodorin gave an RMSD value of 0.252 Å for the full-length sequence (Fig. 3.13a). The N- (1'-53') and C-terminal (60'-118') domains gave values of 0.108 Å and 0.103 Å respectively, while the linker region (53'-60') showed the largest deviation (0.177 Å). The modeled surface structure of savignin shows the two separate domains distinctly with the single

chain linker in-between. The C-terminal domain is shown to consist of a predominantly negative electrostatic potential necessary for its proposed association with the basic fibrinogen binding exosite of thrombin (Fig. 3.13b).

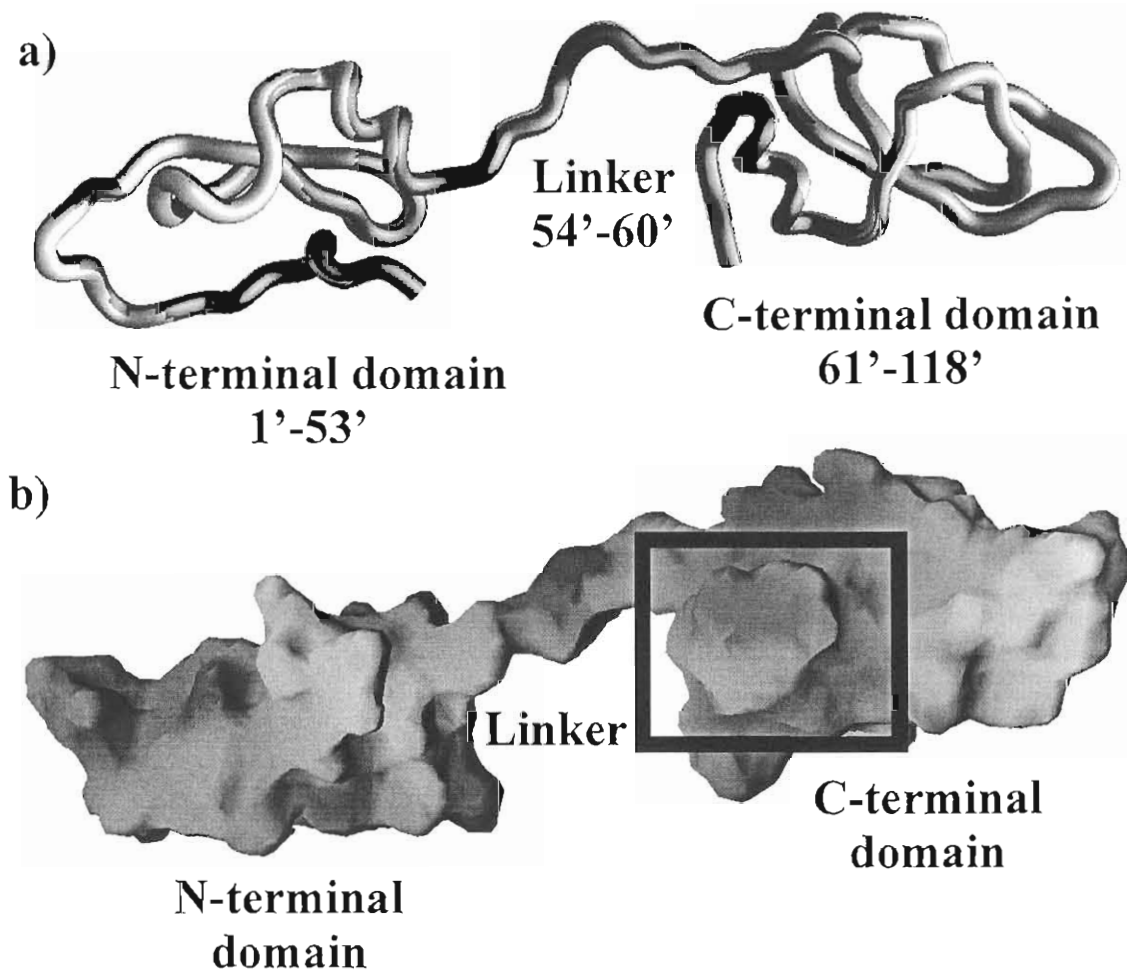


Fig. 3.13: Modeled structure of savignin. (A) Model of the backbone structure of savignin (light gray) superimposed on that of ornithodorin (dark gray). The RMSD value obtained for the two backbone structures is 0.25 Å. (B) A surface model of savignin with the C-terminal surface that interacts with thrombin's basic fibrinogen binding exosite boxed. The darker shadings of gray on the surface indicate acidic surface potential. In the current orientation, no significant basic surface potential is observed.

3.3.6 Interaction of savignin with thrombin

Docking of savignin with thrombin shows that the N-terminal fits inside the active site cleft of thrombin, as is found for ornithodorin (Fig. 3.14a). It also shows the C-terminal domain helix of savignin interacting with the fibrinogen binding exosite of thrombin. A surface model of savignin fitted to thrombin shows insertion of the N-terminal sequence into the active site of thrombin, while interaction with the fibrinogen binding exosite is even more evident (Fig. 3.14b). Prediction of the residues of savignin that interact specifically with thrombin in the modeled structure indicates three main regions (Table 3.2). The first is the N-terminal residues involved in the binding of savignin to the active site cleft of thrombin. The second region is the linker region between the two domains of savignin that is buried inside the structure of thrombin (Fig. 3.14b). The third region is around the C-terminal domain helix that is proposed to bind to the basic fibrinogen-binding site on thrombin. All interactions are mediated via hydrogen bonds or hydrophobic interactions. Residues of savignin involved in interaction with thrombin correlate to those of ornithodorin, suggesting similar mechanisms of action.

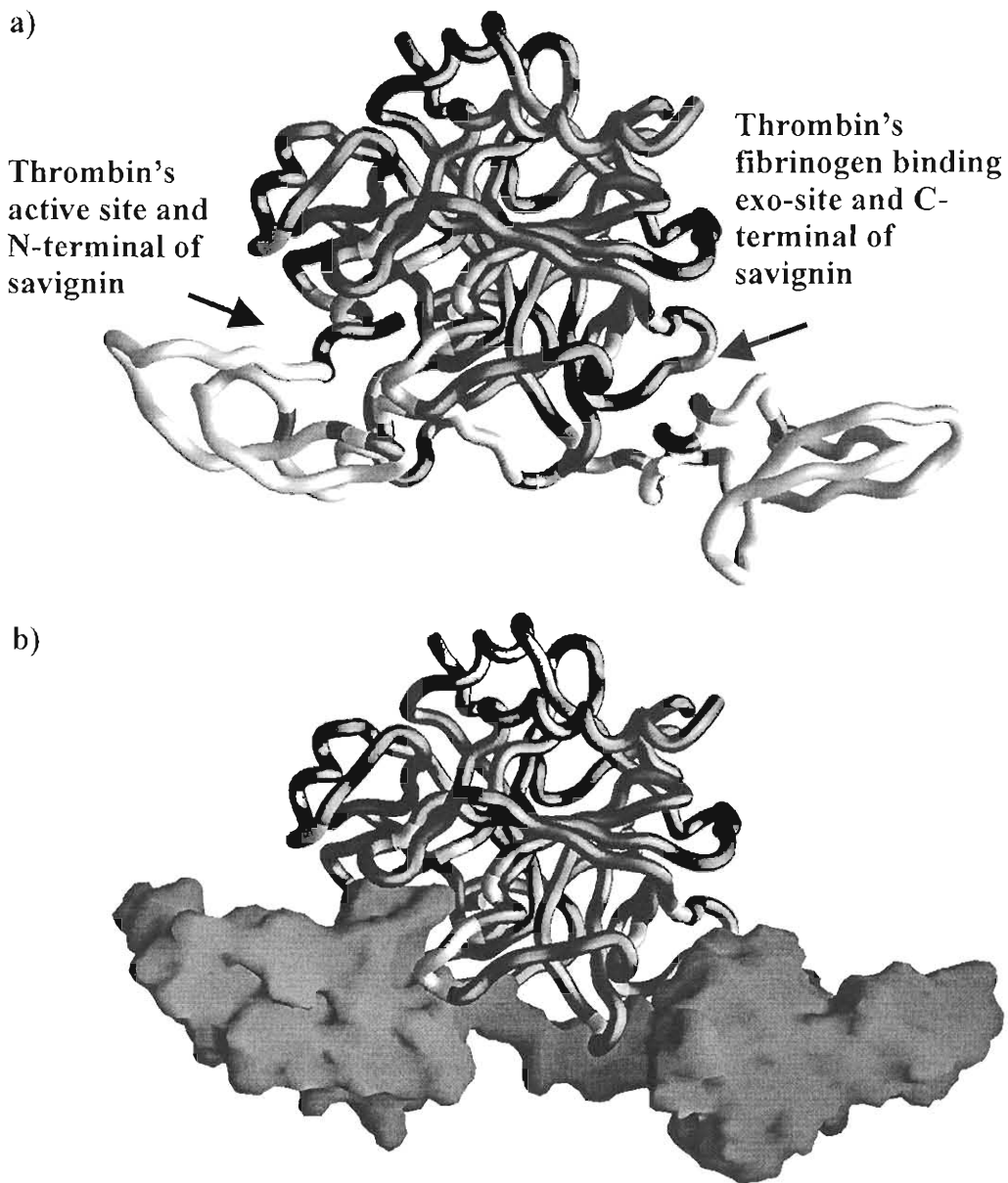


Fig. 3.14: Interaction of savignin with thrombin. (a) Structure of savignin (light gray) fitted into that of thrombin (dark gray). The N-terminal of domain 1 of savignin fits into the active site cleft of thrombin while the C-terminal α -helix of domain 2 of savignin shows close proximity to the fibrinogen-binding exosite of thrombin. Residues that interact between the two structures are indicated as dark and light grey for savignin and thrombin, respectively. (b) Surface model of savignin (gray) shows how the N-terminal residues are accommodated in the active site cleft of thrombin (dark gray), while the Arg4' is excluded from the active site (not shown). The proximity of savignin's C-terminal domain and α -helix region to the basic fibrinogen-binding exosite of thrombin, is evident. Indicated in light gray are the residues of thrombin that interact with savignin.

Table 3.2: Residues of ornithodorin and savignin that interacts with thrombin as predicted by LIGPLOT.

Chain parameters are obtained from the PDB file (1TOC) of ornithodorin and thrombin.

Region of savignin	Ornithodorin	Savignin
N-terminal domain	Leu 1'– Leu 99 Leu 1'– Tyr 60A Leu 1'– Trp 60D Leu 1'– Ser 214 Asn 2'– Gly 219 Val 3'– Gly 216 Val 3'– Glu 217 Val 3'– Trp 215 Leu 4'– Glu 217 Leu 4'– Gly 219 Leu 4'– Arg 221 Cys 5'– Trp 60D Asn 6'– Trp 60D Asn 6'– Tyr 60A	Leu 1'– His 57 Leu 1'– Leu 99 Leu 1'– Tyr 60 A Leu 1'– Trp 60D Leu 1'– Ser 214 Leu 1'– Gly 216 Asn 2 – Gly 219 Val 3'– Ile 174 Val 3'– Gly 216 Val 3'– Glu 217 Val 3'– Trp 215 Arg 4' – Glu 217 Arg 4'– Gly 219 Arg 4' – Arg 221 Cys 5'– Trp 60 D
Linker region	Phe 51'– Glu 192 Asp 56'– Arg 73 Ser 58'– Thr 74 Glu 60'– Thr 74 His 62'– Arg 77A Ser 64'– Arg 77A	Phe 51'– Glu 192 Asp 56'– Arg 73 Ser 58'– Thr 74 Asn 60'– Thr 74
C-terminal domain	Glu 100'– Arg 77 A Thr 102'– Arg 77 A Phe 103'– Arg 77 A Val 107'– Leu 65 Glu 108'– Ile 82 Gln 110'– Gln 38 Val 111'– Gln 38 Val 111'– Leu 65 Val 111'– Ile 82 Ala 112'– Tyr 76 Gly 114'– Gln 38 Ala 115'– Gln 38 Ile 117'– Lys 36 Ile 117'– Leu 65	Glu 100'– Arg 77A Ala 102'– Arg 77A Ile 107'– Leu 65 Ile 107'– Ile 82 Ile 107'– Met 84 Glu 108'– Ile 82 Gln 110'– Gln 38 Val 111'– Gln 38 Val 111'– Leu 65 Val 111'– Ile 82 Gly 114'– Gln 38 Val 115'– Gln 38
Other	Arg 24'– Trp 60 D Glu 25'– Trp 60 D Gly 26'– Pro 60 B Tyr 40'– Glu 146 Gln 48'– Trp 148	Arg 24'– Trp 60 D Gly 26 – Pro 60 B Tyr 40 – Glu 146 Lys 48 – Trp 148

3.3.7 Quality assessment of the modeled structure of savignin

Analysis of the model of savignin using Ramachandran plots at a resolution of 2 Å indicates that 75% of the residues are in regions that are most favored, 20.8% are in additional allowed regions, 1% in generously allowed regions and no residues in disallowed regions (Fig. 3.15 and Table 3.3).

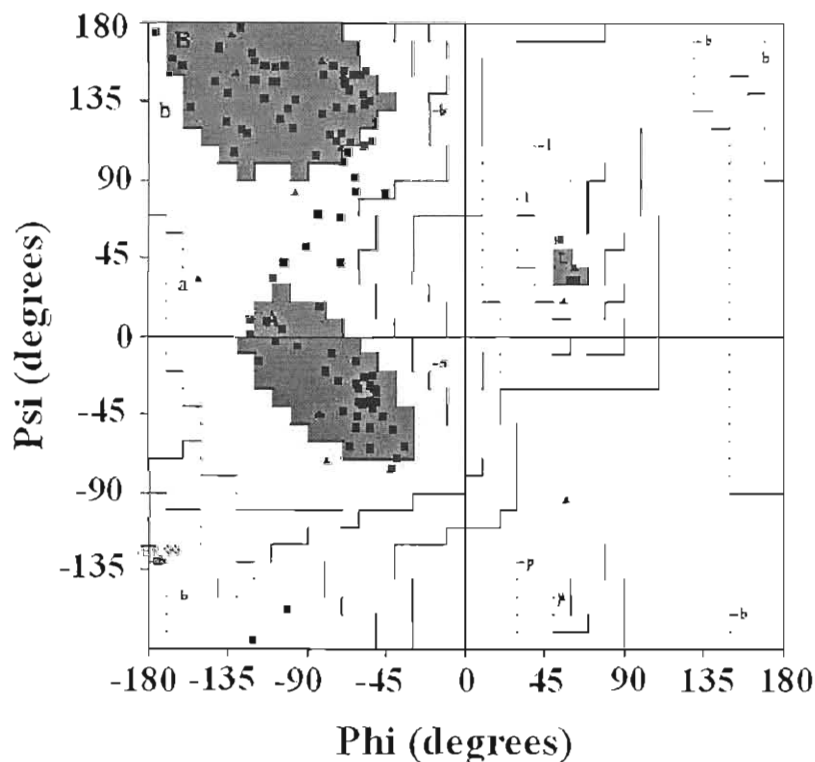


Fig. 3.15: Ramachandran plot of modeled structure of savignin. Position of residues is indicated by squares. Dark gray indicates most favored positions while lighter shades of gray indicate additionally and generally allowed regions respectively.

Compared to the expected values of the main chain parameters (quality assessment, peptide bond planarity, alpha carbon tetrahedral distortion, hydrogen bond energies and overall G-factor) at 2 Å, the obtained values are within the normal distribution or even better (results not shown). It can be concluded that the model proposed for savignin is of high quality.

Table 3.3: Statistics of Ramachandran plots of savignin and ornithodorin. The number and percentage of residues and their localization on the Ramachandran plot is indicated for both ornithodorin and savignin. Values obtained with Procheck.

Characteristics	Ornithodorin	Savignin
Residues in most favoured regions	75 (75%)	75 (78.1%)
Residues in additional allowed regions	25 (25%)	20 (20.8%)
Residues in generously allowed regions	0	1 (1.0%)
Residues in disallowed regions	0	0
Number of non-glycine and non-proline residues	100	96 (100%)
Number of end-residues (exl. Gly and Pro)	3	2
Number of glycine residues	9	13
Number proline residues	6	7
Total	118	118

3.38 Unusual conformation of savignin and thrombin

The number of thermodynamically unfavorable solute-solvent interactions is minimized, by burying hydrophobic residues inside a protein structure. Generally the reduction of a protein's surface that is exposed to solvent is achieved by adoption of a spherical structure and hence the globular nature of most proteins (Jones and Thornton, 1995). In terms of this general observation, the structure of savignin and ornithodorin seems unusual, in that it is not globular but rather extended, with a single amino acid chain being exposed to the solvent in the linker area. To see whether this is truly a deviation from general trends, the relationship between molecular mass and protein volume of various proteins were investigated. It is clear that a direct relationship exists between the molecular mass and volume of globular proteins (Fig. 3.16). A volume of $\sim 28000 \text{ \AA}^3$ is calculated for savignin with its mass as 12430 Da. However, the volume measured (using the Rasmol package) from the structure of ornithodorin and the modeled structure of savignin is $\sim 49000 \text{ \AA}^3$, for which a molecular mass of $\sim 19 \text{ kDa}$ is calculated. This is clearly outside normal deviation. Bikunin, which is also a double BPTI-domain protein, falls well into the expected mass volume relationship (Xu *et al.* 1998).

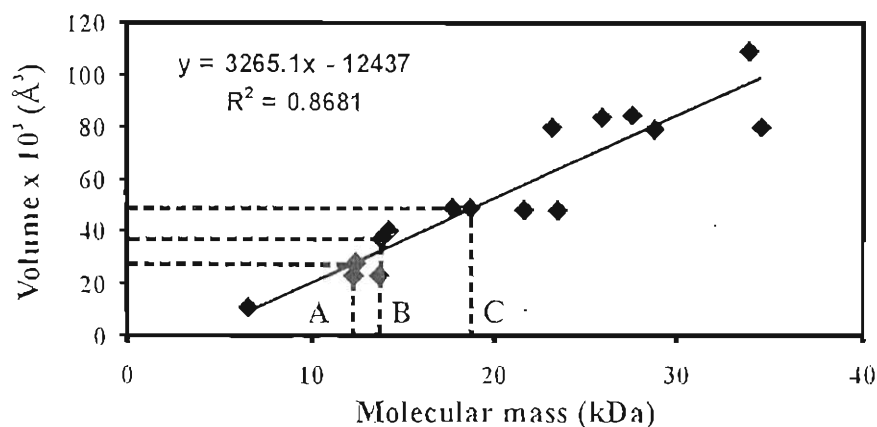


Fig. 3.16: Relationship between molecular mass and volume (\AA^3) of various proteins. Hydrodynamic data of the various proteins were obtained from Creighton (1992). Dashed lines indicate values for (A) savignin's volume calculated from its molecular mass, (B) mass and volume for bikunin and (C) savignin's mass calculated from its measured volume.

Table 3.4: Proteins used for determination of the relationship between molecular mass and volume. Values were obtained from Creighton (1992). Masses and volumes determined from structures and sequence.

Protein	Mr (Da)	Dimensions (\AA)	Volume (\AA^3)
BPTI	6520	29 X 19 X 19	10469
Cytochrome c	12310	25 X 25 X 37	23125
*Savignin	12430	85 X 23 X 25	48875
Ribonuclease A	13690	38 X 28 X 22	23408
*Bikunin	13850	52 X 29 X 24	36192
Lysozyme	14320	45 X 30 X 30	40500
Myoglobin (sperm whale)	17800	44 X 44 X 25	48400
Adenylate kinase	21640	40 X 40 X 30	48000
Bovine trypsin	23200	50 X 40 X 40	80000
Bence Jones REI	23500	40 X 43 X 28	48160
Bovine chymotrypsinogen	23660	50 X 40 X 40	80000
Porcine elastase	25900	55 X 40 X 38	83600
Subtilin	27530	48 X 44 X 40	84480
Carbonic anhydrase	28800	47 X 41 X 41	79007
Superoxide dismutase	33900	72 X 40 X 38	109440
Carboxypeptidase A	34500	50 X 42 X 38	79800

3.4 Discussion

Savignin is the first serine protease inhibitor of thrombin from ticks for which a full-length cDNA has been obtained, which includes 5' and 3' UTR's, a signal peptide sequence and the full-length gene coding for the protein (Fig. 3). Noteworthy is the presence of the less common poly-adenylation site, AATACA (Mason *et al.* 1985) of savignin, which differs from the more commonly found AATAAA sequence (Whale 1992). An important characteristic for assignment of biological significance is the secretion of bio-active components during feeding. The presence of a signal peptide indicates targeting to the endoplasmic reticulum and the salivary gland granules, suggesting that savignin is secreted during feeding (von Heijne 1990). This is supported by the presence of a thrombin inhibitory activity identified in salivary gland secretions (Nienaber, personal communication).

Comparison of the sequences of savignin and ornithodorin and their functions indicate that these proteins are orthologs. Significant is the conserved cysteine pattern characteristic of the Kunitz bovine pancreatic trypsin inhibitor (BPTI) family, which is present in both domains (Laskowski and Kato, 1980).

3.4.1 Interaction of savignin with the thrombin active-site

Savignin is classified as a slow, tight binding competitive inhibitor of thrombin. Kinetic studies indicated that savignin competes with Chromozym TH, a chromogenic substrate for the active site of thrombin (Nienaber, Gaspar and Neitz, 1999). This is confirmed in the modeled structure, where the N-terminal residues (Leu1-Cys5) of savignin bind inside the active site cleft and associates extensively with Ser214-Gly219 of thrombin (Fig. 3.14, Table 3.2) by forming a parallel β -sheet arrangement. Secondary interactions of the N-terminal domain are between the β -hairpin loops of ornithodorin and savignin (Arg24-Gly26) and the thrombin 60loop and residues 40 and 48 of the α -helix with the thrombin 148loop (van de Locht *et al.* 1996).

3.4.2 Interaction of savignin with thrombin's fibrinogen recognition site

Interaction of savignin with the fibrinogen recognition exosite of thrombin is crucial for the potent inhibition observed for α -thrombin ($K_i \sim 5$ pM). γ -thrombin lacks the fibrinogen binding exosite due to excision of Ile68-Arg77 (Stubbs and Bode, 1993). The affinity ($K_i \sim 22.3$ nM) of savignin for γ -thrombin was found to be three orders of magnitude lower than for α -thrombin (Nienaber, Gaspar and Neitz, 1999). Interaction of the C-terminal helix of savignin with the fibrinogen recognition exosite as predicted for the modeled structure would thus seem to be of major importance in the stabilization of savignin-thrombin interactions. Hirudin interacts with the fibrinogen recognition exosite via ionic interactions (Grütter *et al.* 1990), while an increase in ionic strength did not influence the K_i value of savignin (Nienaber, Gaspar and Neitz, 1999). This indicates that ionic interaction between savignin and the fibrinogen recognition exosite of thrombin is not the major type of interaction (Nienaber, Gaspar and Neitz, 1999). The model of savignin interaction with thrombin confirms this, in that the main residues of savignin (Glu100'-Val115'), which includes the C-terminal α -helix of the second BPTI-like domain, interacts with the fibrinogen binding exosite of thrombin (Lys70'-Glu80') with specific H-bond interaction of Glu100' and Asn102' of savignin with Arg77A of thrombin (Table 3.2). Triabin, a thrombin inhibitor from the triatomine bug, also shows hydrophobic interaction with the fibrinogen-binding exosite, rather than ionic interactions (Fuentes-Prior *et al.* 1997).

3.4.3 Unusual conformation of complexed savignin and thrombin

The crystal structure of ornithodorin in complex with thrombin, and the modeled structure of savignin have an unusual conformation. It consists of two globular domains with a flexible linker area that interacts with thrombin via non-bond interactions (Table 3.2). The question is whether it is the normal conformation in which the uncomplexed inhibitors exist, and the conclusion is that it is very unlikely, since a stretch of 9 amino acids (in the linker) face the solvent. A solution to this dilemma would be if the two globular domains fold back upon themselves, with a turn in the linker area. Such a structure has been observed for the uncomplexed form of bikunin (a plasma serine protease inhibitor that also has two BPTI-like domains) (Xu *et al.* 1998).

Association of the globular domains with each other could explain some of the phenomena observed for native savignin. Multiple associations of the globular domains with each other could give rise to conformational isoforms that were observed previously for savignin under both iso-electric focusing conditions as well as non-reducing SDS-PAGE (Nienaber, Gaspar and Neitz, 1999). Under non-reducing conditions savignin also migrated at a much lower mass due to stabilizing disulphide bonds and/or domain association. To resolve this, the structure of the thrombin inhibitors in an uncomplexed form has to be determined.

3.4.4 Kinetic mechanism of thrombin inhibition

Several mechanisms of inhibition by slow binding inhibitors have been proposed (Sculley, Morrison and Cleland, 1996). In mechanism A, interaction is slow due to structural barriers encountered by the inhibitor, while in mechanism B the inhibitor reacts rapidly with the enzyme to form an intermediate, which slowly undergoes a conformational change to form the stable-enzyme-inhibitor complex. Analysis of the omithodorin-thrombin complex indicates no major conformational changes in thrombin's structure upon binding of omithodorin (van de Locht *et al.* 1996). This suggests that any major structure rearrangements would have to take place in the structures of savignin and omithodorin upon binding to thrombin. The time it takes for this conformational change to take place could explain the slow binding kinetics observed for savignin and omithodorin. TAP has been shown to bind to fXa in a two-step fashion, with initial slow-binding to the secondary site and subsequent rearrangement of the N-terminus leading to tight binding into the active site cleft of fXa (Jordan *et al.* 1992; Wei *et al.* 1998).

3.5 Summary

A schematic summary of savignin's proposed mechanism is shown in Fig. 3.17.

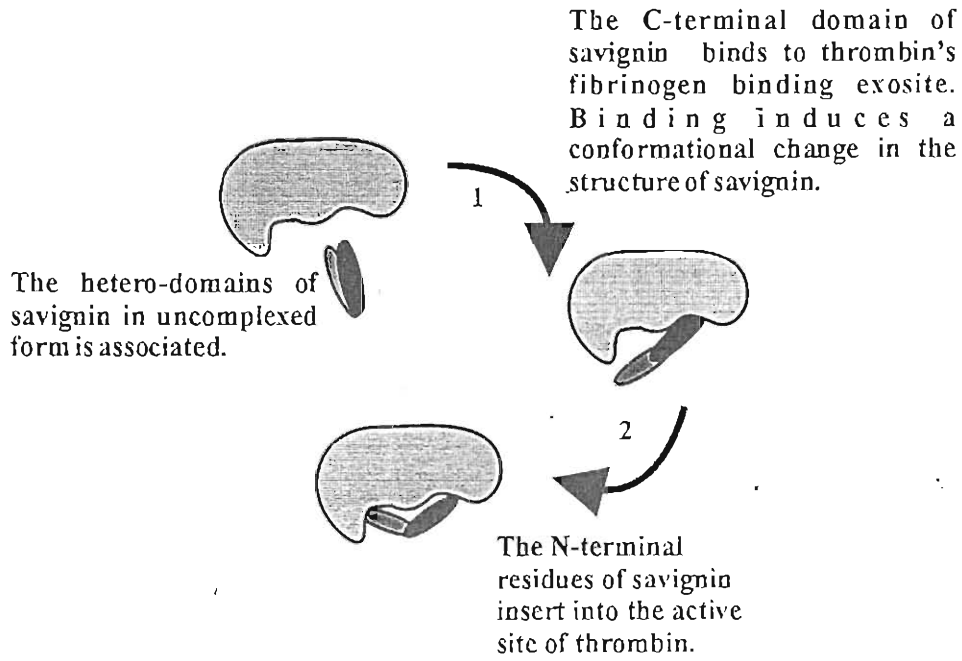


Fig. 3.17: A two-step mechanism for savignin binding to thrombin.