Chapter 4: Evolution of soft tick Kunitz/BPTI anti-hemostatic components*

*Part of the work presented in this chapter has been accepted for publication in the The Journal of Biological Chemistry (Mans, Louw and Neitz, 2002b) and Molecular Biology and Evolution (Mans, Louw and Neitz, 2002c).

4.1.1 The Kunitz/BPTI protein family

The fact that BPTI-proteins are ubiquitous in plants, vertebrates and invertebrates, suggests that they originated at least 500 MYA and diversified through a series of gene duplications, very early in their evolution (Ikeo, Takahashi and Gojobori, 1992). Phylogenetic analysis of these proteins has been problematic, due to evolutionary convergence and gene duplication. Pritchard and Dufton (1999) recently conducted a phylogenetic analysis of 74 Kunitz/BPTI sequences (Fig. 4.1). Rather than grouping into an organismal hierarchy, these inhibitors are grouped into functional classes that indicate their paralogous nature (Dufton, 1985; Ikeo, Takahashi and Gojobori, 1992; Pritchard and Dufton, 1999).

4.1.2 β-bungarotoxins

β-bungarotoxins are the outlier of the whole BPTI family and occur as a fusion protein with phospholipase A_2. This is probably an ancient gene duplication in the ancestor of both mammals and reptiles (Dufton, 1985).

4.1.3 Inter-alpha trypsin inhibitor

Inter-alpha trypsin inhibitors (bikunin) are synthesized as a double Kunitz domain fusion protein with the lipocalin, α1-microglobulin, from which it is cleaved during secretion. Bikunin can inhibit trypsin, elastase and plasmin. It is intracellularly modified by attachment of a chondroitin sulphate chain to Ser_{10} and before secretion is covalently bound to the C-terminal amino acid residue of other larger chains (HC1-3), via a protein-glycosaminoglycan-protein linkage (Åkerström *et al.* 2000). This fusion protein probably plays a role in extracellular matrix binding and stabilization (Bost, Diarra-Mehrpour and Martin, 1998).

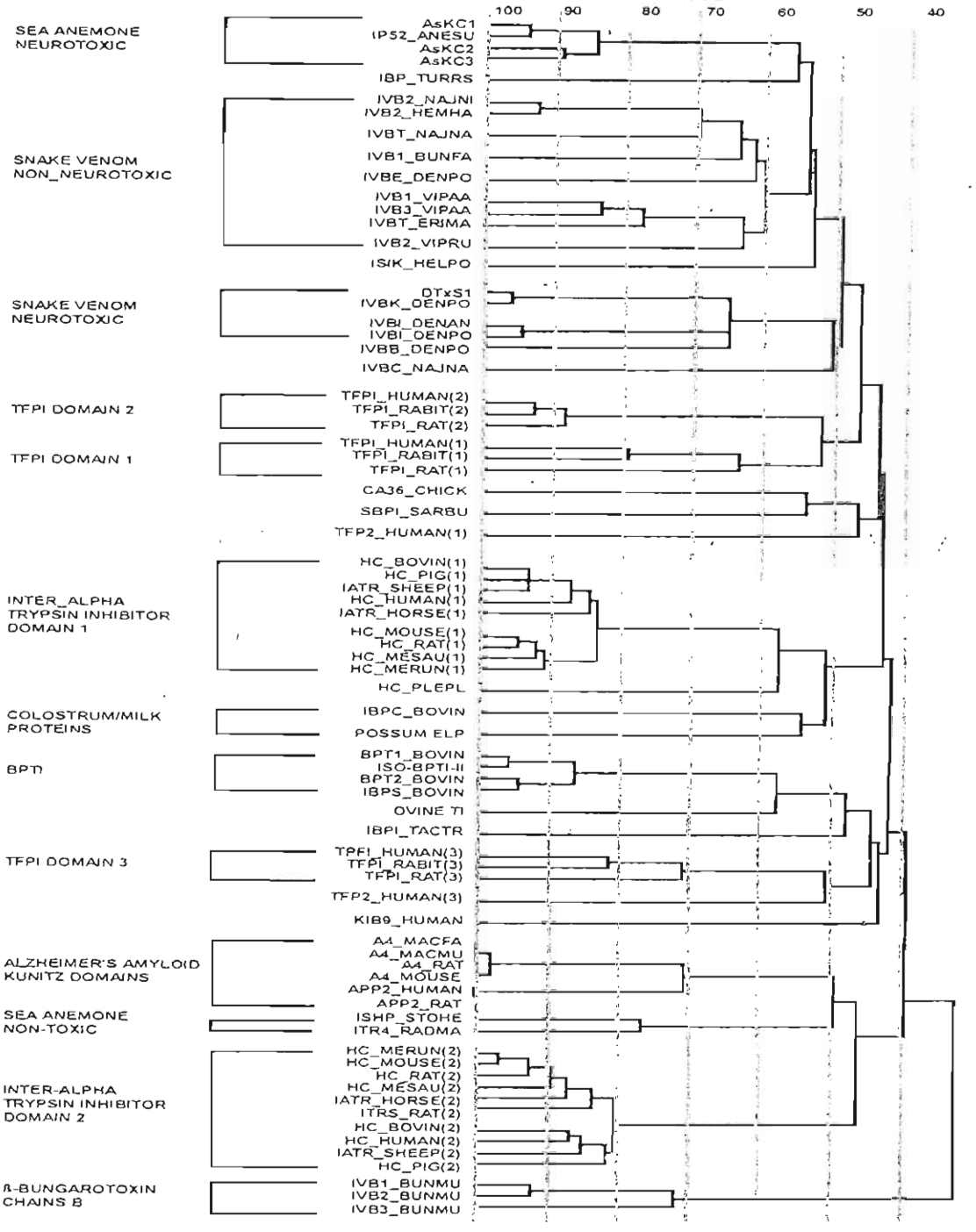


Fig. 4.1: UPGMA dendrogram from 74 Kunitz/BPTI-homologous sequences. The clustering of sequences reflects functionality and class of target rather than species ancestry, with a high degree of paralogy (Pritchard and Dufton, 1999).

4.1.4 Sea anemone Kunitz inhibitors

Kunitz inhibitors from sea anemones show either protease inhibitory activity with a narrow specificity (serine proteases) or a broad protease specificity that include serine, cysteine and aspartic proteases (Minagawa *et al.* 1997; Delfin *et al.* 1996). Some also display neurotoxic as well as trypsin inhibitory activity (Schweitz *et al.* 1995).

4.1.5 Alzheimer's amyloid Kunitz domain

The Alzheimer's amyloid Kunitz domain is part of the β-amyloid precursor protein for which no function has yet been elucidated (Ikeo, Takahashi and Gojobori, 1992).

4.1.6 Tissue factor pathway inhibitor (TFPI)

TFPI exists as three tandemly linked Kunitz domains of which the second domain inhibits fXa and regulates hemostasis by inhibition of the tissue factor-fXa complex (Burgering *et al.* 1997).

4.1.7 BPTI

The BPTI-family functions as basic proteinase inhibitors in serum, although a specific function is not known (Birk, 1987).

4.1.8 Colostrum BPTI inhibitors

Colostrum BPTI inhibitors are found in milk and probably function in allowing intestinal transmission of undegraded macromolecules, such as immunoglobins in young animals (Telemo *et al.* 1987).

4.1.9 Snake venom Kunitz inhibitors

The snake venom Kunitz inhibitors can be divided into those that inhibit proteases and those that show neurotoxicity and it is fairly certain that the gene duplication events giving rise to this family took place within the evolutionary history of the snake family (Dufton, 1985).

4.1.10 Arthropod derived Kunitz inhibitors

Various serine proteases have been identified in insect hemolymph (Sasaki, 1984; Sasaki, 1988; Papayannopoulos and Biemann, 1992). It is thought that these inhibitors function in the immune system of insects, by inhibiting fungal and bacterial serine proteases. Other functions might involve the regulation of endogenous proteases involved in hemolymph coagulation, pro-phenol activation or cytokine activation (Sasaki, 1988; Kanost, 1999).

4.1.11 Tick derived Kunitz inhibitors

Kunitz inhibitors from ticks involved in the regulation of blood coagulation have been described (Chapter 3). Of interest are the recently described double-domain Kunitz inhibitors from larvae of the hard tick *Boophilus microplus* that inhibits trypsin, elastase and kallikrein (Tanaka *et al.* 1999). There is also another BPTI-like sequence from *B. microplus* deposited in the Genbank named carrapatin (P81162), for which a BLAST search indicates very high similarity with the second domain of TFPI. The present chapter investigates the relationship between the platelet aggregation and blood coagulation inhibitors from soft ticks and the functional mechanism of platelet aggregation inhibition.

4.2 Materials and Methods

4.2.1 Protein fold prediction of platelet aggregation inhibitors

Amino acid sequences of disagregin and savignygrin were analyzed using the EMBL's advanced WU-BLAST 2.08 server (BLASTP2) with the non-redundant database (nrdb95) using the default settings (Yuan *et al.* 1998.) Protein fold prediction was performed by submission to the 3D-PSSM Server (Kelley *et al.* 2000). Protein family classification was performed with the Family Pairwise Search v2.0 (Grundy and Bailey, 1999).

4.2.2 Retrieval of BPTI sequences

Tick BPTI-sequences as well as those for the tick platelet aggregation inhibitors were obtained from the literature or generated in our own laboratories (Waxman *et al.* 1990;

Karczewski, Endris and Connolly, 1994; van de Locht *et al.* 1996; Joubert *et al.* 1998; Tanaka *et al.* 1999; Mans *et al.* 2002a). Other Kunitz inhibitor sequences were retrieved from the National Center for Biotechnology Information (NCBI) Genbank database, using the search term Kunitz. A representative subset of 62 sequences was used for multiple sequence alignment (Prichard and Dufton, 1999). Except for some tick-derived BPTI-inhibitors, all sequences were SWISS-PROT database entries. SWISS-PROT entries are followed by the common description and SWISS-PROT accession numbers. For proteins with no SWISS-PROT entry the descriptive names used in the literature are employed while the Genbank accession codes are provided below:

TAP_ORNMO: Tick anticoagulant peptide (P17726), fXaI: fXa inhibitor (AAD09876), ORNT_ORNM1: Ornithodorin domain 1 (P56409), Savignin1: Savignin domain 1 (AAL37210), DISG_ORNMO: Disagregin (g544163), Savignygrin (AF452885), ORNT_ORNM2: ornithodorin domain 2 (P56409), Savignin2: Savignin second domain (AAL37210), A4_RAT: Alzheimer's disease amyloid 4 (P08592), A4_MACFA: Alzheimer's disease amyloid 4 (P53601), A4_SAISC: Alzheimer's disease amyloid 4 (Q95241), A4_MOUSE: Alzheimer's disease amyloid 4 (P12023), APP2_RAT: Amyloid-like protein 2 (P15943), APP2_HUMAN: Amyloid-like protein 2 (Q06481), AMBP_BOV2: bovine alpha-1-microglobulin domain 2 (P00978), IATR_SHEEP2: sheep alpha-1-microglobulin domain 2 (P13371), AMBP_PIG2: pig alpha-1-microglobulin domain 2 (P04366), AMBP_HUMAN2: human alpha-1-microglobulin domain 2 (P02760), AMBP_RAT2: rat alpha-1-microglobulin domain 2 (Q64240), ITR4_RADMA: trypsin inhibitor *Radiantus macrodactylus* (P16344), ISH1_STOHE: serine protease inhibitor from *Stichodactyla helianthus* (P31713), BPTI1_BOVIN: bovine pancreatic trypsin inhibitor (P00974), BPTI2_BOVIN: spleen trypsin inhibitor (P04815), IBPS_BOVIN: serum basic protein (P00975), TFPI_HUMAN3: human tissue factor pathway inhibitor domain 3 (P10646), TFPI_RABIT3: rabbit tissue factor pathway inhibitor domain 3 (P19761), TFPI_RAT3: rat tissue factor pathway inhibitor domain 3 (Q02445), ISC1_BOMMO: silkworm chymotrypsin inhibitor (P10831), ISC2_BOMMO: silkworm chymotrypsin inhibitor (P10832), TIMTC3: silkworm chymotrypsin inhibitor (TIMTC3), SBPI_SARBU: grey flesh fly protease inhibitor (P26228), TIFHBP: flesh fly proteinase inhibitor (TIFHBP), CRPT_BOOMI: carrapatin *Boophilus microplus* (P81162), BMTI-2: *B. microplus* trypsin inhibitor second domain (Tanaka *et al.* 1999), TFPI_HUMAN1: human tissue factor pathway inhibitor domain 3 (P10646), TFPI_RABIT1: rabbit tissue factor pathway inhibitor domain 3 (P19761), TFPI_RAT1: rat tissue factor pathway inhibitor domain 3 (Q02445), AMBP_BOV1: bovine alpha-1-microglobulin domain 1 (P00978), IATR_SHEEP1: sheep alpha-1-microglobulin domain 1 (P13371), AMBP_PIG1: pig alpha-1-microglobulin domain 1 (P04366), AMBP_HUMAN1: human alpha-1-microglobulin domain 1 (P02760), AMBP_RAT1: rat alpha-1-microglobulin domain 1 (Q64240), TIBOC: bovine colostrum inhibitor

(TIBOC), TFPI_HUMAN2: human tissue factor pathway inhibitor domain 3 (P10646), TFPI_RABIT2: rabbit tissue factor pathway inhibitor domain 3 (P19761), TFPI_RAT2: rat tissue factor pathway inhibitor domain 3 (Q02445), AsKC1: kalicludine sea anemone toxin (AAB35413), AsKC2: kalicludine sea anemone toxin (AAB35414), IP52_ANESU: sea anemone protease inhibitor SA5 II (P10280), ISIK_HELPO: roman snail isoinhibitor K (P00994), IVBK_DENPO: dendrotoxin K (P00981), IVBI_DENAN: alpha dendrotoxin (P00980), IVBE_DENPO: dendrotoxin E (P00984), IVB1_VIPAA: venom trypsin inhibitor I (P00991), IVB3_VIPAA: venom basic protease inhibitor 3 (P00992), IVBT_ERIMA: venom trypsin inhibitor (P24541), IVB1_BUNFA: venom basic protease inhibitors IX AND VIIIB (P25660), IVBT_NAJNA: venom trypsin inhibitor (P20229), IVB2_NAJNI: venom basic protease inhibitor II (P00986), IVB2_HEMHA: venom basic protease inhibitor II (P00985), IVB1_BUNMU: beta-1 bungarotoxin chain B (P00987), IVB3_BUNMU: beta-2 bungarotoxin chain B (P00989).

4.2.3 Multiple Sequence Alignment

Sequences were processed to give only a single BPTI core-domain, by truncation of amino acid sequences one amino acid before and after the first and last cysteine of the BPTI-fold, respectively (Pritchard and Dufton, 1999). For BPTI-proteins that exist as multiple-domains, each domain was treated as a single BPTI-fold. Multiple sequence alignment was performed with ClustalX, using the PAM250 matrix and default gap penalty options (Jeanmougin *et al.* 1998). Alignments were manually adjusted based on conserved cysteine positions and secondary structure considerations (Antuch *et al.* 1994).

4.2.4 Neighbor joining Analysis of the BPTI-Family

Phylogenetic analysis on the total BPTI-family was conducted using MEGA version 2.0 (Kumar, Tamura and Nei, 1994). Neighbor joining (NJ) was performed using the number of amino acid differences per site to construct a distance matrix between sequences. Positions that contain gaps were completely deleted so that 39 informative sites were used for analysis. The confidence of the consensus tree obtained was estimated using 10 000 bootstraps. Branches were collapsed below 60% confidence.

4.2.5 Maximum Parsimony Analysis of Tick derived BPTI-Inhibitors

NJ did not completely resolve the relationships within the soft tick BPTI-inhibitor clade. To obtain a more accurate description of the underlying relationships between tick BPTI-

inhibitors, maximum parsimony (MP) using the PHYLIP package, Version 3.2 were performed employing the PROTPARS method using 1000 bootstraps, with or without gapped positions (Felsenstein, 1989). For comparative purposes BPTI-inhibitors from hard ticks as well as insect hemolymph were included, while β-bungarotoxin chain B, proposed to be an outlier of the whole BPTI-family was used as outgroup (Ikeo, Takahashi and Gojobori, 1992; Dufton, 1985).

4.2.6 Phylogeny of Soft Tick Inhibitors based on Protein Structure

Three-dimensional protein structure is generally more conserved in evolution than sequence, so that homologous structures resemble each other closer than more distantly related ones. It has been indicated that structure comparison could indeed resolve phylogenetic relationships (Johnson, Sutcliffe and Blundell, 1990). Construction of a pairwise distance tree based on root mean square deviation (RMSD) of the α-carbon backbone structure could assist in the estimation of distant homologies. As the X-ray diffraction structure of ornithodorin (PDB code: 1TOC; van de Locht *et al.* 1996) and NMR structure of TAP (PDB code: 1TAP: Antuch *et al.* 1994) are known, modeling of their orthologs (savignin and fXaI, respectively) were conducted using the SWISS-MODEL Automated Comparative Protein Modeling Server (Guex *et al.* 1999) and MODELLER (Sali *et al.* 1995). The low percentage identity observed between the platelet aggregation inhibitors (PAI), ornithodorin and TAP complicate their modeling using automated servers. Model structures using the MODELLER package could however, be obtained. The structure of BPTI (PDB code: 1BPI) considered to be the prototype BPTI-fold, was used as outgroup. RMSD values between structure pairs were determined by fitting of the backbone structures using the McLachlan algorithm (McLachlan, 1982), as implemented in the protein least squares fitting program, ProFit V1.8 (http://www.biochem.ucl.ac.uk/~martin/#profit). The phylogenetic tree was constructed by using a pairwise distance matrix of RMSD values and the program NEIGHBOR of the PHYLIP package. The quality of the modeled structures was assessed by construction of Ramachandran plots using Procheck (Laskowski *et al.* 1996). Protein structures were obtained from the RCSB Protein Databank (Berman *et al.* 2000; http://www.rcsb.org/pdb/).

4.2.7 Assay for serine protease inhibitory activity

Serine protease inhibitory activity was assayed as described (Nienaber, Gaspar and Neitz, 1999). Concentrations used were, 0.5 nM fXa, 10 nM plasmin, 50 nM trypsin, and 5 nM thrombin. Enzymes were obtained from Enzyme Research Laboratories, Inc. (South Bend, OR, USA). Enzyme and inhibitor (2.6 μM savignygrin, final concentration) were incubated for 15 min at 37°C before substrate was added and the reaction monitored at 405 nm for 5 minutes. All experiments were performed in triplicate. Thrombin (5 nM) was incubated with savignygrin in 200μl buffer (50mM Tris-HCl, pH 8.3, 227 mM NaCl, 0.1% BSA) before addition of 20μl Chromozym TH (final concentration: 250μM, Roche Molecular Biochemicals). fXa (0.5 nM) was incubated with savignygrin in 200μl buffer (50 mM Tris-HCl, pH 7.4, 0.15M NaCl, 0.1% BSA) before addition of 20μl Chromozym X (final concentration: 1000 μM, Roche Molecular Biochemicals). Bovine trypsin (50nM) was incubated with savignygrin in 200μl buffer (50 mM Tris-HCl, pH 7.4, 20 mM $CaCl_2$) before addition of 20μl BAPNA (N-α-benzoyl-L-arginine-4-nitranilide, final concentration, 500 μM, Roche Molecular Biochemicals). Plasmin (10 nM) diluted in buffer containing PEG6000 and 50mM glycine, pH 2.5, was incubated with savignygrin in 220μl buffer (50 mM Tris-HCl, pH 8.2, 0.1M NaCl) before addition of Chromozym PL (final concentration: 500 μM, Roche Molecular Biochemicals). Chromozym PL was prepared in 100 mM glycine, 0.2% Tween. Enzyme concentrations were determined by active-site titration using 4-NPGB (p-nitrophenyl-p'-guanidinobenzoate in 2.25% dimethyl formamide in acetonitrile) (Chase and Shaw, 1967). 4-NPGB (50 μM) was used in the case of thrombin, trypsin and plasmin and 33.2μM for fXa. Enzymes (100 μl) were added to 400μl veronal buffer (0.1M sodium barbiturate, pH 8.3, 20 mM $CaCl_2$) and incubated for a minute before addition of 5μl of 4-NPGB. Reactions were mixed by inversion before measuring the burst at 410 nm (Hitachi U2000) until a plateau was reached. Enzyme concentrations were calculated from the released p-nitrophenol observed during the burst.

4.3 Results

4.3.1 Protein fold prediction for the platelet aggregation inhibitors

BLASTP2 analysis of disagregin and savignygrin indicated similarity to proteins from the BPTI-family with P(N)-values ranging from 0.011-0.74 for the first fifty hits. The highest scoring protein folds obtained for disagregin and savignygrin with the 3D-PSSM

Server are part of the BPTI-like superfamily (E-values: 0.178-0.855 for the first ten proteins), and contains the functionally diverse proteins BPTI, dendrotoxin, bungarotoxin and knottins. Assignment to a protein family in the structural classification of proteins (SCOP) database (Murzin *et al.* 1995) using the Family Pairwise Search indicated identity to the SCOP BPTI-like superfamily with E-values of 1.6e-15 and 7.97e-16, for disagregin and savignygrin, respectively. The second highest hits gave E-values ranging from 0.09-1.42, indicating the high similarity to the BPTI protein fold relative to other protein folds. This strongly suggested that the platelet aggregation inhibitors exhibit a BPTI-fold. As both fXa and thrombin inhibitors with BPTI-folds have been identified in this tick genus, the hypothesis that all share a common ancestor was advanced.

4.3.2 Multiple alignment of savignygrin with BPTI inhibitors

Alignment with various members of the BPTI family indicated that savignygrin and disagregin exhibit the conserved cysteine pattern characteristic of the BPTI-fold (Fig. 4.2). Significant differences of the soft tick BPTI-inhibitors compared to the rest of the BPTI-family include amino acid insertions that lengthen the loops before the first β-sheet and the C-terminal α-helix. Significant deletions are a single deletion before C14 (BPTI notation) and a deletion of G37 (BPTI notation), a residue conserved throughout the BPTI family. G37 precedes C38 (BPTI notation) that disulphide bonds with C14. This leads to displacement of the disulphide bridges and major distortion of the binding loop conformation (Antuch *et al.* 1994; van de Locht *et al.* 1996). Of interest is the occurrence of the RGD and RED motif in savignygrin and disagregin at the P_1, P'_1, P'_2 positions respectively, normally associated with the substrate binding loop of canonical BPTI-like inhibitors (Laskowski and Kato, 1980). The high and low mass forms of savignygrin have been shown to differ at position R52G that occurs in the α-helix (Chapter 2). Of interest is that no glycine residues occur in the indicated α-helix for the other BPTI-like sequences used in the alignment. Also note the preponderance of serine and threonine at the start of the α-helix, which is in accord with the N-terminal capping preferences of α-helixes (Aurora and Rose, 1998).

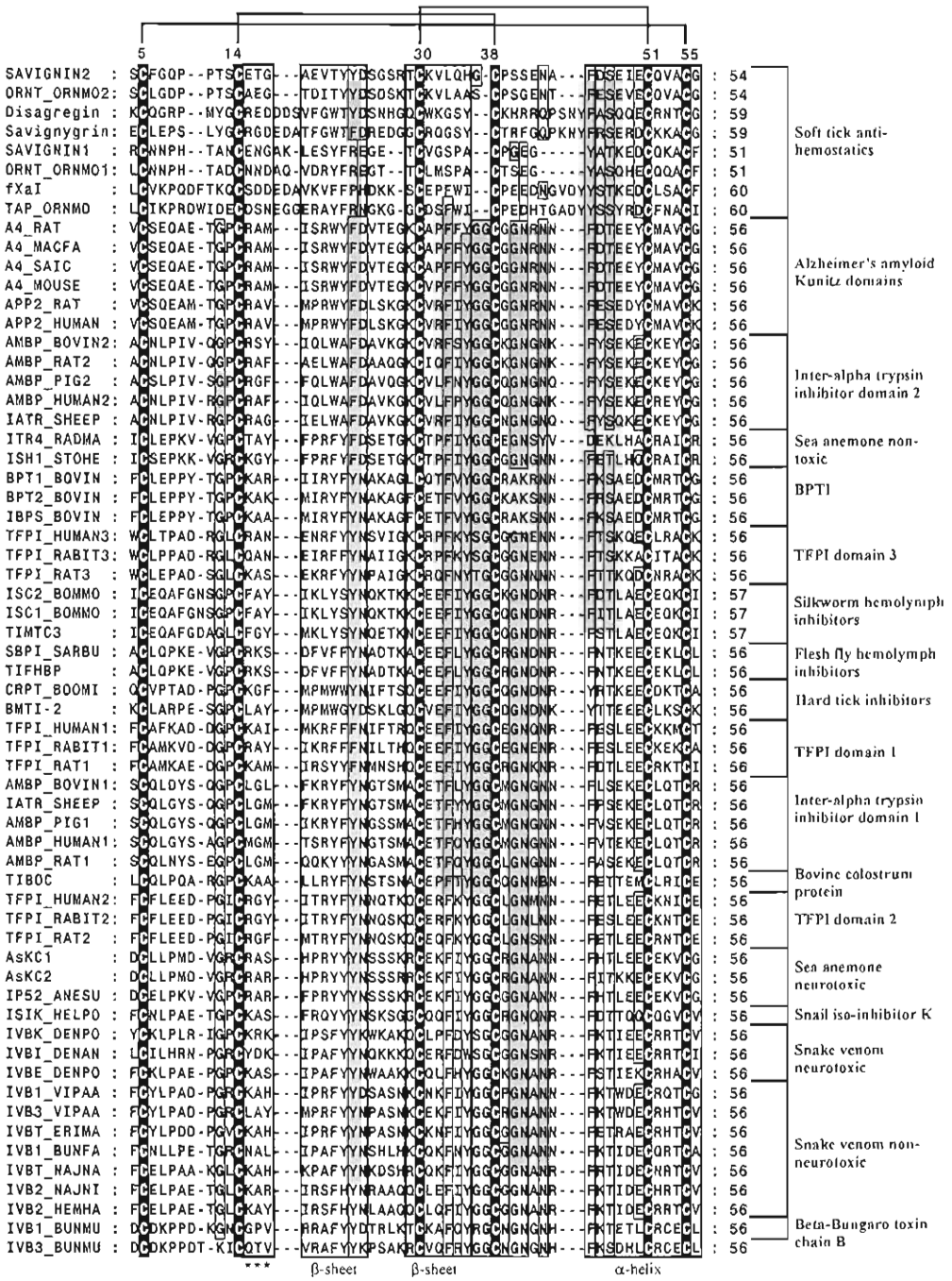


Fig. 4.2. Multiple sequence alignment of BPTI inhibitors with the sequences of disagregin and savignygrin included. Identity (100%) is boxed in black, while 80% similarity using the PAM 250 matrix (DENQH, SAT, KR, FY and LIVM) are boxed in dark gray. The conserved disulphide bond pattern observed for BPTI proteins are indicated by connecting lines. Secondary structure was assigned according to known crystallography structures (Antuch *et al.* 1994). Proteins are grouped according to phylogenetic analysis and the different functional properties are indicated. Sequence names correspond to SWISS-PROT entries. In the case of tick BPTI inhibitors names are indicated as used in the literature. Residue numbering is according to BPTI notation. The *** indicate the P_1, P'_1 and P'_2 (Schechter Berger notation) of the substrate binding loop of canonical BPTI-inhibitors.

4.3.3 Identity and similarity within ortholog groups of the BPTI-family

Comparison of the percent identity/similarity between different ortholog groups indicates that the orthologs of the BPTI-family generally have a constant rate of evolution within ortholog groups with an average percent identity and similarity of 81% and 89%, respectively (Fig. 4.3). The percent identity/similarity of two domains of the thrombin inhibitors (NTI for the N-terminal domains and CTI for the C-terminal domains) is slightly lower than average although this is probably insignificant. In contrast, the percent identity/similarity of the fXa and platelet aggregation inhibitor orthologs fall well below average. While the percent identity and similarity of the family in general is similar, the percent similarity for the tick inhibitors is almost twice that of the identity. These results indicate that soft tick inhibitors show a higher evolutionary rate compared to other BPTI-inhibitors and a higher rate of non-synonymous versus synonymous substitution. This suggests positive Darwinian selection, which could indicate selective pressure on ticks to adapt to a blood-feeding environment (Graur and Li, 2000). It has been shown that the structure of TAP has an increased internal mobility relative to BPTI (Antuch *et al.* 1994). This could indicate a less constrained structure for the fXaI and PAI that might be able to accommodate higher evolution rates. Higher evolutionary rates might also imply a relaxation of structural/functional constraints that is reflected in the fact that both fXaI and PAI are phylogenetically the most divergent of the tick BPTI-like proteins.

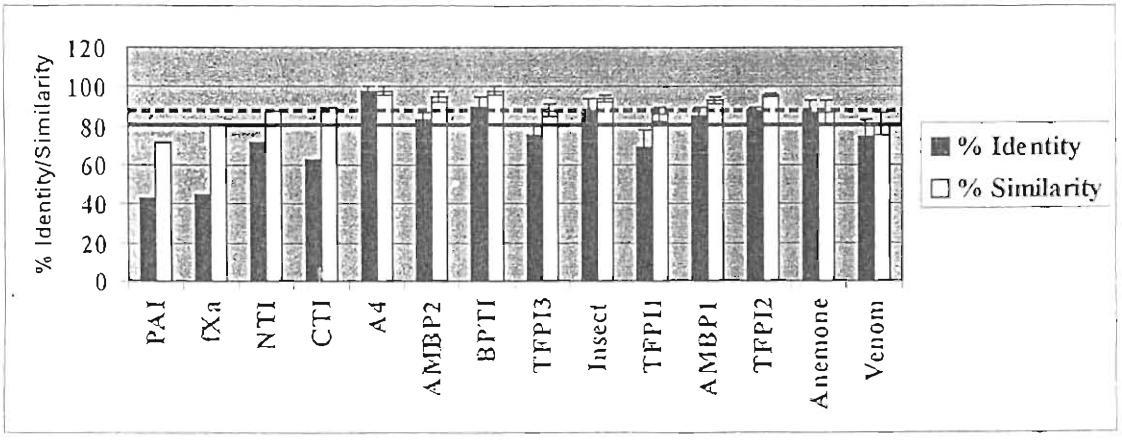


Fig. 4.3. Percent identity and similarity observed between orthologs of the different functional classes from the BPTI-family. Indicated are average values with standard mean deviation for orthologs within a specific functional class. The average values for the whole family are indicated with a solid line for percent identity and dashed line for percent similarity. Percent similarity was obtained using the Dayhoff PAM 250 matrix.

4.3.4 Neighbor joining analysis of the BPTI-family

NJ grouped the BPTI-inhibitors into functional classes previously observed (Dufton, 1985; Ikeo, Takahishi and Gojobori, 1992; Pritchard and Dufton, 1999). These include the β-bungarotoxins which form an outlier group to the whole family, Alzheimer amyloid domains, the two inter-alpha-trypsin inhibitor domains that group into separate clades, the sea anemone BPTI inhibitors that group into non-toxic and neuro-toxic clades, insect hemolymph derived inhibitors, which include inhibitors from the hard tick *B. microplus*, the TFPI domains that group into three separate clades and the snake venom BPTI-like inhibitors that group into neurotoxic and non-neurotoxic clades (Fig. 4.4). It was previously observed that no information on the organismal hierarchy could be obtained from UPGMA constructed trees of the BPTI-like family (Dufton, 1985; Pritchard and Dufton, 1999). This was also observed during the present study. Reasons proposed previously for this problem include the limitation on divergent change by protease inhibitory function and small size that probably led to evolutionary convergence of character states, that does not reflect the total number of changes that have taken place in the past (Dufton, 1985). Numerous gene duplication events further complicate the issue. However, the soft tick blood coagulation and platelet aggregation inhibitors grouped into a monophyletic clade that indicates a common ancestor for these functionally distinct inhibitors and that these paralogous gene duplication events took place within the soft tick family. Disagregin, savignygrin, fXaI, savignin and ornithodorin have all been isolated from salivary gland extracts. The possibility of a common origin for the platelet aggregation and blood coagulation inhibitors due to gene duplication is thus highly probable if it is considered that these inhibitors are all expressed in the tick salivary glands. In contrast, the BPTI-inhibitors from the hard tick *B. microplus* group closer to BPTI inhibitors derived from insect hemolymph that inhibit trypsin and chymotrypsin. BMTI inhibits trypsin, elastase and kallikrein further supporting functional similarity with insect hemolymph derived proteins (Tanaka *et al.* 1999).

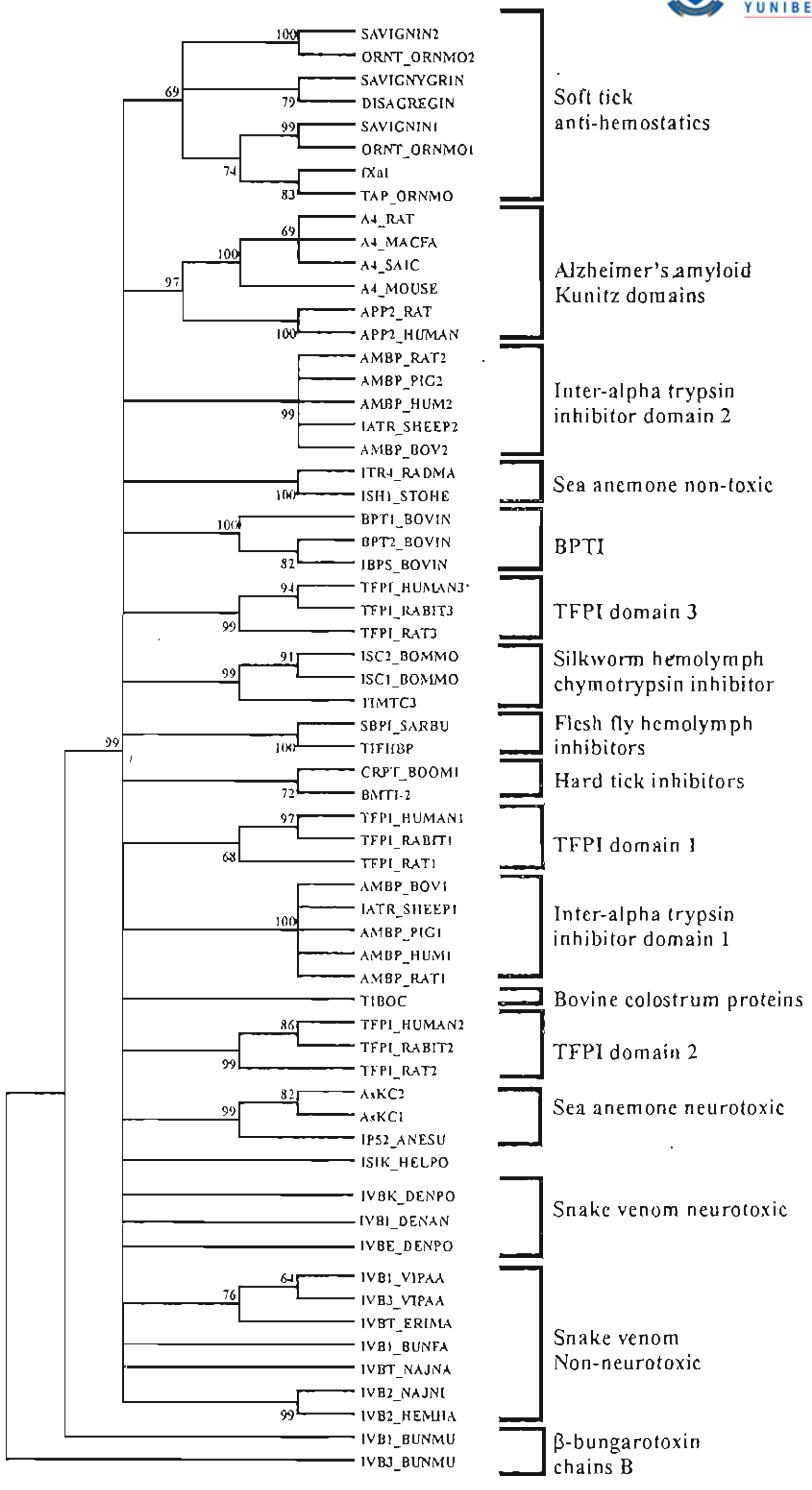


Fig. 4.4: A Neighbor joining dendrogram of 62 BPTI sequences. The tree was constructed based on amino acid differences per site. Indicated is the percent confidence level from 10 000 bootstraps. Branches with confidence levels lower than 60% were collapsed.

4.3.5 Maximum parsimony analysis of the tick-derived BPTI-inhibitors

Maximum parsimony using the alignment where gapped positions were removed, grouped BPTI-inhibitors derived from insect hemolymph and hard ticks into a monophyletic clade, while the soft tick anti-hemostatic inhibitors were grouped into their own monophyletic clade (Fig. 4.5a). The same unresolved relationship between the soft tick-derived inhibitors as for NJ was observed. In this case, fXaI and NTI are grouped together, while CTI is closer to PAI. However, the gapped positions observed in the alignment, especially the insertions before the first β-sheet and C-terminal α-helix could be information rich in terms of structure and functional constraints of these inhibitors. Maximum parsimony analysis using this information gave a more clear relationship between the tick-derived inhibitors (Fig. 4.5b). CTI is basal to this clade, followed by NTI. The fXaI and PAI then group as the terminal clade. This indicates that at least three separate paralogous gene duplication events had occurred, with the evolution of platelet aggregation inhibitory activity being one of the last events.

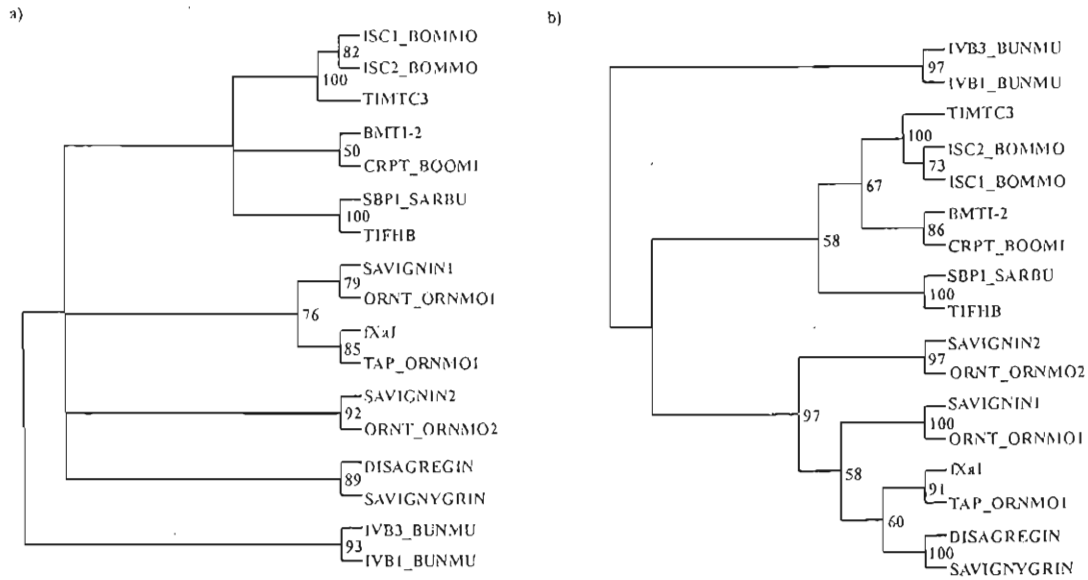


Fig. 4.5: Maximum parsimony analysis of tick BPTI inhibitors. (a) Maximum parsimony analysis of BPTI inhibitors derived from insect hemolymph and hard ticks, as well as soft tick inhibitors. Percent confidence is indicated for 1000 bootstraps. As outgroups the β-bungarotoxins were used. All gapped positions in the alignment used were ignored. (b) Maximum parsimony analysis of the same dataset using the same conditions, with inclusion of the gapped positions. Branches with confidence levels below 50% were collapsed.

4.3.6 Structural comparison of the soft tick-derived BPTI-like inhibitors

To differentiate between the different phylogenies obtained with maximum parsimony, an independent tree based on structural similarities was constructed. The general topology of the structure tree is the same as that obtained with maximum parsimony, using the gapped alignment (Fig. 4.6). The CTI domains show closest structural similarity to BPTI (RMSD: 2.83±0.31Å), followed by the NTI domains (RMSD: 3.4±0.39Å) and lastly the fXa inhibitors (RMSD: 4.3±0.1Å). Because the platelet aggregation inhibitors are paralogs to all three inhibitor folds they were modeled on all three inhibitor folds to test the hypothesis of being closer related to the fXa inhibitors. Models based on the CTI fold (RMSD: 1.8±0.24Å) and NTI fold (RMSD: 3.24±0.52Å) gave higher RMSD values than models obtained based on the fXaI fold (RMSD: 1.13±0.29Å). RMSD values between model pairs for disagregin and savignygrin also indicated lowest RMSD values for the TAP derived models (RMSD: 0.748Å) compared to the models obtained from the thrombin inhibitor folds (RMSD: 2.035 and 1.82 for the N- and C-terminal domain derived models, respectively). These results suggest that the closest structural, as well as ancestral relative to the platelet aggregation inhibitors are the fXa inhibitors.

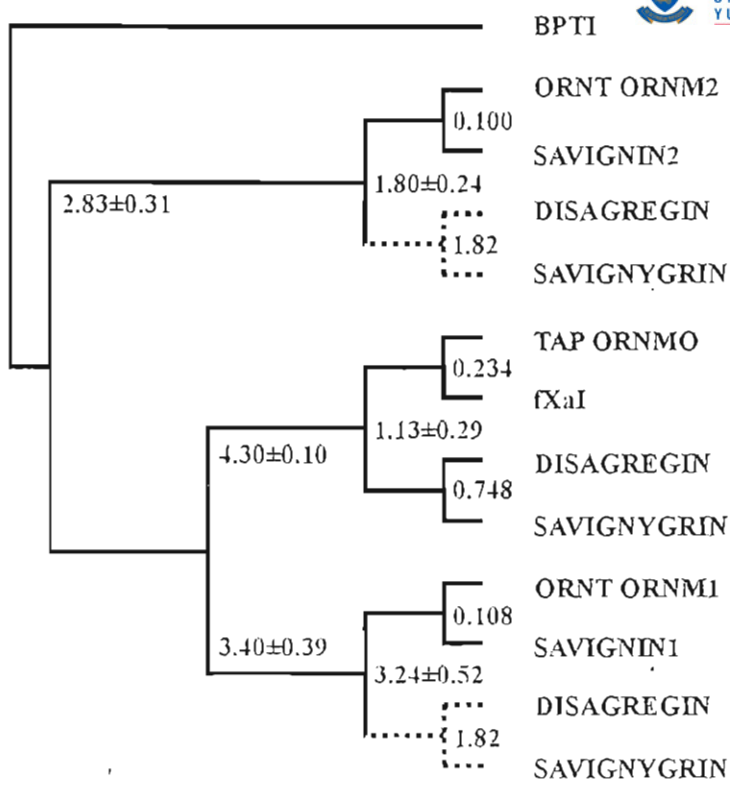


Fig. 4.6: A phylogenetic tree based on pairwise distances of root mean square deviation values (RMSD) obtained from comparison of different structural models. BPTI (PDB code: 1BPI) were used as the prototype BPTI structural outgroup. The structures of TAP (PDB code: 1TAP) and ornithodorin (PDB code: 1TOC) were used to model the structures of fXaI, savignin, disagregin and savignygrin. Values at terminal nodes indicate the pairwise RMSD values between orthologs. Values at internal nodes indicate the average RMSD ± standard deviation for the grouped paralog pairs. Values indicated at horizontal branches designate the respective average RMSD ± standard deviation for the internal branch values in relation to that of BPTI. Branches for platelet aggregation inhibitors shown as dashed lines indicate models with the largest RMSD values and lowest confidence.

4.3.7 Homology modeling of savignygrin

Because the TAP derived model of savignygrin gave the lowest RMSD values, this model was used for structural and functional analysis. Fitting the structures of savignygrin and disagregin onto that of TAP gave RMSD values of 1.0 Å and 1.5 Å respectively (Fig. 4.7). Considering that RMSD values of 1.0-1.5Å are normally obtained for structures of ~40% sequence identity, this model is useful to predict functional parameters (Chothia and Lesk, 1986), especially considering that these proteins show 16-26% identity, in contrast to the assumed ~30% sequence identity needed for building a

structure with molecular modeling (Sali *et al.* 1995). This is probably due to the rather conserved size of these proteins as well as the disulphide bond pattern.

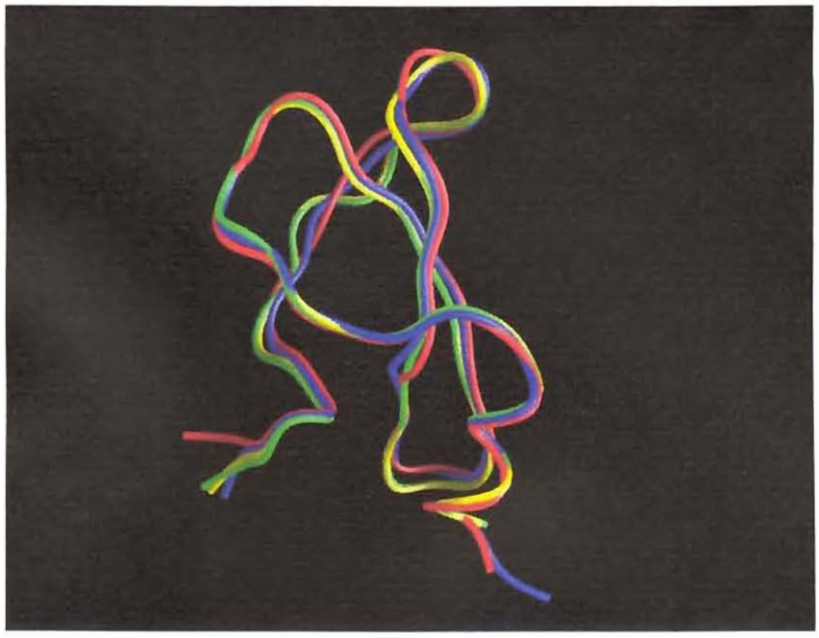


Fig. 4.7: Comparison of the PAI structural models with the fXa inhibitors. Structures of modeled fXai (green), savignygrin (blue) and disagregin (red), superimposed onto the structure of TAP (yellow). RMSD values are 0.2 Å, 1.0 Å and 1.5 Å for fXai, savignygrin and disagregin, respectively.

4.3.8 Analysis of modeled structure

Ramachandran plots showed that 6.2% of the amino acids of savignygrin were in disallowed regions. (Fig.4.8; Table 4.1). Reasons for this high number, is probably due to the low sequence identity/similarity (26/44%) observed between TAP and savignygrin, which complicate modeling procedures. However, at least one disallowed residue (C39) in the structure of TAP is also in the disallowed region for savignygrin (C38). This is probably the reason for the second disallowed residue (C13), which is the corresponding disulphide-bonding partner. The other reason for this distortion in conformation is the presence of an indel (2 residue deletion) in the sequence of savignygrin just before C13, which probably puts a torsion stress on the formation of the loop and disulphide bond leading to the distortion of D16 which also resides on this loop.

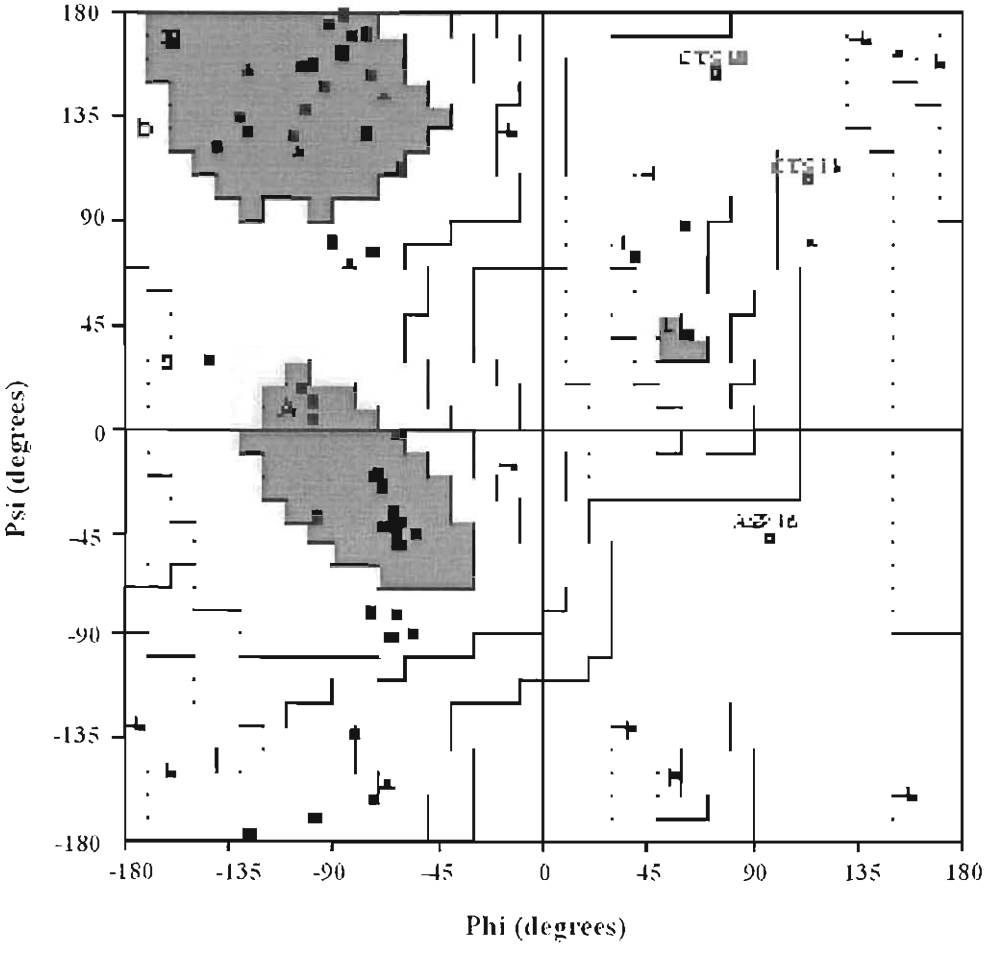


Fig. 4.8: Ramachandran plot of the modeled structure of savignygrin. Position of residues is indicated by squares. Dark gray indicates most favored positions while lighter shades of gray indicate additionally and generally allowed regions respectively.

Table 4.1: Statistics of Ramachandran plots of fXaI and savignygrin. 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	fXaI	Savignygrin
Residues in most favoured regions	23 (60.5%)	30 (62.5%)
Residues in additional allowed regions	14 (36.8%)	15 (31.2%)
Residues in generously allowed regions	0	0
Residues in disallowed regions	1 (2.6%)	3 (6.2%)
Number of non-glycine and non-proline residues	**38**	**48**
Number of end-residues (exl. Gly and Pro)	14	2
Number of glycine residues	6	8
Number of proline residues	2	3
Total	**60**	**61**

4.3.9 Structural implications for savignygrin

The modeled structures indicate that the RGD and RED motifs of savignygrin and disagregin respectively, are located after the second cysteine on the substrate-binding loop associated with canonical Kunitz inhibitors (Fig. 4.9A). The modeled structure also indicates the formation of three disulphide bonds (C5-C58, C13-C38, C32-C54) that corresponds with that of the general Kunitz BPTI-fold. This is important, as disulphide bond partners were not specifically designated during the modeling procedure and as such are completely dependent on the generally allowed stereochemistry of disulphide bridges (Sali and Blundell, 1993). These stereochemical requirements consist of distances between the respective cysteine, α–carbon (4.6-7.4Å) and sulphur (2-3Å) atoms. Dihedral angle (Cβ-S-S-Cβ) constraints are bimodal with peaks at –87.1°C and 93.9°C with a standard deviation of 10°C (Thornton, 1981; Sali and Blundell, 1993). This in itself is a further confirmation of the BPTI-fold of the PAI, but also that the modeled structures are probably very close to the native structure. A surface model of savignygrin, indicates that the RGD motif extends into the surrounding solvent. Of interest though, is the observation that the downstream acidic residues form with the RGD motif a binding epitope (Fig. 4.9B). This suggested that the RGD motif as well as surrounding residues might indeed be involved in the inhibitory activities of these platelet aggregation inhibitors.

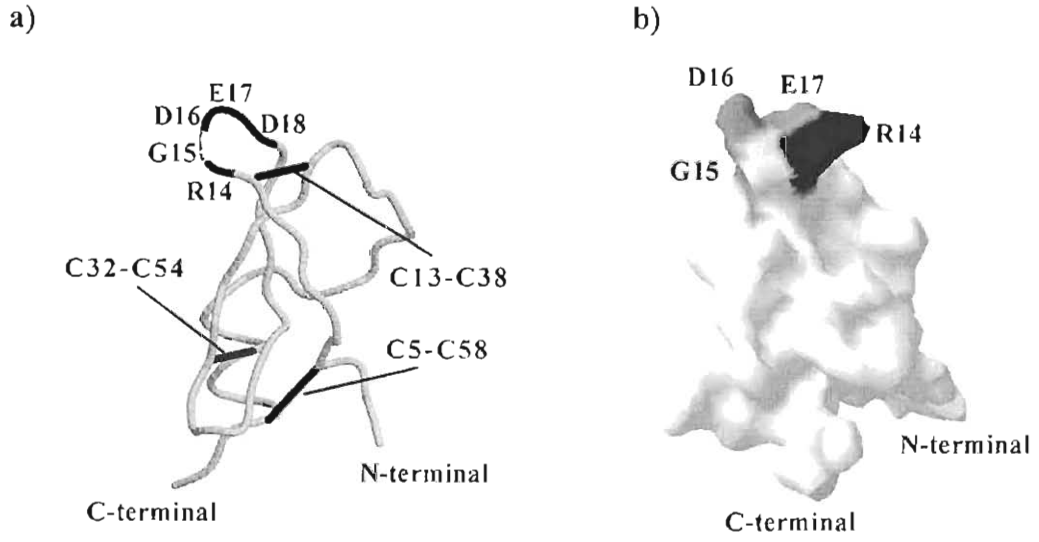


Fig. 4.9: The structure of savignygrin. (a) Intact disulphide bonds (black) predicted by the model and the position of the RGD motif, with R14, G15, D16, E17, D18 indicated. (b) A surface model of savignygrin showing acidic (gray) and basic (black) electrostatic potentials. Indicated is the RGDED binding epitope, projecting upwards from the page.

4.3.10 Serine protease inhibitory activity

The savignygrins have an arginine residue at the P1 position of the canonical BPTI-like inhibitors. As inhibitors with an arginine or lysine at the P1 position inhibit trypsin-like enzymes (Laskowski and Kato, 1980), inhibitory activity against trypsin, thrombin, fXa and plasmin (all recognize arginine at position P1) was investigated. No significant inhibitory activity was observed for any of the proteases tested (Table 4.2).

Table 4.2: Inhibition of serine protease activity. Activity is expressed relative to control values. SD± is indicated for triplicate values. Values in parenthesis indicate the molar ratio of savignygrin to protease.

Serine protease	Activity compared to control
Thrombin (480:1)	114 ± 5%
FXa (436:1)	107 ± 13%
Trypsin (48:1)	95 ± 12%
Plasmin (240:1)	105 ± 11%

4.4 Discussion

4.4.1 A novel BPTI-platelet aggregation inhibitor

Savignygrin is the first platelet aggregation inhibitor described that is part of the Kunitz/BPTI inhibitor family. Presentation of the RGD motif on the canonical substrate-binding loop is a novel way in which this motif is used to antagonize $\alpha_{IIb}\beta_3$. This could open the way for the design of a new class of platelet aggregation inhibitors based on the BPTI-fold. Furthermore, homology with inhibitors of fXa and thrombin in the same tick genus, indicates definite gene duplication events that explains the evolutionary mechanisms of soft tick adaptation to a blood-feeding environment.

4.4.2 Implications of the BPTI-fold for the structure of savignygrin

Assignment of savignygrin to the BPTI family allows the assignment of disulphide bond partners for the previously described disulphide nature of this inhibitor. It also sheds light on some of its properties previously elucidated. Savignygrin was previously shown to be extremely stable in non-reduced form under conditions of high temperature and in the presence of SDS and only unfolded under conditions of high temperature, SDS and urea. The high stability previously observed for the savignygrins are concurrent with a BPTI-like fold, as native BPTI has a melting point of ~95°C (Moses and Hinz, 1983). The high stability of BPTI has been attributed to the stabilizing effect of the disulphide cross-links. BPTI has also been shown to occur as a very disordered polymer in its reduced form, even in the absence of denaturants (Creighton, 1978). This could also be the case for the reduced forms of savignygrin that did not show any activity.

4.4.3 Savignin is a highly conserved protein: implications for structure

It has been shown that a positive Darwinian selection existed for fXaI and PAI. This implied a specific evolutionary pressure applied to the tick during adaptation to a blood-feeding environment. A much lower level of such selection seemed to have operated on the thrombin inhibitors. This would seem strange, as the thrombin inhibitors according to the phylogenetic analysis was a first defense against the host's hemostatic system. However, there could be several reasons for this. The thrombin inhibitors possess a double BPTI domain, which show multiple interactions with thrombin (van de Locht *et*

al. 1996). A higher degree of conservation could be expected due to functional constraints, although this alone could not explain the high degree of conservation (non-selection). Interactions with thrombin are after all still predominantly localized to the N-terminal sequences of the first domain and the C-terminal α-helix of the second domain (van de Locht *et al.* 1996). Structural conservation to maintain the BPTI fold of the two domains could have been a reason for the high degree of conservation, but it has been shown that the structure of the thrombin inhibitors already deviate to some extent from that of the canonical inhibitors (van de Locht *et al.* 1996). It was shown in Chapter 3 that the complexed structure of savignin with that of thrombin is highly unusual and that the two domains might rather associate in their native uncomplexed form as observed for bikunin. This would add another "protein-protein interaction functionality" to this protein, thus placing more selective pressure on it, thereby reducing the rate of divergence.

4.4.4 Soft tick BPTI proteins: a functional paradox

The Kunitz family is a diverse group of proteins that were initially identified as protease inhibitors sharing the common characteristics of a conserved cysteine pattern and canonical inhibition mechanism (Laskowski and Kato, 1980). The soft tick inhibitors are the only serine protease inhibitors of the BPTI-family that do not inhibit their respective enzymes by the canonical mechanism (van de Locht *et al.* 1996; Wei *et al.* 1998). Their similar mechanisms are in fact totally unrelated to the canonical mechanism. The question raised is how could proteins with a very restricted protein fold have evolved a totally different mechanism to perform a similar function (i.e. why did the tick BPTI inhibitors, switch mechanisms)?

4.4.5 Evolution of tick BPTI-proteins: a paradox resolved

It has been suggested that the only way new protein functions could evolve from duplicated genes, is by way of gene sharing for a single domain protein or as an existing bi-functional multi-domain protein, where duplication of each domain leads to acquisition of individual functions (Hughes, 1994). In the light of this a probable evolutionary scenario can be proposed based on considerations of thrombin's structure and function (Fig. 4.10).

(A) Thrombin exerts conformational restrictions on BPTI inhibitors due to the insertion loops (loop 60 and 149) present around its active site that prevents inhibition by the canonical mechanism (Stubbs and Bode, 1993). This probably influenced the ancestral CTI domain, to evolve a new functional mechanism: inhibition of fibrinogen binding to thrombin via its basic fibrinogen binding exo-site (Fig. 4.10). This is reminiscent of the inhibition mechanism of triabin, a lipocalin that inhibits thrombin via its fibrinogen binding exo-site (Fuentes-Prior *et al.* 1997). This newly acquired functional mechanism probably relaxed the functional restrictions on the substrate-binding loop so that CTI lost its original protease inhibitory activity exerted via the canonical mechanism through the acquisition of mutations and indels. Targeting of the fibrinogen-binding exosite of thrombin would have allowed the tick to inhibit both clotting as well as platelet aggregation induced by thrombin (Stubbs and Bode, 1995). It is arguable whether this mechanism of inhibition would have been efficient on its own, as it would not have inhibited catalytic activity completely.

(B) A tandem gene duplication and N-terminal fusion event led to the formation of a homodimeric BPTI protein. Conformational restrictions that the CTI domain placed on the NTI domain probably led to evolution of a new mechanism of thrombin inhibition: insertion of its N-terminal residues into the active site of thrombin. This double-mechanism allows a much more specific mode of inhibition, where both enzyme active site as well as additional substrate binding sites are targeted.

(C) It is highly probable that the next gene duplication occurred from the N-terminal domain and led to an inhibitor that retained and modified the existing mechanism of NTI to target fXa. While this would seem to be a redundant strategy, the inhibition of fXa ensures that even lower concentrations of thrombin would be produced. The secondary interaction sites observed for both NTI as well as fXaI is

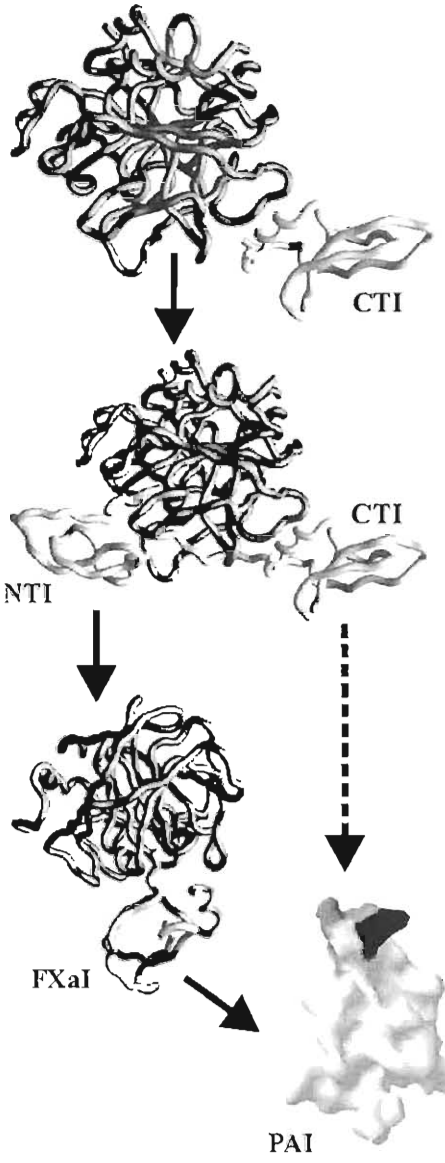
probably gene -sharing remnants of CTI's interaction with thrombin's fibrinogen binding exosite.

(D) Subsequent duplication of PAI from the fXaI domain led to a utilization of the now defunct and probably highly distorted, but still present substrate binding-loop of the canonical BPTI-inhibitors, to evolve a new specificity for the platelet aggregation receptor $\alpha_{IIb}\beta_3$. An alternative scenario supported by Neighbor joining and maximum parsimony (where gapped positions were ignored), suggest that the PAI might have duplicated directly from the CTI domain, probably after the tandem fusion event. In both scenarios the fibrinogen-mimicking site of CTI probably played an important role in the initial targeting to the platelet fibrinogen receptor. This could explain observations that disagregin targets more than one site on $\alpha_{IIb}\beta_3$ (Karczewski *et al.* 1997) and suggests that the C-terminal α-helix of the PAI might also be involved in integrin interactions. The fact that no inhibition of serine proteases are observed for savignygrin fits with a scenario, where PAI evolved after NTI, CTI and fXaI lost the canonical mechanism of serine protease inhibition.

4.4.6 The BPTI-Fold as Evolutionary Unit

It is not so surprising that soft ticks have used the BPTI-like fold to evolve new protein functions. The BPTI fold *per se* is not novel to serine protease inhibitors. A rather large group of these proteins are found in snake venoms, where they can act as toxins by blocking ion channels of the cardiac and nervous systems (Prichard and Dufton, 1999). Furthermore, the substrate-presenting loop of the BPTI-like fold is the ideal place for the presentation of recognition motifs, as that found for RGD-containing proteins. What is perhaps surprising is the absence of a RGD motif in the snake venom BPTI-like proteins, where the disintegrin protein family exploits this motif (Huang, 1998). This might imply that in the case of snake venoms, the disintegrins were already present when the BPTI-like proteins acquired their respective functions. In ticks, such a disintegrin-like protein might have been absent, forcing the tick to rely on protein families already present in its repertoire, to generate functional diversity. The loss of a restricted canonical substrate-

binding loop conformation probably further allowed utilization of this loop for a novel presentation mechanism, while the utilization of the fibrinogen-mimicking binding site of CTI could have been important for targeting to the relevant binding site.



Due to conformational restriction loops around thrombin's active site, canonical BPTI-inhibitors cannot inhibit thrombin. CTI evolved to target the fibrinogen binding exo-site. The canonical BPTI-mechanism is lost due to mutation and indels.

A tandem gene duplication and fusion event to the N-terminal of CTI leads to a homo-dimeric inhibitor. Due to dimeric conformational restrictions a new mechanism of serine protease inhibition evolves: insertion of the N-terminal residues into thrombin's active site, thereby avoiding the restrictive loops.

A gene duplication of NTI leads to utilization of the novel protease inhibitory mechanism, to evolve a new fXa inhibitory capability.

Gene duplication of fXaI leads to utilization of the still existing but redundant BPTI-substrate binding loop to evolve new specificity for the fibrinogen platelet receptor, $\alpha_{IIb}\beta_3$. The alternative evolutionary pathway suggested by Neighbor-Joining and maximum parsimony when gapped positions are ignored, suggest that the platelet aggregation inhibitors might have duplicated directly from CTI.

Fig. 4.10: Evolutionary mechanisms for the acquisition of new anti-hemostatic functions in soft ticks. The solid and broken lines indicate alternative pathways, supported by the different phylogeny approaches. NTI indicates the N-terminal domain of the thrombin inhibitors, while CTI indicates the C-terminal domain of these inhibitors. PAI indicates the platelet aggregation inhibitors and fXaI the fXa inhibitors.

4.4.7 Independent evolution of anti-hemostatic inhibitors in hard and soft ticks

It is interesting that BPTI inhibitors from the hard tick, *B. microplus* is not grouped with the anti-hemostatic factors of soft ticks, but rather with hemolymph derived protease inhibitors, that inhibit their respective enzymes via the classical BPTI mechanism. fXa and thrombin inhibitors from hard ticks have also been described with molecular masses (17-65 kDa) that differ significantly from that of the BPTI fold (Bowman *et al.* 1997). Furthermore, variabilin, a platelet aggregation inhibitor from the hard tick *Dermacentor variabilis* does not resemble the platelet aggregation inhibitors from soft ticks at all (Wang *et al.* 1996). Its cysteine pattern as well as the localization of its RGD-motif differs completely from the observed BPTI-fold and the motif found in savignygrin. This would suggest that soft tick derived BPTI inhibitors only acquired their specific mechanisms of action after the divergence of hard and soft ticks, as well as suggesting independent adaptation to a blood-feeding environment. This is of interest because it would have been expected that ticks, being monophyletic, would have adapted to a blood-feeding environment before divergence. It also raises the question whether anti-hemostatic functions observed in the *Ornithodoros* genus are represented in other soft tick families? However, other signs of probable independent adaptation to a blood-feeding environment include different salivary gland morphologies, feeding behavior and reproductive strategies, that is all intimately linked with blood-feeding (Sonenshine, 1991). However, while this study suggests independent acquisition of novel anti-hemostatic components by the two main tick families, the presence of apyrase in both families have been indicated (Law, Ribeiro and Wells, 1992). Apyrase, (ATP-diphosphohydrolase; EC 3.6.1.5) inhibits platelet aggregation induced by ADP as well as being able to disaggregate platelets aggregated by ADP (Mans *et al.* 1998b; Mans *et al.* 2000). It has been identified in all of the hematophagous arthropod families investigated so far (Ribeiro, 1995), which suggests that this enzyme might have been present in the ancestral non-hematophagous tick. Of interest is the absence of apyrase in the saliva of the tick, *Amblyomma americanum* (Bowman *et al.* 1997).

4.4.8 The driving force behind tick divergence

Independent adaptation to a blood-feeding environment indicates a rapid divergence soon after the origin of the Ixodida (120 MYA), with ticks being adapted to a blood-feeding life by 92 MYA. The question that is raised is what could have triggered such a rapid diversification into different tick families? There is a current controversy around the divergence of early birds and placental mammals, where the fossil records argue for divergence around the Cretaceous-Tertiary (K/T) boundry (~65 MYA), while molecular phylogeny evidence suggest a much earlier divergence in the Early Cretaceous (120-80 MYA) (Benton, 1999; Easteal, 1999; Madsen *et al.* 2001; Murphy *et al.* 2001). A rapid divergence and independent acquisition of hematophagous mechanisms in ticks from ~120-90 MYA fits with the molecular phylogeny hypothesis. Radiation of birds as well as placental mammals would have provided ample opportunity for ticks to find novel niches in which they could excel as blood-feeding arthropods. It could thus be argued that the emergence of hematophagy in ticks was triggered by the divergence of early modern birds and mammals. This provides an interesting counterpoint to suggestions that evolution of ticks was not as much influenced by host specificity as by ecological factors (Klompen *et al.* 1996). While this might be the case for adaptation to the environment, independent evolution of anti-hemostatic strategies would suggest that host diversity could have influenced the adaptation of ticks to the vertebrate hemostatic system. Positive Darwinian selection would indeed suggest that the hemostatic system of the host played a decisive role in the evolution of hematophagy in ticks.

4.4.9 Implications for pharmacological and vaccine development

An intriguing possibility that emerges from this study is the possibility to design a chimeric protein with fXa, thrombin and platelet aggregation inhibitory capacities. Such a protein might be useful as a multi-functional agent to control thrombosis in a regulated manner. Such a multi-epitope protein could also be used as a possible vaccine agent, to generate immune responses that could knock out more than one function necessary for tick feeding.