Chapter 4

Isolation and characterisation of a serine protease from *O. s. scutellatus*

4.1 INTRODUCTION

The abundance and varying antigenicity of the lower molecular weight proteins reported in Chapter 3 suggested these proteins may provide an interesting area of research. In addition to the numerous small proteins, *O. s. scutellatus* proteins form large complexes affecting coagulopathy (Walker, *et al,.* 1980; Speijer, *et al.*, 1986; Lalloo, *et al.*, 1995b; Kini, *et al.*, 2001). Differing isomers of these proteins and peptides may hold the key that may help to describe clinical variations encountered after differing *Oxyuranus* species envenomation. Due to the varying coagulopathy's encountered between *Oxyuranus* envenomations, the precedent of large, complex enzymes effecting these coagulopathy's and the ability to identify specific proteins within the venom, selected large proteins were isolated and characterised.

4.1.1 Coagulant Enzymes

Haemostasis is a balance of two opposing forces: clot formation and dissolution. Correct clotting is essential at the site of a wound in order to maintain haemostasis, however clotting away from the site of a wound must be minimised in order to prevent life-threatening thrombotic events. Blood clotting, which occurs through the production of fibrin resulting from the activation of prothrombin to thrombin, involves a cascade of biochemical events that can be initiated intrinsically or extrinsically. The intrinsic pathway requires the clotting factors VIII, IX, X, XI, and XII, prekallikrein, calcium ions and high molecular-weight kininogen. Each of these constituents leads to the conversion of factor X (inactive) to factor Xa (where 'a' signifies active). Factor V, calcium ions and phospholipids secreted from platelets are required for the conversion of prothrombin to thrombin which in turn leads to the production of fibrin (Figure 4.1). The extrinsic pathway is initiated at the site of injury (eg snake bite) in response to the release of tissue factor (factor III), and occurs at factor Xa (the point where the intrinsic

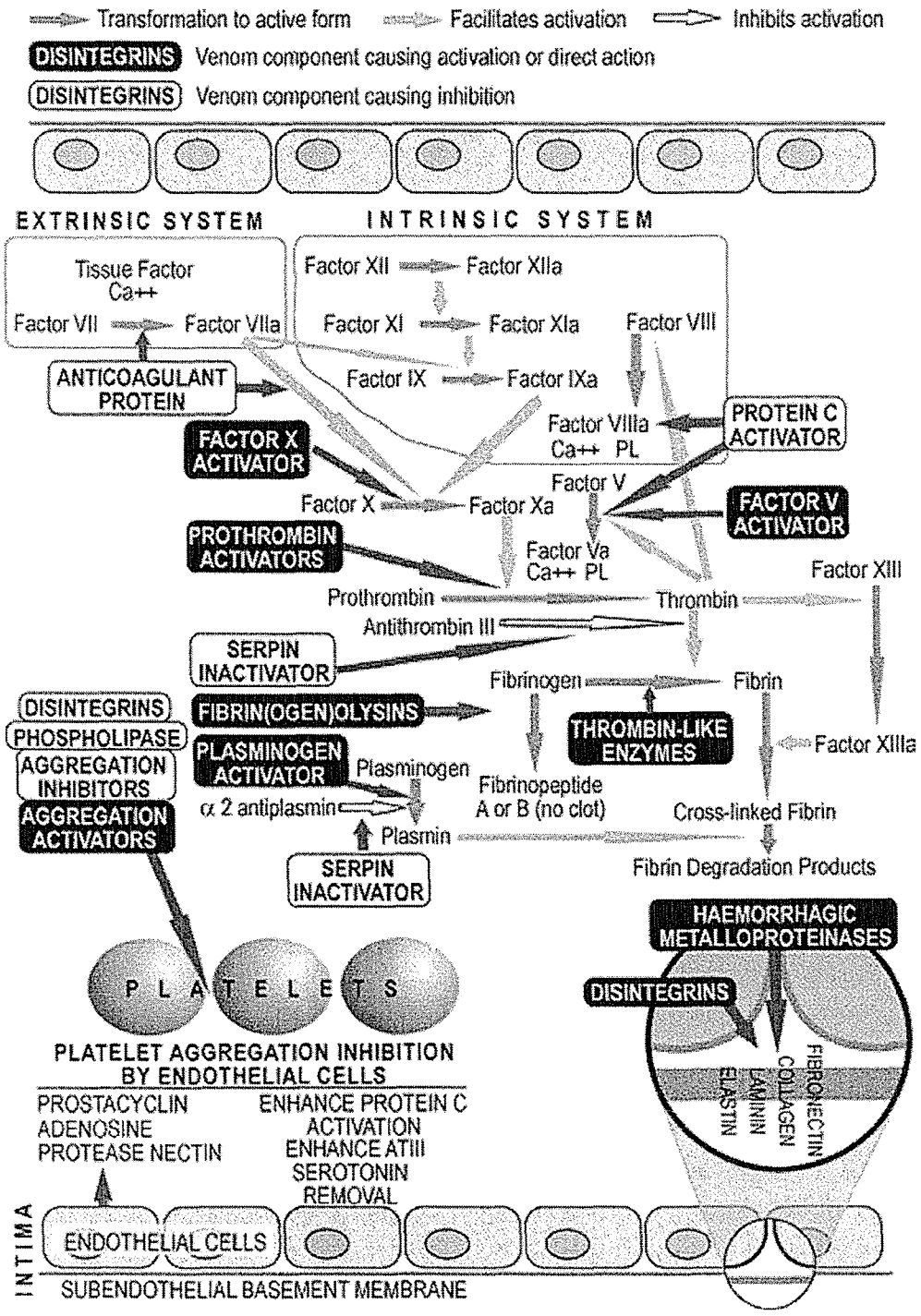
and extrinsic pathways meet). Tissue factor is a cofactor in the factor VIIa-catalysed activation of factor X. Factor VIIa, a serine protease, cleaves factor X to factor Xa in a manner identical to that of factor IXa of the intrinsic pathway (Davie, *et al.*, 1991; King, 2001).

Many species of the Australian elapids show a strong coagulant activity (Chester and Crawford, 1982; Joseph and Kini, 2002; Joseph, *et al.*, 2002). These contain factor Xa-like, and to a lesser extent factor Va, plus prothrombin activating enzymes (Suttie and Jackson, 1977; Rosing and Tans, 1988; Mann, *et al.*, 1990; Davie, *et al.*, 1991; King, 2001) which are responsible for most of the disruption of blood chemistry and haemorrhage seen clinically (DIC or defibrinogenation) (Sutherland, 1976; King and Smith, 1991; Lalloo, *et al.*, 1997), see Figure 4.1.

One group of therapeutically useful venom proteins that cause haemostatic reactions, which have been highly characterised in the venoms of true vipers (Viperinae) and pit vipers (Crotalinae), are the serine proteases. All have shown to contain highly effective serine proteases and act on different steps of the blood coagulation cascade (Ouyang, *et al.* 1992; Serrano, *et al.*, 1995; Markland, 1998a and b; Pirkle, 1998; Stocker, 1998; Samel, *et al.*, 2002). Some of these venoms can specifically hydrolyse fibrinogen to release fibrinopeptide A, fibrinopeptide B or both (Hofmann and Bon, 1987;Lollar, *et al.*, 1987; Markland, 1998a). Some affect other substrates, such as kininogen (Komori and Nikai, 1998; Nikai and Komori, 1998), plasminogen (Zhang, *et al.*, 1995; Zhang, *et al.*, 1997) and protein C (Stocker, 1994). It has been reported the variety of functions this family of venom proteases undertakes may, again, be due to gene duplication and an accelerated evolution which acquired these special functions (Deshimaru, *et al.*, 1996; Wang, *et al.*, 2001). Therapeutic venom proteases are used in areas such as antithrombotics to facilitate tissue oxygenation, prevent arterial embolism (Chang, *et al.*, 1995) and myocardial infarction (Farid, *et al.*, 1989).

The venom of *O. s. scutellatus* contains a potent converter of prothrombin to thrombin in the absence of all other known clotting factors (Denson, 1969) and it has been demonstrated the venom also possesses an activator of Factor VII (Nakagaki, *et al.*, 1992). The multimeric prothrombin activator purified from *O. s. scutellatus* was named scutelarin (EC 3.4.21.60) (Walker, *et al.*, 1980 Speijer, *et al.*, 1986).

Figure 4.1: Diagrammatic representation of principle ways snake venom interacts with human haemostasis. (Illustration copyright © 2005 reproduced with permission of Dr. Julian White).



is activator is approximately 260 kDa and is composed of four subunits: A 110 kDa subunit showing similar enzymic activity to Factor Va, an 80 kDa subunit similar to Factor X and two disulphide linked polypeptides, of which one or both contain the active site of 30 kDa each. Scutelarin is classed as a Group C prothrombin activating protein, requiring calcium and phospholipids for optimal activity. These proteins are also present in species of *Pseudonaja* (Masci, *et al.,* 2000; Williams, *et al.,* 1994; Masci, *et al.,* 1988; Stocker, 1994). Neither the structure nor sequence of the prothrombin activator from *Oxyuranus* has been reported.

4.2 RESULTS

4.2.1 N-terminal sequence

Proteins of *O. s. scutellatus* whole venom were separated on reducing SDS-PAGE and the N-terminal sequences are summarised in Table 4.1.

Table 4.1: N-terminal sequence of four protein bands from *O. s. scutellatus*.

Band	Size(kDa)	Query Sequence	Hit Sequence
B1	~110	?(Y)QLREYR	no matches of known proteins 2003 pseutarin
B2	~ 100	?EPLPLLT	EPLPLL
B3	~75	?LPAKTTF	no matches of known proteins
B4	peak 4 GF	?LLNFANLIE?	OS1= secretary phospholipase

4.2.2 Serine Protease Activity

Sequence data of one protein, B2, a 300 kDa protein under non-reducing conditions and 100 kDa under reducing conditions, showed a close match to a characterised serine protease from *Xenopus laevis* (GenBank accession no. AAB96905). Whole venom of *O. s. scutellatus* and *O. microlepidotus* were tested for activity using chromogenic p-nitroaniline substrates S-2288, S-2222 and S-2266 (Chromogenix). *O. s. scutellatus* showed the highest activity with S-2288, a broad range serine protease substrate (Figure 4.2). Whole venom from *O. s. scutellatus* was fractionated by size exclusion chromatography (SEC), and each fraction assayed for serine protease activity using S-2288. Most serine protease activity was found in peak one (Figure 4.3).

Figure 4.2: Protease activity profile from OS and OM whole venom using chromogenix substrates. Whole venom from OS and OM were assayed for protease activity using chromogenic substrate S2288 (1.74mM -Ile-Pro-Arg-pNA), S2222 (2.03mM -Ile-Glu-Gly-Arg-pNA) and S2266 (2.07mM -Val-Leu-Arg-pNA). Increase in absorbance at A405 is seen on the left over time in seconds.

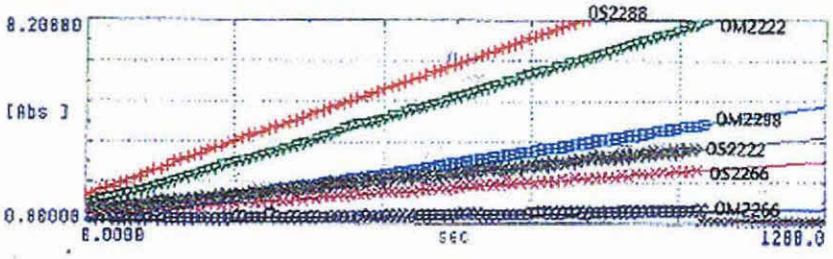
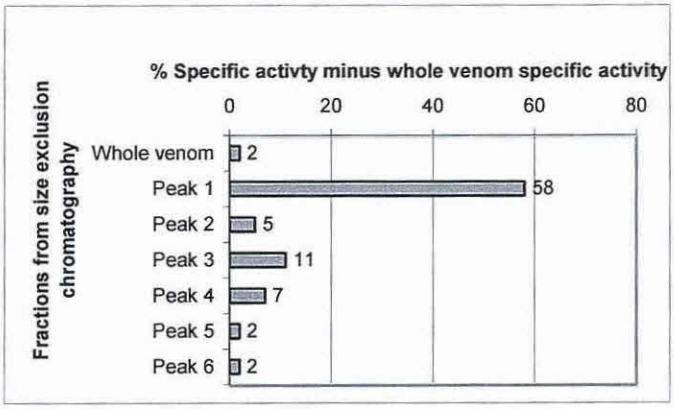


Figure 4.3: Specific activity of OS whole venom fractionated by size exclusion chromatography.

Percent specific activity of SEC fractions compared to whole venom specific activity when assayed for serine protease activity at A_{405} (S-2288).



4.2.3 Stability from fractionated venom

Most of the serine protease activity eluted in peak 1, this fraction contained both B2 and scutelarin. To establish if these proteins underwent autolysis, degradation of this fraction under varying handling conditions was established using electrophoretic analysis prior to the isolation of B2. Stability experiments were mimicked from Chapter 3 and in protocols Chapter 2.

Time

Samples of peak one were incubated at room temperature for up to 24 hours. No difference in electrophoretic patterning of protein banding above 80 kDa was detected, while those proteins below 30 kDa became very faint (gel not shown), similar to the patterning seen in pH studies.

Temperature

Incubation of peak one samples at differing temperatures was visualised on SDS-PAGE. Banding patterns were seen to be identical to the control up to 80°C, whereas at 100°C the protein precipitated. Repetitive freeze/thaw cycles did not affect electrophoretic mobility's of the protein fractions (results not shown).

pH

Differences between treated samples and control samples were detected when samples were stored at pH 6.0. A broad range of proteins, of 6.5 kDa, 10 to 14 kDa, 20 to 30 kDa and 45 kDa were very faint or non-existent compared to the control at pH 7.6. Though some protein recovery did appear at around 45 kDa with samples at pH 6.0 treated with PMSF or benzamidine. Contrasting to this, samples at pH 9.0 showed similarity to the control samples with the exception of an extra band at 14 kDa. Proteins appeared to be stable when stored at pH 9.0 (Figure 4.4).

4.2.4 Two dimensional gel electrophoresis

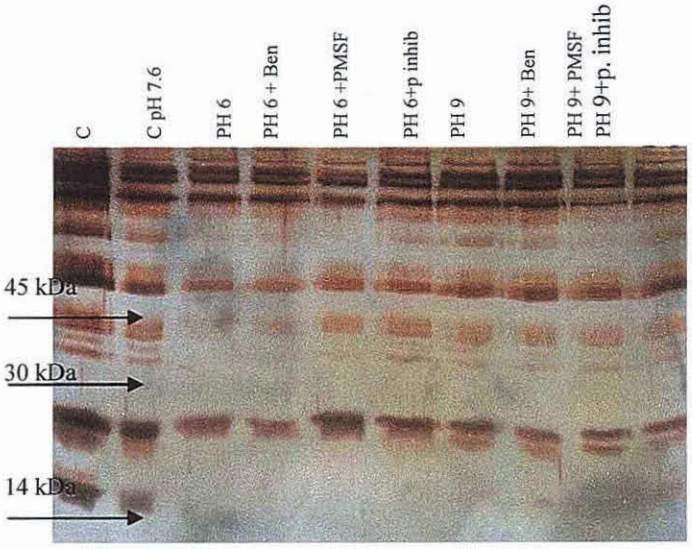
The variety and number of proteins present in peak one was demonstrated using 2DE. The 2DE images of the SEC fractions demonstrated a broad distribution of molecular masses and pI. The major bands observed in SEC peak one at 120, 100, 80, 30 and 14-6 kDa were separated, these smaller proteins migrated to a pI between 5 and 7 (Figure 4.5).

4.2.5 Isolation and characterisation of 300 kDa protein

The fractionated venom was stable under a wide variety of storage conditions. The serine protease activity seen in peak one from SEC could be due to the scutelarin (with factor Xa-like serine protease) or to B2. These proteins needed to be separated to establish their individual activity.

Figure 4.4: Effect of pH 6.0 and 9.0 to OS peak one.

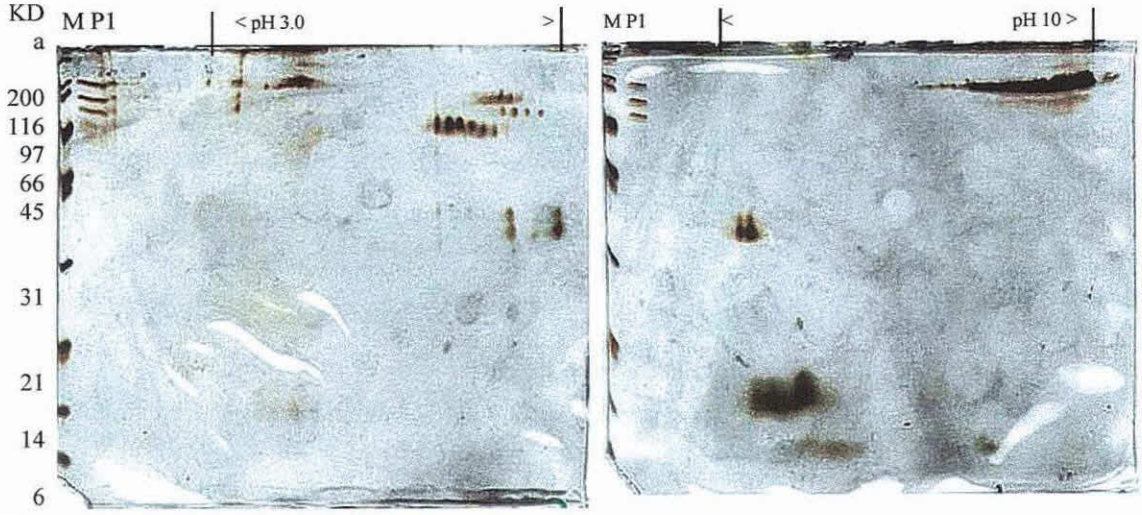
Peak one of OS whole venom, fractionated on size exclusion chromatography, was exposed to pH 6.0 and 9.0. Samples were separated on 15% SD- PAGE and visualised using silver stain.



1 ug at room temperature
C: Control, sample on ice
C pH 7.6: Sample not pH treated, diluted to copy pH samples
pH 6: sample exchanged to pH 6
pH 6 + Ben: Sample at pH 6 + 1mM Benzamidine
pH 6 +PMSF: Sample at pH 6 + 1mM PMSF
pH 6 + P. inhib: Sample at pH 6 + Bacterial Protease Inhibitor Cocktail mixture (2ml/20ml)
pH 9 as above pH 6 samples

Figure 4.5: Resolution of proteins in peak 1 by two dimensional gel electrophoresis.

Peak one (8 µg) was loaded onto an 11cm Amersham IEF strip pH 3-10. These 11 cm focusing strips would not fit on the BioRad mini protean II system, therefore were cut and run on two gels with a protein marker (M, BioRad) and 8 µg of peak one (P1).



Chromatography

Scutelarin and B2 could not be separated by either size exclusion or ion exchange chromatography. Both of these methods of protein fractionation did not adequately isolate B2. Affinity chromatography using a benzamidine HiTrap fast flow column (Amersham) was used to separate these proteins (Figure 4.6). Whole venom (Figure 4.7) and peak one from SEC (Figure 4.8) were eluted from this column. Peak one loaded onto the affinity column resulted in the isolation of a 300 kDa protein; B2.

Characterisation of an isolated 100 kDa protein

Storage and handling parameters of B2 were determined. Optimal buffering of the protein was checked through activity studies using S-2288. A variety of buffers were used within the working conditions for the p-nitroaniline substrate and the highest protein activity was detected using 50mM HEPES (pH 7.6) + 20mM NaCl, and this was used in subsequent assays (Figure 4.9).

The proteinase was active between pH 5 and 11, with the highest activity at pH 8 (Figure 4.10). Exposure of the proteinase to extreme pH's from pH 3 to 4 resulted in abolition of enzyme activity. Activity was detected at 0°C to 50°C, the highest rate of activity seen at 30°C (Figure 4.11). Abolition of activity was seen above 60°C to 80°C. B2 degraded rapidly over several hours at room temperature and also proved to degrade over two freeze thaw cycles, and this was reflected by a faint or absent protein band on SDS-PAGE (results not shown).

The specific activity or fold purification did not indicate B2 was highly active (Table 4.2). This assumption was reflected by Michaelis-Menten kinetics revealing a sigmoidal curve (Figure 4.12). Activities of B2 separated via ion exchange were also low (Table 4.3). Possible activators and inhibitors for B2 were investigated to indicate its molecular properties (Table 4.4). An activator of the protein was not found at this time, therefore further sequence of this protein was sought.

Figure 4.6: Affinity chromatography profile of whole venom.

Whole venom (20 µg) was loaded onto a benzamidine affinity column and run at 1ml/min at 4°C. Peak two seen in this trace was the protein flow through (F) not binding to the column. Peak 3–5 is a 1M NaCl wash (NaCl), he salt gradient, seen as the black line, and was used to remove non-specific binding. Peak 6 and 7 are the proteins eluting (E) from the column using glycine. Not shown in this trace are later column washes to remove bound proteins using Guanidine (G).
Peak one bound to this column and gave an identical pattern, with a lower absorbance in the flow through. Elution of Peak one (SEC) from the Benzamidine column isolated the B2 protein.

All chromatography was analysed using BioLogic software (BioRad) as per gel filtration.

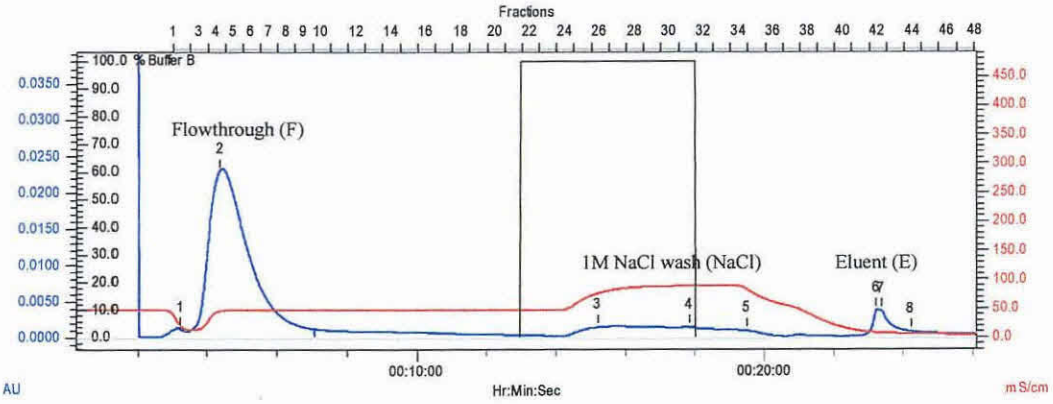


Figure 4.7: Whole venom bound and eluted from a Benzamidine column.
Fractions were separated by SDS-PAGE (15 %).
WV=whole venom

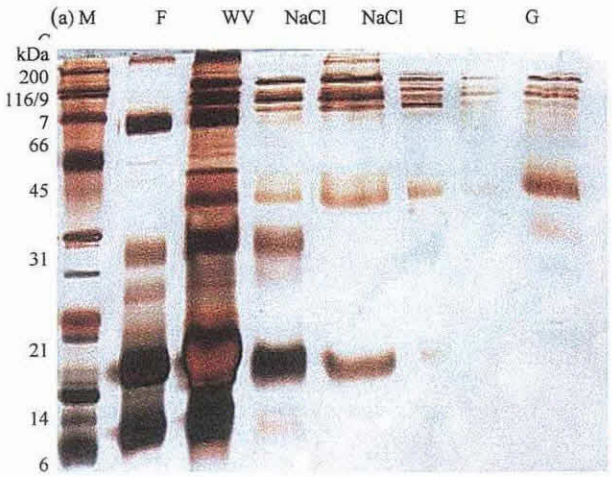


Figure 4.8: Peak 1 bound and eluted from a Benzamidine column.
Fractions separated by SDS-PAGE (12 %).
WV=whole venom

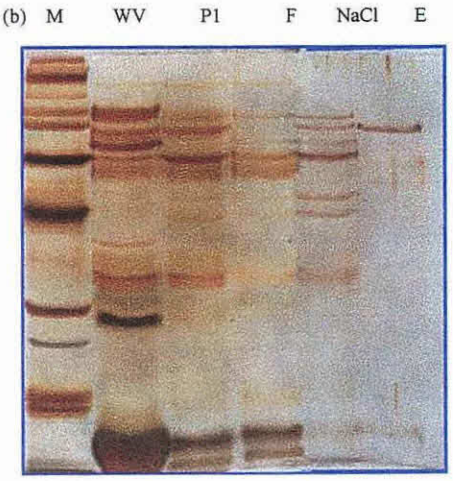
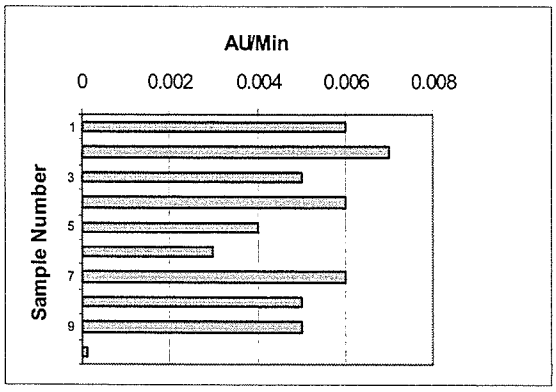


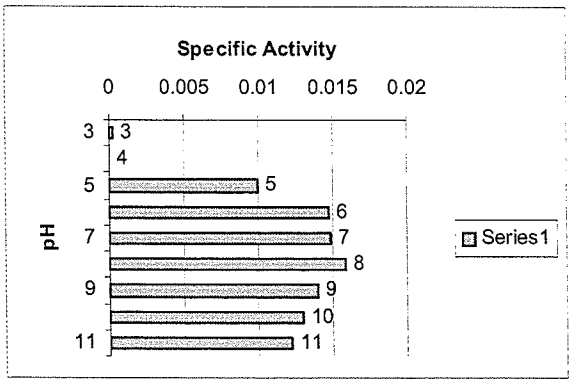
Figure 4.9: The effect of differing buffers to the activity of isolated B2

In a final volume of 100 µl, 10 mM of S-2288 and 0.001 mg/ml of enzyme, varying components were investigated in an attempt to maximise enzyme activity. All future work and assays was undertaken in 50mM HEPES (pH 7.6) + 20mM NaCl.



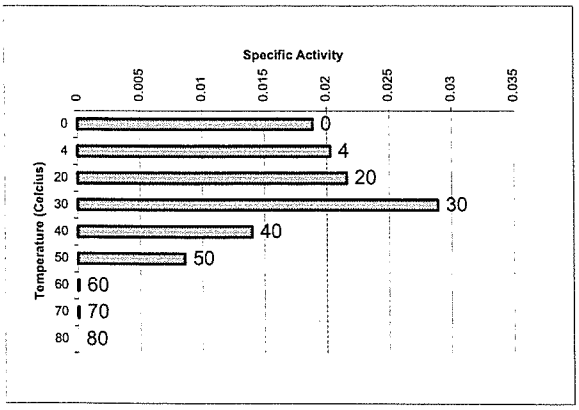
1) 50mM HEPES (pH 7.6) + 100mM NaCl
2) 50mM HEPES (pH 7.6) + 20mM NaCl
3) 50mM HEPES (pH 8.3) + 100mM NaCl
4) 50mM HEPES (pH 8.3) + 20mM NaCl
5) 50mM HEPES (pH 8.3)+0.5% BSA+20mM NaCl
6) 50mM HEPES (pH 8.3) + 0.5% BSA + 130mM NaCl
7) 50mM HEPES (pH 7.6) + 0.5% BSA + 20mM NaCl
8) 50mM HEPES (pH 7.6) + 0.5% BSA + 130mM NaCl
9) 50mM HEPES (pH 7.6) + 0.5% BSA + 20mM NaCl + 1mM CaCl + 1ul Vesicle
10)50mM HEPES (pH 7.6) + 0.5% BSA + 20mM NaCl + 1mM CaCl + 10ul Vesicle

Figure 4.10: Effect of pH upon enzyme activity.



Bar graph reflecting the specific activity (units/mg) of B2 was measured when exposed to various pH.

Figure 4.11: Effect of temperature upon enzyme activity.



Bar graph reflecting the Specific activity (units/mg) of B2 when exposed to various temperatures.

Table 4.2: Purification Table for B2

*U=Activity was determined by pNa release. One unit corresponds to the amount of enzyme, which converts 1 μmole of substrate per min at 37°C under standard conditions (Chromogenix S-2288)

Purification Step	Total Activity (Units*)	Specific Activity (units*/mg)	Purification (fold)	% Yield
Whole Venom	480,000	3,800	1	100
Gel Filtration: Peak one	7 900,000	2 600,000	677	1,625
Benzamidine Flow through	200,000	180,000	46	40
1M Wash	233,000	360,000	92	48
Eluent	11,000	224,000	58	3

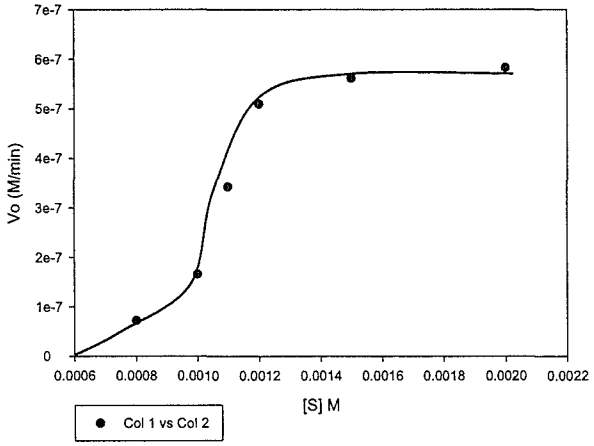
Table 4.3: Specific activity of OS fractions from ion exchange

Peak one from SEC was separated on ion exchange chromatography. Fractions were exchanged into 25mM HEPES+ 20mM NaCl and assayed for activity using S-2288 (10mM)

Column	Details	Specific Activity (units/mg)
Anion		
Peak one	SEC	0.034
Flow through	B2	0.001
Gradient	Scutelarin	0.042
Cation		
Peak one		0.042
Flow through	Scutelarin	0.080
Gradient peak 1		nil activity
Gradient peak 2	B2	0.03

Figure 4.12: Kinetics curve of B2 and comparative catalytic rates

Kinetics conducted using isolated B2 and S-2288 substrate. The rate obtained for the venom protein is compared to the established rates of other enzymes.



Enzyme	K_{cat}/K_m ($M^{-1}s^{-1}$)	K_m (M)	K_{cat} (s^{-1})	V_{max} ($M^{-1}min^{-1}$)
Acetylcholinesterase	$1.5x10^8$	$9.5x10^{-5}$	$1.4x10^4$	
Carbonic Anhydrase	$8.3x10^7$	$1.2x10^{-2}$	$1x10^6$	
Catalase	$4x10^8$	$2.5x10^{-2}$	$1x10^7$	
Fumerase	$1.6x10^8$	$5x10^{-6}$	$8x10^2$	
Scutelarin (OS prothrombin activator, spiejer)		$3x10^{-7}$		$7.1x10^{-9}$ /min/mg
B2 (OS 300kDa) S2288	$7x10^4$	$1.8x10^{-3}$	$8.2x10^1$	$9.1x10^{-7}$
B2 (OS 300 kDa) S2266	$5.8x10^1$	$1.3x10^{-2}$		

Note: all rates for B2 are $K_{0.5}$ and not K
Catalytic Rates of Enzymes (voet and Voet pg 353)

Table 4.4: Effect of various compounds on the activity of purified B2
The effect of various compounds on the activity of purified B2 was expressed as the % specific activity compared to the control. An "*" indicates where the solution precipitated when this compound was added.

Inhibitors	**% activity**	**Inhibitor/Activator**	**% Activity**
PMSF 10mM	10	EDTA 1mM	100
1mM	68	EGTA 1mM	100
Benzamidine 1mM	8	TritonX-100 1%	90
.5mM	18	Tween 20	95
50uM	50	BSA 0.5%	130
10uM	100	SDS 1%	1
NEM 10mM	17	CaCl 10mM	50
Iodoacetamide 10mM	60	CaCl 1mM	100
1mM	70	Mg 10mM	40
DTT 10mM	17	Mg 1mM	100
1mM	100	Mn 10mM	50
1,10-phenonitroline 10mM	40	Mn 1mM	100
1mM	90	Zn *10mM	300
EDTA 10mM	22	1mM	200

4.2.6 Determination of the amino acid sequence

Further sequence data from B2 through MS and later N-terminal sequence was obtained to establish any similarities to characterised proteins.

Mass spectrometry (MS)

Isolated B2 was sent to APAF for MS/MS sequencing. The resulting 5 peptide sequences, summarised in Table 4.5, required manual interpretation and the data was used to interrogate the public NCBInr database using the protein BLAST search program (http://www.ncbi.nlm.nih.gov/BLAST/). No confident matches were obtained using the sequences. Peptide four was used to produce an oligonucleotide primer for the isolation of B2 from the *O. s. scutellatus* cDNA library. MS scans can be found in appendix IV.

Table 4.5: Peptides sequenced from mass spectrometry (Sample RW300S)

Peptides sequences of B2 from mass spectrometry (Sample RW300S). Amino acids in brackets could not be discerned.

Peptide no.	Peptide sequence
1.	YF[L/I]VSNPGR
2.	AA[L/I]FAD[L/I]SPAE[L/I]R
3.	TW[L/I][L/I][L/I][Q/K]R
4.	AV[F/M][F/M]Q[Q/K]R (AVFFQ[Q/K]R)
5.	[L/I]NSHAG[L/I]V[L/I]PR

N-terminal sequence

Further N-terminal sequence of B2 was undertaken attempting to produce a long 20 to 30 amino acid sequence. This sequence would be used for database mining and for use as a 'megaprimer' to isolate B2 from the OS cDNA library. The resulting sequence, AING??ATCPQ??Y, did not return any database matches.

4.2.7 PCR

Due to the lack of sequence matches for this protein, the DNA sequence of B2 was pursued. Oligonucleotide primers were derived from peptide sequences to isolate this protein from an *O. s. scutellatus* cDNA library through PCR. In addition to this,

protein expression of the cDNA library was probed with isolated antibodies for B2 from the taipan antivenom.

PCR 1

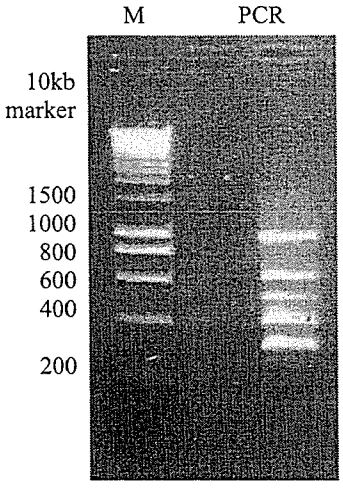
Oligonucleotide primers were produced from MS peptide 4 AVFFQQR (forward and complementary) and matched with T3 and T7 respectively. Two broad protein bands from both PCR resulted in 300 and 500 bp doublets. All sequences (10 for the 300 bp and 10 for the 500 bp band) obtained for these PCR products were different. One of the clones (Clone no. RW6) matched, with 88% accuracy, a short neurotoxin 6 precursor from *Pseudonaja textilis* (brown snake, GenBank accession no. Q9W7J7), and is described in Chapter 5. Nineteen of the 20 sequences obtained from this PCR did not match any of the sequence data held thus far for B2. One sequence matched, with 90% accuracy, peptide four. Subsequently, the *O. s. scutellatus* cDNA library was probed using this DNA sequence. Sequences obtained from this, again, did not match any sequences for B2. Sequences obtained were for HSP 70 (2 kb), myosin (800 bp), OS2 (800 bp) and other sequences of 2.5 kb, 1.5 kb and 1 kb did not have any matches in public databases. All sequences are described in Chapter 5 (RW6) or Appendix VII.

PCR 2

Due to the non-specific results from PCR 1, new primers for MS peptide four and the N-terminal sequence were produced using inosine and a nested PCR protocol. The resultant bands were strong and clear (Figure 4.14). Isolation, cloning and sequencing of these PCR products did not contain any of the N-terminal or MS peptide sequences of B2, nor did the sequences conclusively match DNA or protein sequences on public databases. Each of these bands were used to probe the cDNA library, no plaques isolated using these DNA probes matched any B2

Figure 4.13: PCR products from nested PCR.

PCR products (20 µl) from nested PCR were visualized on a 1% agarose gel.



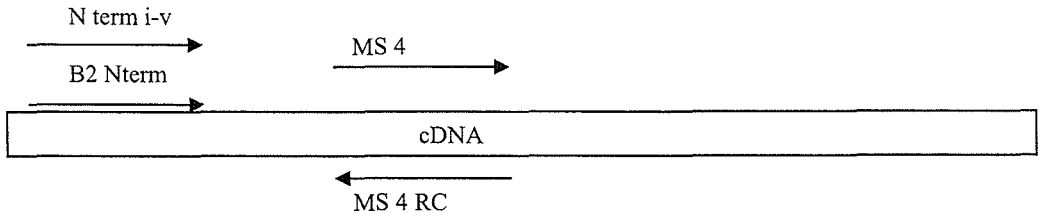
N-terminal or MS data sequences. Internet searches using BLAST matched DNA or translated sequences with 10 to 60% accuracy. From the 1000 bp band these matches were; hypothetical protein, POL protein, CD80, lipoprotein receptor related protein or no matches in the BLAST database.

PCR 3

N-terminal sequence was obtained with the goal to isolate B2 using a 'megaprimer', consisting of 10 to 20 amino acids. Unfortunately, the N-terminal sequence was neither clean nor long. Primers were synthesised to reflect part of the sequence received (N term i-v). With numerous specific primers (summarised in Figure 4.15), nested PCR was used to isolate a specific fragment from B2. Despite the previous success with PCR, and a complete replacement of all PCR reagents and primers, none of the primers would work with the cDNA library. Positive controls worked with established clones in pBS, though not in the library.

Figure 4.14; Primers specific for B2.

Primer sequences were deduced and optimised using MacVector then synthesized by PROLIGO.



4.2.8 Isolation of B2 using specific antibodies

Whole venom controls showed the "B2" antibodies bound strongly to B2 and to proteins smaller than 14 kDa (results not shown). To how see this result translated in the cDNA library, this isolated antibody was used to screen the *O. s. scutellatus* cDNA expressed with IPTG. Over 98 % of plaques were positive for this antibody (approximately 200 plaques), therefore, this technique was not pursued further.

4.3 DISCUSSION

4.3.1 N-terminal sequence and 2DE

Physical data was obtained from the largest proteins visualised under reducing conditions within the venom. Two of the three subunits were assumed to be part of the described prothrombin activator, the first 110-120 kDa band, denoted B1 was assumed to be the factor V subunit while the third 80 kDa band (B3), assumed to be the FX component (Speijer, *et al.,* 1986). The second 100 kDa protein band (B2), under non-reducing conditions was 300 kDa, suggested this protein formed a trimer of three subunits. Separation of this protein by ion exchange chromatography indicated it was not part of the prothrombin activator. A fourth protein from gel filtration fraction three of 14 kDa (B4), thought to be OS1, was sequenced to check the accuracy of protein preparation, BLAST searches and assumptions made of the protein bands.

Clean sequences of the N-terminal sequences were returned for all four protein bands, which was surprising due to the number of protein 'trains' seen on 2DE. This suggested the proteins represented in 2DE (Figure 3.13) as trains of spots are homologues or isoforms of the same proteins, with differences due possibly to variations of the levels of glycosylation. This also indicated B2 was composed of three identical subunits. The peptide mass data were matched against theoretical peptide mass and sequence of all known proteins in NCBI and SWISS-PROT databases. The identification parameters included peptide mass tolerance, number of peptides matched, modifications and protein sequence coverage of identified peptide sequences. Protein identification was based on using five or more matching peptides.

The 120 kDa band or B1, thought to be similar to factor V, did not match any proteins until an mRNA clone was excised for the *O. s. scutellatus* cDNA library matching the N-terminal sequence. In February 2002, Rao described a factor V subunit of *P. textilis* that matched the DNA sequence seen in *O. s. scutellatus* (Rao and Kini, 2002)(see Chapter 5). The 80 kDa (B3), also did not share homology with peptides in the BLAST database. This is not surprising as there is little venom protein data within public databases. The 14 kDa band 4, according to literature (Lambeau, *et al.,* 1990) should have been OS1 and proved to be so, exhibiting 100% identity to OS1 over the

area sequenced. The N-terminal sequence of the protein named B2 (300 kDa non reduced SDS, 100 kDa reducing SDS) was observed to be common with many BLAST matches. These matches were narrowed by searching for this sequence at the N-terminal region of other proteins. This search returned one hit. B2 showed 100% homology with the N-terminal region of a previously identified serine protease from *Xenopus laevis* (African Clawed Frog). Further investigation revealed the N-terminal area of the *Xenopus* sequence was outside the conserved region of the described serine protease. Therefore an investigation of B2 was required to ascertain if this protein was a 300 kDa serine protease.

4.3.2 Serine protease activity

Whole venom of all *O. s. scutellatus* and *O. microlepidotus* were assayed for broad serine protease activity, factor Xa activity and glandular kallikreins and factor XIa using Chromogenix substrates S-2288 and S-2222 and S-2266, respectively. Assays were undertaken using fresh venom and the venom stored for several months with no difference in assay results (see methods Chapter 2). Fractions from gel filtration were also stored at –20°C and showed no variation in electrophoretic patterning or activity over subsequent months (results not shown).

Whole venom from *O. s. scutellatus* showed the highest serine protease activity and *O. microlepidotus* showed strong factor Xa activity. As it was serine protease activity that was sought, *O. s. scutellatus* whole venom was fractionated to ascertain which fractions contained serine protease activity. Fractions from SEC showed the first peak fractionated from whole venom contained the highest serine protease activity. The remaining fractions contained low or no serine activity (Figure 4.3). This was expected, as subunits from the *O. s. scutellatus* prothrombin activator (scutelarin) are known to contain serine-like proteases. Unknown at this time was if B2 was also a serine protease. Thus the prothrombin activator needed to be separated from the 300 kDa protein.

4.3.3 Stability

It was initially hypothesised that the integrity and stability of venom components would be adversely affected by dilution, freeze thawing cycles and changes in storage temperatures. Results from electrophoretic patterns demonstrated that the same venom proteins would predictably degrade with the majority of the major protein banding remaining intact, regardless of treatment.

A potential problem with stability assays using fractionated venom samples was the dilution of the samples. Stability studies undertaken with whole venom in Chapter 3 reflected that degradation could occur through protease activity after dilution. Because of this, inhibitors were used in parallel with the stability studies. Initial stability studies of peak one from SEC under varying handling and storage conditions reflected the whole venom stability study undertaken. Peak one from SEC was placed at room temperature for extended periods of time showing no change in electrophoretic patterning until after 24 hours, after this time minor band patterns between 20 to 30 kDa become very faint, similar banding patterns were seen within the pH study (Figure 3.2). Protein recovery was seen at 45 kDa by the use of inhibitors Benzamidine< PMSF< Protease inhibitor and may have been due to a prevention of serine protease activity. Proteins between 20 to 30 kDa were not recovered by inhibitors, indicating another mechanism was occurring at pH 6.0 or protein degradation. Degradation of these protein bands were consistent between whole venom and peak one. Importantly, throughout all stability studies the protein B2 showed consistency throughout the studies. This protein would now be isolated.

As a point of interest, peak one did not degrade in the same manner at pH 6.0 as pooled fraction one from *O. s. scutellatus* (F1: pooled SEC peak one and two, Chapter 3). At pH 6.0, neither SEC peak one, two (results not shown) or whole venom showed this degradation. It could be suggested that the interaction of proteins from peaks one and two combined allowed the mechanics of an acid protease mechanism to be present. Further work not encompassed within this thesis will confirm or reject this hypothesis.

4.3.4 Isolation of the serine protease (B2) from *O. s. scutellatus*

Chromatography

Isolation of B2, the 300 kDa serine protease, was not achieved by SEC (Chapter 3) due to a number of large proteins eluting within the same fraction. Ion exchange of SEC peak one separated scutelarin (120, 80 and 30 kDa) from B2. The prothrombin activator eluted on the salt gradient of anion exchange, the 300 kDa (B2) protein eluted in the flow through, with smaller proteins eluting within each of these fractions. B2 showed little activity when assayed with S-2288, and it was unknown at this time if these proteins were associated with this protein. Alternatively, when SEC peak one was eluted from a cation exchange column, scutelarin eluted in the flow through as B2 eluted in 25% 1M NaCl along with a 14 kDa protein. The serine protease activity of these fractions revealed the B2 fraction was now active. The activity of B2 eluted from anion exchange, with the addition of varying concentrations of salt from 25 mM up to 100 mM, was not comparable to rates seen from the cation exchange (25% NaCl). This indicated the presence of Na^+ was not the cause of the differences of activity (Dang , 1997).

It is possible the activity was due to the lower pH of the cation exchange buffer at pH 6.0, the Chromogenix substrate activity was not optimal at pH 6.0 and therefore was not pursued. A more probable explanation for the differences in activity between ion exchange is indicated in 2DE. The pI of the 120 kDa, 100 kDa, 80 kDa and 30 kDa proteins overlap in the pI range between 4.5 to 6, due to multiple isomers. It is possible the active subunits from scutelarin were contaminating the other fractions eluting from ion exchange causing the changes in activity seen. This would be possible due to the differing binding and elution profiles of the ion exchange columns. Whereas the anion exchange separated proteins from pI 4.9 to 4.6 cation exchange did not show the same separation with protein separation between 6.9 and 8.7 shown (Appendix V and VI).

Affinity chromatography

Assuming B2 to be a serine protease with a differing Ki to scutelarin, separation of these proteins was undertaken using affinity chromatography. Whole venom and peak one from SEC were loaded onto the column and bound proteins were eluted using a low

pH. Seven protein bands eluted from the column when both whole and peak one venom samples were loaded, including B2. Stringent salt washes and elution of bound proteins from SEC peak one eluted one protein; a 300 kDa protein, B2. The other proteins eluting in the extended salt wash (Figure 4.7 and 4.8). Some bands of 100 kDa were also within the 10 ml salt wash, possibly reflective of a mixture of protein subunits and the carbohydrate moieties attached as indicated by the Concanavalin A staining in Chapter 3 and the visualisation of protein trains seen in 2DE in Figure 3.13. Further characterisation utilising deglycosylation of the proteins and re-running the two-dimensional gels or affinity chromatography would resolve this question.

B2 had been isolated and elution from the affinity column revealed the three 100 kDa subunits of B2 bound benzamidine with more affinity than scutelarin. Subsequent assays for serine protease activity showed both the salt wash and eluent active for serine protease activity. It could be argued that any activity seen in the eluent may have been due some 30 kDa active subunits of the prothrombin activator, which may be bound to the benzamidine column and not be visualised by silver stain, giving B2 false activity. The scutelarin subunits have been shown to be inactive on their own, requiring calcium to catalyse activity. Assays were undertaken with and without calcium. Although sluggish, the 100 kDa (B2) protein did confer activity without calcium. Little activity was seen in the salt wash fractions without calcium. From this it was assumed the 30 kDa subunit was not a factor in the activity seen of B2. These steps resulted in a highly purified end product with a modest but workable yield.

4.3.5 Characterisation of B2

Varying buffer conditions were used to reflect the varying enzyme requirements reflected in *O. s. scutellatus* literature such as calcium and phospholipids (vesicles) required for scutelarin and differing salt concentrations required of enzymes indicated by Chromogenix. Protease activity was greatest in 50 mM HEPES (pH 7.6) + 20 mM NaCl and showed an average physiological working range for pH and temperature.

The isolation of B2 revealed it degraded rapidly. Storage at -20°C for over 1 week, or greater than two freeze-thaw cycles resulted in the loss of activity and protein integrity when separated using SDS-PAGE and visualised with silver stain. This

confirmed the opinion held in Chapter 3 comparing protein-banding patterns to protein degradation. The degradation of B2 was not reflected in fractionated samples or whole venom, where these samples were stable for over 12 months (results not shown). The degradation of isolated B2 may have been due to a lack crowding agents required to maintain folding or to prevent aggregation (Berg, *et al.*, 1999). Assays using BSA, glycerol or Tween 80 did not change activity or prevent the degradation of B2. Dilutions of both whole venom and peak one to the same concentration of B2 did not initiate the rapid degradation of B2. The stability noted with whole and fractionated venom may be due to an inherent inhibitor within the venom, a factor absent within the affinity isolated protein. Although not undertaken at this time, further stability studies should encompass the identification of SEC fractions or specific proteins (especially in peak one) within the venom that may prevent the degradation of B2.

Purification of B2 and activity assays did not indicate the 300 kDa protein was highly active. This lack of activity may have been due to sample degradation, yet kinetics suggested this protein underwent allosteric control (Figure 4.12). The sigmoidal curves obtained from assays using B2 were not reflected in assays undertaken using peak one from SEC or fractions containing scutelarin, which resulted in a hyperbola. The broad range serine protease substrate S-2288 gave a $K_{0.5}$ $1.8x10^{-3}M$ and $kcat/k_{0.5}$ $7x10^{4}$ $M^{-1}s^{-1}$ and $k_{cat(S-1)}$ 8.2 x $10^{1}sec^{-1}$. The S-2266 gave a $K_{0.5}$ $1.3x10^{-2}M$ and $kcat/k_{0.5}$ $5.76x10^{1}$ $M^{-1}s^{-1}$. No activity was seen for Chromogenix substrate S-2222. These rates were far lower than those seen of scutelarin. As B2 bound tighter to the benzamidine column this may suggest B2 has higher substrate specificity than scutelarin, yet this was not demonstrated. Either the broad range serine substrate was not optimal for B2 or perhaps the allosteric activation observed with the kinetics was real.

A possible activator for this protein was sought, or, if an activator could not be found inhibitors affecting the activity of B2 may divulge a clue its molecular properties (Table 4.4). The protein was strongly inactivated by the serine/trypsin protease inhibitor benzamidine and higher concentrations of PMSF. Inhibition was also observed for metal chelator 1, 10-phenonitroline although high concentrations were required for an effect. Both alkylating reagents NEM and Iodoacetamide and reducing agent DTT appeared to modify enzyme activity, high concentrations of these were also required. Non-ionic detergents such as Triton X-100 and Tween 20 had a negligible effect on

enzyme activity. In contrast, SDS effectively prevented any activity of the protein. BSA showed to have a small effect on the enzyme activity. Proteinase activity was insensitive to the bivalent cations Mg^{++} and Mn^{++} and Ca^{++}, while the addition of Zinc increased enzyme activity drastically. This was found to be due to a precipitate forming when zinc was added to the assay. The decrease in activity at 10 mM may be due to crowding of the protein or a decrease in the flexibility of the protein, thus activity. An increase in Na^{+} concentrations did not confer increased activity.

Further sequence data was sought through MS to see if internal sequence would match any known serine proteases and give an idea of a possible physiological role for this venom protein.

4.3.5.1 Sequencing

To further characterise the isolated protein, rather than N-terminal sequence, internal peptide sequence data was sought through mass spectrometry (MS). Internal MS sequence may match proteins within public databases and give an indication of the family this protein may belong. Six peptides were retrieved from the glycosylated B2 through MS/MS. The sequences required manual interpretation and the data was used to interrogate the public NCBInr database using the protein BLAST search program (http://www.ncbi.nlm.nih.gov/BLAST/). No confident matches were obtained using the sequences. It was stated by APAF with reasonable confidence that peptide 4 contained two phenylalanines (FF), but this could not be guaranteed. The second last amino acid is Lysine and Glutamine, as the Lysine may be a tryptic miscleavage. Although MS has become a central element for proteomic analysis (Siigur, *et al.*, 1998; Sun, *et al.*, 1999; Kamiguti, *et al.*, 2000), mass mapping relies on databases of known peptide sequences (such as predicted trypsin digest fragment sequences of known proteins)(Wilkins, *et al.*, 1999). This approach can be successfully applied for organisms where genomes of which are fully sequenced and annotated, or for proteins having sequences well conserved for cross species reference (Wang, *et al.*, 2003).

This approach to sequencing specific *Oxyuranus* venom proteins was not productive due to the paucity of knowledge of venom proteins and the low representation of venom proteins within public databases. This problem has also been described within other Elapid research (Kinter and Sherman, 2000; Jonsson, 2001; Kazmi and Krull, 2001; Fry

et al., 2003a). If the entire protein were subjected to MS, without sequence similarities to align the MS peptides against the sequence would not be able to be assembled.

The nucleotide sequence of B2 was pursued using *O. s. scutellatus* cDNA library using primers derived from the N-terminal and MS data using PCR.

4.3.5.2 PCR

Oligonucleotide primers were chosen from protein sequences that gave the least number of codon usage. Inosine was later used to increase the concentration of primers due to this codon usage variation. It was unknown if the three subunits were translated as one 100 kDa peptide, associating to form a trimer, or, if all three protein subunits were coded on one cDNA producing a 300 kDa protein. Therefore the expected size of a PCR product for B2 using the N-terminal sequence at one end of a cDNA clone and T7 at the other was between 1.5 to 6 kbp. Glycosylation of the mature protein also hindered an estimation of the correct protein size, thus an estimation of DNA base pairs. Deglycosylation of the protein was attempted, though the reagents did not work, from this all PCR products over 1 kbp were potentially the target DNA sequence. This size variation was compounded through the use of mass spectrometry peptides for PCR primers. As it was unknown where in the protein the MS peptides were sequenced, therefore the size of the product was ultimately unknown. All PCR products from 100 bp to 6 kbp may have contained the target DNA sequence.

Nevertheless, PCR was pursued using variations in PCR reagents and cycling parameters to increase specificity, while variations of specific primers were used to isolate a DNA sequence containing one or more of the N-terminal or MS peptide sequences once translated. This approach was not successful. The minimal sequence data and unknown protein size could mean the target DNA sequence may have been isolated but not recognised. It is possible the PCR products isolated and sequenced were DNA segments between the N-terminal or MS peptides. The inability to find the target DNA sequence after probing the cDNA library with PCR products from nested PCR (PCR 2) was a disappointment. It is not understood why so many non-specific proteins were isolated. It is possible there are multiple conserved regions through out

the venom peptides, a possibility that could be attributed in the results seen using antivenom later.

Alternatively, perhaps the cDNA encoding the B2 protein was not present in the cDNA library. Described in Chapter 5, the largest protein band (B1) was shown to be the factor V subunit of scutelarin. This protein band appears, through electrophoretic patterning, to be a prominent band similar to the bands at 100 kDa (B2) and 80 kDa (B3). This 120 kDa protein comprised 10% of the 115 clones isolated from the *O. s. scutellatus* cDNA library. Surprisingly other highly expressed characterised proteins were also not found eg. the subunits of taicatoxin (Possani, 1992) or any factor X-like subunits (Fohlman, Eaker *et al.* 1976). As it is known *O. s. scutellatus* species contain the major proteins mentioned above, perhaps there is some, as yet undiscovered, regulation of expression of the differing venom toxins, causing an abundance of some mRNA sequences and a rareness of others.

4.3.6 Antivenom binding studies to isolate B2

Antibodies bound to B2 were isolated from taipan monovalent antivenom (CSL). These antibodies were used to probe an expressed cDNA library and were also bound to *O. s. scutellatus* whole venom separated on reducing 15% SDS-PAGE as a control. Antibody binding was visualised using DAB and showed bands at 100 kDa and a large, broad band at 6 to 15 kDa. These antibodies, while binding strongly to B2 also bound to lower molecular weight proteins. This was of concern as these proteins make up to 80% of the venom components, decreasing the probability of isolating B2. Expression of the cDNA library was continued, unfortunately the resultant isolated B2 taipan antibody reflected the scenario of the control, and approximately 98% from 200 plaques counted bound the isolated antibody.

Chapter 5

Sequences of *O. s. scutellatus* venom toxins

5.1 INTRODUCTION

Snake venoms contain a large number of biologically active substances, and much work has been conducted with the most abundant and lethal proteins of *Oxyuranus*. However, there has been little work conducted in regard to increasing a broad general base of understanding of this venom and its constituents. It is from a general base of understanding from which interesting or novel proteins molecular interactions may become apparent.

Chapter 4 introduced the use of a venom gland cDNA library from *O. s. scutellatus* in an attempt to isolate a novel protein identified from whole venom. This cDNA library was expressed and probed using taipan monovalent antivenom (CSL) to discover the antigenic nucleotide and potential peptide sequences that were present within whole venom.

5.2 RESULTS

The sequences obtained from the 5' and/or 3' ends of isolated antigenic clones were analysed and homologous sequences identified using the nucleotide (BLASTN) and protein (BLASTX) programs. The identification, isolation, sequencing and identification of open reading frames and subsequent putative peptides of these sequences are described in sections 2.2.2.12 and 13). Many of the clones analysed revealed significant similarity with sequences from different sources (Table 5.1, for sequences see Appendix VIII). A variety of venom proteins matching, for example, polymerases, carboxylases, dehydrogenases and myosin sequences have been found within venoms, belying their assumed origins (Nawarak, *et al.*, 2003). Therefore, the venom or non-venomous origins or designated roles of many proteins could not be allocated through sequence homologies.

The majority of sequences isolated shared homology with proteins used for protein folding or as chaperones (protein disulphide isomerases (PDI) and HSP 70 and 90). The translation of some nucleotide sequences also required the use of vertebrate mitochondrial codon usage to obtain an ORF. These sequences shared homology with cytochrome C and NADH dehydrogenase. For a minority of sequences an ORF could not be deduced regardless of the codons used.

Of the clones isolated from the cDNA library, 17% of the nucleotide sequences from a total of 115 clones were identified and showed homology with characterised toxins. Due to the large number of nucleotide sequences and implied proteins, thirty clones were chosen to be fully sequenced. These clones were chosen based on the results of BLAST searches of their 5'-sequences or matches of interest.

A group of eight sequences were selected for discussion within this chapter based upon observed toxic components from *Oxyuranus* and elapid envenomation (Table 5.2). Four of the putative peptide sequences shared identity with PLA_2 and pre-synaptic neurotoxins. These are likely isomers of taipoxin subunits and OS2. Two of the sequences putatively encode the β-taipoxin subunit and the OS1 peptide with one amino acid residue difference to established peptide sequences. Potentially, two new post-synaptic short-chain neurotoxins have also been identified. One of these sequences showed homology with *O. scutellatus* post-synaptic neurotoxins, toxin 1 and 2 and may have been encoded by two different genes. The final sequence represents a class of post-synaptic toxins not previously described within *Oxyuranus*. The complete factor V component of scutelarin is also described (Table 5.2). The full nucleotide and putative translated sequences of these clones are included in an annexure at the end of this Chapter.

This is the first time these nucleotide sequences have been reported.

Table 5.1: Summary of mRNA sequences from *O. scutellatus* venom gland cDNA library and the sequences they share homology with.

Summary of sequences an *O. scutellatus* venom gland cDNA library and their alignment to sequences to which they share homology.
Database comparisons were carried out using the National Centre for Biotechnology Information website (www.ncbi.nlm.nih.gov). The nucleotide (BLASTN), translated and protein sequence (BLASTX) databases were mined while searching for homologous sequences.
Nucleotide sequences of the *O. s. scutellatus* clone are listed as the clone reference no. and/or the name of the peptide to which the sequence shared homology. This is followed by the length of the nucleotide sequence, GenBank accession number and the size of the potential mature peptide are shaded.
Sequences with the highest homology to these sequences are listed by description, size and their GenBank accession number.
To the right of the GenBank accession number is the sequence coverage of the *O. s. scutellatus* sequence to sequences from the BLAST database expressed as a percentage, and is followed by the matching residues within the sequence, again listed as a percentage. Due to the paucity of toxin nucleotide sequences, most matches were made using the BLASTX database. The nucleotide sequences are included in Appendix VIII.

(ME: Clone isolated using mass excision)

***O. s. scutellatus* clone reference no.**	**No. base pairs**	**GenBank accession no.**	**Length of translated protein**	**Match from BLAST database displaying homology to *O. s. scutellatus* clones**	**Length of protein**	**BLAST match accession no.**	**% Sequence coverage**	**% Match**	**Notes**
Toxins & serpent DNA matches									
R 15, R 24 mega: RW 18, 19, 20, 22	690	AY691657	182	Beta taipoxin subunit	118	AF332697	100	100	Signal peptide not published
R 120	625	AY691659	151	similar to gamma taipoxin subunit	133	P00616	100	94	Signal peptide not published
R 125	634	AY691661	148	similar to OS2	119	AAB33760	100	94	Signal peptide not published
R56/Ron 4, Wel14/R4	615	AY691662	147	similar to OS2, 1aa diff to R125	119	AAB33760	100	95	Signal peptide not published
RW 26, 27, 28,22	467	AY691658	146	Similar to α taipoxin	119	AAB33760	100	86.5	α taipoxin and OS2 very similar
R 132	621	AY691660	155	OS1	118	AAB33759	100	100	Signal peptide not published
R 121	687	AY691664	84	post-synaptic neurotoxin (*Naja sputatrix*), short neurotoxin *(Lapemis hardwickii*)	83 81	AAL54894 AAD08813	100 100	90 90	121&147 similar for two stop codons then diverge
R 147	468	AY691664	84	As above		As above			
R 27	806	AY691663	144	natriuretic peptide (O. *scutellatus*)	35	P83228	100	100	
	806		144	Natriuretic peptide (*Micurus corallin natriuretic*)	185		33 to 41%*	100	Dependent on amino acid match or similarity
RW 6	445	AY691665	78	Post synaptic neurotoxin (*pseudonaja textilis*)	79	Q9WYJ7	100	70	Possible there are other proteins coded within this sequence

***O. s. scutellatus* clone reference no.**	**No. base pairs**	**GenBank accession no.**	**Length of translated protein**	**Match from BLAST database displaying homology to *O. s. scutellatus* clones**	**Length of protein**	**BLAST match accession no.**	**% Sequence coverage**	**% Match**	**Notes**
R5	3317		No ORF	cobratoxin *(Naja atra)* PLA_2*(Bungarus multisinctus)*	2387 bp 2358 bp	Y13399 AJ416991	2 2	87 87	Many ORF Genomic DNA match
R8	1928		No ORF	Gene (*Elaphe obsolete*) *Atractcospis microlepidosi*	535 bp 469 bp	AF544661 D13522	8.5 7	89 93	Genomic DNA Many ORF
R11	1501		No ORF	*Trimeresurus Flavoviridis*	301 bp	D31777	3	84	Many ORF
R 51	657			*Elaphe bimaculata*	812 bp	AF236671	60	84	genomic DNA
R65 (incomplete)	1251		No ORF	*Vipera ammodytes* (Bov B) *Bungarus multisinctus* PLA_2 inhibitor (*Elaphe quadrivirgata)* PLA2 gene (*Laticauda colubrine*)	4606 bp 1203 bp 2433 bp 4304 bp	AF332697 AJ25122 AB060638 AB062448	1.5 0.1 105 0.8	93 93 97 91	Genomic DNA matches all in the same area of all sequences.
Complete									
FV	4725	AY691656	1460	Pseutarin (*pseudonaja textilis*)	1460	AAO38805	100	97	
	4725		1460	FV (*Homo sapiens*)	2251	P12259	100	50	
PDI	2495	AY691666	531	PDI (*Gallus gallus*)	515	P09102	100	78 to 88	
HSP 70	2340	AY691667	666	HSP 70 (*Gallus gallus*)	634	AAP37959	100	97	
EF2	1315	AY691668	406	EF2 (*Gallus gallus*)	858	AAA87587	47	98	Clone complete
				EF 2 (*Homo sapiens*)	357	P13639	100	98	
				EF2 (*Rattus norvegicus*)	342	NP058941.1	100	98	
CGI protein	682	AY691669	182	CGI protein (*Homo sapiens*)	224	NP861974	81	98	Clone complete
2.19 gene	763	AY691670	165	MemberA (*Homo sapiens*)	230	AAH08912	71	90	Clone complete
Bet 3	757	AY691670	182	BET 3 (*Homo sapiens*) BET 3 homologue (*Mus musclus*) 180	180 180	NP055223.1 NP038746.1	100 100	97 97	
Polyposis-R 117/ TB2	772		214	Polyposis (*Homo sapiens*) TB2 (*Homo sapiens*)	185 198	NP005660 AAA66351	100	71	ORF + signal
Cytochrome C subunit IV (R144)	700	AY691674	180	Cyt C (IV) (*Meleagris gallopavo*) Cyt C (IV) *(Bos taurus)*	171 169	AAQ14274 AA93149	100 100	71 64	

***O. s. scutellatus* clone reference no.**	**No. base pairs**	**GenBank accession no.**	**Length of translated protein**	**Match from BLAST database displaying homology to *O. s. scutellatus* clones**	**Length of protein**	**BLAST match accession no.**	**% Sequence coverage**	**% Match**	**Notes**
R 19	1504	AY691673	171	Polydenylate binding protein Ribonuclease protein	126	AAH01716 XP352454 AAG01716	100	94	+ signal peptide ? polycistronic
RW 21				*Gallus gallus* C-term end	1948 bp	Bx934062			
R 26	811			Archease (*Rattus norvegicus*) (*Mus musculus*)	167 146	XP216347 BAB26341	42	80	Many ORF, some overlap
R 36	1308			(*Gallus gallus*)	1070 bp	BX930134	46	81	Many ORF; DNA match
R 42 ME	2013			NBM					Many ORF
R 46, RW32 ME	1851		No ORF	*Gallus gallus* *Homo sapiens*		CR354147 BX537527	1-1110		Genomic DNA ←
R 50 ME	2609	AY02468		Endocrine regulator (*Homo sapiens*)	2099 bp	AAD17298 AK025505	43	55	ORF, matches mRNA
R 54 ME	623			NBM					Many ORF, several match 2 seq join
Incomplete									
HSP 90 ME	2523	5' AY02457 3'AY702458	680	HSP 90 (*Mus musculus*)		XP142222			Internal could not be sequenced
	2523		680	HSP 90 (*Homo sapiens*)	2567bp/ 725aa	NM007355	93	96.5	
EF1			112	EF 1 (*Bos taurus*)	181	BAC5675.1	61	94	Clone not complete
Oncogene	1433	5' AY702460 3' AY702461		B Abelson oncogene(*Homo sapiens)* Tyrosine Kinase(*Homo sapiens*)	1182	NP009298 P42684	21	100	
Disease resistance protein (R 31)	1333	AY691672	→ ←	Lecithin retinol or disease resistance protein					
G prot binding (R32)	1305	5' AY702469 3' AY702470	207	G protein binding (*homo sapien*)	634	NP036473	31	100	
Ca ATPase (R 122)	797	AY702462	123	Ca transportingATPase (fast twitch skeletal muscle) *Gallus gallus*	994	A32792	12	91	
Golgi assoc (R62)		AY702463		Golgi associated (*Homo sapien*)	900	S35342	16	78	
Lecithin Retinol (R20)	2326		667	(*Mus musculus*)	228	NP076113	100	68	
Carboxypepsidase	740	AY702464	247	Carboxypepsidase A_2 (pancreatic (*Homo sapiens*)	417	NP001860	59	80	

***O. s. scutellatus* clone reference no.**	**No. base pairs**	**GenBank accession no.**	**Length of translated protein**	**Match from BLAST database displaying homology to *O. s. scutellatus* clones**	**Length of protein**	**BLAST match accession no.**	**% Sequence coverage**	**% Match**	**Notes**
Carboxylate	903	AY702465	214	Carboxylate synthase (*Homo sapiens*)	501	NP115981.1	43	66	
Riborphorin (R45-RW31)	1439	5'AY702474 3' AY702475	343	*Mus musculus* *Homo sapiens*	608 608	AAH16080 NP002941	56.5 56.5		Internal not sequenced
Protoadherin				*Homo sapiens*	932	O9Y5H0			
B Actin	796	AY702473	259	*Sigmodon hispidus* *Gallus gallus*	375 375	AAL16942 CAA25004	69	100 100	
A actin	726	AY702472	242	Skeletal (*Homo sapiens*)	377	NP001091.1	60	100	+ signal
myosin light	594	AY702471	177	Myosin regulatory light chain-(*Homo sapiens*) *Felis catus*	173	NP002468 AAK00755	100	91	
Myosin heavy	664	AY702466	221	Superfast myosin heavy(*Felis catus*)	1945	BAB15219	11	85.5	
Ribonuclease HI sub	877	AY702467	218	Ribonuclease HI lge(zebra fish) and *Xenopus laevis*	308	AAH46061	70	78	Clone complete
Cytochrome C subunit IV	700		180	Cytochrome C oxidase subunit IV (*Bos taurus*)	169	AAA30461.1	100		
Mitochondrial sequences									
Cytochrome C subunit 1	744	AY691675	247	*Dinodon semicarinatus*	533	NP008421.1	46	86	Incomplete, N term seq starts 70bp into seq.
Cytochrome C subunit II	669	AY691676	223	*Dinodon semicarinatus* *Laepemis hardwickii*	228 192	NP008422.1 AAL55554	100 100	83 90	Complete
Cytochrome C subunit III	665	AY691677	221	*Dinodon semicarinatus*	261	NP008425.1	85	94	Incomplete, C term would not sequence
NADH dehydrogenase (RW7) R 40	700	AY70256	91 142	*Dinodon semicarinatus* *Dinodon semicarinatus*	95 445	BAA33030 NP008428	100 32	81 85	Incomplete, **Two ORF, two different proteins on one sequence**
Many ORF									
Wel 12, R2 (PCR)	1273	-		Similar matches to R 65					Similar to R 65 matches
Wel 16-R6 (PCR)	1247	-		NBM					Internal not sequenced
Wel 10 24/3	744	-		NBM					No ORF
R 131	741	-		NBM					Good sequence, no ORF

***O. s. scutellatus* clone reference no.**	**No. base pairs**	**GenBank accession no.**	**Length of translated protein**	**Match from BLAST database displaying homology to *O. s. scutellatus* clones**	**Length of protein**	**BLAST match accession no.**	**% Sequence coverage**	**% Match**	**Notes**
R 138, 139	741	-		NBM					
Surface protein R 148	451	-		NBM					

Table 5.2: Summary of clones isolated from the *O. scutellatus* venom gland cDNA library sharing homology with characterised toxins.
Summary of complete cDNA clones isolated from *O. scutellatus* venom gland selected for analysis compared to homologous published sequences (below shaded).
Φ: undertaken using MacVector software, - not undertaken

Toxin or *O. s. scutellatus* clone no.	MW (Da)	pI	Residues	No. Cysteine residues	Comment	Reference
PLA_2/pre-synaptic						
RW26	13,520	9.37	119	14		
α-taipoxin	14,600	basic	119	14	toxic	Fohlman *et al.*, 1976
R24	13,234	5.45	119	14		
β- taipoxin	14,300	neutral	119	14	non-toxic	Fohlman *et al.*, 1976
R120	14,601	3.88	133	16		
γ- taipoxin	26,900	acidic	133	16	CHO	Fohlman *et al.*, 1976
R132	17,025	5.61	127	14		
OS1	15,900	4.73Φ	127	14	non-toxic	Lambeau, *et al.*, 1990
R56	13,313	9.36	119	14		
R125	13,371	9.36	119	14		
OS2	13,600	8.65Φ	119	14	toxic	Lambeau, *et al.*, 1990
Post-synaptic						
R147	6,776	10.10	62	8		
Toxin 1	6,726Φ	8.47Φ	62	8	toxic	Zamudio, *et al.*,1996
Toxin 2	6,781Φ	8.88Φ	62	8	toxic	Zamudio, *et al.*,1996
RW6	6,668	8.56	65	9		
OS FV	179, 062	6.15	1,460	-		
Pseutarin	-	-	1,460	-		Rao & Kini, 2003

5.2.1 Nucleotide sequences of *O. s. scutellatus*

The analysis of the cDNA nucleotide sequences, (summarised in Annex 1 of this chapter) using MacVector software was undertaken to identify eukaryotic promoter sequences. Some sequences contained a larger 5'-upstream region than others. This may have possibly been due to the lack of a secondary structure within the mRNA.

Homology searches indicated the *Oxyuranus* nucleotide sequences shared homology with the nucleotide sequences from other Elapid species. These species included *Austrelaps superbus, Notechis scutatus*, *Laticauda semifasciata*, *Pseudonaja textilis*, *Bungarus flaviceps* and *Ophiphagus hannah*. These comparisons indicated the sequences were conserved across active sites and conformed to structural restrictions described within the specific toxin groups (data not shown). Comparisons of the *Oxyuranus* nucleotide sequences sharing homology with pre- and post- synaptic

neurotoxins revealed point mutations in the first and second bases of codons (data not shown). Although further analysis of nucleotide sequences would not be pursued within this chapter, this trend confirmed the directional mutations (accelerated evolution) suggested, as seen is PLA_2 (Ogawa *et al.*, 1992), cardiotoxins (Lachumanan, *et al.*, 1998), and neurotoxins (Afifiyan *et al.*, 1999) instead of random mutations within all sequences.

Due to the low representation of nucleotide sequences within the *Oxyuranus* species, the putative peptide sequences were used for further analysis.

5.2.2 Potential peptides isolated from *O. s. scutellatus*

The analysis of the translated peptides (see Chapter 2, 2.2.2.13) indicated the sequences could be classified into distinctly different classes. Therefore, sequences were discussed within the classes to which they shared homology; pre-synaptic neurotoxins, post-synaptic neurotoxins and the Factor V component of scutelarin.

5.2.2.1 Oxyuranus sequences sharing homology to PLA_2/pre-synaptic neurotoxins

The nucleotide sequences of the cDNA clones were translated and identified using MacVector or Sequencher programs. The clones designated R24, RW26 and R120 shared homology with taipoxin subunits, whereas clone R132, R125 and R56 shared homology with OS1 and OS2. These sequences and homologies were summarised in Annex 1 at the end of this Chapter, Figures 5.10 to 5.19.

The *Oxyuranus* clones demonstrated high homology with the PLA_2 pre-synaptic neurotoxins from other elapid species, displayed in Figures 5.2 and 5.3. High sequence homology was reflected over the active and calcium-binding sites with 90 to 100% sequence identity (Figure 5.1). These peptides could also be segregated into group IA and group IB phospholipases based on these homologies. This suggests these sequences may be isomers of the characterised *Oxyuranus* peptides and evolutionary derivatives of the other homologous Elapid sequences. Interestingly, the signal peptides of these sequences were also highly conserved (Figure 5.4).

Additionally, the *O. s. scutellatus* putative peptides shared similar size, hydrophobicity and cysteine positions, thus potentially structure, to the characterised

Oxyuranus peptides to which they shared homology. The pIs were also similar. The putative protein sequences of RW26, R56 and R125 were basic. The peptide encoded by RW26 showed the highest homology with the basic α-taipoxin subunit (80%), OS2 (74%) and β-taipoxin (67%) peptide sequences. Similarly, the basic peptides R56 and R125, differed by only one amino acid, and shared homology with OS2 (93%), α-taipoxin (83%), RW26 (81.4%) and β-taipoxin (65%).

In contrast, the putative peptide sequences of R120 and R132 were acidic. These shared close alignment with the pancreatic group IB peptides, γ-taipoxin and OS1. R120 shared the closest homology with the γ-taipoxin chain (94%), and contained a propeptide, pancreatic loop and a potential carbohydrate moiety. Whereas R132, although it did not contain a propeptide or potential carbohydrate binding site, contained a pancreatic loop and differed by only one amino acid from OS1. These sequences contained larger hydrophobic areas at the N- and C-terminals and differient charges compared to the other peptides described (RW26, R125 and R56), indicating they may have different binding properties.

Regions outside the active and calcium-binding sites displayed a similarity of approximately 50% with the C-terminus (30 aa) showing 65% similarity. Despite the high homology of the PLA_2 catalytic and calcium binding domains, few residues suggested for toxicity (approx 13%, Gly^{30} and Arg^{43}) within other genera were present within the *Oxyuranus* sequences (see introduction and Table 5.3).

A preliminary pair-wise, neighbour joining alignment of the putative isolated *Oxyuranus* peptides and homologous sequences was undertaken to display the relationships between the sequences (Figure 5.5 and 5.6). Data sets of the *Oxyuranus* putative peptide sequences were bootstrapped (x 10,000) to check how robust the clusters were (Appendix VII).

Figure 5.1: Pairwise alignment of putative *O. scutellatus* peptides.

The group IA and IB (γ-taipoxin, R120, OS1, R132) peptide and putative peptide sequences of *O. scutellatus* sequences using the ClustalW program (Higgins *et al.*, 1994) on the EMBL-EBI site (http://www.ebi.ac.uk/). The conserved cysteine residues are indicated (*).
The variant amino acids are displayed as symbols denoting the degree of conservation observed in each column: "*" means that the residues in that column are identical in all sequences in the alignment. ":" means that conserved substitutions have been observed. "." means that semi-conserved substitutions are observed. Gaps (-) have been inserted for optimal alignment.
The first box outlines the PLA_2 calcium binding domain from amino acid residues 25 to 33, the second box outlines the active site from amino acid residues 42 to 52. A pancreatic loop is described within group 1B peptides (γ-taipoxin, R120, OS1, R132) from amino acid residues 62 to 66; (N/S/K) (Y/V/L/P) (L/M/Q/V/P) (M/P/L/V) (S/N/D) and a propeptide is seen at the N-terminal of some of these peptides.
The pairwise alignment used the parameters: aln w/numbers, PHYLIP tree type, distance correction off and ignore gaps off.

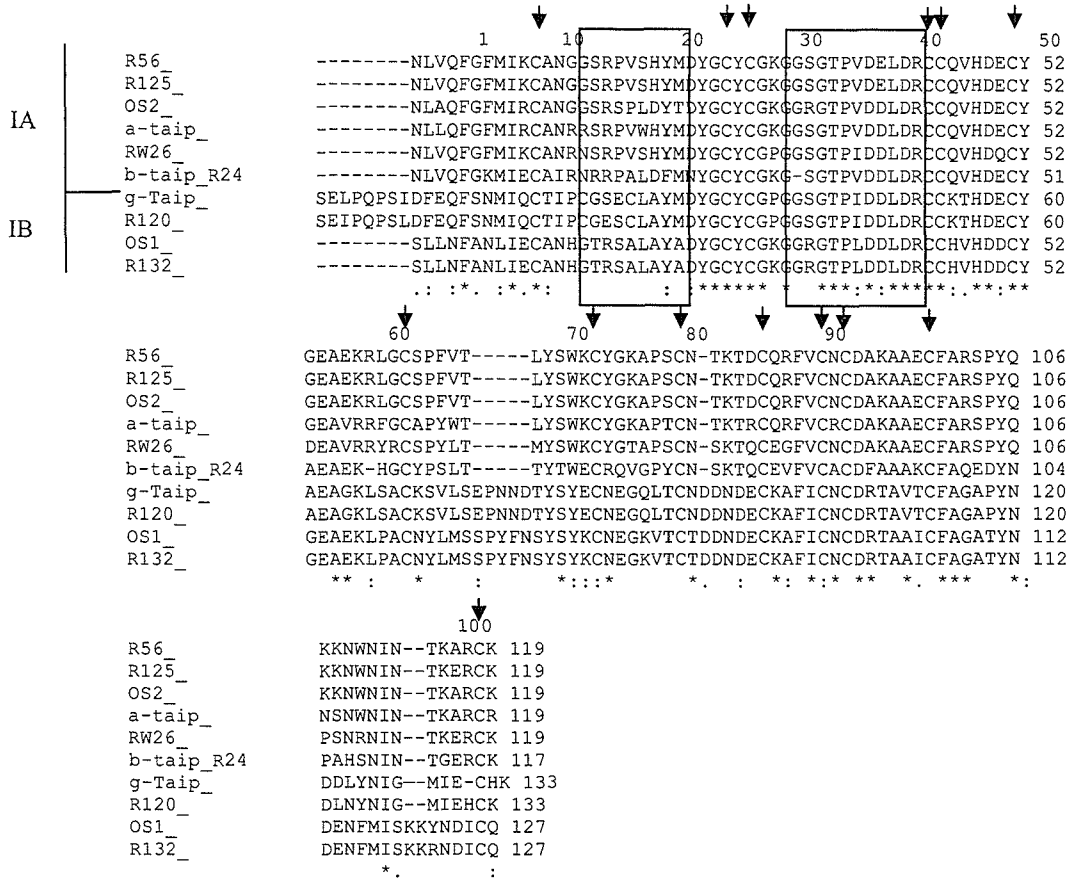


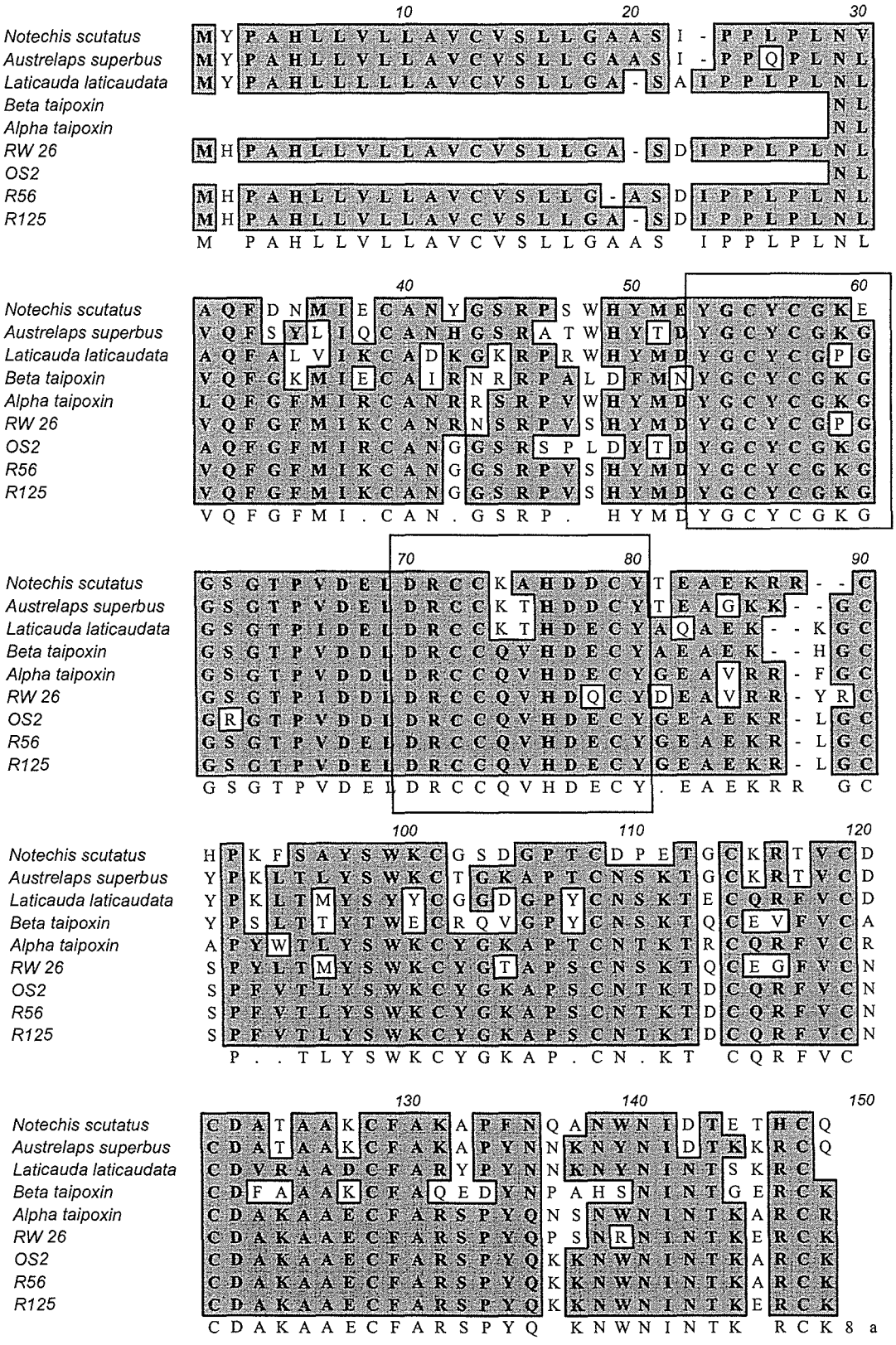
Figure 5.2: Sequence alignment of chosen group IA peptides.

Group IA *O. scutellatus* cDNA clones shared alignment with *Laticauda laticaudata* (GenBank Accession no. BAB72249), *Austrelaps superbus* (GenBank Accession no. AAD56557) and *Notechis scutatus* (GenBank Accession no. X14043).
The peptide sequence of *Oxyuranus scutellatus* α-taipoxin subunit (GenBank Accession no. POO614), β-taipoxin subunit (GenBank Accession no. POO615) and OS2 (GenBank Accession no. AAB33760) are included. The nucleotide and signal peptide of these sequences are unknown at this time.
Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text. Gaps (-) have been inserted for optimal alignment.
The first box outlines the PLA_2 calcium binding domain from amino acid residues, the second box outlines the active site.
Page 132

Figure 5.3: Sequence alignment of chosen group IB peptides.

Group IB *O. scutellatus* cDNA clones shared alignment with *Notechis scutatus* (GenBank Accession no. AAB34122), *Lapemis hardwicki* (GenBank Accession no. AAL55555), *Pseudonaja textilis* (GenBank Accession no. AF144319), *Laticauda semifasciata* (GenBank Accession no. ABO78348) and *Austrelaps superbus* (GenBank Accession no. AAD56410), These clones shared the highest homology with the *Oxyuranus scutellatus* δ-taipoxin subunit (GenBank Accession no. POO616), and OS1 (GenBank Accession no. AAB33759). The nucleotide and signal peptides of these sequences are unknown.
Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text. Gaps (-) have been inserted for optimal alignment. The first box outlines the PLA_2 calcium binding domain from amino acid residues, the second box outlines the active site.
Page 133

Figure 5.2: Clustal alignment of chosen peptides within group IA PLA_2.



Formatted Alignments

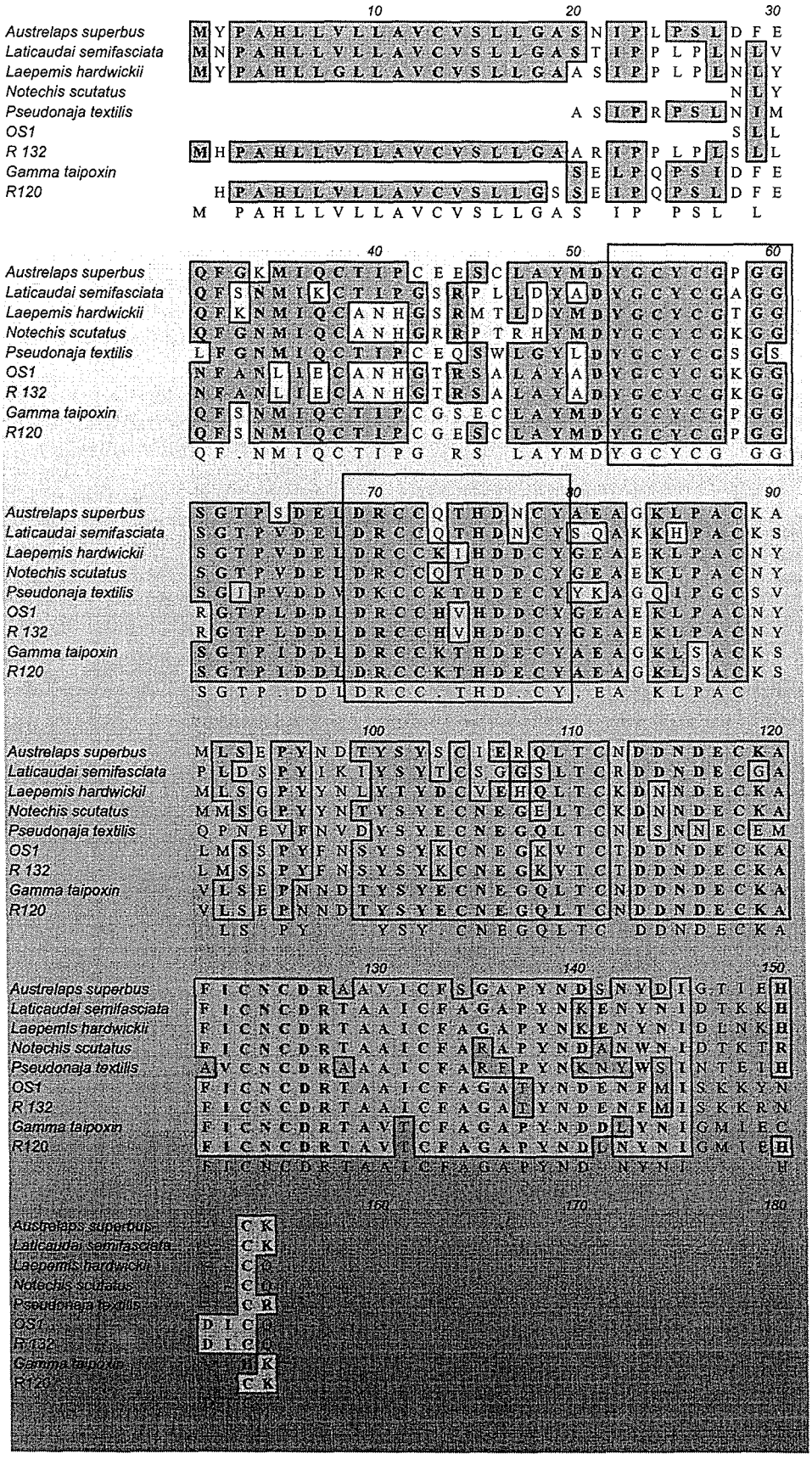


Figure 5.4: Alignment of the signal peptides isolated from an *O. s. scutellatus* cDNA library and the homology with other Elapid species.

A) Pre-synaptic neurotoxins

i) Pre-synaptic neurotoxin signal peptides from clones isolated

R24(β):	MHPAHLLVLLAVCVSLLGASDIPPLPLNLVQ
RW26,(α)	MHPAHLLVLLAVCVSLLGASDIPPLPLNLVQ
R 125(OS2)	MHPAHLLVLLAVCVSLLGASDIPPLPLN
R 132 (OS1)	MHPAHLLVLLAVCVSLLGAARIPPLPLSLLN
R 120 (γ)	MHPAHLLVLLAVCVSLLGSSEIPQPSLDFEQF

i) Homology of pre-synaptic neurotoxin signal peptides

O. scutellatus R120 (γ-taipoxin)	M	HPAHL	LVLLA	VCVSL	LG - S	
O. scutellatus (α,β, OS2)	M	HPAHL	LVLLA	VCVSL	LGASD	IPPLPL
O. scutellatus R132 (OS1)	M	HPAHL	LVLLA	VCVSL	LGAAR	IPPLPL
B. multicinctus (Kondo, 1982)	M	YPAHL	LVLSA	VCVSL	LGAAN	IPPYPL
A. Laevis (Ducance, 1988)	M	YPAHL	LVLLA	VCVSL	LGASD	IPPLPL
L. semifasciata (Fujimi, 2000)	M	NPAHL	LVLLA	VCVSL	LGASA	IPPLPL
A. superbus (Singh, 2000)	M	YPAHL	LVLLA	VCVSL	LGASD	IPPQPL
	-27	-21	-16	-11	-6	-1

Table 5.3: Annotation of suggested toxin residues within Elapidae pre-synaptic neurotoxins.

Annotation of toxin residues from *O. scutellatus* pre-synaptic and putative neurotoxins compared to the suggested residues necessary for toxicity within Elapidae.
The numbering incorporated within this table does not include the residues encompassing the pancreatic loop of the group 1B peptides, as the many 'toxic' residues are numbered using the group 1A (toxic) peptides that do not include a pancreatic loop. Areas of high homology are coloured in blue. After residue 55 the numbering becomes difficult due to the numerous insertions and deletions

Sites suggested causing activity/ toxicity within pre-synaptic neurotoxins	α taip Toxic	RW26 (α)	β taip Non-toxic	γ taip Non-toxic	R120 (γ)	OS1 Non-toxic	R132 (OS1)	OS2 Toxic	R56 (OS2)	R125 (OS2)	Notes
30 G	30 G	30 G	30 G	30 G	30 G	30 G	30 G	30 G	30 G	30 G	
31 W	31 K	31 P	31 K	31 P	31 P	31 K	31 K	31 K	31 K	31 K	
43 R	43 R	43 R	43 R	43 R	43 R	43 R	43 R	43 R	43 R	43 R	
46 K	46 Q	46 Q	46 Q	46 K	46 K	46 H	46 H	46 Q	46 Q	46 Q	
50 D	50 E	50 Q	50 E	50 E	50 E	50D	50 D	50 E	50 E	50 E	
54 G	54 E	54 E	54 E	54 E	54 E	54 E	54 E	54 E	54 E	54 E	
57	57 R	57 R	57 K	57 K	57 K	57 K	57 K	57 K	57 K	57 K	
58K	58 R	58 R	58 H	58 L	58 L	58 L	58 L	58 R	58 R	58 R	
63 Wor F	63 P	63 P	63 P	63 P	63 P	63 P	63 P	63 P	63 P	63 P	
64 Y	64 Y	64 Y	64 S	64 V	64 V	64 L	64 L	64 F	64 F	64 F	
69 W	69 S	69 V	69 W	69 N	69 N	69 Y	69 Y	69 S	69 S	69 S	All toxic peptides contain W @ 70
70	70 W	70 W	70 Y	70 Y	70 Y	70 Y	70 W	70W	70 W	70 W	
74 N	74 Q	74 G	74 G	74 E	74 E	74 E	74 E	74 G	74 G	74 G	No N in vicinity
90 D	90 C	90 C	90 C	90 C	90 C	90 C	90 C	90 C	90 C	90 C	All D at 93
94 G	94 A	94 A	94 A	94 V	94 V	94 A	94 A	94 A	94 A	94 A	Used residue 97

Figure 5.5; Alignment of *O. s. scutellatus* pre-synaptic PLA2s and putative *O. s. scutellatus* amino acid sequences.

The predicted protein sequences of the mature, processed peptides were aligned by using Clustal. The phylogenetic relationships of the proteases were assessed by the neighbour joining method using the ClustalW program (http://www.ebi.ac.uk, Higgins, 1994). Trees were generated and robustness was assessed by bootstrap analysis using 10,000 replicates; clades with more than 80% support are depicted. Actual bootstrap phylogeny and values can be viewed in Appendix VII. The branch lengths shown to the right of the peptide labels are proportional to the amount of inferred evolutionary change.
This clustering was mimicked when additional elapid PLA_2 sequences were included within the analysis (Figure 5.6).

Figure 5.5 indicated the Type IB (γ-taipoxin, OS1 R120, R132) and IA peptides evolved differently, with type IB *Oxyuranus* clones and subsequence putative peptides an older lineage.

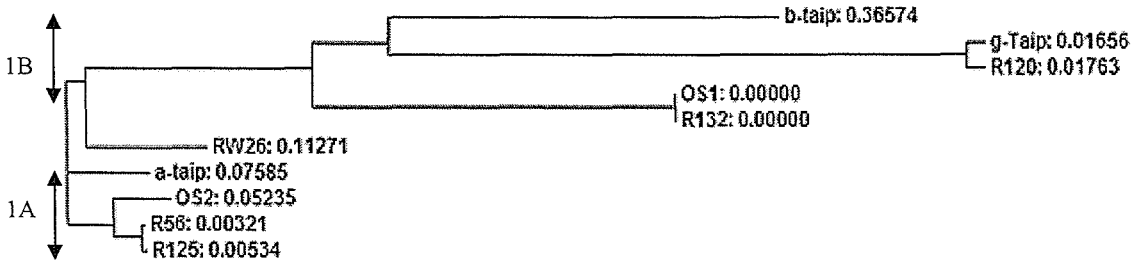
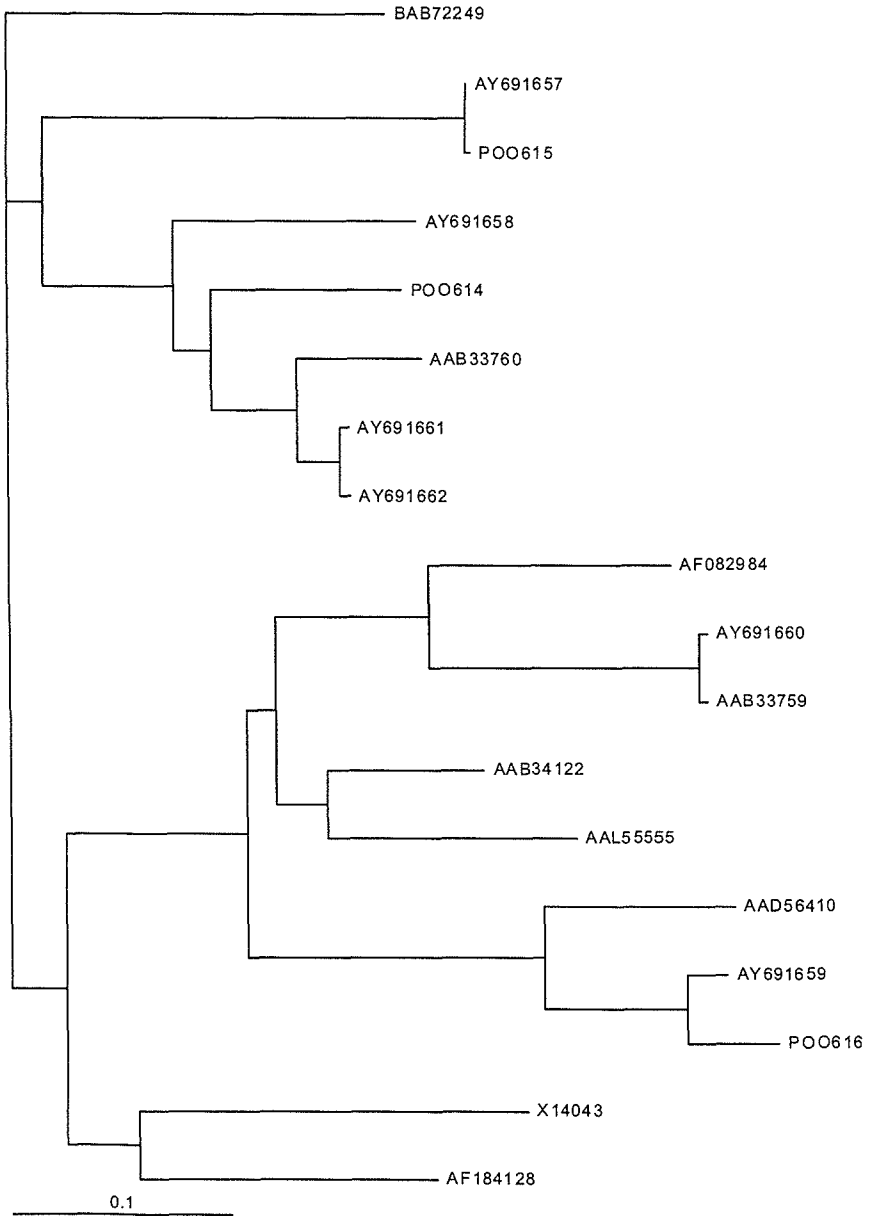


Figure 5.6 Alignment of Elapid pre-synaptic PLA_2s and putative amino acid sequences

The phylogram below is assumed to be an estimate of phylogeny for the aligned elapid PLA_2 sequences. The analysis was undertaken using the parameters in Figure 5.5.

The toxic type IA peptides clustered predictively; *O. s. scutellatus* α-taipoxin (GenBank accession no. POO614) and putative peptide RW26 (GenBank accession no. AY691658) grouped together next to *O. s. scutellatus* OS2 (GenBank accession no. AAB33760), R 125 (GenBank accession no. AY691661, similar to OS2) and R56 (GenBank accession no. AY691662, similar to OS2). *O. s. scutellatus* β-taipoxin (GenBank accession no. POO615) and R24 (GenBank accession no. AY691657, similar to β taipoxin) were outliers within the Type IA clustered group, differing to the analysis within Figure 5.5. *Laticauda laticaudata* (GenBank accession no. BAB72249 appeared to bean outlier of these groups.

Pseudonaja textilis (GenBank accession no. AF082984) shared homology with the type IB R132 (GenBank accession no. AY691660, similar to OS1), *O. s. scutellatus* OS1 (GenBank accession no. AAB33759) and *Lapemis hardwickii* (GenBank accession no. AAL55555). Surprisingly, *Notechis scutatus* (GenBank accession no. AAB34122, toxic PLA_2) was included within this group. *Austrelaps superbus* (GenBank accession no. AAD56410) clustered with R120 (GenBank accession no. AY691659, similar to γ-taipoxin) and *O. s. scutellatus* γ-taipoxin (GenBank accession no. POO616). *Notechis scutatus scutatus* (GenBank accession no. X14043), and *Austrelaps superbus* (GenBank accession no. AF184128) were outliers within this group.



5.2.2.2 Oxyuranus sequences sharing homology to post-synaptic neurotoxins

Short chain post-synaptic neurotoxins

Similar to the sequences matching the pre-synaptic neurotoxin sequences, the translated peptide from the nucleotide sequences of the R147, R121 and RW6 showed homology with the peptide sequences of post-synaptic short chain neurotoxins.

R147 and R121 appeared to encode the same protein at the 5'-end of the nucleotide sequence. The 485 bp sequence of R147 included a 249 bp ORF that translated into an 83 amino acid peptide including a 21 amino acid signal peptide (Annex 1, Figure 5.20). R121 consisted of 369 nucleotides of which the initial 249 base pairs were open reading frame. This sequence was followed by a 3'-untranslated sequence of 114 nucleotides (Annex 1 Figure 5.21) and did not contain a poly A tail, seen in R147.

When translated, R121 diverged from R147 by 20 amino acid residues from the first TAG stop codon (Figure 5.22). The molecular mass and pI of the peptide encoded by R121 and R147 was 6,776 and 10.1 respectively and contained a 21 residue signal peptide. Two possible N-glycosylation sites were seen at 180 bp (NRT) and 211 bp (NFS). The algorithm of Kyte and Doolittle (Kyte, 1982) of R121/R147 showed this peptide was an overall hydrophilic peptide with three short hydrophobic regions within the protein.

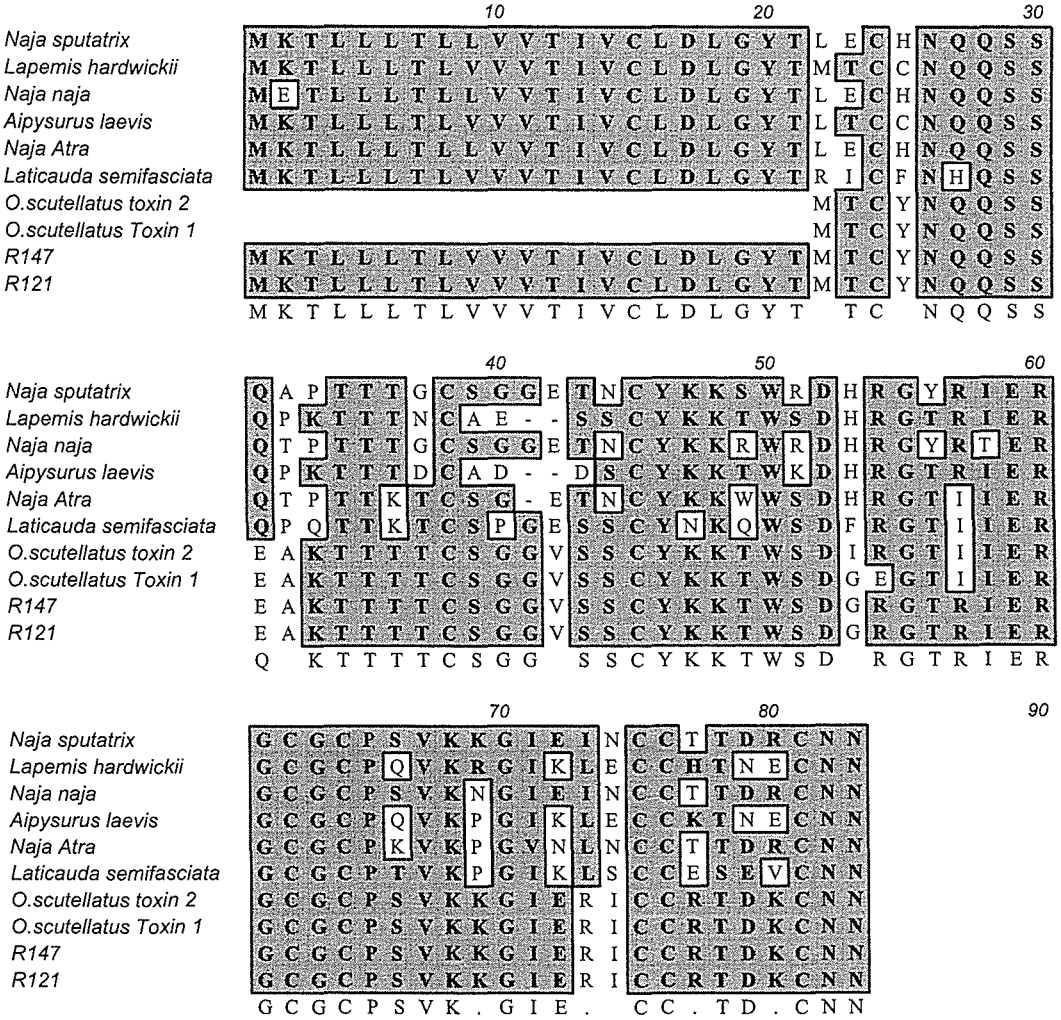
Comparisons of the amino acid sequences of cDNAs encoding short chain neurotoxins of *O. s. scutellatus* showed R147 shared sequence homology with *O. scutellatus* toxin 1 (91% homology) and toxin 2 (90%), *Naja sputatrix*, 63% (GenBank Accession no. Q9YGJ5) and 71% (GenBank Accession no. AAD08812) and *Laticauda semifasciata* erabutoxin a (56%), b (56%), and c (56%) (Figure 5.7). These sequences shared identical cysteine positions and homology to all residues required for toxicity and binding (Tyr^{25}, Lys^{27}, Trp^{29}, Asp^{31}, Cys^{41}, Gly^{42}, Cys^{43}, Pro^{44} and Gly^{49}).

These peptide and putative peptide sequences also appear to conform to a similar structure. This is inferred from the similar pI, hydrophobicity and cysteine residues identified for each of the sequences (data not shown).

Figure 5.7: Alignment of R147 and R121 with post-synaptic neurotoxins.

Putative protein R147 and R121 from cDNA clones isolated from *O. s. scutellatus* were aligned with *Naja sputatrix* (GenBank Accession no. AAD08813), *Lapemis hardwickii* (Novel neurotoxin precursor, GenBank Accession no. AAL54894), *Aipysurus laevis* (B34019), *Naja atra* (cobrotoxin precursor, P80958, *O. s. scutellatus* (toxin 1, acc), *O. s. scutellatus* (Toxin 2, Zamudio, *et al.*, 1996), *Laticauda semifasciata* (erabutoxin 1A, Zamudio, *et al.*, 1996) and *Naja naja* preprotein (GenBank Accession no. AAF21774) sequences. Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text.
Gaps (-) have been inserted for optimal alignment.

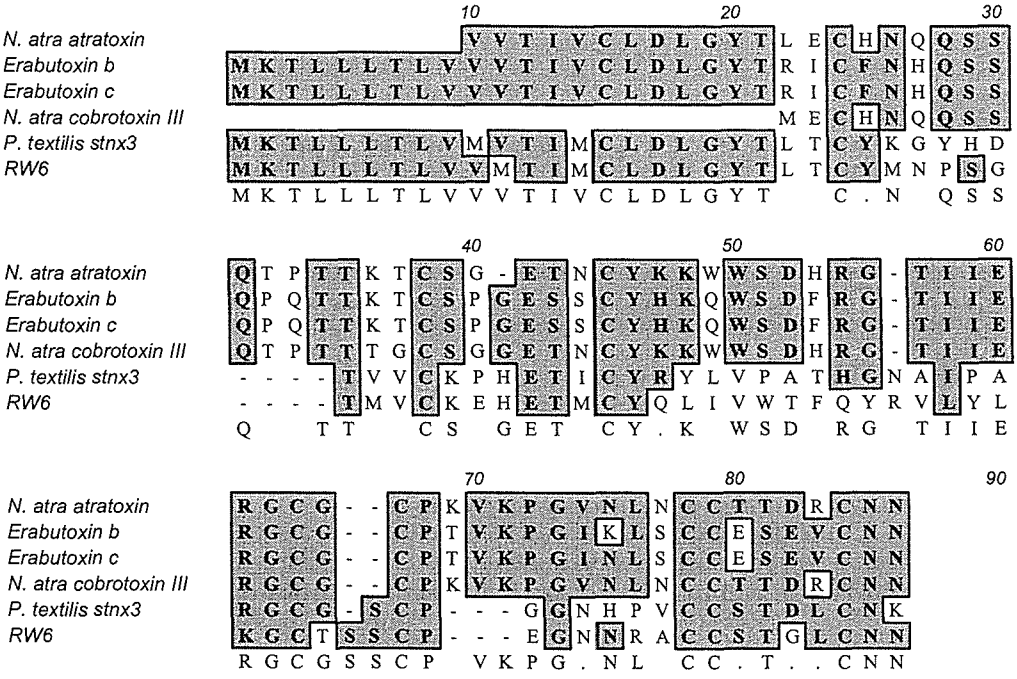
Formatted Alignments



The final clone matching a neurotoxin, was RW6 and was the only sequence to be isolated using PCR. PCR products were synthesised using the *O. s. scutellatus* cDNA library as a template and primers originally produced to isolate a 300 kDa peptide in Chapter 4. PCR using primers (MS4) 5'-GCNGNTTTYTTYCAR-3' and the T7 primer (forward and reverse primer respectively) resulted in a 300 bp product which was subsequently subcloned into pGEMT. The full sequence of this clone identified a 350 nucleotide sequence that included an ORF of 234 base pairs. This sequence translated into a 79 amino acid residue protein including a 21 residue signal peptide (Figure 5.23). The ORF in started at nucleotide 328.

Figure 5.8: Alignment of RW6 with post-synaptic neurotoxins.

RW6 isolated by PCR from the *O. scutellatus* cDNA library shared homology with sequences from *Pseudonaja textilis*, stnx3 (92%, AF204971), *Naja atra*, atratoxin b (94%, AY471579), *Naja atra*, cobrotoxin III (93%, AF088998) and *Laticauda semifasciata*, erabutoxin genes b (92%, X16950) and c (92%, X51410). Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text. Gaps (-) have been inserted for optimal alignment.



The molecular mass and pI of the deduced protein from RW6 after GYT was 6,668 and 8.56 respectively with a potential N-glycosylation site at amino acid residue 27. Contrasting to R147/R121, translated RW6 was markedly more hydrophobic. The algorithm of Kyte and Doolittle (Kyte, 1982) demonstrated this protein contained a hydrophobic mid region and hydrophilic N- and C-terminal regions. The putative peptide sequence aligned with a group of post-synaptic neurotoxins not characterised within *Oxyuranus*. The sequence shared homology with the characterised post-synaptic neurotoxins of *Pseudonaja textilis* 92% (GenBank Accession no. AF204971) and 92% (GenBank Accession no. AF204970), *Naja atra*, atratoxin b 94% (GenBank Accession no. AY471579), cobratoxin III 93% (GenBank Accession no. AF088998) and *Laticauda colubrina* 93% (GenBank Accession no. AB17937)(Figure 5.8). While the cysteine residues were highly conserved, the sequence homology of suggested toxic and binding sites were not. Only residues Tyr, Pro, Gly and Cys were conserved, though due to the alignment of sequences the numbering was different to that discussed within the Introduction.

5.2.2.3 *Oxyuranus sequences sharing homology to Factor V*

Expressed proteins from the *O. scutellatus* venom gland cDNA library were screened with taipan monovalent antivenom (CSL). Positive clones were isolated and the nucleotide sequences of *in vivo*-excised plasmids determined (Annex 1, Figure 5.24, pg 164). Analysis of a number of nucleotide sequences revealed similarities with Factor V (FV) sequences from *Homo sapiens* (GenBank Accession number NP000121) and *Bos taurus* (GenBank Accession number AAA30513). This clone was the most abundant isolated and represented 13% of the total clones characterised from the cDNA library. Thus far, the only described Factor V component from *O. s. scutellatus* venom was incorporated in the prothrombin activator, scutelarin, described in Chapter 4.

However, this sequence did not contain a full open reading frame of FV. It was not until Rao (2002) published the sequence of the non-enzymic subunit from the *P. textilis* prothrombin activator (pseutarin C)(Rao, *et al.*, 2002) that an open reading frame was detected. Closer examination of the taipan sequence revealed an additional A at

nucleotide 3,020 when compared to Pseutarin C FV. Removal of this additional 'A' resulted in a full ORF. It is still not known if this was a cloning artefact as duplicate clones contained this 'double A'. The complete 4,690 nucleotide sequence, minus the additional A, resulted in 4,380 bp of ORF which translated into a 1,460 amino acid residue protein including a 30-residue signal peptide. This ORF was preceded by a 5'-untranslated region of 284 nucleotides and contains a polyadenylation signal 19 bp upstream of the poly $(A)^+$ tail (Annex, Figure 5.24).

The molecular mass and pI of the deduced Factor V was 179,062 and 6.15, respectively. The algorithm of Kyte and Doolittle (Kyte and Doolittle, 1982) showed this potential FV contained both hydrophilic and hydrophobic regions alternating throughout the protein. The deduced amino acid sequence shared identity with the non enzymic subunit of the *P. textilis* prothrombin activator, Pseutarin C (95%)(GenBank Accession no. AAO38805)(Figure 5.9). The homology between the species and the match of this to the N-terminal sequence QLREYR (B1) from chapter 3 (B1 was a glycosylated 110-120 kDa protein band from whole venom assumed to be the FV subunit from the *Oxyuranus* prothrombin complex), confirmed this cDNA clone was the FV subunit from scutelarin. These two sequences maintained a highly conserved structure, size and charge.

This non-enzymic subunit of scutelarin showed identical domain architecture to mammalian FV consisting of domains A1-A2-B-A3-C1-C2. Similar to the non enzymic subunit of pseutarin C, the B domain was markedly truncated compared to mammalian FV enzyme. The *O. s. scutellatus* and *P. textilis* sequences maintained a highly conserved structure, size and charge with cleavage sites conserved across the two species. The predicted sites across the domains, including Cu oxidase domains, PLA_2 binding areas, spacer and repeat regions, and post-translational modifications such as phosphorylation and sulphation and disulphide bonds were identical. Importantly, the activated protein C (APC) proteolytic cleavage sites were conserved, with two of the three proteolytic sites cleaved by APC mutated.

Twelve potential N-glycosylation sites were detected within the *O. s. scutellatus* sequence using MacVector. The potential sites at amino acid residues 173, 259, 317, 422, 487, 573, 588, 959, 1016, 1195, 1405 and one tyrosine kinase at residue 60

(EDWDY) mimicked that of *P. textilis* FV except for an extra N-glycosylation site at 1502. This potential N-glycosylation site is not present in the *P. textilis* FV sequence, although it is described within bovine FV.

Most of the proteolytic activation sites, when compared to the *P. textilis* FV sequence were conserved. Of interest was two of the three proteolytic sites cleaved by activated protein C (arginine) were mutated.

Figure 5.9: ClustalW alignment of *Oxyuranus* FV and pseutarin (*Pseudonaja*).

The deduced amino acid sequences showed identity to a Factor V toxin characterised from *Pseudonaja textilis*, pseutarin C (GenBank Accession no. AAO38805). Amino acids sharing similarities are shaded. The pairwise alignment was undertaken using the EMBL-EBI site (http://www.ebi.ac.uk/) using the ClustalW program. APC cleavage sites are listed in red. Two of the three APC cleavage sites are absent. The position where the arginine is cleaved by APC is underlined.

```
Pseutarin                                MGRYSVSPVPKCLLLMFLGWSGLKYYQVNAAQLREYHIAAQLE  43
Oxyuranus                                MGRYSVSPVPKCLLLMFLGWSGLKYYQVNAAQLREYRIAAQLE  43

Pseutarin        DWDYNPQPEELSRLSESDLTFKKIVYREYELDFKQEEPRDALSGLLGPTLRGEVGDSLII 103
Oxyuranus        DWDYNPQPEELSRLSESDLTFKKIVYREYELDFKQEKPRDPLSGLLGPTLRGEVGDSLII 103

Pseutarin        YFKNFATQPVSIHPQSAVYNKWSEGSSYSDGTSDVERLDDAVPPGQSFKYVWNITAEIGP 163
Oxyuranus        YFKNFATQPVSIHPQSAVYNKWSEGSSYSDGTSDVERLDDAVPPGQSFKYVWNITAEIGP 163

Pseutarin        KKADPPCLTYAYYSHVNMVRDFNSGLIGALLICKEGSLNANGSQKFFNREYVLMFSVFDE 223
Oxyuranus        KKADPPCLTYAYYSHVNMVRDFNSGLIGALLICKEGSLNADGAQKFFNREYVLMFSVFDE 223

Pseutarin        SKNWYRKPSLQYTINGFANGTLPDVQACAYDHISWHLIGMSSSPEIFSVHFNGQTLEQNH 283
Oxyuranus        SKNWYRKPSLQYTINGFANGTLPDVQACAYDHISWHLIGMSSSPEIFSVHFNGQTLEQNH 283

Pseutarin        YKVSTINLVGGASVTADMSVSRTGKWLISSLVAKHLQAGMYGYLNIKDCGNPDTLTRKLS 343
Oxyuranus        YKVSTINLVGGASVTANMSVSRTGKWLISSLVAKHLQAGMYGYLNIKDCGNPDTLTRKLS 343

Pseutarin        FRELMKIKNWEYFIAAEEITWDYAPEIPSSVDRRYKAQYLDNFSNFIGKKYKKAVFRQYE 403
Oxyuranus        FREWRRIMKWEYFIAAEEITWDYAPEIPSSVDRRYKAQYLD-FSNFIGKKYKKAVFRQYE 402

Pseutarin        DCNFTKPTYAIWPKERGILGPVIKAKVRDTVTIVFKNLASRPYSIYVHGVSVSKDAEGAI 463
Oxyuranus        DSNFTKPTYAIWPKERGILGPVIKAKVRDTVTIVFKNLASRPYSIYVHGVSVSKDAEGAV 462

Pseutarin        YPSDPKENITHGKAVEPGQVYTYKWTVLDTDEPTVKDSECITKLYHSAVDMTRDIASGLI 523
Oxyuranus        YPSDPKENITHGKAVEPGQVYTYKWTVLDTDEPTVKDSECITKLYHSAVDMTRDIASGLI 522

Pseutarin        GPLLVCKHKALSVKGVQNKADVEQHAVFAVFDENKSWYLEDNIKKYCSNPSAVKKDDPKF 583
Oxyuranus        GPLLVCKRKALSIRGVQNKADVEQHAVFAVFDENKSWYLEDNIKKYCSNPSSVKKDDPKF 582

Pseutarin        YKSNVMYTLNGYASDRTEVLRFHQSEVVQWHLTSVGTVDEIVPVHLSGHTFLSKGKHQDI 643
Oxyuranus        YKSNVMYTLNGYASDRTEVWGFHQSEVVEWHLTSVGTVDEIVPVHLSGHTFLSKGKHQDI 642

Pseutarin        LNLFPMSGESATVTMDNLGTWLLSSWGSCEMSNGMRLRFLDANYDDEDEGNEEEEEDDGD 703
Oxyuranus        LNLFPMSGESATVTMDNLGTWLLSSWGSCEMSNGMRLRFLDANYDDEDEGNEEEEEDDGD 702

Pseutarin        IFADIFIPSEVVKKKEEVPVNFVPDPESDALAKELGLIDDEGNPIIQPRREQTEDDEEQL 763
Oxyuranus        IFADIFNPPEVVIKKEEVPVNFVPDPESDALAKELGLFDDEDNPK-QSRSEQTEDDEEQL 762

Pseutarin        MKASMLGLRSFKGSVAEEELKHTALALEEDAHASDPRIDSNSARNPDDIAGRYLRTINRG 823
Oxyuranus        MIASMLGLRSFKGSVAEEELKHTALALEEDAHASDPRIDSNSAHNSDDIAGRYLRTIYRR 822

Pseutarin        NKRRYYIAAEEVLWDYSPIGKSQVRSRAAKTTFKKAIFRSYLDDTFQTPSTGGEYEKHLG 883
Oxyuranus        NKRRYYIAAEEVLWDYSPIGKSQVRSLPAKTTFKKAIFRSYLDDTFQTPSTGGEYEKHLG 882
```

```
Pseutarin     ILGPIIRAEVDDVIEIQFKNLASRPYSLHAHGLLYEKSSEGRSYDDKSPELFKKDDAIMP 943
Oxyuranus     ILGPIIRAEVDDVIEVQFRNLASRPYSLHAHGLLYEKSSEGRSYDDNSPELFKKDDAIMP 942

Pseutarin     NGTYTYVWQVPPRSGPTDNTEKCKSWAYYSGVNPEKDIHSGLIGPILICQKGMIDKYNRT 1003
Oxyuranus     NGTYTYVWQVPPRSGPTDNTEKCKSWAYYSGVNPEKDIHSGLIGPILICQKGMIDKYNRT 1002

Pseutarin     IDIREFVLFFMVFDEEKSWYFPKSDKSTCEEKLIGVQSLHTFPAINGIPYQLQGLTMYKD 1063
Oxyuranus     IDIREFVLFFMVFDEEKSWYFPKSDKSTCEEKLIGVQSRHTFPAINGIPYQLQGLMMYKD 1062

Pseutarin     ENVHWHLLNMGGPKDIHVVNFHGQTFTEEGREDNQLGVLPLLPGTFASIKMKPSKIGTWL 1123
Oxyuranus     ENVHWHLLNMGGPKDVHVVNFHGQTFTEEGREDNQLGVLPLLPGTFASIKMKPSKIGTWL 1122

Pseutarin     LETEVGENQERGMQALFTVIDKDCKLPMGLASGIIQDSQISASGHVGYWEPKLARLNNTG 1183
Oxyuranus     LETEVGENQERGMQALFTVIDKDCKLPMGLASGIIQDSQISASGHVGYWEPKLARLNNTG 1182

Pseutarin     KYNAWSIIKKEHEHPWIQIDLQRQVVITGIQTQGTVQLLQHSYTVEYFVTYSEDGQNWIT 1243
Oxyuranus     MFNAWSIIKKEHEHPWIQIDLQRQVVITGIQTQGTVHLLKHSYTVEYFVTYSKDGQNWIT 1242

Pseutarin     FKGRHSETQMHFEGNSDGTTVKENHIDPPIIARYIRLHPTKFYNRPTFRIELLGCEVEGC 1303
Oxyuranus     FKGRHSKTQMHFEGNSDGTTVKENHIDPPIIARYIRLHPTKFYNTPTFRIELLGCEVEGC 1302

Pseutarin     SVPLGMESGAIKNSEITASSYKKTWWSSWEPSLARLNLEGGTNAWQPEVNNKDQWLQIDL 1363
Oxyuranus     SVPLGMESGAIKDSEITASSYKKTWWSSWEPFLARLNLKGRTNAWQPKVNNKDQWLQIDL 1362

Pseutarin     QHLTKITSIITQGATSMTTSMYVKTFSIHYTDDNSTWKPYLDVRTSMEKVFTGNINSDGH 1423
Oxyuranus     QHLTKITSIITQGATSMTTSMYVKTFSIHYTDDNSTWKPYLDVRTSMEKVFTGNINSDGH 1422

Pseutarin     VKHFFKPPILSRFIRIIPKTWNQYIALRIELFGCEVF                         1460
Oxyuranus     VKHFFNPPILSRFIRIIPKTWNQYIALRIELFGCEVFN                        1460
```

5.3 DISCUSSION

There are a growing number of nucleotide and peptide sequences reported thus far within elapid species, to which the sequences described within this report can be added. Large variations of sequences were represented within the venom, which raises more questions than answers.

A large number of sequences isolated were composed of protein disulphide isomerases (PDI). In light of the number of cysteines described within the abundant phospholipase group this should not be surprising, as they most likely fold these amino acids, with many disulphide bridges, into functional peptides. Yet the presence of the antigenic PDI's leads to an inquiry regarding the expression, role and transport of the venom peptides. The observation of the PDI's containing an apparent endoplasmic reticulum retention/retrieval tetra peptide at the C-terminal (KDEL) (Appendix VIII) indicates the PDI's may not contain a venomous role (as displayed with a Factor V subunit), despite the antigenicity displayed. Future investigation may provide an answer to this.

The role of other sequences present within the venom remain more subjective, their presence or function subject to speculation. Though an in-depth analysis of all sequences isolated was not within the scope of this thesis, the sequences submitted to the GenBank database will extend a working knowledge of venom proteins. Through access to an increasing number of sequence data across species, genus and even families or ongoing expression studies it will be possible to map specific evolutionary traits or gain an understanding of venom mechanisms in the future.

Snake venoms have also exhibited a diversity of secreted PLA_2 and three finger toxins that are part of a large family of naturally occurring, structurally related enzymes. Many species have varied populations of these distinct peptides (Dennis, 1994; Takasaki, *et al.*, 1990). The level of identity between these enzymes can range from 40% to 99% and the *Oxyuranus* species sequences were no exception. The peptide sequences showed a strong correlation with major venom proteins previously described within the Elapid genus, and specifically within *Oxyuranus*.

5.3.1 Pre-synaptic neurotoxins

A group of sequences isolated through protein expression, shared sequence homology, and resembled a group of PLA_2 peptides. The sequences displayed a conservation of size, PLA_2 active and binding sites and cysteine residues, inferring a conservation of structure. This may indicate a fundamental structure is necessary for their mechanism of action, such as requiring a specific conformation to bind to a receptor site. These PLA_2 sequences, in turn, shared homology with *O. s. scutellatus* and other elapid pre-synaptic neurotoxins. The pre-synaptic neurotoxins reported appeared to be closely related, which is common for the group, although sequence homology is not necessarily indicative of their enzymic activities. The hydrophilic properties of these peptides are consistent with soluble peptides such as peptide hormones (insulin, gastrin) or biogenic amines (serotonin, melatonin). Therefore it may be assumed these peptides cannot diffuse through plasma membranes but enter cells through receptors.

These sequences were grouped into type IA or IB based upon their sequence homology, pI, hydrophobicity, size and cysteine residues, and will be discussed accordingly.

5.3.1.1 Type IA peptides (RW26, R125, R56)

Sequence comparisons of the basic and hydrophilic *O. scutellatus* putative peptide sequences aligned with peptide sequences exhibiting toxicity. These sequences aligned with the active and calcium binding sites of the PLA_2 from other species, and showed some correlation with suggested binding sites (Phe^5, Lys^7, Ile^9, Lys^{10}, Gly^{30}).

Variations of the 'toxic' residues between the *Oxyuranus* sequences and other elapids species were apparent. To establish potential toxic sites within *Oxyuranus* peptides, and due to the variation of 'toxic' residues when compared to other species, the peptide sequences from *O. scutellatus* were compared. This comparison showed specific amino residue differences between the toxic (type 1A) and non-toxic (type 1B) sequences (at positions 6, 7, 19, 14, 16, 19, 21, 46, 38, 59, 60, 70, 71, 72, 75, 78, 79, 80, 81, 82, 83, 86, 87, 88, 92, 93, 95, 100, 101, 104, 108, 109, 112 and 113; the region of the pancreatic loop included in the residue numbering, Figure 5.1). Of interest from these sites were the residues that differed between the type 1A peptides; OS2 and the α-taipoxin chain (sharing 82% homology). The residue variations between these sequences may be responsible for the differences in toxicity described between the two peptide chains; OS2 is reported to have an LD_{100} of 17.5 μg/kg[1] whereas the isolated α-taipoxin subunit has a lower toxicity with an LD_{50} of 300 μg/kg. Distinctions between these two chains may indicate specific residues utilised for the binding and toxicity of *Oxyuranus* pre-synaptic neurotoxins. The residues varying between the two chains were (taipoxin-residue position-OS2) R-14-G, V-19-P, H-21-D, F-59-L, W-70-V, T-83-S, N-113-K respectively. α-taipoxin appeared to retain a higher number of aromatic residues (taipoxin-residue position-OS2, W20L, F59L, Y69F) compared to OS2, and both contained a similar number of hydrophobic residues within their sequences. As both aromatic and hydrophobic residues are reported to be extremely important for binding and toxicity, the differences between these two peptides does not explain their different toxicities.

[1] Minimal lethal dose (LD_{100}) by intracisternal injection (5μl) into mice (20g body weight). LD_{100} was the lowest quantity of toxin that killed the mice (n=3) (Lambeau, *et al.*, 1989)

Due to this, it is possible the charges of the residues may be attributed to the toxicity and binding of these peptides. The differences between OS2 and α-taipoxin also did not resolve this.

Through sequence comparisons, the lack of identified residues to potentially cause toxicity suggests that specific sites are not responsible for toxicity, but that a combination of cited residues and properties affects the toxicity of a peptide. Site-directed mutagenesis studies with α-taipoxin would help resolve this question.

5.3.1.2 Neutral peptide

The deduced amino acid sequence of β-taipoxin (R24) was the only cDNA sequence to show 100% homology with a characterised *Oxyuranus* protein. This nucleotide sequence represented 30% of the total number of PLA_2s isolated.

Fohlman (1976) reported taipoxin as a trimer consisting of equal numbers of subunits and it has been reported that two isoforms of the β-subunit from taipoxin (β-1, β-2) can be separated by chromatography (Lipps, 2000). To date, only one β-subunit peptide sequence has been published, to which the nucleotide sequence and signal peptide can be added from this work. It may be possible the isoform β-2 contains a varying sequence, instead of possible post-translational modifications. The presence of this sequence and its relationship with taipoxin indicates isomers are present within this peptide group, reflecting the homologous sequences described within this thesis.

In fact, isoforms of taipoxin sequences may have been previously demonstrated though not reported. Lipps (2000) displayed differing chromatography fractions reported to contain α- and β-subunits. Curiously, differing fractions of β-subunits were identified as isoforms, but differing chromatography fractions of α-subunits did not receive any documented attention. One may speculate that these patterns may represent isomers of the α-subunit.

5.3.1.3 Type IB peptides (R120, R132)

These sequences also maintained high homology between species. The type IB peptides were observed overall to be acidic and more hydrophobic than the group IA

peptide sequences. These acidic toxins were also so closely related, that it is likely one sequence was a derivative of the other.

The γ-taipoxin chain, a group IB peptide, is part of one of the most lethal neurotoxins described, taipoxin, though the interaction of the three taipoxin chains has not yet been described. The cause of the toxicity of the α-chain and the purpose of the β-chain is unknown in taipoxin, while several suggestions have been postulated regarding the action of the γ-chain. Fohlman (1979) suggested this chain, specifically the carbohydrate moiety, may have a two fold function, to protect the toxic chain from non-specific adsorption of less relevant targets in the organism, and perhaps orientating the γ-chain away from the membrane in order to allow good cell contact for the α-chain (Fohlman, *et al.*, 1979).

Alternatively, binding may be initiated through another mechanism. The single hydrophilic peptides of OS2 and the α-taipoxin chain (as a single chain) show similar hydrophilicity but have differing toxicities as discussed above. This indicates that if OS2 does act as a single subunit, it is efficiently capable of binding to its target. Similarly, the more hydrophobic peptides of OS1 and the single γ-chain, which would be expected to be capable of surface binding, are non-toxic. Speculating from this information, this may indicate the group of *Oxyuranus* pre-synaptic peptides do not reach their targets through entering cell membranes. These peptides possibly bind through the formation of a secondary structure to receptors as the key to toxicity within this group of peptides. These acidic peptides may aid specifically with the insertion of peptides into receptors.

Speculatively, the non-toxic γ-chain may play a role in the toxicity of the taipoxin complex (higher than OS2). This may be caused by the structural interactions of the subunits allowing some of the 'toxic' residues to become available. This interaction has been suggested of the acidic chain of crotoxin[2], indicating the acidic, non-toxic subunit

[2] Crotoxin from the venom of the South American rattlesnake *Crotalus durissus terrificus* is a complex of two different chains: a basic and weakly toxic phospholipase A_2, component B (CB), and an acidic, non-toxic protein, component A (CA). Crotoxin exerts its pathophysiological action by blocking neuromuscular transmission. It acts primarily at the pre-synaptic level by altering neurotransmitter release although it is also able to act post-synaptically by stabilizing the acetylcholine receptor in an inactive state.

acts as more than a chaperone for the second and toxic chain (Kaiser and Aird, 1987) (Kaiser and Middlebrook, 1988). The specific mechanism is not understood.

Structural changes in mammalian pancreatic lipases (3.1.1.3) have also been described after specific binding to reveal an active site (Winkler, 1990). These phospholipases (PL) contain two sections, an N-terminal domain containing the active site and a surface loop traversing two cysteines which forms the so-called lid. In the closed formation, this 'lid' prevents the substrate from reaching the active site. The non-catalytic C-terminal domain plays an important role in interactions with another protein named a colipase, an amphiphilic protein with a comparable "three finger" topology to that of snake toxins. The tips of the fingers of colipase bind to the C-terminal domain of PL (van Tilbeurgh, *et al.*, 1992). When colipase binds to PL a drastic change in the conformation of the lid occurs (van Tilbeurgh, *et al.*, 1993), as the active site becomes accessible to a substrate and the lid binds to the N-terminal part of colipase (Chahinian and Carriere, 1999).

A structural mechanism similar to this may occur within taipoxin, though most likely through a different mechanism due to the venom PLA_2s being shorter than the long mammalian phospholipases (449 residues). It is possible each chain of taipoxin may lead a specific role in the binding, insertion and action of this toxin.

Significantly receptor proteins, neuronal pentraxins, have been identified and are suggested to facilitate the binding of taipoxin to neurones (Omesis *et. al.* 1996, Dodds *et. al.* 1997, Kirkpatrick, *et. al.*, 2000). Hydropathy plots of each sequence did not indicate areas of potential interaction (data not shown). Future protein modelling may indicate specific areas of binding for both taipoxin, specifically the α-subunit, and OS2. This in turn may indicate a specific binding mechanism and the residues causative of toxicity.

Of interest was the formation of molecular complexes between peptide subunits, or the lack of, with some peptide subunits remaining as monomers though maintaining apparently similar sequences. An example of this is the acidic γ-taipoxin, which forms a complex whilst another similar sequence, OS1, purportedly remains as a monomer.

Comparative analysis of peptide sequences of the two subunits to outlying genera did not indicate specific regions that would cause this binding variation. That is, amino acid residues from OS1 (monomer) were shared in sequences from subunits known to form complexes (eg *N. scutatus* sequences are subunits of a multimeric protein). Accumulative or charge discrepancies may cause this difference in binding. Or, these sequences are potentially located on different genes. Again, further work involving mutation studies or the production of a genomic library of the venom sequences and phylogenetic interpretation may answer this question.

5.3.2 Preliminary alignment of pre-synaptic neurotoxins

In the absence of expression studies, the preliminary sequence alignment produced from the *O. s. scutellatus* peptide and putative peptide sequences was aimed to identify the relationships between these sequences. Based on the comparative analysis of the putative and reported peptide sequences of *Oxyuranus* and other elapid sequences the PLA_2s seem to have originated from a common ancestor by gene duplication, and evolved separately. Similar speculation has been put forward by Housset and Fontecilla-Camps (1996).

Phylogenetic analysis indicated a strong relationship between the sequences reflecting the type and pI of the peptides. The basic *Oxyuranus* peptides (α-chain, OS2, R125, R56), of which two have been proven to be toxic, and acidic peptides (γ taipoxin, R120, OS1 and R132) clustered into groups indicating strong relationships. Phylogenetic trees were generated and robustness was assessed by bootstrap analysis using 10,000 replicates. As indicated in Figure 5.5, RW26 appeared to be an outlier to the group consisting of the alpha chain, OS2, R56 and R126. The phylogenetic trees also indicated the group IB peptide sequences appeared earlier in evolution, suggesting that they could have served as the ancestral molecules for group IA peptides. Outlying groups added in Figure 5.6 confirmed the relationships between the sequences suggested in Figure 5.5 (Figure 5.5).

The preliminary neighbour joining trees were useful in demonstrating peptide sequence similarities and supported the analysis and the suggested relationships of the

amino acid sequences. Yet this analysis may not accurately reflect phylogeny due to a number of problems. Within the analysis of *Oxyuranus*, it was unknown if the clustered amino acid sequences represented different loci or were different alleles on the same loci. This may have significant repercussions on the accuracy of the analysis. Other problems with constructing accurate phylogenies using venom sequences include the paucity of knowledge regarding *Oxyuranus* toxins and, with the small sample size it would be difficult to identify a statistically viable and accurate analysis without a high degree of expertise. An expert analysis was not only outside the scope of this investigation, but was beyond the proficiency of the author.

If however, this were to be pursued there are several routes that could be undertaken to extend this data. It would be interesting to undertake a rooted phylogeny with maximum likelihood trees, especially of genomic sequences. These aim to reconstruct the history of successive divergence from a common ancestor which took place during their evolution (Davidson and Dennis, 1990; Nakashima, *et al.,* 1993; John, *et al.,* 1994; Nakashima, *et al.,* 1995; Kordis and Gubensek, 2000; Armugam, *et al.,* 2004). It would also be interesting to undertake an analysis of the DNA substitution patterns across species, which may indicate which sequences are under selection pressures.

An analysis of the *Oxyuranus* toxin sequences, together with a variety of available PLA_2 toxin sequences from different snakes could show a number of things. In particular to what extent the different *Oxyuranus* toxins evolved from a common ancestry solely within taipans. It may also show to what extent they are representative of older lineages that are also present in other elapid snakes; in effect, how was the PLA_2 arsenal of *Oxyuranus* assembled. Analysis could also indicate how the polymeric toxins (e.g., taipoxin) were assembled, or may describe the phylogenetics of the constituent protein chains. It would be interesting to investigate, for example, clone OS1/R132, as this sequence appeared to contain a chimeric relationship with the N-terminal of the toxic α-chain and the C-terminal of the non-toxic γ-chain. This may indicate R132 represents an evolutionary intermediate, through gene duplication, between groups IA and IB. Further analysis of these sequences, due to their similar structures and isomers could also prove to be useful as markers for understanding population genetics or for use within conservation genetics of the *Oxyuranus* species.

One problem that may arise from aligning venom sequences is due to an inconstant evolutionary clock. The accelerated evolution together with reported gene duplication or crossing over (e.g., exon swapping) of venom proteins occurring over short time periods (Moura da Silva, *et al.,* 1995; Chang, *et al.,* 1997; Slowinski, *et al.,* 1997; Afifiyan, *et al.,* 1999; Kordis and Gubensek, 2000; Takacs, *et al.,* 2001) may cause molecular phylogenies to not represent a true divergence.

Analyses of the signal peptides also indicate a potential evolutionary pathway, such as the recruitment of peptides into venom. Classical leader sequences vary, and the described signal peptides are presumably involved in the secretion of neurotoxins. The signal peptides of the pre-synaptic neurotoxins from *O. scutellatus* shared strong homology to other elapid species, indicating these sequenced most probably diverged from a common recruitment event, supporting the NJ trees (Fig 5.5). *O. s. scutellatus* sequences contained a His at –26 where an aromatic Tyr and an acidic Asn are found in the other elapid sequences compared. This may suggest this residue is not under selective pressure as the residue characteristics at this site are highly varied. Similarly, residues at positions –3 and –7 also follow this trait. Residues at –6 and possibly –7 contain either an Ala or Ser residue. Interestingly, the species with the closest sequence homology is *L. semifasciata* (Fujimi, *et al.*, 2002). This sequence maintained residues opposite to the other sequences at -3 and -26 (acidic residue where the others are basic and vice versa). The significance of these substitutions is not known.

5.3.3 Post-synaptic neurotoxins

Toxin 1 isomer (R147 and R121)

Initial sequence analysis of R121 suggested it was identical to R147. Subsequent sequencing showed that R121 differed by sixty nucleotides after the first stop codon from R147 and lacked a poly-A tail. This divergence may suggest the similar toxin is coded my either different genes or may possibly be due to different gene splicing (Figure 5.19), further analysis is required to confirm this suggestion.

The translated proteins of R147 and R121 maintained the highly conserved structures representative of post-synaptic or three finger toxins (3FTx). These sequences exhibited similar hydrophobic, charge and structural properties maintaining the position of the eight cysteine residues, inferring a similar structure. Whereas all residues suggested for toxicity and binding were also conserved. This indicated these sequences potentially represent isomers within an established class of post-synaptic toxins.

In vitro and *in vivo* studies of peptides sharing this homology have shown these are pharmacologically active neurotoxins capable of exerting muscle paralysis, spasms and increased respiration. These peptides bound with high affinity to muscle AchR thus inhibiting the physiological function of the receptor by blocking the binding of acetylcholine. At times the similarity in sequence belies the differing actions, for example, erabutoxin demonstrates distictive pharmacological actions between the three similar forms of chains named a, b and c, with one amino acid difference between each chain (Low, 1979). The enzymic action of these clones cannot be inferred until expression studies are undertaken.

The nucleotide sequence from RW6

This was the only toxin sequence to be isolated from the *O. scutellatus* library using PCR. The deduced amino acid sequence showed distinctly different characteristics to other deduced amino acid sequences with few conserved toxic residues, sharing 50% of the conserved sites. This protein maintained the conserved cysteine residues required for 3FTx structure and was hydrophobic with large trans-membrane regions. These large hydrophobic regions suggest this peptide may target cell membranes. This was also the only post-synaptic peptide sequence with possible glycosylation sites. This sequence may represent a new class of toxin within *Oxyuranus*. Future expression and activity assays will help to categorise this peptide sequence.

Due to the varying pharmacological effects of the toxins described, it can be summarised that sequence comparisons of all venom toxins described within this Chapter may not accurately infer a physiological action. Further structural analysis such as CD spectra, NMR and X-ray crystallography will allow an understanding of

how these toxins interact, whereas site-directed mutagenesis may pinpoint specific residues or groups of residues causing activities.

The cloning of α-neurotoxins has so far been carried out on sea snakes (Tamiya *et al.,* 1985; Obara *et al.,* 1989; Fuse *et al.,* 1990) and on many land snakes (Chu *et al.,* 1995; Afifyan *et al.,* 1998) including some Australian snakes (Gong *et al.,* 1999; Gong *et al.,* 2000; Judge *et al.,* 2002). Further cloning of the cDNAs and genes encoding *Oxyuranus* α-neurotoxins should be pursued, as interpretation of the sequences isolated using phylogenetic interpretation, similarly to the pre-synaptic neurotoxins, may provide a useful tool in the evolutionary placement of these sequences.

5.3.4 Factor V

The complete cDNA sequence of the FV subunit from *O. s. scutellatus* was determined. This sequence shared strong homology with the non enzymic (FV) subunit of the prothrombin activator from an Australian snake toxin, *P. textilis* (Pseutarin C). The structural and cleavage sites across the two species were conserved, including a unique modification allowing an apparent resistance to activated protein C (APC) which makes it highly suitable for its role as a toxin. Normal factor Va is broken down by APC, in the presence of protein S, turning off coagulation. Although most of the proteolytic activation sites are conserved, similar to *P. textilis* FV, two of the three proteolytic sites cleaved by activated protein C are mutated, thus APC is not able to inactivate this pro-coagulant toxin.

One variation observed from this comparison that may be of consequence was an additional potential glycosylation site. Glycosylation is known to play an important role in the stability as well as the folding of proteins. Altering either the carbohydrate moiety or the glycosylation site can either cause loss or enhance the function of the protein. For example, α-neurotoxins from *Naja naja* venom have adapted to resist the action of its own venom. α-neurotoxins are targeted against nicotinic acetylcholine receptors (nAChR), yet in *N. naja* these toxins fail to bind to their own receptors. It was found that an N-glycosylation signal within the ligand-binding domain prevented the venom protein from binding to its own receptors (Takacs, *et al.*, 2001). Due to this, the presence of an extra glycan attachment within *O. s. scutellatus* may have repercussions

regarding its activity. It would be interesting to examine the role of this glycosylation in pseutarin C and the scutelarin FV non-enzymic subunit.

It has been suggested a similar, if not identical, Factor V protein might undertake two roles, with a peptide as a toxin in the venom and as a haemostatic factor when present in the plasma (Rao, *et al.*, 2002). Intriguingly, the signal peptide from both snake species was identical. It would be interesting to examine the signal peptides of the *Pseudonaja* and *Oxyuranus* FV sequences from plasma. The comparisons of the plasma and toxin sequences may indicate a method of recruitment of these toxin sequences.

In conclusion, this is the second sequence of an FV-like protein from a non-hepatic and non-mammalian source to be described. The non-enzymic subunit of scutelarin shares a similar domain structure to mammalian FV, consisting of domains A1-A2-B-A3-C1-C2 and an identical structure to the non-enzymic subunit of *P. textilis* pseutarin C. The complete sequence of this protein revealed the importance of this sequence when compared with the clinically significant brown snake prothrombin activator. Future structure and function studies may aid in understanding relationships of FVa in prothrombin activation, whereas mutation of active sites and lethal dose studies will help characterise this peptide. Phylogenetic analysis of future venom and haemostatic sequences will aid in understanding the evolution and recruitment mechanism used for venom proteins.

5.4 Annex of Chapter 5

5.4.1 Details of sequences isolated from *O. s. scutellatus* cDNA library matching venom PLA_2 neurotoxins

Putative β-taipoxin subunit (Clone no. R24)

Six identical sequences were isolated matching the R24 clone. This was a 690 nucleotide sequence of which 438 nucleotides composed an open reading frame (ORF). This ORF translated into a 146 amino acid including a 27 amino acid signal peptide (Figure 5.10). There were 113 bp of 5'-upstream non-coding sequence. The putative peptide encoded by R24 had a molecular mass and pI of 13,234 Da and 4.96, respectively, and matched β-taipoxin (GenBank Accession No. POO615). This was the only peptide sequence to exhibit 100% homology with a toxin, β-taipoxin, a characterised *O. scutellatus* protein and closely resembled group IA PLA_2s (Figure 5.2).

Putative α-taipoxin isomer (Clone no. RW 26)

Four matching clones contained a nucleotide sequence that included 438 base pairs of ORF translating to a 146 amino acid residue protein including a 27 amino acid signal peptide (Figure 5.11). The peptide encoded by RW26 shared homology with α-taipoxin (GenBank Accession No. POO614)(Figure 5.12). The molecular mass and pI of the deduced 119 residue RW26 was 13,520 Da and 9.37, respectively and aligned with group 1A PLA_2s (Figure 5.2).

Putative γ-taipoxin isomer (Clone no. R120)

This clone was a 625 nucleotide sequence and included 453 nucleotide residues of ORF translating into a 151 amino acid residue protein including an 18 amino acid residue signal peptide. This ORF was followed by a 3'-non coding region of 165 nucleotides (Figure 5.13). Only one clone was found from the sequences isolated. The molecular mass and pI of the deduced protein from R120 was 14,601 Da and 3.88 respectively with one potential glycosylation site at 97 (NDT). This sequence shared

the greatest homology with the *O. s. scutellatus* γ-taipoxin (94%)(GenBank Accession No. POO616)(Figure 5.14), and closely resembled group IB PLA_2s (Figure 5.3).

Putative OS1 isomer (Clone no. R132)

Clone R132 contained a 623 nucleotide sequence that included a 462 bp ORF. This ORF translated into a 154 amino acid residue protein that included a 27 amino acid residue signal peptide. This ORF was followed by a 3'-untranslated region of 155 nucleotides (Figure 5.15). The molecular mass and pI of the protein encoded by R132 was 13,786 Da and 4.69, respectively. This peptide shared homology with OS1 (GenBank Accession No. AAB33759)(Figure 5.16) and group IB proteins (Figure 5.3).

Putative OS2 isomer (Clone no. R56)

The complete nucleotide sequence of this clone was comprised of 626 nucleotides that included 438 nucleotides of ORF that translated into a 146 amino acid residue protein and differed by one amino acid residue to R125 (Figure 5.17).

Putative OS2 isomer (clone no. R125)

Sequence of these clones revealed a 613 nucleotide sequence that included 438 nucleotides of ORF translating into a 146 amino acid residue protein including a 27 amino acid residue signal peptide (Figure 5.18). The molecular mass and pI of the deduced protein was 13,371 Da and 9.36, respectively. The deduced peptide sequence from R56 and R125 isolated from the *O. s. scutellatus* cDNA library matched *O. s. scutellatus* OS2 peptide (GenBank Accession no. AAB33760, Figure 5.19)(94%) with the greatest homology with group IA proteins (Figure 5.2).

Figure 5.10: Nucleotide and deduced amino acid sequence of β-taipoxin (R24).

The putative peptide sequence (clone R24) isolated from *O. s. scutellatus* cDNA library that matched β-taipoxin (GenBank Accession No. POO615). The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined. A potential TATA box and CCGAACTC are boxed.

```
1    ATTATAGTACCAAGGTGGCCGGGGGACTGAGAAAGTTTGGGAACCACTGCTATAAAAGGCCGAA    64

65   CTCGCCAGGTATCCAAATTCATCTTGCTTGCAGCTTCACCACTGACAAA ATG CAT CCT GCT    125
1                                                      M   H   P   A      4

125  CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC CTC TTA GGA GCC TCC GAC  175
5     H   L   L   V   L   L   A   V   C   V   S   L   L   G   A   S   D   21

176  ATT CCT CCG CTG CCT CTC AAC CTC GTA CAA TTC GGC AAA ATG ATT GAA TGT  226
22    I   P   P   L   P   L   N   L   V   Q   F   G   K   M   I   E   C   38

227  GCC ATC CGT AAC AGG CGA CCT GCT TTG GAT TTT ATG AAC TAC GGT TGC TAC  277
39    A   I   R   N   R   R   P   A   L   D   F   M   N   Y   G   C   Y   55

278  TGT GGC AAA GGA GGT AGC GGG ACA CCG GTA GAC GAC TAT TTG GAT AGG TGC  328
56    C   G   K   G   G   S   G   T   P   V   D   D   L   D   R   C   C   72

329  TGC CAG GTT CAT GAC GAG TGC GCT GAA GCC GAA AAG CAT GGA TGC TAC CCC  379
73    Q   V   H   D   E   C   Y   A   E   A   E   K   H   G   C   Y   P   89

380  AGT TTG ACG ACG TAT ACA TGG GAA TGT CGC CAA GTT GGA CCC TAC TGC AAT  430
90    S   L   T   T   Y   T   W   E   C   R   Q   V   G   P   Y   C   N   106

431  TCA AAA ACG CAG TGT GAA GTT TTT GTG TGT GCT TGT GAC TTC GCG GCA GCC  481
107   S   K   T   Q   C   E   V   F   V   C   A   C   D   F   A   A   A   123

482  AAG TGC TTT GCC CAA GAA GAT TAC AAC CCT GCC CAC TCC AAT ATC AAC ACC  532
124   K   C   F   A   Q   E   D   Y   N   P   A   H   S   N   I   N   T   140

533  GGT GAA CGT TGC AAA TGA TATTTGAGAGGCTTCAGCGCGAGGACTGTGGCAGTTACTCACC  594
141   G   E   R   C   K   *                                              145
595  TGCGCCTGGCAATTCTCTGGACGGGCCTCTATTATATATATAAAAATAGAAAATTATATATATATAAT 662
663  TATTAAAAACAAAAGGAACCGTTTCTGAAAAAAAA                                690
```

Figure 5.11: Nucleotide and deduced amino acid sequence of an α-taipoxin isomer (RW26). The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    CACGAGGA ATG CAT CCT GCT CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC   53
1             M   H   P   A   H   L   L   V   L   L   A   V   C   V   S   15

54   CTC TTA GGA GCC TCC GAC ATT CCT CCG CTG CCT CTC AAC CTC GTA CAA TTC   104
16    L   L   G   A   S   D   I   P   P   L   P   L   N   L   V   Q   F    32

105  GGC TTC ATG ATT AAA TGT GCC AAC CGT AAC AGT CGA CCA GTT TCG CAT TAT   155
33    G   F   M   I   K   C   A   N   R   N   S   R   P   V   S   H   Y    49

156  ATG GAC TAC GGT TGC TAC TGC GGC CCT GGA GGT AGT GGG ACA CCG ATA GAT   206
50    M   D   Y   G   C   Y   C   G   P   G   G   S   G   T   P   I   D    66

207  GAC TTG GAT AGG TGC TGC CAG GTT CAT GAC CAG TGC TAT GAT GAA GCC GTA   257
67    D   L   D   R   C   C   Q   V   H   D   Q   C   Y   D   E   A   V    83

258  AGA CGC TAC AGA TGC TCC CCC TAC TTG ACG ATG TAT AGT TGG AAA TGT TAT   308
84    R   R   Y   R   C   S   P   Y   L   T   M   Y   S   W   K   C   Y   100

309  GGA ACA GCA CCC TCC TGC AAT TCA AAA ACG CAG TGT GAA GGT TTT GTG TGT   359
101   G   T   A   P   S   C   N   S   K   T   Q   C   E   G   F   V   C   117

360  AAT TGT GAT GCC AAA GCA GCC GAG TGC TTC GCC AGA TCC CCT TAC CAG CCC   410
118   N   C   D   A   K   A   A   E   C   F   A   R   S   P   Y   Q   P   134

411  TCG AAC CGG AAT ATC AAC ACC AAG GAA CGT TGC AAA TGA TATTT            455
135   S   N   R   N   I   N   T   K   E   R   C   K   *                 146
```

Figure 5.12: Alignment of α-taipoxin isomer (RW26) with α-taipoxin subunit. Putative peptide sequence from RW26 isolated from *O. s. scutellatus* cDNA library matched α-taipoxin (GenBank Accession No. POO614). Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text.

Formatted Alignments

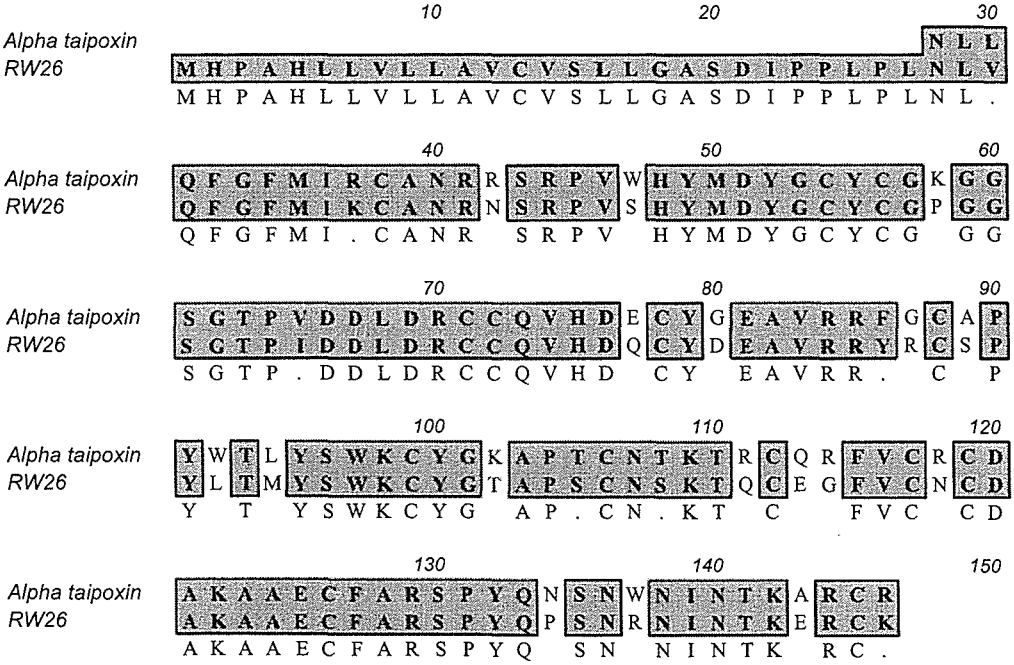


Figure 5.13: Nucleotide and deduced amino acid sequence of γ-taipoxin isomer (R120). The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    ATG CAT CCT GCT CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC CTC TTA  50
1     M   H   P   A   H   L   L   V   L   L   A   V   C   V   S   L   L  16

51   GGA TCC TCC GAG ATT CCT CAG CCA TCT CTT GAC TTC GAA CAA TTC AGC AAC  101
17    G   S   S   E   I   P   Q   P   S   L   D   F   E   Q   F   S   N  33

102  ATG ATT CAA TGT ACC ATC CCC TGC GGA GAA TCT TGC TTG GCT TAT ATG GAC  152
34    M   I   Q   C   T   I   P   C   G   E   S   C   L   A   Y   M   D  50

153  TAC GGT TGC TAC TGC GGC CCT GGA GGT AGT GGG ACA CCG ATA GAT GAC TTG  203
51    Y   G   C   Y   C   G   P   G   G   S   G   T   P   I   D   D   L  67

204  GAT AGG TGC TGC AAA ACA CAT GAC GAG TGC TAT GCT GAA GCT GGA AAA CTT  254
68    D   R   C   C   K   T   H   D   E   C   Y   A   E   A   G   K   L  84

255  TCT GCA TGT AAA TCC GTG CTG AGT GAG CCC AAC AAC GAC ACC TAT TCA TAT  305
87    S   A   C   K   S   V   L   S   E   P   N   N   D   T   Y   S   Y 101

306  GAA TGT AAT GAA GGC CAA CTC ACC TGC AAC GAT GAC AAC GAT GAG TGT AAA  356
102   E   C   N   E   G   Q   L   T   C   N   D   D   N   D   E   C   K 118

357  GCC TTT ATT TGT AAC TGT GAC CGC ACA GCA GTC ACC TGT TTC GCT GGA GCC  407
119   A   F   I   C   N   C   D   R   T   A   V   T   C   F   A   G   A 135

408  CCT TAC AAC GAC TTG AAC TAC AAT ATC GGC ATG ATT GAA CAT TGC AAA TGA  458
136   P   Y   N   D   L   N   Y   N   I   G   M   I   E   H   C   K   * 152

459  TATTTGAGAGTCTTCAGTACAAGGACTGTGGCAGTTACCCACCTGCGCGTGGCAATTCTCTGGACGG  525
526  GCCTCTAATATATATATATATATATATATAATAGAAAATTATAATAGAAAATTTATATATATATATA  592
593  TAAAGTAACTGTTTCCTGAACAATAAAGTGAGG                                    625
```

Figure 5.14: Alignment of clone γ-gamma taipoxin isomer (R120) with γ-taipoxin. Putative peptide sequence from R120 isolated from *O .s. scutellatus* cDNA library matched δ-taipoxin (GenBank Accession No. POO616). Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text.

Formatted Alignments

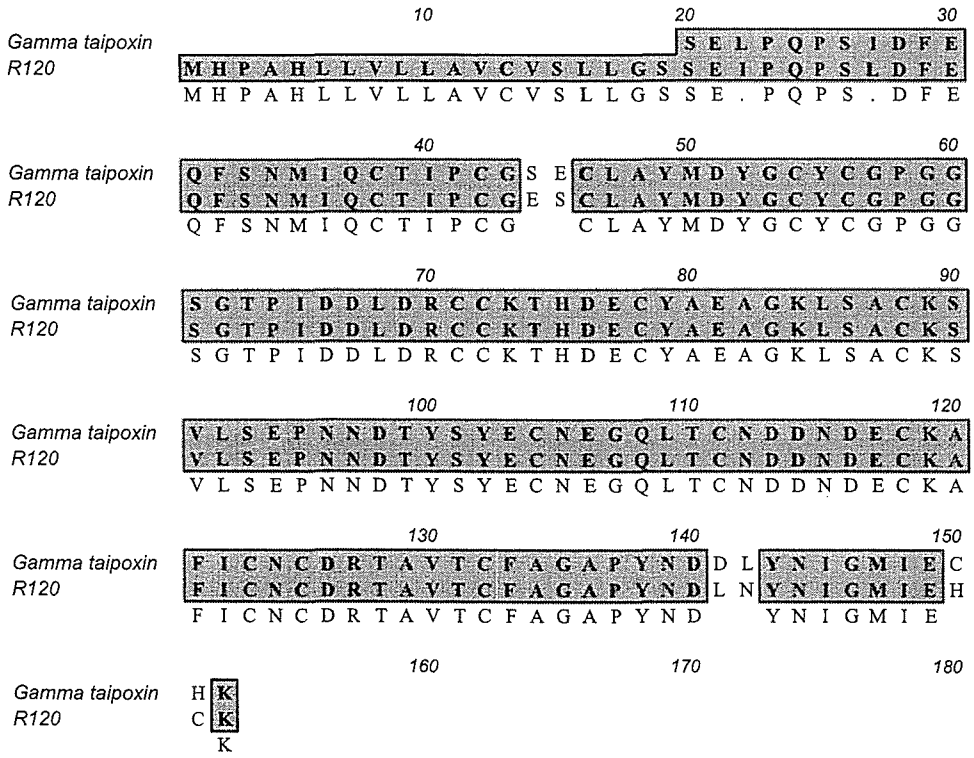


Figure 5.15: Nucleotide and deduced amino acid sequence of OS1 isomer (R132).
The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    AAA ATG CAT CCT GCT CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC CTC  51
1         M   H   P   A   H   L   L   V   L   L   A   V   C   V   S   L  16

52   TTA GGA GCT GCC AGA ATT CCT CCG CTG CCT CTC AGC CTC CTA AAT TTC GCC  102
17    L   G   A   A   R   I   P   P   L   P   L   S   L   L   N   F   A  33

103  AAC TTG ATT GAA TGT GCT AAC CAT GGC ACT CGA AGT GCG TTG GCT TAT GCG  153
34    N   L   I   E   C   A   N   H   G   T   R   S   A   L   A   Y   A  50

154  GAC TAC GGT TGC TAC TGC GGC AAA GGA GGT CGC GGG ACA CCG CTA GAC GAC  204
51    D   Y   G   C   Y   C   G   K   G   G   R   G   T   P   L   D   D  67

205  TTG GAT AGG TGC TGC CAT GTT CAT GAC GAC TGC TAT GGT GAA GCC GAA AAA  255
68    L   D   R   C   C   H   V   H   D   D   C   Y   G   E   A   E   K  84

256  CTT CCT GCA TGT AAT TAC CTG ATG AGT AGC CCC TAC TTC AAC AGC TAT TCA  306
85    L   P   A   C   N   Y   L   M   S   S   P   Y   F   N   S   Y   S  101

307  TAC AAA TGT AAT GAA GGC AAA GTC ACC TGC ACA GAT GAC AAC GAT GAG TGT  357
102   Y   K   C   N   E   G   K   V   T   C   T   D   D   N   D   E   C  118

358  AAA GCC TTT ATT TGT AAT TGT GAC CGC ACG GCA GCC ATC TGT TTC GCC GGA  408
119   K   A   F   I   C   N   C   D   R   T   A   A   I   C   F   A   G  135

409  GCC ACT TAC AAC GAC GAA AAC TTC ATG ATC TCC AAG AAG AGA AAT GAT ATT  459
136   A   T   Y   N   D   E   N   F   M   I   S   K   K   R   N   D   I  152

460  TGC CAA TGA TATTTGAGAGGCTTCAGTGCAAGGACTGTGGCAGTTACTCACTTGCACATGGCAA  523
153   C   Q   *                                                          155

524  TTCTCTGGACAGGCCTTTATTATATATATATCAAAATAGAAAATTATATATATATATATATAATTAT  590
591  TAAAAAACAAAAGGAACCATTTCCTGAACAATA                                    623
```

Figure 5.16: Alignment of R132 with *O. scutellatus* OS1
Putative peptide sequence from R132 isolated from *O. s. scutellatus* cDNA library shared homology with OS1 (GenBank Accession No. AAB33759). Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text.

Formatted Alignments

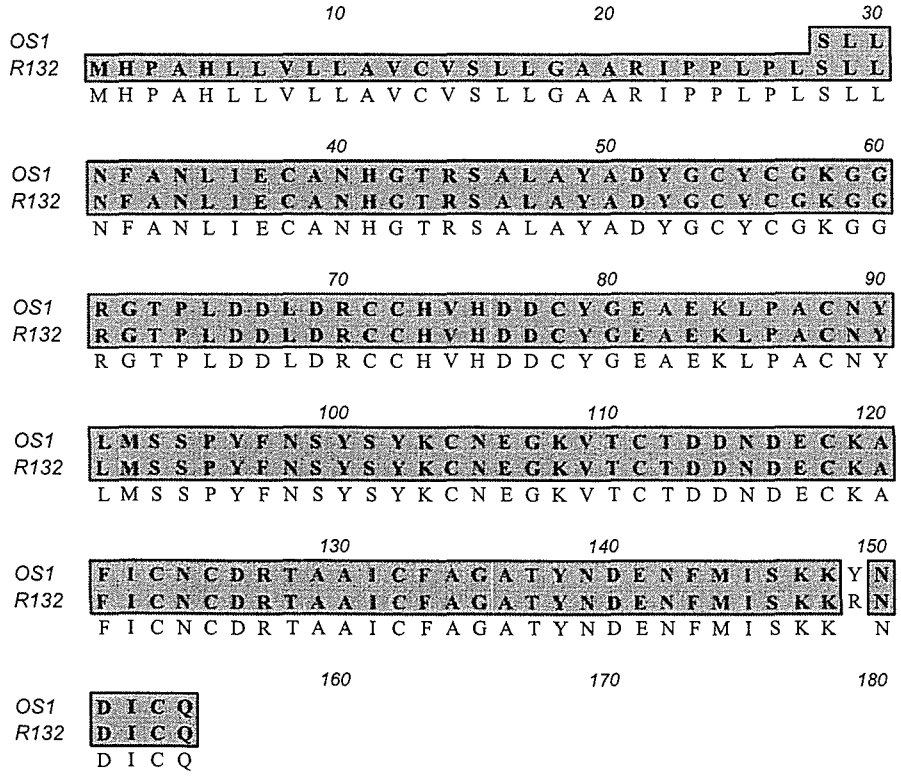


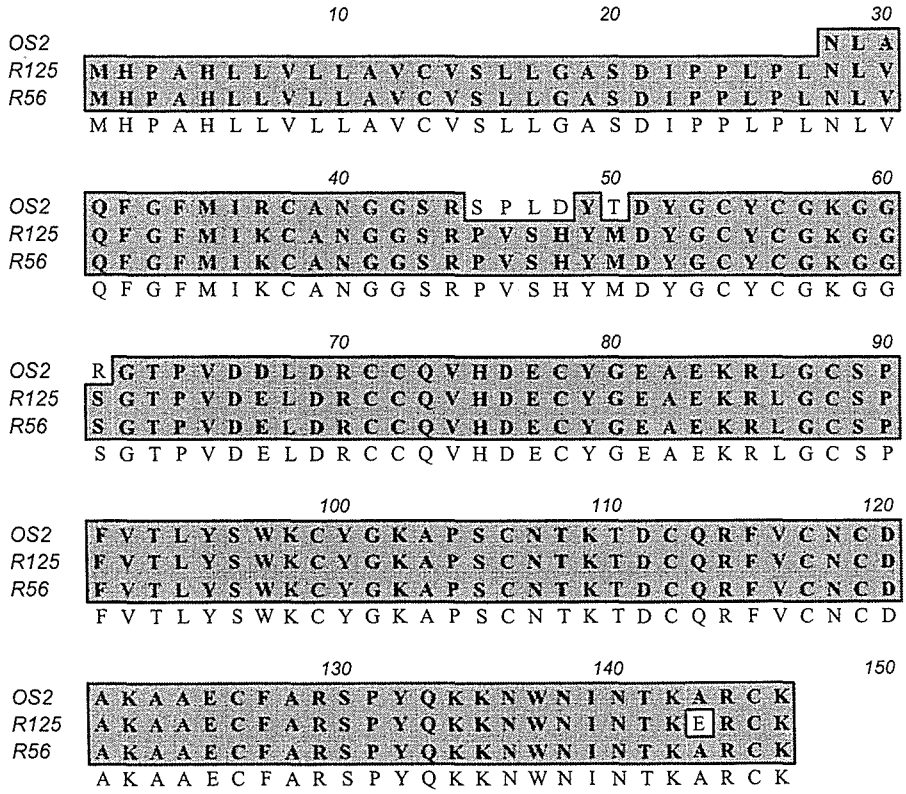
Figure 5.17: Nucleotide and deduced amino acid sequence of OS2 isomer (R56).
The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    C AAA ATG CAT CCT GCT CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC CTC 52
1          M   H   P   A   H   L   L   V   L   L   A   V   C   V   S   L  16

53   TTA GGA GCC TCC GAC ATT CCT CCG CTG CCT CTC AAC CTC GTA CAA TTC GGC  103
17    L   G   A   S   D   I   P   P   L   P   L   N   L   V   Q   F   G   33

104  TTC ATG ATT AAA TGT GCC AAC GGT GGC AGT CGA CCA GTT TCG CAT TAT ATG  154
34    F   M   I   K   C   A   N   G   G   S   R   P   V   S   H   Y   M   50

155  GAC TAC GGT TGC TAC TGT GGC AAA GGA GGT AGC GGG ACA CCG GTA GAC GAG  205
49    D   Y   G   C   Y   C   G   K   G   G   S   G   T   P   V   D   E   67

206  TTG GAT AGG TGC TGC CAG GTT CAT GAC GAG TGC TAT GGT GAA GCC GAA AAA  256
68    L   D   R   C   C   Q   V   H   D   E   C   Y   G   E   A   E   K   84

257  CGC TTG GGA TGC TCC CCC TTC GTG ACG TTG TAT AGT TGG AAA TGT TAT GGA  307
85    R   L   G   C   S   P   F   V   T   L   Y   S   W   K   C   Y   G  101

308  AAA GCA CCC TCC TGC AAT ACG AAA ACG GAC TGT CAA CGT TTT GTG TGT AAT  358
102   K   A   P   S   C   N   T   K   T   D   C   Q   R   F   V   C   N  118

359  TGT GAT GCC AAA GCA GCC GAG TGC TTC GCC AGA TCC CCT TAC CAG AAA AAA  409
119   C   D   A   K   A   A   E   C   F   A   R   S   P   Y   Q   K   K  135

410  AAC TGG AAT ATC AAC ACC AAG GCA CGT TGC AAA TGA TATTTGAGAGGCTTCAGCG  464
136   N   W   N   I   N   T   K   A   R   C   K   *                     147

465  CGAGGACTGTGGCAGTTACTCACCTGCGCCTGGCAATTCTCTGGACGGGCCTCTATTATATATATAA  531
532  AAATAGAAAATTATATATATATATAATTATTAAAAAACAAAAGGAACCATTTCCTGAACAATAAAGT  598
599  GAGGTGCCGATAAAAAAAAAAAAAAAA                                         610
```

Figure 5.18: Nucleotide and deduced amino acid sequence of OS2 isomer (R125).
The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    GAC AAA ATG CAT CCT GCT CAC CTT CTG GTC CTG TTG GCA GTT TGT GTC TCC  51
1          M   H   P   A   H   L   L   V   L   L   A   V   C   V   S   15

52   CTC TTA GGA GCC TCC GAC ATT CCT CCG CTG CCT CTC AAC CTC GTA CAA TTC  102
16    L   L   G   A   S   D   I   P   P   L   P   L   N   L   V   Q   F   32

103  GGC TTC ATG ATT AAA TGT GCC AAC GGT GGC AGT CGA CCA GTT TCG CAT TAT  153
33    G   F   M   I   K   C   A   N   G   G   S   R   P   V   S   H   Y   49

154  ATG GAC TAC GGT TGC TAC TGT GGC AAA GGA GGT AGC GGG ACA CCG GTA GAC  204
50    G   T   P   V   D   M   D   Y   G   C   Y   C   G   K   G   G   S   66

205  GAG TTG GAT AGG TGC TGC CAG GTT CAT GAC GAG TGC TAT GGT GAA GCC GAA  255
67    E   L   D   R   C   C   Q   V   H   D   E   C   Y   G   E   A   E   83

256  AAA CGC TTG GGA TGC TCC CCC TTC GTG ACG TTG TAT AGT TGG AAA TGT TAT  306
84    K   R   L   G   C   S   P   F   V   T   L   Y   S   W   K   C   Y  100

307  GGA AAA GCA CCC TCC TGC AAT ACG AAA ACG GAC TGT CAA CGT TTT GTG TGT  357
101   G   K   A   P   S   C   N   T   K   T   D   C   Q   R   F   V   C  117

358  AAT TGT GAT GCC AAA GCA GCC GAG TGC TTC GCC AGA TCC CCT TAC CAG AAA  408
118   N   C   D   A   K   A   A   E   C   F   A   R   S   P   Y   Q   K  134

409  AAA AAC TGG AAT ATC AAC ACC AAG GAA CGT TGC AAA TGA TATTTGAGAGGCTTCA 463
135   K   N   W   N   I   N   T   K   E   R   C   K   *                  146

464  GAGGACTGTGGCAGTTACTCACCTGCGCCTGGCAATTCTCTGGACGGGCCTCTATTATATATATAAAA 531
532  ATAGAAAATTATATATATATATAATTATTAAAAAACAAAAGGAACCGTTTCCTGAACAATAAAGTGAG 599
600  GTGCCGATACCGGAAAA                                                   613
```

Figure 5.19: Alignment of clones R125 and R56 with OS2
Putative peptide sequence from R 56 and R125 isolated from *O. s. scutellatus* cDNA library matched *O. s. scutellatus* OS2 (GenBank Accession no. AAB33760). Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text.



5.4.2 Details of sequences isolated from *O. s. scutellatus* cDNA library matching post-synaptic neurotoxins

Figure 5.20: Nucleotide and deduced amino acid sequence of R147. The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    GCAAG ATG AAA ACT CTG CTG CTG ACC TTG GTG GTG GTG ACA ATC GTG TGC    50
1          M   K   T   L   L   L   T   L   V   V   V   T   I   V   C    15

51   CTG GAC TTA GGG TAC ACC ATG ACA TGT TAC AAC CAA CAG TCA TCG GAA GCT  101
16    L   D   L   G   Y   T   M   T   C   Y   N   Q   Q   S   S   E   A   32

102  AAA ACC ACT ACA ACT TGT TCA GGT GGG GTG AGC TCT TGC TAT AAA AAG ACT  152
33    K   T   T   T   T   C   S   G   G   V   S   S   C   Y   K   K   T   49

153  TGG AGT GAT GGC CGT GGA ACT AGA ATT GAA AGG GGA TGT GGT TGC CCT AGC  203
50    W   S   D   G   R   G   T   R   I   E   R   G   C   G   C   P   S   66

204  GTG AAG AAA GGT ATTGAACGTATATGTTGCAGAACAGACAAATGCAACAATTA GCT CTA    262
67    V   K   K   G   I   E   R   I   C   C   R   T   D   K   C   N   N   *    84

263  CGA GTG GCT AAA TTC CTT GAG TTT TGC TCT CAT CCA TCG ACG ACC ATC CTT  313

314  GAA AAT TTA TGC TTC TGG CCT TTA CCA CCA GAT GGT CCA TCA TCC CCC TCT  364

365  CCC CTG CTG TCT TTG ACA CCT CAA CAT CTT TCC CTT TTC TCT CAT TCT GTA  415

416  AGT TTC CTT CTG CTA GTT CTG TAG TTTGAGAATCAAATAAACCTCAGCATTCAAAAAA  473

474  AAAAAAAAAAAA
```

Figure 5.21: Nucleotide and deduced amino acid sequence of R121.
The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    GC AAG ATG AAA ACT CTG CTG CTG ACC TTG GTG GTG GTG ACA ATC GTG TGC   50
1          M   K   T   L   L   L   T   L   V   V   V   T   I   V   C   15

51   CTG GAC TTA GGG TAC ACC ATG ACA TGT TAC AAC CAA CAG TCA TCG GAA GCT  101
16    L   D   L   G   Y   T   M   T   C   Y   N   Q   Q   S   S   E   A   32

102  AAA ACC ACT ACA ACT TGT TCA GGT GGG GTG AGC TCT TGC TAT AAA AAG ACT  152
33    K   T   T   T   T   C   S   G   G   V   S   S   C   Y   K   K   T   49

153  TGG AGT GAT GGC CGT GGA ACT AGA ATT GAA AGG GGA TGT GGT TGC CCT AGC  203
50    W   S   D   G   R   G   T   R   I   E   R   G   C   G   C   P   S   66

204  GTG AAG AAA GGT ATT GAA CGT ATA TGT TGC AGA ACA GAC AAA TGC AAC AAT  254
67    V   K   K   G   I   E   R   I   C   C   R   T   D   K   C   N   N   83

255  TAG CTCTACGAGTGGCTAAATTCCTTGAGTTTTGCTCTCATCCATCGACGACCATCCTTGAAAGGT  320
86    *

321  CGCCGCCCCGTAACCTGTCGGATCACCGGAAAGGACCCGTAAAGTGATA                   369
```

Figure 5.22: Alignment of R147 and R121

The alignment of R147 and R121 which shared a common deduced peptide sequence diverging after the first stop codon. Amino acids sharing similarities are shaded and the common sequence is included beneath the shaded text. Gaps (-) have been inserted for optimal alignment.

Formatted Alignments

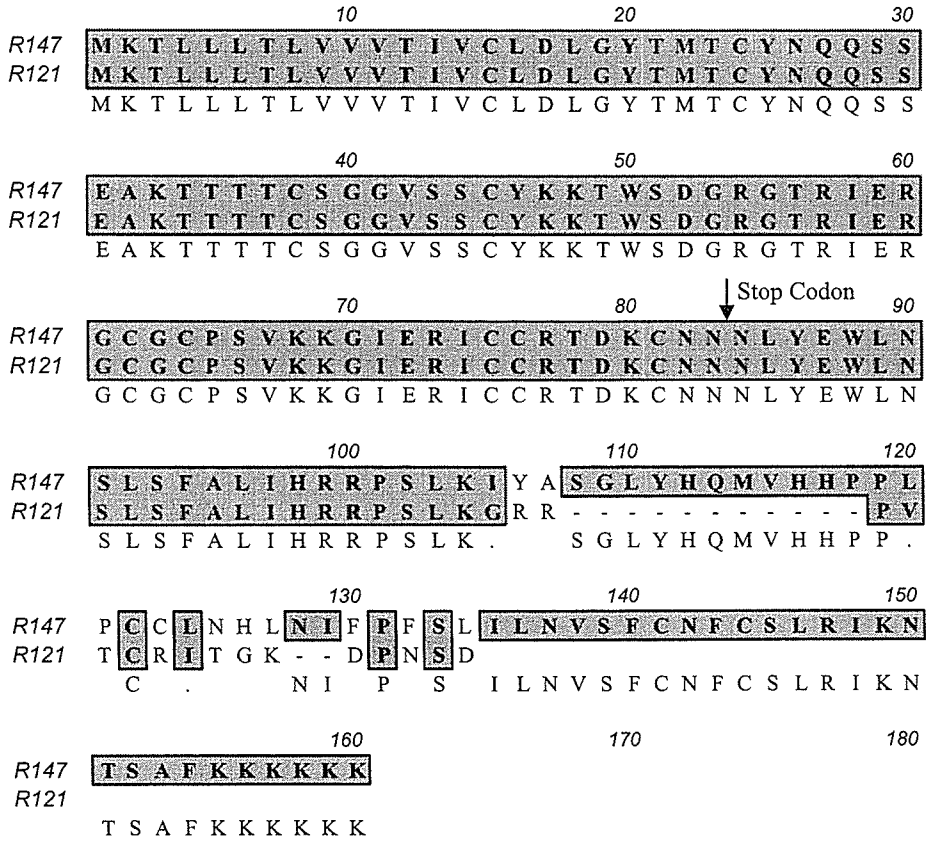


Figure 5.23: Nucleotide and deduced amino acid sequence of RW6.
The deduced amino acid sequence is shown below the nucleotide sequence. The transit peptide is underlined.

```
1    ATG AAA ACT TTG CTG CTG ACC TTG GTG GTG ATG ACA ATC ATG TGC CTG GAC  51
1     M   K   T   L   L   L   T   L   V   V   M   T   I   M   C   L   D   17

52  TTA GGA TAC ACC CTA ACA TGT TAC ATG AAT CCG TCT GGT ACT ATG GTT TGT  102
18   L   G   Y   T   L   T   C   Y   M   N   P   S   G   T   M   V   C   34

103 AAA GAA CAC GAG ACC ATG TGC TAT CAA CTT ATT GTT TGG ACA TTT CAA TAT  153
35   K   E   H   E   T   M   C   Y   Q   L   I   V   W   T   F   Q   Y   51

154  CGT GTG CTA TAC CTC AAG GGA TGT ACT TCT TCT TGC CCT GAGGGAAACAATCGT 207
52    R   V   L   Y   L   K   G   C   T   S   S   C   P   E   G   N   N   R   69

208 GCGTGTTGCAGTACAGGCCTGTGCAACAATTAGCT CTA CAA GTG GCT AAA TTC CTT GAG  266
70   A   C   C   S   T   G   L   C   N   N   *                          79

267  TTT TAC TCT CAT CCA TCA TGG ACC ATC CTT AAA AAT TTA TGG TTG TGG CCT  317

318  TTA CCA CCA GAT GGT CCA TCA TCC CTC TCT CCC                          350
```

5.4.3 Sequence sharing homology to Factor V

Figure 5.24: Complete nucleotide and deduced amino acid sequence of a clone isolated from an *O. s. scutellatus* venom gland cDNA library. The deduced amino acid sequence is shown below the nucleotide sequence and the transit peptide is underlined. The beginning and end of each domain is marked by an arrow. A potential adenylation signal is indicated in bold.

```
        10        20        30        40        50        60        70
AGTTGCTGCAGGCAGAACTGACTTCCGTGTATCTTTCAGCACATTAGCATCAATGGGAAGATACAGTGTG
                                                    M   G   R   Y   S   V

        80        90       100       110       120       130       140
AGCCCTGTCCCCAAATGTCTTCTACTGATGTTCCTGGGTTGGTCAGGGCTGAAGTATTACCAAGTGAATG
 S   P   V   P   K   C   L   L   L   M   F   L   G   W   S   G   L   K   Y   Y   Q   V   N

↓A1   150       160       170       180       190       200       210
CAGCTCAGCTCAGGGAGTACCGTATAGCTGCTCAGCTGGAAGACTGGGATTACAACCCCCAACCTGAGGA
A   A   Q   L   R   E   Y   R   I   A   A   Q   L   E   D   W   D   Y   N   P   Q   P   E   E

        220       230       240       250       260       270       280
GCTATCCAGATTATCAGAGTCAGATCTTACGTTTAAAAAAATTGTCTATAGAGAATATGAACTAGATTTC
  L   S   R   L   S   E   S   D   L   T   F   K   K   I   V   Y   R   E   Y   E   L   D   F

        290       300       310       320       330       340       350
AAACAAGAGAAGCCAAGAGATGAGCTCTCAGGGCTCCTAGGGCCAACACTACGTGGAGAAGTGGGAGACA
 K   Q   E   K   P   R   D   E   L   S   G   L   L   G   P   T   L   R   G   E   V   G   D

        360       370       380       390       400       410       420
GCCTCATAATTTATTTCAAGAATTTTGCTACTCAGCCTGTGAGCATTCACCCGCAGAGTGCCGTGTACAA
S   L   I   I   Y   F   K   N   F   A   T   Q   P   V   S   I   H   P   Q   S   A   V   Y   N

        430       440       450       460       470       480       490
CAAATGGTCAGAAGGTTCTTCATATTCTGATGGAACATCAGATGTGGAAAGACTGGATGATGCTGTGCCT
  K   W   S   E   G   S   S   Y   S   D   G   T   S   D   V   E   R   L   D   D   A   V   P

        500       510       520       530       540       550       560
CCAGGCCAGTCGTTCAAGTATGTGTGGAATATCACTGCAGAAATTGGGCCAAAGAAAGCTGATCCTCCCT
 P   G   Q   S   F   K   Y   V   W   N   I   T   A   E   I   G   P   K   K   A   D   P   P
```

```
      570       580       590       600       610       620       630
GTCTCACTTATGCGTACTACTCACATGTAAACATGGTGCGAGACTTTAATTCTGGTCTCATTGGTGCTTT
C  L  T  Y  A  Y  Y  S  H  V  N  M  V  R  D  F  N  S  G  L  I  G  A  L

      640       650       660       670       680       690       700
GCTGATATGTAAAGAAGGAAGCCTGAATGCAGATGGTGCACAAAAATTCTTCAACAGAGAATATGTGCTG
  L  I  C  K  E  G  S  L  N  A  D  G  A  Q  K  F  F  N  R  E  Y  V  L

      710       720       730       740       750       760       770
ATGTTTTCTGTGTTTGATGAAAGCAAGAACTGGTACAGAAAGCCCTCATTACAGTACACAATTAATGGGT
 M  F  S  V  F  D  E  S  K  N  W  Y  R  K  P  S  L  Q  Y  T  I  N  G

      780       790       800       810       820       830       840
TTGCCAATGGAACATTGCCTGATGTTCAGGCTTGTGCTTATGATCATATTAGCTGGCATTTGATAGGAAT
F  A  N  G  T  L  P  D  V  Q  A  C  A  Y  D  H  I  S  W  H  L  I  G  M

      850       860       870       880       890       900       910
GAGTTCCAGTCCTGAGATCTTCTCTGTTCACTTCAATGGACAAACCTTGGAACAAAACCATTACAAAGTG
  S  S  S  P  E  I  F  S  V  H  F  N  G  Q  T  L  E  Q  N  H  Y  K  V

      920       930       940       950       960       970       980
TCAACCATCAACCTTGTCGGAGGTGCCTCAGTAACAGCCAACATGTCAGTGAGCAGGACAGGAAAATGGC
 S  T  I  N  L  V  G  G  A  S  V  T  A  N  M  S  V  S  R  T  G  K  W

      990      1000      1010      1020      1030      1040      1050
TAATATCTTCTCTGGTTGCAAAGCATCTACAAGCTGGGATGTATGGTTATCTTAATATCAAAGACTGTGG
L  I  S  S  L  V  A  K  H  L  Q  A  G  M  Y  G  Y  L  N  I  K  D  C  G

     1060      1070      1080      1090      1100      A21110    1120
AAATCCAGATACTTTAACAAGAAAGTTATCCTTTAGAGAATGGAGGAGGATTATGAAATGGGAATATTTC
  N  P  D  T  L  T  R  K  L  S  F  R  E  W  R  R  I  M  K  W  E  Y  F

     1130      1140      1150      1160      1170      1180      1190
ATTGCTGCAGAAGAAATCACCTGGGATTATGCTCCAGAAATTCCTAGCAGTGTTGACAGAAGATACAAAG
 I  A  A  E  E  I  T  W  D  Y  A  P  E  I  P  S  S  V  D  R  R  Y  K

     1200      1210      1220      1230      1240      1250      1260
CTCAGTATCTGGATTTTTCAAATTTTATTGGCAAGAAATACAAAAAGGCAGTTTTCAGGCAATATGAAGA
A  Q  Y  L  D  F  S  N  F  I  G  K  K  Y  K  K  A  V  F  R  Q  Y  E  D

     1270      1280      1290      1300      1310      1320      1330
CAGCAATTTCACTAAACCGACCTATGCCATTTGGCCCAAAGAACGTGGAATTCTGGGCCCCGTTATCAAA
  S  N  F  T  K  P  T  Y  A  I  W  P  K  E  R  G  I  L  G  P  V  I  K

     1340      1350      1360      1370      1380      1390      1400
GCTAAAGTCAGAGACACAGTAACAATTGTATTCAAAAATCTGGCCAGTCGACCTTACAGCATTTATGTGC
 A  K  V  R  D  T  V  T  I  V  F  K  N  L  A  S  R  P  Y  S  I  Y  V

     1410      1420      1430      1440      1450      1460      1470
ATGGAGTTTCCGTTTCAAAAGATGCAGAAGGAGCTGTTTATCCTTCAGATCCCAAAGAGAATATAACTCA
H  G  V  S  V  S  K  D  A  E  G  A  V  Y  P  S  D  P  K  E  N  I  T  H

     1480      1490      1500      1510      1520      1530      1540
TGGCAAAGCAGTTGAACCAGGACAGGTCTACACATATAAATGGACTGTGCTGGATACAGATGAACCTACA
  G  K  A  V  E  P  G  Q  V  Y  T  Y  K  W  T  V  L  D  T  D  E  P  T

     1550      1560      1570      1580      1590      1600      1610
GTAAAGGATTCTGAGTGCATTACTAAATTATATCATAGTGCTGTGGACATGACAAGAGATATTGCTTCAG
 V  K  D  S  E  C  I  T  K  L  Y  H  S  A  V  D  M  T  R  D  I  A  S

     1620      1630      1640      1650      1660      1670      1680
GACTTATTGGGCCACTTCTGGTTTGTAAACGCAAGGCACTCAGCATCAGGGGGGTACAGAATAAAGCTGA
G  L  I  G  P  L  L  V  C  K  R  K  A  L  S  I  R  G  V  Q  N  K  A  D

     1690      1700      1710      1720      1730      1740      1750
TGTGGAACAGCATGCAGTCTTCGCAGTGTTTGATGAAAACAAGAGCTGGTACTTGGAAGACAATATCAAG
  V  E  Q  H  A  V  F  A  V  F  D  E  N  K  S  W  Y  L  E  D  N  I  K

     1760      1770      1780      1790      1800      1810      1820
AAATACTGCAGCAATCCTTCCAGTGTTAAGAAAGATGACCCTAAATTTTACAAGTCCAATGTTATGTACA
 K  Y  C  S  N  P  S  S  V  K  K  D  D  P  K  F  Y  K  S  N  V  M  Y

     1830      1840      1850      1860      1870      1880      1890
CACTCAATGGCTATGCATCAGATAGAACAGAGGTTTGGGGGTTTCATCAGTCTGAAGTTGTTGAATGGCA
```

```
T  L  N  G  Y  A  S  D  R  T  E  V  W  G  F  H  Q  S  E  V  V  E  W  H

      1900       1910       1920       1930       1940       1950       1960
CCTCACCAGCGTAGGTACAGTGGATGAGATTGTTCCAGTACATCTTTCTGGTCACACCTTCTTATCCAAG
  L  T  S  V  G  T  V  D  E  I  V  P  V  H  L  S  G  H  T  F  L  S  K

      1970       1980       1990       2000       2010       2020       2030
GGAAAACATCAAGATATTTTAAATCTTTTTCCCATGAGTGGTGAATCCGCTACTGTAACAATGGACAATC
  G  K  H  Q  D  I  L  N  L  F  P  M  S  G  E  S  A  T  V  T  M  D  N

      2040       2050       2060       2070       2080       2090       2100
TAGGAACCTGGCTTCTGTCATCATGGGGCTCCTGTGAGATGAGCAATGGCATGAGATTGAGATTTTTGGA
L  G  T  W  L  L  S  S  W  G  S  C  E  M  S  N  G  M  R  L  R  F  L  D

      2110       2120       2130       2140       2150       2160       2170
                    ▼B
TGCCAATTATGATGATGAAGATGAGGGAAATGAAGAAGAGGAAGAAGATGATGGTGATATTTTTGCCGAC
  A  N  Y  D  D  E  D  E  G  N  E  E  E  E  E  D  D  G  D  I  F  A  D

      2180       2190       2200       2210       2220       2230       2240
ATTTTCAATCCTCCAGAAGTAGTAATAAAGAAAGAAGAGGTTCCCGTAAATTTTGTACCAGACCCAGAAT
  I  F  N  P  P  E  V  V  I  K  K  E  E  V  P  V  N  F  V  P  D  P  E

      2250       2260       2270       2280       2290       2300       2310
CGGATGCGCTAGCAAAAGAATTAGGATTATTTGATGACGAGGATAATCCAAAACAGTCACGCAGTGAACA
S  D  A  L  A  K  E  L  G  L  F  D  D  E  D  N  P  K  Q  S  R  S  E  Q

      2320       2330       2340       2350       2360       2370       2380
GACAGAGGATGATGAAGAACAGCTAATGATAGCTTCAATGCTTGGGCTTCGATCATTTAAGGGGTCAGTT
  T  E  D  D  E  E  Q  L  M  I  A  S  M  L  G  L  R  S  F  K  G  S  V

      2390       2400       2410       2420       2430       2440       2450
GCTGAAGAAGAATTGAAACACACAGCTCTAGCTTTAGAAGAAGATGCCCATGCTTCTGATCCTCGAATTG
 A  E  E  E  L  K  H  T  A  L  A  L  E  E  D  A  H  A  S  D  P  R  I

      2460       2470       2480       2490       2500▼A3    2510       2520
ACAGTAATAGTGCACATAATTCTGACGACATAGCTGGACGCTACCTGCGTACTATCTACCGCAGAAATAA
D  S  N  S  A  H  N  S  D  D  I  A  G  R  Y  L  R  T  I  Y  R  R  N  K

      2530       2540       2550       2560       2570       2580       2590
AAGGAGGTACTACATTGCAGCAGAAGAAGTTTTGTGGGACTACTCACCGATCGGAAAAAGTCAAGTGAGA
  R  R  Y  Y  I  A  A  E  E  V  L  W  D  Y  S  P  I  G  K  S  Q  V  R

      2600       2610       2620       2630       2640       2650       2660
AGTCTCCCAGCCAAGACCACATTCAAAAAAGCTATTTTCCGAAGTTATCTTGATGATACTTTCCAGACAC
 S  L  P  A  K  T  T  F  K  K  A  I  F  R  S  Y  L  D  D  T  F  Q  T

      2670       2680       2690       2700       2710       2720       2730
CTAGCACTGGAGGAGAATATGAAAAGCATCTTGGTATACTGGGTCCTATCATTAGGGCTGAGGTGGATGA
P  S  T  G  G  E  Y  E  K  H  L  G  I  L  G  P  I  I  R  A  E  V  D  D

      2740       2750       2760       2770       2780       2790       2800
TGTAATCGAAGTTCAGTTCAGAAATTTGGCCTCTAGACCATACTCACTTCATGCTCATGGCCTTCTCTAT
  V  I  E  V  Q  F  R  N  L  A  S  R  P  Y  S  L  H  A  H  G  L  L  Y

      2810       2820       2830       2840       2850       2860       2870
GAGAAATCTTCTGAAGGCAGAAGCTATGACGACAACTCTCCTGAATTGTTCAAAAAGGATGATGCTATCA
 E  K  S  S  E  G  R  S  Y  D  D  N  S  P  E  L  F  K  K  D  D  A  I

      2880       2890       2900       2910       2920       2930       2940
TGCCAAACGGCACATACACATATGTCTGGCAAGTCCCTCCACGGTCAGGACCAACAGACAATACAGAAAA
M  P  N  G  T  Y  T  Y  V  W  Q  V  P  P  R  S  G  P  T  D  N  T  E  K

      2950       2960       2970       2980       2990       3000       3010
ATGTAAATCATGGGCCTATTACTCTGGTGTAAATCCGGAAAAAGATATTCACTCTGGGCTTATTGGACCT
  C  K  S  W  A  Y  Y  S  G  V  N  P  E  K  D  I  H  S  G  L  I  G  P

      3020       3030       3040       3050       3060       3070       3080
ATTTTGATCTGCCAGAAAGGCATGATTGACAAGTACAACAGGACAATAGACATAAGGGAATTTGTCTTGT
 I  L  I  C  Q  K  G  M  I  D  K  Y  N  R  T  I  D  I  R  E  F  V  L

      3090       3100       3110       3120       3130       3140       3150
TTTTTATGGTCTTTGATGAGGAGAAAAGCTGGTACTTTCCAAAATCTGACAAAAGCACTTGTGAAGAGAA
```

```
F  F  M  V  F  D  E  E  K  S  W  Y  F  P  K  S  D  K  S  T  C  E  E  K

      3160       3170       3180       3190       3200       3210       3220
ACTTATAGGAGTCCAATCTCGCCACACATTTCCTGCAATTAATGGGATCCCTTATCAGCTGCAAGGCTTG
  L  I  G  V  Q  S  R  H  T  F  P  A  I  N  G  I  P  Y  Q  L  Q  G  L

      3230       3240       3250       3260       3270       3280       3290
ATGATGTACAAAGATGAGAATGTCCACTGGCATTTGCTGAACATGGGTGGGCCCAAAGATGTCCATGTTG
 M  M  Y  K  D  E  N  V  H  W  H  L  L  N  M  G  G  P  K  D  V  H  V

      3300       3310       3320       3330       3340       3350       3360
TTAATTTTCATGGTCAGACATTCACTGAAGAGGGAAGGGAAGATAATCAACTTGGAGTCCTTCCTCTTCT
V  N  F  H  G  Q  T  F  T  E  E  G  R  E  D  N  Q  L  G  V  L  P  L  L

      3370       3380       3390       3400       3410       3420       3430
TCCTGGTACATTCGCCTCCATCAAAATGAAACCATCCAAAATTGGCACATGGCTTTTAGAAACAGAAGTT
  P  G  T  F  A  S  I  K  M  K  P  S  K  I  G  T  W  L  L  E  T  E  V

      3440       3450       3460       3470       3480 C1    3490       3500
GGTGAAAATCAGGAAAGAGGAATGCAGGCTCTCTTTACTGTCATTGACAAAGATTGTAAATTACCAATGG
 G  E  N  Q  E  R  G  M  Q  A  L  F  T  V  I  D  K  D  C  K  L  P  M

      3510       3520       3530       3540       3550       3560       3570
GACTGGCAAGTGGGATAATACAAGACTCACAGATCAGTGCTTCAGGTCATGTTGGATATTGGGAGCCTAA
G  L  A  S  G  I  I  Q  D  S  Q  I  S  A  S  G  H  V  G  Y  W  E  P  K

      3580       3590       3600       3610       3620       3630       3640
GCTAGCAAGACTGAATAATACTGGAATGTTTAATGCTTGGAGCATCATAAAGAAGGAACATGAACATCCG
  L  A  R  L  N  N  T  G  M  F  N  A  W  S  I  I  K  K  E  H  E  H  P

      3650       3660       3670       3680       3690       3700       3710
TGGATCCAGATCGACCTACAAAGACAAGTTGTCATCACAGGCATTCAGACCCAGGGAACCGTGCACCTAC
 W  I  Q  I  D  L  Q  R  Q  V  V  I  T  G  I  Q  T  Q  G  T  V  H  L

      3720       3730       3740       3750       3760       3770       3780
TGAAACATTCGTATACTGTGGAATATTTTGTTACCTACAGCAAAGATGGGCAAAACTGGATTACTTTTAA
L  K  H  S  Y  T  V  E  Y  F  V  T  Y  S  K  D  G  Q  N  W  I  T  F  K

      3790       3800       3810       3820       3830       3840       3850
AGGAAGACATTCCAAAACACAAATGCATTTTGAGGGTAATTCAGATGGCACCACAGTAAAAGAAAACCACAT
  G  R  H  S  K  T  Q  M  H  F  E  G  N  S  D  G  T  T  V  K  E  N  H  I

      3860       3870       3880       3890       3900       3910       3920
TGATCCTCCTATTATTGCCAGATATATTAGGCTGCATCCAACCAAGTTCTACAACACACCTACTTTCCGCAT
  D  P  P  I  I  A  R  Y  I  R  L  H  P  T  K  F  Y  N  T  P  T  F  R  I

      3930 C2    3950       3960       3970       3980       3990
TGAACTGTTAGGTTGTGAAGTTGAAGGTTGCTCAGTGCCATTGGGAATGGAAAGTGGGGCTATCAAGGATT
  E  L  L  G  C  E  V  E  G  C  S  V  P  L  G  M  E  S  G  A  I  K  D

      4000       4010       4020       4030       4040       4050       4060
CAGAGATTACAGCCTCTTCTTATAAAAAGACTTGGTGGAGTTCATGGGAACCATTCCTTGCACGACTCAAT
S  E  I  T  A  S  S  Y  K  K  T  W  W  S  S  W  E  P  F  L  A  R  L  N

      4070       4080       4090       4100       4110       4120       4130
CTGAAAGGACGAACAAATGCTTGGCAACCAAAGGTAAACAACAAAGATCAATGGCTACAAATTGACCTGCAA
 L  K  G  R  T  N  A  W  Q  P  K  V  N  N  K  D  Q  W  L  Q  I  D  L  Q

      4140       4150       4160       4170       4180       4190       4200
CATCTTACAAAAATAACAAGCATAATAACTCAAGGAGCCACATCAATGACTACATCAATGTATGTGAAAACA
 H  L  T  K  I  T  S  I  I  T  Q  G  A  T  S  M  T  T  S  M  Y  V  K  T

      4210       4220       4230       4240       4250       4260       4270
TTCTCCATCCATTATACTGATGACAATTCAACATGGAAGCCTTATTTGGATGTTCGCACTTCCATGGAAAAG
 F  S  I  H  Y  T  D  D  N  S  T  W  K  P  Y  L  D  V  R  T  S  M  E  K

      4280       4290       4300       4310       4320       4330       4340
GTTTTCACAGGAAATATTAACAGTGATGGTCATGTCAAACATTTTTTCAACCCCCCTATATTGTCCAGGTTC
 V  F  T  G  N  I  N  S  D  G  H  V  K  H  F  F  N  P  P  I  L  S  R  F

      4350       4360       4370       4380       4390       4400       4410
ATTCGTATCATCCCTAAAACATGGAATCAATATATTGCACTCCGGATAGAATTGTTTGGTTGTGAAGTTT
 I  R  I  I  P  K  T  W  N  Q  Y  I  A  L  R  I  E  L  F  G  C  E  V
```

```
      4420       4430       4440       4450       4460       4470       4480
TTTAAGGCTTGGACAGAAGACTGTCAAATCAAGCAACTTCAATGTTTCAAGTTTTCTTATTACTAACTCTG
F  *
      4490       4500       4510       4520       4530       4540       4550
CTTTTTAAAAGGAAACAAAAACAAAAGCATAATAAAACTGTCTTAGCATAAAAAAAACTATCCTTCTCAAT

      4560       4570       4580       4590       4600       4610       4620
TTTCAGCCATAGCTTTCAAATAGCTTTGAAAATATCAATCAAAATATCATAACTGAAGTGACGTTTACAA

      4630       4640       4650       4660       4670       4680       4690
TGATTAATTCGTAGTGCCACGTTTAATCATGAATGTAATCCTAATACAATAAAATCGTTATTGTTTTTGCC

      4700
CCAAAAAAAAAAAAAAAAAA
```

Chapter 6

Conclusions

Objective 1:

The identification of proteins and protein expression patterns under given physiological conditions by proteomic analysis has gained fundamental importance for the functional study of cellular processes. To date, *Oxyuranus* species venom proteins have been studied within the context of whole venom including how these proteins interact within the venom as a whole, and the venom differences between similar species indicating possible reasons for the specific toxicities or the pharmaceutical differences of venoms. In this study a combination of protein separation methods and antigen binding studies were used to separate and characterise venom proteins.

The lethality of snake venom is caused by a complex interplay of venom components, including enzymes, peptides and specific toxins, and venoms from *Oxyuranus* showed remarkable complexity and stability. The frozen storage of whole venom over time (and absence of lyophilisation), and the dilution of whole and fractionated venom did not affect the electrophoretic patterning of the venom. In contrast, isolated protein components were less stable, despite the use of commercial inhibitors and crowding factors. The lack of whole venom degradation suggests there is potentially an undescribed endogenous inhibitor present within whole venom preventing the autolysis of this venom. From this theory, it may be suggested that venom proteins may become active upon envenomation due to the dilution of endogenous inhibitors away from toxins once within the body.

Future research to identify a potential endogenous inhibitor within the taipan venom together with the characterization of potential methods of toxin 'activation', for example, subtle changes in the environmental pH may potentially cause these proteins to activate, would make an interesting project in the future.

O. s. scutellatus and *O. s. canni* venoms shared similar if not identical protein patterning throughout electrophoretic, column fractionation and immunobinding analysis. The similarity between *O. s. scutellatus* and *O. s. canni* venoms was unexpected, as these species appear dissimilar, at least superficially, with differences in colour and pattern. Nevertheless, the venom similarities, together with previous comparative research using mitochondrial DNA sequencing and morphology studies indicated *O. canni* may not be a subspecies but is the same species; *O. scutellatus*. Despite, or perhaps due to this, the data did not resolve specific proteins causative of the varying coagulopathy reported of these species, nor did it provide a potential molecular explanation for the differences in death tolls between *O. s. scutellatus* and *O. s. canni* envenomations. It is possible that isomers of similar proteins could cause the differences in the severity of the signs from envenomation.

Further study of the subtle differences between the profiles of the non-antigenic peptides, the extra proteins banding described or the numerous protein isomers displayed may increase our understanding of the pharmacological differences between the Australian and Papuan taipan species. Alternatively, the variations of the clinical presentations described for humans envenomated by these species may be indicative of the length of time the individual is left untreated before medical or antivenom therapy is applied, causing an extension of observed signs and symptoms. There is also a potential that the people indigenous to PNG may have a slightly more susceptible physiology.

O. microlepidotus on the other hand, though sharing major proteins, revealed distinctly different patterns compared to *O. s. scutellatus*. Electrophoretic and chromatographic analysis indicating possible proteins responsible for the differences in toxicity of *O. microlepidotus* venom compared to the other species. The peptide disparities between the species were further enhanced by the absence of taipan antivenom binding of some 6 to 14 kDa proteins within *O. microlepidotus*. Despite the differences noted between *O. microlepidotus* and *O. s. scutellatus* venoms, the major protein bands within *O. microlepidotus* were antigenic against taipan antivenom (CSL). This result is reflected clinically where antivenom helps to resolve acute envenomation symptoms.

Within the bounds of this project, it was accepted the repetitive banding patterns and isomers observed with the venoms collected was a strong general trend for these snake species but, was by no means a definitive representation of the species as a whole. From venom similarities, together with research from mtDNA sequencing and morphology, several interpretations may be suggested. Most likely, it seems taipans underwent a recent genetic bottleneck (ie, reduced to a small population), and their present range is the result of recent range expansion. Other possibilities include introgressive hybridisation. Whatever the case, it was clear there was extensive genetic exchange between the New Guinea and Australian taipans very recently (tens rather than hundreds of thousands of years ago, most likely during the last major drop in sea levels during northern hemisphere ice ages, -15,000 years ago) (W.Wüster personal communication). This also indicates that diet, feeding and habitat (captive/wild) do not play a major role in contributing to the venom compositions of the taipan, opposing literature stating the contrary in other species (van der Weyden, *et al.,* 2000; Magro, *et. al*, 2001). Variation within the venom composition between the species may be due instead to a common ancestry producing similar venoms, the geographic distances and habitat pressures causing these variations

Despite the relative similarity of the *Oxyuranus* venoms reported in this study, antivenoms should still be raised from the venom of each *Oxyuranus* species across the entire target region. This is because an antivenom raised from the venom of one snake population has the potential to be less effective against the venom of another population. This suggestion is put forward due to the small sample size within this project potentially causing biased results, the intra-specific geographic variations reported in other snake species' venom compositions (it is logical to presume the variations described in other snake species may be found within *Oxyuranus*), and the variation of clinical observations after *Oxyuranus* envenomations. If possible, a low cost antivenom, produced specifically from *O. s. canni,* would make this currently rare commodity more easily accessible. The production of an antivenom not requiring refrigeration would also make it an invaluable tool for saving lives in Papua New Guinea

Objective 2:

In this study I have isolated and purified a proteolytic 300 kDa serine protease. To my knowledge, this protein has never been previously purified. This antigenic 300 kDa glycoprotein from *O. s. scutellatus* is comprised of a homo trimer, with 100 kDa subunits joined by disulphide bonds. This protein was separated on SDS-PAGE and the three overlapping subunits sequenced via N-terminal Edman degradation. This clean sequence indicated the three subunits were identical and matched a previously identified serine protease. It is unknown if one or all subunits contain active sites. Kinetics revealed this protein to undergo allosteric interactions.

Although an activator was not found for B2, there are several serine proteases that have been isolated from snake venoms effecting blood coagulopathy (Chow, 1998; Kini *et al*, 2001). Blood proteins such as thrombin-like factors, Factors VIIa and Xa are serine proteases with described allosteric interaction (Dittmar, *et al.*, 1997). Therefore it would not be unreasonable to suggest that this protein may have haemolytic effects in the body through allosteric interactions. Further characterisation of B2 would need to incorporate a broad range of physiological or chromogenic assays to address the many actions within the blood coagulation cascade. These could include studies such as platelet aggregation, D-dimer test, zymograms to test thrombin breakdown and the addition of ATP to enzyme assays to check if the activity is due to phosphorylation. Attempts were made to determine the DNA sequence encoding this protein from an *O. s. scutellatus* cDNA library using primers derived from the glycoprotein sequence through PCR. This technique failed to isolate the cDNA clones of interest.

O. s. scutellatus has already shown a propensity for producing large and complex haemolytic proteins (scutelarin) and future investigations are necessary to determine the biological significance and genetic sequence of this highly abundant protein. Indeed as this protein was so abundant in the venom it is assumed to play a prominent role, while possibly not a toxic one, perhaps in a supportive role. Further characterisation of this protein will provide an insight into the structure-activity relationship of this venom protease. From this a model could reveal a novel mechanism of action or a new structure within this well described protease group. Due to the importance of *O. s.*

scutellatus venom to antivenom production future description of this protein should be pursued.

Objective 3:

There are a growing number of nucleotide and peptide sequences reported thus far within elapid species, to which the sequences described within this report can be added. Large variations of sequences were represented within the venom, which raises more questions than answers. These sequences, submitted to the GenBank database, will add to a growing database of venom proteins. Future analysis and interpretation of these sequences will aid in extending knowledge and understanding of not only venom proteins but this data may provide new information in varying fields. Through access to an increasing number of sequence data across species, genus and even families, and ongoing expression or structural studies it may be possible to map specific evolutionary traits or gain an understanding of venom mechanisms in the future.

Throughout the sequence comparisons of the potential pre- and post-synaptic sequences, *Oxyuranus* maintained strong sequence and structural homology within the genus Elapidae. Alignments of isolated nucleotide and putative amino acid sequences indicated a common structure may be conserved with a highly repetitive cysteine residue structure represented across sequences. This signifies a structural constraint has been maintained during toxin evolution within specific groups and is required for their mechanism of action. Additionally, within the analysis undertaken using pre-synaptic neurotoxins, sequences conformed to the evolutionary model put forward within standing literature.

Despite the sequence and structural homology displayed between sequences, specific diversity was displayed with *Oxyuranus* pre-synaptic sequences in regard to residues required for toxicity. The 'toxic' residues identified within other species were not replicated within the *Oxyuranus* sequences indicating the evolutionary isolation and diversity of this genus. An understanding of these differing residues, through future mutational studies, may indicate whether specific residue characteristics or accumulated charges are the cause of the peptide and possibly membrane binding.

It is possible, though unlikely (see Chapter 3), that the PLA_2 nucleotide sequences determined within this study were a product of mutations from different *Oxyuranus* species due to environmental pressures and geographical isolation. These pressures may have altered the venom constituents giving rise to different proteins between each study and subsequent reports, similar to the the situation reported for *Notechis* species. For example, α-taipoxin and RW26 may not be isomers but may be a product of regional variation; RW26 may be the α-subunit within this sample. Until a detailed study is undertaken analysing differing field samples of the vast amount of PLA_2s from *Oxyuranus* venoms this question will not be resolved. Although viewing the precedents set by other Australian elapids it may be proposed with some confidence that the sequences reported are isomers of the established taipan toxins.

Future work to understand the specificities of toxin binding and the derivation of toxin sequences within *Oxyuranus* in comparison to other genera will aid in understanding possible common underlying mechanisms between toxins and receptors. Future phylogeny investigations of the genomic nucleotide, peptide and transit peptide sequences may also reveal relationships between the evolutionary process and functional divergence of *Oxyuranus*.

I hope the results of this study will stimulate more interest in the venoms from *Oxyuranus*. The medically important venomous Australo-Papuan snakes hold evolutionary, pharmaceutical and molecular mechanisms of significance. Future work involving this species will further an understanding of one of the most co-ordinated bio-weapons, yet to be matched by modern science.