In Silico Structural and Functional Analysis of Cobra Snake Venom Component- Kaouthiagin

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE IN BIOTECHNOLOGY

Submitted by: DOLA KHANDAKER

Student ID: 13136009

Biotechnology Program

Department of Mathematics and Natural Sciences

BRAC University

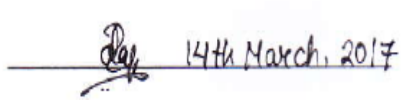
February 2017

Dedicated to

My parents and loved ones.

DECLARATION

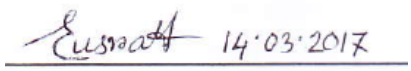
I hereby solemnly declare that the research work embodying the analysis and results reported in the following thesis entitled **"*In Silico* Structural and Functional Analysis of Cobra Snake Venom Component- Kaouthiagin"**, submitted by the undersigned has been carried out under the supervision of **Ms. Eusra Mohammad**, Lecturer, BSc. in Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University, Dhaka. It is further declared that the research work presented here is original and no part of this thesis has been submitted to any other institution for any degree or diploma.



(Dola Khandaker)

Candidate

Certified



(Ms. Eusra Mohammad)
Supervisor
Lecturer
BSc. In Biotechnology Program
Department of Mathematics and Natural Sciences
BRAC University, Dhaka

ACKNOWLEDGEMENTS

Above all, I wish to express my most sincere and utmost gratitude to the Almighty, to have blessed me with His mercy, and providing me with the strength, perseverance and understanding, needed for the successful completion of this thesis project.

My deepest appreciation and special thanks to **Professor A.A.Z. Ahmad,** the Chairperson of the Department of Mathematics and Natural Sciences, and **Professor Naiyyum Choudhury,** former coordinator of the Biotechnology and Microbiology Program, for giving me their exemplary guidance, endless provision and support, during my tenure as a student in BRAC University. I would also thank the authority of BRAC University for providing me with many opportunities and facilities.

I am truly indebted to my supervisor, Lecturer **Ms. Eusra Mohammad**, Department of Mathematics and Natural Sciences, BRAC University, for believing in me and allowing me to work on what I desired. She has my earnest appreciation. Throughout the journey of this research, Ms. Eusra has constantly encouraged me to explore new ideas and bring out the best from this project. Her helpful advice has refined and enhanced my capability and the success of this project immensely.

I am much obliged to acknowledge and express gratitude to **Ms. Abira Khan**, Lecturer, Department of Genetic Engineering and Biotechnology, University of Dhaka for aiding and providing me with valuable insights regarding the technical aspects of this project.

I am grateful to all the faculty members of the Department of Mathematics and Natural Sciences for their unwavering support and guidance throughout the entire period of my bachelor's degree.

Finally, I would like thank my parents for their sheer devotion to my education, future and happiness. They offered me their undying support, encouragement along with their patience. They are indeed, the most precious blessings in my life.

I perceive this opportunity as a big milestone in my career development. I will strive to use the gained skills and knowledge in the best way possible, and I will continue to work on their improvement, in order to attain my desired career objectives.

Sincerely,

Dola Khandaker

ABSTRACT

Snake venoms have been gaining a lot of attention in research, which has provided not only new tools to decipher molecular details of various physiological processes, but also the inspiration to design and develop many therapeutic agents, over the last few decades. Particularly, thrombosis and haemostasis are the major processes that are targeted by the snake venom proteins. Among them, anticoagulant proteins exhibit various enzymatic activities that interfere in the normal blood coagulation mechanisms. Such studies, enable us to fight unwanted clot formations and contribute to the treatment of cardiac arrests and strokes in patients with cardiovascular diseases, arteriosclerosis and hypertension. The ability of snake venom toxins to cause toxicity is associated with their high specificity and affinity for cell or tissues. This observation stimulated the development of many chemotherapeutic drugs based on snake venom toxins, which have the capacity to be highly cytotoxic. One of the targets investigated were integrins, which are cell surface receptors that play critical roles in cell adhesion and migration during cancer. A metalloproteinase, named kaouthiagin, from the venom of the snake *Naja kaouthia* was selected to study its structure-function relationships and mechanisms to comprehend its significance as a new anticoagulant and antagonist to integrins involved in cancer. Several *in-silico* tools have been used in this project to analyze the protein's structure and function relationships. Kaouthiagin has three specific domains: Zn^{2+} metalloprotease, disintegrin and an ADAM-CR domain. A predicted 3D structure was elucidated for this protease. This project provides an insight to how kaouthiagin could be a potential therapeutic as an anticoagulant and anticancer agent, given that further research is performed to prove its acceptance and efficiency.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND:

It is estimated that approximately 2,500,000 human beings suffer from snakebites per year, worldwide. 125,000 of which result in deaths. The most lethal snakebites predominantly occur in Africa and Asia (Gasanov, Dagda, & Rael, 2014). The location of Bangladesh, being in a humid and tropical zone, gives it a rich diversity in species of both flora and fauna. According to the report funded by UNDP and ICUN, Bangladesh has 125 known species of reptiles ("Snake Database," n.d.-a). Snakes in general, are considered a social and cultural threat. Due to the danger and lack of knowledge of handling these reptiles, the tremendous importance in biopharmaceutical research of snake venom has been compromised. Particularly in the rural areas of Bangladesh, snake bites are a major cause of mortality, which has a significant impact on human health and poses a substantial economic burden, due to the expenditures of its treatment. An epidemiological study estimated about 8000 snakebites per year in Bangladesh, with a mortality rate of 22%, which has been identified to be one of the highest in the world (Kadir et al., 2015). However, existence of these snakes is essential for conservation of the biodiversity. It is imperative to collect and combine all information on the indigenous snakes of Bangladesh and their venom components to be constructed into an organized database, because venom components and proteins have remarkable potentials in the biopharmaceutical industry due to their varied effects on different biological compounds and how they function. Many drugs derived from snake venom protein have been introduced in the market and many are still under development. Despite being rich in the diversity of snake species, Bangladesh still has a lot more to explore. Nonetheless, there are approximately 80 species of snakes found in Bangladesh, among which only few are venomous. Bites by green pitvipers (*Cryptelytrops erythrurus* and other species), cobras (*Naja* species) and kraits (*Bungarus*) are the most commonly identified ones in Bangladesh ("Snake Database," n.d.-a). The species of snakes found in Bangladesh are listed as follows, by their scientific names, according to the database of snakes in Bangladesh ("Snake Database," n.d.-a):

- *Naja naja*
- *Naja kaouthia*
- *Bungarus niger*
- *Bungarus caeruleaus*
- *Bungarus walli*
- *Bungarus lividus*
- *Bungarus fasciatus*
- *Callophis melanurus*
- *Callophis macclelellandi*
- *Ophiophagus hannah*
- *Hydrophis cantoris*
- *Typhlina bramina*
- *Hydrophis fasciatus*
- *Hydrophis stricticollis*
- *Hydrophis ornatus*
- *Trimeresurus albolabris*

- *Trimeresurus gramineus*
- *Trimeresurus erythrurus*
- *Trimeresurus popeorum*
- *Daboia russelii*
- *Ovophis monticola*
- *Protobothrops jerdonii*
- *Hydrophis nigrocinctus*
- *Laticauda laticaudata*
- *Laticauda colubrine*
- *Pelamis platurus*
- *Hydrophis obscurus*
- *Astrotia stokesii*
- *Hydrophis gracilis*
- *Hydrophis caerulescens*
- *Elaphe helena*
- *Hydrophis cyanocinctus*
- *Lapemis curtus*
- *Amphiesma stolata*
- *Macropisthidon plumbicolor*
- *Dendrolaphis tristis*
- *Ahaetulla nasutus*
- *Coelognathus helenus*
- *Boiga ochracea*
- *Boiga multomaculata*
- *Boiga trigonata*
- *Boiga cyanea*
- *Boiga cynodon*
- *Boiga gokool*
- *Enhydris dussimieri*
- *Enhydris sieboldi*
- *Enhydris enhydris*
- *Gerada prevostiana*
- *Fordonia leukobalia*
- *Elachistodon westermanni*
- *Crysopelea ornata*
- *Coluber mucosus*
- *Coluber nigromarginatus*
- *Lycodon aulicus*
- *Lycodon fasciatus*
- *Oligodon albocinctus*
- *Oligodon cinereus*
- *Oligodon taeniolatus*
- *Oligodon dorsalis*
- *Oligodon arnensis*
- *Oligodon theobaldi*
- *Oligodon cyclurus*
- *Sibynopis subpunctatus*
- *Sibynopis sagittaricus*
- *Liopeltis calamaria*
- *Xenochrophis cerasogaster*
- *Atretium Schistosum*
- *Argyrogena fasciolata*
- *Python reticulatus*
- *Python molurus*
- *Acrochordus granulatus*
- *Elaphe radiate*
- *Cerberus rinchops*
- *Eryx conicus*
- *Pareas moticola*
- *Dendrelaphis pictus*
- *Psammodynastes pulverulentus*

The species selected for this project was a local cobra snake, *Naja kaouthia*, to analyze and identify the probable therapeutic properties of its snake venom metalloprotease (SVMP), which is a component of its venom.

1.2 AIM OF THE PROJECT:

The project had the following objectives:

- Study the venom composition of *Naja kaouthia*
- Analyze the SVMP of *Naja kaouthia* for platelet aggregation, anticoagulant and antitumoral properties.
- Search for homologous and conserved regions of the sequence in other species.
- Find the functional motifs and annotate the sequence of the SVMP.
- Deduce a phylogenetic relationship from the homologous sequences.
- Predict a 3D structure of the SVMP and its binding site.
- Study the importance of its practical uses.

CHAPTER 2: *Naja Kaouthia*

2.1 *Naja Kaouthia:*

Naja kaouthia is locally known as "Gokhra Shap" and more commonly known as the monocellate cobra snake, referring to its monocellate hood pattern. It is a species of venomous cobra, belonging to the Elapidae family. They have an average size of 140 - 220cm, and in some locality they can grow up to 270cm. Their body is slender with oval shaped smooth scales, with a wide range of body colours including dark yellow, brown, light brown, reddish-brown, black with reddish or greyish tint. The females lay a clutch of 8 to 45 eggs and incubate them for a period of 60 days, till hatching ("Snake Database," n.d.-a). Young ones mostly feed on amphibians, whereas the adults prey on small mammals, fishes and other snakes. Like other cobras, its major toxic component include postsynaptic neurotoxins. However, the neurotoxins in this species are considerably weaker than other species belonging to this family (Ogay, Rzhevsky, Murashev, Tsetlin, & Utkin, 2005).

Figure 2.1.1: Images of the monocellate cobra snake, *Naja kaouthia*. [Taken from the Snake Database of Bangladesh ("Snake Database," n.d.-b)]

2.2 TAXONOMY OF *Naja Kaouthia*:

Taxonomy ("Naja kaouthia (Monocled Cobra)," n.d.)

- Kingdom : Animalia
  - Phylum : Chordata
    - Sub-phylum : Vertebrata
      - Class : Reptilia
        - Order : Squamata
          - Sub-order : Serpentes
            - Family : Elapidae
              - Genus : *Naja*
                - Species : *Naja kaouthia*

2.3 DISTRIBUTION OF *Naja Kaouthia*:

The monocellate cobra are widespread among South and Southeast Asian countries. They are distributed throughout India, west China, Vietnam, Cambodia, Malay Peninsula, Bhutan, Myanmar, Laos, Nepal and are also native to Bangladesh (A. K. Mukherjee & Maity, 2002; "Naja kaouthia | The Reptile Database," n.d.). *Naja kaouthia* is expected to occur throughout the country and to cause the majority of cobra bites. They prefer wet habitats, but are also found in grasslands and forests. It is the only species of *Naja* found in southeastern Bangladesh (Chittagong District, Cox's Bazar District, and the three Chittagong Hill Tract districts) ("Snake Database," n.d.-b).

Figure 2.3.1: Map illustrating the geographical distribution of *Naja kaouthia* ("Naja kaouthia | The Reptile Database," n.d.).

2.4 SNAKE VENOM COMPOSITION OF *Naja Kaouthia*:

Snake venoms are a mixture of complex pharmacologically active biological proteins and different polypeptides that target vital physiological processes (Kini & Koh, 2016). It is the venom components that incapacitates, immobilizes and even digests the prey. Over 90% of snake venom components are proteins that bring about pharmacological effects on their victims, which are mostly enzymes and the rest are non-enzymatic. Non-protein components, which comprise about 5-10% of the venom, include carbohydrates, lipids, free amino acids, nucleotides and metals in the form of glycoprotein or metalloproteins (Rodnight, 1979). Key venom components include neurotoxins, myotoxins, cardiotoxins, hematoxins and catalytic enzymes (Chaisakul, Hodgson, Kuruppu, & Prasongsook, 2016).

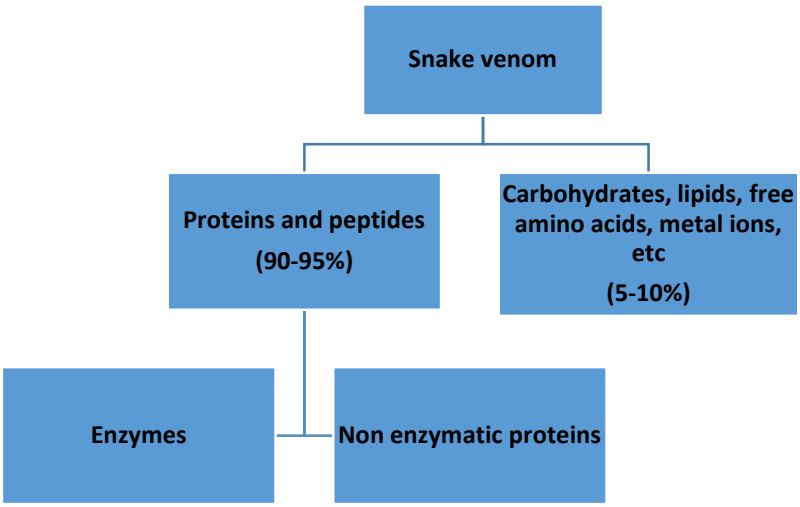


Figure 2.4.1: Snake venom composition

Naja kaouthia venom contains neurotoxins, cytotoxins, myotoxins and cardiotoxins. Signs of bites exhibit pain and swelling at bite site often followed by blistering and extensive necrosis. Neurotoxic symptoms may include ptosis, drowsiness, dysphagia, dysphonia, and generalized weakness. There is a high mortality rate following envenomation of this species ("Snake Database," n.d.-b). The polypeptide toxins and enzymes, especially hydrolases make up 25-70% of elapid venoms. These enzymes include digestive hydrolases; L-amino acid oxidase; phospholipases; thrombin-like pro-coagulant, and metalloproteinase (hemorrhagins), that break down biological molecules including proteins; nucleic acids and phospholipids, and are responsible for almost all of the biological effects on the prey (Mahanta & Mukherjee, 2001). Neuromuscular and circulatory systems are the two main physiological systems that are targeted by these toxins, as interruptions in these systems make the prey succumb to the venom very fast (Kini & Koh, 2016; Mahanta & Mukherjee, 2001).

The list of the major venom toxins, found in *Naja kaouthia*, and their functions is tabulated below with their accession numbers:

Table 2.4.1: Toxins found in *Naja kaouthia* and their functions. [Information collected from Snake Database of Bangladesh ("Snake Database," n.d.-b)]

Name of the component	Function	Accession ID
Cytotoxin 2	Shows cytolytic activity on many different cells by forming pores in lipid membranes. In vivo, increases heart rate or kills the animal by cardiac arrest.	Q9DGH9
Cytotoxin 1	Produces complete blockade of auricular contraction, which is irreversible at high concentrations. Induces apoptosis in leukemic cells. Possess anti-arthritic and anti-inflammatory potential.	P0CH80
Cytotoxin 4	Shows cytolytic activity on many different cells by forming pore in lipid membranes. In vivo, increases heart rate or kills the animal by cardiac arrest.	P60303
Weak toxin CM-9a	Binds with low affinity to muscular and very low affinity to neuronal (alpha7) nicotinic acetylcholine receptor	P25679
Phospholipase A2 inhibitor	Inhibits the enzymatic activity of phospholipase A2.	Q7LZI2
Tryptophan-containing weak neurotoxin	Modulates both nicotinic and muscarinic acetylcholine receptors on nAChR, acts as an orthosteric (competitive) antagonist. On mAChR, in presence of ACh, it partially inhibits the effect of ACh (allosteric antagonist), whereas in the absence of ACh, it activates the receptor (allosteric agonist). Exerts a myorelaxant effect, induces a dose-dependent decrease in blood pressure and an increase in hear rate in mice and rats.	P82935
Cobrotoxin-c	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. Binds to the nicotinic acetylcholine receptor.	P59276

Muscarinic toxin-like protein 1	Binds weakly to the muscarinic acetylcholine receptor	P82462
Elapitoxin-Nk2a	Causes paralysis by preventing acetylcholine binding to the nAChR. In mice lung cancer, causes reduction of tumour growth	P01391
Cytotoxin 5	Shows cytolytic activity on many different cells by forming pore in lipid membranes. In vivo, increases heart rate or kills the animal by cardiac arrest.	P24779
Acidic phospholipase A2 2	PLA2 catalyses the calcium-dependent hydrolysis of the 2-acyl groups in 3-sn-phosphoglycerides.	P00597
Cobrotoxin	Binds to muscle nicotinic acetylcholine receptor (nAChR and inhibits acetylcholine from binding to the receptor, thereby impairing neuromuscular transmission.	P60771
Cobrotoxin II	Binds to muscle nicotinic acetylcholine receptor (nAChR) and inhibits acetylcholine from binding to the receptor, thereby impairing neuromuscular transmission.	P82849
Cobra venom factor (Zn metalloproteinase cobrin)	Complement-activating protein in cobra venom. It is a structural and functional analogue of complement component C3b, the activated form of C3. It binds factor B (CFB), which is subsequently cleaved by factor D (CFD) to form the bimolecular complex CVF/Bb. CVF/Bb is a C3/C5 convertase that cleaves both complement components C3 and C5. Structurally, it resembles the C3b degradation product C3c, which is not able to form a C3/C5 convertase. Unlike C3b/Bb, CVF/Bb is a stable complex and completely resistant to the actions of complement regulatory factors H (CFH) and I (CFI). Therefore, CVF continuously activates complement resulting in the depletion of complement activity.	Q91132
Cysteine-rich venom protein kaouthin-1	Inhibits calcium-activated potassium channels (KCa), voltage-gated potassium channel (Kv), and the calcium release channel/ryanodine receptor.	P84805

Disintegrin trigramin-beta-2	Inhibits fibrinogen interaction with platelets. Acts by binding to the alpha-IIb/beta-3 receptor (ITGA2B/ITGB3) on the platelet surface and inhibits aggregation induced by ADP, thrombin, platelet-activating factor and collagen.	P17495
Acidic phospholipase A2 1	PLA2 catalyses the calcium-dependent hydrolysis of the 2-acyl groups in 3-sn-phosphoglycerides.	P00596
Phospholipase A2 inhibitor	Inhibits the enzymatic activity of phospholipase A2.	Q7LZI1
Snake venom metalloproteinase-kaouthiagin	Snake venom zinc protease that inhibits haemostasis by binding and cleaving the vWF in humans. Has also and inhibitory effect on the collagen-induced platelet aggregation.	P82942
Cobrotoxin-b	Produces peripheral paralysis by blocking neuromuscular transmission at the postsynaptic site. Binds to the nicotinic acetylcholine receptor.	P59275
Short neurotoxin 1	Binds to muscle nicotinic acetylcholine receptor (nAChR) and inhibits acetylcholine from binding to the receptor, thereby impairing neuromuscular transmission.	P14613
Muscarinic toxin-like protein 3	Antagonist of muscle and neuronal nicotinic acetylcholine receptors (nAChR) with highest affinity for neuronal alpha-7 nAChRs.	P82464

CHAPTER 3: APPLICATIONS OF SNAKE VENOM

3.1 APPLICATIONS OF SNAKE VENOM:

Despite being highly toxic, from the previous section it is evident that snake venom consists of a vast array of biologically active compounds, which target an immense number of receptors, membrane proteins as well as coagulation proteins. Thus, they can be used as:

- Therapeutic agents
- Research tools for use in the diagnosis of several diseases
- For understanding the physiological and pathological changes it brings about in an organism

Over the past few decades, investigations have shown that a myriad of snake venom proteins have the potential to be therapeutic agents to be used for the treatment of various medical issues, such as cardiovascular ailments, thrombosis, arthritis and even cancer, along with many other diseases. Many of them have been employed as research tools to study the coagulation cascade and all its related factors and proteins that are involved in the process. Venom toxins have a high specificity for their target molecules, which makes them valuable for drug development. Studying venom proteins have also contributed to the knowledge about various molecular mechanisms involved in the physiological processes. The focus on snake venom has led to the development of novel therapeutic compounds (Ashis K Mukherjee, Saikia, & Thakur, 2011). The following table shows a list of snake venom proteins that have been put to medical uses according to their functions:

Table 3.1.1: Therapeutic applications of snake venom and snake venom components. [Taken from (Ashis K Mukherjee et al., 2011)]

Snake venom Component	Example	Source	Biological functions	Applications
Snake venom thrombin like enzymes	Ancrod	Agkistrodon rhodostoma	Therapeutic defibrination	Treatment of ischaemic stroke, HATT syndrome, deep vein thrombosis

Plasminogen activating enzymes	TSV-PA	Trimeresurus stejnegeri	Dissolution of fibrin clot via activation of plasminogen to plasmin	Treatment of vascular diseases, cancer
Direct fibrinolytic enzymes	Alfimeprase	Agkistrodon contortrix contortrix	Dissolution of fibrin	
Disintegrin	Contortrostatin	Agkistrodon contortrix contortrix	Blocks integrins during tumor progression	Applicable as antitumor agents.
	Rhodostatin	Calloselesma rhodostoma	Inhibits angiogenesis	
Platelet glycoprotein IIb/ IIIa antagonists	Integrilin	Sisturus miliarius	Inhibits platelet aggregation	For reducing the risk of acute cardiac diseases.
Thrombin inhibitors	Bothrojaracin	Bothrops jaraca	Anticoagulant	
Plasmin inhibitors	Textilinin-1	Pseudonaja textilis	Inhibitor of plasmin catalyzed fibrinolysis	Anti-bleeding agent in the treatment of chronic inflammatory diseases such as rheumatoid arthritis and arteriosclerosis.

CHAPTER 4: ANALYSIS OF SNAKE VENOM METALLOPROTEASE

4.1 METALLOPROTEASE:

Metalloproteinase or metalloprotease is a protease enzyme that needs a metal counterpart to carry out its function. The metal ion is organized with the enzyme protein with three ligands, which varies according to the amino acid residue involved (Rawlings & Barrett, 1995). They are widely distributed from bacteria to mammals (Ito et al., 2001). For example, meltrin, is a metalloproteinase which plays an important role in the fusion of muscle cells during embryo development, during the process of myogenesis (Abe, Mocharla, Yamate, Taguchi, & Manolagas, 1999). Human fertilin beta contains pro-metalloprotease-like, disintegrin-like, cysteine-rich, epidermal growth factor-like (EGF) repeat, transmembrane, and cytoplasmic domains. Due to this domain organization, human fertilin beta has been identified as a member of the ADAM family, which is composed of membrane-anchored proteins having A Disintegrin And Metalloprotease domain (Blobel, 2000). These domains will be discussed and elaborated later in this report. Metallproteases can be classified into two groups—metalloendopeptidases and metalloexopeptidases (Rawlings & Barrett, 1995).

4.2 SNAKE VENOM METALLOPROTEASE OR METALLOPROTEINASE (SVMP):

Metalloproteinases are important and among the most abundant compounds in most snakes of the viperid and crotalid families (Calderon et al., 2014; Sarray, Luis, Ayeb, & Marrakchi, 2013). SVMPs are endoproteolytic enzymes that require divalent cations, such as Zn^{2+} or Ca^{2+} ions for their enzymatic activity or structural conformation (Ito et al., 2001). Mostly, they are Zn^{2+} dependent, thus they are called monozinc endopeptidases, varying in size from 20-100kDa (Sarray et al., 2013). They are closely related to ADAM (A Disintegrin And Metalloproteinase) family of proteins (Kini & Koh, 2016). Structural analysis of the SVMPs isolated revealed that they can be classified into 3 categories, based on size and domain organization: P-I, P-II, P-III. P-I are the simplest class of SVMPs, with low molecular mass, that contain only a metalloproteinase domain. Class P-II has a metalloprotease and a disintegrin domain at the C terminal of the metalloprotease domain. P-III class is a group of higher molecular mass compounds, which have a metalloprotease domain; a disintegrin-like domain followed by a cysteine-rich domain. Previously called P-IV, the heterotrimeric class of SVMPs that contain an extra snake disulphide linked C-type lectin-like domain, is now included in the P-III group as its subclass (Ito et al., 2001; Q. Lu, Clemetson, & Clemetson, 2005; Sarray et al., 2013).

A variety of SVMPs have been isolated and found as a haemorrhagic factor, specific activator, or the inhibitor of coagulation factors or the inhibitor of platelet aggregation (Ito et al., 2001). They selectively cleave a small number of key proteins involved in the blood coagulation cascade and in platelet aggregation. Such limited proteolysis leads to either activation or inactivation of the protein involved in the process, thus resulting in haemorhhagic, pro-coagulant, anticoagulant and antiplatelet effects (Kini, 2006; Kini & Koh, 2016). As procoagulant proteases, SVMPs activate only two key coagulation factors- factor X (FX) and prothrombin to exhibit their procoagulant effects. Metalloproteases, that are prothrombin activators, convert prothrombin to meizothrombin. Antagonistically, for fibrinolytic effects, fibrinogen is cleaved by metalloproteases. SVMPs selectively cleave the Aα chain of fibrinogen, but does not cleave Bβ and γ chains and are thus classified as α- fibrinogenases. Some of these fibrinogenases inhibit platelet aggregation. Catalytic cleavage in SVMPs is through Zn^{2+} ions coordinated by three His side chains and a water molecule anchored to a conserved Glu residue. This water molecule acts as the general base that catalyzes the peptide bond cleavage (Kini & Koh, 2016). SVMPs also show anticancer activities, which involve pro-inflammatory effect and apoptotic activity (Chaisakul et al., 2016). It was reported that several SVMPs inhibited integrin-mediated adhesion of cancer cells on extracellular matrix proteins as described in the following table.

Table 4.2.1: SVMPs affecting tumor cells. [Table taken from (Sarray et al., 2013)]

Proteins	Snake	Integrins	Effects
VAP1, VAP2	*Crotalus atrox*	α3,α6,β1	Induce apoptosis of HUVEC
HV1	*Trimeruserus flavoviridis*	-	Inhibits adhesion of HUVEC and induces apoptosis
Halysase	*Gloydius halys*	α1β1;α5β1	Inhibits proliferation and Induces apoptosis of HUVEC
VLAIPs	*Vipera lebetina*	-	Inhibits proliferation and Induces apoptosis of HUVEC
Graminelysin	*Trimeresurus gramineus*	α1β1;α5β1	Inhibits proliferation and Induces apoptosis of HUVEC
BaG	*Bothrops alternatus*	α5β1	Inhibits adhesion of K562 cells
TSV-DM	*Trimeresurus stejnegeri*	-	Inhibits cell proliferation and induces transient cell morphologic changes of endothelial cells.

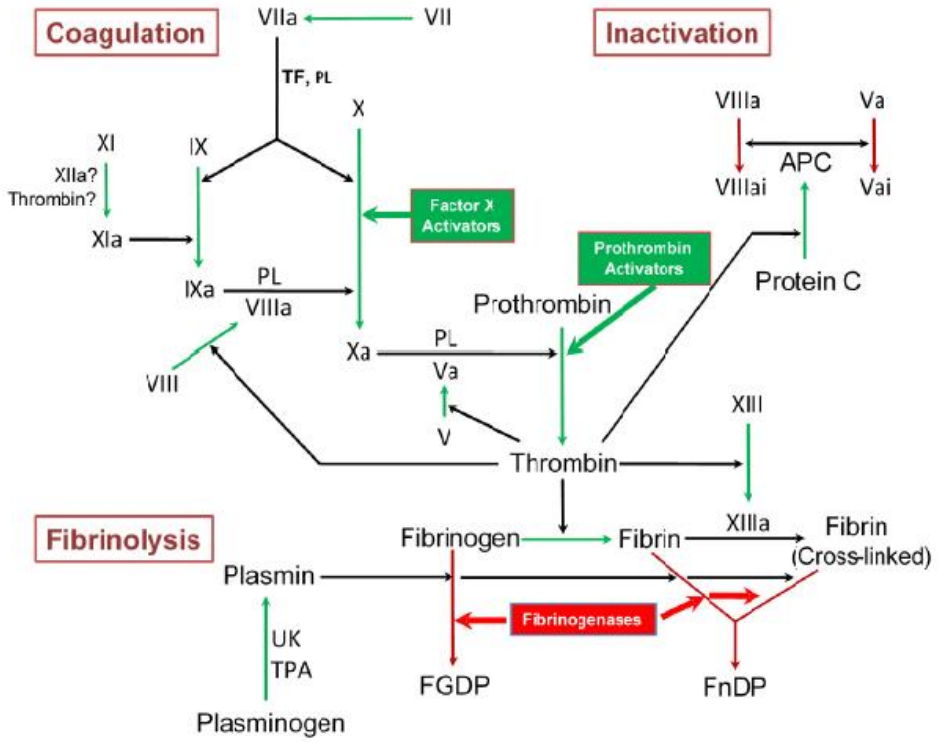


Figure 4.2.1: SVMPs affecting blood coagulation. Proteolytic activity interferes with specific activators (thick arrow head). Green boxes- procoagulant SVMPs; red boxes- fibrinogenases that cleave fibrinogen and fibrin; APC- activated protein C; FGDP- fibrinogen degradation products; FnDP- firbin degradation prodcucts; PL- phospholipids; TF- tissue factor; TPA- tissue plasminogen activator; UK- urokinase. [Figure taken from (Kini & Koh, 2016)]

4.3 DISINTEGRINS:

Disintegrins are a family of non-enzymatic, low molecular weight (fewer than 100 amino acids), cysteine-rich and Arg-Gly-Asp (RGD) residue containing peptides, found in snake venom (Ito et al., 2001; Q. Lu et al., 2005). Snake venom disintegrins are mostly derived from proteolytic precursors of SVMPs (Selistre-de-Araujo, Pontes, Montenegro, & Martin, 2010). Disintegrins can be fit into five different categories, according to their length and number of disulfide linkages present. The first group comprises of short disintegrins, composed of 49-51 amino acids with four disulfide bridges. The second group includes medium sized disintegrins with about 70 amino acids and six disulfide bridges. The third group includes long disintegrins of 83 amino acids with seven disulfide bridges. Disintegrins belonging to the fourth group are the domains of P-III class SVMPs, being composed of approximately 100 amino acid residues

with 16 Cys residues which are involved in the formation of eight disulfide bonds. The aforementioned groups of disintegrins are single chain molecules, unlike the fifth group which is composed of homo and heterodimers. The dimeric disintegrin subunits contain about 67 residues with four disulfide intra-chain bridges and two inter-chain bridges (Sarray et al., 2013).

Originally disintegrins were identified as inhibitors of platelet aggregation and were successively shown to be antagonistic to fibrinogen binding to platelet integrin, αIIbβ3 (platelet membrane glycoprotein) (Ito et al., 2001; Sarray et al., 2013). The RGD sequence is highly mobile, which allows rapid binding to the integrin binding site within the 217-302 residues of GPIIIa (Q. Lu et al., 2005). Disintegrins potently block the binding of fibrinogen and VWF (von Willebrand factor) to GPIIb/IIIa complexes in ADP or thrombin activated platelets (Calderon et al., 2014; Chaisakul et al., 2016). VWF is a multimeric protein, found in blood plasma, which is essential for platelet adhesion to the damaged subendothelial matrices to form a haemostatic plug (Bergmeier & Hynes, 2012). P-III class metalloproteases are significantly more haemorrhagic than the P-I class, due to the presence of a disintegrin domain (Calderon et al., 2014; Hammouda et al., 2016). They have also been found to interact with integrins α5β1 and αvβ3lls, expressed by a number of cells including those involved in tumor development and proliferation (Calderon et al., 2014; Chaisakul et al., 2016). Integrins are heterodimeric transmembrane proteins linked by non-covalent bonds between α and β subunits. They play important roles to promote the major mechanisms during tumor development, including cell-extracellular matrix interaction; cytoskeleton organization; signal transduction; epithelial cell adhesion; growth; proliferation; invasion and migration. Mechanisms involved behind the antitumour functionality of disintegrins include inhibition of cell adhesion; attenuation of cancer cell migration, invasion of normal cells and antimetastatic activity (Chaisakul et al., 2016; Sarray et al., 2013). Thus, disintegrins isolated from several snake venoms, have revealed potentials of uses not only to treat cardiovascular diseases, but also as an effective inhibitor of integrin and in turn tumour cells. Studies on the peptides of metalloproteases and disintegrins, containing the RGD sequence, have been proven to act on the extracellular matrix and have antitumour effects involving angiogenesis and cancer metastatic dissemination. A number of RGD containing snake venom disintegrins have also been used to illuminate target receptors in a wide variety of primary cultured tumour cells (Chaisakul et al., 2016). A number of purified disintegrins have been demonstrated for their potential to inhibit platelet aggregation and

abolish cancer growth, and led to the development of new therapeutic agents for arterial thrombosis; osteoporosis; angiogenesis and metastasis (Sarray et al., 2013).

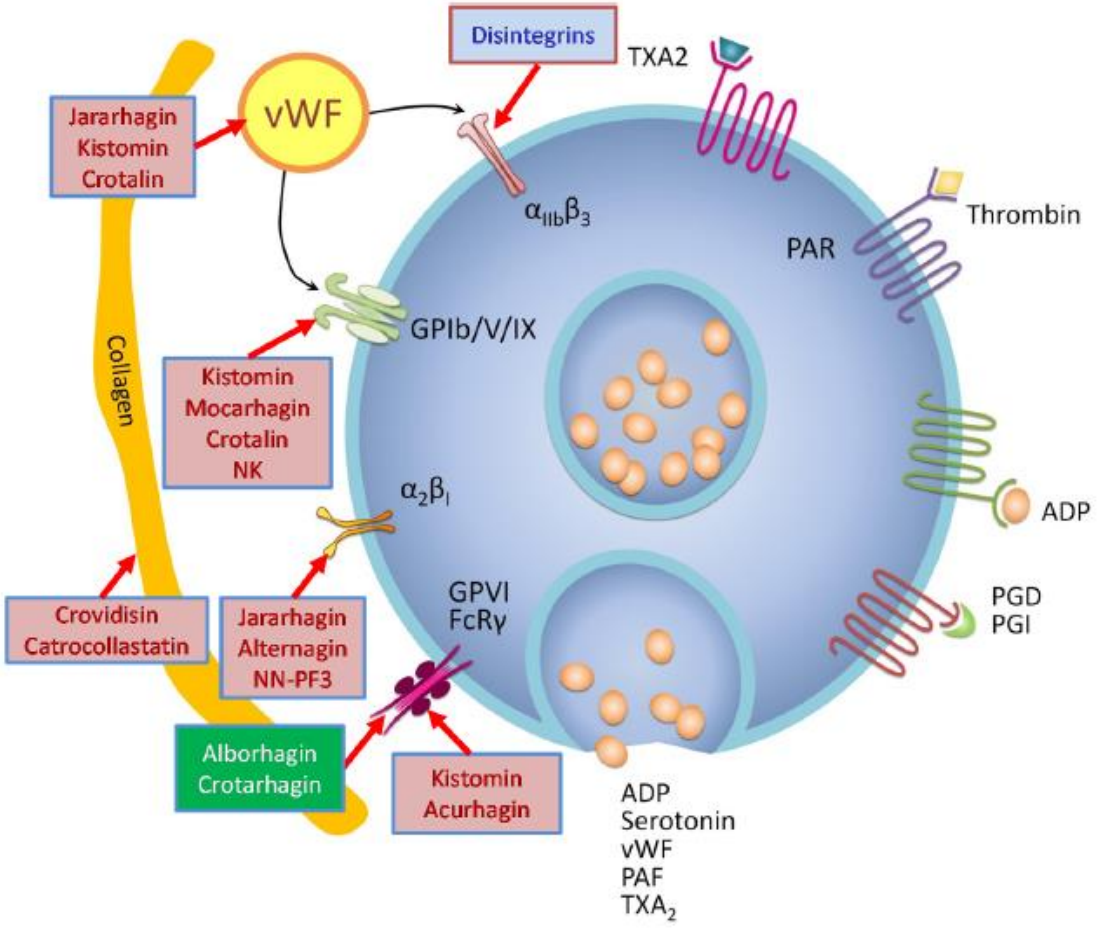


Figure 4.3.1: SVMPs affecting platelet aggregation. Proteases that induce or inhibit platelet aggregation are shown in green or red boxes, respectively. Disintegrins that inhibit platelet aggregation are shown in the blue box. PAF- platelet activating factor; PAR- protease activated receptor; PGD- prostaglandin D; PGI- prostaglandin I; TXA2- thromboxane A_2. [Figure taken from (Kini & Koh, 2016)]

Table 4.3.1: Effects of disintegrins on cancerous cells [Taken from (Sarray et al., 2013)]

Proteins	Snake	Integrins	Effects
Triflavin	*Trimeresurus flavoviridis*	α5β1,αvβ3, α3β1	Inhibits adhesion of tumor cells to matrix proteins, cell migration and angiogenesis *in vitro* and *in vivo*
Rhodostomin	*Agikistrodon rhodostoma*	αvβ3,αvβ5	Inhibits cell migration, invasion of endothelial cells; inhibits angiogenesis *in vivo* and *in vitro*
Contortrostatin	*Agkistrodon contortrix contortrix*	αvβ3,α5β1, αvβ5, αIIββ3	Blocks adhesion, migration invasion of different type of tumor cells
Lebestatin	*Macrovipera lebetina*	α1β1	Inhibits migration and angiogenesis
Accurhagin-C	*Agkistrodon acutus*	αvβ3	Prevents migration and invasion of endothelial cells; anti-angiogenic activity *in vitro* and *in vivo*; elicites anoïkis
Eristostatin	*Eritocophis macmahoni*	α4β1,other integrin not yet determined	Inhibits cell motility; no effect on cell proliferation or angiogenesis
DisBa-01	*Bothrops alternatus*	αvβ3	Anti-angiogenic and anti-metastatic effect on melanoma cells
Leberagin-C	*Macrovipera lebetina*	αvβ3	Inhibits cell adhesion of melanoma tumor cells
Accutin	*Agkistrodon acutus*	αvβ3	Inhibits angiogenesis *in vitro* and *in vivo*; induces apoptosis

4.4 FINDINGS FROM THE SVMP IN *Naja Kaouthia*:

The SVMP protein named kaouthiagin, has been found in the venom of the cobra snake *Naja kaouthia*.

Kaouthiagin has the following characteristics:

- It falls under the group of the higher molecular mass metalloprotease P-III class, which comprises of a metalloprotease domain in the N-terminal; a disintegrin like domain and a Cys-rich domain in the C-terminal.
- The protein sequence of kaouthiagin is 401 amino acid residues in length.
- The metalloprotease has a zinc-binding motif (HEXXHXXGXXH).
- Uniquely, kaouthiagin has an extra disintegrin-like sequence (RGD) in its Cys-rich domain.
- It is sensitive to EDTA or *o*- phenanthroline, treatment with which makes kaouthiagin lose its proteolytic activity.
- Kaouthiagin is an endopeptidase (Ito et al., 2001).

Kaouthiagin has the following known specific functions:

- Kaouthiagin binds to and cleaves VWF at a peptide bond between Pro708 and Asp709, resulting in the loss of the platelet and collagen binding activities of VWF by degrading the multimeric structure of VWF.
- The additional disintergrin-like sequence in the Cys-rich domain has an inhibitory effect on platelet aggregation (Ito et al., 2001).
- Kaouthiagin cleaves GPIbα, which is a platelet membrane glycoprotein (Kini & Koh, 2016).

CHAPTER 5: SOFTWARE TOOLS AND METHOD USED TO ANALYSE KAOUTHIAGIN

5.1 PROTEIN SEQUENCE:

The target protein sequence was obtained from GenPept database, via NCBI, in FASTA format.

5.2 PROTEIN CHARACTERISTIC ANALYSIS:

The sequence of the protein in question was analyzed with ProtParam.

- ProtParam- This tool allows us to analyze and compute various physical and chemical parameters of a given protein, which includes molecular weight; isoelectric point; atomic composition; amino acid composition; estimated half-life; instability index and aliphatic index and GRAVY.

5.3 HOMOLOGY:

Homology signifies the condition of being homologous, that is, to be similar in sequence or structure. Homology between different species can help us observe their functional relatedness. If sequences are homologous, we can say that they belong to a certain protein family and serve the same function. The tools used to find out the homology of different species with reference to the kaouthiagin SVMP were PSI-BLAST and PRALINE.

- PSI-BLAST- Position Specific Iterated Basic Local Alignment Search Tool, is a tool in NCBI that gives results based on BLOSUM62 matrix. The BLOSUM matrix is a substitution matrix that is used for sequence alignment of protein sequences that are evolutionarily divergent. They are based on local alignments. With the help of PSI-BLAST we can find out the similarities between distant related proteins of different species. Iterations or repetitive searching and adjusting the scoring matrix of the alignments, brings out the homologous matching sequences significant to our target sequence as well as among other proteins.

  Up to 3 iterations were performed with maximum 500 entries. Entries were selected according to Identity match percentage of greater than and equal to 65%. For the first iteration, 10 entries were selected, 12 were selected for the second and finally 15 entries selected for the third iteration, each time including the target protein sequence.
- PRALINE- PRofile ALigNEment is a fully customizable multiple sequence alignment application. PRALINE can integrate information from database homology searches to

generate a homology-extended multiple alignment. The web tool is designed to facilitate the comprehensive visualization of the generated alignments by means of five default colour schemes based on: residue type, position conservation, position reliability, residue hydrophobicity and secondary structure, depending on the options set. The purpose of using this tool was to calculate the best match for the selected sequences from PSI-BLAST and line them up to identify the similarities and differences in detail.

Selected protein sequences from PSI-BLAST were inserted in their FASTA format to perform the MSA through PRALINE. BLOSUM62 exchange weight matrix was used and the progressive alignment strategy was kept to "PSI-BLAST pre profile processing (Homology-extended alignment)". The other program parameters were as set default.

5.4 PROTEIN MOTIF:

Structural motifs of proteins are short segments of its 3D structure which are usually spatially close but not necessarily adjacent to each other. Motifs may be conserved in many different proteins and their role may be structural or functional. Therefore, we can say that motifs are different functional domains of a protein. Identifying these domains allows us to identify known and unknown, distantly related proteins and how they are related. Motifs arise among different proteins due to the particular requirements of specific regions of the proteins which may be important structurally or functionally. For example the residues of an enzyme active binding site, on which the enzyme activity depends.

The SVMP was analysed for its protein motifs by the following softwares:

- Pfam- The Pfam database shows a large collection of protein families, which are represented by multiple alignment sequences and Hidden Markov Models (HMMs). HMMs are considered as a generalized version of a mixture of models with hidden or latent variables, which control the mixture component selected for each observation and are rather independent of each other. The data presented on Pfam is based on the *UniProt Reference Proteomes,* after an accession number of the protein sequence is inserted. Pfam shows the domains of the protein sequence and their boundaries. Clicking on specific domains gives an elaborate description of the domain.

- InterProScan- This tool allows us to search for a protein, inserted in FASTA format that matches against sequences based on the InterPro Protein Signature database. InterPro enables us to analyse the functionality of proteins by classifying them into specific families and predicting domains. They combine protein signatures from a number of databases into a single resource tool, producing a powerful, integrated database and diagnostic tool.

5.5 PHYLOGENETICS:

Phylogenetics, or evolutionary biology is used to study evolutionary history of an organism and their relationships with other groups of organisms or populations. Construction of a phylogenetic tree can be done with various in-silico tools, which involve computational approaches to implement the methods of parsimony and maximum likelihood. For this project, a phylogenetic tree was established using PHYLOGENY.FR.

- Phyolgeny.fr- Is a tool that is dedicated to reconstruct and analyse phylogenetic relationships from different bioinformatics programs to create a robust phylogenetic tree from a set of different molecular sequences. The 15 protein sequences were copy-pasted in their FASTA format to construct the phylogenetic tree by the "One-click" method on the server menu, where the parameters are pre-selected.

5.6 SIGNAL PEPTIDE AND TRANSMEMBRANE REGIONS:

A signal peptide is a short amino acid sequence at the N-terminal of the peptide.

- SignalP- Is an online server predicts the presence and locations of signal peptide cleavage sites in the provided protein sequence. FASTA sequence of the protein is pasted on the search tab, with Eukaryotes as the selected group of organism.

Presence of transmembrane regions were found with TMpred.

- TMpred- Is an online tool provided by ExPasy, which predicts membrane spanning regions and their orientation. The algorithm is based on the statistical analysis of

TMbase, a database of naturally occurring transmembrane proteins. The FASTA format of the protein sequence was inserted to run the TMpred search.

5.7 MOLECULAR STRUCTURE:

A three-dimensional molecular structure can be predicted with the sequence of amino acids in the protein. That is, it predicts the folding of the protein along with its secondary and tertiary structure by the information provided in its primary structure—the amino acid sequence. The molecular structure prediction of the kaouthiagin SVMP was done using SwissModel and I-TASSER. Furthermore SOPMA was used for secondary structure analysis of the protein.

- SwissModel- It is a fully automated protein structure and homology modelling online server, accessible via ExPASy. Inserting the target sequence in FASTA format will result in many different templates from which a few models will be developed.
- I-TASSER- This tool not only predicts a 3D molecular structure but also functional properties of the protein. Structural templates are first identified from the PDB by multiple threading approach. Full length atomic models are then constructed by iterative template fragment assembly simulations. Then, function insights of the target protein are derived by threading 3D models through the protein function database.
- SOPMA- Self-Optimized Prediction Method with Alignment is a method of prediction of secondary structures of a protein. The binding of a protein with other molecules is very specific to carry out its function properly. For this reason every protein has a particular structure. Protein structures are classified into primary, secondary, tertiary, and quaternary. The protein sequence was pasted on the server's search tab with its output width pre-selected at 70. Parameter for number of conformational states was selected as 4.

CHAPTER 6: RESULTS

6.1 PROTEIN SEQUENCE:

The target protein sequence was retrieved from NCBI in GenPept, under the protein database in FASTA format, with the accession number- **P82942.1**.

>P82942.1 RecName: Full=Hemorrhagic metalloproteinase-disintegrin-like kaouthiagin; AltName: Full=Snake venom metalloproteinase; Short=SVMP

TNTPEQDRYLQAEKYIEFYVIVDNRMYRYYNYDKPAIKIRVYEMINAVNTKFRPLKI
HIALIGLEIWSNEDKFEVKPAASVTLKSFREWRQTVLLPRKRNDNAQLLTGINLNGT
AVGIAYPGSLCTQRSVFVVQDYNRRMSLVASTMTHELGHNLGIHHDEASCICIPGPCI
MLKKRTAPAFQFSSCSIRDYQEYLLRDRPQCILNKPLSTDIVSPAICGNYFVEEGEECD
CGSPAACQSACCDAATCKFNGAGAECRAAKHDCDLPELCTGQSAECPTDSLQRNGH
PCQNNQGYCYNGKCPTLTNQCIALLGPHFTVSPKGCFDLNMRGDDGSFCRMEDGTK
IPCAAKDVKCGRLYCTEKNTMSCLIPPNPDGIMAEPGTKCGDGMVCSKGQCVDVQT
AY

6.2 PROTEIN CHARACTERISTIC ANALYSIS:

- ProtParam- Analysis of the protein sequence saved from NCBI, through the ProtParam server shows the following results:

  Number of amino acids: 401

  Molecular weight: 44492.86

  Theoretical pI: 6.66

  Amino acid composition:

Ala (A) 29	7.2%	His (H) 8	2.0%
Arg (R) 21	5.2%	Ile (I) 23	5.7%
Asn (N) 23	5.7%	Leu (L) 28	7.0%
Asp (D) 23	5.7%	Lys (K) 22	5.5%
Cys (C) 30	7.5%	Met (M) 10	2.5%
Gln (Q) 17	4.2%	Phe (F) 12	3.0%
Glu (E) 21	5.2%	Pro (P) 24	6.0%
Gly (G) 28	7.0%	Ser (S) 21	5.2%

Thr (T) 23	5.7%	Val (V) 19	4.7%
Trp (W) 2	0.5%	Pyl (O) 0	0.0%
Tyr (Y) 17	4.2%	Sec (U) 0	0.0%

Total number of negatively charged residues (Asp + Glu): 44
Total number of positively charged residues (Arg + Lys): 43

Atomic composition:

Carbon	C	1922
Hydrogen	H	3026
Nitrogen	N	544
Oxygen	O	591
Sulfur	S	40

Formula: $C_{1922}H_{3026}N_{544}O_{591}S_{40}$
Total number of atoms: 6123

Estimated half-life:
The N-terminal of the sequence considered is T (Thr).
The estimated half-life is: 7.2 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 42.97

Aliphatic index: 70.57

Grand average of hydropathicity (GRAVY): -0.341

- The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell (Zhou, 2004).
- Aliphatic index is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine) (Ikai, 1980).
- Instability index classifies the protein as stable or unstable. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad, Reddy, & Pandit, 1990).
- GRAVY is the average of the hydropathy values of all the amino acids. Hydropathy is a measure of hydrophobicity or hydrophilicity (Kyte & Doolittle, 1982).

6.3 HOMOLOGY:

- PSI-BLAST was used to find sequences homologous to the kaouthigagin SVMP, through the Standard Protein BLAST suite on NCBI. PSI-BLAST was done up to 3 iterations with maximum 500 sequence entries at a time. This search was performed against non-redundant protein sequence database. The search result showed the putative conserved domains containing ZnMc adamalysin II; Disintegrin and ADAM_CR.

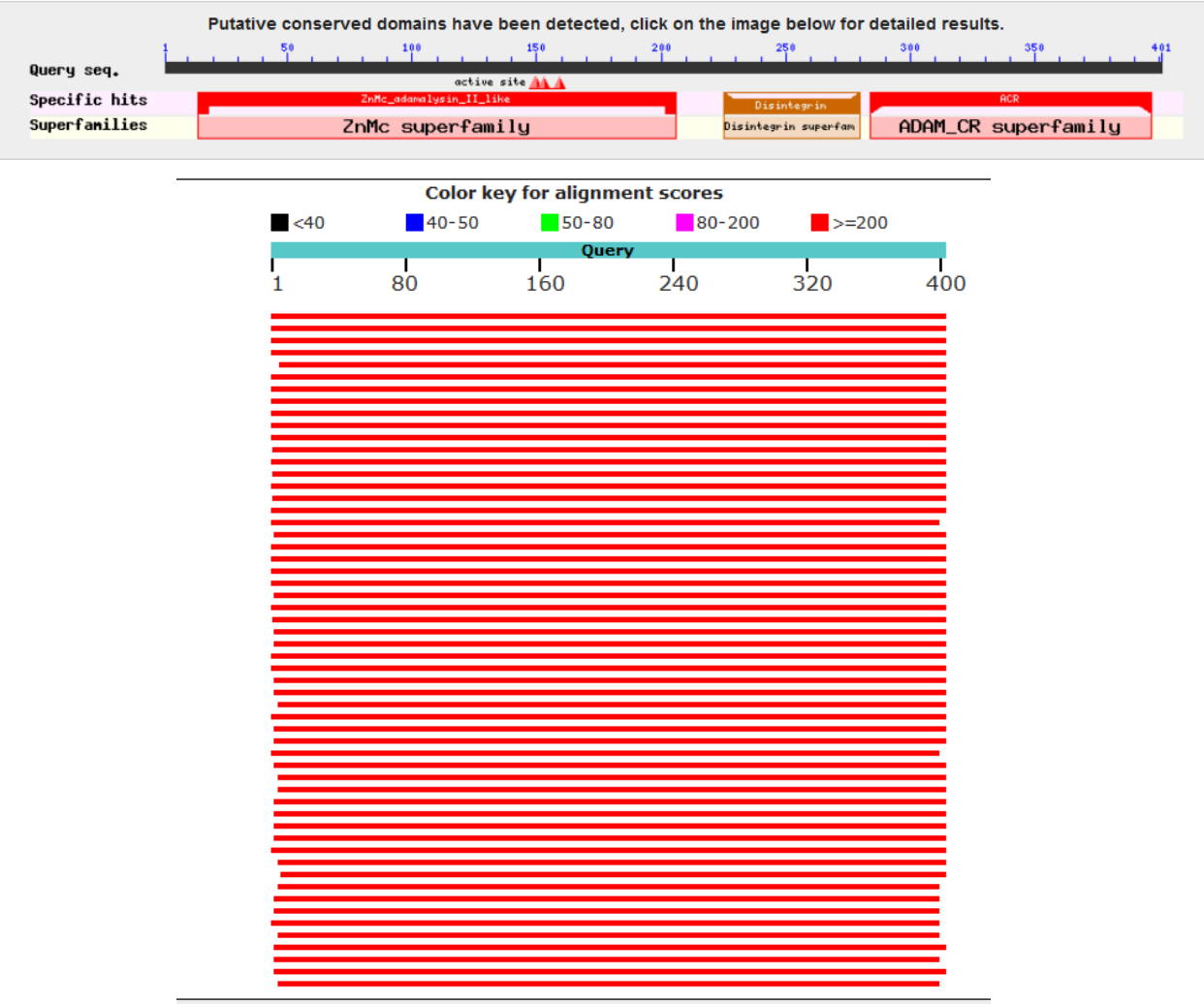


Figure 6.3.1 Results from PSI-BLAST, showing putative conserved domains and alignment of the homologous sequences.

- Regions with sequence similarity were found, which will yield functional and evolutionary clues about the structure and function of the protein in question.

- Three conserved domains were found in the kaouthiagin sequence, against which other proteins were compared and aligned with.
- The conserved domains were: Zn metalloprotease superfamily; disintegrin superfamily and ADAM CR superfamily.
- It also shows possible active sites present in the metalloprotease domain.
- The metalloprotease domain has a Zn-binding motif.

- PRALINE- After the 3rd iteration in PSI-BLAST, 15 sequence entries, including the target protein sequence were selected for multiple sequence alignment, which had 65% and above match with the target. The sequences selected are tabulated as follows:

Table 6.3.1: Homologous sequences selected from PSI-BLAST for Multiple Sequence Alignment

Name	Accession ID	Ident Match
Naja kaouthiagin [*Naja kaouthia*]	P82942.1	100%
Metalloproteinase III 2 [*Micrurus tener*]	JAS05093.1	65%
Metalloproteinase III 1 [*Micrurus fulvius*]	JAS04981.1	65%
Asrin [*Austrelaps superbus*]	ABH10621.1	65%
Zn Metalloproteinase atragin [*Naja atra*]	D3TTC2.1	66%
Zn Metalloproteinase MTP9 [*Drysdalia coronoides]*	F8RKV9.1	65%
Zn Metalloproteinase MTP4 [*Drysdalia coronoides*]	F8RKW1.1	65%
Zn Metalloproteinase cobrin [*Naja kaouthia*]	Q9PVK7.1	66%
Zn Metalloproteinase MTP8 [*Drysdalia coronoides*]	F8RKW0.1	65%

Metalloproteinase III 2 [*Micrurus tener*]	JAS05092.1	65%
Zn Metalloproteinase atrase-B [*Naja atra*]	D6PXE8.1	91%
Zn Metalloproteinase kaouthiagin [*Naja kaouthia*]	D3TTC1.1	91%
Chain A [*Naja atra*]	3K7L_A	66%
Metalloproteinase atrase B [*Naja atra*]	ADD14036.1	91%
Metalloproteinase III 2b [*Micrurus fulvius*]	JAS04979.1	65%

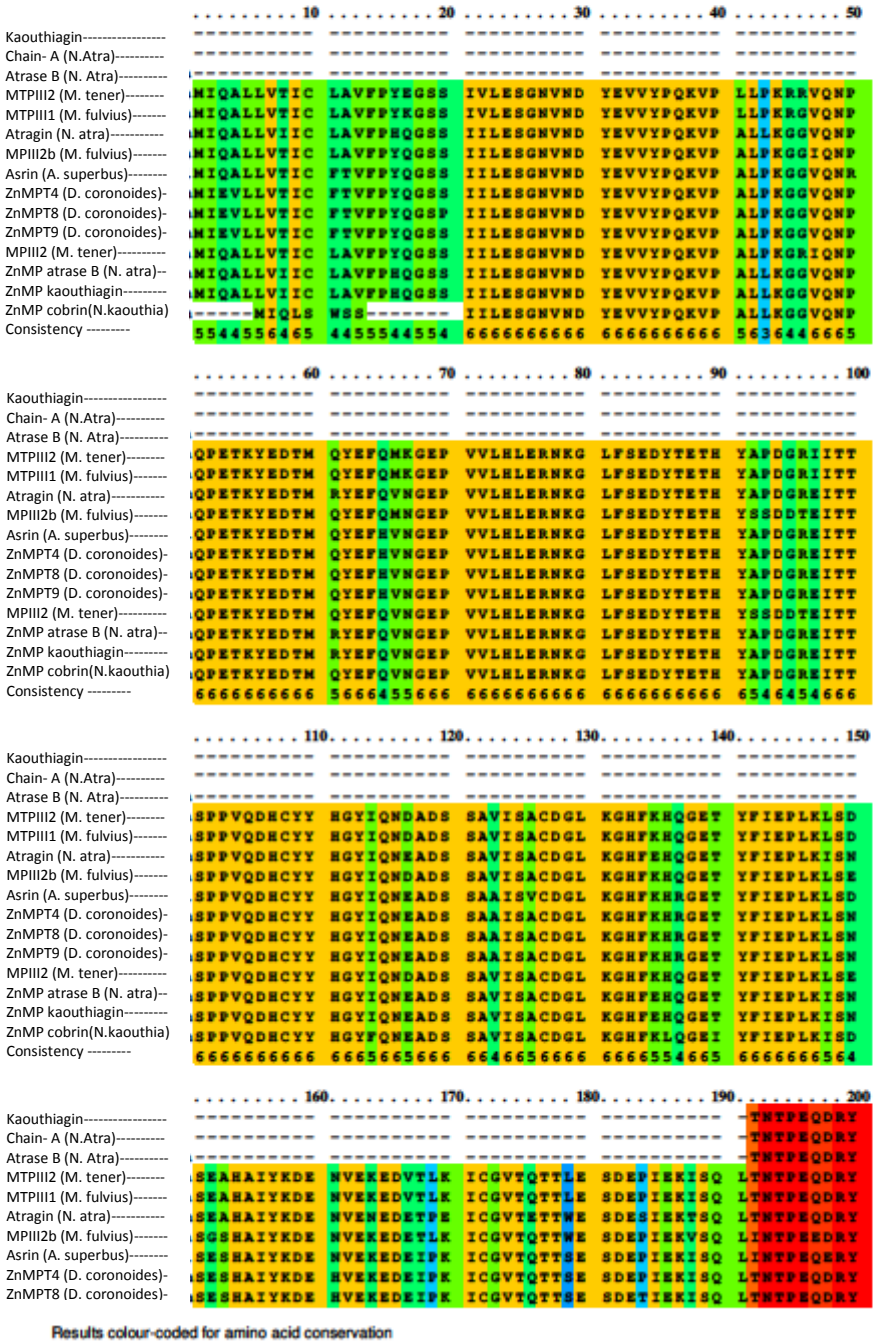
MSA of the kaouthigin SVMP was done with these selected protein sequences on the PRALINE program. Results from PRALINE showed similarities among the proteins in terms of amino acid residues and hydrophobicity of the sequences.

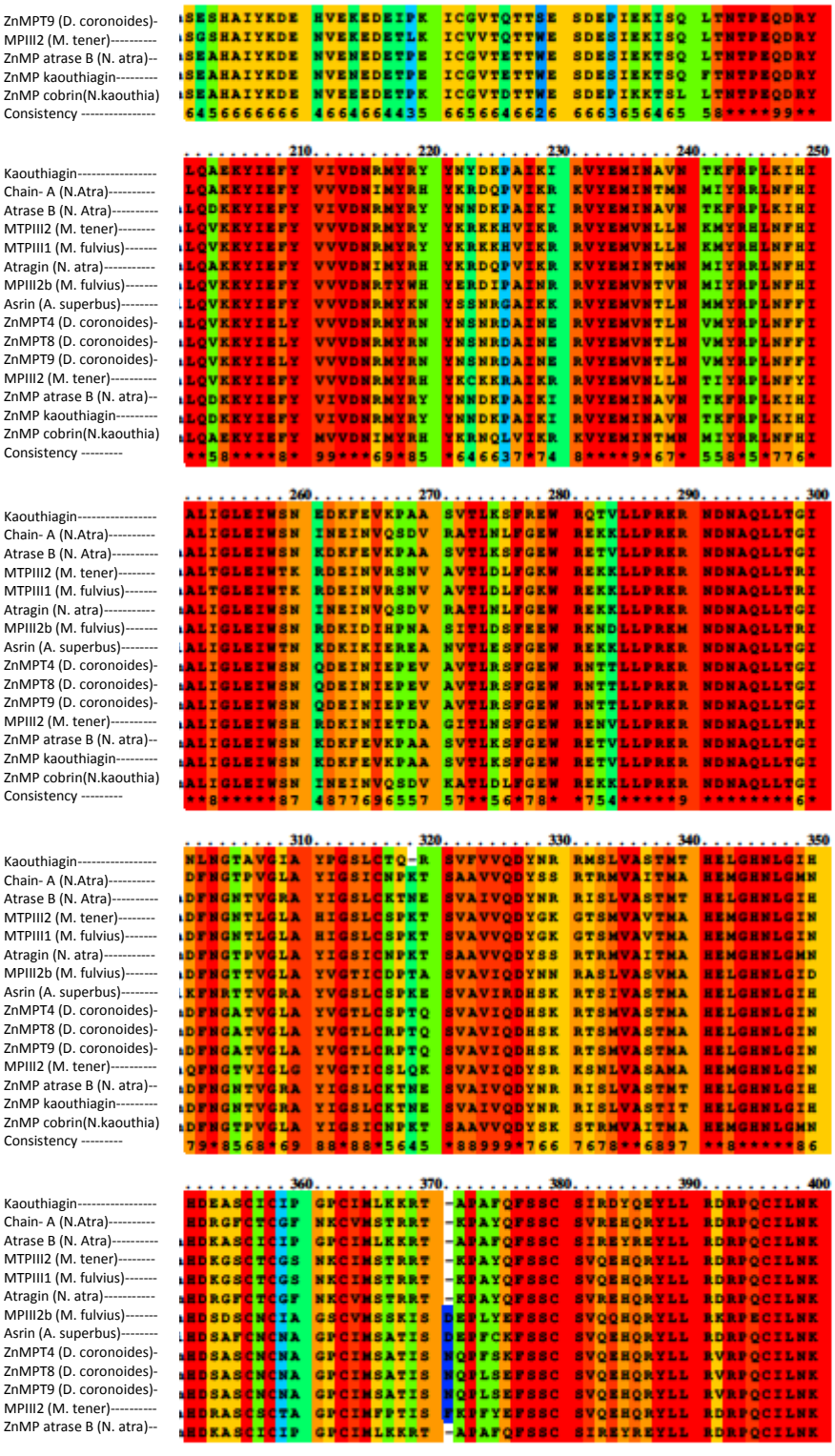
Results colour-coded for amino acid conservation

The current colourscheme of the alignment is for **amino acid conservation**.

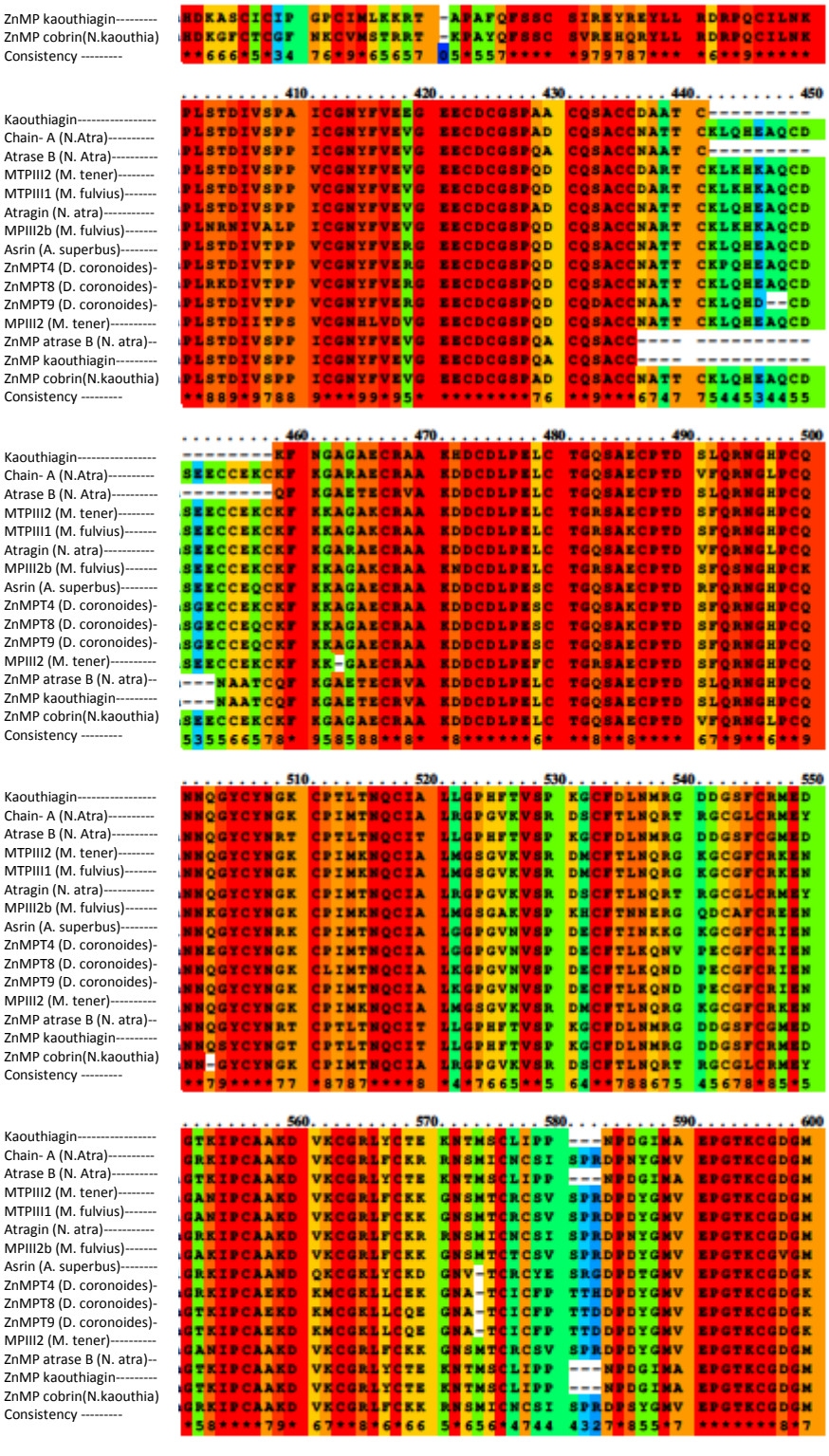
The conservation scoring is performed by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position. The colour assignments are:

Unconserved 0 1 2 3 4 5 6 7 8 9 10 **Conserved**





Results colour-coded for amino acid conservation



Results colour-coded for amino acid conservation

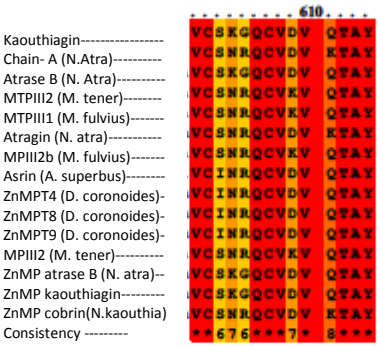


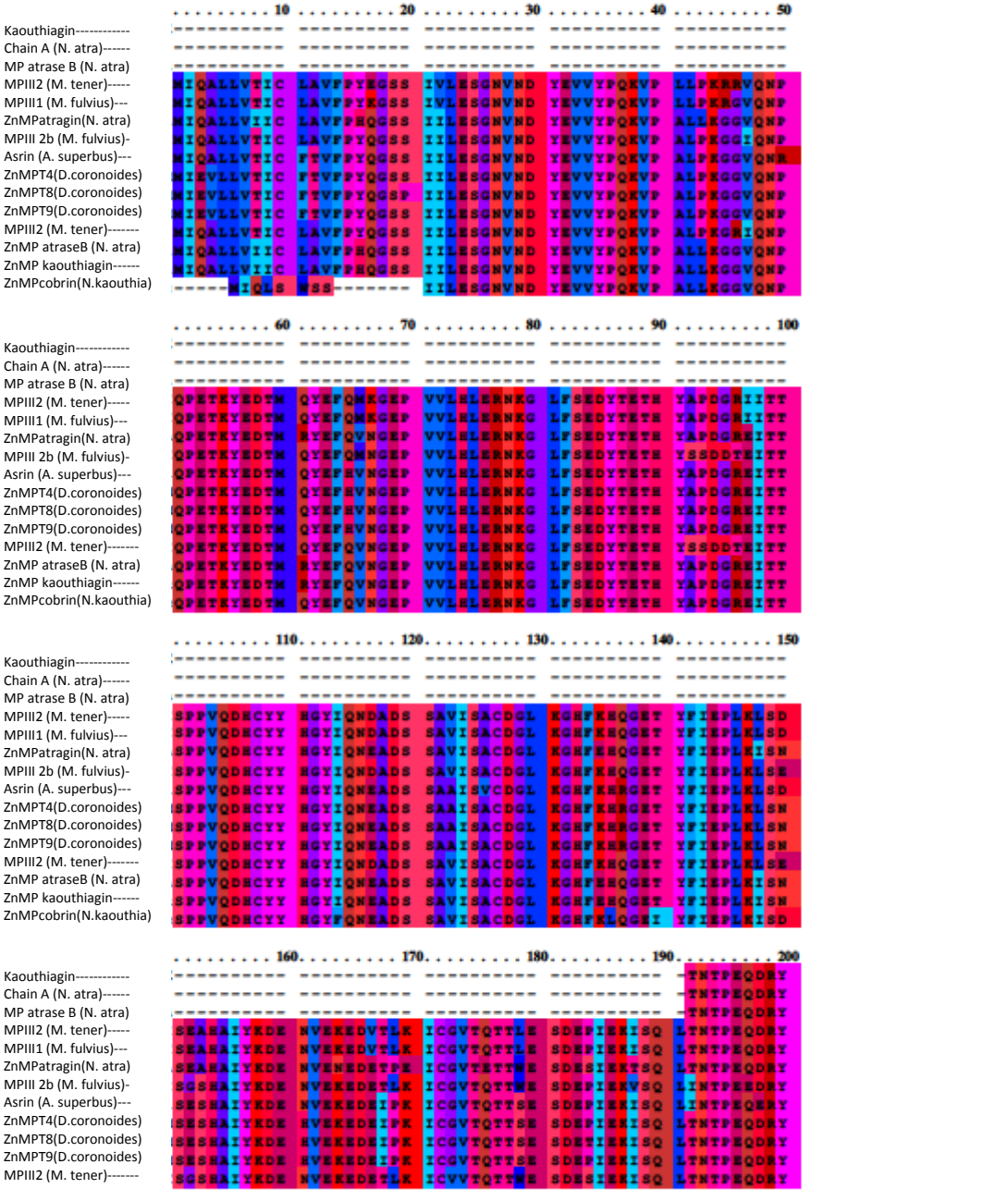
Figure 6.3.2: Result from the MSA by PRALINE showing conservation of amino acid residues among the homologous sequences.

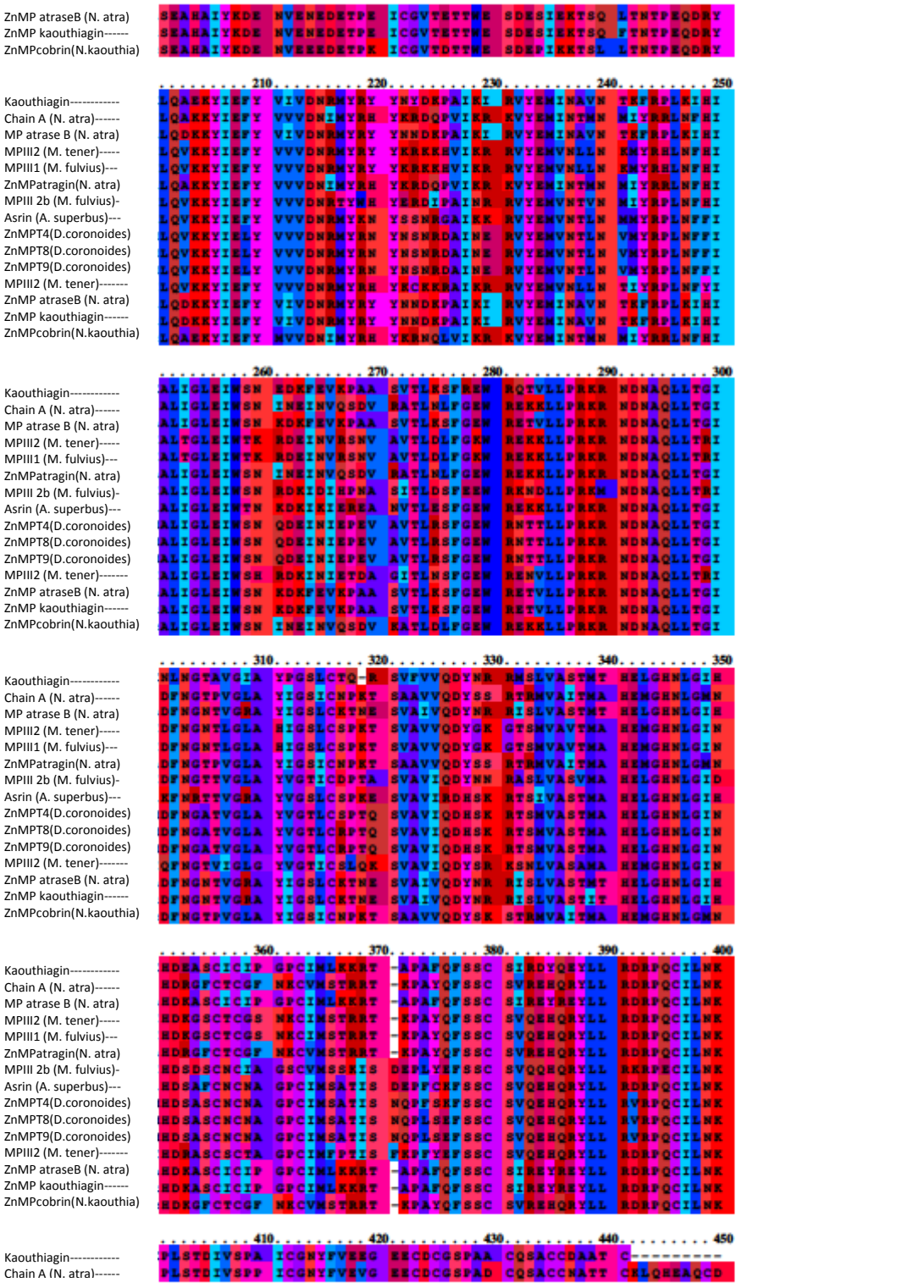
Results colour-coded for hydrophobicity

The current colourscheme of the alignment is for **hydrophobicity**.

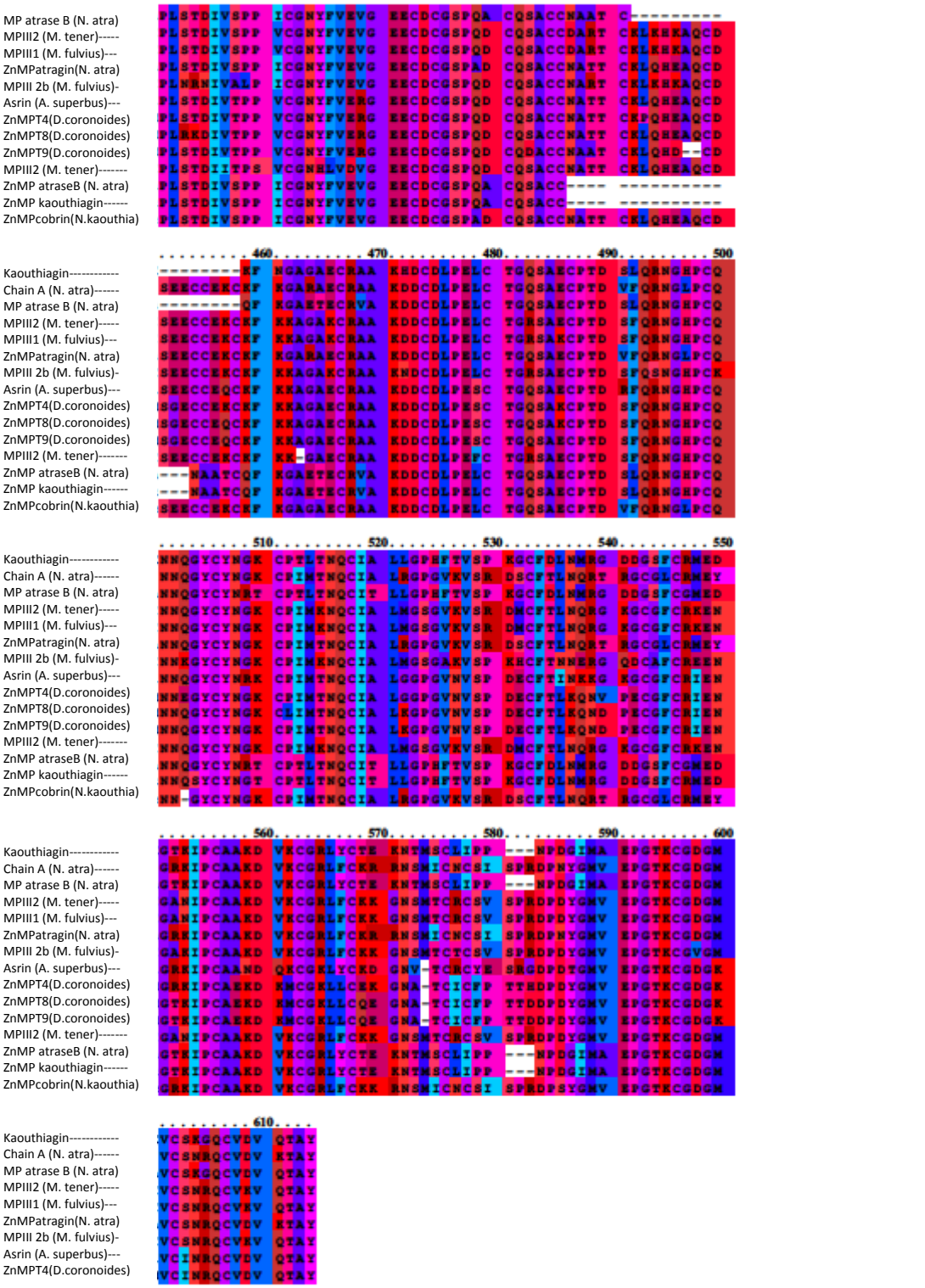
The hydrophobicity scale used is from Eisenberg *et al* (1984) Abstract
The colour assignments **from hydrophobic to hydrophilic** are:

HYDROPHOBIC ILE (I) PHE (F) VAL (V) LEU (L) TRP (W) MET (M) ALA (A) GLY (G) CYS (C) TYR (Y) PRO (P) THR (T) SER (S) HIS (H) GLU (E) ASN (N) GLN (Q) ASP (D) LYS (K) ARG (R) HYDROPHILIC





Results colour-coded for hydrophobicity



Results colour-coded for hydrophobicity

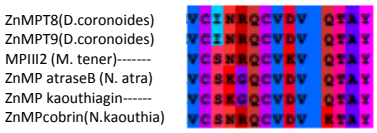


Figure 6.3.3: Result of hydrophobicity of the compared homologous sequences, by PRALINE.

- In the conservation results, the sequences were aligned to display the similarities and dissimilarities between all the inserted protein sequences by each amino acid residue. This is presented by a colour schemed scoring, where red shows the most conserved regions (with a score of 10) and blue shows the least conserved regions (with a score of 0). The amino acid conservation among the sequences show they are highly conserved, from the beginning of the kaouthiagin protein sequence.
- Hydrophobicity of the sequences are also presented by a colour scheme, where a specific colour is designated for each of the 20 amino acids. From the PRALINE results for hydrophobicity, it can be stated that there is a good mixture of both hydrophobic and hydrophilic amino acids.

6.4 PROTEIN MOTIF:

- Pfam- The arrangement of domains on the target protein was shown as—

Download the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End	Gathering threshold (bits)		Score (bits)		E-value	
				Sequence	Domain	Sequence	Domain	Sequence	Domain
Pfam	Reprolysin	14	208	20.80	20.80	199.70	198.90	1.5e-55	2.7e-55
Pfam	ADAM_CR	285	369	23.00	23.00	62.40	62.40	1.7e-13	1.7e-13

Figure 6.4.1: Results of the domains from Pfam.

- InterProScan-

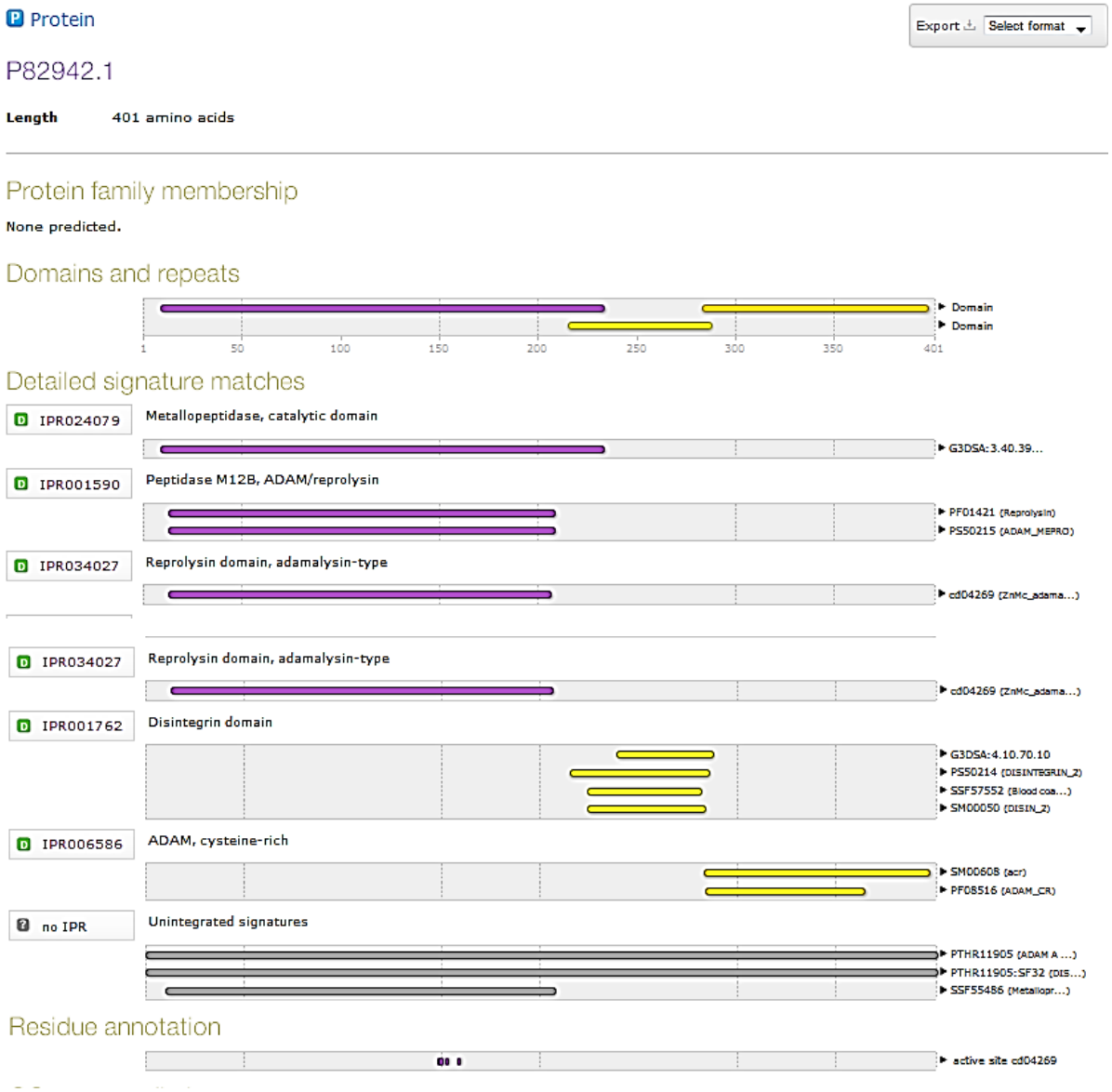


Figure 6.4.2: Results of domains from InterProScan.

- The following table shows the results acquired from both the tools used to find the domains of kaouthiagin.

Table 6.4.1: Results of the protein motifs found in kaouthiagin

Tool/ server used	Domains found	Position in sequence (amino acid residue)
Pfam	Reprolysin	14-208
	ADAM CR	285-369
InterProScan	Metalloprotease	10-233
	Disintegrin	216-287
	ADAM CR	284-397

- Reprolysin are proteins of the family that cleave peptides. These proteases require Zn^{2+} ions for their catalysis. Most members are snake venom endopeptidases (Rawlings & Barrett, 1995). Thus, they are Zn metalloproteases.
- Disintegrins are a family of proteins that obstruct the activity of integrins.
- An ADAM is a transmembrane protein that contains a disintegrin and a metalloprotease domain. It is bound to zinc containing metalloproteinase, which is considered to be crucial modulators of physiological and pathological processes (Chellapandi, 2014). All members of the ADAM family display a common domain organisation: a pro-domain, the metalloprotease, disintegrin, cysteine-rich, epidermal-growth factor like and transmembrane domains. They possess four potential functions: proteolysis, cell adhesion, cell fusion, and cell signalling. ADAMs are membrane-anchored proteases that proteolytically modify cell surface and extracellular matrix (ECM) in order to alter cell behaviour. They are responsible for the proteolytic cleavage of transmembrane proteins and release of their extracellular domain. The adamalysins are zinc dependent endopeptidases found in snake venom (Smith et al., 2002).

6.5 PHYLOGENY:

- Phylogeny.fr- The phylogenetic tree constructed was as follows:

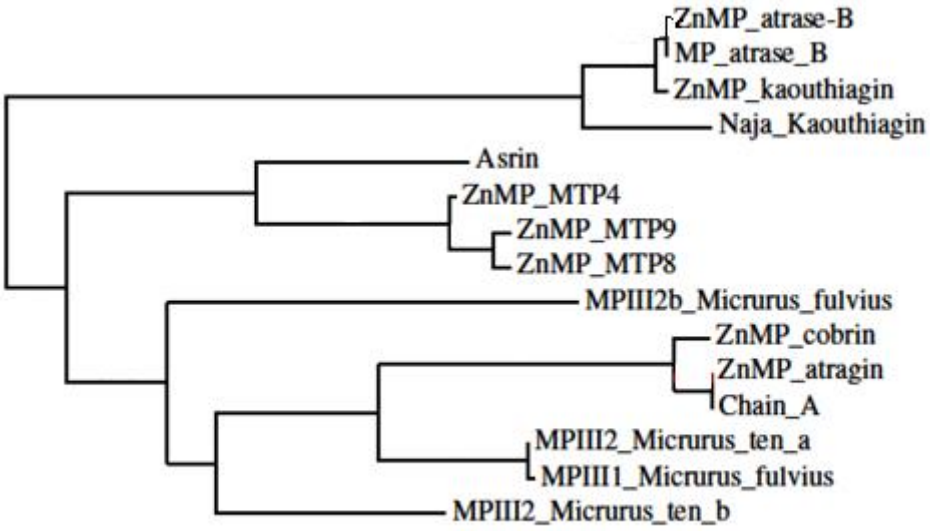


Figure 6.5.1 Result of phylogenetic tree obtained from Phylogeny.fr

Table 6.5.1: The phylogenetic results are shown on the table below for clearer understanding.

Result	Organism
ZnMP_atrase-B	Zn Metalloproteinase atrase-B [*Naja atra*]
MP_atrase_B	Metalloproteinase atrase B [*Naja atra*]
ZnMP_kaouthiagin	Zn Metalloproteinase cobrin [*Naja kaouthia*]
Naja_kaouthiagin	*Naja kaouthia*gin [*Naja kaouthia*]
Asrin	Asrin [*Austrelaps superbus*]
	Zn Metalloproteinase MTP4

ZnMP_MTP4	[*Drysdalia coronoides*]
ZnMP_MPT9	Zn Metalloproteinase MTP9 [*Drysdalia coronoides*]
ZnMP_MPT8	Zn Metalloproteinase MTP8 [*Drysdalia coronoides*]
MPIII2b_Micrurus fulvius	Metalloproteinase III 2b [*Micrurus fulvius*]
ZnMP_cobrin	Zn Metalloproteinase cobrin [*Naja kaouthia*]
ZnMP_atragin	Zn Metalloproteinase atragin [*Naja atra*]
Chain A	Chain A [*Naja atra*]
MPIII2_Micrurus_ten_a	Metalloproteinase III 2 [*Micrurus tener*]
MPIII1_Micrurus_fulvius	Metalloproteinase III 1 [*Micrurus fulvius*]
MPIII2_Micrurus_ten_b	Metalloproteinase III 2 [*Micrurus tener*]

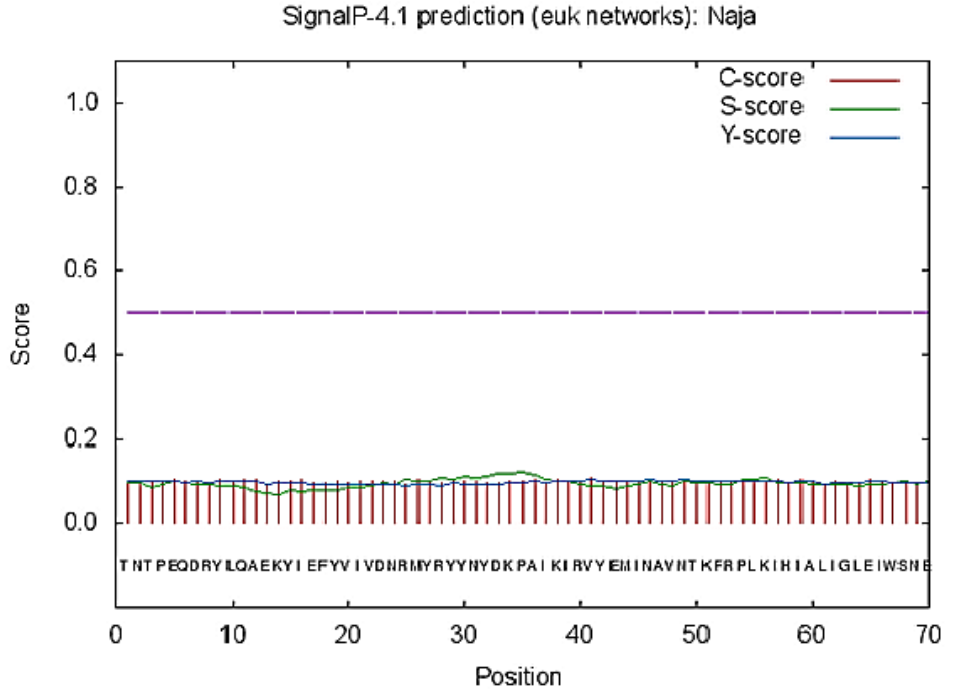
- The phylogenetic tree shows how the species are related.
- The tree primarily divides into two clades, one comprising of Zn metallproteinases from the genus *Naja* only and the other comprising of Zn metalloproteinases from other Elapid snakes.
- All the other SVMPs were also from snakes belonging to the Elapidae family.

6.6 SIGNAL PEPTIDE AND TRANSMEMBRANE REGIONS:

- SignalP (Version 4.1) - Shows the cleavage sites in terms of C-score (raw cleavage site score), S- score (signal peptide score) and Y-score (combined cleavage site score). C-score is the output from the CS networks, which are trained to distinguish signal peptide cleavage sites from everything else. S-score is the output from the SP networks, which are trained to distinguish positions within signal peptides from positions in the mature part of the proteins and from proteins without signal peptides. Y-score is the

combination of the C- score and the slope of the S- score, resulting in a better cleavage site prediction than the raw C- score alone ("SignalP 4.1 Output format," n.d.). SignalP result for kaouthiagin was negative.

```
# SignalP-4.1 euk predictions
>Naja  Kaouthiagin
```

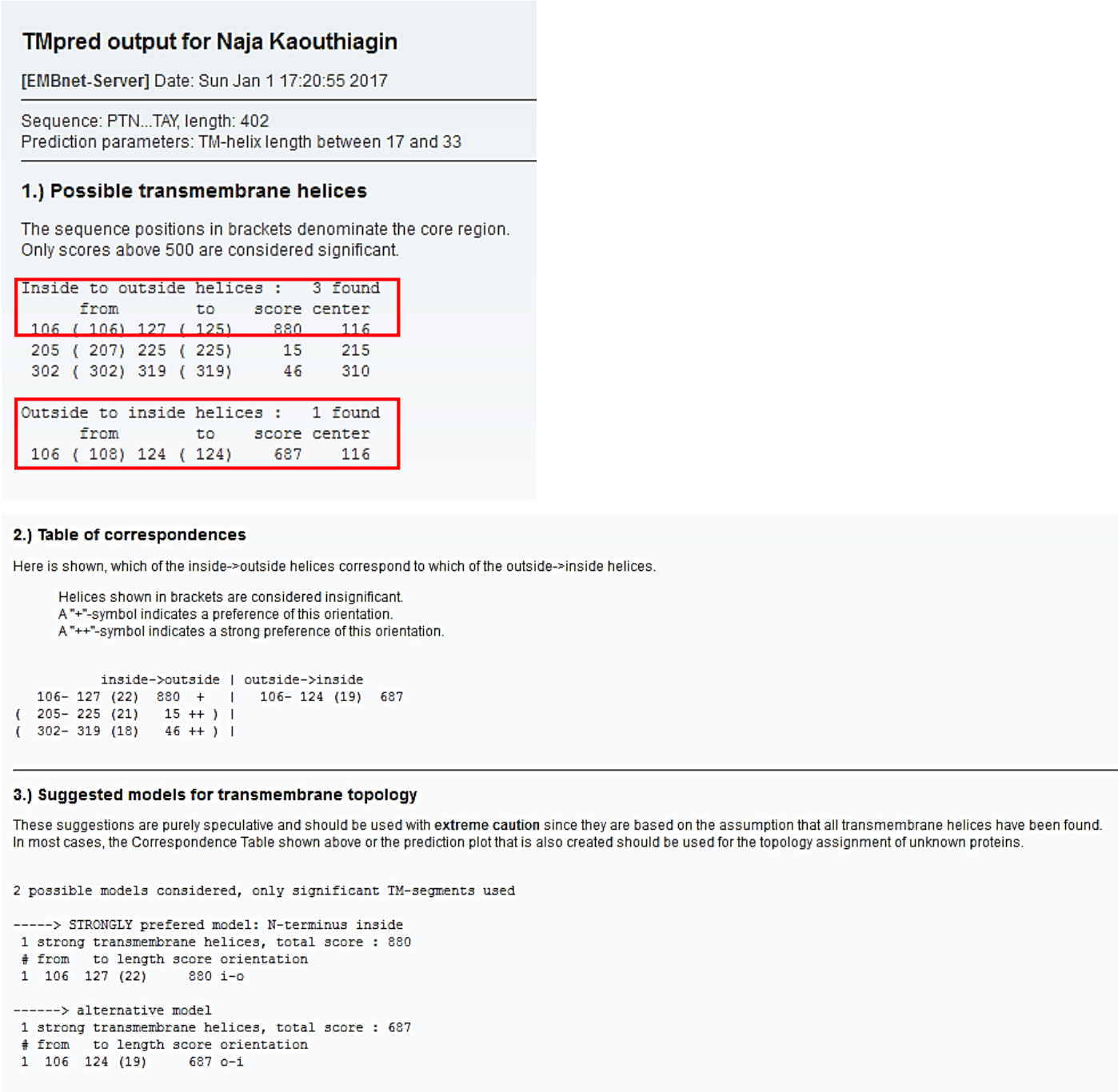



```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    41       0.108
  max. Y    41       0.106
  max. S    35       0.122
  mean S     1-40    0.096
       D     1-40    0.101   0.450    NO
Name=Naja       SP='NO' D=0.101 D-cutoff=0.450 Networks=SignalP-noTM
# data
# gnuplot script
```

Figure 6.6.1: Result from SignalP

- The C-score is supposed to be the highest right after the cleavage site. C-score here remains more or less unchanged.
- The S-score and the Y-score too, remains somewhat steady.
- The results suggest there is no cleavage of signal peptide. From this, we see that this protein is already active and does not require to be activated by accessory factors or precursors.

- TMpred was used to make a prediction of the presence of transmembrane regions on the query protein.



TMpred output for Naja Kaouthiagin

[EMBnet-Server] Date: Sun Jan 1 17:20:55 2017

Sequence: PTN...TAY, length: 402
Prediction parameters: TM-helix length between 17 and 33

1.) Possible transmembrane helices

The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

```
Inside to outside helices :   3 found
      from        to     score center
 106 ( 106) 127 ( 125)     880    116
  205 ( 207) 225 ( 225)      15    215
  302 ( 302) 319 ( 319)      46    310

Outside to inside helices :   1 found
      from        to     score center
 106 ( 108) 124 ( 124)     687    116
```

2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant.
A "+"-symbol indicates a preference of this orientation.
A "++"-symbol indicates a strong preference of this orientation.

```
           inside->outside | outside->inside
   106- 127 (22)  880  +   |   106- 124 (19)  687
(  205- 225 (21)   15 ++ ) |
(  302- 319 (18)   46 ++ ) |
```

3.) Suggested models for transmembrane topology

These suggestions are purely speculative and should be used with **extreme caution** since they are based on the assumption that all transmembrane helices have been found.
In most cases, the Correspondence Table shown above or the prediction plot that is also created should be used for the topology assignment of unknown proteins.

```
2 possible models considered, only significant TM-segments used

-----> STRONGLY prefered model: N-terminus inside
 1 strong transmembrane helices, total score : 880
 # from   to length score orientation
 1  106  127 (22)     880 i-o

------> alternative model
 1 strong transmembrane helices, total score : 687
 # from   to length score orientation
 1  106  124 (19)     687 o-i
```

Figure 6.6.2: Results from TMpred

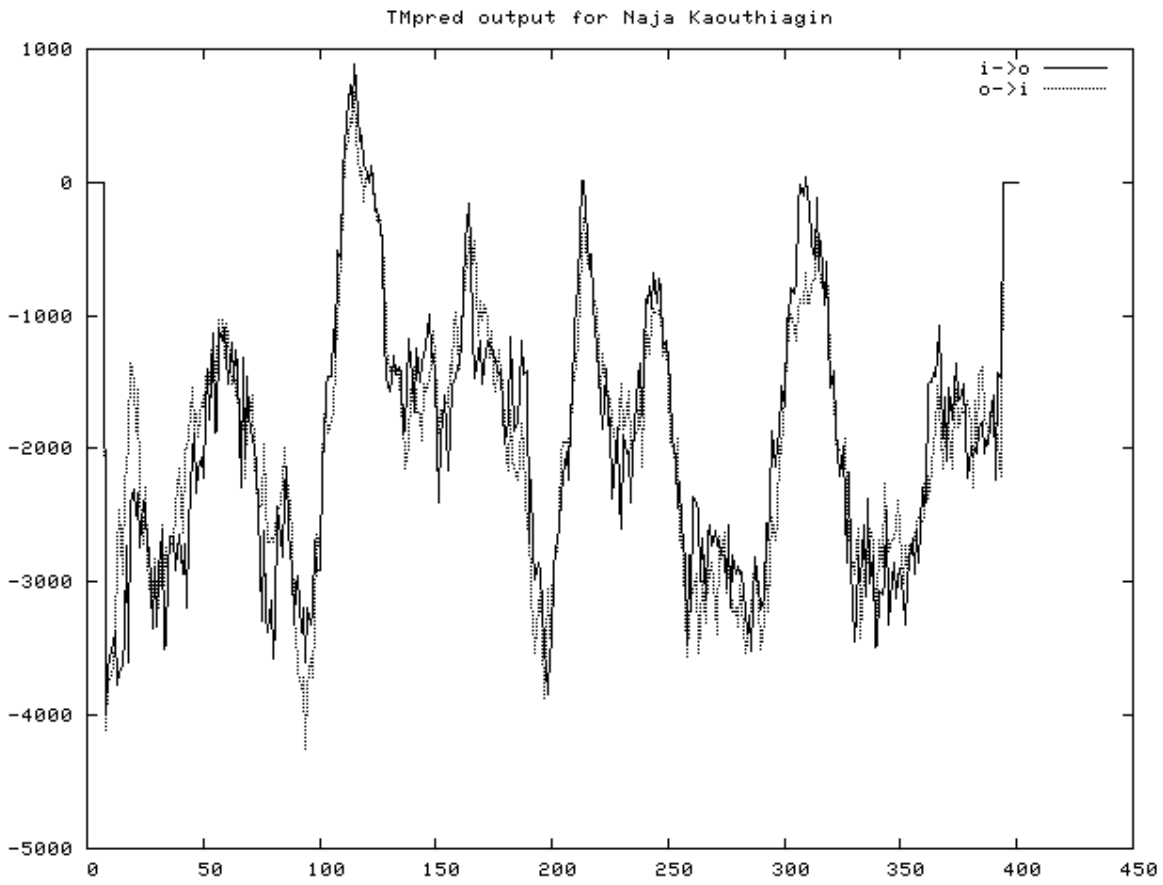


Figure 6.6.3: Graphical presentation of TMpred output

- The result from TMPred displays a positive result.
- Kaouthiagin is a transmembrane protein, as two transmembrane helices have been found with scores above 500, both in the same region.
- The graph also shows peaks above threshold value at amino acid residue 106.
- One membrane bound region has been predicted.

6.7 MOLECULAR STRUCTURE:

Homology 3D structure modelling results from I-TASSER and SwissModel are shown as follows:

- I-TASSER-

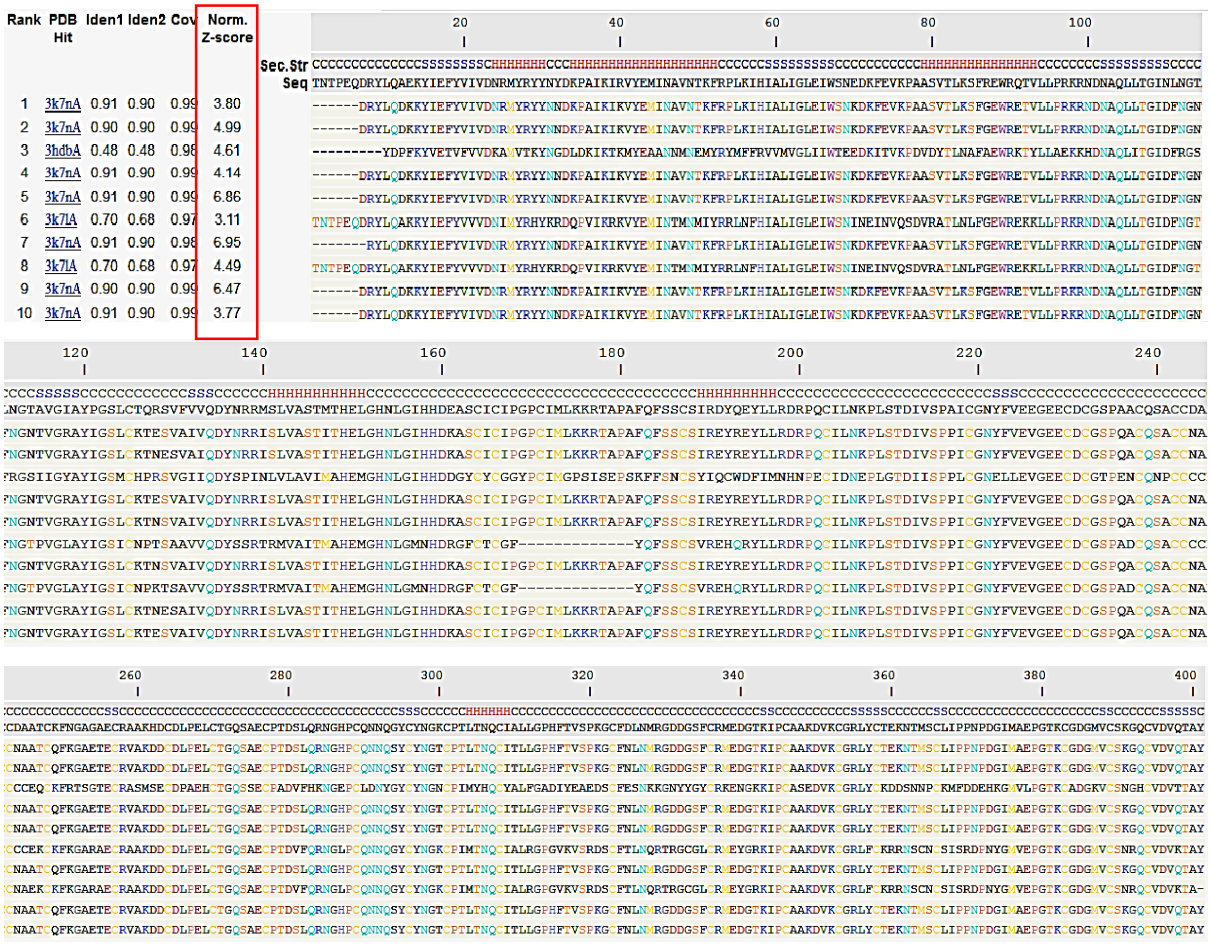


Figure 6.7.1: The top 10 template- query alignments generated by LOMETS

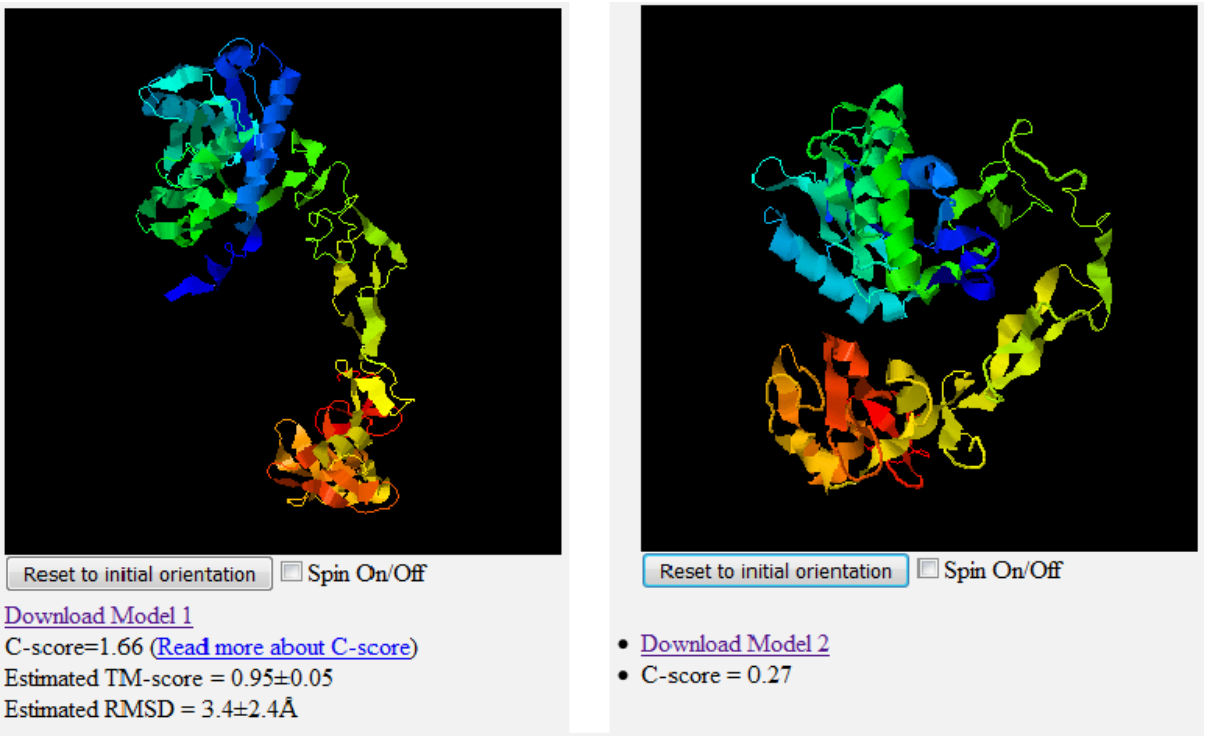


Figure 6.7.2: Top 2 final models predicted by I-TASSER

Table 6.7.1: Predicted function using COACH

Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
1	0.38	52	3hdbA	PEPTIDE	Rep, Mult	114,115,116,117,136,146,149,150,159,174,175,176,177
2	0.31	33	3b2zB	CA	Rep, Mult	17,101,203,206
3	0.08	12	2dw0A	CA	Rep, Mult	218,219,221,223,225,228,231
4	0.04	6	2dw0A	ZN	Rep, Mult	149,153,159
5	0.04	6	3dslA	CA	Rep, Mult	265,266,268,280,281

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column. **Mult** is the complex structures with all potential binding ligands in the cluster.

Table 6.7.2: Enzyme Comission numbers and active sites

Rank	Cscore^{EC}	PDB Hit	TM-score	RMSD^{a}	IDEN^{a}	Cov	EC Number	Active Site Residues
1	0.464	3aigA	0.484	1.26	0.467	0.496	3.4.24.46	80,120,146,149,153,172
2	0.438	1htdA	0.482	1.23	0.455	0.494	3.4.24.42	83,122,146,149,153,156
3	0.432	1wniA	0.479	0.97	0.487	0.486	3.4.24.53	120,149,156
4	0.428	3b2zF	0.467	2.35	0.207	0.501	3.4.24.82	47
5	0.330	3dslA	0.610	1.91	0.337	0.636	3.4.24.49	149,152,172

(a) Cscore^{EC} is the confidence score for the EC number prediction. Cscore^{EC} values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD^{a} is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^{a} is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

- Results from I-TASSER showed the top 10 threading templates used by the server to predict the models.
- From these templates, 2 final 3D structure models were developed by the server.
- If Norm. Z-score is higher than the value of 1, it means the alignment is good. The higher the score, the better.
- C-score for model 1 is 1.66 and for model 2 it is 0.27.
- I-TASSER also showed 5 predicted ligand binding sites of the protein, along with the respective ligands it would bind to.
- The server also predicted active site residues for the query protein along with predicted EC (enzyme commission) numbers for the query protein.

- SwissModel-

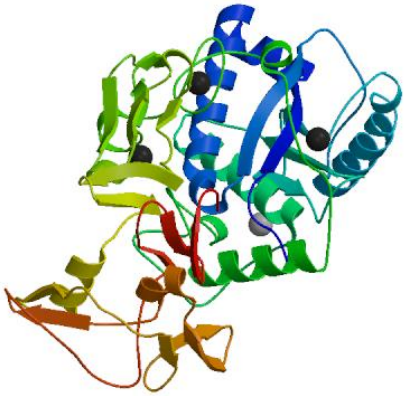


Figure 6.7.2: Model 1 from SwissModel

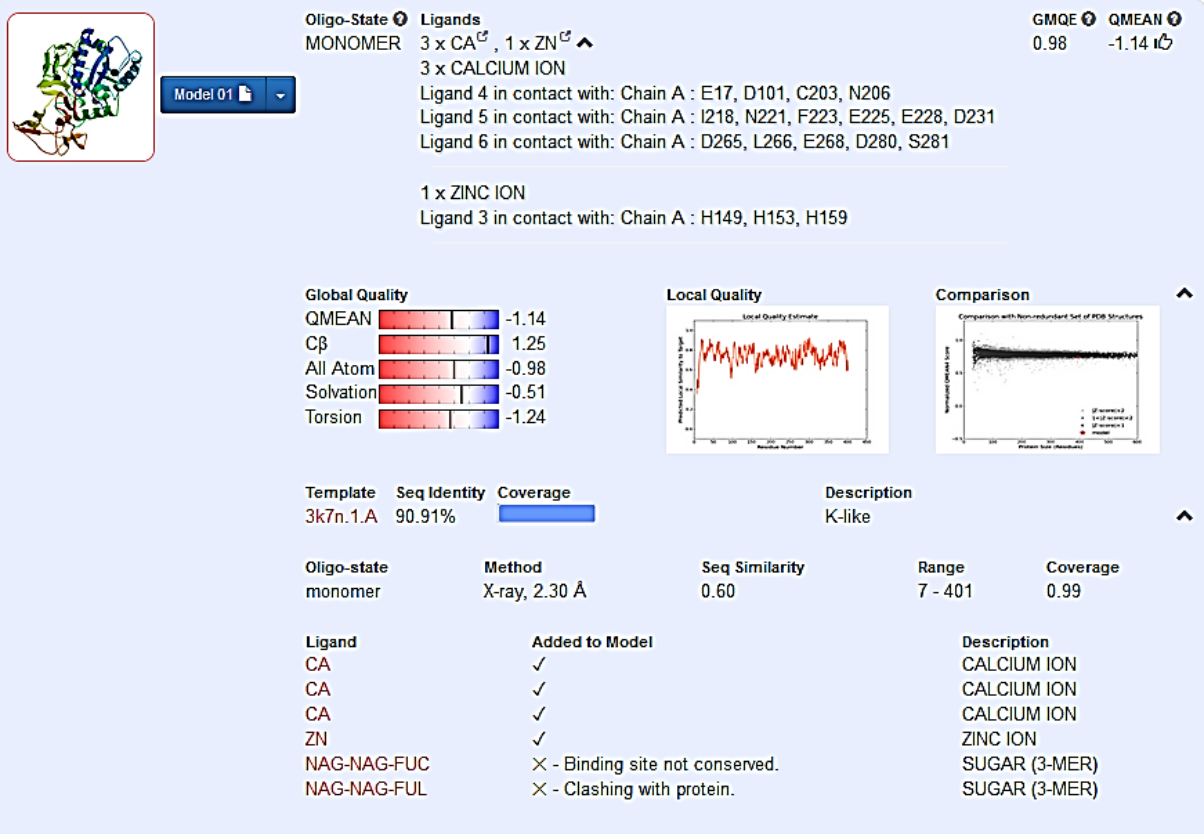


Figure 6.7.3: SwissModel results for Model 1

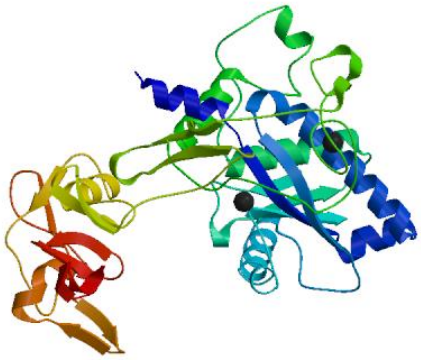


Figure 6.7.4: Model 2 from SwissModel

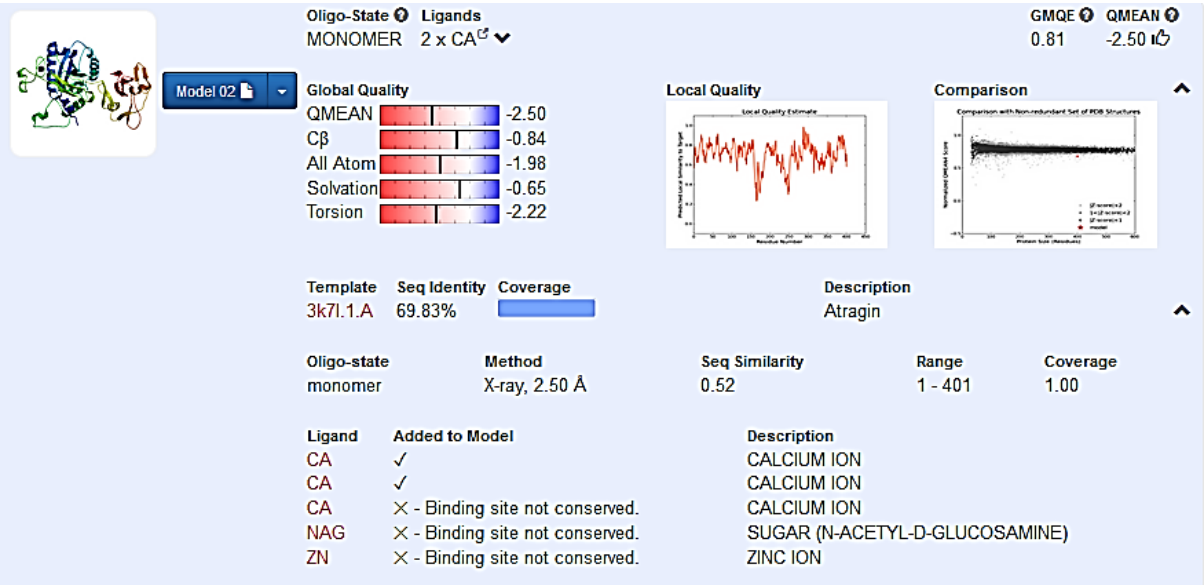


Figure 6.7.5: SWISS MODEL results for Model 2

- SWISS MODEL used 50 templates for the homology modelling.
- It also generated 2 probable final 3D structure models.
- Model 1 has a QMEAN score of -1.14, and model 2 has a score of -2.50.
- The two models also show the possible ligands that may be able to bind with the protein.

- SOPMA- The secondary structure prediction by SOPMA shows a percentage of the type of secondary structure the sequence forms.

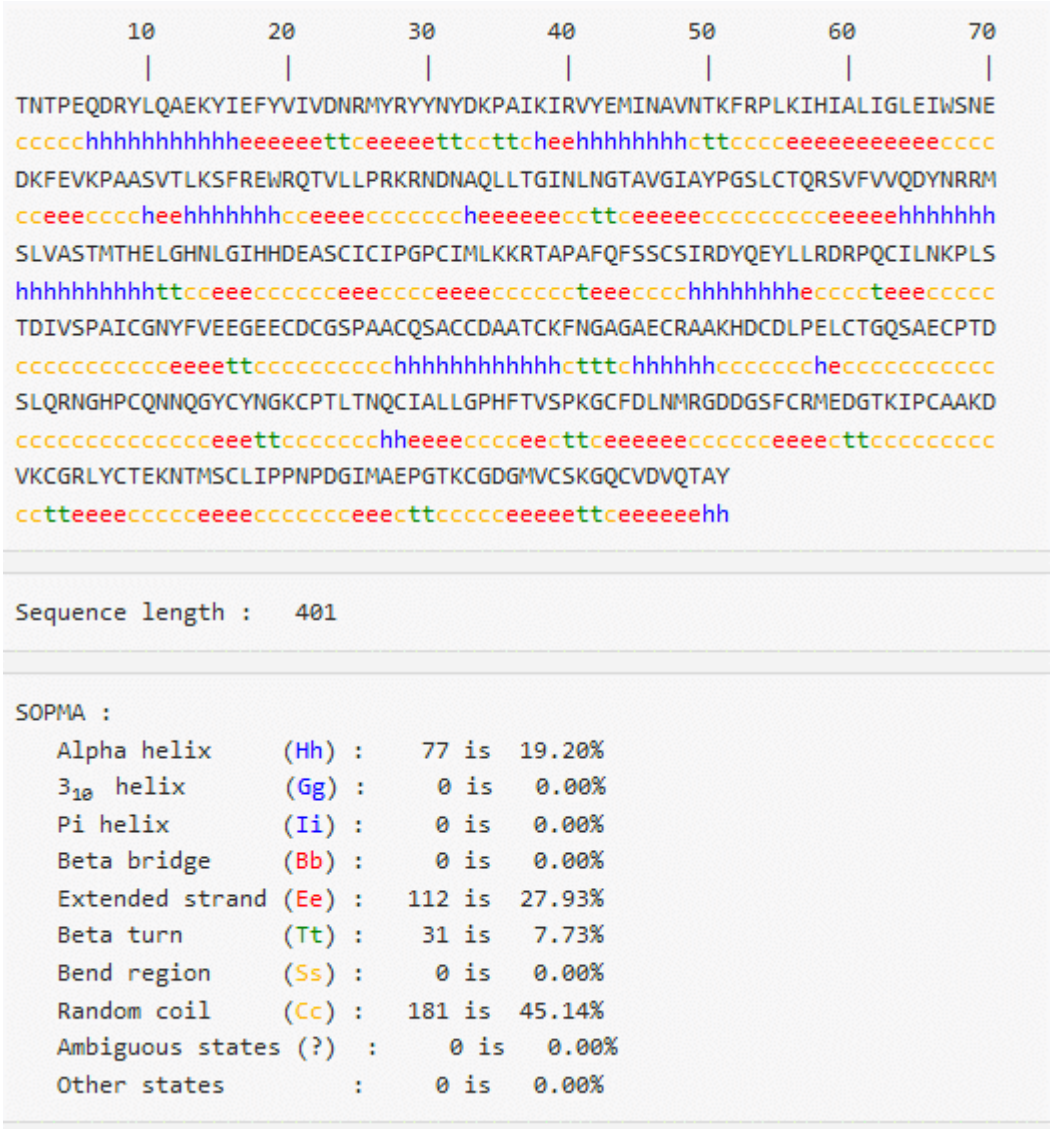


Figure 6.7.6: Results from SOPMA

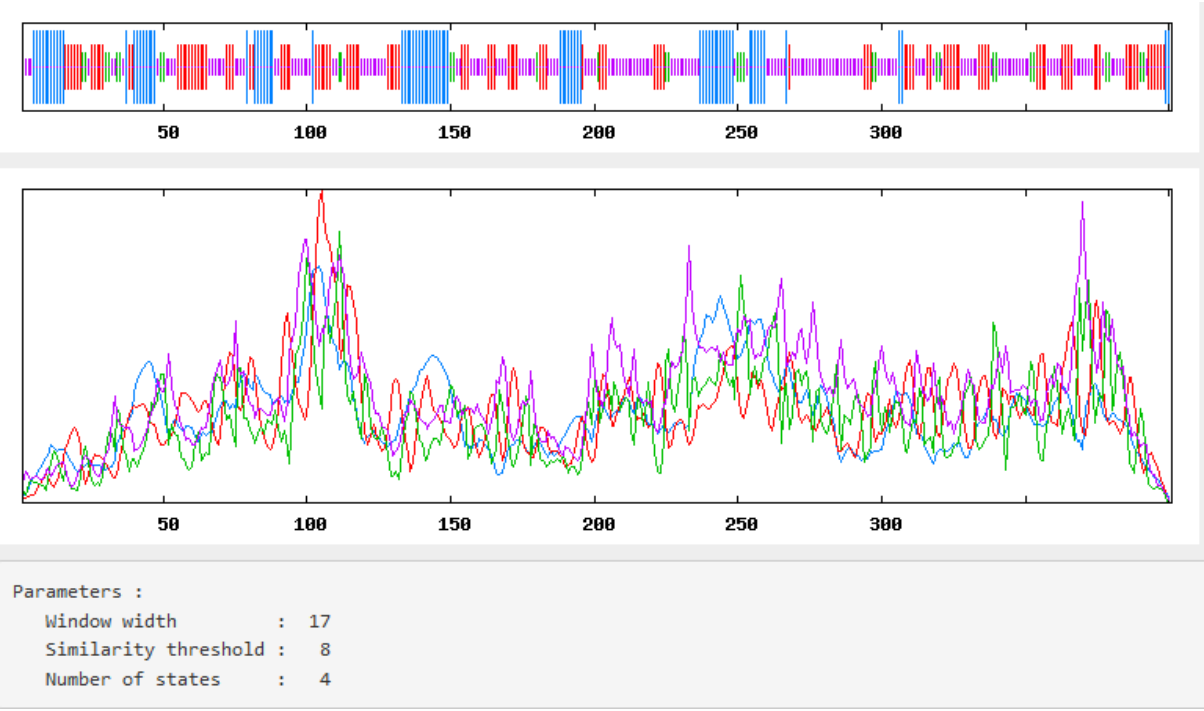


Figure 6.7.7: Graphs generated from SOPMA. The first graph helps to visualize the prediction and the second graph contains the score curves for all predicted states.

- Majority of the protein is in random coils (45.14%).
- The protein also has alpha helix structures (19.2%); extended strands (27.93%) and beta turns (7.7.3%).

CHAPTER 7: DISCUSSION OF RESULTS AND FUTURE PROSPECTS

7.1 DISCUSSION:

From the analysis of data and results attained from this study, kaouthiagin was found to be a sequence of 401 amino acids, with the accession ID: P82942.1.

The protein has a high molecular weight resembling the P-III class of SVMPs, as it also consists of 3 different domains. The estimated half-life of this protein is 7.2 hours in mammalian cells, 20 hours in yeast cells and 10 hours in E.coli cells. The instability index of the protein shows its value to be 42.97. This indicates that the protein in unstable. This poses as the major limitation in translating the possible use of kaouthiagin, due to its instability and limited bioavailabilty. Further research is essential to stabilize the protein by manipulation of its residues, without altering its function or efficacy. Average hydropathicity of the protein sequence is -0.341. A negative GRAVY value indicates the protein as non-polar.

Knowing hydropathy of a protein is essential to decipher the structure and thus the function of the protein. Hydrophobicity or hydrophilicity will determine the folding of the protein, which in turn will define its substrate specificity and binding, its membrane domains and solubility.

PSI-BLAST was used to find its distant homologues, which would not have been possible using the BLASTp tool. Finding homologues, based on the sequence of the protein of interest is important. Homologous sequences give us insights of the possible function of the protein and also may identify proteins with a developed three-dimensional structure that can serve as models for the structure of the protein of interest. The conserved domain database groups the sequences having strong similarities into domain fingerprints, which allows us to search these groups individually. The results enable function predictability from known functions of the homologous proteins, to explore evolutionary relationships or to identify structural features.
PSI-BLAST iteratively searches one or more protein databases for sequences similar to the query sequence. The sequences selected for a second round of the search are above specified score (*e*- value) threshold. Similarity between two sequences can be expressed as a percentage of sequence identity (identity match). The profile is refined for another round of searching, using sequences with high identity match with the protein of interest. This process is iteratively continued until desired or until convergence, where no new sequences are detected above the defined threshold. The iterative profile generation process of PSI-BLAST makes it far more capable of detecting distant sequence similarities than a single query alone in BLASTp, as it

combines the underlying conservation information from a range of related sequence into a single score matrix (Bhagwat & Aravind, 2007). 3 iterations were taken as it was observed that the resulting sequences from the second and third round were quite similar.

Both local alignment (PSI-BLAST) and global alignment (PRALINE) were used to find near and distant relationships of *Naja kaouthia* with other species. PSI-BLAST finds out the segments of the sequence or domains that are conserved among other organism species, whereas PRALINE enhances this search and produces a more refined result, displaying the resemblances among the corresponding sequences of amino acids of each of the compared proteins. Using these results, we can deduce that the kaouthiagin domains have similar functions as the protein sequences compared with. From the conservation results obtained from PRALINE, it is evident that the residues that are important for stabilizing domains architecture are strictly conserved throughout the primary structure among SVMPs. The PSI-BLAST results also suggest that kaouthiagin does not have resemblances with other organisms other than snakes. The homologous sequences selected were greatly similar and all belonged to snakes species of the Elapidae family.

Kaouthiagin has a Zn-binding motif (HELGHNLGIHHD), from amino acid residue 148 to 160, in its metalloprotease domain (10-233). The disintegrin-like domain showed the HDCD sequence (262-265) at the position corresponding to the RGD sequence of disintegrins (321-331), in the Cys-rich domain. Disintegrins rich in Cys residues are mainly involved in disulfide bonds, resulting in proteolysis- resistant molecules (Selistre-de-Araujo et al., 2010). Disintegrins, containing a signature RGD or KGD sequence, signifies the presence of the motif that specifically binds to integrin IIb- IIIa receptors on the platelet surface, thereby blocking the binding of fibrinogen to the receptor-glycoprotein complex of activated platelets. Disintegrins act as receptor antagonists, inhibiting aggregation induced by ADP, thrombin, platelet-activating factor and collagen. The role of disintegrin in preventing blood coagulation renders it of medical interest, particularly with regard to its use as an anti-coagulant (X. Lu, Lu, Scully, & Kakkar, 2006; Xu & Rahman, 2001). The XXCD (HDCD) is a disulphide bonded cysteine sequence (Jia, Wang, Shannon, Bjarnason, & Fox, 1997). Although the exact function of this motif in this domain has been unknown, further research can be performed to unleash if this motif has any therapeutic properties, such as having anticancer or antitumour properties, as it was previously mentioned in this report, disintegrins containing the RGD sequence have been proven to act on ECM and show anticancer effects.

Construction of a phylogenetic tree delivers a visual representation of how the species of organisms, selected by their homologous protein sequences, are evolutionarily related. From the phylogenetic tree, it can be seen that kaouthiagin primarily formed two adjacent clades. One of these branches was exclusively for the zinc metalloproteinases from *Naja atra* and *Naja kaouthia*, snakes from the *Naja* genus. The other branch formed included other Elapid snakes. This suggests that kaouthiagin is evolutionarily distant from the snake venom metalloproteinases of other families of snakes but close to that of *Naja atra*.

The results from SignalP reveals that kaouthiagin is an active protein that is non-secretory. Thus, it has to be first isolated from inside the cell and purified to be used. TMpred predicted a membrane bound region on the protein in its reprolysin or metalloprotease domain.

The 3D structure models produced by both of the homology modelling server, I-TASSER and SwissModel were in consensus. Two servers were used to produce more reliable and refined results. I-TASSER is an online server for structure and function prediction. I-TASSER retrieves templates from the PDB library by LOMETS. For predicting the biological function of the protein, the I-TASSER server matches the predicted 3D models to the proteins in 3 independent libraries which consist of proteins of known enzyme classification (EC) number, gene ontology (GO), and ligand-binding sites. The final results of function predictions are deduced from the consensus of top structural matches with the function scores calculated based on the confidence score of the I-TASSER structural models, the structural similarity between model and templates as evaluated by TM-score, and the sequence identity in the structurally aligned regions (Roy, Kucukural, & Zhang, 2010; Yang & Zhang, 2015). C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5 to 2], where a C-score of higher value signifies a model with a high confidence and vice-versa. Since, model 1 had a higher C-score (1.66) between the two models obtained from I-TASSER, it can be asserted that model 1 was more accurate than model 2. Also, model 1 had a good TM-score of greater than 0.5, which indicates a model of correct topology. Modelling requests are computed by the SwissModel server homology modelling pipeline for the top-ranking templates using ProMod3. Submitting the amino acid sequence of the protein of interest will automatically identify suitable templates based on BLAST and HHbits. The template quality is estimated and ranked from its properties. In the SwissModel Workspace the QMEAN4 score is used to

evaluate the generated models. A higher QMEAN score indicates better quality of the predicted model, which is why model 1 from SwissModel is the most likely model for the query protein. The ligands that may bind to the protein were identical in the results produced from the two tools as well. However, model 2 generated from SwissModel, showed that the Zn ligand binding site was not conserved. From this we can conclude that this model is not reliable. The results from SOPMA generated the secondary structure of the entire sequence of the protein. Predicting the secondary structure of a protein from its primary structure is important to form protein structures.

7.2 FUTURE PROSPECTS:

As SVMPs and their domains possess the property of high specificity and selectivity, they are used in various applications. They have evolved to bind to various integrins, receptors and ECM proteins. The understanding of their structure-function relationships and mechanism of action has significantly contributed to protein chemistry, enzymology, haematology and cancer biology, and thus helped in the development of diagnostic and therapeutic agents. Abberations in normal blood coagulation functions can result in thrombotic disorders or haemorrhage. In thrombosis, largely unknown conditions promote the apparently spontaneous formation of blood clots, large enough to block circulation. Such occurrence of blood clots, block arteries supplying blood to vital organs, in turn causing myocardial infarction or strokes. Anticoagulants are pivotal for the prevention and treatment of thromboembolic disorders (Kini, 2006). Further studies on these proteins will help to unlock several complex lead molecules (Ito et al., 2001). Disintegrins have helped to design drugs for anti-platelet plug and provoked efficient anti-haemostatic effects in several therapeutic trials. More diverse clinical trials should be focused towards a variety of diseases (Matsui, Hamako, & Titani, 2010). Since integrin receptors are quite indiscriminate as they support cell adhesion to several substrates, it seems highly fitting that the RGD-disintegrin scaffold of the integrin binding motif could be employed as a prototype for drug designing for novel anti-metastatic therapies by blocking both tumour cell adhesion and tumour angiogenesis (Sarray et al., 2013). Despite having a high content of disulfide bonds, RGD disintegrin can be produced in their active form in bacteria, thus allowing the production of large quantities needed for tests *in vivo* (Selistre-de-Araujo et al., 2010). Formulation of a tablet form of disintegrins, which is more stable, may be

feasible through coupling to digestive proteins (Matsui et al., 2010). With the increasing aging population and patients suffering from cancer worldwide, more people will require antithrombotic and anticancer therapies in the future. Therefore, creating a recombinant kaouthiagin protein can open windows for more advanced research and the development of novel drugs against haemorrhage and cancer.

CHAPTER 8: CONCLUSION

8.1 CONCLUDING REMARKS:

Snake venoms are a rich source of compounds having diverse pharmacological effects. Toxins from these venoms are selective and potent to vital physiological processes in their prey. As a result, venoms are a veritable mixture of biologically active components, including neurotoxins, mytoxins, enzymes and substances which induce pain, paralysis or death in the prey. The intervention of the scientists in the pharmaceutical development field would employ these molecules as therapeutic agents for several human conditions such as cancer, thrombosis and diabetes. Keeping this in mind, a metalloproteinase known as kaouthiagin, from the local cobra snake *Naja kaouthia* was selected for research in this project. Many of the other species of the Naja family have been worked on previously, for different research purposes on their snake venom. The motive behind choosing *Naja kaouthia* was to find a novel therapeutic agent, from a locally available snake, which would make further research easy and cost efficient. Snake venom metalloprotease and disintegrin were the focal components for this project, since their functions are known to have potential therapeutic properties against blood clotting; haemorrhage and cancer, and have been isolated and put to use from other snake species. Metalloproteases are enzymes that require a divalent metal ion as its cofactor. The SVMP from *Naja kaouthia*, kaouthiagin, specifically works on the VWF. The SVMP had an elucidated amino acid sequence, and defined domains and motifs. The objective was mainly to focus on these domains and identify lead compounds. Kaouthiagin has been identified as a class P-III metallprotease, uniquely with a disintergin and a corresponding disintergin-like domain. Most of the Zn^{2+} ion dependent enzymes induce haemorrhage and some have the ability to degrade protein aggregates like fibrin clots. SVMPs have been applied to the treatment of human conditions involving abnormal blood clot formation. Kaouthiagin was not found to have the property of platelet aggregation which would inhibit haemorrhage. The role of disintegrins as well as metalloproteases in preventing blood coagulation renders them of medical interest, particularly with regard to its use as an anti-coagulant. Along with metalloproteases, disintegrins from snake venom have also shown a high potential for treatment against cancer. A search for new molecule that impairs other survival mechanisms of tumour cells may be necessary to achieve improved results in antimetastatic therapy. Also fascinating is the possible ability of disintegrins to interfere with collective cell migration, which differs from the single cell process mainly by the fact that cells remain coupled by cell to cell junctions while moving.

Thus the intention of this project has been to bring attention to kaouthiagin as a new therapeutic. Using bioinformatics or dry lab methods, the process of these findings were made labour efficient and time efficient. The risk of handling crude snake venom was eliminated, and the length of procedure reduced, as there would have been a broad range of prospective and objectives, but it was minimized as there was a specific motive to work on a very specific protein, the amino acid sequence for which was already discovered. On a positive note, the functions and application or employment of kaouthiagin, as a novel therapeutic still needs to be validated by further extensive wet lab research.

Bibliography

1. Abe, E., Mocharla, H., Yamate, T., Taguchi, Y., & Manolagas, S. C. (1999). Meltrin-alpha, a fusion protein involved in multinucleated giant cell and osteoclast formation. *Calcified Tissue International*, *64*(6), 508–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10341023
2. Bergmeier, W., & Hynes, R. O. (2012). Extracellular matrix proteins in hemostasis and thrombosis. *Cold Spring Harbor Perspectives in Biology*, *4*(2). https://doi.org/10.1101/cshperspect.a005132
3. Bhagwat, M., & Aravind, L. (2007). PSI-BLAST tutorial. *Methods in Molecular Biology (Clifton, N.J.)*, *395*, 177–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17993673
4. Blobel, C. P. (2000). Functional processing of fertilin: evidence for a critical role of proteolysis in sperm maturation and activation. *Reviews of Reproduction*, *5*(2), 75–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10864851
5. Calderon, L. A., Sobrinho, J. C., Zaqueo, K. D., De Moura, A. A., Grabner, A. N., Mazzi, M. V., … Soares, A. M. (2014). Antitumoral activity of snake venom proteins: New trends in cancer therapy. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/203639
6. Chaisakul, J., Hodgson, W. C., Kuruppu, S., & Prasongsook, N. (2016). Effects of animal venoms and toxins on hallmarks of cancer. *Journal of Cancer*, *7*(11), 1571–1578. https://doi.org/10.7150/jca.15309
7. Chellapandi, P. (2014). Structural , Functional and Therapeutic Aspects of Snake Venom Metal- loproteinases, 28–44.
8. Gasanov, S. E., Dagda, R. K., & Rael, E. D. (2014). Snake Venom Cytotoxins, Phospholipase A2s, and Zn(2+)-dependent Metalloproteinases: Mechanisms of Action and Pharmacological Relevance. *Journal of Clinical Toxicology*, *4*(1), 1000181. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24949227
9. Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *"Protein Engineering, Design and Selection," 4*(2), 155–161. https://doi.org/10.1093/protein/4.2.155
10. Hammouda, M. B., Montenegro, M. F., S??nchez-Del-Campo, L., Zakraoui, O., Aloui, Z., Riahi-Chebbi, I., … Essafi-Benkhadir, K. (2016). Lebein, A snake venom

disintegrin, Induces apoptosis in human melanoma cells. *Toxins*, *8*(7). https://doi.org/10.3390/toxins8070206

11. Ikai, A. (1980). Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.*, *1898*(6), 1895–1898. https://doi.org/10.1017/CBO9781107415324.004
12. Ito, M., Hamako, J., Sakurai, Y., Matsumoto, M., Fujimura, Y., Suzuki, M., … Matsui, T. (2001). Complete amino acid sequence of kaouthiagin, a novel cobra venom metalloproteinase with two disintegrin-like sequences. *Biochemistry*, *40*(14), 4503–4511. https://doi.org/10.1021/bi0022700
13. Jia, L. G., Wang, X. M., Shannon, J. D., Bjarnason, J. B., & Fox, J. W. (1997). Function of disintegrin-like/cysteine-rich domains of atrolysin A. Inhibition of platelet aggregation by recombinant protein and peptide antagonists. *The Journal of Biological Chemistry*, *272*(20), 13094–102. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9148922
14. Kadir, M. F., Karmoker, J. R., Alam, M. R., Jahan, S. R., Mahbub, S., & Mia, M. M. K. (2015). Ethnopharmacological survey of medicinal plants used by traditional healers and indigenous people in chittagong hill tracts, bangladesh, for the treatment of snakebite. *Evidence-Based Complementary and Alternative Medicine : eCAM*, *2015*, 871675. https://doi.org/10.1155/2015/871675
15. Kini, R. M. (2006). Anticoagulant proteins from snake venoms: structure, function and mechanism. *The Biochemical Journal*, *397*(3), 377–87. https://doi.org/10.1042/BJ20060302
16. Kini, R. M., & Koh, C. Y. (2016). Metalloproteases affecting blood coagulation, fibrinolysis and platelet aggregation from snake venoms: Definition and nomenclature of interaction sites. *Toxins*, *8*(10), 1–27. https://doi.org/10.3390/toxins8100284
17. Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7108955
18. Lu, Q., Clemetson, J. M., & Clemetson, K. J. (2005). Snake venoms and hemostasis. In *Journal of Thrombosis and Haemostasis* (Vol. 3, pp. 1791–1799). https://doi.org/10.1111/j.1538-7836.2005.01358.x
19. Lu, X., Lu, D., Scully, M. F., & Kakkar, V. V. (2006). Integrins in drug targeting-RGD templates in toxins. *Current Pharmaceutical Design*, *12*(22), 2749–69. https://doi.org/10.2174/138161206777947713
20. Mahanta, M., & Mukherjee, A. K. (2001). Neutralisation of lethality, myotoxicity and

toxic enzymes of Naja kaouthia venom by Mimosa pudica root extracts. *Journal of Ethnopharmacology*, *75*(1), 55–60. https://doi.org/10.1016/S0378-8741(00)00373-1

21. Matsui, T., Hamako, J., & Titani, K. (2010). Structure and function of snake venom proteins affecting platelet plug formation. *Toxins*, *2*(1), 10–23. https://doi.org/10.3390/toxins2010010
22. Mukherjee, A. K., & Maity, C. R. (2002). Biochemical composition, lethality and pathophysiology of venom from two cobras - Naja naja and N. kaouthia. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, *131*(2), 125–132. https://doi.org/10.1016/S1096-4959(01)00473-0
23. Mukherjee, A. K., Saikia, D., & Thakur, R. (2011). Medical and diagnostic applications of snake venom proteomes, *2*(June), 31–40.
24. Naja kaouthia (Monocled Cobra). (n.d.). Retrieved February 21, 2017, from http://www.iucnredlist.org/details/177487/0
25. Naja kaouthia | The Reptile Database. (n.d.). Retrieved February 19, 2017, from http://reptile-database.reptarium.cz/species?genus=Naja&species=kaouthia
26. Ogay, A. Y., Rzhevsky, D. I., Murashev, A. N., Tsetlin, V. I., & Utkin, Y. N. (2005). Weak neurotoxin from Naja kaouthia cobra venom affects haemodynamic regulation by acting on acetylcholine receptors. *Toxicon*, *45*(1), 93–99. https://doi.org/10.1016/j.toxicon.2004.09.014
27. Rawlings, N. D., & Barrett, A. J. (1995). Evolutionary families of metallopeptidases. *Methods in Enzymology*, *248*, 183–228. https://doi.org/10.1016/0076-6879(95)48015-3
28. Rodnight, R. B. (1979). Venoms: Chemistry and Molecular Biology. *Journal of the Royal Society of Medicine*, *72*(1), 79.
29. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5
30. Sarray, S., Luis, J., Ayeb, M. El, & Marrakchi, N. (2013). Snake Venom Peptides: Promising Molecules with Anti-Tumor Effects. *Bioactive Food Peptides in Health and Disease*, 219–238. https://doi.org/10.5772/3318
31. Selistre-de-Araujo, H. S., Pontes, C. L. S., Montenegro, C. F., & Martin, A. C. B. M. (2010). Snake venom disintegrins and cell migration. *Toxins*, *2*(11), 2606–2621. https://doi.org/10.3390/toxins2112606
32. SignalP 4.1 Output format. (n.d.). Retrieved February 22, 2017, from

http://www.cbs.dtu.dk/services/SignalP/output.php

33. Smith, K. M., Gaultier, A., Cousin, H., Alfandari, D., White, J. M., & DeSimone, D. W. (2002). The cysteine-rich domain regulates ADAM protease function in vivo. *The Journal of Cell Biology*, *159*(5), 893–902. https://doi.org/10.1083/jcb.200206023
34. Snake Database. (n.d.-a). Retrieved February 19, 2017, from http://www.snakebd.com/overview.php
35. Snake Database. (n.d.-b). Retrieved February 19, 2017, from http://www.snakebd.com/snakeProfile.php
36. Xu, C.-S., & Rahman, S. (2001). Identification by Site-directed Mutagenesis of Amino Acid Residues Flanking RGD Motifs of Snake Venom Disintegrins for Their Structure and Function. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao Acta Biochimica et Biophysica Sinica*, *33*(2), 153–157. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12050803
37. Yang, J., & Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Curr. Protoc. Bioinform.*, *52*, 5.8.1-5.8.15. https://doi.org/10.1002/0471250953.bi0508s52
38. Zhou, P. (2004). Determining Protein Half-Lives. In *Signal Transduction Protocols* (Vol. 284, pp. 067–078). New Jersey: Humana Press. https://doi.org/10.1385/1-59259-816-1:067