

**Molecular mechanism involved in regulation of Serine Proteinase isoforms  
expression in *Sistrurus catenatus edwardsii***

**A**

**Dissertation**

Submitted in the partial fulfillments of the requirements  
for the award of degree of

**Masters of Science  
In  
Biotechnology**

**By:**

**SUBHI MARWARI**

**(Regn no. - 300901018)**

**Under the supervision of:**

**Dr. R. Manjunatha Kini**

**Internal guide:**

**Dr. M. Sudhakara Reddy**



**Department of Biotechnology and Environmental Sciences**

**Thapar University**

**Patiala-147004**

**July 2011**

## CANDIDATE'S DECLARATION

I hereby declare that the work presented in the dissertation entitled "**Molecular mechanism involved in regulation of Serine Proteinase isoforms expression in *Sistrurus catenatus edwardsii***" in partial fulfilment of the requirement for the award of the degree of Masters in Biotechnology, Department of Biotechnology and Environmental Sciences, Thapar university, Patiala, (Punjab) India, is an authentic record of my own work during the period of six months from Jan 2011 to June 2011, under the supervision of **Prof. R. Manjunatha Kini** (Department of Biological Sciences, National University of Singapore, Singapore) and internal guide **Prof. M. Sudhakara Reddy** (Department of Biotechnology and Environmental Science, Thapar University, Patiala). The report has not been submitted for the award of any other degree or certificate in this or any other University.

Date: July 2011

Place: Patiala

  
Subhi Marwari

# CERTIFICATE

This is to certify that the thesis entitled "Molecular mechanism involved in regulation of Serine Proteinase Isoforms expression in *Sistrurus catenatus edwardsii*" submitted by Subhi Marwari in partial fulfilment of the requirement for the award of Degree of Masters of Science in Biotechnology to Thapar University, Patiala, is a record of student's own work carried out by her. The report has not been submitted for the award of any other degree or certificate in this or any other University or Institute.



Dr. R. Manjunatha KINI

Professor

Department Of Biological Sciences

National University Of Singapore

14 Science Drive 4

Singapore 117543

Professor R. Manjunatha KINI  
Dept. Biol. Sci., National Univ. Singapore  
Singapore 117543  
65-6516-5235

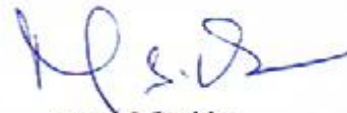


Dr M.S. Reddy

Professor, Internal Guide, Head

DBTES, TU

Patiala



Dr M.S.Reddy

Head of Department

DBTES, TU

Patiala

Dean   
Academic Affairs  
Thapar University  
Patiala

## ACKNOWLEDGEMENTS

Life of not only of human being but also of every living organism is full of struggle. Struggle is not only to compete with external environmental needs but to compete with itself and then, a passion to exist itself according to the need of environment. For me too, life gave me signaling for self sustaining processes which characterize and distinguishes the objects for proceeding ahead. The writing of this section is a time when all the students look forward to fulfill their dreams. The time of reflection causes me to recall all the individuals who have guided and encouraged me along the way. Without their support, this journey was not only difficult but impossible. Hence, I would like to take this opportunity to express my heartfelt gratitude to these people.

I am very thankful to my supervisor, **Professor R. Manjunatha Kini** for his guidance and support. Without his acceptance into the **Protein Science Laboratory, Department of Biological Sciences, National University of Singapore, Singapore** this thesis would not have materialized. Prof Kini has always been an open channel of communication which is a luxury for any student. His door is always open and he would be behind the desk ready for discussion. He has not only taught me to do good quality research, but more importantly, guided me to think critically and independently as an individual.

I am fortunate to have **Professor M. Sudhakara Reddy** as my internal guide and Head of Department of Biotechnology & Environmental Sciences, **Thapar University**. I have deep sense of gratitude towards his guidance and continuous support which has been invaluable in journey of this project. His zest for science will be a constant source of inspiration throughout my scientific career. I attribute the level of my Masters degree to his encouragement and effort and without him this thesis, too, would not have been completed or written.

I owe a great many thanks to **Dr. Niranjan Das**, Associate Professor, without his acceptance, I would not have even thought of taking an initiative to go for this project in NUS. I am highly grateful for his consistent support and suggestions to me.

I would like to thanks to **Thapar University** and faculty members without whom this project would have been distant reality.

The Protein Science Laboratory is like a big extended family. I have made many friends and it has certainly enriched my experience. Firstly, I would like to thank my senior, **Sindhuja Sridharan**, who was my mentor when I was doing my project. Sindhuja imparted many molecular biology techniques which she had painstakingly learnt by herself. I would like to thank postdoctoral fellow, **Ryan McCleary**, my senior, for sharing his snake experiences and guidance for all my queries. I am also thankful to **Guna, Ritu, Varuna, Summer, Bhaskar, Sheena, Bidhan, Girish, Angelina, Amrita, Aldo** for their support and making the lab nice place to work in. I like to give special mention to my lunch and coffee buddies **Anugraha, Kishore, Savitha, Hoi Yee, Vuin, Vick, Remya**. Besides my lab members, special mention has to be given to **Bee Ling**. She has been a great source of support as our lab officer in the lab.

Besides working place in NUS, it was my home in Choa Chu Kang where I had fun with my friends and they provide me a sharing of knowledge of cultural heritage of different countries. I am thankful to **Ana Lee** and **Priscille** for their all caring and loving behavior towards me all the time which never let me realize that I was far away from my home. I am very thankful to my friends, **Farah, Ritzel, Zell, Hurvey, Dennis, Anthony, Michael** and **Suprie** for the all quite friendly environment there. I am also very thankful to **Uncle Tony Sim, Aunty Trita** and **Evon** for providing me the totally family environment there. I feel myself very blessed person of having these people in my life.

I am highly indebted to my dear friend **Meenakshi** who has been keep on updating me regarding all the necessary info of my college work and helped me throughout the way of this long journey. I am also thankful to all my classmates for all the fun we had in these 2 years.

I am very blessed to have my family who has been there to support me. My parents have been a constant source of inspiration and motivation. They have been my harbor where I seek refuge when I feel like giving up. Most importantly, I thank God for always providing and blessing me along this beautiful journey.



Subhi Marwari

July-2011

## ABSTRACT

The relationship between snakes and humans is a turbulent one that has evolved through the millenia. In some cultures, snakes are revered and worshipped, in others they are reviled. Snake venoms are cocktails of pharmacologically active proteins and peptides. Some of these toxins help us as highly specific molecular scalpels in deciphering molecular mechanisms of normal physiological processes. A number of these toxins could also help in developing therapeutic agents for the treatment or prevention of human diseases. A large proportion of the biologically active proteins and peptides present within snake venoms interact with components of the haemostatic system to promote or inhibit the normal sequence of events that lead to clot formation. The venom proteins achieve their effects through interaction with various components of the coagulation cascade, endothelial matrix and platelets. Within the latter group, a number of venom proteins target the interaction of platelets with the major adhesive proteins, Von Willebrand factor and collagen. Snake venoms which are complex mixtures of biologically active proteins and polypeptides, belong to a small number of superfamilies of proteins. The members in a single family show remarkable similarities in their primary, secondary and tertiary structures. At times, however, they differ from each other in their biological targeting and hence their pharmacological effects. Some of the well-recognized families of venom proteins are: (1) Phospholipase A<sub>2</sub> (PLA<sub>2</sub>) (2) Serine Proteinases (3) Metallo Proteinase (4) Three-finger toxin (3FTs) (5) Proteinase Inhibitor and (6) lectin. **Snake Venom Serine Proteinases (SVSPs)** comprise an enzyme superfamily with multifunctional activities that may have diverged or have undergone gene duplication resulting in alteration of their biological properties during the process of evolution thus acquiring special functions (Calvete *et al.*, 2005; Sanz *et al.*, 2006). Although considerable amount of data is now available, no standardised grouping of these venom Serine Proteinases has yet been documented. However, in 2001 Wang et al compared sequences of 40 Serine Proteinases isolated from different snake venoms, using a constructed phylogram in which such sequences were clustered into three groups designated as coagulating enzymes, kininogenases, and plasminogen activators. No Serine Proteinases have yet been purified from venom of the Edward's rattle snake *Sistrurus catenatus edwardsii* in particular or

for members of the *Echis* genus in general. However, the fact that the Serine Proteinase superfamily was important in the venom of the Viperidae, suggested that such enzymes should be present in the venom of *S. c. edwardsii* and that Serine Proteinase-specific antibodies are likely to be an important factor in *S. c. edwardsii* envenoming. Therefore *S. c. edwardsii* cDNA library was screened in order to isolate and characterize different isoforms or variants of this enzyme.

Proteins expressed in the venom gland are highly similar to the physiological protein, and yet are able to inhibit various physiological processes. Hence, it is a well knit mechanism which drives their differential expression in the venom gland. Gene expression is controlled in various levels; transcriptional, translational and post-translational. Transcriptional control resides on the highest in hierarchy of gene regulation as it aids the regulation in the copy of messenger RNA made. The mode of regulation involves the presence of *cis/trans* elements in the DNA which are recognized by the transcription factor to up/down regulate a particular gene.

Total mRNA was isolated from the venom gland of *Sistrurus catenatus edwardsii* and cDNA of *Sistrurus catenatus edwardsii* (Dessert massasauga rattlesnake), reveals the expression of 11 isoforms of Snake Venom Serine Proteinase (SVSPs) were made. SVSPs perturb the hemostatic mechanism of prey. They act on diverse protein substrates such as fibrinogen, kininogen and platelet receptors. Some SVSPs exhibit more than one activity. Further, Real-Time polymerase chain reaction showed the differential expression of these genes. Thus, the present study focuses to understand the intricate mechanism which drives this expression pattern. This would lead to comprehend how genetic element control gene expression and evolution. Identification of 11 isoforms of Serine Proteinase and their differential expression in the venom gland has lead to speculate the presence of novel *cis* elements control the entire pattern. Control of gene transcription is commonly used in biological systems to regulate protein expression. Transcriptional regulation in eukaryotes depends upon a series of complex signal transduction networks that ultimately control gene promoter activity via *cis*-acting elements like enhancers, Matrix Attachment Regions (MARs), locus Control Regions (LCRs), and *trans*-acting elements (transcription factors). Promoters are generally defined as the region of a few hundred base pairs located directly upstream of the site of initiation of transcription. More distal regions and parts of the 5' UTR may also contain regulatory elements and may be part of the promoter. The start of

the CDS (**C**oding **S**equences) only corresponds to the translation start site and gives no hint on the localization of the promoter. Eukaryotic genes usually have 5' untranslated regions (5' UTRs) of variable length in a range of a few base pairs up to several kb. The **five prime untranslated region (5' UTR)** can contain elements for controlling gene expression by way of regulatory elements. It begins at the transcription start site and ends one nucleotide (nt) before the start codon (usually ATG) of the coding region. Our lab with the help of 5'RACE technique extensively identified the 5' UTR region for all the 11 isoforms of Serine proteinases in *S. c. edwardsii*. Identification of 5'UTR region helps us to identify the transcription start site and provide us a better clue to understand the gene organization as well. Therefore, identification of the transcription start site directly leads to the location of the promoter of a gene. Multiple sequence alignment of all the 11 isoforms within their 5'UTR regions reveals the significant similarity and makes the basis for identifying the promoter region of the genes in order to study the varying expression pattern of these isoforms.

# TABLE OF CONTENTS

	<b>Page</b>
<b>Acknowledgement</b>	<b>I</b>
<b>Abstract</b>	<b>III</b>
<b>Table of Contents</b>	<b>VI</b>
<b>List of Figures</b>	<b>VII</b>
<b>List of Tables</b>	<b>VIII</b>
<b>Abbreviations</b>	<b>IX</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
<b>Chapter 2: Review of Literature</b>	<b>19</b>
<b>Chapter 3: Working Methodology</b>	<b>ss 30</b>
<b>Chapter 4: Results and Discussions</b>	<b>65</b>
<b>Conclusion</b>	<b>94</b>
<b>References</b>	<b>96</b>
<b>Annexure</b>	<b>104</b>

# **LIST OF FIGURES**

## **Chapter 1**

Figure 1.1: Crystal structure of trypsin; a typical Serine Proteinase.

Figure 1.2: Catalytic triad in 2 Serine Proteinases.

Figure 1.3: Serine Proteinase Reaction mechanism.

Figure 1.4: Image of snake.

Figure 1.5: Schematic diagramme of venom gland apparatus.

Figure 1.6: Phylogenetic tree of Snake venom Prothrombin Activators and blood coagulation factors.

Figure 1.7: A cis element of human, Murine, TrFX, TrocarinD and PCCS promoter regions.

## **Chapter 2**

Figure 2.1: Transcriptome profile of venom gland of *S. c. edwardsii*.

Figure 2.2: Phylogenetic tree of SVSPs.

## **Chapter 3**

Figure 4.1: Mechanism of SMART cDNA synthesis.

Figure 4.2: Overview of SMART RACE procedure.

Figure 4.3: pJET vector map.

Figure 4.4: Flow chart of genome walking protocol.

## **Chapter 4**

Figure 5.1: Isolation of total RNA from venom gland.

Figure 5.2-5.10: Isoforms gel images.

Figure 5.10: Restriction digested samples of Isoform 5, 7.

Figure 5.11: Isolation of genomic DNA.

Figure 5.12: Restriction digestion of genomic DNA.

Figure 5.13 Genome walker amplified products.

## **LIST OF TABLES**

### **Chapter 1**

Table 1.1: Classification of Snake Venom Prothrombin Activators.

### **Chapter 2**

Table 2.1: Distribution of 576 ESTs sequenced from *S. c. edwardsii*.

## ABBREVIATIONS

APC	activated protein C
A	adenosine
CDS	<b>C</b> oding <b>S</b> equence
FX	blood coagulation factor X
FV	blood coagulation factor
cDNA	complementary deoxyribonucleic acid
DNA	deoxyribonucleic acid
EST	expressed sequence tag
mRNA	messenger ribonucleic acid
PFV	<i>Pseudonaja textilis</i> blood coagulation factor V
PFX	<i>Pseudonaja textilis</i> blood coagulation factor X
PCCS	pseutarin C catalytic subunit
PCNS	pseutarin C nonenzymatic subunit
RT-PCR	real-time polymerase chain reaction
RP-HPLC	reverse-phase high performance liquid chromatography
RNA	ribonucleic acid
SVSP	Snake venom Serine Proteinases
3FTX	three-finger toxin
TLE	thrombin-like enzyme
TSS	transcription start site
TrFX	<i>Tropidechis carinatus</i> blood coagulation factor X
UTR	Untranslated Region
VERSE	Venom Recruitment/Switch Element

---

# *Chapter 1*

---

## **Introduction**

## Chapter 1: Introduction

---

**What is snake venom and how it works and what can be done about it?** Easy questions to ask—very difficult to answer succinctly and simply. Snake venom is saliva, a highly modified saliva, and is produced by modified salivary glands. The proteinaceous nature of snake venom was established by Napoleon Bonaparte's brother, Lucien in 1843. Proteins constitute the major portion of venom's dry weight—90% or more. Snake venom is a cocktail of hundreds, sometimes thousands, of different proteins and enzymes. Snake venom is one of the most amazing and unique adaptations of animal evolution. Venomous serpents have developed one of the most effective and efficient weapon systems of the animal kingdom. Snake venom is one of the most amazing and unique adaptations of animal evolution. Venomous serpents have developed one of the most effective and efficient weapon systems of the animal kingdom. Venom is a prey-immobilizing substance in snakes that is used secondarily as a defense system. Venom is not composed of a single substance, but is a toxic saliva consisting of a complex mixture of chemicals called enzymes. Almost all venoms are composed of approximately 90% proteins. Two general types of toxins are known, neurotoxins and hemotoxins. Neurotoxic venom attacks the victim's central nervous system and usually result in heart failure and/or breathing difficulties. Hemotoxic venom attacks the circulatory system and muscle tissue causing excessive scarring, gangrene, permanent disuse of motor skills, and sometimes leads to amputation of the affected area. The Viperidae family such as rattlesnakes, copperheads and cottonmouths are good examples of snakes that employ mostly hemotoxic venom. Some snakes contain venom that contains combinations of both neurotoxins and hemotoxins. Although no venomous snake has all of these toxins, most snakes employ between six to twelve of these enzymes in their venom. Each of these enzymes has its own special function. Some aid in the digestive process, while others specialize in paralyzing the prey. Scientists believe they have identified the following chemicals from snake venom and the specific purpose of each as follows:

- **Cholinesterase:** attacks the nervous system, relaxing muscles to the point where the victim has very little control.
- **Amino acid oxidase:** plays a part in digestion and the triggering of other enzymes, (is responsible for venom's characteristic light yellowish coloring).
- **Hyaluronidase:** causes other enzymes to be absorbed more rapidly by the victim.
- **Proteinase:** plays a large part in the digestive process, breaking down tissues at an accelerated rate. (causes extensive tissue damage in human victims.
- **Adenosine triphosphatase:** believed to be one of the central agents resulting in the shock of the victim and immobilizing smaller prey. (probably present in most snakes)
- **Phosphodiesterase:** accounts for the negative cardiac reactions in victims, most notably a rapid drop in blood pressure. These are only a few of the enzymes found in the chemistry of snake venom known today.

**Serine Proteinases** or **Serine Endopeptidases** (newer name) are proteases (enzymes that cut peptide bonds in proteins) in which one of the amino acids in the active site of the enzyme is serine. Serine proteinases are grouped into clans that share structural similarities (homology) and are then further subgrouped into families with similar sequences. The major clans found in humans include the chymotrypsin-like, the subtilisin-like, the alpha/beta hydrolase, and signal peptidase clans. In evolutionary history, Serine Proteinases were originally digestive enzymes. In mammals, they evolved by gene duplication to serve functions in blood clotting, the immune system, and inflammation (Conant 1957; Wright *et al.*, 1957). Thrombin and activated Factors 10 ("Xa") and 11 ("XIa") are also **Serine Proteinases**.

The main player in the catalytic mechanism in the chymotrypsin and subtilisin clan enzymes mentioned is the *catalytic triad*. The triad is located in the active site of the enzyme, where catalysis occurs, and is preserved in all Serine Proteinase enzymes. The triad is a coordinated structure consisting of three essential amino acids: histidine (His 57), serine (Ser 195) (hence the name "Serine Proteinase") and aspartic acid (Asp 102). Located very near to one another and near the heart of the enzyme, these three key amino acids each play an essential role in the cleaving ability of the proteinases (Fig 1.2).

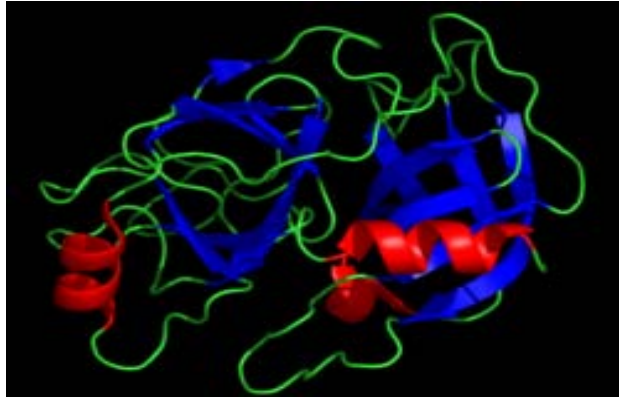
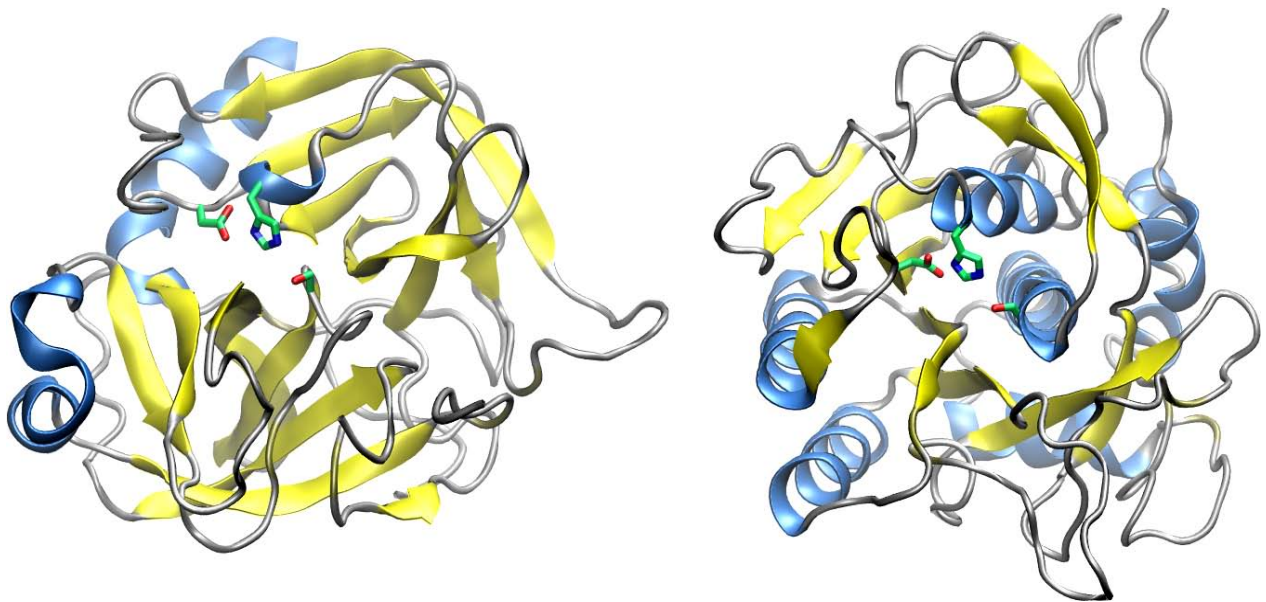


Fig 1.1: Crystal structure of Trypsin, a typical **Serine Proteinase**



**Fig. 1.2: Catalytic triad site in two Serine Proteinases.** Structure of Chymotrypsin (Bleyins 1985) (left) and Subtilisin (Takeuchi *et al.*, 1991) (right) in cartoon format with beta sheets in yellow and alpha helices in blue. Catalytic triad of aspartic acid, histidine and serine (left to right) is shown in stick format with carbons green, nitrogens blue, and oxygens red.

In the event of catalysis, an ordered mechanism occurs in which several intermediates are generated. The catalysis of the peptide cleavage can be seen as a ping-pong catalysis, in which a substrate binds (in this case, the polypeptide being cleaved), a product is released (the N-terminus "half" of the peptide), another substrate binds (in this case, water), and another product is released (the C-terminus "half" of the peptide).

Each amino acid in the triad performs a specific task in this process:

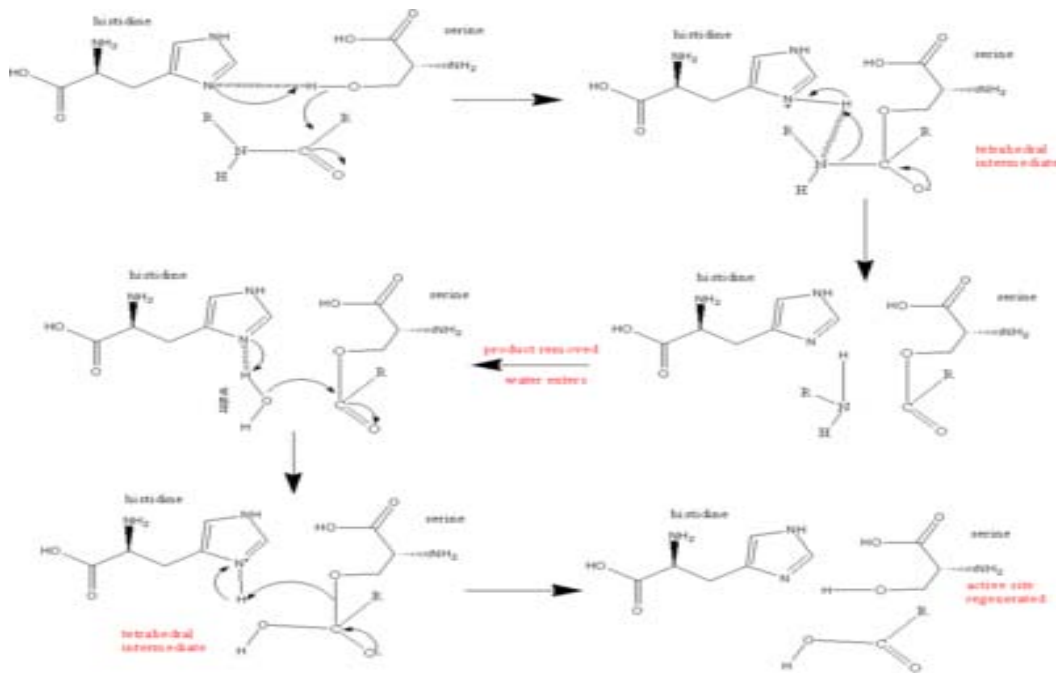


Fig 1.3: The Serine Proteinase reaction mechanism, Source: Transferred from (<http://en.wikipedia.org> en.wikipedia)

The serine has an -OH group that is able to act as a nucleophile, attacking the carbonyl carbon of the scissile peptide bond of the substrate. A pair of electrons on the histidine nitrogen has the ability to accept the hydrogen from the serine -OH group, thus coordinating the attack of the peptide bond. The carboxyl group on the aspartic acid in turn hydrogen bonds with the histidine, making the pair of electrons mentioned above much more electronegative (Fig 1.3).

## **THE SNAKE**

### *Sistrurus catenatus edwardsii*

It is a venomous snake pitviper subspecies found in the southwestern United States and northern Mexico. In places its range overlaps that of *S. c. tergeminus* and intergrading of the two subspecies is not unknown.



**Fig 1.4:** Picture taken from the Image Library of the US-CDC (<http://phil.cdc.gov/Phil/home.asp>)

**Common names:** *desert massasauga, Edward's massasauga, Edward's rattlesnake.*

This subspecies is more slender and smaller than *S. c. tergeminus*, reaching a maximum length of 53 cm. The color pattern consists of a light gray or white base color, with dark gray or gray-brown blotches. They have a distinctive, dark stripe that runs along the side of the head which passes over the eye. Their rattles are significantly higher pitched than those of larger species of rattlesnake, sometimes giving them the nickname **buzztail**.

Compared to *S. c. tergeminus*, it is paler in color and its belly is nearly white. Midbody, it has 23 rows of dorsal scales instead of 25, as well as fewer ventral scales and dorsal blotches (Fig 1.4).

## **Geographic range**

Found in the United States extreme southeastern Arizona, central and southern New Mexico, western Texas about as far north and east as the Colorado River, in the Rio Grande Valley, in many of the Gulf Coast counties about as far north as Brazoria, and on several barrier islands including North Padre Island, Matagorda Island and San José Island. In addition, isolated populations have been reported in northeastern Mexico.

## **Habitat**

Primarily found in rocky, semi-arid and arid areas. According to Conant (1975), it is mostly found in desert grasslands.

## **Behavior**

They are primarily nocturnal, especially during the summer months when it is too hot for them to be active, but they will sometimes be found out sunning themselves.

## **Feeding**

Their diet consists primarily of rodents, lizards and frogs.

## **Conservation status**

The Desert Massasauga is listed as a species of concern in Colorado, due to its limited range in the state, and it is protected by Arizona state law. It is listed as a sensitive species by the United States Forest Service.

Populations of *S. catenatus* generally are threatened or declining rangewide, primarily as a result of habitat loss and human encroachment, and therefore endangered species status has been recommended (Spencer 1985; Donahue *et al.*, 1994). In a systematic study, Holycross and Mackessy showed that among Colorado, Arizona and New Mexico populations of *S. c. edwardsii*, lizards are the major prey, followed by small mammals and centipedes. In the present work, the venom gland has been collected from snakes originating from the Colorado population.

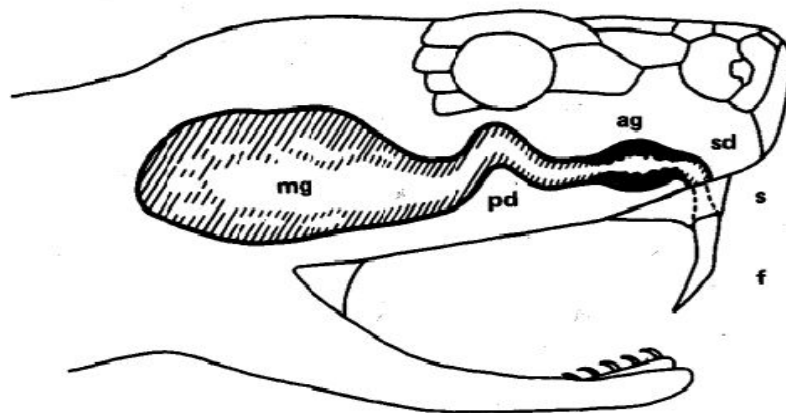
**Snake venom** is highly modified saliva that is produced by special glands of certain species of snakes. The gland which secretes the zootoxin is a modification of the parotid salivary gland of other vertebrates, and is usually situated on each side of the head below and behind the eye, invested in a muscular sheath. It is provided with a large alveoli in which the venom is stored before being conveyed by a duct to the base of the channeled or tubular fang through which it is ejected. Snake venom is a combination of many different proteins and enzymes. Many of these proteins are harmless to humans, but some are toxins.

Snake venom consists of proteins, enzymes, substances with a cytotoxic effect, neurotoxins and coagulants.

- Phosphodiesterases are used to interfere with the prey's cardiac system, mainly to lower the blood pressure.
- Phospholipase A2 causes hemolysis by lysing the phospholipid cell membranes of red blood cells (Wright & Wright 1957).
- Snake venom inhibits cholinesterase to make the prey lose muscle control.
- Hyaluronidase increases tissue permeability to increase the rate that other enzymes are absorbed into the prey's tissues.
- Amino acid oxidases and proteases are used for digestion. Amino acid oxidase also triggers some other enzymes and is responsible for the yellow color of the venom of some species.
- Snake venom often contains ATPase, an enzyme which catalyzes the hydrolysis of ATP to ADP and a free phosphate ion. The presence of enzymes in snake venom has led to the belief that it was an adaptation to assist in the digestion of prey, but studies of the western diamondback rattlesnake, a snake with highly proteolytic venom, show that envenomation has no impact on the time food takes to pass through the gut. More research is needed to determine the selective pressures that have armed snakes in this way (Fry *et al.*, 2005).

***Bioweapon synthesis and storage: The venom gland of *Sistrurus catenatus edwardsii****

The anatomy of venomous snakes is widely varied, but some aspects are universal. All snakes have similar venom delivering apparatus systems comprised of venom glands, a duct with an accessory gland and fangs for venom delivery (Fig 1.5). The venom glands are where venom is produced and they consist of three major cell types: basal cells, conical mitochondria-rich cells, and secretory cells, which produce the venom. Venom production appears to be regulated by the gland itself and is independent of neural control. Kochva determined that the gland contains some baro-sensitive (pressure sensitive) cells, and that fluid pressure inside the gland was the main determinant in rate for venom production.



**Fig 1.5: Schematic drawing of the venom gland apparatus of the venom apparatus of *Crotalidae*. Venom is synthesized and stored in the venom gland (mg) and transported by the primary duct (pd) and secondary duct (sd) which exists into the fang sheath pocket (s), diverting the venom into and through the fangs (f). The loop in the primary duct accommodates the movement of the fang (f). All of these features occur bilaterally. Image taken from *Rattlesnake Venoms*, copyright 1982. Marcel Decker Inc.**

Not all components of venom are produced at the same rate. Venom that is fresh has a different composition than venom that has been allowed to mature because some components of the

venom are not present until the venom is several days old. Venom is carried from the gland to the fang by a duct that flows through an accessory gland. The accessory gland was noted by Mitchell to prevent wasteful flow of the secretions. This gland lacks smooth muscle and is strictly regulated by striated muscle. Implications are that the accessory gland is under voluntary control and that snakes can and do control the amount of venom that is released during a bite. The secretions from the accessory gland are generally non-toxic and mucous-like, but often these secretions mix with the venom and act as an activating complex, which increases the potency of the venom.

Fangs are cone-shaped and tapered and are usually curved and are essentially hollow venom delivering tubes. They can be replaced if lost or damaged and some species even have reserve fangs that remain in a suspended, immature state until stimulated to develop. The fang is the site of envenomation. Some snakes have the ability to spit venom.

### **Envenomation:**

**Envenomation** is the process by which venom is injected into some animal by the bite (or sting) of a venomous animal. Many kinds of animals, including mammals (e.g., the Northern Short-tailed Shrew, *Blarina brevicauda*), reptiles (e.g., the King Cobra), spiders (e.g., Black widows), insects (e.g., wasps, honey bees and caterpillars), employ venom for hunting and for self defense. *Sistrurus catenatus edwardsii* is an extremely venomous snake. Rattle snake venom appears to function primarily as a hemotoxin.

### **Hemotoxic Venoms:**

- Causes abnormal bleeding by interfere with the normal clotting mechanism of blood
- Attacks the circulatory system damages blood platelets and affects normal clotting, causing bleeding signs and symptoms
- Produce local or systemic hemorrhage
- Damages muscle tissue causing excessive Scarring, Gangrene, permanent disuse of muscle, affect motor skill and sometime leads to amputation of affected area.

### ***Body Proteins to Venom Toxins: Changes in Expression Pattern***

For a body protein to be recruited into the venom gland transcriptome, one of the major modifications it must undergo is changes in expression pattern, i.e. expression levels and tissue-specific expression. Unlike the expression pattern of the body proteins which is usually constitutive and low, the expression of the venom toxins are inducible and high. The expression of the venom toxins has to be inducible as the toxins need to be expressed only when the glands are empty and stopped when the glands are full. Also, as the venom is usually an important component of the snake's defense and offense, the venom toxins need to be replenished quickly when the venom gland is emptied (Paine *et al.*, 1992). The venom toxins have elevated levels of expression so that the toxins would be present in sufficient concentrations within the venom to bring about their toxic effects. Besides these, as the venom toxins could be potentially toxic to the snake itself, one of the preventive measures is for the venom toxins to be exclusively expressed in the venom glands. For these changes of expression pattern to occur, there must be modifications to the regulatory *cis* elements of the gene. Previous studies have focused on documenting the evolutionary relationships between the venom toxin proteins and their ancestral parent protein by using phylogenetic analysis and comparing the gene sequences. However, little focus has been given to investigate and verify how these genetic differences actually bring about changes in expression pattern. Therefore, it would be interesting to advance this field of toxicology to its next level by investigating this area. Transcription and its regulation play important roles in controlling expression levels and tissue specific expression. The next section will briefly describe the process of transcription and the mechanisms which regulate it.

## **TRANSCRIPTION IN EUKARYOTES AND ITS REGULATION**

Transcription and translation are two crucial processes for a eukaryotic cell to get instructions for its survival and function from its genetic material. During transcription, genetic material DNA is copied into a complementary RNA nucleotide sequence within the nucleus. This RNA sequence is further processed and matured in mRNA. After maturation, the mRNA is then exported out of the nucleus into the cytoplasm. During translation, based on the information that is encoded by the mRNA, proteins are synthesized and folded in the cytoplasm. Protein expression is tightly regulated from the transcriptional to posttranslational stages to prevent unwanted protein synthesis. If unregulated or incorrect protein synthesis occurs, the cell will not be able to function correctly and it often results in cell death.

### **Transcription factors**

There are different groups of transcription factors that have different properties with regards to regulating transcription. (1) General or basal transcription factors (such as TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH) which form the preinitiation complex. This complex functions to position the RNA polymerase II over the transcription start site and assist in transcription initiation by denaturing the DNA (Lee and Young 2000). (2) Activating transcription factors which “turn on” genes by promoting transcription. They recruit the general transcription factors to the promoter, thereby promoting transcription (Choudhuri 2004). (3) Silencing transcription factors which “turn off” genes by deterring transcription. Some of these factors inhibit transcription by preventing the assembly of general transcription factors (Choudhuri 2004). The availability of these transcription factors also plays a role in transcription regulation. If a gene is to be suppressed, expression of silencing transcription factors may get elevated and/or the activating transcription factors may get sequestered. Vice versa, if a gene is to be expressed, the expression of the activating transcription factors may get elevated and/or the silencing transcription factors may get sequestered.

## **Transcription Regulation**

Although the same set of genes is present in all the cells of an organism, not all the genes are expressed. Each individual cell expresses a distinct subset of genes and one main mechanism that determines this is transcription regulation. During this process, the *cis*-elements of a gene and their corresponding transcription factors are the determining factors which regulate if and when a gene is transcribed, and how much RNA is produced.

### **Regulatory *cis*-elements**

The *cis*-elements determine which transcription factors are attracted to the regulatory regions of the gene. Each transcription factor brings along its influence on the transcription initiation complex, i.e. some promote transcription by increasing the stability the complex and some deter transcription by destabilizing the complex. Whether a gene gets eventually transcribed depends on the overall influence of the transcription factors. These *cis*-elements reside in the non-coding regions of a gene, i.e. from the promoter (or further upstream) to the translational start site, intron regions, and from the stop codon to the 3' end of the gene (untranslated or untranscribed regions). Enhancers are also *cis*-elements which are located upstream or downstream of a gene. Some enhancers are not necessarily near the gene it acts on and can be even found on another chromosome (Spillanakis 2005). Although these enhancers may be far from the gene of interest, the ability of the DNA to fold and coil allows these enhancers to be brought geometrically near to the initiation complex to elicit its effect. Similar to the transcription factors, the steric availability of the *cis* elements also plays a role in transcription regulation. The folding of the gene within a chromosome may cause its *cis*-elements to be made sterically unavailable. When a gene is to be expressed, the chromosome may unfold and expose the regulatory *cis*-elements. Hence, transcription factors are able to interact with the *cis*-elements and promote transcription initiation.

## ***RECRUITMENT OF BLOOD COAGULATION FACTOR INTO THE VENOM TO FUNCTION AS TOXINS***

One group of venom proteins that have been recruited from body proteins by gene duplication are the group C and D prothrombin activators whose ancestral proteins are blood coagulation factor proteins. This thesis aims to investigate how the genetic differences have enabled the blood coagulation factors to change their expression levels and tissue specific expression so that they can function as venom toxins. This section aims to describe how the evolutionary relationship between these two groups of proteins was established.

### **Group C and D Prothrombin Activators**

Snake venoms comprise of a wide range of toxins that target two major systems, namely the neuromuscular and the haemostatic system. The proteins targeting the haemostatic system include procoagulants, anticoagulants, fibrinolytic components, agonists and antagonists of platelet aggregation, and hemorrhagic proteins (Braud *et al.*, 2000). Procoagulant proteins promote blood coagulation and in general are proteinases which activate specific blood coagulation factors (Davie *et al.*, 2003). Prothrombin activators are procoagulant proteins that specifically activate prothrombin and induce coagulation. A wide range of prothrombin activators from snake venoms have been identified and characterized (Gao *et al.*, 2002). This group of proteins differs in their structure, cofactor requirements and end-products formed. Based on these differences, snake venom prothrombin activators are classified into four groups (Kini *et al.*, 2001) (Table 1.1). Group A and B prothrombin activators, such as ecarin from *Echis carinatus* venom (Kornalik *et al.*, 1975) and multactivase from *E. multisquamatus* venom (Yamada *et al.*, 1997) respectively, are metalloproteinases and convert prothrombin to meizothrombin. Group C and D prothrombin activators, such as oscutarin from *Oxyranus scutellatus* venom (Owen *et al.*, 1973) and notecarin from *Notechis scutatus scutatus* venom (Tans *et al.*, 1985) respectively, are serine proteinases and convert prothrombin to thrombin. Our laboratory extensively characterized one representative each of group C (pseutarin C) and group

D (trocarin D) prothrombin activators, and showed that they are very similar to blood coagulation factors. Therefore, it was postulated that these venom prothrombin activators have evolved from blood coagulation factors by gene duplication and accelerated evolution.

**Table 1.1: Classification of snake venom prothrombin activators.**

Group	Cofactor Requirements	Type of proteinase	Product formed	Estimated Size	Similar to Blood coagulation Factors	Examples
A	None	Metalloproteinase	Meizothrombin	~47 kDa	None	Ecarin
B	Ca <sup>2+</sup>	Metalloproteinase	Meizothrombin	Two subunits of ~25 and ~60 kDa	None	Carinactivase, multactivase
C	Ca <sup>2+</sup> plus phospholipids	Serine proteinase	Thrombin	Two subunits of ~60 and ~220 kDa	“FXa-FVa” complex	Oscutarin, pseutarin C
D	Ca <sup>2+</sup> plus phospholipids plus factor Va	Serine proteinase	Thrombin	~60 kDa	FXa	Trocarin D, hopsarin D, notanarin D

### **Venom Prothrombin Activators and Mammalian Blood coagulation Factors are Structurally and Functionally Similar**

Group C and D prothrombin activators are Serine Proteinases, which show functional similarities in their cofactor requirements and products produced to endogenous prothrombin activators such as factor Xa (FXa) (Table 1.1). The structures of group C and D procoagulant proteins were determined to get a detailed understanding of the structure-function relationships of not only snake venom prothrombin activators, but also that of mammalian blood coagulation factors. The size of group D prothrombin activators was smaller and their quaternary structure was simpler than group C prothrombin activators, trocarin D from *Tropidechis carinatus* venom was characterized as a representative of group D prothrombin activators (Joseph 1999). Subsequently, pseutarin C from *Pseudonaja textilis* venom was characterized as an example of group C prothrombin activators (Rao 2003b).

### **Group C Prothrombin Activators:**

This group of Serine Proteinases requires only  $\text{Ca}^{2+}$  ions and negatively charged phospholipids, but not factor Va, for maximal activity (Kochva 1959). They have been purified and characterized from *Oxyuranus scutellatus* (Francis *et al.*, 1992; Grams *et al.*, 1993; Fenton *et al.*, 1995) and *Pseudonaja textilis* venoms (Straight *et al.*, 1976; Servent *et al.*, 1998; Kochva *et al.*, 1996; Freitas *et al.*, 1992). The native activators have a molecular mass of 300 kDa and consist of a factor Xa-like catalytic subunit (60 kDa) and a factor Va like cofactor subunit (200 kDa). The two chains of the factor Xa-like catalytic subunit are disulfide-linked, whereas the two chains of the factor Va-like cofactor are held together by non covalent interactions. The catalytic subunit, similar to factor Xa, has weak catalytic activity in isolation, which is greatly stimulated by the presence of the factor Va-like subunit (Fenton *et al.*, 1995; Straight *et al.*, 1976; Servent *et al.*, 1998). These activators cleave at both Arg 271 -Thr 272 and Arg 320 -Ile 321 bonds of prothrombin (Fenton *et al.*, 1995), converting it to mature thrombin, in contrast to groups A and B activators, which only convert prothrombin to meizothrombin. We have shown that the group C prothrombin activators structurally and functionally resemble the mammalian factor Xa-Va complex (Servent *et al.*, 1998; Kochva *et al.*, 1996; Freitas *et al.*, 1992).

### **Group D Prothrombin Activators:**

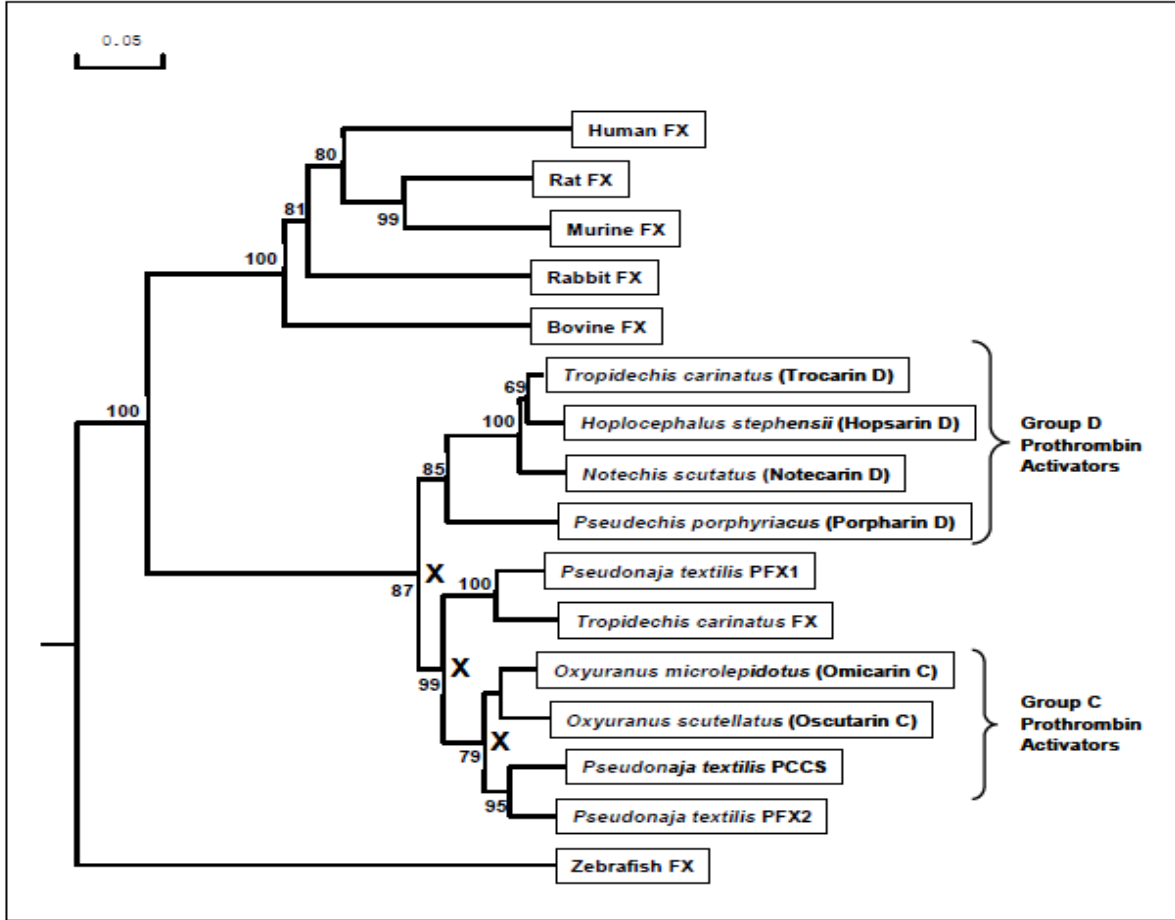
The activities of these Serine Proteinases are strongly stimulated by  $\text{Ca}^{2+}$  ions, factor Va and negatively charged phospholipid vesicles (Fry 2005; Strydom 1973; Patthy 2003; Wang *et al.*, 2001). They are Serine Proteinases with molecular mass ranging from 45 to 47 kDa (Xu *et al.* 1982; Li *et al.*, 2001a). They have two chains held together by disulfide bonds.

The Serine Proteinase active site is located in the heavy chain (Wang *et al.*, 2001; Xu *et al.*, 1982). They exhibit potent procoagulant effects, comparable to mammalian factor Xa, through activation of prothrombin. Similar to factor Xa, they cleave prothrombin at two sites, Arg 271 -Thr 272 and Arg 320 -Ile 321. Their poor activator activity is stimulated by more than a million-fold by negatively charged phospholipids, factor Va and  $\text{Ca}^{2+}$  ions. They also hydrolyze factor Xa-specific chromogenic substrates. Recently, we determined the complete amino acid sequences of trocarin D (Xu *et al.*, 1982) and hopsarin D (Li *et al.*, 2001a). They share high

identity (53–60%) and similarity (62–70%), and hence similar domain architecture, with factor Xa. Their light chains consist of an N-terminal Gla domain (residues 1–39), followed by two epidermal growth factor-like domains, EGF-I (residues 50–81) and EGF-II (residues 89–124). Thus, group D venom prothrombin activators are true structural and functional homologues of blood coagulation factors.

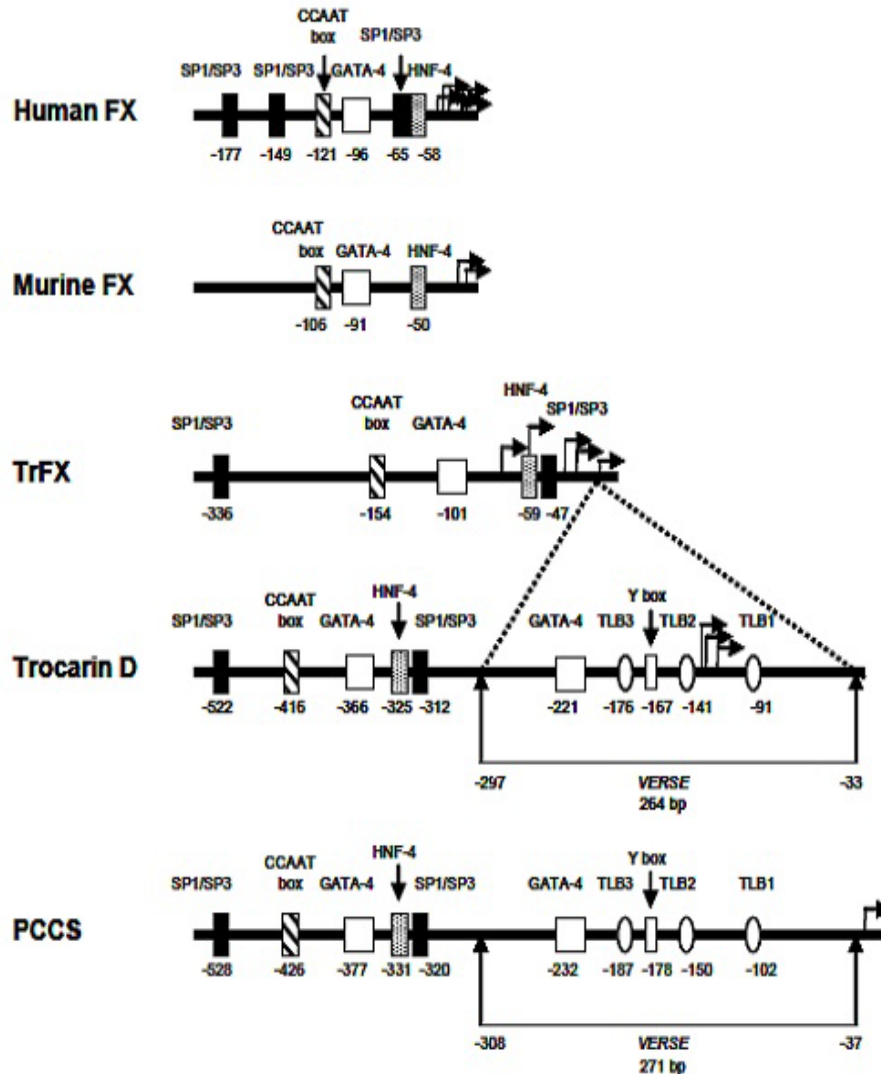
## **Evolutionary relationships of venom prothrombin activators and blood coagulation factors**

A detailed sequence analysis of the venom prothrombin activators (trocarin D and pseutarin C) and the plasma prothrombin activators (PFX1, PFX2 and TrFX) reveals evolutionary relationships between them (Reza *et al.*, 2005a; Reza *et al.*, 2006). PFX1 has a higher identity (~94%) to TrFX than PFX2, whereas PFX2 shows higher identity (~90%) to PCCS and trocarin D than PFX1. A comparison of the activation peptides from the deduced amino acid sequences shows that, while the size of PFX1 activation peptide (57 residues) is identical to that of TrFX, the size of the shorter PFX2 activation peptide (27 residues) is identical to its venom counterparts, PCCS and trocarin D (Reza *et al.*, 2006). PFX2 also has an insertion in its heavy chain which is the exact position as that in PCCS and trocarin D. However, the insertion in PFX2 is nine residues long while the insertions in PCCS and trocarin D are 13 and 12 residues long respectively. These patterns observed suggest that PFX2 is an intermediate between PFX1 and PCCS (Reza *et al.*, 2006). Therefore, PFX2 is proposed to be an evolutionary intermediate and a “molecular fossil” of the evolution of venom prothrombin activators from blood coagulation factors (Reza *et al.*, 2006). A phylogenetic tree of the snake venom and plasma prothrombin activators with other known FX sequences reveals more evolutionary relationships (Reza *et al.*, 2006). All the reptilian sequences come out in a monophyletic group (Reza *et al.*, 2005b). Within the reptilian clade, group C and D prothrombin activators appear as two separate clades on the tree. Interestingly, the PFX2 sequence is found nested within the group C prothrombin activators. This enhances the theory that PFX2 is an evolutionary intermediate of PCCS from PFX1. Based on the topology of the phylogenetic tree, it is suggested that to the existence of snake venom prothrombin activators could have been established through three independent evolutionary events (Fig. 1.6) (Reza *et al.*, 2005b).



**Fig 1.6: Phylogenetic tree of snake venom prothrombin activators and blood coagulation factors. “X” indicates possible independent gene duplication events.**

## A – Promoter region



**Fig 1.7: A Cis element of human, murine, TrFX, trocarin D and PCCS Promoter regions. The identification of *cis* elements in human and murine FX are experimentally supported, whereas *cis* elements in TrFX, trocarin D and PCCS promoter regions are predicted by sequence comparison (Huang *et al.*, 1992; Reza *et al.*, 2007; Wilberding and Castellino 2000).**

---

# *Chapter 2*

---

## **Review of Literature**

## Chapter 2: Review of Literature

---

The advanced snakes (superfamily Colubroidea) consist of a monophyletic group of four families: Atractaspididae, "Colubridae", Elapidae and Viperidae. These snakes have evolved biochemical weapon (toxins), rather than mechanical means of handling prey. Phylogenetic studies show that the venom gland (where toxins are produced) evolved once at the base of the Colubroidea about 60–80 million years ago and has undergone extensive "evolutionary tinkering" of delivery systems and compositions of venom (Mackessy 2002; Vidal *et al.*, 2002). Phylogenetic reconstruction between toxin genes and snake families showed that the recruitment of toxin families into the venom gland has occurred multiple times by both basal (e.g. metalloproteinases, CRISP, Kunitz-type serine protease inhibitors, NGF) and independent (e.g. PLA2, natriuretic peptides) recruitment events (Fry *et al.*, 2004). Approximately 26 families of toxins have been catalogued in snake venom proteomes, and several families appear to be specific to a particular family of venomous snakes. Sarafotoxins are found only in venoms of Atractaspididae; Serine Proteinases related to blood coagulation factors Xa, cobra venom factor, waprins and AVIT (prokineticin) family peptides appear to be limited to the Elapidae; and vascular endothelial growth factor (VEGF), disintegrins, waglerins, dipeptidyl peptidase IV and crotamine occur primarily in venoms of the Viperidae. The occurrence, relative abundance and pharmacological potency of various members of these toxin families in venom make envenomation remarkably complex. Envenomation by elapid snakes is usually characterized by rapid neurotoxic complications due to presence of large amounts of postsynaptic neurotoxins (Hodgson *et al.*, 1991), while envenomation by viperid snakes evokes complex hemorrhagic, hypotensive and inflammatory effects caused by the actions of numerous serine proteinases, metalloproteinases and C-type lectins (CLP) (Meir *et al.*, 1991; Braud *et al.*, 2000; Matsui *et al.*, 2000; Morita 2004). Effects of envenomation by snakes in the genus *Atractaspis* can include vasoconstriction, resulting in cardiac arrest (Kloog *et al.*, 1988). Despite overall similarity in clinical symptoms exhibited after envenomation by members of a particular family of snakes, there exists considerable species-specific variation in absolute effects within each group, contributing to the difficulty in assessing and treating envenomated victims.

Previously, identification and characterization of venom components relied primarily on various methods in protein chemistry or on cloning of individual genes. However, neither approach is well-suited to detect toxins that are found in low abundance. Therefore, the apparent absence of a particular family of toxins from venom could be due to either to their very low abundance or to the lack of expression in the venom gland. The genes of low abundance toxins are best discovered by the construction of a cDNA library and sequencing of a sizeable number of ESTs. Using this approach, new toxin genes in known families as well as several completely new families of toxins have been discovered, and the spectrum of snake toxin proteome is gradually expanding (Bazza *et al.*, 2005; Birrel *et al.*, 2006; Chinag *et al.*, 2006; Cidade *et al.*, 2006; Francischetti *et al.*, 2004; Juraz *et al.*, 2004; Junqueira-de-Azevedo *et al.*, 2002; Kashima *et al.*, 2004; Li M. *et al.*, 2005; Li S. *et al.*, 2005; Nawark *et al.*, 2003; Pierre *et al.*, 2005; Serrano *et al.*, 2005; Tsai *et al.*, 2004; Quinghua *et al.*, 2006; Wagstaff *et al.*, 2006; Zhang *et al.*, 2006;). To search for novel and low abundance toxin genes or new families of toxins, we constructed a cDNA library and sequenced ESTs from the venom gland of *Sistrurus catenatus edwardsii* (Desert Massasauga).

*Sistrurus catenatus* (Massasauga Rattlesnake) is a small pitviper broadly distributed across the North American prairies from Ontario, Canada and New York to extreme southeastern Arizona, with an apparently disjunct population in northern Chihuahua, Mexico (Mackessy *et al.*, 2005; Stebbins *et al.*, 1985). One subspecies, *S. c. edwardsii* (Desert Massasauga), occurs primarily in arid and desert grasslands, occasionally occurring in dune formations and desert scrub (Degenhardt *et al.*, 1996; Hobert *et al.*, 2004; Holycross *et al.*, 2002). Populations of *S. catenatus* generally are threatened or declining rangewide, primarily as a result of habitat loss and human encroachment, and therefore endangered species status has been recommended (Allen 1988; Seigel 1986). In a systematic study, Holycross and Mackessy showed that among Colorado, Arizona and New Mexico populations of *S. c. edwardsii*, lizards are the major prey, followed by small mammals and centipedes. In the present work, the venom gland has been collected from snakes originating from the Colorado population.

General symptoms of envenomation resulting from many North American pitvipers bite are pain, local tissue effects (progressive edema, erythema and necrosis) with coagulopathy (hypofibrinogenemia and prolongation of prothrombin time) and thrombocytopenia as systemic

effects. However, there is no specific report to date in the literature concerning envenomation by *S. c. edwardsii*. Profiling of toxin expression of this threatened snake species will give a global view for the expression of all genes of toxins, including variation in coding/noncoding sequences and their evolutionary trends. The results of this study will also help in the understanding of envenomation processes of rattlesnake bites, which in turn will be important for more effective clinical treatment and antivenom management in cases of snakebite.

### **Results obtained in background of research:**

A total of 518 out of 576 ESTs produced readable sequences. The sizes of sequences showed a distribution between 300 and 2000 base pairs, with an average of 800 base pairs. A total of 232 clusters were obtained and subsequently all clusters checked manually (Table 1; Figures 2.1a and 2.1b). A large number of ESTs matched with snake toxins (360 ESTs in 76 clusters; 69.4%). Others code for cellular (nontoxin) proteins (107 ESTs in 106 clusters; 20.65%), and 42 hypothetical ESTs (8.1%). Nine ESTs (1.7%) matched with mitochondrial genes. Fifteen ESTs did not significantly match with any sequence available in non-redundant databases. Further, they do not have any ORFs and may represent either long UTRs (3' or 5') or regulatory RNAs and may have important functions in the rapidly expressing gland tissues. The library contains a large portion of putative toxin genes (69.4%) compared to the cellular EST population (20.6%). We determined the complete sequence of the longest EST of each cluster and sequences were confirmed by repeated sequencing. We completed the sequencing for all major toxins and two low abundant toxin-like transcripts. The existence of genes of particular interest, especially singletons, was confirmed by RT-PCR, using a separate pool of RNA that was used to make cDNA library as template, followed by sequencing.

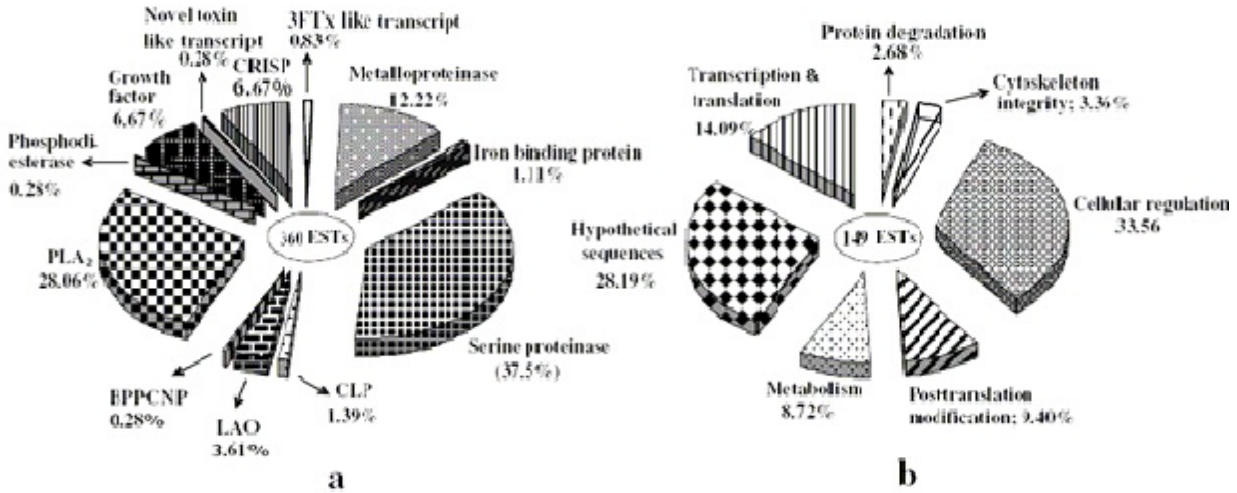
**Table 2.1: Distribution of 576 ESTs sequenced from *S. c. edwardsii* venom gland in toxin and toxin like transcripts, cellular proteins and hypothetical sequence clusters.**

Transcripts category	Number of ESTs	Number of clusters	Redundancy (clones/cluster)	Representation over total clones (%)	Representation over matching clones (%)
Toxin	360	76	4.74	69.40	77
Cellular	107	106	1.04	20.65	23
Mitochondrial	9	9	1.00	1.73	-
Hypothetical	42	42	1.00	8.10	-

The variations observed among toxin isoforms in this library are not due to population level diversity as they are derived from the venom gland of a single snake

### Confirmation of species:

Taxonomic identification at the molecular level is essential to ensure species identity (Pook *et al.*, 2005), and 12S and 16S mitochondrial ribosomal RNAs are commonly used in the classification of snakes (Wuster *et al.*, 1999). Three ESTs for the 12S RNA gene in the library show 100% identity to the reported *S. c. edwardsii* ribosomal sequence (AF057227) (Parkinson *et al.*, 1999), confirming the venom gland used to make the library is of *S. c. edwardsii* origin. Interestingly, we observed that the ribosomal RNA sequence has poly (A) tail and therefore they appeared in the cDNA library. Polyadenylation of ribosomal RNA has been observed in yeast (*Candida albicans*), fungus (*Saccharomyces cerevisiae*), protistan parasites (*Leishmania braziliensis* and *L. donovani*) and human (*Homo sapiens*) cells, and it is proposed to have a quality control role in rRNA degradation (Decuyper *et al.*, 2005; Fleishmann *et al.*, 2004; Kuai *et al.*, 2004; Decuyper *et al.*, 2005; Fleishmann *et al.*, 2004; Kuai *et al.*, 2004). This is a preliminary report showing the possibility of polyadenylation of ribosomal RNA in a reptilian system. On closer examination, we found a putative polyadenylation signal (Wahle and Kuhn 1997) sequence six bases upstream of the poly (A) tail.



**Figure 2.1:** Transcriptome profile of the venom gland of *S. c. edwardsii*. The transcriptome profile of the venom gland of *S. c. edwardsii*. Abundance of (a) toxins and toxin-like transcript clusters, and (b) cellular proteins and hypothetical sequences clusters. Percentage of total ESTs for each category are shown.

## Identification of toxin family:

### ***Serine Proteinase***

The Serine Proteinases in the venom gland library of *S. c. edwardsii* are expressed with the highest transcript abundance (38% of 360 ESTs) (Figure 2.1a) and belong to 19 clusters. Multiple clones appeared in 12 clusters, while 7 were singletons. One representative EST from each cluster was completely sequenced (DQ464238–DQ464248, DQ439973). One of the clusters (DQ439973) contains only 3'UTR (2 ESTs). This cluster shows 90% similarity with the 3'UTR of a serine proteinase from *Bothrops jararaca* venom gland (Saguchi *et al.*, 2005). Most snake venom serine proteinases (SVSPs) to date are single polypeptide chains, except for two fibrinolytic enzymes from the venom of a Korean Viper, *Agkistrodon blomhoffi brevicaudus* (brevinase, AJ243757 and salmonase, AF176679). In both cases, a single chain precursor is most likely cleaved by proteolysis (Lee *et al.*, 1999). In our library, one cluster (DQ464244) shows

90% and 83% sequence identity at the nucleotide and amino acid levels respectively to salmonase. It is not clear whether or not it is also processed to form a heterodimeric Serine Proteinase in *S. c. edwardsii* venom. The SVSPs are found in all families of snakes and in general, they perturb the hemostatic mechanisms of prey. They act on diverse protein substrates such as fibrinogen, kininogen and platelet receptors (Serrano *et al.*, 2005; Wang *et al.*, 2001). Some SVSPs exhibit more than one activity. For example, in addition to their thrombin-like activity, both thrombin, crotalase and LM-TL induce platelet aggregation, kinin release and gyratory activities, respectively (Markland *et al.*, 1982; Nishida *et al.*, 1994; Silveira *et al.*, 1989). We constructed a neighbor-joining (NJ) phylogenetic tree with 11 newly identified SVSP isoforms from *S. c. edwardsii*, to assign putative functions and to examine trends in the evolution of new isoforms (Wang *et al.*, 2001) (Figure 2.2). The phylogenetic tree showed a scattered distribution of various isoforms with different pharmacological activities from several species of pitvipers. This pattern indicates that SVSPs diverged after snake lineages speciated. Many SVSPs are commonly considered as thrombin-like enzymes (TLEs) because they mimic the fibrinolytic function of thrombin, promoting blood coagulation. Therefore, in most cases only fibrinolytic function of SVSPs is tested and the SVSP is categorized as a TLE. However, some thrombin-like enzymes, in addition to releasing fibrinopeptide A and/or B from fibrinogen, also activate protein C (Kisiel *et al.*, 1987), complement C3 (Yamamoto *et al.*, 2002) and platelets (Niewiarowski *et al.*, 1979). Therefore, it would be interesting to determine the specific pharmacological properties of various SVSP isoforms within each group and map these on their evolutionary relationships. SVSP genes belong to a multigene family, and the protein coding regions have been shown to be experiencing accelerated evolution within the venom glands of pitvipers (Deshimaru *et al.*, 1996). Such accelerated evolution could lead to the changes in surface loops surrounding the substrate binding site, resulting in the variation of substrate recognition and hence, the function of the protein. The ratio between nonsynonymous and synonymous substitution ( $dN/dS$ ) of the protein coding sequences of serine proteinase isoforms of this species was found to be 0.99, indicating a trend toward accelerated evolution and therefore divergence in pharmacological function during envenomation.

## **Genetics of origin of Serine Proteinase toxin:**

The mechanism for the creation of these genes needs to be studied further, it clearly indicates that other genetic processes (gene shuffling or TIC) are also operating in the venom gland to create novel toxin genes. Genes originating by other genetic processes such as exon shuffling are recent (Long *et al.*, 2003), and therefore the addition of this fused toxin-like transcript to the venom proteome is perhaps new. At this stage, it is tempting to speculate that the origin of modular organization of different classes of SVMPs, which appears to be the result of gene fusion events, may be due to a genetic process other than gene duplication. SVMPs are very abundant toxins and carry out a principal role in envenomation by viperid snakes, and therefore studies of their genetic origin and organization will be of great interest. Circumstantial evidence of trans-splicing for the generation of serine proteinase isoforms in the venom gland of *V. lebetina* has been presented (Santamaria *et al.*, 2005; Siigur *et al.*, 2001).

Kopelman *et al.*, (2005) have shown that alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. According to Parra *et al.*, only 4–5% of the tandem gene pairs in the human genome can produce chimeric proteins. It is obvious that these alternative genetic processes responsible for expanding functional proteomes are uncommon among biological systems, and it is therefore not surprising in our case to have just a singleton of the fused transcript out of 576 ESTs (transcript abundance 0.28%). This also demonstrates that to detect rare genetic processes operating in the venom gland, the library generated must be of high quality and that subsequent analyses must be performed very carefully. In turn, these analyses help elucidate in detail the principles of evolution of snake venom transcriptome which have led to the evolutionary success of the advanced snakes (Mackessy *et al.*, 2006).

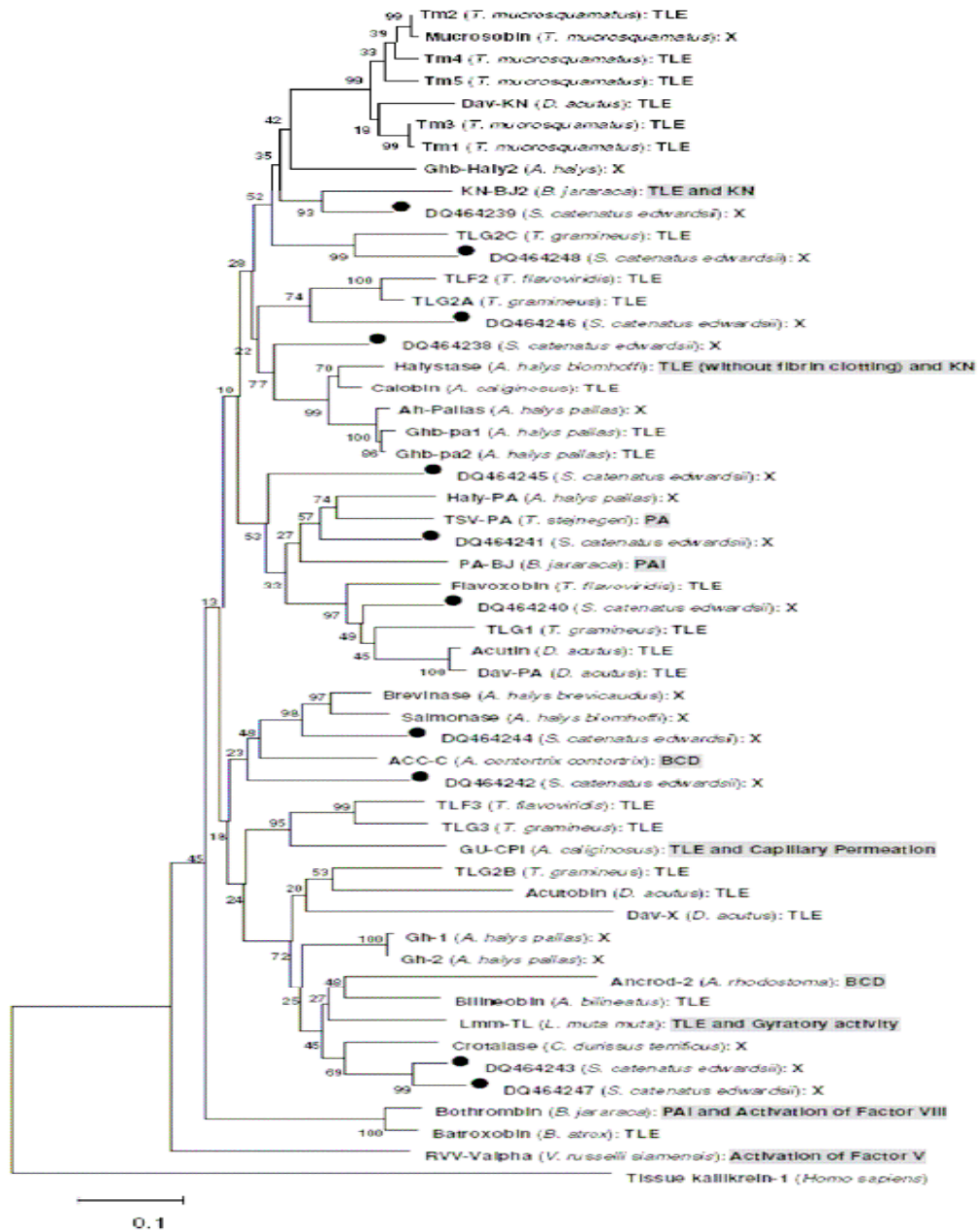


Figure 2.2: Phylogenetic (NJ) tree of SVSPs. Sequences (complete ORF) available from other pit vipers (Yamada *et al.*, 1997) and 11 isoforms (DQ464238–DQ464248) from this study (filled circle) were used. Tissue kallikrein-1 (P06870) is used as outgroup. The numbers on the branches indicate the bootstrap support values for nodes, and the horizontal bar represents number of substitutions per site. TLE, thrombin-like enzymes; KN, kininogenase; PA, plasminogen activator; PAI, platelet aggregation inducer; BCD, blood clot dispersion; X, activity unknown. Experimentally verified activities are shaded.

Along this identification of 5'UTR region reveals the information of Transcription Start Site in all the isoforms. Despite high sequence similarity in the cDNAs of all isoforms, there are significant differences to indicate the variable level of expression in the particular, giving the better understanding of molecular mechanisms involved. Real time PCR reveals the differences in expression patterns. The gene organizations will be determined to understand the underlying causes of these differences in expression patterns.

The thesis aims to identify the 5'UTR region which can reveal the information about transcription start site and thus can help in understanding the gene organization. Further this work aims to identify the promoter region for their role in regulating elevated and tissue specific expression. The objectives of this thesis are:

1. To identify the 5'UTR region of all the isoforms which is required for the proximal promoter activity.
2. To identify the Transcription start site from 5'UTR region which gives the hint in understanding the gene organization of all the isoforms.
3. To establish the phylogenetic relationship of all the isoforms in the course of evolution.
4. To identify the promoter region for its role in venom recruitment using genome walking and bioinformatics tools.
5. To identify and characterize the *cis/trans* factor elements which interact with the promoter cis elements.

On the whole, the scope of this thesis covers the approaches used to fulfill these aims, the results obtained and the discussion of their implications. Explicitly, we have characterized the 5' UTR sequences of all the isoforms using 5'RACE techniques and bioinformatics tools. Further using the genome walking experiments and bioinformatics tools, promoter region of all the isoforms can be depicted. *Cis/trans* factor elements can be characterized which integrate with promoter region,

---

# *Chapter 3*

---

## **Materials and Methods**

## Chapter 3: Materials and Methods

---

### **Materials**

Specimens of *Sistrurus catenatus edwardsii* and liver tissues were provided by Venom Supplies Pty Ltd, Colorado, USA.

The DNeasy Tissue kit and RNeasy Mini kit were purchased from Qiagen (Valencia, CA, USA).

The Long PCR Enzyme Mix was purchased from Fermentas (Hanover, MD, USA).

The cell culture media were purchased from Sigma (St. Louis, MO, USA).

The pJET®-T Easy vector, The GeneAll Plasmid Quick kit was purchased from GeneAll Biotechnology (Seoul, Korea).

Oligonucleotides primers were ordered from 1st BASE (Singapore).

The SMART RACE cDNA Amplification kit and Advantage 2 PCR Enzyme system were purchased from Clontech Laboratories Inc. (Palo Alto, CA, USA).

The ABI PRISM® BigDye® terminator cycle sequencing ready reaction kit was purchased from Perkin Elmer (Foster City, CA, USA).

Gel extraction kit includes QIAquick spin columns, Buffer QG, Buffer PE, Buffer EB, collection tubes.

GenomeWalker Universal kit were purchased from Perkin Elmer (A Takara Bio Company 1290 Terra Bella Ave. Mountain View, CA 94043 Technical Support (US).

## **WORKING METHODOLOGY**

### **(1) Venom extraction and collection:**

Specimens of *Sistrurus c. edwardsii* (Desert Massasauga) were collected in Lincoln County, USA under permits granted by the Colorado Division of Wildlife to SPM. Venom was extracted from adult snakes using standard manual methods (Mackessy *et al.*, 1988); venoms were then centrifuged to remove particulates, frozen and lyophilized. Prior to gland removal, snakes were extracted of venom. Four days later, when mRNA levels are presumed maximal (Rotenberg *et al.*, 1971), two snakes were anesthetized with isoflurane and then sacrificed by decapitation. Glands were then rapidly dissected from the snakes, cut into small pieces and placed in approximately 0.5 ml RNA later and frozen at -80°C until used.

### **(2) Isolation of mRNA from venom gland:**

#### **(Guanidinium thiocyanate Phenol Chloroform extraction)**

Tissue sample from venom gland was excised and 10 mgm of tissue was cut into small pieces. Tissue was lysed in 700 µl of RLT buffer (contains Guanidinium thiocyanate [GITC]). After that 7 µl of β mercaptoethanol was added to it [β-mercaptoethanol is prepared in 70% ethanol]. The above solution was homogenized and 700 µl of RLT buffer was again added to the tissue. The solution was centrifuged at 14,000 rpm for 3 min. Supernatant was carefully removed and transferred into new microcentrifuge tube. One volume of 70% ethanol was added to the clear lysate and was mixed immediately by pipetting. 700 µl of sample was transferred to RNeasy spin column and was centrifuged at 14, 000 rpm for 3 min. Flow through was discarded and 350 µl of wash buffer (Buffer RPE) was added to remove excess GITC. After that DNase enzyme was added with the buffer to remove DNA. Again the 350 µl of RPE buffer was added and was incubated for 15 min. After that solution was centrifuged at 14,000 rpm for 1 min. Flow through was discarded and wash buffer was added in spin column. It was spin down again at 14,000 rpm for 30 sec. Then empty spin was given again to get rid of ethanol. 30 µl of RNase free water was added to elute the RNA. It was spin down again at 14K rpm for 30 sec.

### **RNA Analysis:**

RNA was examined by electrophoresing a sample on a denaturing formaldehyde agarose/EtBr gel. Mammalian total RNA typically exhibits two bright bands at 4.5 and 1.9 Kb; these bands correspond to 28S and 18S ribosomal RNA, respectively. The ratio of intensities of these bands should be about 1–2:1. Poly A<sup>+</sup> RNA samples from mammalian cells should produce smears from 0.5–12 Kb with much weaker ribosomal RNA bands. Size distribution may be smaller with nonmammalian tissue sources.

### **(3) RACE ( Rapid Amplification Of cDNA Ends) :**

RACE is a technique used in molecular biology to obtain the full length sequence of an RNA transcript found within a cell. RACE results in the production of a cDNA copy of the RNA. Sequence of interest, is produced through reverse transcription, followed by PCR amplification of the cDNA copies. The amplified cDNA copies are then sequenced and, if long enough, should map to a unique mRNA already described, the full sequence of which is known. RACE can provide the sequence of an RNA transcript from a small known sequence within the transcript to the 5' end (5' RACE-PCR) or 3' end (3' RACE-PCR) of the RNA. This technique is sometimes called *one-sided PCR* or *anchored PCR*.

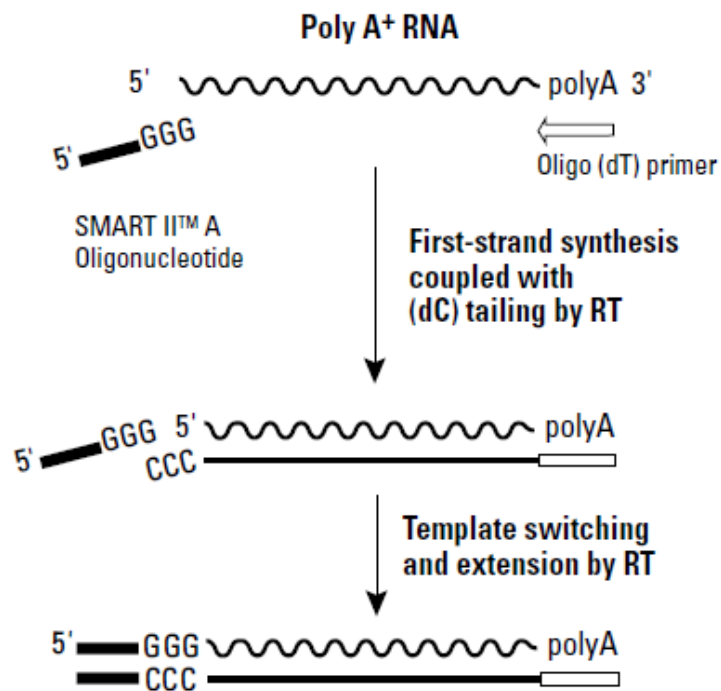
### **SMART™ 5' RACE cDNA Amplification:**

#### **(a) First-Strand cDNA synthesis:**

##### **Introduction & Protocol Overview:**

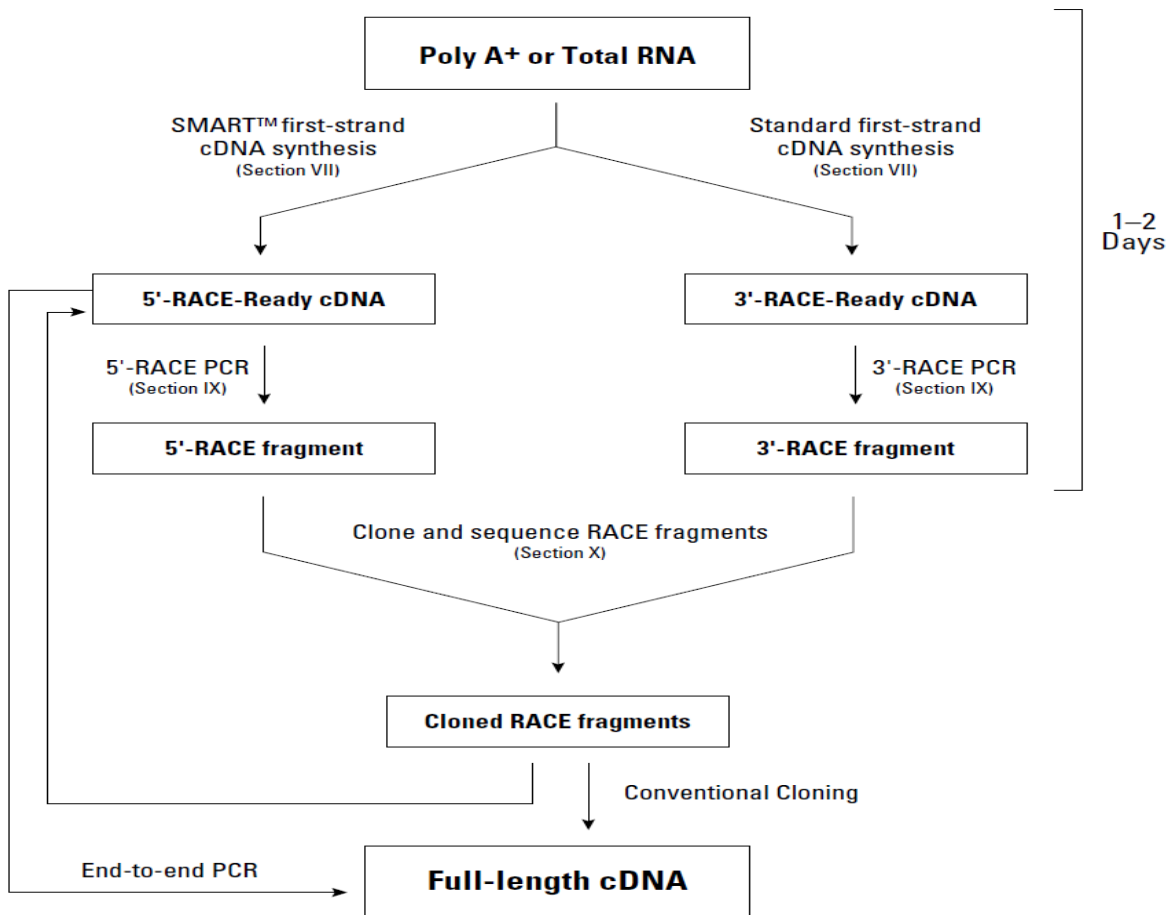
The SMART™ RACE cDNA Amplification Kit provides a method for performing both 5'- and 3'-rapid amplification of cDNA ends (RACE). This kit integrates MarathonR cDNA Amplification Kit (Chenchik, A. et al. 1995; 1996) with SMART (Switching Mechanism At 5' end of RNA Transcript) cDNA synthesis technology. This powerful combination allows us to isolate the complete 5' sequence of your target transcript more consistently than ever before. Furthermore, SMART technology eliminates the need for problematic adaptor ligation and lets us use first-strand cDNA directly in RACE PCR, a benefit that makes RACE far less complex and much faster (Chenchik *et al.*, 1998). The SMART RACE Kit also includes advances in PCR technology that both increase the sensitivity and reduce the background of the RACE reactions.

As a result we can use either poly A<sup>+</sup> or total RNA as starting material for constructing full-length cDNAs of even very rare transcripts. SMART technology provides a mechanism for generating full-length cDNAs in reverse transcription reactions (Zhu *et al.*, 2001). This is made possible by the joint action of the SMART II™ A Oligonucleotide and the Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV RT). The MMLV RT, upon reaching the end of an RNA template, exhibits terminal transferase activity, adding 3–5 residues (predominantly dC) to the 3' end of the first strand cDNA (Figure 4.1). The SMART oligo contains a terminal stretch of G residues that anneal to the dC-rich cDNA tail and serves as an extended template for RT. MMLV RT switches templates from the mRNA molecule to the SMART oligo, generating a complete cDNA copy of the original RNA with the additional SMART sequence at the end. Since the dC-tailing activity of RT is most efficient if the enzyme has reached the end of the RNA template, the SMART sequence is typically added only to complete first strand cDNAs. This process guarantees that the use of high quality RNA will result in the formation of a set of cDNAs that have a maximum amount of 5' sequence.



**Figure 4.1: Mechanism of SMART cDNA synthesis.**

First-strand synthesis is primed using a modified oligo (dT) primer. After reverse transcriptase reaches the end of the mRNA template, it adds several dC residues. The SMART II A Oligonucleotide anneals to the tail of the cDNA and serves as an extended template for MMLV RT. Following reverse transcription, the first-strand cDNA is used directly in 5'- and 3'-RACE PCR reactions, without the need for tedious second-strand synthesis and adaptor ligation. The incorporation of SMART technology also permits the use of “universal priming” in the RACE PCR amplification. This method, along with the techniques of suppression PCR and step-out PCR ensure high specificity in amplifying our target cDNA. The only requirement for SMART RACE cDNA amplification is that we know at least 23–28 nucleotides (nt) of sequence information in order to design gene-specific primers (GSPs) for the 5'- and 3'-RACE reactions.



**Figure 4.2: Overview of the SMART™ RACE procedure.**

With the cloned RACE fragments we can use a restriction site in an overlapping region to construct a full-length cDNA by subcloning. Alternatively, you can sequence the 5' end of the 5' product and the 3' end of the 3' product to obtain the sequences of the extreme ends of the transcript. Using this information, we can design 5' and 3' gene-specific primers to use in LD PCR with the 5'-RACE-Ready cDNA as template to generate the full-length cDNA.

This limited requirement makes SMART RACE ideal for characterizing genes identified through diverse, including cDNA subtraction, differential display, RNA fingerprinting, ESTs, library screening, and more. SMART RACE cDNA amplification is a flexible tool—many researchers use this kit in place of conventional kits to amplify just the 5' or 3' end of a particular cDNA. Others perform both 5'- and 3'-RACE, and many then go on to clone full-length cDNAs using one of the two methods described in the latter part of this protocol. In many cases, researchers obtain full-length cDNAs without ever constructing or screening a cDNA library.

- **Primer Design** (Section V)

I have designed gene-specific primers for the 5' RACE PCR. As described they will facilitate analysis of our RACE products. We can also be used for nested RACE PCR if necessary.

- **First-strand cDNA synthesis** (Section VII)

Since the 5' elongation benefits of SMART technology are only relevant for 5'-RACE, the SMART RACE Kit includes a protocol for the synthesis of two separate cDNA populations: 5'-RACE-Ready cDNA and 3'-RACE-Ready cDNA. The cDNA for 5'-RACE is synthesized using a modified lock-docking oligo (dT) primer and the SMART II A oligo as described above. The modified oligo (dT) primer, termed the 5'-RACE CDS Primer A (5'-CDS), has two degenerate nucleotide positions at the 3' end. These nucleotides position the primer at the start of the poly A+ tail and thus eliminate the 3' heterogeneity inherent with conventional oligo (dT) priming (Borson, N.D., et al. 2001).

- **Positive Control RACE Experiment**

Prior to performing RACE with our template, we have performed the positive control RACE experiment using the Control Human Placental Total RNA provided in the kit.

## • RACE PCR Reactions

After we generate RACE-Ready cDNAs, we have enough material to perform 5' RACE with many different genes, simply by using different gene-specific primers. All PCR reactions in the SMART RACE protocol are optimized for use with the AdvantageR 2 Polymerase Mix.

The Polymerase Mix is comprised of TITANIUM™ Taq DNA Polymerase—a nuclease-deficient N-terminal deletion of Taq DNA polymerase plus TaqStartR Antibody to provide automatic hot-start PCR (Kellogg et al., 1994) and a minor amount of a proofreading polymerase. Advantage 2 technology enables us to perform long distance PCR (LD PCR) reactions with confidence that my products will have high fidelity to the original sequences (Barnes 1994). As a result, we will be able to amplify longer templates than were possible in traditional RACE procedures.

## Primer Design

### A. Primer Sequence:

Gene-Specific Primers (GSPs) should be:

- 23–28 nt
- 50–70% GC
- $T_m \geq 65^\circ\text{C}$ ; best results are obtained if  $T_m > 70^\circ\text{C}$  (enables the use of touchdown PCR)

For the complete SMART RACE protocol, we do have need at least two GSPs: an antisense primer for the 5'-RACE PCR and a sense primer for the 3'-RACE PCR.

But as we have continued with only 5'- RACE, we were needed only one GSP. All primers were 23–28 nt long; there is generally no advantage to using primers longer than 30 nt. GSPs should have a GC content of 50–70% and a  $T_m$  of at least  $65^\circ\text{C}$ ; whenever possible the  $T_m$  should be greater than  $70^\circ\text{C}$ , as determined by nearest neighbor analysis (Frier *et al.*, 1986). we use the GENERUNNER software to calculate  $T_m$ 's. longer primers with annealing temperatures above  $70^\circ\text{C}$  give more robust amplification in RACE, particularly from difficult samples.  $T_m$ 's over  $70^\circ\text{C}$  allow us to use “touchdown PCR”. Additionally, designing GSP1 and GSP2 so that they have similar  $T_m$ 's will facilitate their use in the SMART RACE protocol.  $T_m$ 's of GSP1 and GSP2 can be calculated or determined experimentally by performing PCR at different temperatures. Avoid using self-complementary primer sequences which can fold back and form

intramolecular hydrogen bonds. Similarly, we avoid primers that have complementarity to the primers in the Universal Primer Mix, particularly in their 3' ends.

1. Combine the following in separate microcentrifuge tubes:

For preparation of 5'-RACE-Ready cDNA

1–3  $\mu\text{l}$  RNA sample\*

1  $\mu\text{l}$  5'-CDS primer A

1  $\mu\text{l}$  SMART II A oligo

Added sterile  $\text{H}_2\text{O}$  to a final volume of 5  $\mu\text{l}$  for each reaction. Contents were mixed and spinned down the tubes briefly in a microcentrifuge. Tubes were incubated at 70°C for 2 min. Tubes were cooled on ice for 2 min. Then the tubes were spinned briefly to collect the contents at the bottom.

The following were added to each reaction tube (already containing 5  $\mu\text{l}$ ):

2  $\mu\text{l}$  5X First-Strand Buffer

1  $\mu\text{l}$  DTT (20 mM)

1  $\mu\text{l}$  dNTP Mix (10 mM)

1  $\mu\text{l}$  MMLV Reverse Transcriptase\*

**10  $\mu\text{l}$  Total volume**

Contents of the tubes were mixed by gently pipetting. The tubes were spinned briefly to collect the contents at the bottom. The tubes were incubated at 42°C for 1.5 hr in an air incubator or a hot-lid thermal cycler. Diluted the first-strand reaction product with Tricine-EDTA Buffer:

Tubes were heated at 72°C for 7 min. Samples were stored at –20°C for up to three months.

At this point, we had 5'-RACE-Ready cDNA samples. The RACE reactions in Section IX use only a fraction of this material for each RNA of interest.

We prepared enough Master Mix for all PCR reactions and one extra reaction to ensure sufficient volume. For each 50- $\mu\text{l}$  PCR reaction, mixed the following reagents:

34.5  $\mu\text{l}$  PCR-Grade Water

5  $\mu\text{l}$  10X Advantage 2 PCR Buffer

1  $\mu\text{l}$  dNTP Mix (10 mM; in SMART RACE or Advantage 2 PCR Kit)

1  $\mu\text{l}$  50X Advantage 2 Polymerase Mix

**41.5  $\mu\text{l}$  Total volume**

They were mixed well by vortexing (without introducing bubbles), then briefly spun the tube in a microcentrifuge and the following cycling conditions were given.

- 5 cycles:
  - 94°C 30 sec
  - 72°C 3 min
- 5 cycles:
  - 94°C 30 sec
  - 70°C 30 sec
  - 72°C 3 min
- 25 cycles:
  - 94°C 30 sec
  - 68°C 30 sec
  - 72°C 3 min
- 1 Cycle:
  - 72°C 5 min
  - 16°C ∞

We analyzed 5 µl of each sample on a 1.2 % agarose/EtBr gel. Store the remaining 45 µl of each reaction at –20°C until we were sure that the control experiment has worked.

We designed Gene Specific Reverse Primers for all the isoforms using GENERUNNER software.

Primers were designed for all the 11 Serine Proteinase Isoforms:

**Isoform 1:**

Unique region in cDNA: 5' TGAGGTGTGTGTAGCAGCTTACCCA 3'  
Reverse Primer made: 5' TGGGTAAGCTGCTACACACACCTCA 3' (T<sub>m</sub>= 65.5,  
% GC=75.8, bp= 25)

**Isoform2:**

Unique region in cDNA: 5' TGTGGTTACGTGCGTGAGCCTG 3'  
Reverse Primer made: 5' CAGGCTCACGCACGTAACCACA 3' (T<sub>m</sub>= 65.6, %GC=  
75.0, bp= 22)

**Isoform3:**

Unique region in cDNA: 5' CCTTATCCAGAATTGCCGGCGAC 3'  
Reverse Primer made: 5' GTCGCCGGCAATTCTGGATAAGG 3' (T<sub>m</sub>= 66.3,  
%GC= 56.5, bp= 23)

**Isoform4:**

Unique region in cDNA: 5' CATTGACGGCGACAAGCG 3'  
Reverse Primer made: 5' CGCTTGTCGCCGTCAATG 3' (T<sub>m</sub>= 60.0, %GC= 69.1,  
bp=18)

**Isoform5:**

Unique region in cDNA: 5' ACAGCAGTACACACATCGCGCC 3'  
Reverse Primer made: 5' GCGCGATGTGTGTACTGCTGT 3' (T<sub>m</sub>= 64.8,  
%GC= 75, bp= 22)

**Isoform6:**

Unique region in cDNA: 5' CAATTTGGGTTGCCAGCGACAAGC 3'  
 Reverse Primer made: 5'GCTTGTCTGCTGGCAACCCAAATTG 3' (Tm= 71.3,  
 %GC= 75.6, bp= 24)

**Isoform7:**

Unique region in cDNA: 5' AGGTGTGTCAAGCAGCTTACGGAG 3'  
 Reverse Primer made: 5' CTCCGTAAGCTGCTTGACACACCT 3' (Tm= 64.4,  
 %GC= 75.6, bp= 24)

**Isoform8:**

Unique region in cDNA: 5' GGTGGCCGGTGACAACCA 3'  
 Reverse Primer made: 5'TGGTTGTCACCGGCCACC 3' (Tm= 61, %GC= 71.3,  
 bp= 18)  
2<sup>nd</sup> set of primer:

Reverse Nested Primer (Isoform8):

1) ATGGGTATGCTGCTCGGCACACTGCATA (bp=28, Tm=75.2, %GC=79.4) [for  
 2<sup>nd</sup> set of nested pcr]

(unique sequence in cDNA= TATGCAGTGTGCCGAGCAGCATACCCAT)

2) CACAATATTCTGGTTGTCACCGGCCACC (bp=28, Tm= 74.1, %GC= 79.4) [for  
 1<sup>st</sup> set of nested pcr]

(Unique sequence in cDNA= GGTGGCCGGTGACAACCAGAATATTGTG)

**Isoform 9:**

Unique region in cDNA: 5' ATGGGGCGCAACCACATCTCC 3'  
 Reverse Primer made: 5' GGAGATGTGGTTGCGCCCCAT 3' (bp= 21, Tm= 66.5,  
 %GC= 61.9)

**Isoform10:**

Unique region in cDNA: 5' GTCGAGCAGCTTACTCATCGTTGC 3'  
 Reverse Primer made: 5' GCAACGATGAGTAAGCTGCTCGAC 3' (bp= 24, Tm= 65,  
 %GC= 75.6)

**Isoform11:**

Unique region in cDNA: 5' TTAGTAGCAAAAACAATAAAGAATG 3'  
 Reverse Primer made: 5' CATTCTTTATTGTTTTTGCTACTAA 3' (bp= 25, Tm= 54.3,  
 %GC=64.3)

2<sup>nd</sup> set of primer:

Reverse Primer (Isoform11):

1) CAATACTCTGCTTTTCTCCAGCAACCCT (bp=28, Tm=69, %Gc=76.4) (for 1<sup>st</sup>  
 round of nested pcr)

(Unique region in cDNA= AGGGTTGCTGGAGAAAAGCAGAGTATTG)

2) ATCAGGCGGTTTGTGCGAGAAATTTACCC (bp=28, Tm=72, %GC=76.4) [for 2<sup>nd</sup>  
 round of nested pcr]

(Unique region in cDNA= GGGTAAATTTCTCGACAAACCGCCTGAT)

## cDNA sequences for all Isoforms:

Sistrurus catenatus edwardsi serine proteinase **isoform 1 mRNA**, complete cds

```
ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTTCTTATGCACAAAAGTCTTCTGAACTGGTCGTTGGAGGT  
GATGAATGTAACATAAATGAACATCGTTTCCTTGTAGCCTTGTACTCTAGGTCTAAGACTTTTCTCTGCGGTGGGACTTTGC  
TCAATGAGGAATGGGTGCTCACTGCTGCACACTGTGACAGGAGAAATATGCAGATAAAGCTTGGTATGCATAGCAAACCTGT  
ACCAAATGAGGATGAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTAGCAAAAACCTATACCTTTGGGACAAGG  
ACATCATGTTGATCAGGCTGGACAGACCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCCAGCCCTCCTAGTG  
TGGGCTCAGTTGCCGTATTATGGGATGGGGCAGAATCTCACCTAATAAAGAGACTTATCCCAATGTCCCTCATTGTGCTAACA  
TTAACATACTCAATTATGAGGTGTGTGTAGCAGCTTACCCAGAATATATGTTGCCAGCAACAAGCAGAACATTGTGTGCAGGT  
ATCCTGGAAGGAGGCAAAGATTCTGTAAGGGTGACTCTGGGGGA
```

Sistrurus catenatus edwardsi serine proteinase **isoform 2 mRNA**, complete cds

```
ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTTCTTATGCACAAAAGTCTTCTGAACTGATCATTGGAGGT  
GATGAATGTAACATAAATGAACATCGTTTCCTTGCACCTTGTACTCTGATGGGAATCAATGCGGTGGGACTTTGATCAATGAG  
GAATGGGTGCTCACTGCTGCACACTGCGAAGGGAACAAAATGAAGATACATCTTGGTATGCATAGCAAAAAGGTACCAAATA  
AGGATAAACAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTAGTAGCAAAAACCTACACCTTTGGGACAAGGACATCATG  
TTGATCAGGCTGGACAGACCTGTTGGCAACAGTGAACACATTGCGCCTCTCAGCTTGCCTTCCAGCCCTCCAGTGTGGCTCA  
GTTTCCGTATTATGGGATGGGGCACAATCTCACCTACTAAAGTATTTTGCCTGATGTCCCTCACTGTGTTAACATTAACCTAC  
TCAATTATCGGTGTGTGCGAGCAGCTTACCAGAATATGGGTTGCCAGCGACAAGCAGAACATTGTGTGCAGGTATTCTGGAA  
GGAGGCAAAGATAACATGTGTGGGAGACTCTGGGGGACCCCTCATTTGTAATGGACAATCCAGGGAATTGCATCTGGGGGA  
GCCCTAACTGTGGTTACGTGCGTGAGCCTGGCCTCTACACCAAGGTCTTCGATCATCTTACTGGATCCAGAGCATTATTGCAG  
GAAATACAGATGCGACCTGCCCCCATGA
```

Sistrurus catenatus edwardsi serine proteinase **isoform 3 mRNA**, complete cds

```
ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTGACAGCTTTCTTATGCACAAAAGTCTTCTGAACTGGTCATTGGAGGT  
GATGAATGTAACGTAATGAACATCGTTCCCTTGTAGCCATCTTAACTCTACCGGGTTTTCTGCGCTGGGACTTTGATCAACA  
AGGAATGGGTGCTCACTGCTGCCACTGCGACAGTACAAATTTCCAAATGAAGCTTGGTGTGCATAGCAAAAAGGTATCAAAT  
GAGGATGAGCAGACAAGAAACCCAAAGGAGAAGTCTTTTGTCCGATAAAAAAGATGACGTAAGGACATCATGT  
TGATCAAGCTGGACAGTCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCCAGCCCTCCAGTGTGGCTCAG  
TTTGCCATATTATCGGATGGGGATCAATCACACCTATTGAAAAGACTTTGCCTGATGTCCCTTATTGTGCTAACATTAACCTACT  
CGATGATGCGGTGTGTCAACCCTTATCCAGAATTGCCGGCGACAAGCAGAACATTGTGTGCAGGTATCCTGGAAGGAGGC  
AAAGATAACATGTGGGGTGACTCTGGGGGACCCCTCATCTGTAATGGACAATCCAGGGCATTGATTTTATGGGGCGCATCC  
TTGTGGCCAAGCTCTTAAGCCTGGTGTCTACACCAAAGTCTTTGACTATAATGACTGGATTGAGCATTATTGCAGGAAATAC  
AGCTGCAACGTGCCCCCGTGA
```

Sistrurus catenatus edwardsi serine proteinase **isoform 4 mRNA**, complete cds

```
ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTTCTTATGCACAAAAGTCTTCTGAACTGGTCATTGGAGGT  
GAAGAATGTAACATAAATGAACATCGTTCCCTTGTAGTCTTCTTAACTCTACCAATTTCTCTGCGCTGGGACTTTGATCAACC  
AGGAATGGGTGCTCACTGCTGCCACTGCGACAGTAAAAATTTCCAGATGCAGCTTGGTGTGCATAGCAAAAAGGTACTAAAT  
GAGGATGAGCAGACAAGAGACCCAAAGGAGAAGTCTTTTGTCCCAATAGAAAAAAGATGATGAAAAGGACAAGGACATC  
ATGTTGATCAGGCTGGACAGACCTGTGAGCAATAGTGAACACATCGCGCCTCTCAGCTTACCTTCCAGCCCTACCAGTGTGGC  
TCAGTTTCCGTATTATGGGATGGGGCATAAACTCACCTACTAAAGAGACTTATCCCGATGTCCCTCATTGTGCTAACATTAAC  
ATACTCCATCATGAGGTGTGTAAGCAGCTTATCCATCATTGACGGCGACAAGCGGAACACTGTGTGCAGGTATCCTGGAAGG  
AGGCAAAGATGCATGTGGGGTGACTCTGGGGGACCCCTCATCTGTAATGGAGAAATCCAGGGCATTGATCTGGGGGGGA  
GCATCCTTGTGGCCAAGTCTTACTCTGGCTTCTACACCACGGTCTTCGATTACATTGACTGGATCCAGAGCATTATTGCAGGA  
AATACAGATGCGACCTGCCCCCATGA
```

Sistrurus catenatus edwardsi serine proteinase **isoform 5 mRNA**, complete cds

```
ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTTCTTATCGGACAAAAGTCTTCTGAACTGGTCATTGGAGGT  
GATGAATGTAACATAAATGAACATCGTTCCCTTGCACCTGCTGTATATCACTAACGGTTTTCTCTGCGGTGGGACTTTGATCAACC  
AGGAATGGGTGCTCACTGCTGCACACTGCGACAGGGGAAATATGCTCATATTCTTCGGTGTGCATAGACTAAAGGGACTAAAT
```

AAGGATGCGCAGACAAGAGTCGCAAAGGAGAAGTTCATTTGTCCCAATAGGAAAAAATGATGAAAAGGACAAGGACATC  
ATGTTGATCAGGCTGGACAGTCCTGTTAACAGCAGTACACACATCGCGCCTATCAGCTTGCCTTCCAACCCTCCAGTGTGGGC  
TCAGTTTGCCATATTATGGGATGGGGCGCAATCACATCTCTAATGTGACTTTACCCGGTGTCCCTCATTGTGCTAACATTAACA  
TACTCGATCATGAGGTGTGCAAGCAGCTTATCCATCATTGACCGCGACAAGCAGAACATTGTGTGCAGGTATCCTGGAAGGA  
GGCAAAGTTCATGTGACGGTGACTCTGGGGGACCCCTCATCTGTAATGGAGAAATCCAGGGCATTGTATCTTGGGGGGCG  
ATATTTGTGCCCAACCCCGTGAGCCTGGCGTCTACACCAAGGTCTTCGATTATACTGATTGGATCCAGAACATTATTGCAGGAA  
ATACAGATGCGACCTGCCCCCATGA

Sistrurus catenatus edwardsi serine proteinase **isoform 6 mRNA**, partial cds

TCTTCTGAACTGGTCACTGGAGGTGATGAATGTAACATAAATGAACATCGTTTCCTTGTAGCCTTGTATCACTCTAGGTCTAAGA  
CTTTTCTCTGCGGTGGGACTTTGCTCAACGAGGAATGGGTTCTCACTGCTGCACACTGCAACAGGGTATTTTTGTACATCAGGC  
TTGGTATGCATAACAAAAATGTAAATTTGACGATGACCAGAGAAGATACCCAAAGAAGAAGTACTTTTTCGCTGCACAATA  
ACTTTACAAACGGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGTTACAGTGAACACATTGCGCCTCTCAGCT  
TGCCTTCCAGCCCTCCAGTGTGGGCTCAGTTTGGCGTGAATGGGATGGGGCCAAACCACATCTCTCAAGAGACTTTGCCCG  
ATGTCCTCATTGTGCTAACATTAACCTACTCGATTATGAAGTGTGTCGAGCAGCTACCCACAATTTGGGTTGCCAGCGACAA  
GCAGAATATTGTGTGCAGGTGTCTTGAAGGAGGCATAGATACATGTCATCGTACTCTGGGGACCCCTCATCTGTAATGGG  
CAATTCAGGGCATTGTATCTTGGGGAGATGGTCTTGTGCCAACCGGATAAGCCTGCCCTCTACAGCAAGGTCTTTGATCAT  
CTTACTGGATCCAGAACATTATTGCAGGAAATACAACCTGTGAATTGCCCCCGTGA

Sistrurus catenatus edwardsi serine proteinase **isoform 7 mRNA**, complete cds

ATGGTATCTCAAGTTAAAGCAAACAACCTTGCTGCACAGAATTGAAGCTATGGCGCTGATCAGACTGCTAGTAAACCTTCTGATA  
CTACAGCTTTCTTACGCACAAAAGTCTTCTGAACTGGTCACTGGAGGTGATGAATGTAACGTAATGAACATCGTTTCCTTGC  
CTCTTGTATTCTGACAAGTTTCAATGCGGTGGGACTTTGATCAATGAAGAATGGGTGCTCACTGCTGCACACTGCAA  
CAGAAG  
ACACATGTACATATTCCTTGGTGTGCATAACATAAGCGTAAAAATGATGATGAGCAGAGAAGATCCCAAAGAAGAAGTACT  
TTTGCCTGAGTAGCAGAACTATACCAAAGGGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGGAAAAGTGA  
ACACATCGCGCCTCTCAGCTTGCCTTCCAGTCTCCAGTGTGGCTCAGTTTGGCGTGTATGGGATGGGGACAATCACATC  
TCTAATGTGATGTTGCCCGATGTCCCTTACTGTGCTAACATTAACATACTTGAATTATGAGGTGTGCAAGCAGCTTACGGAGG  
ATCGCCAGCAACAAGCAGAACATTGTGTGCAGGTATCCTGGAAGGAGGCAAAGATTGATGTGACGGTGACTCTGGGGGACCC  
CTCATCTGTAATGGACAATTCCAGGGCATTGTATCTTGGGGAGCCGATACTTGTGCCAACCGCGTGCCTGGCTTCTACACC  
AATGTCTTTGATTATACTGACTGGATCCAAAGCATTATTCAGGAAATACAGATGCGACCTGCCCTCCGTGA

Sistrurus catenatus edwardsi serine proteinase **isoform 8 mRNA**, complete cds

ATGGTGCTGATCAGAGTGCTAGCAACCCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTTCTGAACTGGTTCGTTGGAGGT  
CGTCCATGTAACATAAATGAACATCGTTCCCTTGTGTCTTATTAACCTAGCGGTTTTCTCTGCGCTGGGACTTTGATCAACCA  
GGAATGGGTGCTCACTGCTGCACACTGCGACATTAATAAATTTCCAGATACAGCTTGGTGTGCATAGCAAAAAGGTACGAAATG  
AGGATGAGCAGACAAGAGACCCAAGGGAGAAGTCTTTTGTCTCGGTAGCAAAAACCAATAACGAATGGGAGAAGGACATCAT  
GTTGATCAGGCTGAACAATCCATTGCGCCTCTCAGCTTGCCTTCTAGCCCTCCAGTGTGGCTCACTTTGCCGTATTATGGGAT  
GGGGACAATCTCATCTACTAAAGAAATTTATCCCAATGTCCCTCGTTGTGCTAACATTAACATACTCGATTATGCAAGTGTGCC  
GAGCAGCATACCCATGGTGGCCGGTGACAACCAGAATATTGTGTGCAGGTATTCTGGAAGGAGGCAAAGATTGATGTCAAG  
GTGACTCTGGGGGACCCCTCATCTGTAATGGAGAAATCCAGGGTATTGTATCTTGGGGGGCGCATCCTTGTGGCCGACGTCTT  
AATCTGGCTTCTACACCAAGGTCTTTGATTACATTGACTGGATCCAAAGCATTACTGCAGGAAATACAACCTGTGACTTGCCCC  
CAGTGA

Sistrurus catenatus edwardsi serine proteinase **isoform 9 mRNA**, complete cds

ATGGTTCTAATCAGAGTGCTGGCAAACCTTCTGATACTGCAGCTTTCTTACGCACAAAAGTCTTCTGAACTGAT  
CATTGGAGGTGAAGAATGTATCATAAATGAACATCGTTTCCTTGTAGCCTTGTACTCCTTAAATCTAAGAGGTTTC  
ACTGCAGTGGGACTTTGATCAACCAGGAATGGGTGCTGACTGCTGCACACTGCAACAGGAAAAATATCCGGATAA  
AGCTTGGTACGCATAGCACAAAGGTAACAAATGAGGATGTGCAGACAAGAGTCCCAAAGAAGAAGTCTTTTGTCT  
TCAGTAGCAAAAACCTACACCAGATGGGACAAGGACATCATGTTGATCAGGCTGAAAAGACCTGTTAACAACAGTG  
AACACATCGTGCCTGTGAACTTGCCTTCCAACCCTCCAGTCTGGCTCAGTTTGCATATTATGGGATGGGGCGC

**AACCACATCTCCTAATGTGACTTTACCCGGTGTCCCTCATTGTGCTAACATTAACATACTCGATCATGAGGTGTGTC**  
**AAGCAGCTTATCCATCATTGACCGCGACAAGCAGAACATTGTGTGCAGGTATCCTGGAAGGAGGCAAAAAGTTCAT**  
**GTGACGGTGACTCTGGGGGACCCCTCATCTGTAATGGAGAATTCCAGGGCATTGTATTTTGGGGACCCGATACTT**  
**GTGCCCAACCGGATAAGCCTAGCCTCTACACCAAGGTCTTTGATCATCTTGACTGGATCCAGAGCATTATTGCAGG**  
**AAATAAAACTGTGAATTGCCCCCGTGA**

Sistrurus catenatus edwardsi serine proteinase **isoform 10 mRNA**, complete cds

**ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTCTTATGCACAAAAGTCTTCTGAACTGATCATTGGAGGT**  
**GAAGAATGTAACATAAATGAACATCGTTTCTTGTAGCCTTGTACTCTAGGTCTAAGACTTTTCTCTGCGGTGGGACTTTGC**  
**TCAACGAGGAATGGGTGCTCACTGCTGCACACTGCAACAGG****GTATTTATGTACATCAGGCTTGGTATGCATAACAAAAATGTA**  
**ATATTTGACGATGACCATATAAGATACGCAAAGGAGAAGTACTTTTTTCGCTGTCGCAATAACTTTACCAAATGGGACAAGGA**  
**CATCATGTTGATCAGGCTGAACAAACCTGTTAGTTACAGTGAATACATCGCGCCTCTCAGCTTGCCTTCCAGCCCTCCCAGTGTG**  
**GGCTCAGTTGCCGTGTTATGGGATGGGGCCAAACCACATCTCCTCAAG****AGACTTTGCCCGATGTCCCTCATTGTGCTAACATT**  
**AACCTACTTGATTAGAAGTGTGTCGAGCAGCTTACTCATCGTTGCCGGCGACAAGCAGAGTATTGTGTGCAGGTGTCCTGGA**  
**AGGAGGCATAGATACATGTAATCGTGACTCTGGGGGACCCCTCATCTGTGATGGGCAATTCCAGGGCATTGTATCTTGGGGAG**  
**ATGGTCTTGTGCCCAACCGGATAAGCCTGCCCTCTACAGCAAGGTCTTTGATCATCTTGACTGGATCCAGAACATTATTGCAG**  
**GAAATACAACCTGTGAATTGCCCCCGTGA**

Sistrurus catenatus edwardsi serine proteinase **isoform 11 mRNA**, complete cds

**ATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTCTTATCGCACAAAAGTCTTCTGAACTGGTTCGTTGGAGGT**  
**GATGAATGTAACATAAATGAACATCGTTCCCTTGTACTCGTGTACTCTGACGGGATTCAATGTGGTGGGACTTTGATCAACCAG**  
**GAATGGATGCTCACGGCTGCACACTGTGACGGG****AAAAGAATGAAGTTACAGTTTGGTCTGCATAGCAAAAATGTACCAAACA**  
**AGGATAAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTAGCAAAAACAATAAAGAATGGGACAAGGACATCA**  
**TGTTGATCAGGCTGAACCGTCTGTTAACAACAGTAAACACATTGCGCCTCTCAGCTTGCCTTCCAAGCCTCCCAGTCAGGACA**  
**CAGTTTGCAATATTATGGGATGGGGC****ACAATCTCACCTACTAAGAGATTATCCTGATGTCCCTCATTGTGCTAACATTAATAT**  
**ACTCGATCAGGCGTTTTGTCGAGAAATTTACCCAGGGTTGCTGGAGAAAAGCAGAGTATTGTGTGCAGGTATCCTGGAAGGA**  
**GGCAAAGATACATGTGGAGGTGACTCTGGGGGACCTCTCATCTGTAATGGAGAAATCCAGGGCATTATCTGTGGGGGGCG**  
**ATCCTTGTGCCCAATCTCATGTGCCTGCCCTCTACATCAAGGTCTTTGATTATACTGAGTGGATCCAGAGCATTATTACAGGAAA**  
**TACAGCTGCAACTTGCCCCCATGA**

Exon 1 – Yellow Exon 2 – Green Exon 3 – Red Exon 4 – Blue Exon 5 – Dark yellow

## Amino acid sequences of all the isoforms:

>SCEP1

MVLIRVLANLLILQLSYAQKSELVVGDECNINEHRFLVALYHSRKTFLCGGTLNNEEWVLTAAHCDRRNMQIKLGMHSKTVPNEDEQTRVPKEKFFCLSSKNYTLW  
DKDIMLIRLDRPVSNSEHIAPLSLSPSPSVGVCVCRIMGWGRISPKNKETYPNVPHCANINILNVEVCVAAYPEYMLPATSRTLCAILEGGKDSCKGDSGGPLICNGQFQ  
GILSWGDDPCGYVLQPALYTGVDHLDWIQSIAGNTDATCPP

>SCEP2

MVLIRVLANLLILQLSYAQKSELIIGDECNINEHRFLALVSDGNQCGGTLNNEEWVLTAAHCEGNKMKIHLGMHSHKVPNKDKQTRVPKEKFFCVSSKNYTFWDDK  
IMLIRLDRPVGNSEHIAPLSLSPSPSVGVCVCRIMGWGTISPTKVLDPVPHCVNINLLNYSVCRAAYPEYGLPATSRTLCAILEGGKDKTCVGDSSGGPLICNGQFQGIAS  
WGSNPGYVREPGLYTKVDFDHLWDWIQSIAGNTDATCPP

>SCEP3

MVLIRVLANLLILQLSYAQKSELVIGGDECNVNEHRSLVAIFNSTGFFCAGTLINKEWVLTAAHCDSTNFQMKLVHSHKVSNEDEQTRNPKEKFFCPDKDDVLDKDI  
MLIKLSDPVSNSEHIAPLSLSPSPSVGVCVCHIGWGSITPIEKLDPVPCANINLLDVAVCQPPYELPATSRTLCAILEGGKDKTCGGDSGGPLICNGQFQGIYFYGAHP  
CGQALKPGVYTKVFDYNDWIQSIAGNTAATCPP

>SCEP4

MVLIRVLANLLILQLSYAQKSELVVGEECNINEHRSLVFFNSTNFLCAGTLINQEWVLTAAHCDKSNFQMLGVHSHKVLNEDEQTRDPKEKFFICPNRKKDDEKDK  
DIMLIRLDRPVSNSEHIAPLSLSPSPSVGVCVCRIMGWGINSPTKETYDPVPHCANINILHHEVCQAAYPSLTATSRTLCAILEGGKDACGGDSGGPLICNGEIQGIIVSG  
GEHPCGQVLTGPFYTVFDYIDWIQSIAGNTDATCPP

>SCEP5

MVLIRVLANLLILQLSYGQKSELVIGGDECNINEHRSLVYITNGFLCGGTLINQEWVLTAAHCDRGNMLIFFGVHRLKGLNKDAQTRVAKEKFFICPNRKKNDEKDKDI  
MLIRLSDPVSNSSTHIAPLSLSPSPSVGVCVCHIMGWGAIITSPNVTLPVPHCANINILDHEVCQAAYPSLTATSRTLCAILEGGKSSCDGDSGGPLICNGEIQGIIVSWG  
DICAQPREPGVYTKVFDYIDWIQSIAGNTDATCPP

>SCEP6

SSELVIGGDECNINEHRFLVALYHSRKTFLCGGTLNNEEWVLTAAHCNRVFLYIRLGMHNKNVNFDDQRRYPKPKYFFRCHNNFTKRDKDIMLIRLNKPVSYSEHIAP  
LSLSPSPSVGVCVCRAMGWGQTTSPQETLPVPHCANINLLDYEVCAAHPPQFLPATSRLCAGVLEGGIDTCHRDSGGPLICNGQFQGIIVSWGDSQCAQPKPALY  
SKVFDHLDWIQSIAGNTTVCNCP

>SCEP7

MVSQVKANLLHRIEAMALIRLLVNLILQLSYAQKSELVIGGDECNVNEHRFLALLYSDFKFCGGTLNNEEWVLTAAHCNRHMYIFLGVHNISVKYDDEQRRFPKKK  
YFCLSSRNYTKGDKDIMLIRLNKPVKSEHIAPLSLSPSPSVGVCVCRVMGWGTITSPNVMLPDPVPCANINILDYEVCAAYGGSPATSRTLCAILEGGKDSCKGDSG  
GPLICNGQFQGIIVSWGADTCAQPRAPGFYTNVFDYIDWIQSIAGNTDATCPP

>SCEP8

MVLIRVLATLLILQLSYAQKSELVVGGRPCNINEHRSLVLFNSSFGLCAGTLINQEWVLTAAHCDIKNFQIQLGVHSHKVRNEDEQTRDPREKFFCLGSKTNNEWEKDI  
MLIRLNPNVNSAHIAPLSLSPSPSVGSLCRIMGWGTISSTKEIYPNVPRCANINILDYAVCAAYPWVPTTRILCAGILEGGKDSCKGDSGGPLICNGEIQGIIVSWG  
AHPGRRRLNPGFYTKVFDYIDWISITAGNTTVCPCQ

>SCEP9

MVLIRVLANLLILQLSYAQKSELIIGGEECIINEHRFLVALYFSKSRFHCSGTLINQEWVLTAAHCNRNIRIKLGTHTKVTNEDVQTRVPKFFCLSKTYTRWDKDI  
MLIRLKRPNVNSEHIVPNLSPNPPSLGVCVCHIMGWGATTSNVTLPVPHCANINILDHEVCQAAYPSLTATSRTLCAILEGGKSSCDGDSGGPLICNGEFQGIIVFW  
GPDTCAPDKPSLYTKVFDHLDWIQSIAGNTTVCNCP

>SCEP10

MVLIRVLANLLILQLSYAQKSELIIGGEECNINEHRFLVALYHSRKTFLCGGTLNNEEWVLTAAHCNRVFMIRLGMHNKNVIFDDDHIRYAKEKYFFRCRNFTKWWDK  
DIMLIRLNKPVSYSEIAPLSLSPSPSVGVCVCRVMGWGQTTSPQETLPVPHCANINLLDYEVCAAYSSLPATSRLCAGVLEGGIDTCNRDSGGPLICNGQFQGIIVS  
WGDGSCAQPKPALYSKVDHLDWIQSIAGNTTVCNCP

>SCEP11

MVLIRVLANLLILQLSYAQKSELVVGDECNINEHRSLVLYSDGICQCGGTLINQEWMLTAAHCDGKRMKLFGLHSHKVPNKDKQTRVPKEKFFCLSSKNKEWWDK  
DIMLIRLNRPVNSKHIAPLSLSPSPSQDTCVNCIMGWGTISPTKEIYPVPHCANINILDQAVCREIYPGLLEKSRVLCAGILEGGKDKTCGGDSGGPLICNGEIQGILSVG  
GDPCAQSHVPALYIKVFDYTEWISITGNTAATCPP

**Multiple sequence alignment to show the sequence similarity in amino acid sequences:**  
 CLUSTAL format alignment by MAFFT (v6.857b)

```

SCESP1 -----MVLIRVLANLLILQLSYAQKSELVVGDECNINEHRFLVALYH
SCESP2 -----MVLIRVLANLLILQLSYAQKSELIIGGDECNINEHRFLVALVY-
SCESP11 -----MVLIRVLANLLILQLSYAQKSELVVGDECNINEHRSLVLVY-
SCESP3 -----MVLIRVLANLLILQLSYAQKSELVIGGDECNVNEHRSLVAIF-
SCESP4 -----MVLIRVLANLLILQLSYAQKSELVVGGEENINEHRSLVVFV-
SCESP8 -----MVLIRVLATLLILQLSYAQKSELVVGGRPCNINEHRSLVVF-
SCESP5 -----MVLIRVLANLLILQLSYGQKSELVIGGDECNINEHRSLALVY-
SCESP9 -----MVLIRVLANLLILQLSYAQKSELIIGGEECIINEHRFLVALYS
SCESP7 MVSQVKANLLHRIEAMALIRLLVNLILQLSYAQKSELVIGGDECNVNEHRFLALLY-
SCESP6 -----SSELVIGGDECNINEHRFLVALYH
SCESP10 -----MVLIRVLANLLILQLSYAQKSELIIGGEECNINEHRFLVALYH
                                     ****::** * :**** * . :
  
```

```

SCESP1 SRSKTFLCGGTLLEEWVLTAAHCDRRNMQIKLGMHSTKVPNEDEQTRVPKEKFFCLSSK
SCESP2 --SDGNQCGGTLINQEWVLTAAHCEGNKMKIHLGMHKKVVPNKDKQTRVPKEKFFCVSSK
SCESP11 --SDGIQCGGTLINQEWMLTAAHCDGKRMKLFQGLHKNVVPNKDKQTRVPKEKFFCLSSK
SCESP3 -NSTGFFCAGTLINKEWVLTAAHCDSTNFQMKLGVHSHKVSNEDEQTRNPKEKFFCPDKK
SCESP4 -NSTNFLCAGTLINQEWVLTAAHCDKSNFQMLGVHSHKVLNEDEQTRDPKEKFIQPNRK
SCESP8 -NSSGFLCAGTLINQEWVLTAAHCDIKNFQIQLVHSHKVRNEDEQTRDPREKFFCLGSK
SCESP5 -ITNGFLCGGTLINQEWVLTAAHCDRGNMLIFFGVHRLKGLNKDAQTRVAKEKFIQPNRK
SCESP9 FKSFRFHCSGTLINQEWVLTAAHCNRKNIRIKLGTHTSKVTNEDVQTRVPKFFCLSSK
SCESP7 --SDKFCQCGGTLINQEWVLTAAHCNRHRMYIFLGVHNI SVKYDDEQRRFPKKKYFCLSSR
SCESP6 SRSKTFLCGGTLLEEWVLTAAHCNRVFLYIRLGMHNKNVFNDDQRRYPKKKYFFRCHN
SCESP10 SRSKTFLCGGTLLEEWVLTAAHCNRVFMYIRLGMHNKNVIFDDDHIRYAKEKYFFRCHN
      : * .***:*.***:*****: : : :* * . . * : * .:***:
  
```

```

SCESP1 NYTLWDKDIMLIRLDRPVSNSEHIAPLSLSPSPSPVSGSVCRIMGWGRISPNETYPNVPH
SCESP2 NYTFWDKDIMLIRLDRPVGNSEHIAPLSLSPSPSPVSGSVCRIMGWGTISPTKVILPDVPH
SCESP11 NNKEWKDIMLIRLNRPVNSKHIAPLSLSPKPPSQDTCVNIIMGWGTISPTEIYDPVPH
SCESP3 D-DVLDDKDIMLIKLDSPVNSEHIAPLSLSPSPSPVSGSVCHIIGWGSITPIEKTLPDVPY
SCESP4 KDEKDKDIMLIRLDRPVSNSEHIAPLSLSPSPSPVSGSVCRIMGWINSPTKETYDPVPH
SCESP8 TNEWKDIMLIRLNNPVSNSAHIAPLSLSPSPSPVSGSLCRIMGWGTISPTEIYPNVPR
SCESP5 KNDEKDKDIMLIRLDPVNSSTHIAPLSLSPNPPVSGSVCHIMGWGAITSPNVTLPGVPH
SCESP9 TYTRWDKDIMLIRLKRPNVNSEHIVPNLSPNPPSLGVSCHIMGWGAITSPNVTLPGVPH
SCESP7 NYTKGDKDIMLIRLNKPVKSEHIAPLSLSPSPSPVSGSVCRVMGWGITSPNVMLPDVPH
SCESP6 NFTKRDKDIMLIRLNKPVSYSEHIAPLSLSPSPSPVSGSVCRAMGWGQTTSPQETLPDVPH
SCESP10 NFTKWDKDIMLIRLNKPVSYSEYIAPLSLSPSPSPVSGSVCRVMGWGQTTSPQETLPDVPH
      :*****:*. * * :*.*.***.*. * .:*. :*** :. : * .**
  
```

```

SCESP1 CANINILNYESVCAAYPEYMLPATSRITLCAGILEGGKDSCKGDSGGPLICNGQFQGIISW
SCESP2 CVNINLLNYSVCRAAYPEYGLPATSRITLCAGILEGGKDTCVGDSGGPLICNGQFQGIASW
SCESP11 CANINILDQAVCREIYP--GLEKSRVLCAGILEGGKDTCCGDSGGPLICNGEIQGILSV
SCESP3 CANINLLDDAVCQPPYP--ELPATSRITLCAGILEGGKDTCCGDSGGPLICNGQFQGIIVF
SCESP4 CANINILHHEVCQAAYP--SLTATSGTLCAGILEGGKDACGDSGGPLICNGEIQGIVSG
SCESP8 CANINILDYAVCRAAYP--WWPVTRILCAGILEGGKDSQCGDSGGPLICNGEIQGIVSW
SCESP5 CANINILDHEVCQAAYP--SLTATSRITLCAGILEGGKSSCDGDSGGPLICNGEIQGIVSW
SCESP9 CANINILDHEVCQAAYP--SLTATSRITLCAGILEGGKSSCDGDSGGPLICNGEIQGIVFW
SCESP7 CANINILDYEVCAAYG--GSPATSRITLCAGILEGGKDSQCGDSGGPLICNGQFQGIIVSW
SCESP6 CANINLLDYEVCAAAHPQFGLPATSRITLCAGVLEGGIDTCHRDSSGGPLICNGQFQGIIVSW
SCESP10 CANINLLDYEVCAAAYS--SLPATSRVLCAGVLEGGIDTCNRDSSGGPLICDQFQGIIVSW
      * .***:*. * * :. : *****:*** .:* *****:*.***
  
```

```

SCESP1 GDDPCGYVLQPALYTGVDHLDWIQSI IAGNTDATCPP
SCESP2 GSPNCGYVREPLYTKVFDHLDWIQSI IAGNTDATCPP
SCESP11 GGDPCAQSHVPALYIKVFDYTEWISIIITGNTAATCPP
SCESP3 GAHPCGQALKPGVYTKVFDYNDWIQSI IAGNTAATCPP
SCESP4 GEHPCGQVLTGPFYTTVFDYIDWIQSI IAGNTDATCPP
SCESP8 GAHPCGRRLNPGFYTKVFDYIDWIQSI IAGNTTVTCPQ
SCESP5 GGDICAQPREPGVYTKVFDYTDWIQNI IAGNTDATCPP
SCESP9 GPDTCAPDKPSLYTKVFDHLDWIQSI IAGNTVNCPP
SCESP7 GADTCAQPRAPGFYTNVFDYTDWIQSI IAGNTDATCPP
SCESP6 GDGSCAQPDKALYSKVFDHLDWIQNI IAGNTVNCPP
SCESP10 GDGSCAQPDKALYSKVFDHLDWIQNI IAGNTVNCPP
      * * . * . * * * : * * * . * * * . * *
  
```

Multiple sequence alignment for all the isoforms shows the high similarity. Region near to 5' translation start site have highly conserved region in most of the isoforms. So most of the primers were designed in exon 4 region which were unique for the particular isoform.

**Multigradient Temperature PCR:**  
**(For Isoform 8 and Isoform 11)**

The **Gradient Thermal Cycler** is a PCR device used in molecular biological research and diagnostics. Based on the general function of standard PCR product, the Gradient PCR is further enhanced by thermal gradient technology. It's thermal gradient module allows simultaneous incubation at up to 12 different annealing temperatures, to optimizing reactions in a single run and improve efficiency of your research.

To get the amplification of expected base pair size in Isoform 8 and Isoform 11, both were put at 5 different gradient temperatures.

**Gel electrophoresis to check the amplification on the gel:**

To pour a gel, agarose powder is mixed with electrophoresis buffer to the desired concentration, then heated in a microwave oven until completely melted. Gel Red was added to the gel (final concentration 0.5 µg/ml) at this point to facilitate visualization of DNA after electrophoresis. After cooling the solution to about 60<sup>0</sup>C, it was poured into a casting tray containing a sample comb and was allowed to solidify at room temperature. Samples containing DNA were mixed with loading buffer and then pipette into the sample wells, the lid and power leads were placed on the apparatus, and a current was applied.

**Gel extraction:**

DNA was excised from the agarose gel with a clean sharp gel cutter. Gel slice was weighed and 3 volumes of Buffer QG was added to 1 volume of gel. Eppendorf tube was incubated at 50<sup>0</sup>C for 10 min and to facilitate in dissolving the gel tube was vortexed after every 2-3 min during incubation. After get dissolves completely the colour of the mixture becomes yellow. 1 volume of isopropanol was added to the sample and was mixed. The above sample was applied in a 2 ml QIAquick spin column. The spin column was centrifuged for 1 min at 14000 rpm. The flow through was discarded and was placed back in the same collection tube. 700 ul of Buffer PE was

added to wash the column. The above solution was spun again for 1 min. The flow through was discarded and was centrifuged again at 13K rpm for 1 min to get rid of residual ethanol in wash buffer. QIAquick spin column was placed in a clean 1.5 ml eppendorf tube. After that DNA was eluted by adding 30ul of elution buffer (EB Buffer) in the centre of the QIAquick spin column. It was centrifuged again at 13K rpm for 30 sec.

#### **Quantification Of DNA using nanodrop:**

DNA concentration was quantified on nanodrop .

Absorbance A<sub>260/280</sub>, Absorbance 260/230 was calculated.

#### **Ligation In pJET vector:**

PCR Cloning Kit is an advanced positive selection system for the highest efficiency cloning of PCR products generated with Pfu DNA polymerase, Taq DNA Polymerase, DNA polymerase or other thermostable DNA polymerases.

Additionally, any other DNA fragment, either blunt or sticky-end, can be successfully cloned. The kit features the novel positive selection cloning vector pJET1.2/blunt. This vector contains a lethal gene which is disrupted by ligation of a DNA insert into the cloning site. As a result, the vector contains an expanded multiple cloning site, as well as a T7 promoter for in vitro transcription. Sequencing primers are included for convenient sequencing of the cloned insert.

#### **Sticky end cloning method:**

For cloning PCR products with 3' overhangs generated by *Taq* DNA polymerase, DreamTaq DNA polymerase or enzymes mixtures containing *Taq* DNA polymerase.

For cloning PCR products when DNA end structure of the generated PCR products is not specified by the supplier of the polymerase. For cloning DNA fragments with 5'- or 3'-overhangs generated by restriction enzyme digestion. Gel purifies the DNA fragment prior to ligation and use in a 3:1 molar ratio with pJET 1.2/blunt.

[The DNA Blunting Enzyme is a proprietary thermostable DNA polymerase with proofreading activity. It will remove 3'-overhangs and fill-in 5'overhangs. Nucleotides for the blunting reaction are included in the reaction buffer.

1. Set up the blunting reaction on ice:

Component	Volume
2X Reaction Buffer	10 $\mu$ l
Non-purified PCR product or purified PCR product/other sticky-end DNA fragment	1 $\mu$ l 0.15 pmol ends
Water, nuclease-free	to 17 $\mu$ l
DNA Blunting Enzyme	1 $\mu$ l
Total volume	18 $\mu$ l

The above solution was vortexed briefly and centrifuged for 3-5 sec. The mixture was incubated at 70<sup>0</sup>C for 5 min. and then chilled on ice for 5 min. After that the ligation reaction was set on ice and the following were added to the blunting reaction mixture.

Component	Volume
pJET1.2/blunt Cloning Vector (50 ng/ $\mu$ l)	1 $\mu$ l (0.05 pmol ends)
T4 DNA Ligase	1 $\mu$ l
Total volume	20 $\mu$ l

The above solution was vortexed briefly and centrifuged for 3-5 s to collect the drops. The ligation mixture was incubated at room temperature i.e. at 22<sup>0</sup>C for 5 min. And this ligation mix was used directly for the transformation.

**Map and features of pJET1.2/blunt cloning vector**

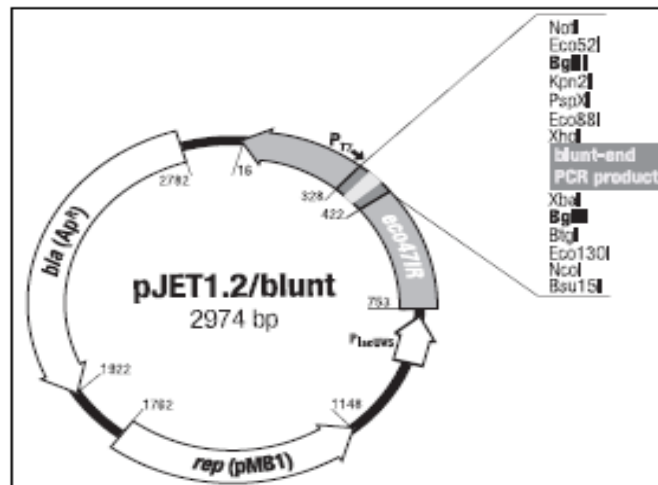


Figure 4.3: pJET1.2/blunt vector map

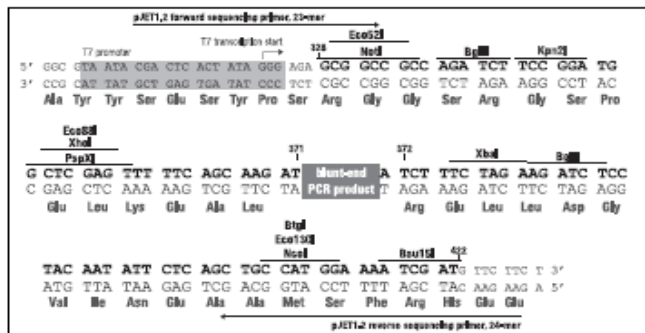


Figure 4.4: DNA sequence of MCS region

**CLONING PRINCIPLE:**

pJET1.2/blunt is a linearized cloning vector, which accepts inserts from 6 bp to 10 kb. The 5'-ends of the vector cloning site contain phosphoryl groups, therefore, phosphorylation of the PCR primers is not required. Blunt-end PCR products generated by proofreading DNA polymerases can be directly ligated in just 5 min with the pJET1.2/blunt cloning vector. PCR products with 3'-dA overhangs generated using Taq DNA polymerase or other non-proofreading thermostable DNA polymerases are blunted in 5 min with a proprietary thermostable DNA blunting enzyme (included in the kit) prior to ligation. All common laboratory E.coli strains can be directly transformed with the ligation product. We have used E. coli JM109 strain for transformation purposes. Recircularized pJET1.2/blunt vector expresses a lethal restriction enzyme after transformation and is not propagated. As a result, only recombinant clones containing the insert appear on culture plates. Therefore, blue/white screening is not required.

**Preparation Of Competent cells:**

1. **Day1:** E. coli JM109 bacterial cells were streaked on LB agar media containing plate and was put in 37<sup>0</sup>C shaker at 200 rpm for overnight.
2. **Day2:** single colony was picked up and was put in 10 ml LB Broth and was kept on 37<sup>0</sup>C shaker at 200 rpm for overnight.
3. **Day3:** 1 ml of culture was taken from 10 ml mass culture and was put in 100 ml LB Broth. It was incubated at 37<sup>0</sup>C for 200 rpm. After 1 hr. OD was checked. When it reached to 0.4-0.6, then cells were spin down and centrifuged at 3500rpm for 10 min at 40C. Supernatant is removed and 50 ml 100 mM ice cold CaCl<sub>2</sub> was added to resuspend the pallet. The above

mixture was kept on ice for 30 min-40 min. The above solution was spin down at 3500 rpm for 10 min at 40°C. Supernatant was discarded completely. Pallet was resuspended again in 6 ml 100 mM CaCl<sub>2</sub> and 2ml ice cold 50 % Glycerol. The above solution was aliquoted into 30 µl into premarked prechilled eppendorf tube. It was quick chilled using liquid N<sub>2</sub>. It was stored back in -80°C.

### **Transformation:**

JM 109 E.coli strain was taken and thawed on ice for 5 min. All the 20 ul ligation mix was added and it was incubated at 4°C. Above solution was kept on ice for 20-45 min. After that heat shock was given at 42°C for 90 sec. After that it was again incubated in ice for 5 min. 200 µl of LB broth was added to above mix and was kept on shaker at 37°C at 200 rpm for 45 min. After that in 50 ml eppendorf tube 50 µl ampicilline and 50 ml LB agar was added and mixed properly. It was allowed to solidify for 10-15 min. After that Ligation mix was spreaded with the help of glass beads. The plates were incubated in 37°C incubator for 16 hrs.

### **Inoculation:**

Colonies were picked up and were inoculated in 2 ml LB Broth containing 2 µl Ampicilline and were kept on shaker for 37°C at 200 rpm for 16 hrs.

### **MiniPrep Plasmid isolation:**

2ml LB culture was taken and spin at 6000 rpm for 3 min. Supernatant was discarded and pellet was resuspended in 170 µl suspension buffer [S1 buffer]. The above solution was vortexed for complete suspension. After that 170 µl of S2 Buffer was added and was shaken smoothly. 250 µl of G3 neutralizing buffer was added with the 5 min of addition of former. The above solution was shaken 5-6 times and was transferred to mini prep spin column. The above solution was spin down at high speed of 14K for 1 min. and the first column was discarded. Wash Buffer PW was added 700 µl to the second column. It was kept for 5 min for better washing. The above solution was spin down at 14K rpm for 1 min. 30 µl of EB buffer was added and incubated at room temperature for 5 min. The above solution is spin down at high speed of 14 K for 1 min. The concentration of DNA sample was checked using nanodrop.

**Restriction digestion:**

The above samples were restricted digested with **BgIII**.

Preparation of Restriction mix:

DNA- 2 $\mu$ l

Restriction enzyme- 0.5  $\mu$ l

10 X Buffer: 1  $\mu$ l

Water: 6.5  $\mu$ l

The above solution was incubated at 37<sup>0</sup>C for 30 min.

After that samples were loaded on 1% Agarose gel to confirm if plasmid has released the insert.

**Big dye terminator sequencing:**

Sequencing was performed to determine the order of the bases and also to confirm if the insert was in proper frame. Sanger's chain terminaton method was adopted where the dideoxynucleotide triphosphates (ddNTPs) are used as DNA chain terminators. This reaction also requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleotidephosphates (dNTPs), and modified nucleotides (dideoxyNTPs) that terminate DNA strand elongation. These ddNTPs were radioactively or fluorescently labelled for detection in automated sequencing machines. To perform the reaction, Bigdye was used. Perkin-Elmer' BIGDYE is a ready reaction mix which contains all the components necessary to the sequencing reaction. Only template and primer needs to be added. The principle of the reaction is that successive rounds of denaturation, annealing and extension of the mix result in incorporation of one of the four dideoxy terminators (ddNTPs), each tagged with a different fluorescent dye, in each extension product. The first step to sequencing is sequencing PCR.

**SEQUENCING PCR:**

PCR sequencing involves determination of the nucleotide sequence within a DNA fragment amplified by the PCR, using primers specific for a particular genomic site. The method that has been most commonly used to determine nucleotide sequences is based on the termination of in vitro DNA replication. The primers for the promoter and terminator sequences of the vector were designed. In a reaction only one of the primers is to be used since the sequencer can take-up only

single fragments. With dye terminator labeling, each of the four dideoxy terminators (ddNTPs) is tagged with a different fluorescent dye. The growing chain is simultaneously terminated and labeled with the dye that corresponds to that base.

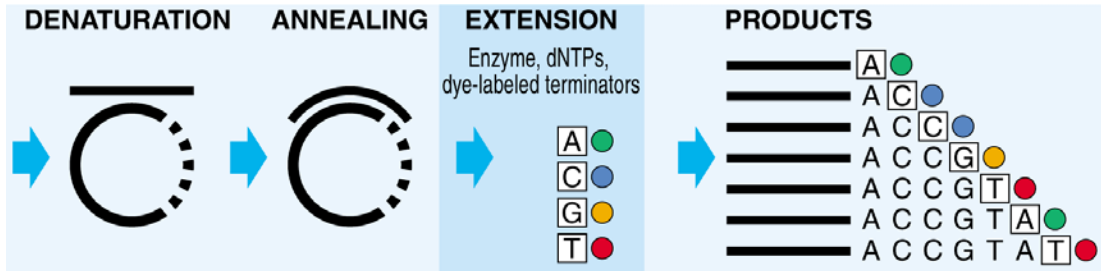


Figure 4.5: Sequencing PCR method

### **Sequence PCR for the positive clones:**

Seq PCR sample Preparation:

Master mix for Forward Primer:

DNA- 2  $\mu$ l

Big Dye: 1  $\mu$ l

Forward Primer: 0.5  $\mu$ l

Water: 1.5  $\mu$ l

Master mix for Reverse Primer:

DNA- 2  $\mu$ l

BigDye: 1  $\mu$ l

Reverse Primer: 0.5  $\mu$ l

Water: 1.5  $\mu$ l

The reaction was performed in a 96 well plate with 16 replicates with primer to forward and 16 replicates with primer to reverse. This approach helps in getting the full length sequence of 5' RACE DNA.

### **Cycling Conditions:**

1 cycle:

96<sup>0</sup>C- 1 min

25 cycles:

96<sup>0</sup>C- 10 sec

50<sup>0</sup>C- 5 sec

60<sup>0</sup>C- 2 min

1 cycle:

16<sup>0</sup>C- hold

### **Cleaning of Seq PCR samples:**

Plate was kept on ice for 2-3 min and short spined it.

10 µl of clean seq sample was added to each well. 35 µl of 85% ethanol was added to precipitate DNA. The plate was placed on the magnetic plate for 5 min. The solution was aspirated. 100 µl of 85 % ethanol was added and was kept for 1 min. Then the solution was aspirated again. The step 6 was repeated for better washing of ethanol. The solution was aspirated back and it was kept for air dry for 15 min. 40 µl of autoclaved water was added again.

### **Data analysis of sequencing results:**

Sequences obtained both for forward and reverse from sequencer machine were copied and regions with Ns were neglected. And with the help of vec screen (<http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>) vector specific regions were deleted. Reverse primer containing clone was put for reverse complement. Both forward and reverse sequences were aligned together to make the complete sequence. After making the complete sequence for the given isoform, it was aligned with the corresponding template cDNA. And the procedure was repeated for all the clones of given isoform. Complete sequence containing 5'UTR is properly aligned with template cDNA from its translation start site, i.e. ATG. Sequence ranging from 60-190 base pairs were obtained. The clone which is most prominent in the all sequenced clone of given isoform is taken as representative clone for 5'UTR.

Further genome walking was performed to identify the promoter region which may be 250-1000bp upstream sequence to 5'UTR

### **Genome Walking:**

GenomeWalker DNA walking is a simple method for finding unknown genomic DNA sequences adjacent to a known sequence such as a cDNA (Siebert et al., 1995). Using our genomic DNA of

interest, the first step is to construct pools of uncloned, adaptor-ligated genomic DNA fragments, which are referred to for convenience as GenomeWalker “libraries.” The starting DNA must be very clean and have a high average molecular weight, requiring a higher quality preparation than the minimum suitable for Southern blotting or conventional PCR. To ensure that our genomic DNA is of adequate quality, the kit includes controls for comparison. Separate aliquots of DNA are completely digested with different restriction enzymes that leave blunt ends. The GenomeWalker Universal Kit comes with a set of four restriction enzymes; however, alternative restriction enzymes that leave blunt ends may be substituted. Each batch of digested genomic DNA is then ligated separately to the GenomeWalker Adaptor.

After the libraries have been constructed, the protocol takes just two days and consists of two PCR amplifications per library (Figure 4.6). The first or primary PCR uses the outer adaptor primer (AP1) provided in the kit and an outer, gene-specific primer (GSP1) provided by the researcher. The primary PCR mixture is then diluted and used as a template for a secondary or “nested” PCR with the nested adaptor primer (AP2) and a nested gene-specific primer (GSP2). This generally produces a single, major PCR product from at least three of the four libraries (and often in all four; Figure 4.6). Each of the DNA fragments—which begin in a known sequence at the 5' end of GSP2 and extend into the unknown adjacent genomic DNA—can then be cloned and further analyzed.

#### **Long-distance PCR with the Advantage® 2 PCR Kit:**

GenomeWalker reactions should be performed with a 50X polymerase mix containing a combination of DNA polymerases suitable for long-distance PCR (Barnes, W. M. et al. 1994). In LD PCR, a combination of two thermostable DNA polymerases is used to increase the range and accuracy of PCR amplification. Most of the extension is carried out by a primary polymerase, while a secondary polymerase provides the critical 3' 5' exonuclease or "editing" function that corrects misincorporated nucleotides. In the GenomeWalker protocol, the use of LD PCR extends the range of possible PCR products to about 6 kb.

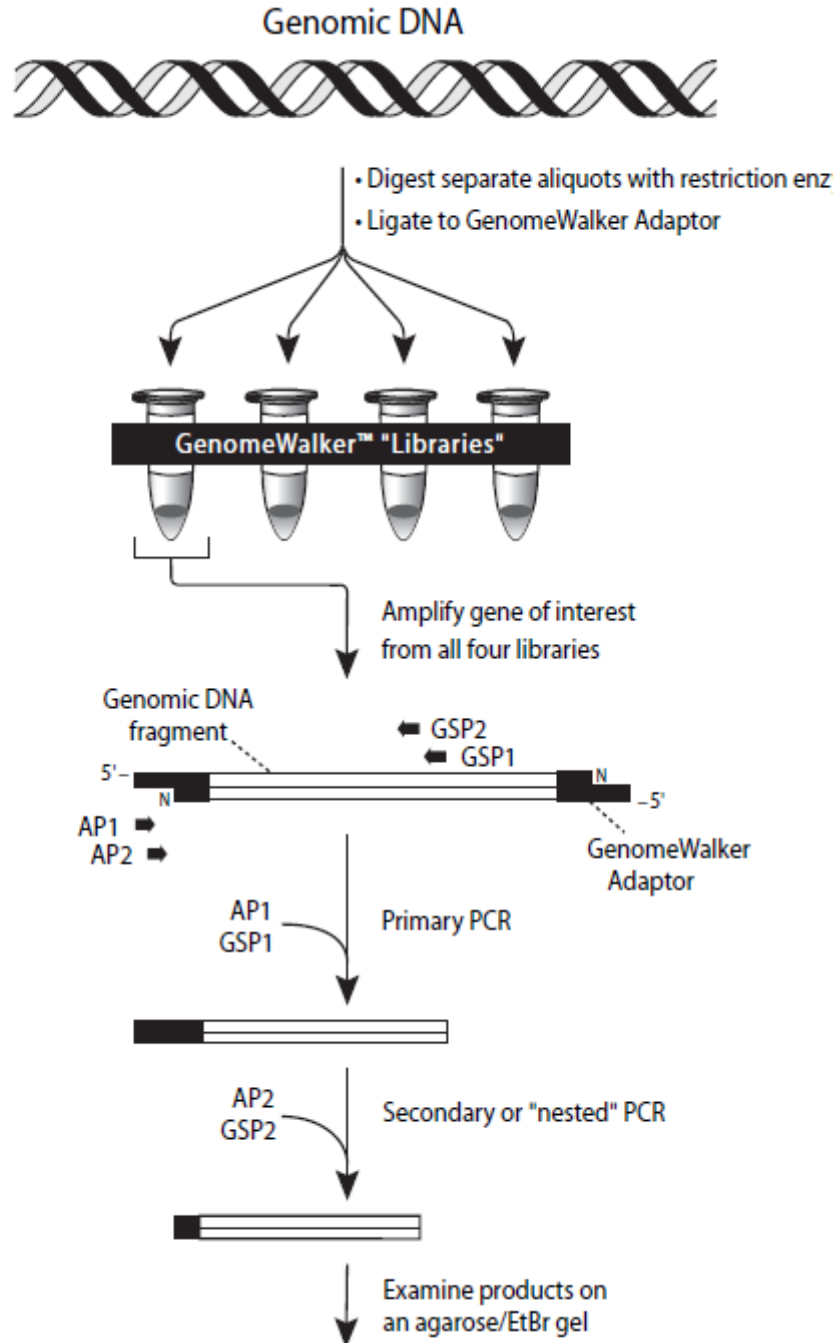
## **Isolation Of Genomic DNA:**

### **The basic principle**

With the **NucleoSpin® Tissue** method genomic DNA can be prepared from tissue, cells (e.g., bacteria), and many other sources. Lysis is achieved by incubation of the sample material in a proteinase K / SDS solution. Appropriate conditions for binding of DNA to the silica membrane in the **NucleoSpin® Tissue Columns** are created by the addition of chaotropic salts and ethanol to the lysate. The binding process is reversible and specific to nucleic acids. Contaminations are removed by subsequent washing with two different buffers. Pure genomic DNA is finally eluted under low ionic strength conditions in a slightly alkaline elution buffer.

### **Procedure for the isolation of genomic DNA:**

25 mg animal tissue is cut into small pieces and placed in a microcentrifuge tube and samples are prelysed by adding Buffer T1 and Proteinase K solution. Samples are vortexed and then Buffer B3 is added and then incubated at 70°C for 10 min. After that Ethanol is added for complete precipitation. Then each sample is placed on NucleoSpin Tissue column and then centrifuged for 1 min at high speed. The flow through is discarded and column is placed back on collection tube. The gDNA binded to silica membrane is washed 2-3 times with Wash Buffer BW. And then column is centrifuged again and flow through is discarded. Residual ethanol is removed during this step. Nucleospin column is placed into 1.5 ml microcentrifuge tube and 100 µl Buffer BE is added and incubated at room temperature for 5 min and finally centrifuged at high speed for 1min to get eluted DNA.



**Figure 4.6: Flow chart of the GenomeWalker™ protocol.** The gel shows a typical result generated by walking with GenomeWalker human libraries and gene-specific primers. Lane 1: EcoR V Library. Lane 2: Dra I Library. Lane 3: Pvu II Library. Lane 4: Ssp I Library. Lane M: DNA size markers. The absence of a major product in one of the libraries is not unusual. In our experience, there is no major band in one or more lanes in approximately half of the Genome Walker experiments.

The Genome-Walker Universal Kit enables us to create uncloned libraries for walking by PCR in any genomic DNA. In less than a week, the method provides access to the genomic DNA sequences adjacent to a known DNA sequence in any species. Using both the SMART RACE™ cDNA Amplification Kit and the Genome-Walker Universal Kit, we can clone full-length cDNAs and the surrounding genomic sequences without ever screening a library. In addition to obtaining promoters, Genome-Walker DNA walking can also be used to map intron/exon junctions and to walk bidirectionally from any sequence-tagged site (STS) or expressed sequence tag (EST). Although individual steps are limited to about 6 kb, multiple steps can be strung together to create longer walks. Consequently, this method is useful for filling in gaps in genome maps, particularly when the missing clones have been difficult to obtain by conventional library screening methods. In all applications, GenomeWalker PCR products are generally pure enough to allow restriction mapping without cloning. Nevertheless, a discussion of cloning PCR products and testing them for promoter activity is included at the end of this manual.

### **Construction of GenomeWalker™ Libraries:**

Construction of GenomeWalker DNA libraries was begun with very clean, high-molecular weight genomic DNA. This requires a higher quality preparation than the minimum suitable for Southern blotting or conventional PCR. Isolation procedures for genomic DNA can be found in various reference manuals (Ausubel, F. M. et al. 1994).

### **Quality of Genomic DNA:**

1. Checked the size of genomic DNA on a 0.6% agarose/EtBr gel.

Loaded 1 µl of experimental genomic DNA (0.1 µg/µl) and 1 µl of control genomic DNA (0.1 µg/µl) on a 0.6% agarose/EtBr gel in 1X TAE, along with DNA size markers, such as a 1 kb ladder or λ/Hind III digest. Genomic DNA should be bigger than 50 kb with minimum smearing.

2. Checked the purity of genomic DNA by Dra I digestion.

**a.** In a 0.5-ml reaction tube, combined the following:

5 µl Experimental genomic DNA

1.6 µl Dra I (10 units/µl)

2  $\mu$ l 10X Dra I Restriction Buffer

11.4  $\mu$ l Deionized H<sub>2</sub>O

Also set up a control digestion without enzyme.

**b.** Mixed gently by inverting tube, vigorous mixing will shear genomic DNA.

**c.** Incubated at 37°C overnight.

**d.** Run 5  $\mu$ l of each reaction on a 0.6% agarose/EtBr gel along with 0.5  $\mu$ l of experimental genomic DNA as a control. At this point, we have seen a smear, indicating that our DNA is digested by restriction enzymes.

### **Digestion of Genomic DNA**

For each library construction, set up for a total of five reactions was kept. For our experimental genomic DNA, set up four blunt-end digestions—one for each blunt-end restriction enzyme provided. Additionally, set up one Pvu II digestion of human genomic DNA as a positive control.

Labelled five 1.5-ml tubes: DL1, DL2, DL3, DL4, and positive control.

For each reaction, combine the following in a separate 1.5-ml tube:

25  $\mu$ l genomic DNA (0.1  $\mu$ g/ $\mu$ l)

8  $\mu$ l restriction enzyme (10 units/ $\mu$ l)

10  $\mu$ l restriction enzyme buffer (10X)

57  $\mu$ l deionized H<sub>2</sub>O

Mixed gently by inverting tube. Incubated at 37°C for 2 hr.

Vortexed the reaction at slow speed for 5–10 sec. Return to 37°C overnight (16–18 hr). From each reaction tube, removed 5  $\mu$ l and run on a 0.6% agarose/EtBr gel to determine whether digestion is complete.

### **Purification of DNA**

To each reaction tube, added an equal volume (95  $\mu$ l) of phenol. Vortexed at slow speed for 5–10 sec. Spined briefly at room temperature to separate the aqueous and organic phases. Using a pipet, transferred the upper (aqueous) layer into a fresh 1.5-ml tube. Discarded the lower

(organic) layer properly into the chlorinated hazardous waste. To each tube, added an equal volume (95  $\mu$ l) of chloroform. Vortexed at slow speed for 5–10 sec. Spined briefly at room temperature to separate the aqueous and organic phases, transferred the upper (aqueous) layer into a fresh 1.5-ml tube. Discard the lower (organic) layer properly into the chlorinated hazardous waste. To each tube, add 2 volumes (190  $\mu$ l) of ice cold 95% ethanol, 1/10 volume (9.5  $\mu$ l) of 3 M NaOAc (pH 4.5), and 20  $\mu$ g of glycogen. Vortexed at slow speed for 5–10 sec. Centrifuge at 14,000 rpm for 15 min at 4°C. Decant supernatant and wash pellet in 100  $\mu$ l of ice cold 80% ethanol. Centrifuged at 14,000 rpm for 10 min. Decant supernatant and air dry the pellet. Dissolve pellet in 20  $\mu$ l of TE (10/0.1, pH 7.5). Vortexed at slow speed for 5–10 sec. From each reaction tube, remove 1  $\mu$ l and run on a 0.6% agarose/EtBr gel to determine the approximate quantity of DNA after purification.

### **Ligation of Genomic DNA to GenomeWalker™ Adaptors**

For each library construction, we put a set up a total of five ligation reactions. We had blunt-end digestions of our experimental genomic DNA and one positive control Pvu II digestion of human genomic DNA.

From each tube, transfer 4  $\mu$ l of digested, purified DNA to a fresh 0.5-ml tube. To each, added the following:

1.9  $\mu$ l GenomeWalker Adaptor (25  $\mu$ M)

1.6  $\mu$ l 10X Ligation Buffer

0.5  $\mu$ l T4 DNA Ligase (6 units/ $\mu$ l)

Incubated at 16°C overnight.

To stop the reactions, it was incubated at 70°C for 5 min. To each tube, added 72  $\mu$ l of TE (10/1, pH 7.5). it was vortexed at slow speed for 10–15 sec.

### **A. Primer Design**

We had designed two gene-specific primers—one for primary PCR (GSP1) and one for secondary PCR (GSP2). The nested PCR primer should anneal to sequences beyond the 3' end of the primary PCR primer (i.e., upstream of the primary PCR primer when walking upstream and downstream of the primary PCR primer when walking downstream). Whenever possible, the

outer and nested primers should not overlap; if overlapping primers must be used, the 3' end of the nested primer should have as much unique sequence as possible. In general, the gene-specific primers were derived from sequences as close to the end of the known sequence as possible. For walking upstream from cDNA sequence, the primers were close to 5' ends. The primers were derived from the 5'UTR region and from Intron 1 region for Isoform 3 and 4. Gene-specific primers should be 26–30 nucleotides in length and have a GC content of 40–60% (we typically use 27-mers). This will ensure that the primers will effectively anneal to the template at the recommended annealing and extension temperature of 67°C. Primers should not be able to fold back and form intramolecular hydrogen bonds, and sequences at the 3' end of your primers should not be able to anneal to the 3' end of the adaptor primers. There should be no more than three G's and C's in the last six positions at the 3' end of the primer.

*Primers designed for genome walking for Isoform 1, 3, 4 and 5 from corresponding 5'UTR and Intron 1 region. (In-Intron, Is- Isoform, GW- Genome walking, R-Reverse)*

<b>Is1In1GW-R;</b>	<b>5' TCATTTGCAATTCATATATATGTCTTTGCCCATG 3'</b>
<b>Is1UTR-Nested1-R;</b>	<b>5' CTTTAAGATCCTAGTCCCCTACTTAAAC 3'</b>
<b>Is1UTR-Nested2-R;</b>	<b>5' ATCTGGAACACTGGCAGAGACACGTCC 3'</b>
<b>Is3In1GW-R;</b>	<b>5' ATTTATTCCCCATTCTTCAAATATGG 3'</b>
<b>Is3In1-nested1-R;</b>	<b>5' TGGTATCACTGCCAGTCAAATTACGCT 3'</b>
<b>Is3In1-nested2-R;</b>	<b>5' CCTCCTTAACACTGAAAACCACTCCCC 3'</b>
<b>Is4In1-GW-R;</b>	<b>5' AGGAAACAACAGACAGCACGCTGTGGG 3'</b>
<b>Is4In1-nested1-R;</b>	<b>5' CCCACAGCGTGCTGTCTGTTGTTTCCT 3'</b>
<b>Is4UTR-nested2-R;</b>	<b>5' AACTGTTTGCCTGTAAGATCCCAGTCCCA 3'</b>
<b>Is5In1-GW-R;</b>	<b>5' CTCCCCCTTTCCTTGCATGTCTCACAG 3'</b>
<b>Is5In1-nested1-R;</b>	<b>5' ATTAGTGTCTCATAGCTTACCAACCG 3'</b>
<b>Is5UTR-nested2-R;</b>	<b>5' AACTCTGTGCGACAAACTGTTTACCTG 3'</b>

#### **Procedure for PCR-based DNA Walking in GenomeWalker Libraries:**

The Genome-Walker DNA walking protocol consists of eight primary and secondary PCR amplifications: four experimental libraries, two positive controls (Genome-Walker Human Positive Control Library and one positive control library constructed from Control Human Genomic DNA), and two negative controls (without templates). For both positive controls, use the positive control gene-specific primers, PCP1 and PCP2 (provided). For primary PCR, use 1 µl of each library. For secondary PCR, use 1 µl of a 50X dilution of the primary PCR product.

Labeled the 0.5-ml PCR tubes. For convenience, I used the plan in Table I (GSP1 and GSP2 indicate my gene-specific primers):

<b>TABLE I. SUGGESTED LABELING PLAN</b>				
<b>DNA Library (DL)</b>	<b>Tube No.</b>	<b>1° PCR Primers</b>	<b>Tube No.</b>	<b>2° PCR Primers</b>
DL-1	1A	GSP1 & AP1 <sup>a</sup>	1B	GSP2 & AP2 <sup>b</sup>
DL-2	2A	"	2B	"
DL-3	3A	"	3B	"
DL-4	4A	"	4B	"
<b>Negative control No.1</b>				
None	5A	"	5B	"
<b>Positive control No.1</b>				
Control library <sup>c</sup>	6A	PCP1 & AP1 <sup>a</sup>	6B	PCP2 & AP2 <sup>b</sup>
<b>Negative control No.2</b>				
None	7A	"	7B	"
<b>Positive control No.2</b>				
Pre-constructed <sup>d</sup> control library	8A	"	8B	"

<sup>a</sup> Primer contained in primary PCR master mix.  
<sup>b</sup> Primer contained in secondary PCR master mix.  
<sup>c</sup> Positive control for library construction. You construct this library from the control human genomic DNA provided in the kit (see Section IV).  
<sup>d</sup> Positive control for PCR. This preconstructed library is included in the kit.

Prepared enough **primary PCR master mix** for all eight reactions plus one additional tube. Combine the following reagents in a 0.5-ml tube:

9 rxns per rxn:

360 µl 40 µl deionized H<sub>2</sub>O

45 µl 5 µl 10X Advantage 2 PCR Buffer

9 µl 1 µl dNTP (10 mM each)

9 µl 1 µl AP1 (10 µM)

9 µl 1 µl Advantage 2 Polymerase Mix (50X)

432 µl 48 µl Total volume

Mixed well by vortexing (without introducing bubbles) and briefly spin the tube in a microcentrifuge. Added 48  $\mu$ l of the primary PCR master mix to the appropriately labeled tubes. For reactions 1A through 5A, add 1  $\mu$ l of GSP1 to each tube. For reactions 6A through 8A, added 1  $\mu$ l of PCP1 to each tube. Added 1  $\mu$ l of each DNA library (including the positive control library) to the appropriately labeled tubes. Added 1  $\mu$ l of H<sub>2</sub>O to each negative control. Briefly spinned the tubes in a microcentrifuge. Commenced the thermal cycling using the following two-step cycle parameters:

- 7 cycles: 94°C 25 sec,  
72°C 3 min
- 32 cycles: 94°C 25 sec  
67°C 3 min
- 67°C for an additional 7 min after the final cycle.

Analyzed the 5  $\mu$ l of the primary PCR products on a 1.5% agarose/EtBr gel, along with DNA size markers such as a 1 kb ladder.

**Expected results of primary PCR:** In all lanes except for negative controls, we should observe your predicted banding patterns. If we didn't not observe **any** product or smear with our gene-specific primers, then diluted 1  $\mu$ l of each primary PCR (including positive and negative controls) into 49  $\mu$ l of deionized H<sub>2</sub>O. Prepared a **secondary PCR master mix** for all eight reactions plus one additional tube. Combined the following reagents in 0.5-ml tube:

9 rxns per rxn

360  $\mu$ l 40  $\mu$ l deionized H<sub>2</sub>O

45  $\mu$ l 5  $\mu$ l 10X Advantage 2 PCR buffer

9  $\mu$ l 1  $\mu$ l dNTP (10 mM each)

9  $\mu$ l 1  $\mu$ l AP2 (10  $\mu$ M)

9  $\mu$ l 1  $\mu$ l Advantage 2 Polymerase Mix (50X)

432  $\mu$ l 48  $\mu$ l Total volume

**Mixed well** by vortexing (without introducing bubbles) and briefly spinned the tube in a microcentrifuge. Added 48  $\mu$ l of the secondary PCR master mix to the appropriately labeled tubes (Table I). For reactions 1B through 5B, added 1  $\mu$ l of GSP2 to each tube. For reactions 6B

through 8B, add 1  $\mu$ l of PCP2 to each tube. Added 1  $\mu$ l of each diluted primary PCR product (from Step 11) to the appropriately labeled tubes. I have included the positive and negative controls. Briefly spined tubes in a microcentrifuge. Commenced the thermal cycling using the following two-step cycle parameters:

- 5 cycles: 94°C 25 sec

72°C 3 min

- 20 cycles: 94°C 25 sec

67°C 3 min

- 67°C for an additional 7 min after the final cycle.

Analyzed the 5  $\mu$ l of the secondary PCR products on a 1.5% agarose/EtBr gel, along with DNA size markers such as a 1 kb ladder. If you do not see any product, perform four additional cycles. Stored the unused portion of each secondary PCR at 4°C until we have confirmed that the procedure has been successful. At that point, we proceed with analyzing and cloning the fragments of interest.

---

# *Chapter 4*

---

## **Results and Discussions**

## Chapter 4: Results and Discussions

---

### Results

**(1) Isolation of total RNA:** Total RNA isolated from venom gland shows the 2 distinct bands of 18s and 28s RNA of 750bp and 1Kb respectively (Figure 5.1).

**(2) Molecular weight of respective isoforms:**

Isoform 1 has the expected base pair size of 750 bp and it shows the specific band at 750 bp. Similarly the specific amplification was found for Isoform 4,5,6,10 was at 750 bp. Isoform 2 has amplification at 1Kb and Isoform 3 has specific amplification at 700bp. Isoform 9 shows the specific amplification at 800 bp. Primers designed for Isoform 8 showed no amplification for the size of expected base pair due to its low  $T_m$  value. Further nested primers were designed for Isoform 8 and the RACE PCR reaction mix was put for multigradient temperature PCR. Then it shows the specific amplification at 750 bp. Similarly nested primers were designed for Isoform 11 and it shows the specific amplification at 700 bp.

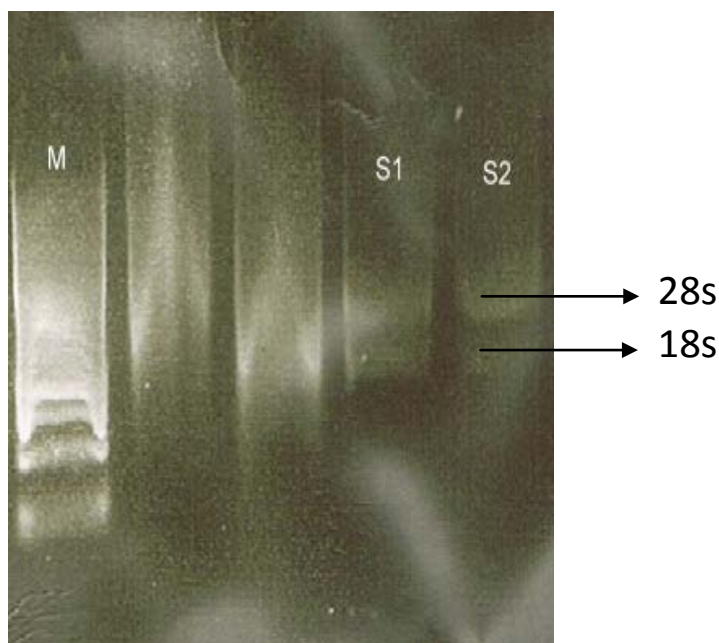
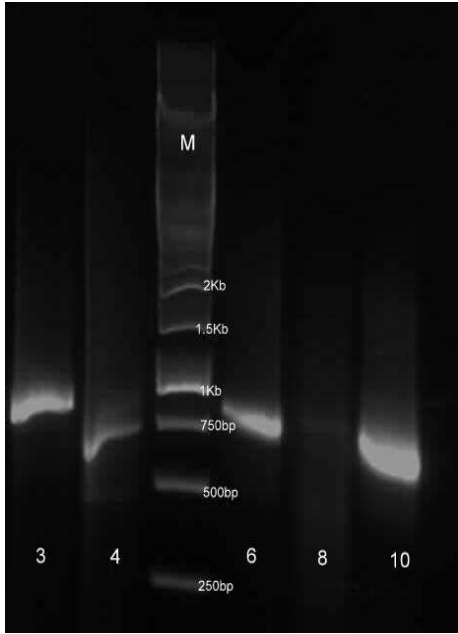
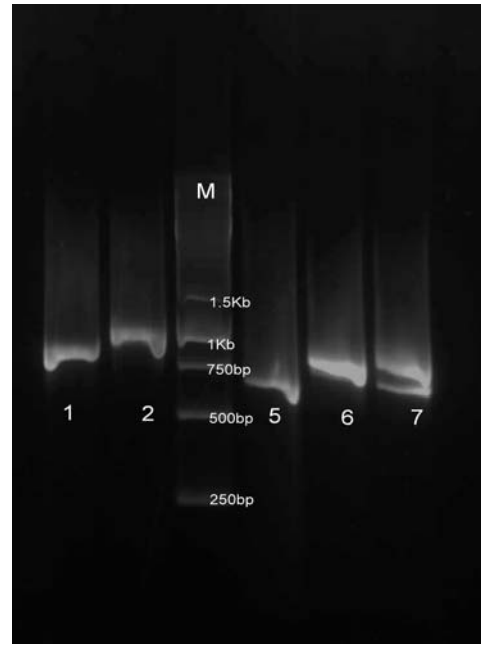


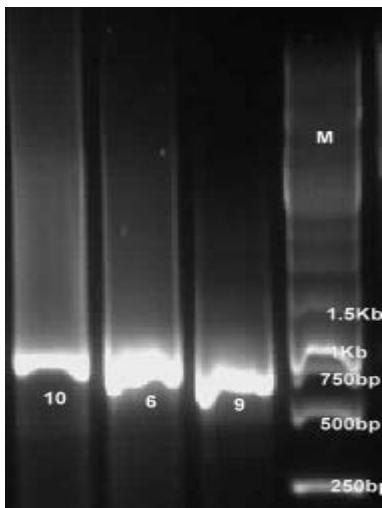
Figure 5.1: Isolation of total mRNA from venom gland



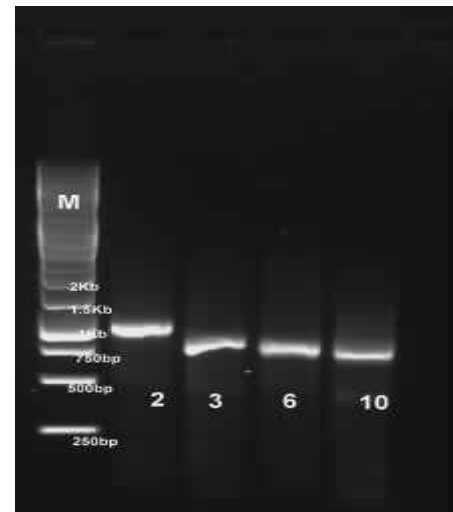
**Figure 5.2;** Isoform 3: 800 bp, Isoform 4: 750 bp, Isoform 6: 750bp, Isoform 10: 750 bp



**Figure 5.3:** Isoform 1: 750 bp, Iso 2: 1Kb, Iso 5: 750 bp Isoform 6: 750 bp, Isoform 7: 750 bp



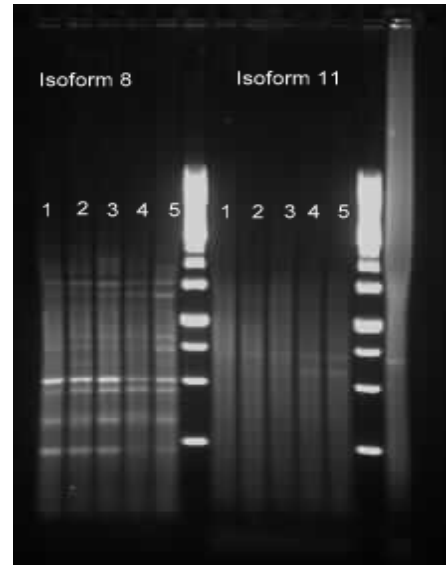
**Fig 5.4:** Isoform 10: 750 bp, Isoform 6: 750 bp, Isoform 9: 750 bp



**Fig 5.5:** Isoform 2: 1Kb, Isoform 3: 750 bp, Isoform 6: 750 bp, Isoform 10: 750 bp

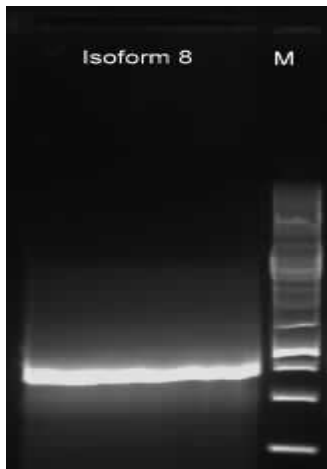


**Fig 5.6:** Isoform 9: Multigradient Temperature PCR  
 1: Tm= 70/68, 2: Tm= 68/65, 3: Tm= 67/63, 4: Tm=65/  
 5: Tm= 63/61;

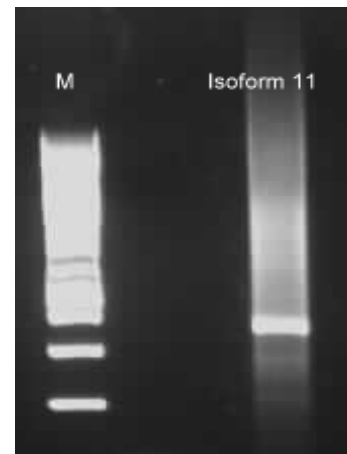


**Fig5.7:** Isoform 8: Multigrad Temp PCR  
 1: Tm= 65/60, 2: Tm= 63/57, 3: Tm=63/60  
 4: Tm= 60/55. 5: Tm= 60/53, Isoform 11: 60/55

Gel amplification results for all the isoforms were reproduced very well. Multigradient temperatures PCR were set for Isoform 8, 9, 11. But it didn't give the specific amplification and in many instances bands were too close to run it for larger quantity as they start merging in each other. As alternative, nested primers were designed for Isoform 8 and 11 and specific amplification at 750 bp was get for both the isoforms (Fig 5.8, Fig 5.9).



**Fig 5.8:** Isoform 8 (Tm=72/70)

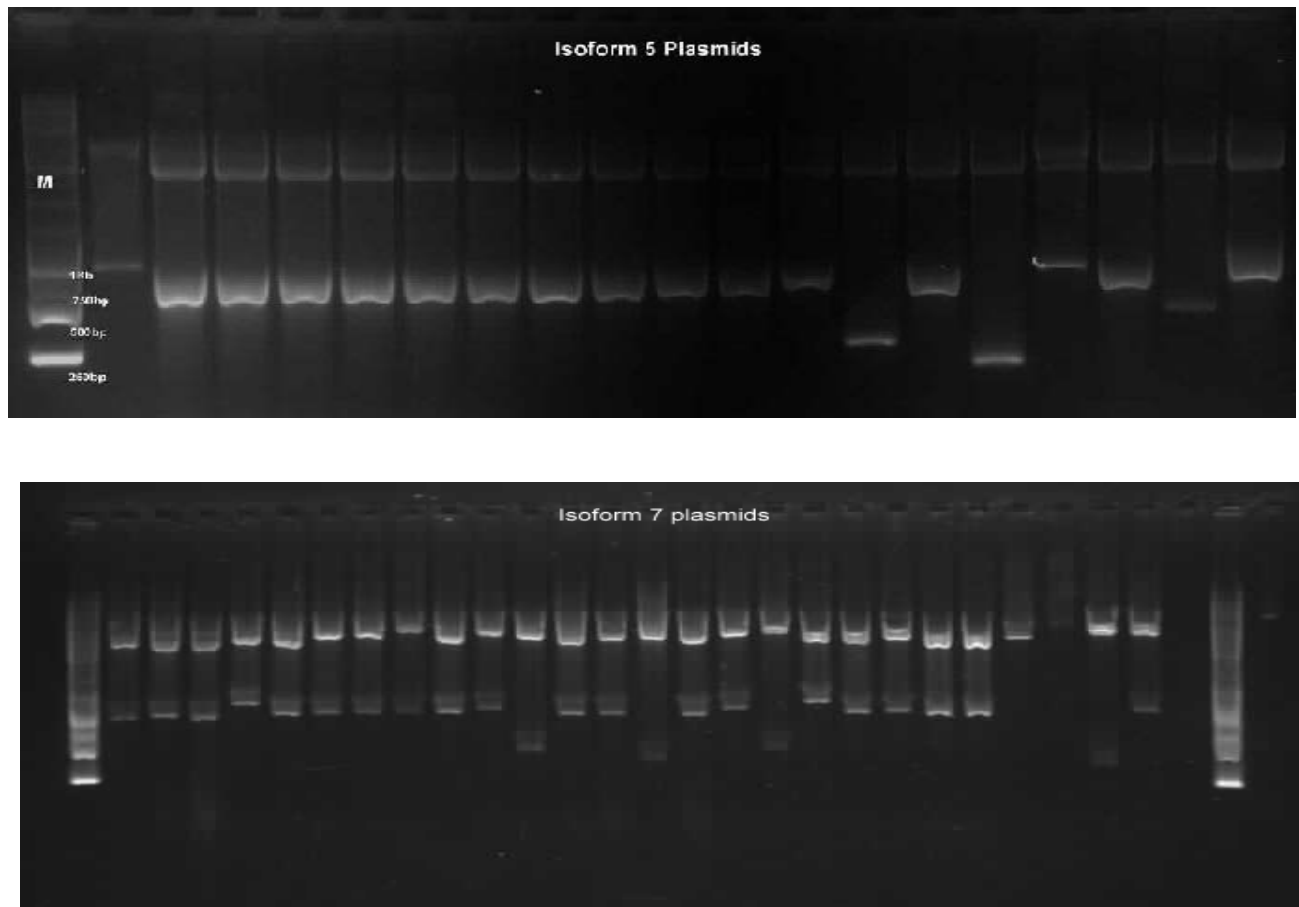


**Fig 5.9:** Isoform 11 (Tm=60/55)

### **(3) Restriction digestion results for the Isoforms:**

After ligating the insert of RACE PCR amplified product to pJET work, restriction digestion was performed with BglII restriction enzyme. To check whether the insert is in proper frame and in order to screen the clones for the most prominent 5'UTR regions, about 50-60 colonies were picked up for each isoform and were digested with BglII.

All of the isoforms shows the release of insert at expected base pair size. So, only 2 Isoforms are shown for the release of insert.



**Fig 5.10: Isoform 5 and Isoform 7 plasmids after digesting with BglII.**

#### **(4) 5'UTR sequences identified:**

After doing the sequence alignment of RACE amplified PCR product to the template cDNA, 5'UTR regions were identified. 5'UTR region starts with base "G", which is also known as **5' G capping** in mRNA. The clone which starts with 'G' and have maximum number of similarity with other identified clone is chosen to be the best clone of certain length of 5'UTR.

### **Isoform 1**

Complete sequence with 5'UTR was obtained for Isoform 1 and was aligned with template cDNA. 40 clones with different sequence of 5'UTR were obtained. The clone showing the maximum similarity with other clones was taken as a representative clone for 5'UTR. 195bp sequence was identified as 5'UTR region for Isoform 1.

#### **Complete sequence:**

>2  
GGATGACACTCTGGACGTGCTCTGCCAGTGTCCAGATTTTTGGCCATCCAGCTGCTTAATTTGATCAAAATAAGTGTCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTATGATACGGTATCTCAAGTT  
AAGTAGGGGACTAGGATCTTAAAGCAAACAGCTTCCGCTGCAGAGTTGAAGCTATGGTCTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTTTCTTATGCACAAAAGTCTTGAACCTGGTCGTTGGAGGTG  
ATGAATGTAACATAAATGAACATCGTTTCCTGTAGCCTTGATCACTTAGGTCTAAGACTTTTCTCGCGGTGGGACTTTGCTCAATGAGGAATGGGTGCTCACTGCTGCACACTGTGACAGGAGAAATATGCAGATA  
AAGCTTGGTATGCATAGCAAACCTGTACCAAATGAGGATGAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTAGCAAAAACCTATACCCCTTGGGACAAGGACATCATGTTGATCAGGCTGGACAGACCT  
GTTAGCAACAGTGAACACATCGCGCTCTCAGCTTGCCTCCAGCCCTCTAGTGTGGCTCAGTTGCCGTATTATGGGATGGGGCAGAATCTCACCTAATAAAGAGACTTATCCAATGCTCCCTATTGTGCTAACATT  
AACATACTCAATTATGAGGTGTGTG

Multiple sequence alignment with isoform 1:

```
1cDNA -----
clone2  GGATGACACTCTGGACGTGCTCTGCCAGTGTCCAGATTTTTGGCCATCCAGCTGCTT 60

1cDNA -----
clone2  AATTTGATCAAAATAAAGTGTCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA 120

1cDNA -----
clone2  TTGATACGGTATCTCAAGTTTAAGTAGGGGACTAGGATCTTAAAGCAAACAGCTTGCCG 180

1cDNA -----ATGGTGTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTT 45
clone2  TGCAGAGTTGAAGCTATGGTGTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTT 240
*****

1cDNA TCTTATGCACAAAAGTCTTCTGAACTGGTCGTTGGAGGTGATGAATGTAACATAAATGAA 105
clone2 TCTTATGCACAAAAGTCTTCTGAACTGGTCGTTGGAGGTGATGAATGTAACATAAATGAA 300
*****

1cDNA CATCGTTTCCTTGTAGCCTTGTATCACTTAGGTCTAAGACTTTTCTCTGCGGTGGGACT 165
clone2 CATCGTTTCCTTGTAGCCTTGTATCACTTAGGTCTAAGACTTTTCTCTGCGGTGGGACT 360
*****

1cDNA TTGCTCAATGAGGAATGGGTGCTCACTGCTGCACACTGTGACAGGAGAAATATGCAGATA 225
clone2 TTGCTCAATGAGGAATGGGTGCTCACTGCTGCACACTGTGACAGGAGAAATATGCAGATA 420
*****

1cDNA AAGCTTGGTATGCATAGCAAACCTGTACCAAATGAGGATGAGCAGACAAGAGTCCCAAAG 285
clone2 AAGCTTGGTATGCATAGCAAACCTGTACCAAATGAGGATGAGCAGACAAGAGTCCCAAAG 480
*****

1cDNA GAGAAGTCTTTTGTCTTAGTAGCAAAAACCTATACCCCTTGGGACAAGGACATCATGTTG 345
clone2 GAGAAGTCTTTTGTCTTAGTAGCAAAAACCTATACCCCTTGGGACAAGGACATCATGTTG 540
*****

1cDNA ATCAGGCTGGACAGACCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCC 405
clone2 ATCAGGCTGGACAGACCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCC 600
*****
```

The brown region is 5'UTR.

5'UTR region identified for all the 14 clones of Isoform 1 cDNA:

```
>1
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTGATACGGTATCTCAAGTTT
AAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>2
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>3
ACGGGGGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTG
ATTAGATTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>4
GATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTGATACGGTATCTCAAGTTTAA
GTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>5
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>6
GGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTG
ATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>7
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTGATACGGTATCTCAAGTTT
AAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>8
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTGATACGGTATCTCAAGTTT
AAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>9
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>10
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>11
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>12
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>13
GGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTG
ATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>14
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGA
TTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
```

CLUSTAL 2.1 multiple sequence alignment

```
6          -----GGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
9          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
12         -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
4          -----GATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
13         -----GGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
14         -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
10         -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
11         -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
8          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
7          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
5          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
1          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
3          ACGGGGGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
2          -----GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAG
*****
```

The 5'UTR region of Isoform 1:

```
GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGC
TGCTTGATCAAGAAGTCTCCACTGGGTTATCTGATTAGATTGATACGGTATCTCAAGTTTAAAGTAGGGGACTAGGATC
TTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
```

## Isoform 2

141 bp sequences upstream to ATG was obtained as 5'UTR region for Isoform 2 by aligning the 20clones together. And all the clones were similar. No different 5'UTR sequence was found in case of Isoform2.

### Complete sequence

>2  
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCAGCTGCTTAATTGAGCAACAGCTTGCCTGTCAG  
AGTTGAAGCTATGGTCTGATCAGAGTGTAGCAAACTTCTGATACACAGCTTTCTATGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGATGAATGTAACATAAATGAAACATCGTTTCTTGCACCT  
GTGTAAGCTGATGGGAATCAATGCGGTGGGACTTTGATCAATGAGGAATGGGTGCTCACTGCTGCACACTGCGAAGGGAACAAAATGAAGATACATCTTGGTATGCATAGCAAAAAGGTACCAAAATAG  
GATAAACAGACAAGAGTCCCAAGGAGAAGTTCTTTGTGTCAGTAGCAAAAACACACCTTTGGGACAAGGACATCATGTTGATCAGGCTGGACAGACCTGTTGGCAACAGTGAACACATTCGCGCTCT  
CAGCTTGCCTTCCAGCCTCCAGTGGGCTCAGTTTCCGATATTGGGATGGGGACAATCTCACACTAAAGTATTTGCTGATGTCCTCACTGTGTTAACATTAACTACTCAATTATTCGGGTG  
GTCGAGCAGCTTACCAGAATATGGTTGCCAGCGACAAGCAGAATTTGTGTCAGGATTCTGGAAGGAGGCAAGATACATGTTGGGAGACTCTGGGGACCCCTCATTTGTAATGACAATTCCA  
GGGAATTGCATCTTGGGGAGCCCTAACTGTGGTTACGTGCGTGAGCTGGCCTCACCAAGGTTCTCGATCATCTTGACTGGATCCAGAGCATTATTGAGCAAAATACAGATGCGACCTGCCCCCAT  
GA

### Multiple sequence alignment with Isoform 2 cDNA:

CLUSTAL W (1.81) multiple sequence alignment

```
2. cDNA -----  
clone2 GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGTGACACTTCTGGA  
  
2. cDNA -----  
clone2 CGTAACTCTGTGACGCTCCAGATTGTTGGCCACCAGCTGCTTAATTGAGCAACAGC  
  
2. cDNA -----ATGGTGTGCTGATCAGAGTGTAGCAAACTTCTGATACTA  
clone2 TGGCCTGTCAGAGTTGAAGCTATGGTGTGCTGATCAGAGTGTAGCAAACTTCTGATACTA  
*****  
  
2. cDNA CAGCTTTCTTATGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGATGAATGTAACATA  
clone2 CAGCTTTCTTATGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGATGAATGTAACATA  
*****  
  
2. cDNA AATGAACATCGTTTCTTGCACCTTGTGACTCTGATGGGAATCAATGCGGTGGGACTTTG  
clone2 AATGAACATCGTTTCTTGCACCTTGTGACTCTGATGGGAATCAATGCGGTGGGACTTTG  
*****  
  
2. cDNA ATCAATGAGGAATGGGTGCTCACTGCTGCACACTGCGAAGGGAACAAAATGAAGATACAT  
clone2 ATCAATGAGGAATGGGTGCTCACTGCTGCACACTGCGAAGGGAACAAAATGAAGATACAT  
*****  
  
2. cDNA CTTGGTATGCATAGCAAAAAGGTACCAATAAGGATAAACAGACAAGAGTCCCAAGGAG  
clone2 CTTGGTATGCATAGCAAAAAGGTACCAATAAGGATAAACAGACAAGAGTCCCAAGGAG  
*****  
  
2. cDNA AAGTTCTTTGTGTCAGTAGCAAAAACACACCTTTGGGACAAGGACATCATGTTGATC  
clone2 AAGTTCTTTGTGTCAGTAGCAAAAACACACCTTTGGGACAAGGACATCATGTTGATC  
*****  
  
2. cDNA AGGCTGGACAGACCTGTTGGCAACAGTGAACACATTGCGCCTCTCAGCTTGCCTTCCAGC  
clone2 AGGCTGGACAGACCTGTTGGCAACAGTGAACACATTGCGCCTCTCAGCTTGCCTTCCAGC  
*****  
  
2. cDNA CCTCCAGTGTGGGCTCAGTTTGCCTGATTTATGGGATGGGGACAATCTCACCTACTAAA  
clone2 CCTCCAGTGTGGGCTCAGTTTGCCTGATTTATGGGATGGGGACAATCTCACCTACTAAA  
*****  
  
2. cDNA GTGATTTGCTGATGTCCTCACTGTGTTAACATTAACCTACTCAATTATTCGGTGTGT  
clone2 GTGATTTGCTGATGTCCTCACTGTGTTAACATTAACCTACTCAATTATTCGGTGTGT  
*****  
  
2. cDNA CGAGCAGCTTACCAGAAATATGGGTTGCCAGCGACAAGCAGAACATTGTGTCAGGTATT  
clone2 CGAGCAGCTTACCAGAAATATGGGTTGCCAGCGACAAGCAGAACATTGTGTCAGGTATT  
*****  
  
2. cDNA CTGGAAGGAGGCAAGATACATGTGTGGGAGACTCTGGGGGCCCTCATTGTAATGGA  
clone2 CTGGAAGGAGGCAAGATACATGTGTGGGAGACTCTGGGGGCCCTCATTGTAATGGA  
*****  
  
2. cDNA CAATCCAGGGAATTGCATCTTGGGGAGCCCTAACTGTGGTTACGTGCGTGAGCCTGGC  
clone2 CAATCCAGGGAATTGCATCTTGGGGAGCCCTAACTGTGGTTACGTGCGTGAGCCTGGC  
*****  
  
2. cDNA CTCTACACCAAGGTCTTCGATCATCTTACTGGATCCAGAGCATTATTGAGCAAAATACA  
clone2 CTCTACACCAAGGTCTTCGATCATCTTACTGGATCCAGAGCATTATTGAGCAAAATACA  
*****
```

The cyan region is 5'UTR.

5'UTR region identified for all the 10 clones of Isoform 2:

```
>1
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>2
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>3
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>4
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>5
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>6
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>7
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>8
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>9
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
>10
GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAG
AGTTGAAGCT
```

5' UTR region of Isoform 2:

GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGACGCTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT

## Isoform 3

194 bp sequence was obtained as 5'UTR for Isoform 3 by aligning the 25 clones together.

### Complete Sequence:

```
>3
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTGATCAAGTAGTCTCCGCTTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCTATGGTGCTGATCAGAGTGCTAGCAAATCTTGTACTGCAGCTTCTTATGCACAAAAGTCTTCTGAACCTGG
TCATTGGAGGTGATGAATGTAACGTAACGTAACATCGTTCCCTGTAGCCATCTTAACTCTACCGGGTTTTTCTGCGCTGGGACTTTGATCAACAAGGAATGGGTGCTCACTGCTGCCACTGCGACAGTAC
AAATTTCCAATGAAGCTTGGTGTGATAGCAAAAAGTATCAAATGAGGATGAGCAGACAAGAAACCCAAAGGAGAAGTTCTTTTGCCGATAAAAAAGATGACGTACTGGACAAGGACATCATGTTGA
TCAAGCTGGACAGTCTGTAGCAACAGTGAACACATCGCCTCTCAGCTTGCCCTCCAGCCCTCCAGTGTGGGCTCAGTTTGCCATATTATCGGATGGGATCAATCACACCTATTGAAAAGACTTTGCC
TGATGCCCTTATTGTCTAACATTAACCTACTCGATGATCGGGTGTCAACCACTTATCCAGAATTGCCGCGACAAGCAGAACATTGTGTCAGGTATCCTGGAAGGAGGCAAAGATACATGTGGGG
TGACTCTGGGGACCCCTCATCTGTAATGGACAATTCAGGGCATTGATTTTATGGGGCGCATCTTGTGGCAAGCTCTTAAGCTGGTGTCTACACCAAAGCTTTGACTATAATGACTGGATTACAGAGC
ATTATTGAGGAAATACAGCTGCAACGTGCCCCCGTGA
```

### Multiple Sequence Alignment with Isoform 3 cDNA:

CLUSTAL W (1.81) multiple sequence alignment

```
3. cDNA
clone3      -----
            GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTA
3. cDNA
clone3      -----
            AATTTGATCAAATAAAGTGCTGCTGATCAAGTAGTCTCCGCTTGGGTTATCTGATTAGAC
3. cDNA
clone3      -----
            TGATACGGTATCTCAATTTTAAGTAAGGACTGGGATCTTACAGGCAAACAGCTTTCCTT
3. cDNA
clone3      -----ATGGTGCTGATCAGAGTGCTAGCAAATCTTCTGATACTGCAGCTTT
            CGAGAGTTGAAGCTATGGTGCTGATCAGAGTGCTAGCAAATCTTCTGATACTGCAGCTTT
            *****
```

```

3. cDNA      CTTATGCACAAAAGTCTTCTGAACTGGTCATTGGAGGTGATGAATGTAACGTAATGAAC
clone3      CTTATGCACAAAAGTCTTCTGAACTGGTCATTGGAGGTGATGAATGTAACGTAATGAAC
*****

3. cDNA      ATCGTTCCCTTGTAGCCATCTTTAACTCTACCGGGTTTTTCTGCGCTGGGACTTTGATCA
clone3      ATCGTTCCCTTGTAGCCATCTTTAACTCTACCGGGTTTTTCTGCGCTGGGACTTTGATCA
*****

3. cDNA      ACAAGGAATGGGTGCTCACTGCTGCCACTGCGACAGTACAAATTTCCAATGAAGCTTG
clone3      ACAAGGAATGGGTGCTCACTGCTGCCACTGCGACAGTACAAATTTCCAATGAAGCTTG
*****

3. cDNA      GTGTGCATAGCAAAAAGGTATCAAATGAGGATGAGCAGACAAGAAACCCAAAGGAGAAGT
clone3      GTGTGCATAGCAAAAAGGTATCAAATGAGGATGAGCAGACAAGAAACCCAAAGGAGAAGT
*****

3. cDNA      TCTTTTGTCCCGATAAAAAGATGACGTAAGGACATCATGTTGATCAAGCTGG
clone3      TCTTTTGTCCCGATAAAAAGATGACGTAAGGACATCATGTTGATCAAGCTGG
*****

3. cDNA      ACAGTCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCCAGCCCTCCCA
clone3      ACAGTCTGTTAGCAACAGTGAACACATCGCGCCTCTCAGCTTGCCTTCCAGCCCTCCCA
*****

3. cDNA      GTGTGGGCTCAGTTTGCCATATTATCGGATGGGATCAATCACACATTGAAAAGACTT
clone3      GTGTGGGCTCAGTTTGCCATATTATCGGATGGGATCAATCACACATTGAAAAGACTT
*****

3. cDNA      TGCTGATGTCCTTATTGTGCTAACATTAACTACTCGATGATGCGGTGTGCAACCAC
clone3      TGCTGATGTCCTTATTGTGCTAACATTAACTACTCGATGATGCGGTGTGCAACCAC
*****

```

The cyan region is 5'UTR.

5'UTR regions of all the clones

```

>1
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>2
GACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTCAATTTAAGTAAG
GGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>3
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>4
GTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>5
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>6
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>7
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>8
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>9
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>10
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>11
GACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTCAATTTAAGTAAG
GGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT
>12
GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTC
AATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT

```

5'UTR region of Isoform 3:

```

GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCT
GCTTGATCAAGTAGTCTCCGCTGGGTTATCTGATTAGACTGATACGGTATCTCAATTTAAGTAAGGGACTGGGATCTT
ACAGGCAAACAGCTTTCCTTGCAAGTTGAAGCT

```

## Isoform 4

Only 57 bp sequence of 5'UTR region was obtained by aligning the 40 clones together for Isoform 4.

### Complete Sequence:

>3

```
GGACGTTTGGGACTGGGATCTTACAGGCAACAGTTTCTGCGCAGAGTTGAAGCTATGGTGTGATCAGAGTGTAGCAACCTTCTGATACTACAGCTTTCTACGCACAAAAGTCTTGAATTGGTCG
TTGGAGGTGAAGAATGAACATAAATGAACATCGTTCCCTGTAGTCTTTAACTTACCAATTTCTGCGCTGGGACTTTGATCAACCAGGAATGGGTGCTCACTGCTGCCACTGCGACAGTAAAAAT
TCCAGATGACAGCTTGGTGTGCATAGCAAAAAGGTAATAATGAGGATGAGCAGACAAGAGACCCAAAGGAGAAGTTCATTTGTCCTCAATAGGAAAAAGATGATGAAAAGGACAAGGACATCATGTGA
TCAGGCTGGACAGACCTGTGCAATAGTGAACACATCGCGCTCTCAGCTTACCTTCCAGCCCTACCAAGTGGGGCTCAGTTTCCGCTATTATGGGATGGGCAATAACTCACCTACTAAAGAGACATTCAC
CGATGTCCTCATTGTGCTAACATTAACATACTCATCATGAGGTGTGCAAGCAGCTTATCCATCATTGACGGCGACAAGCGA
```

### Multiple Sequence Alignment with Isoform 4 cDNA:

CLUSTAL W (1.81) multiple sequence alignment

```
4. cDNA -----ATGG
clone6 GGACGTTTGGGACTGGGATCTTACAGGCAACAGTTTCTGCGCAGAGTTGAAGCTATGGTGTGATCAGAGTGTAGCAACCTTCTGATACTACAGCTTTCTACGCACAAAAGTCTTGAATTGGTCG
*****

4. cDNA TGCTGATCAGAGTGTAGCAACCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTT
clone6 TGCTGATCAGAGTGTAGCAACCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTT
*****

4. cDNA CTGAATTGGTCGTTGGAGGTGAAGAATGTAACATAAATGAACATCGTTCCTTGTAGTCT
clone6 CTGAATTGGTCGTTGGAGGTGAAGAATGTAACATAAATGAACATCGTTCCTTGTAGTCT
*****

4. cDNA TCTTTAACTCTACCAATTTCTCTGCGCTGGGACTTTGATCAACCAGGAATGGGTGCTCA
clone6 TCTTTAACTCTACCAATTTCTCTGCGCTGGGACTTTGATCAACCAGGAATGGGTGCTCA
*****

4. cDNA CTGCTGCCACTGCGACAGTAAAAATTTCCAGATGCAGCTTGGTGTGCATAGCAAAAAGG
clone6 CTGCTGCCACTGCGACAGTAAAAATTTCCAGATGCAGCTTGGTGTGCATAGCAAAAAGG
*****

4. cDNA TACTAAATGAGGATGAGCAGACAAGAGACCCAAAGGAGAAGTTCATTTGTCCCAATAGGA
clone6 TACTAAATGAGGATGAGCAGACAAGAGACCCAAAGGAGAAGTTCATTTGTCCCAATAGGA
*****

4. cDNA AAAAAGATGATGAAAAGGACAAGGACATCATGTTGATCAGGCTGGACAGACCTGTGACGA
clone6 AAAAAGATGATGAAAAGGACAAGGACATCATGTTGATCAGGCTGGACAGACCTGTGACGA
*****

4. cDNA ATAGTGAACACATCGCGCCTCTCAGCTTACCTTCCAGCCCTACCAGTGTGGGCTCAGTTT
clone6 ATAGTGAACACATCGCGCCTCTCAGCTTACCTTCCAGCCCTACCAGTGTGGGCTCAGTTT
*****

4. cDNA GCCGTATTATGGGATGGGCATAAACTCACCTACTAAAGAGACTTATCCCGATGTCCTC
clone6 GCCGTATTATGGGATGGGCATAAACTCACCTACTAAAGAGACTTATCCCGATGTCCTC
*****

4. cDNA ATTGTGCTAACATTAACATACTCCATCATGAGGTGTGTCAGCAGCTTATCCATCATTGA
clone6 ATTGTGCTAACATTAACATACTCCATCATGAGGTGTGTCAGCAGCTTATCCATCATTGA
*****

4. cDNA CGGCGACAAGCGAACACTGTGTGCAAGTATCCTGGAAGGAGGCAAGATGCATGTGGGG
clone6 CGGCGACAAGCGA-----
*****

4. cDNA GTGACTCTGGGGACCCCTCATCTGTAATGGAGAAATCCAGGCGATTGTATCTGGGGGG
clone6 GTGACTCTGGGGACCCCTCATCTGTAATGGAGAAATCCAGGCGATTGTATCTGGGGGG
-----

4. cDNA AGCATCCTGTGGCCAAGTCTTACTCCTGGCTTCTACACCAGGTTCTTCGATTACATTG
clone6 AGCATCCTGTGGCCAAGTCTTACTCCTGGCTTCTACACCAGGTTCTTCGATTACATTG
-----

4. cDNA ACTGGATCCAGAGCATTATTGCGAGAAATACAGATGCGACCTGCCCCCATGA
clone6 ACTGGATCCAGAGCATTATTGCGAGAAATACAGATGCGACCTGCCCCCATGA
-----
```

**5'UTR region of all the clones:**

>1  
ATCAACGTAGAGTACGCGGGGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>2  
GCGGGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCTGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>3  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>4  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>5  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>6  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>7  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>8  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>9  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>10  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>11  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>12  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>13  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>14  
ATCAACGTAGAGTACGCGGGGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>15  
GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT  
>16  
GCGGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT

**5'UTR region Identified:**

**GCGGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGCTGCGCAGAGTTGAAGCT**

**Isoform 5**

193 bp 5'UTR sequence was obtained for Isoform 5 with the of multiple sequence alignment of 20clones together. And similarly the sequence showing the maximum similarity with other clones was taken as a representative clone for identifying 5'UTR for Isoform 5.

**Complete Sequence:**

>4  
GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG  
GTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCTATGGTGTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTTCTTACGGACAAAAGTCTTCTGAAGTGGTTCATTGGAGGTGAT  
GAATGTAAACATAAATGAACATCGTTCCCTTGCACCTGTATATCACTAACGGTTTTCTCTGCGGTGGGACTTTGATCAACCAGGAATGGGTGCTCACTGCTGCACACTGCGACAGGGGAAATGCTCATATCTCTCGG  
TGTGCATAGACTAAAGGGACTAAATAAGGATGCGCAGACAAGAGTCGCAAAGGAGAAGTTTCATTTGTCCTCAATAGGAAAAAATGATGAAAAGGACAAGGACATCATGTTGATCAGGCTGGACAGTCTGTAAACA  
GCAGTACACACAT

Multiple Sequence Alignment with Isoform 5 cDNA:  
 CLUSTAL W (1.81) multiple sequence alignment

```

5. cDNA -----
clone4 GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAA

5. cDNA -----
clone4 TTTGATCAAATAAAGTGTGCTTGATCAAGAACTCTCCGCTTGGGTTATCTGGTTAGATT

5. cDNA -----
clone4 GATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGA

5. cDNA -----ATGGTCTGATCAGAGTCTAGCAAACCTTCTGATACTACAGCTTTC
clone4 CAGAGTTGAAGCTATGGTCTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTTTC
*****

5. cDNA TTACGGACAAAAGTCTTCTGAACTGGTCATTGGAGTGATGAATGTAACATAAATGAACA
clone4 TTACGGACAAAAGTCTTCTGAACTGGTCATTGGAGTGATGAATGTAACATAAATGAACA
*****

5. cDNA TCGTTCCCTTGCACCTCGTATATCACTAACGGTTTTCTCTCGGTGGGACTTTGATCAA
clone4 TCGTTCCCTTGCACCTCGTATATCACTAACGGTTTTCTCTCGGTGGGACTTTGATCAA
*****

5. cDNA CCAGGAATGGGTGCTCACTGCTGCACACTGCGACAGGGGAAATATGCTCATATTCTTCGG
clone4 CCAGGAATGGGTGCTCACTGCTGCACACTGCGACAGGGGAAATATGCTCATATTCTTCGG
*****

5. cDNA TGTGCATAGACTAAAGGGACTAAATAAGGATGCGCAGACAAGAGTCGCAAAGGAGAAGTT
clone4 TGTGCATAGACTAAAGGGACTAAATAAGGATGCGCAGACAAGAGTCGCAAAGGAGAAGTT
*****

5. cDNA CATTGTCCCAATAGGAAAAAATGATGAAAAGGACAAGGACATCATGTTGATCAGGCT
clone4 CATTGTCCCAATAGGAAAAAATGATGAAAAGGACAAGGACATCATGTTGATCAGGCT
*****

5. cDNA GGACAGTCTGTAAACAGCAGTACACACATCGCGCCTATCAGCTTGCCTTCCAACCCCTCC
clone4 GGACAGTCTGTAAACAGCAGTACACACAT-----
*****

5. cDNA CAGTGTGGGCTCAGTTTGCATATTATGGGATGGGCGCAATCACATCTCCTAATGTGAC
clone4 -----

5. cDNA TTTACCCGGTGTCCCTCATTGTGCTAACATTAACATACTCGATCATGAGGTGTCAAGC
clone4 -----

5. cDNA AGCTTATCCATCATTGACCGCACAAGCAGAACATTGTGTGACAGGTATCCTGGAAGGAGG
clone4 -----

5. cDNA CAAAAGTTTATGTGACGGTGACTCTGGGGACCCCTCATCTGTAATGAGAGAAATCCAGGG
clone4 -----
  
```

5'UTR Regions Of all clones:

```

>clone1
GACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGGATCTGGTTAGATTGA
TACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone2
GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGTATCTGGTTAGATT
GATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone3
GACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGGATCTGGTTAGATTGA
TACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone4
GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGTATCTGGTTAGATTGATACAGTATCTCAATTTAGG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone5
GGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGGATCTGNTNNATTG
ATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone6
GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGTATCTGGTTAGATTGATACAGTATCTCAATTTAGG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone7
TGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGTATCTGGTTAGATTGATACATATCTCAATTTAGGTTTGGGACTGGGATCTTACAGG
TAAACAGTTTGTGCGACAGAGTTGAAGCT
>clone8
GGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGGATCTGGTTAGATTGATACAGTATCTCAATTTAGG
TTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone9
GGGACACTTCTGGATGTAACCTGTCTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGTCTTGTGATCAAGAAGTCTCCGCTTGGTATCTGGTTAGATTGATACAGTATCTCAATTTAGG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone10
  
```

```

GCGGGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCGCTTGGGTTATCTGGTTAGATTGATACAGTATCTCAA
TTTTAGTTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone11
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone12
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone13
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone14
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone15
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT
>clone16
GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAG
TTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGACAGAGTTGAAGCT

```

CLUSTAL W (1.81) multiple sequence alignment

```

clone1      -----GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone9      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone13     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone16     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone6      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone10     GCGGGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone11     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone14     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone4      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone7      -----TGTTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone12     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone15     -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone2      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone3      -----GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone5      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
clone8      -----GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTT
*****
clone8      TAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTT

```

5' UTR region identified:

GGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTG  
ATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAAC  
AGTTTGTGCGACAGAGTTGAAGCT

## Isoform 6

**Identification Of exon 1 region in Isoform 6:** Before identifying the 5'UTR region for isoform 6, to identify the exon 1 region was 1<sup>st</sup> task. With the help of 5' RACE technique incomplete sequence for the template cDNA was obtained. And then the isoform 6 template cDNA molecule was aligned with all the isoforms. Its sequence was found to be similar with isoform 10 and thus exon 1 region was predicted.

After this identification of 5'UTR was done as in previous isoforms. 20 clones showed the maximum similarity with each other thus 96 bp sequence for isoform 6 was obtained as 5'UTR.

### Isoform 6 template cDNA molecule:

```
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGC
AGAGTTGAAGCTATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCTTTCTTATGCACAAAAGCTTCTGAACTG
GTCATTGGAGGTGATGAATGTAACATAAATGAACATCGTTTCCTTGATAGCCTTGTATCACTCTAGGTCTAAGACTTTTCTCTGCG
GTGGGACTTTGCTCAACGAGGAATGGGTTCTACTGCTGCACACTGCAACAGGGTATTTTTGTACATCAGGCTTGGTATGCATA
ACAAAAATGTAATTTTGTACGATGACCAGAGAAGATAACCAAGAAGAAGTACTTTTTTTCGCTGTACAATAACTTTACCAAAC
GGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGTTACAGTGAACACATTGCGCCTCTCAGCTTGCCTTCCAGCC
CTCCAGTGTGGGCTCAGTTTGCCGTGCAATGGGATGGGGCCAAACCACATCTCCTCAAGAGACTTTGCCCGATGTCCCTCATT
GTGCTAACATTAACCTACTCGATTATGAAGTGTGTCGAGCAGCTCACCCACAATTTGGGTTGCCAGCGACAAGCAGAATATTG
TGTGCAGGTGTCCTGGAAGGAGGCATAGATACATGTCATCGTACTCTGGGGGACCCCTCATCTGTAATGGGCAATTCCAGGG
CATTGTATCTTGGGGAGATGGTTCTTGTCGCCAACCGGATAAGCCTGCCCTCTACAGCAAGGTCTTTGATCATCTTGACTGGAT
CCAGAACATTATTGCAGGAAATACAACCTGTGAATTGCCCCCGTGA
```

Pink region: exon1, Green: 5'UTR.

5'UTR:

```
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGC
AGAGTTGAAGCT
```

### Complete Sequence:

>2

```
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCTATGGTGCTGATCAGAGTGCTAGCAAACCTTCTGATACTACAGCT
TTCTTATGCACAAAAGCTTCTGAACTGGTCAATTGGAGGTGATGAATGTAACATAAATGAACATCGTTTCCTTGATAGCCTTGTATCACTCTAGGTCTAAGACTTTTCTGCGGTGGGACTTTGCTCAACGAGGAATGGGT
TCTCACTGCTGCACACTGCAACAGGGTATTTTTGTACATCAGGCTTGGTATGCATAACAAAATGTAATTTTACAGTACAGGAGAAAGATACCAAGAAGAAGTACTTTTTTCGCTGTACAATAACTTTACCAAAC
GGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGTTACAGTGAACACATTGCGCCTCTCAGCTTGCCTTCCAGCCCTCCAGCTTGCCTTCCAGCCCTCCAGCTTGCCTTCCAGCCCTCCAGCTTGCCTTCCAGCC
TCAAGAGACTTTGCCCGATGTCCTCATTGTGCTAACATTAACCTACTCGATTATGAAGTGTGTCGAGCAGCTCACCCACAATTTGGGTTGCCAGCGACAAGCAGAATATTGTGTCAGGTGTCCTGGAAGGAGGCATA
GATAC
```

Multiple Sequence Alignment with Isoform 6 cDNA:

CLUSTAL 2.1 multiple sequence alignment

```
6. cDNA
clone2      -----
            GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTA

6. cDNA
clone2      -----
            AATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCTATGGTGCTGATCAGAGTGCTAGCA

6. cDNA
clone2      -----TCTTCTGAACTGGTCAATTGGAGGT
            AACCTTCTGATACTACAGCTTCTTATGCACAAAAGCTTCTGAACTGGTCAATTGGAGGT
            *****

6. cDNA
clone2      GATGAATGTAACATAAATGAACATCGTTTCCTTGATAGCCTTGTATCACTCTAGGTCTAAG
            GATGAATGTAACATAAATGAACATCGTTTCCTTGATAGCCTTGTATCACTCTAGGTCTAAG
            *****
```

```

6. cDNA      ACTTTTCTCTGCGGTGGGACTTTGCTCAACGAGGAATGGGTTCTCACTGCTGCACACTGC
clone2      ACTTTTCTCTGCGGTGGGACTTTGCTCAACGAGGAATGGGTTCTCACTGCTGCACACTGC
*****

6. cDNA      AACAGGGTATTTTGTACATCAGGCTTGGTATGCATAACAAAAATGTAATTTTGACGAT
clone2      AACAGGGTATTTTGTACATCAGGCTTGGTATGCATAACAAAAATGTAATTTTGACGAT
*****

6. cDNA      GACCAGAGAAGATACCCAAGAAGAAGTACTTTTTCGCTGTCAACAATACTTTACCAA
clone2      GACCAGAGAAGATACCCAAGAAGAAGTACTTTTTCGCTGTCAACAATACTTTACCAA
*****

6. cDNA      CGGGACAAGGACATCATGTTGATCAGGCTGAACAACTGTTAGTTACAGTGAACACATT

```

The Yellow region is 5'UTR.

**5'UTR region of all 8 clones:**

```

>1
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>2
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>3
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>4
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>5
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>6
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>7
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
>8
GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT

```

5'UTR region identified:

**GGTGACACTTCTGGACGTAACCTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT**

**Isoform 7**

125 bp sequence of 5'UTR was obtained for isoform 7 by aligning the 20 clones together.

**Complete Sequence:**

```

>1
GGGGGGACACTTCTGGATGTGACTTCCAGCGTTCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTGGCTTATCTGATTAATGATATGATATCTCAAGTTA
AAGCAAACACTTCTGCACAGAATTGAAGCTATGGCGTGTACAGCTGTAGTAAACCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTCTGAACCTGGTCATTGGAGGTGATGAATGTAACGTAATGAACAT
CGTTTCTTGCACTTTGATTTCTGCAAGTTTCAATGCGGTGGGACTTTGATCAATGAAGAATGGGTGCTCACTGTCACACTGCAACAGAAGACACATGTACATATTCTTGGTGTGCATAACATAAGCGTAAAAAT
GATGATGAGCAGAGAAGATCCCAAGAAGAAGTACTTTTGCCTGAGTAGCAGAAACTATACCAAAGGGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTAGGAAAAGTGAACACATCGCGCTCTCAGC
TTGCCTCCAGTCTCCAGTGTGGGCTCAGTTTCCGCTGTTATGGGATGGGCAACAATCACATCTCTAATGTATGTTGCCGATGTCCTTNTGTGCTAACATTAACATACTTGATTATGAGGTGTGTCAAGCAGCT
TACGGAAA

```

**Multiple Sequence Alignment with Isoform 7 cDNA:**

CLUSTAL 2.1 multiple sequence alignment

```

7. cDNA      -----
1. cDNA      GGGGGACACTTCTGGATGTGACTTCCAGCGTTCAGATTGTTGGTCATCAGCTGCTT

7. cDNA      -----
1. cDNA      AATTTGATCAAAATAAAGTGTGCTTGTGATCAAGAAGTCTCCACTTGGCTTATCTGATTA

7. cDNA      ----ATGGTATCTCAAGTTAAAGCAAACAACCTTGCTGCACAGAATTGAAGCTATGGCGC
1. cDNA      TTGATATGGTATCTCAAGTTAAAGCAAACAACCTTGCTGCACAGAATTGAAGCTATGGCGC
*****

7. cDNA      TGATCAGACTGCTAGTAAACCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTTCTG
1. cDNA      TGATCAGACTGCTAGTAAACCTTCTGATACTACAGCTTTCTTACGCACAAAAGTCTTCTG
*****

```

```

7 . cDNA      AACTGGTCATTGGAGGTGATGAATGTAACGTAATGAACATCGTTTCCTTGCACTCTTGT
1 . cDNA      AACTGGTCATTGGAGGTGATGAATGTAACGTAATGAACATCGTTTCCTTGCACTCTTGT
*****

7 . cDNA      ATTCTGACAAGTTTCAATGCGGTGGGACTTTGATCAATGAAGAATGGGTGCTCACTGCTG
1 . cDNA      ATTCTGACAAGTTTCAATGCGGTGGGACTTTGATCAATGAAGAATGGGTGCTCACTGCTG
*****

7 . cDNA      CACACTGCAACAGAAGACACATGTACATATTCCTTGGTGTGCATAACATAAGCGTAAAT
1 . cDNA      CACACTGCAACAGAAGACACATGTACATATTCCTTGGTGTGCATAACATAAGCGTAAAT
*****

7 . cDNA      ATGATGATGAGCAGAGAAGATTCCCAAAGAAGAAGTACTTTTGCCTGAGTAGCAGAACT
1 . cDNA      ATGATGATGAGCAGAGAAGATTCCCAAAGAAGAAGTACTTTTGCCTGAGTAGCAGAACT
*****

7 . cDNA      ATACCAAAGGGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGGAAAAGTG
1 . cDNA      ATACCAAAGGGGACAAGGACATCATGTTGATCAGGCTGAACAAACCTGTTAGGAAAAGTG
*****

7 . cDNA      AACACATCGGCCTCTCAGCTTGCCTTCCAGTCTCCAGTGTGGGCTCAGTTTGCCGTG
1 . cDNA      AACACATCGGCCTCTCAGCTTGCCTTCCAGTCTCCAGTGTGGGCTCAGTTTGCCGTG
*****

7 . cDNA      TTATGGGATGGGGACAATCACATCTCCTAATGTGATGTTGCCCGATGTCCTTACTGTG
1 . cDNA      TTATGGGATGGGGACAATCACATCTCCTAATGTGATGTTGCCCGATGTCCTTACTGTG
*****

7 . cDNA      CTAACATTAACATACTTGATTATGAGGTGTGTCAAGCAGCTTACGGAGGATGCCAGCAA
1 . cDNA      CTAACATTAACATACTTGATTATGAGGTGTGTCAAGCAGCTTACGGAAA-----
*****

7 . cDNA      CAAGCAGAACATTGTGTGAGGTATCCTGGAAGGAGCAAGATTTCATGTGACGGTGACT
1 . cDNA      -----

7 . cDNA      CTGGGGACCCCTCATCTGTAATGGACAATCCAGGGCATTGTATCTTGGGGAGCCGATA
1 . cDNA      -----

7 . cDNA      CTTGTGCCAACCCGCTGCGCTGGCTTCTACCAATGTCTTTGATTATACTGACTGGA
1 . cDNA      -----

7 . cDNA      TCCAAAGCATTATTTCCAGAAATACAGATGCGACCTGCCCTCCGTGA
1 . cDNA      -----

```

5'UTR regions of all clones:

```

>1
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>2
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>3
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>4
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>5
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>6
GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>7
GACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>8
GGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>9
GGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
>10
TTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
5'UTR region identified:

```

**GGGGGGACACTTCTGGATGTGACTCTGCCAGCGTCCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT**

# Isoform 8

139 bp sequence of 5'UTR was obtained for isoform 8 by aligning the 20 clones together.

## Complete Sequence:

>1

```
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTAGGTTGGGACTGGGATCTTACAGGTAAACAGTTTGTGCACAG
AGTTGAAGCTATGGTGTGATCAGAGTCTAGCAACCCCTCTGATACTACAGCTTTCTTACGCACAAAAGTCTTCTGAACCTGGTCTGGAGGTCGTCCATGTAACATAAATGAACATCGTTCCTTGT
TGCTTTAATTAACCTAGCGGTTTCTCTGCGCTGGGACTTTGATCAACCAGGAATGGGTGCTCAGTCTGCACACTGGGACATTAATAATTTCCAGATACAGCTTGGTGTGCATAGCAAAAAGGTACG
AAATGAGGATGAGCAGACAAGACCCAAAGGAGAAGTTCTTTTGTCTCGGTAGCAAAACCAATAACGAATGGGAGAAGGACATCATGTTGATCAGGCTGAACAATCCATTGCGCCTCTCAGCTTGCC
TCTAGCCCTCCAGTGTGGGCTCACTTTGCGGTATTATGGGATGGGGCACAATCTCATCTACTAAAGAAATTTATCCCAATGCCCTCGTTGTGCTAAACATACTCGATTATGCAAGTGTGCCGA
GCAGCATACCCATGGTGGCCGTGACAACCAAGAAATATTGTGTGCAAGTATTCTGGAAGGAGGCAAAAGATTTCATGTCAAGGTGACTCTGGGGACCCC
```

Multiple Sequence Alignment with Isoform 8 cDNA:

clustalw.aln

CLUSTAL W (1.81) multiple sequence alignment

```
8. cDNA
clone1 -----
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGT

8. cDNA
clone1 -----
TAGATTGATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAAACAGTTT

8. cDNA
clone1 -----ATGGTGCTGATCAGAGTCTAGCAACCCCTCTGATACTACA
GTCGCACAGAGTTGAAGCTATGGTGTGATCAGAGTCTAGCAACCCCTCTGATACTACA
*****

8. cDNA
clone1 GCITTCTTACGCACAAAAGTCTTCTGAACCTGGTCTGGAGGTCGTCCATGTAACATAAA
GCTTTCTTACGCACAAAAGTCTTCTGAACCTGGTCTGGAGGTCGTCCATGTAACATAAA
*****

8. cDNA
clone1 TGAACATCGTTCCTTGTGTCTTATTAACTCTAGCGGTTTCTCTGCGCTGGGACTTT
TGAACATCGTTCCTTGTGTCTTATTAACTCTAGCGGTTTCTCTGCGCTGGGACTTT
*****

8. cDNA
clone1 GATCAACCAGGAATGGGTGCTCACTGCTGCACACTGCGACATTAATAATTTCCAGATACA
GATCAACCAGGAATGGGTGCTCACTGCTGCACACTGCGACATTAATAATTTCCAGATACA
*****

8. cDNA
clone1 GCTTGGTGTGCATAGCAAAAAGGTACGAAATGAGGATGAGCAGACAAGAGACCCAAGGGA
GCTTGGTGTGCATAGCAAAAAGGTACGAAATGAGGATGAGCAGACAAGAGACCCAAGGGA
*****

8. cDNA
clone1 GAAGTCTTTTGTCTCGGTAGCAAAACCAATAACGAATGGGAGAAGGACATCATGTTGAT
GAAGTCTTTTGTCTCGGTAGCAAAACCAATAACGAATGGGAGAAGGACATCATGTTGAT
*****

8. cDNA
clone1 CAGGCTGAACAATCCATTGCGCCTCTCAGCTTGCCTTCTAGCCCTCCCAGTGTGGGCTCA
CAGGCTGAACAATCCATTGCGCCTCTCAGCTTGCCTTCTAGCCCTCCCAGTGTGGGCTCA
*****

8. cDNA
clone1 CTTTGGCGTATTATGGGATGGGGCACAATCTCATCTACTAAAGAAATTTATCCCAATGTC
CTTTGGCGTATTATGGGATGGGGCACAATCTCATCTACTAAAGAAATTTATCCCAATGTC
*****

8. cDNA
clone1 CCTCGTTGTGCTAACATTAACATACTCGATTATGCAGTGTGCCGAGCAGCATACCCATGG
CCTCGTTGTGCTAACATTAACATACTCGATTATGCAGTGTGCCGAGCAGCATACCCATGG
*****

8. cDNA
clone1 TGCCCGGTGACAACCAAGAAATATTGTGTGCAAGTATTCTGGAAGGAGGCAAGATTTCATGT
TGCCCGGTGACAACCAAGAAATATTGTGTGCAAGTATTCTGGAAGGAGGCAAGATTTCATGT
*****

8. cDNA
clone1 CAAGGTGACTCTGGGGACCCCTCATCTGTAATGGAGAAATCCAGGTATTGTATCTTGG
CAAGGTGACTCTGGGGACCCC-----
*****

8. cDNA
clone1 -----
GGGGCGCATCCCTTGTGCGGACGTCTTAATCCTGGCTTCTACCAAGGTCTTTGATTAC
-----
```

**5' UTR region of all the clones:**

>1  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>2  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>3  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>4  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>5  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>6  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>7  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT  
>8  
GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT

**5'UTR region identified:**

**GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT**

**Isoform 9**

195 bp sequence upstream to ATG was obtained as 5'UTR for Isoform 9 after aligning the 40 clones together.

**Complete Sequence:**

>1  
GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCTATGTTCTAATCAGAGTGTGCAACCTTCTGACTGCAGCTTCTACGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGAAGAATGTATCATAAATGAACATCGTTTCTTGTAGCCTGTACTCCTTAAATCTAAGAGGTTTCACTGCAGTGGGACTTTGATCAACCAGGAATGGGTGCTGACTGTGCACATGCAACAGGAAAAATATCCGGATAAAGCTTGGTACGCATAGCACAAAGGTAACAAATGAGGATGTGCAGACAAGAGTCCCAAAGAAGAAGTTCTTTTGTCTCAGTAGCAAAACCTACACAGATGGGACAAAGGACATCATGTTGATCAGGCTGAAAAGACCTGTTAACAACAGTGAACACATCGTGCCCTGTGAACCTGCTTCCAACCTCCAGTCTGGGCTCAGTTTGCCATATTATGGGATGGGGCCAAACACATCTCTAATGTACTTACCCTGCTCCTCATTTGTGTAACATACTACTGATCATGAGGTGTGCAAGCAG

**Multiple Sequence Alignment with Isoform 9 cDNA:**

CLUSTAL 2.1 multiple sequence alignment

```
9. cDNA -----
clone1      GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT

9. cDNA -----
clone1      AATTTGATCAAAATAAAGTGTCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGA

9. cDNA -----
clone1      TTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCG

9. cDNA -----ATGGTTCTAATCAGAGTGTGCAACCTTCTGATACTGCAGCTT
clone1      CACAGAGTTGAAGCTATGGTTCTAATCAGAGTGTGCAACCTTCTGATACTGCAGCTT
*****

9. cDNA TCTTACGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGAAGAATGTATCATAAATGAA
clone1 TCTTACGCACAAAAGTCTTCTGAACTGATCATTGGAGGTGAAGAATGTATCATAAATGAA
*****

9. cDNA CATCGTTTCCCTGTAGCCTTGTACTCCTTTAAATCTAAGAGGTTTCACTGCAGTGGGACT
clone1 CATCGTTTCCCTGTAGCCTTGTACTCCTTTAAATCTAAGAGGTTTCACTGCAGTGGGACT
*****

9. cDNA TTGATCAACCAGGAATGGGTGCTGACTGCTGCACACTGCAACAGGAAAAATATCCGGATA
clone1 TTGATCAACCAGGAATGGGTGCTGACTGCTGCACACTGCAACAGGAAAAATATCCGGATA
*****

9. cDNA AAGCTTGGTACGCATAGCACAAAGGTAACAAATGAGGATGTGCAGACAAGAGTCCCAAAG
clone1 AAGCTTGGTACGCATAGCACAAAGGTAACAAATGAGGATGTGCAGACAAGAGTCCCAAAG
*****
```

```

9. cDNA      AAGAAGTCTCTTTGTCTCAGTAGCAAACCTACACCAGATGGGACAAGGACATCATGTTG
clone1      AAGAAGTCTCTTTGTCTCAGTAGCAAACCTACACCAGATGGGACAAGGACATCATGTTG
*****

9. cDNA      ATCAGGCTGAAAAGACCTGTTAACAACAGTGAACACATCGTGCCTGTGAACTTGCCTTCC
clone1      ATCAGGCTGAAAAGACCTGTTAACAACAGTGAACACATCGTGCCTGTGAACTTGCCTTCC
*****

9. cDNA      AACCTCCAGTCTGGGCTCAGTTTGCCATATTATGGGATGGGGCGCAACCACATCTCCT
clone1      AACCTCCAGTCTGGGCTCAGTTTGCCATATTATGGGATGGGGCGCAACCACATCTCCT
*****

9. cDNA      AATGTGACTTTACCCGGTGTCCCTCATTGTGCTAACATTAACATACTCGATCATGAGGTG
clone1      AATGTGACTTTACCCGGTGTCCCTCATTGTGCTAACATTAACATACTCGATCATGAGGTG
*****

9. cDNA      TGTC AAGCAGCTTATCCATCATTGACCCGACAGCAGAACATTGTGTGCAAGTATCCCTG
clone1      TGTC AAGCAG-----
*****

9. cDNA      GAAGGAGGCAAAAGTTCATGTGACGGTGACTCTGGGGACCCCTCATCTGTAATGAGAGAA
clone1      -----

9. cDNA      TTCCAGGGCATTGTATTTGGGGACCCGATACTTGTGCCAACCCGATAAGCCTAGCCTC
clone1      -----

9. cDNA      TACACCAAGGCTTTGATCATCTTGACTGGATCCAGAGCATTATTGCAGGAAATAAAACT
clone1      -----

9. cDNA      GTGAATTGCCCCCGTGA
clone1      -----

```

**5'UTR region of all the clones:**

>1

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>2

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>3

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>4

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>5

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>6

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAG  
TATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAAGCT

>7

GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATC  
TCAATTTTAGGTTTGGGACTGGGAT

>8

GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATC  
TCAATTTTAGGTTTGGGACTGGGAT

>9

GACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATC  
TCAATTTTAGGTTTGGGACTGGGAT

**5'UTR region identified:**

GGGGGACACTTCTGGATGTAACCTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATT  
TGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACA  
GTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTCCGACAGAGTTGAA  
GCT

# Isoform 10

121 bp sequence was found to be upstream to ATG after aligning the 20 clones together.

## Complete Sequence:

>1

```
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCTATGGTGTGATCAGAGTGC
TAGCAAACCTTCTGATACTACAGCTTCTTATGCACAAAAGTCTTGAAGTATCATTGGAGGTGAAGAATGAACATAAATGAACATCGTTTCCTGTAGCCTTGTACTACTAGGCTAAGACTTTTCTCTGCGGTGG
GACTTTGCTCAACGAGGAATGGGTGCTCACTGCTGCACACTGCAACAGGGTATTTATGTACATCAGGCTTGGTATGCATAACAAAATGTAATTTGACGATGACCATATAAGATACGAAAGGAGAAGACTTTTT
TCGCTGTCGAATAACTTTACAAATGGGACAAGGACATCATGTTGATCAGGCTGAACAACCTGTTAGTTACAGTGAATACATCGCCCTCAGCTTGCCTTCCAGCCCTCCAGTGTGGCTCAGTTTCCGCTGTTAT
GGGATGGGGCAAACCATCTCTCAAGAGACTTTGCC CGATGTCCCTATTGTGCTAACATTAACCTACTTGATTATGAAGTGTGTCGAGCAGCT
```

CLUSTAL W (1.81) multiple sequence alignment

```
10.cDNA
clone3      -----
             GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCT

10.cDNA
clone3      -----
             CAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGC

10.cDNA
clone3      -ATGGTGTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTTTCCTTATGCACAAAA
             TATGGTGTGATCAGAGTGTAGCAAACCTTCTGATACTACAGCTTTCCTTATGCACAAAA
             *****

10.cDNA
clone3      GTCTTCTGAACTGATCATTGGAGGTGAAGAATGTAACATAAATGAACATCGTTTCCTTGT
             GTCTTCTGAACTGATCATTGGAGGTGAAGAATGTAACATAAATGAACATCGTTTCCTTGT
             *****

10.cDNA
clone3      AGCCTTGTACTCACTCTAGGTCTAAGACTTTTCTCTGCGGTGGGACTTTGCTCAACGAGGA
             AGCCTTGTACTCACTCTAGGTCTAAGACTTTTCTCTGCGGTGGGACTTTGCTCAACGAGGA
             *****

10.cDNA
clone3      ATGGGTGCTCACTGCTGCACACTGCAACAGGGTATTTATGTACATCAGGCTTGGTATGCA
             ATGGGTGCTCACTGCTGCACACTGCAACAGGGTATTTATGTACATCAGGCTTGGTATGCA
             *****
```

## 5'UTR region of all the clones:

>1

```
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>2
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>3
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>4
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>5
GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>6
ACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>7
GTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>8
GTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>9
GTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
>10
ACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT
```

## 5' UTR region identified for Isoform 10:

**GGTGACACTTCTGGACATAACTCTGCCAAGTTCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAG  
TAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGACAGATTGAAGCT**

## Isoform 11

115 bp sequence was identified as 5' UTR sequence after aligning the 20 clones together.

### Complete Sequence:

>clone 5

```
GGGCACTTCTGGACGTGACTCTCCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGTTGATCAAGAAGTCCCGCTGGGTTATCTGATTAATGGTCTGATCAGAGTGTAGCAA
ACCTTCTGATACTACAGCTTCTTACGCACAAAAGTCTTCTGAACCTGGTCTGGAGGTGATGAATGTAACATAAAATGAACATCGTTCCTTGTACTCGTGTACTCTGACGGGATTCAATGTGGTGGGACTTGTATCAACC
AGGAATGGATGCTCAGCGTGCACACTGTGACCGGAAAAGAAATGAAGTTACAGTTTGGCTGTCATAGCAAAAATGTACCAAAGGATAAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTACAAAA
ACAATAAAGAATGGGACAAGGACATCATGTTGATCAGGCTGAACCCGCTCTGTTAAACAACAGTAAACACATTGCGCTCTCAGCTTCCCTTCAAGCCTCCAGTCAGGACACAGTTTGTCAATATTATGGGATGGGGCAC
AATCTCACCTACTAAAGAGATTTATCTGATGCTCCCTCATTGTGCTAACATTAATACTCGATCAGGCGGTTTGTGAGAAAATTTACCCAGGGTTGCTGGAGAAAAGCAGAGATTGTGTGCGAGTATCTGGAAGGAG
GCAAAGATACA
```

CLUSTAL 2.1 multiple sequence alignment

```
11.cDNA -----
clone5  GGGCACTTCTGGACGTGACTCTCCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAAT

11.cDNA -----ATGGT
clone5  TTGATCAAATAAAGTCTGCTTGTATCAAGAAGTCTCCGCTTGGGTTATCTGATTAATGGT
      *****

11.cDNA  GCTGATCAGAGTCTAGCAAACCTTCTGATACTACAGCTTCTTACGCACAAAAGTCTTC
clone5  GCTGATCAGAGTCTAGCAAACCTTCTGATACTACAGCTTCTTACGCACAAAAGTCTTC
      *****

11.cDNA  TGAAGTGGTCTGGAGGTGATGAATGTAACATAAATGAACATCGTTCCTTGTACTCGT
clone5  TGAAGTGGTCTGGAGGTGATGAATGTAACATAAATGAACATCGTTCCTTGTACTCGT
      *****

11.cDNA  GTACTCTGACGGGATCAATGTGGTGGGACTTTGATCAACAGGAATGGATGCTCACGGC
clone5  GTACTCTGACGGGATCAATGTGGTGGGACTTTGATCAACAGGAATGGATGCTCACGGC
      *****

11.cDNA  TGCACACTGTGACGGGAAAAGAATGAAGTTACAGTTTGGCTGTCATAGCAAAAATGTACC
clone5  TGCACACTGTGACGGGAAAAGAATGAAGTTACAGTTTGGCTGTCATAGCAAAAATGTACC
      *****

11.cDNA  AAACAAGGATAAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTAGCAAAAA
clone5  AAACAAGGATAAGCAGACAAGAGTCCCAAAGGAGAAGTCTTTTGTCTTAGTAGCAAAAA
      *****

11.cDNA  CAATAAAGAATGGGACAAGGACATCATGTTGATCAGGCTGAACCGTCTGTTAACACAG
clone5  CAATAAAGAATGGGACAAGGACATCATGTTGATCAGGCTGAACCGTCTGTTAACACAG
      *****

11.cDNA  TAAACACATTGCGCCTCTCAGCTTGCCTTCCAAGCCTCCAGTCAGGACACAGTTTGCAA
clone5  TAAACACATTGCGCCTCTCAGCTTGCCTTCCAAGCCTCCAGTCAGGACACAGTTTGCAA
      *****

11.cDNA  TATTATGGGATGGGGACAATCTCACCTACTAAAGAGATTTATCCTGATGTCCTCATTG
clone5  TATTATGGGATGGGGACAATCTCACCTACTAAAGAGATTTATCCTGATGTCCTCATTG
      *****

11.cDNA  TGCTAACATTAATATACTCGATCAGGCGGTTTGTGAGAAAATTTACCCAGGGTTGCTGGA
clone5  TGCTAACATTAATATACTCGATCAGGCGGTTTGTGAGAAAATTTACCCAGGGTTGCTGGA
      *****

11.cDNA  GAAAAGCAGAGTATTGTGTGAGGATCCTGGAAGGAGGCAAGATACATGTGGAGGTGA
clone5  GAAAAGCAGAGTATTGTGTGAGGATCCTGGAAGGAGGCAAGATACATGTGGAGGTGA
      *****

11.cDNA  CTCTGGGGACCTCTCATCTGTAATGGAGAAATCCAGGGCATTATCTGTGGGGGGCGA
clone5  -----

11.cDNA  TCCTTGTGCCAATCTCATGTGCTGCCCTCTACATCAAGTCTTTGATTATACTAGTGTG
clone5  -----

11.cDNA  GATCCAGAGCATTATTACAGGAAATACAGCTGCAACTTGCCCCCATGA
clone5  -----
```

**5'UTR region of all clones:**

>1  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>2  
GATCAAAGTCCCACCACATTGAATCCCGTCAGAGTACACGAGTACAAGGGAACGATGTTCAATTTATGT  
>3  
GGTGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTAGATTGACATGG  
>4  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>5  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>6  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>7  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>8  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA  
>9  
GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA

**5'UTR region identified:**

**GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA**

### (3) 5' Cap and Identification of Transcription Start Site in 5' UTR regions:

<b>Isoform 1: [195bp]</b> GGATGACACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGTTATCTGATTAGATTGATACGGTATCTCAAGTTTAAGTAGGGGACTAGGATCTAAAGGCCAAACAGCTTGCCGTGCAGAGTTGAAGCT
<b>Isoform 2: [141bp]</b> GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGACGTAACCTGTGCAGCGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
<b>Isoform 3: [194bp]</b> GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGTAGTCTCCGCTTGGTTATCTGATTAGACTGATACGGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCCAAACAGCTTTCCTTGACAGAGTTGAAGCT
<b>Isoform 4: [57bp]</b> GGGACGTTTGGGACTGGGATCTTACAGGCCAAACAGTTTGTGCGCAGAGTTGAAGCT
<b>Isoform 5: [193bp]</b> GGGACACTTCTGGATGTAACCTGTGCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT
<b>Isoform 6: [96bp]</b> GGTGACACTTCTGGACGTAACCTGTGCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGAGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
<b>Isoform 7: [125bp]</b> GGGGGACACTTCTGGATGTGACTCTGCCAGCGTTCAGATTGTTGGTCATCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAATTTGAT
<b>Isoform 8: [139bp]</b> GACACTTCTGGATGTAACCTGTGCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAGGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT
<b>Isoform 9: [195bp]</b> GGGGGACACTTCTGGATGTAACCTGTGCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGTTATCTGTTAGATTGATACAGTATCTCAATTTTAGGTTTGGGACTGGGATCTTACAGGTAACAGTTTGTGCGCACAGAGTTGAAGCT
<b>Isoform 10: [121bp]</b> GGTGACACTTCTGGACATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCGTATCTCAATTTAAGTAAGGGACTGGGATCTTACAGGCCAAACAGCTTTCC TTGACAGAGTTGAAGCT
<b>Isoform 11: [115bp]</b> GGGCACTTCTGGACGTGACTCTCCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAAGTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGATTA

**Pink region shows the 5' Cap and Cyan region shows the Transcription start site.**

The 5' *cap* is a modified guanine nucleotide added to the "front" (5' end) of the pre-mRNA using a 5'-5'-triphosphate linkage. This modification is critical for recognition and proper attachment of mRNA to the ribosome, as well as protection from 5' exonucleases. It may also be important for other essential processes, such as splicing and transport. Just after the G capping, there are 3 bases which designates the +1 or Transcription Start Site for the all respective isoforms.

#### (4). Multiple Sequence Alignment of all 5'UTR isoforms:

CLUSTAL W (1.81) multiple sequence alignment

```
1 -----GGATGACACTTCTGGA
7 -----GGGGGACACTTCTGGA
6 -----GGTGACACTTCTGGA
9 -----GGGGGACACTTCTGGA
2 GACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGGTGACACTTCTGGA
4 -----
5 -----GGGACACTTCTGGA
8 -----GACACTTCTGGA
10 -----GGTGACACTTCTGGA
3 -----GGTGACACTTCTGGA
```

```
1 CGTGTCTCTGCCAGTGTTCAGATTTTTGGCCATCCAGCTGCTTAATTTGATCAAATAAA
7 TGTGACTCTGCCAGCGTTCAGATTGTTGGTCATCAG-CTGCTTAATTTGATCAAATAAA
6 CGTAACTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTG-----
9 TGTAACTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAA
2 CGTAACTCTGTCAGCGTTCAGATTGTTGGCCACCCAGCTGCTTAATTTG-----
4 -----
5 TGTAACTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTAATTTGATCAAATAAA
8 TGTAACTCTGTCAGAGTTCGAGATTGTTGGCCACCCAGTTGCTTA-----
10 CATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAG-----
3 CATAACTCTGCCAACGTTCCAGATTGTTGGCCACCCAGCTGCTTAATTTGATCAAATAAA
```

```
1 GTGCTGCTTGATCAAGAAGTCTCCACTTGGGTTATCTGATTAGATTGATACGGTATCTCA
7 GTGCTGCTTGATCAAGAAGTCTCCACTTGGCTTATCTGATTAAATTGAT-----
6 -----
9 GTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATCTCA
2 -----
4 -----
5 GTGCTGCTTGATCAAGAAGTCTCCGCTTGGGTTATCTGGTTAGATTGATACAGTATCTCA
8 -----GGTTAGATTGATACAGTATCTCA
10 -----CGTATCTCA
3 GTGCTGCTTGATCAAGTAGTCTCCGCTTGGGTTATCTGATTAGACTGATACGGTATCTCA
```

```
1 AGTTTAAGTAGGGGACTAGGATCTTAAAGGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
7 -----
6 -----AGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
9 ATTTTAGGTTTGGGACTGGGATCTTACAGGTAAACAGTTTGTGCGCACAGAGTTGAAGCT
2 -----AGCAAACAGCTTGCCGTGCAGAGTTGAAGCT
4 --GGGACGTTTGGGACTGGGATCTTACAGGCAAACAGTTTGTGCGCACAGAGTTGAAGCT
5 ATTTTAGGTTTGGGACTGGGATCTTACAGGTAAACAGTTTGTGCGCACAGAGTTGAAGCT
8 ATTTTAGGTTTGGGACTGGGATCTTACAGGTAAACAGTTTGTGCGCACAGAGTTGAAGCT
10 ATTTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAGAGTTGAAGCT
3 ATTTTAAGTAAGGGACTGGGATCTTACAGGCAAACAGCTTTCCTTGCAGAGTTGAAGCT
```

- 1) Isoform 1, 3, 5, 9 shows the extensive similarity in the yellow region except very few bases quite near to 5'UTR region.

```
GACTTCTGGACGTGTCTCTGCCAGTGTCCAGATTTTGGCCATCCAGCTGCTTAATTTGATCAAATAA
AGTGCTGCTTGATCAAGAAGTCTCCAATTGGGTTATCTGATTAGATTGATACGGTATCTCAGCAAACAGC
TTGCCGTGCAGAGTTGAAGCT
```

- 2) All the isoform (except Isoform 4) show extensive similarity in the region near to 5' cap.

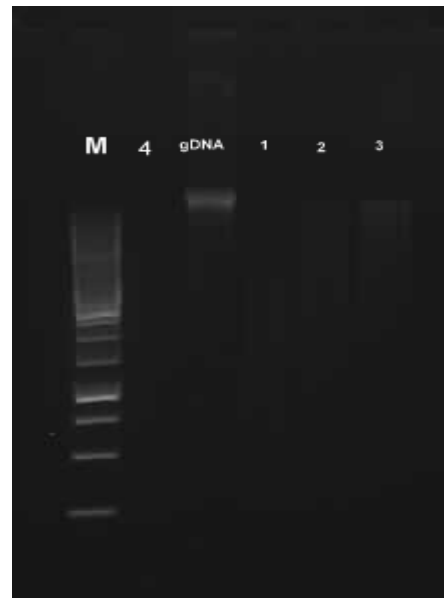
```
GACTTCTGGA
```

### Genome Walking:

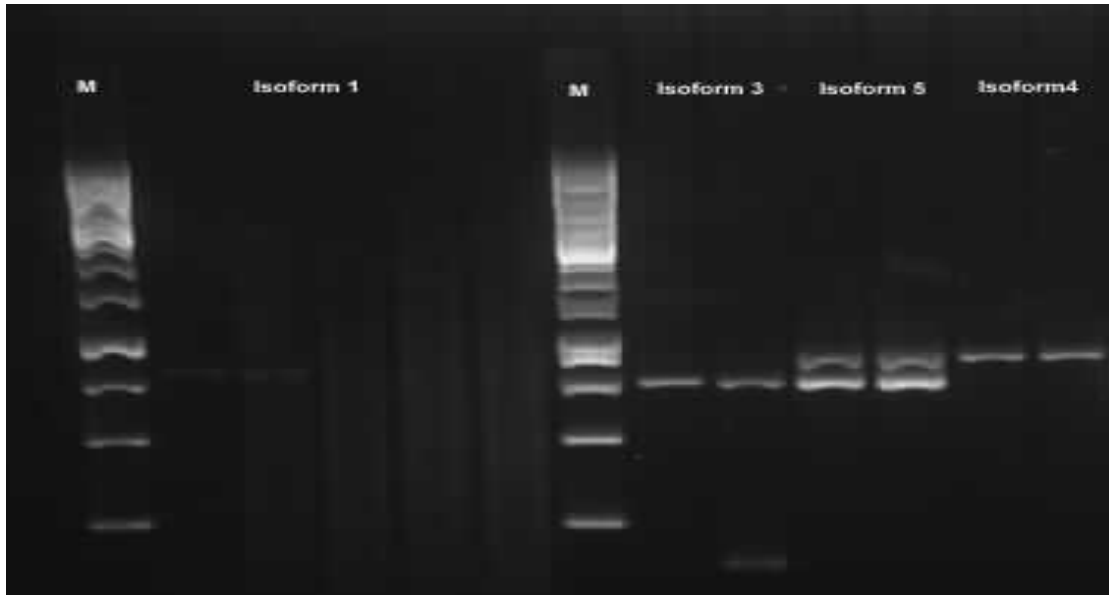
Genomic DNA was isolated from liver tissue and its quality was check by running it on gel. Its molecular weight is approx 7Kb. (Fig 5.11). Genomic DNA was digested with 4 different Restriction enzyme Dra1, EcoRV, PvuII, Stu 1 (Fig 5.12).



**Figure 5.11: isolation of genomic DNA**



**Fig 5.12: digested genomic DNA**



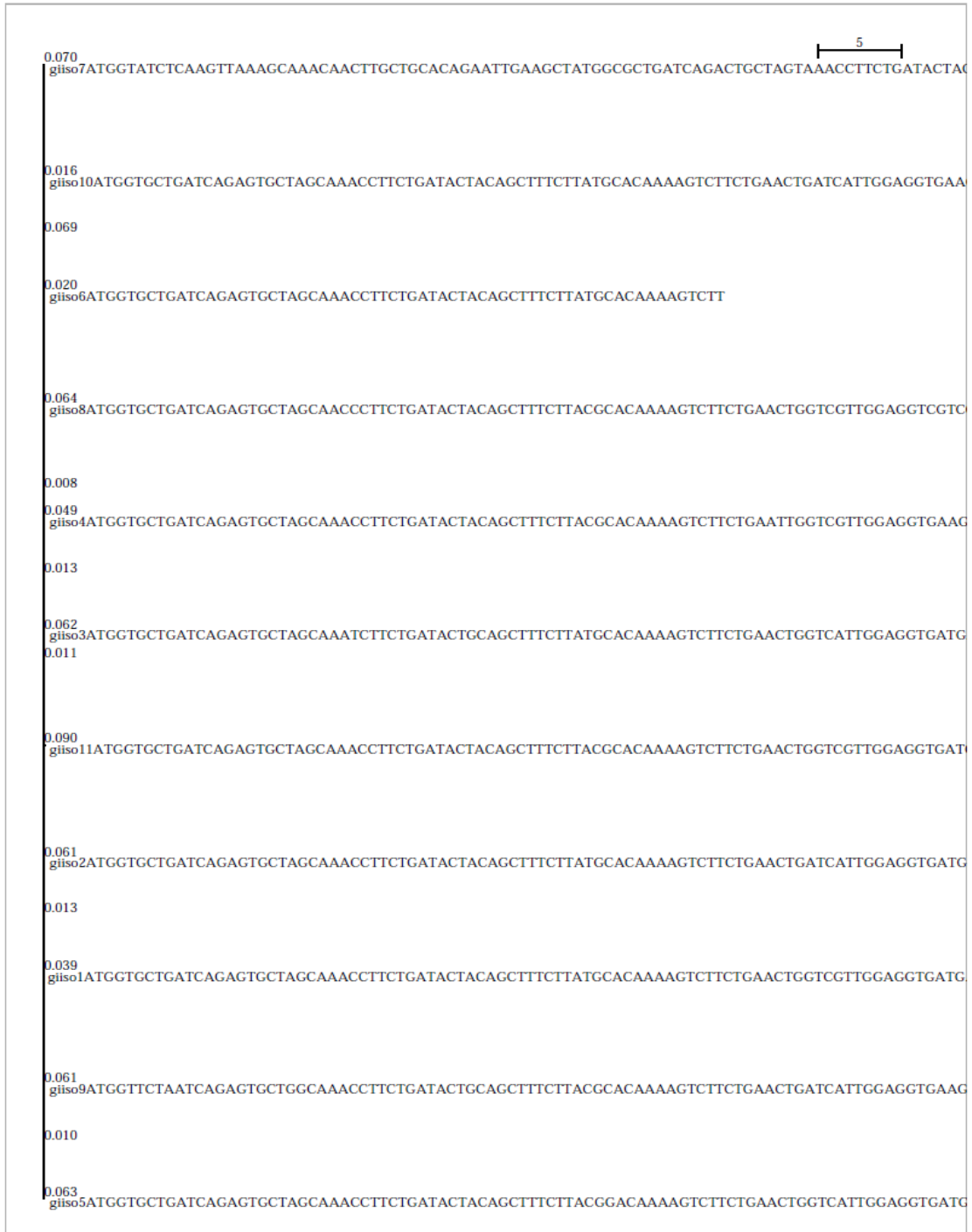
**Fig 5.13: Genome walker amplified products of Isoform 1,3,4 and 5. Isoform 1 didn't show any amplification at all**

Genomic DNA digested with all the 4 restriction enzyme shows the smearing confirming its complete digestion. After that Genomic DNA was diluted and used as a template for genome walking. Gene specific primers were designed from the 5' UTR of the above isoforms and from intron region of the above isoforms. Genome Walker PCR was set for the Isoform 1,3,4,5 altogether. And the specific amplification was got for Isoform 3, 4 and 5.

All the 3 amplified products were cloned in pJET vector and transformed colonies were checked by running the gel. All the insert released plasmids were taken to proceed the sequencing for identifying the promoter region.

## (5) Evolutionary Relationship of all Isoforms:

iso tree Thu May 26 15:24:39 2011 Page 1 of 1



## Discussion

It was interesting to understand the molecular mechanisms behind the differential expression of Serine Proteinase isoforms which shares the extensive similarity in the cDNA sequences. The above story is surely incomplete till the time we come to know about the promoter region revealing the differential expression of Serine Proteinase 11 isoforms and if could recognize the *cis/trans* elements which can give the region behind it. But it has become more interesting to know the mechanism behind this differential expression after identifying the 5'UTR region which again shows high similarity with other isoforms. All the isoform except isoform 7 shows the 100 % similarity in their 30 bp upstream to exon1. All the isoforms except Isoform 4 shows the good similarity of around 12 bases in their 5' region of 5' UTR region except few bases which are different and were used to make the primers for genome walking. Transcription Start site are different in all the 11 isoforms.

A messenger ribonucleic acid (mRNA) molecule codes for a protein through translation. The mRNA also contains regions that are not translated: in eukaryotes these include the **5' untranslated region**, 3' untranslated region, 5' cap and poly-A tail. Untranslated regions (UTRs) are sections of the mRNA before the start codon and after the stop codon that are not translated, termed the five prime untranslated region (5' UTR) and three prime untranslated region (3' UTR), respectively. These regions are transcribed with the coding region and thus are exonic as they are present in the mature mRNA. Several roles in gene expression have been attributed to the untranslated regions, including mRNA stability, mRNA localization, and translational efficiency. The ability of a UTR to perform these functions depends on the sequence of the UTR and can differ between mRNAs. The stability of mRNAs may be controlled by the 5' UTR and/or 3' UTR due to varying affinity for RNA degrading enzymes called ribonucleases and for ancillary proteins that can promote or inhibit RNA degradation. Translational efficiency, including sometimes the complete inhibition of translation, can be controlled by UTRs. Proteins that bind to either the 3' or 5' UTR may affect translation by influencing the ribosome's ability to bind to the mRNA.

The **five prime untranslated region (5' UTR)**, can contain elements for controlling gene expression by way of regulatory elements. It begins at the transcription start site and ends one

nucleotide (nt) before the start codon (usually ATG) of the coding region. In prokaryotes, the 5' UTR usually contains a ribosome binding site (RBS), also known as the Shine Dalgarno sequence (AGGAGGU). The 5' UTR has a median length of ~150 nt in eukaryotes, but can be as long as several thousand bases. Some viruses and cellular genes have unusually long and structured 5' UTRs which may impact gene expression.

## **Conclusion**

Using the 5'-rapid amplification of cDNA ends, we located the transcription start initiation site upstream to 30-100 nucleotides upstream of the translation start site. In the present work, 5'UTR for all the 11 Serine Proteinase isoforms were obtained. Several regulatory sequences like binding sites for proteins, that may affect the mRNA's stability or translation and sequences that promote or inhibit translation initiation can be found. Along this Introns within 5'UTR have been linked to regulation of gene expression and mRNA transport can be speculated to be found. And their multiple sequence alignment shows the extensive sequence similarity in the all the 11 5'UTR regions as well except few bases which were different in Isoform 1,3,4, 5 and 7. Primers designed for genomic walking by using the few difference in 5'UTR sequences among the isoforms will be used for the isolation of flanking genomic sequences (e.g.promoter regions) adjacent to a known sequence.

## References

Allen WB JR. (1988). State lists of endangered and threatened species of reptiles and amphibians and laws and regulations covering collecting of reptiles and amphibians in each state Chicago, Illinois, Chicago Herpetological Society.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1994) In Current Protocols in Molecular Biology (Greene Publishing Associates and John Wiley & Sons, Inc., NY) Vol. 1, Ch. 2.

Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ, (2005). Snake venomomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes*, *Cerastes vipera* and *Macrovipera lebetina*. *Proteomics*. 5:4223-4235.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from  $\lambda$  bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91:2216–2220.

Birrell GW, Earl S, Masci PP, de Jersey J, Wallis TP, Gorman JJ, Lavin MF (2006). Molecular diversity in venom from the Australian Brown Snake, *Pseudonaja textilis*. *Mol Cell Proteomics*. 5:379-389.

Blevins, R. A. & Tulinsky, A. (1985). Refinement and the structure of the dimer of alpha-chymotrypsin at 1.67-Å<sup>0</sup> resolution. *J Biol Chem*. 260, 4264-75.

Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson (1994). *Molecular Biology of the Cell*. Garland Publishing.

Braud S., Bon, C., and Wisner, A. (2000). "Snake venom proteins acting on hemostasis." *Biochimie*. 82:851-859.

Borson, N. D., Sato, W. L. & Drewes, L. R. (1992). A lock-docking oligo (dT) primer for 5' and 3' RACE PCR. *PCR Methods Applic*. 2:144–148.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from  $\lambda$  bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91:2216–2220.

Calvete J. J., Marcinkiewicz, C., Monle, D. (2005). "Snake venom disintegrins: evolution of structure and function," *Toxicon*. vol. 45, no. 8, pp. 1063–1074.

Chenchik, A., Moqadam, F. & Siebert, P. (January 1995). Marathon cDNA amplification: A new method for cloning full-length cDNAs. *Clontechniques X* (1):5–8.

Chenchik, A., Moqadam, F. & Siebert, P. (1996). A new method for full-length cDNA cloning by PCR. In *A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis*. Ed. Krieg, P. A. (Wiley- Liss, Inc.), pp. 273–321.

Chenchik, A., Zhu, Y., Diatchenko, L., Li, R., Hill, J. & Siebert, P. (1998). Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. In *RT-PCR Methods for Gene Cloning and Analysis*. Eds. Siebert, P. & Larrick, J. (BioTechniques Books, MA), pp. 305–319.

Ching AT, Rocha MM, Paes Leme AF, Pimenta DC, de Fatima DF, Serrano SM, Ho PL, Junqueira-de-Azevedo IL (2006). Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. *FEBS Lett.* 580:4417-4422.

Choudhuri, S. (2004). "Gene regulation and molecular toxicology." *Toxicol.Mech.Methods.* 15:1-23.

Cidade DA, Simao TA, Davila AM, Wagner G, de LMJ, Lee HP, Bon C, Zingali RB, Albano RM (2006). Bothrops jararaca venom gland transcriptome: Analysis of the gene expression pattern. *Toxicon.* 48:437-461.

Conant R. (1975). *A Field Guide to Reptiles and Amphibians of Eastern and Central North America*. Second Edition. First published in 1958. Houghton Mifflin Company Boston. 429 pp. 48 plates. ISBN 0-395-19979-4. ISBN 0-395-19979-8 (pbk.).

Davie,E.W. (2003). "A brief historical review of the waterfall/cascade of blood coagulation." *J.Biol.Chem.* 278:50819-50832.

Decuypere S, Vandesompele J, Yardley V, De Donckeri S, Laurent T, Rijal S, Llanos-Cuentas A, Chappuis F, Arevalo J, Dujardin JC (2005). Differential polyadenylation of ribosomal RNA during post-transcriptional processing in *Leishmania*. *Parasitology.* 131:321-329.

Degenhardt WG, Painter CW, Price AH (1966). *The Amphibians and Reptiles of New Mexico* University of New Mexico Press, Albuquerque, NM.

Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M (1996). Accelerated evolution of crotalinae snake venom gland serine proteases. *FEBS Lett.* 397:83-88.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991). "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19:4008.

Donahue, B.A., Yin, S., Taylor, J.S., Reines, D. and Hanawalt, P.C. (1994). "Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template." *Proc.Natl.Acad.Sci. U.S.A.* 91:8502-8506.

- Dvir, A. (2002). "Promoter escape by RNA polymerase II." *Biochim.Biophys.Acta.* 1577:208-223.
- Fenton, A.W., West, P.R., Odell, G.V., Hudiburg, S.M., Ownby, C.L., Mills, J.N., Scroggins, B.T., Shannon, S.B., (1995). Arthropod venom citrate inhibits phospholipase A2. *Toxicon* 33, 763–770.
- Fleischmann J, Liu H, Wu CP (2004). Polyadenylation of ribosomal RNA by *Candida albicans* also involves the small subunit. *BMC Mol Biol.* 5:17.
- Freitas, M.A., Geno, P.W., Sumner, L.W., Cooke, M.E., Hudiburg, S.A., Ownby, C.L., Kaiser, I.I., Odell, G.V., (1992). Citrate is a major component of snake venoms. *Toxicon* 30, 461–464.
- Francis, B., Seebart, C., Kaiser, I.I., (1992). Citrate is an endogenous inhibitor of snake venom enzymes by metal-ion chelation. *Toxicon* 30, 1239–1246.
- Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM (2004). *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene.* 337:55-69.
- Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Tumer, D. H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* 83:9373–9377.
- Fry BG, Wuster W (2004). Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol*, 21:870-883.
- Fry, B. G. (2005). "From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins." *Genome Res.* 15:403-420.
- Gao, R., Kini, R. M., and Gopalakrishnakone, P. (2002). "A novel prothrombin activator from the venom of *Micropechis ikaheka*: isolation and characterization." *Arch.Biochem. Biophys.* 408:87-92.
- Goldman, S. R., Ebright, R. H., and Nickels, B. E. (2009). "Direct detection of abortive RNA transcripts in vivo." *Science.* 324:927-928.
- Grams, F., Huber, R., Kress, L.F., Moroder, L., Bode, W., (1993). Activation of snake venom metalloproteinases by a cysteine switch-like mechanism. *FEBS Lett.* 335, 76–80.
- Hobert JP, Montgomery CE, Mackessy SP (2004). Natural History of the Massasauga *Sistrurus catenatus edwardsii*, in Southeastern Colorado. *The Southwestern Naturalist.* 49:321-326.

Hodgson WC, Wickramaratna JC (2002). In vitro neuromuscular activity of snake venoms. *Clin Exp Pharmacol Physiol.* 29:807-814.

Holycross AT, Mackessy SP (2002). Variation in the diet of *Sistrurus catenatus* (Massasauga), with emphasis on *Sistrurus catenatusedwardsii* (Desert Massasauga). *Journal of Herpetology.* 36:454-464.

Huang, M.N., Hung, H.L., Stanfield-Oakley, S.A. and High, K.A. (1992). "Characterization of the human blood coagulation factor X promoter." *J.Biol.Chem.* 267:15440-15446.

Humphrey H., Dalke, A. & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 14, 33-38.

Junqueira-de-Azevedo IL, Ho PL (2002). A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). *Gene.* 299:279-291.

Joseph, J.S., Chung, M.C., Jeyaseelan, K. and Kini, R.M. (1999). "Amino acid sequence of trocarin, a prothrombin activator from *Tropidechis carinatus* venom: its structural similarity to coagulation factor Xa." *Blood.* 94:621-631.

Juarez P, Sanz L, Calvete JJ: Snake venomomics (2004): characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* 2004. 4:327-338.

Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M Jr., Xavier MA, Fontes MR, Giglio JR, Franca SC: Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects: I--gene expression profile of highly expressed phospholipases A2. *Biochimie* 2004, 86:211-219.

Kini, R.M. and Chow, G. (2001). "Exogenous inhibitors of platelet aggregation from animal sources." *Thromb.Haemost.* 85:179-181.

Kloog Y, Ambar I, Sokolovsky M, Kochva E, Wollberg Z, Bdolah A (1988). Sarafotoxin, a novel vasoconstrictor peptide: phosphoinositide hydrolysis in rat heart and brain. *Science.* 242:268-270.

Kornalik, F. and Blomback, B. (1975). "Prothrombin activation induced by Ecarin - a prothrombin converting enzyme from *Echis carinatus* venom." *Thromb.Res.* 6:57-63.

Kochva, E., (1959). An extended venom gland in the Israel mole viper, *Atractaspis engaddensis* Haas 1950. *Bull. Res. Council Israel* B 8, 31-34.

Kochva, E., Gans, C., (1966). Histology and histochemistry of venom glands of some crotaline snakes. *Copeia* 1966, 506-515.

- Kochva, E., Oron, U., Bdolah, A., Allon, N., (1975). Regulation of venom secretion and injection in viperid snakes. *Toxicon* 13, 104.
- Kochva, E., Oron, U., Ovadia, M., Simon, T., Bdolah, A., (1980). Venom glands, venom synthesis, venom secretion and evolution. In: Eaker, D., Wadstrom, T. (Eds.), *Natural Toxins*. Pergamon Press, Oxford, pp. 3–12.
- Kochva, E., (1987). The origin of snakes and evolution of the venom apparatus. *Toxicon* 25, 65–106.
- Kellogg, D. E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P. & Chenchik, A. (1994). TaqStart Antibody: Hot start PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *BioTechniques* 16:1134–1137.
- Kisiel W, Kondo S, Smith KJ, McMullen BA, Smith LF (1987). Characterization of a protein C activator from *Agkistrodon contortrix contortrix* venom. *J Biol Chem.* 262:12607-12613.
- Kopelman NM, Lancet D, Yanai I (2005). Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat Genet.* 37:588-589.
- Kuai L, Fang F, Butler JS, Sherman F (2004). Polyadenylation of rRNA in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, 101:8581-8586.
- Lee, T.I. and Young, R.A. (2000). "Transcription of eukaryotic protein-coding genes." *Annu.Rev.Genet.* 34:77-137.
- Lee JW, Seu JH, Rhee IK, Jin I, Kawamura Y, Park W (1999): Purification and characterization of brevinase, a heterogeneous twochain fibrinolytic enzyme from the venom of Korean snake, *Agkistrodon blomhoffii brevicaudus*. *Biochem Biophys Res Commun.* 260:665-670.
- Long M, Betran E, Thornton K, Wang W (2003). The origin of new genes: glimpses from the young and old. *Nat Rev Genet.*
- Li, M., Bullock, C.M., Knauer, D.J., Ehlert, F.J., and Zhou, Q.Y. (2001a). "Identification of two prokineticin cDNAs: recombinant proteins potently contract gastrointestinal smooth muscle." *Mol.Pharmacol.* 59:692-698.
- Li M, Fry BG, Kini RM (2005). Eggs-only diet: its implications for the toxin profile changes and ecology of the marbled sea snake (*Aipysurus eydouxii*). *J Mol Evol.* 60:81-89.
- Li M, Fry BG, Kini RM (2005). Putting the brakes on snake venom evolution: the unique molecular evolutionary patterns of *Aipysurus eydouxii* (Marbled sea snake) phospholipase A2 toxins. *Mol Biol Evol.* 22:934-941.

- Li S, Wang J, Zhang X, Ren Y, Wang N, Zhao K, Chen X, Zhao C, Li X, Shao J, Yin J, West MB, Xu N, Liu S (2004). Proteomic characterization of two snake venoms: *Naja naja atra* and *Agkistrodon halys*. *Biochem J.* 384:119-127.
- Markland FS, Kettner C, Schiffman S, Shaw E, Bajwa SS, Reddy KN, Kirakossian H, Patkos GB, Theodor I, Pirkle H (1982). Kallikrein-like activity of crotalase, a snake venom enzyme that clots fibrinogen. *Proc Natl Acad Sci U S A.* 79:1688-1692.
- Matsui T, Fujimura Y, Titani K (2000). Snake venom proteases affecting hemostasis and thrombosis. *Biochim Biophys Acta.* 1477:146-156.
- Mackessy SP (1988). Venom ontogeny in the Pacific rattlesnakes *Crotalus viridis helleri* and *Crotalus viridis oreganus*. *Copeia.* 92-101.
- Mackessy SP (2002). Biochemistry and pharmacology of colubrid snake venoms. *J Toxicol Toxin Rev.* 21:43-83.
- Mackessy SP, Baxter LM (2006). Bioweapons synthesis and storage: the venom gland of front-fanged snakes. *Zoologischer Anzeiger.* 245:147-159.
- Meier J, Stocker K (1991). Effects of snake venoms on hemostasis. *Crit Rev Toxicol,* 21:171-182.
- Morita T (2004). C-type lectin-related proteins from snake venoms. *Curr Drug Targets Cardiovasc Haematol Disord.* 4:357-373.
- Nawarak J, Sinchaikul S, Wu CY, Liau MY, Phutrakul S, Chen ST (2003). Proteomics of snake venoms from Elapidae and Viperidae families by multidimensional chromatographic methods. *Electrophoresis.* 24:2838-2854.
- Nishida S, Fujimura Y, Miura S, Ozaki Y, Usami Y, Suzuki M, Titani K, Yoshida E, Sugimoto M, Yoshioka A. (1994). Purification and characterization of bothrombin, a fibrinogen-clotting serine protease from the venom of *Bothrops jararaca*. *Biochemistry.* 33:1843-1849.
- Kloog Y, Ambar I, Sokolovsky M, Kochva E, Wollberg Z, Bdolah A (1988). Sarafotoxin, a novel vasoconstrictor peptide: phosphoinositide hydrolysis in rat heart and brain. *Science.* 242:268-270.
- Munekiyo, S.M., Mackessy, S.P., (1998). Effects of temperature and storage conditions on the electrophoretic, toxic, and enzymatic stability of venom components. *Comp. Biochem. Physiol.* 119B, 119–127.
- Niewiarowski S, Kirby EP, Brudzynski TM, Stocker K (1979). Thrombocytin, a serine protease from *Bothrops atrox* venom. 2. Interaction with platelets and plasma-clotting factors. *Biochemistry.* 18:3570-3577.

- Owen, W.G. and Jackson, C.M. (1973). "Activation of prothrombin with *Oxyranus scutellatus* scutellatus (Taipain snake) venom." *Thromb.Res.* 3:705-714.
- Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, Thomson TM, Antonarakis SE, Guigo R (2006). Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.* 16:37-44.
- Paine, M.J., Desmond, H.P., Theakston, R.D. and Crampton, J.M. (1992). "Gene expression in *Echis carinatus* (carpet viper) venom glands following milking." *Toxicon.* 30:379-386.
- Parkinson CL (1999). Molecular systematics and biogeographical history of pitvipers as determined by mitochondrial ribosomal DNA sequences. *Copeia*, 3:576-586.
- Pathy, L. (2003). "Modular assembly of genes and the evolution of new functions." *Genetica.* 118:217-231.
- Pierre LS, Woods R, Earl S, Masci PP, Lavin MF (2005). Identification and analysis of venom gland-specific genes from the coastaltaipan (*Oxyranus scutellatus*) and related species. *Cell Mol Life Sci.* 62:2679-2693.
- Pook CE, McEwing R (2005). Mitochondrial DNA sequences from dried snake venom: a DNA barcoding approach to the identification of venom samples. *Toxicon.* 46:711-715.
- Qinghua L, Xiaowei Z, Wei Y, Chenji L, Yijun H, Pengxin Q, Xingwen S, Songnian H, Guangmei Y (2006). A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: identification of the toxins potentially involved in coagulopathy. *Biochem Biophys Res Commun.* 341:522-531.
- Rao, V.S., Swarup, S. and Kini, R.M. (2003b). "The nonenzymatic subunit of pseutarin C, a prothrombin activator from eastern brown snake (*Pseudonaja textilis*) venom, shows structural similarity to mammalian coagulation factor V." *Blood.* 102:1347-1354.
- Reza, M.A., Minh, L.T., Swarup, S. and Kini, R.M. (2006). "Molecular evolution caught in action: gene duplication and evolution of molecular isoforms of prothrombin activators in *Pseudonaja textilis* (brown snake)." *J.Thromb.Haemost.* 4:1346-1353.
- Reza, M.A., Swarup, S. and Kini, R.M. (2005a). "Two parallel prothrombin activator systems in Australian rough-scaled snake, *Tropidechis carinatus*. Structural comparison of venom prothrombin activator with blood coagulation factor X." *Thromb.Haemost.* 93:40-47.
- Reza, M.A., Swarup, S. and Kini, R.M. (2005b). "Gene structures of trocarin D and coagulation factor X, two functionally diverse prothrombin activators from Australian rough scaled snake." *Pathophysiol. Haemost. Thromb.* 34:205-208.

- Rotenberg D, Bamberger E.S., Kochva E (1971). Studies on ribonucleic acid synthesis in the venom glands of *Vipera palaestinae* (Ophidia, Reptilia). *Biochem J.* 121:609-612.
- Saguchi K, Hagiwara-Saguchi Y, Murayama N, Ohi H, Fujita Y, Camargo AC, Serrano SM, Higuchi S (2005). Molecular cloning of serine proteinases from *Bothrops jararaca* venom gland. *Toxicon.* 46:72-83.
- Santamaria C, Larios S, Quiros S, Pizarro-Cerda J, Gorvel JP, Lomonte B, Moreno E (2005). Bactericidal and antiendotoxic properties of short cationic peptides derived from a snake venom Lys49 phospholipase A2. *Antimicrob Agents Chemother.* 49:1340-1345.
- Sanz L.A., Bazaá, N., Marrakchi et al., (2006). "Molecular cloning of disintegrins from *Cerastes vipera* and *Macrovipera lebetina transmediterranea* venom gland cDNA libraries: insight into the evolution of the snake venom integrin-inhibition system," *Biochemical Journal.* vol. 395, no. 2, pp. 385–392.
- Seigel, A.R. (1986). Ecology and conservation of an endangered rattlesnake, *Sistrurus catenatus*, in Missouri, USA. *Biological Conservation.* 35:333-346.
- Servent, D., Mourier, G., Antil, S., Me´nez, A., (1998). How do snake curaremimetic toxins discriminate between nicotinic acetylcholine receptor subtypes. *Toxicol. Lett.* 102–103, 199–203.
- Sato,S. and Tamiya, N. (1971). "The amino acid sequences of erabutoxins, neurotoxic proteins of sea-snake (*Laticauda semifasciata*) venom." *Biochem.J.* 122:453-461.
- Serrano SM, Shannon JD, Wang D, Camargo AC, Fox JW (2005). A multifaceted analysis of viperid snake venoms by two-dimensional gel electrophoresis: an approach to understanding venom proteomics. *Proteomics.* 5:501-510.
- Silveira AM, Magalhaes A, Diniz CR, De Oliveira EB (1989). Purification and properties of the thrombin-like enzyme from the venom of *Lachesis muta muta*. *Int J Biochem.* 21:863-871.
- Slomovic S, Laufer D, Geiger D, Schuster G (2005). Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. *Mol Cell Biol.* 25:6427-6435.
- Spencer,C.A. and Groudine M (1990)."Transcription elongation and eukaryotic gene regulation." *Oncogene.* 5:777-785.
- Spilianakis, C.G., Lalioti, M.D, Town T, Lee G.R, and Flavell R.A (2005). "Interchromosomal associations between alternatively expressed loci." *Nature.* 435:637-645.
- Straight, R., Glenn, J.L., Snyder, C.C (1976). Antivenom activity of rattlesnake blood plasma. *Nature.* 261, 259–260.

- Strydom, D.J. (1973). "Snake venom toxins. Structure-function relationships and phylogenetics." *Comp Biochem. Physiol B*. 44:269-281.
- Stebbins, R.C. (1985). *A field guide to western reptiles and amphibians* 2nd edition. Edited by: Mifflin H. New York.
- Takeuchi, Y., Noguchi, S., Satow, Y., Kojima, S., Kumagai, I. et al. (1991). Molecular recognition at the active site of subtilisin BPN': crystallographic studies using genetically engineered proteinaceous inhibitor SSI (Streptomyces subtilisin inhibitor). *Protein Eng*, **4**, 501-8
- Tans, G., Govers-Riemsma, J.W., Van Rijn, J.L., Rosing, J. (1985). "Purification and properties of a prothrombin activator from the venom of *Notechis scutatus scutatus*." *J.Biol.Chem.* 260:9366-9372.
- Tsai IH, Chen YH, Wang YM (2004). Comparative proteomics and subtyping of venom phospholipases A2 and disintegrins of Protobothrops pit vipers. *Biochim Biophys Acta*. 1702:111-119.
- Vidal N, Hedges SB (2002). Higher-level relationships of snakes inferred from four nuclear and mitochondrial genes. *C R Biol*. 325:977-985.
- Wagstaff SC, Harrison RA (2006). Venom gland EST analysis of the saw-scaled viper, *Echis ocellatus*, reveals novel alpha9beta1 integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like aspartic proteases. *Gene*. 377:21-32.
- Wahle E, Kuhn U (1997). The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acid Res Mol Biol*. 57:41-71.
- Wang YM, Wang SR, Tsai IH (2001). Serine protease isoforms of *Deinagkistrodon acutus* venom: cloning, sequencing and phylogenetic analysis. *Biochem J*. 354:161-168.
- Wright AH, Wright AA. (1957). *Handbook of Snakes*. Comstock Publishing Associates. (7th printing, 1985). 1105 pp. ISBN 0-8014-0463-0.
- Wuster W, Golay P, Warrell DA (1999). Synopsis of recent developments in venomous snake systematics, No. 3. *Toxicon*. 37:1123-1129.
- Xu, X., Wang X. S., Xi, X.T., Liu, J., Huang, J.T., Lu, Z.X. (1982). "Purification and partial characterization of hyaluronidase from five pace snake (*Agkistrodon acutus*) venom." *Toxicon*. 20:973-981.
- Yamada, D. and Morita, T. (1997). "Purification and characterization of a Ca<sup>2+</sup> - dependent prothrombin activator, multactivase, from the venom of *Echis multisquamatus*." *J.Biochem. (Tokyo)*. 122:991-997.

Yamamoto C, Tsuru D, Oda-Ueda N, Ohno M, Hattori S, Kim ST (2002). Flavoxobin, a serine protease from *Trimeresurus flavoviridis* (habu snake) venom, independently cleaves Arg726-Ser727 of human C3 and acts as a novel, heterologous C3 convertase. *Immunology*. 107:111-117.

Zhang B, Liu Q, Yin W, Zhang X, Huang Y, Luo Y, Qiu P, Su X, Yu J, Hu S, Yan G (2006). Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC Genomics*. 7:152.

Zhu, Y.Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. M. (2001). Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *BioTechniques* 30:892–897.

## **ANNEXURE**

**SMART™ RACE cDNA Amplification Kit component:**

**First-strand cDNA Synthesis:**

- 7 µl **SMART II™ A Oligonucleotide** (12 µM)  
5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'
- 7 µl **3'-RACE CDS Primer A** (3'-CDS; 12 µM)  
5'-AAGCAGTGGTATCAACGCAGAGTAC(T)30V N-3'  
(N = A, C, G, or T; V = A, G, or C)
- 7 µl **5'-RACE CDS Primer A** (5'-CDS; 12 µM)  
5'-(T)25V N-3'  
(N = A, C, G, or T; V = A, G, or C)
- 20 µl **5X First-Strand Buffer**  
250 mM Tris-HCl (pH 8.3)  
375 mM KCl  
30 mM MgCl<sub>2</sub>
- 20 µl **Dithiothreitol** (DTT; 20 mM)
- 1 ml **Deionized H<sub>2</sub>O**

**5'- & 3'-RACE PCR**

- 400 µl **10X Universal Primer A Mix** (UPM)  
Long (0.4 µM):  
5'-CTAATACGACTCACTATAGGGCCAAGCAGTGGTATCAACGCAGAGT-3'
- Short (2 µM):  
5'-CTAATACGACTCACTATAGGGC-3'
- 50 µl **Nested Universal Primer A** (NUP; 10 µM)  
5'-AAGCAGTGGTATCAACGCAGAGT-3'

**Control Reagents**

- 5 µl **Control Human Placental Total RNA** (1 µg/µl)
- 25 µl **Control 5'-RACE TFR Primer** (10 µM)
- 25 µl **Control 3'-RACE TFR Primer** (10 µM)
- **MMLV Reverse Transcriptase**

- **Advantage® 2 PCR Kit** (Cat. Nos. 639206 & 639207)
- **PCR reaction tubes**

Human Placental Total RNA and SMART II A Oligonucleotide were stored at  $-70^{\circ}\text{C}$ .

NucleoTrapR Gel Extraction Kit was stored at room temperature.

All other reagents were stored at  $-20^{\circ}\text{C}$ .

### **Nucleospin Genomic DNA Isolation requirement:**

#### **1 Components**

##### **1.1 Kit contents**

<b>Cat. No.</b>	<b>10 preps 740952.10</b>
Lysis Buffer T1	5 ml
Buffer B1*	6 ml
Buffer B2*	1.5 ml
Wash Buffer BW	6 ml
Wash Buffer B5 (Concentrate)*	4 ml
Elution Buffer BE**	3 ml
Proteinase K (lyophilized)*	6 mg
Proteinase Buffer PB	0.8 ml
NucleoSpin® Tissue Columns (light green rings)	10
Collection Tubes (2 ml)	20
Labels for Lysis Buffer B3	1
User Manual	1

### **Genome walking components:**

#### **List of Components**

Store human genomic DNA at  $4^{\circ}\text{C}$ ; all other components at  $-20^{\circ}\text{C}$ .

#### **• Restriction enzymes and buffers:**

- 30  $\mu\text{l}$  *Dra* I (10 units/ $\mu\text{l}$ )
- 100  $\mu\text{l}$  10X *Dra* I Restriction Buffer
- 25  $\mu\text{l}$  *EcoR* V (10 units/ $\mu\text{l}$ )
- 50  $\mu\text{l}$  10X *EcoR* V Restriction Buffer
- 50  $\mu\text{l}$  *Pvu* II (10 units/ $\mu\text{l}$ )
- 100  $\mu\text{l}$  10X *Pvu* II Restriction Buffer
- 25  $\mu\text{l}$  *Stu* I (10 units/ $\mu\text{l}$ )
- 50  $\mu\text{l}$  10X *Stu* I Restriction Buffer
- 75  $\mu\text{l}$  **Control Human Genomic DNA** (0.1  $\mu\text{g}/\mu\text{l}$ )
- 10  $\mu\text{l}$  **T4 DNA Ligase** (6 units/ $\mu\text{l}$ )
- 40  $\mu\text{l}$  **10X Ligation Buffer**

- 36  $\mu$ l **GenomeWalker Adaptor** (25  $\mu$ M)
  - 250  $\mu$ l **Adaptor Primer 1** (AP1; 10  $\mu$ M)
  - 250  $\mu$ l **Nested Adaptor Primer 2** (AP2; 10  $\mu$ M)
- See Figure 3 on the next page for the sequences of AP1 & AP2.
- 10  $\mu$ l **GenomeWalker Human Positive Control Library**
  - 50  $\mu$ l **Positive Control tPA Primer** (PCP1; 10  $\mu$ M)
  - 50  $\mu$ l **Positive Control tPA Nested Primer** (PCP2; 10  $\mu$ M)

The following reagents are required:

- **Phenol**
- **Chloroform**
- **Glycogen** (10  $\mu$ g/ $\mu$ l)
- **3 M sodium acetate**
- **95% ethanol**
- **80% ethanol**
- **TE** 10 mM Tris, 0.1 mM EDTA (10/0.1, pH 7.5)
- **TE** 10 mM Tris, 1 mM EDTA (10/1, pH 7.5)
- **0.5X TBE buffer or 1X TAE buffer** (See Note in Section VII.B.1)
- **Advantage 2 Polymerase Mix** (50X)
- **10X PCR reaction buffer:** If we are using a DNA polymerase mix other than Advantage 2 Polymerase Mix, use the PCR reaction buffer provided with the enzyme mix.
  - **dNTP mix:** 10 mM each of dATP, dCTP, dGTP & dTTP. Store at  $-20^{\circ}\text{C}$ .
- **PCR reaction tubes**
  - **Deionized H<sub>2</sub>O** (Milli-Q<sup>TM</sup>-filtered or equivalent)
  - **1 kb ladder of DNA size markers**

# snake catching assignment in NUS

*Molecular Mechanisms Involved in regulation of Serine Proteinase Isoforms Expression in Sistrurus catenatus edwardsii*

