Spring 2022

Comparative Modeling and Evolutionary Comparison of Serine Protease, A Timber Rattlesnake Venom Protein

Qawer Ayaz
aqawer@stu.jsu.edu

THESIS APPROVAL

Candidate: Qawer Ayaz

Major: Biology

Thesis Title: Comparative Modeling and Evolutionary Comparison of Serine Protease, A Timber Rattlesnake Venom Protein

Jenna L W Ridlen — 4/4/22

Jenna Lee Ridlen — Date

Assistant Professor of Biology

Major Professor

Chris A. Murdock — 4/4/22

Chris Murdock — Date

Professor of Biology

Richard Watkins — 4/8/22

Richard Watkins — Date

Instructor of Biology

Channing R. Ford — 4/8/22

Channing R. Ford — Date

Senior Director, Graduate Studies

COMPARATIVE MODELING AND EVOLUTIONARY COMPARISON OF VG35 SERINE PROTEASE, A TIMBER RATTLESNAKE VENOM PROTEIN

A Thesis Submitted to the
Graduate Faculty
of Jacksonville State University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
With a Major in Biology

By

QAWER AYAZ

Jacksonville, Alabama

May 06, 2022

Qawer Ayaz 04/09/2022

Qawer Ayaz Date

ABSTRACT

The aim of this study is to create homology model of VG35 serine protease and evaluate the evolutionary comparison of secondary structure on basis of protein model using YASARA. This method was furthermore used to predict the potential epitopes which can helps in the investigation of future studies.

The VG35 was used to run BLAST search which gave most resembled serine protease of different species which was then translated and modeled in YASARA. The modeled protein data was then used to determine the secondary structure. This was then used for evolutionary comparison of all proteins to VG35. And then potential epitope was found using DiscoTope 2.0.

Homology ribbon model were generated using YASARA for VG35 and 12 other serine protease proteins. This was then used to make a dataset of secondary structure which shows sheet, helix, turn, and coil of protein structure. This dataset was used to determine the conservation of other proteins to VG35. And the homology ribbon models were used to find epitopes of VG35, VG172, and serine protease of *Crotalus tigris*.

The homology model helps to understand the difference in the structure of proteins and how it has impact on the enzymatic activity of the proteins. This also helps in understanding the variants which lead to change in the secondary structure of proteins which can be caused during post-transcriptional process. The potential epitopes found can be used to future studies to design and develop anti-venom to target specific venom.

Preliminary pages: 9. Manuscript pages: 54

ACKNOWLEDGEMENTS

I would like to thank and dedicate this research in the memory of Dr. Benjie Blair who proposed the idea initially and helped me understand the research. And in memory of Dr. George Cline, who guided me in making the right decision for the research.

I would also like to acknowledge and thank my committee Mr. Richard Watkins, Dr. Jenna Lee Ridlen, and Dr. Chris Murdock for all the support and guidance which made me reverse engineer the complete proposed idea which made this research possible.

Qawer Ayaz

TABLE OF CONTENTS

PAGE

LIST OF TABLES

Page

LIST OF FIGURES

Page

I. INTRODUCTION

History of Snake Venom

The history of the venom study traces back more than two millennia. The earliest survived detailed synopsis of venomous animals and their bites were given as early as 322 BC by Aristotle in Historia Animalium. Modern scientific understanding of venomous snakes and venoms began to take shape thanks to the Italian scientists Francesco Redi and Felice Fontana, who worked in the city of Pisa in the 17 and 18 centuries, respectively. Physician, biologist, linguist, and poet Francesco Redi (1664) published a treatise on poisonous snakes entitled "Osservazioni intorno alle vipere". He found that snake bile is not toxic, as it was accepted at that time, and toxicity is in the venom released from the teeth at the bite. Redi is considered as one of the founders of "toxinology," a specialist area of science dealing with animal, microbial, and plant venoms, poisons, and toxins. A hundred years later, in the 18th century, another Italian, Felice Fontana, discovered snake venom glands and obtained snake venom, which he used for a variety of experiments with animals. "Fontana corrected Redi: venom acts on the animals not by getting in stomach as Redi believed, but in the blood" (Redi 1664).

General Information

Snake venom is a highly modified saliva containing zootoxins that facilitate the immobilization and digestion of prey, and defense against threats. It is injected by unique fangs during a bite, and some species are also able to spit their venom (Bauchot 1994).

The glands that secrete the zootoxins are a modification of the parotid salivary glands found in other vertebrates and are usually situated on each side of the head, below and behind the eye, and encapsulated in a muscular sheath. The glands have large alveoli in which the synthesized venom is stored before being conveyed by a duct to the base of channeled or tubular fangs through which it is ejected (Bottrall et al. 2010). Venoms contain more than 20 different compounds, mostly proteins and polypeptides (Halliday and Kraig 2002). A complex mixture of proteins, enzymes, and various other substances with toxic and lethal properties serves to immobilize the prey animal (Mattison 2007). Enzymes play an important role in the digestion of prey, and various other substances are responsible for important but non-lethal biological effects. Some of the proteins in snake venom have very specific effects on various biological functions including blood coagulation, blood pressure regulation, and transmission of the nervous or muscular impulses, and have been developed for use as pharmacological or diagnostic tools, and even useful drugs (Bauchot 1994).

Physiological effects of Snake Venom

Snake venom has two main purposes: to impair the prey so that they can be consumed and to improve the digestibility of the prey. The four distinct types of venom act on the body differently. Proteolytic venom dismantles the molecular surroundings, including the bite, and enables the breakdown of proteins into smaller polypeptides or amino acids. Hemotoxic venom acts on the heart and cardiovascular system, destroys red blood cells (that is, causes hemotoxin), disrupts blood clotting, and/or causes organ degeneration and generalized tissue damage. Neurotoxic venom acts on the nervous system and brain, alters the normal activity of the nervous system in such a way as to cause permanent or reversible damage to nervous tissue. This can

eventually disrupt or even kill neurons, which are cells that transmit and process signals in the brain and other parts of the nervous system. Cytotoxic venom has a localized action at the site of the bite (Martin 1907).

Clinical applications of venom

Given that snake venom contains many biologically active ingredients, some may be useful to treat disease. For instance, phospholipases type A2 (PLA2s) from the Tunisian vipers *Cerastes cerastes* and *Macrovipera lebetina* have been found to have antitumor activity. Anticancer activity has been also reported for other compounds in snake venom. PLA2s hydrolyze phospholipids, thus could act on bacterial cell surfaces, providing novel antimicrobial (antibiotic) activities. The analgesic (painkilling) activity of many snake venom proteins has been long known. The main challenge, however, is to deliver protein to the nerve cells as proteins usually are not applicable as pills. Captopril emulates the function of the toxin found in Brazilian pit viper venom; Captopril is an ACE inhibitor (angiotensin converting enzyme) that was approved by the FDA in April 1981. It lowers blood pressure by inhibiting the production of angiotensin II which acts in a pathway that leads to vasoconstriction which raises blood pressure (Slagboom et al. 2017). Eptifibatide was modeled after a component in southeastern pygmy rattlesnake venom and is used in anticoagulation therapies to reduce the risk of heart attacks. Batroxobin, is a serine protease found in snake venom produced by *Bothrops atrox* and *Bothrops moojeni*, venomous species of pit viper found east of the Andes in South America. It cleaves fibrinogen, similarly to thrombin (Maroun 2004).

Serine protease

Serine proteases also known as serine endopeptidases are enzymes that cleave peptide bonds in proteins, in which serine serves as the nucleophilic amino acid at the (enzyme's) active site (Hedstrom 2002). They are found ubiquitously in both eukaryotes and prokaryotes. Serine proteases are widely distributed in nature and found in all kingdoms of cellular life as well as many viral genomes. However, significant differences exist in the distribution of each clan across species. For example, clan PA proteases are highly represented in eukaryotes, but rare constituents of prokaryotic and plant genomes. Vertebrates boast an array of clan PA proteases responsible for a variety of extracellular processes. SB and SC clans are most represented in other organisms. Serine proteases are usually endoproteases and catalyze bond hydrolysis in the middle of a polypeptide chain. However, several families of exoproteases have been described as removing one or more amino acids from the termini of target polypeptide chains (Page and Di Cera. 2008). Barrett and coworkers have devised a classification scheme based on statistically significant similarities in sequence and structure of all known proteolytic enzymes and term this database MEROPS (Rawlings et al. 2013). The classification system divides proteases into clans based on catalytic mechanism and families based on common ancestry. Over one third of all known proteolytic enzymes are serine proteases grouped into 13 clans and 40 families (Hedstrom. 2002).

Homology modeling

Homology modeling, also known as comparative modeling of protein, refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "*template*").

Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure (Chothia and Lesk 1986). Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure. It has been shown that three-dimensional protein structure is evolutionarily more conserved than would be expected based on sequence conservation alone. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity. The quality of the homology model is dependent on the quality of the sequence alignment and template structure. Homology modeling can produce high-quality structural models when the target and template are closely related, which has inspired the formation of a structural genomics consortium dedicated to the production of representative experimental structures for all classes of protein folds (Williamson 2000).

Model Assessment

Assessment of homology models without reference to the true target structure is usually performed with two methods: statistical potentials or physics-based energy calculations. Both methods produce an estimate of the energy (or an energy-like analog) for the model or models being assessed; independent criteria are needed to determine acceptable cutoffs. Neither of the two methods correlates exceptionally well with true structural accuracy, especially on protein

types underrepresented in the PDB, such as membrane proteins. Statistical potentials are empirical methods based on observed residue-residue contact frequencies among proteins of known structure in the PDB. They assign a probability or energy score to each possible pairwise interaction between amino acids and combine these pairwise interaction scores into a single score for the entire model. Some such methods can also produce a residue-by-residue assessment that identifies poorly scoring regions within the model, though the model may have a reasonable score overall. These methods emphasize the hydrophobic core and solvent-exposed polar amino acids often present in globular proteins. A very extensive model validation report can be obtained using “What If" software package YASARA; it produces a many pages' document with extensive analyses of nearly 200 scientific and administrative aspects of the model.

Yet Another Scientific Artificial Reality Application (YASARA)

YASARA creates a new level of interaction with 'artificial reality', that allows you to focus on your goal and forget about the details of the program. YASARA is powered by PVL (Portable Vector Language), a new development framework that provides performance way above traditional software. PVL allows you to visualize even the largest proteins and enables true interactive real-time simulations with highly accurate force fields on standard PCs, making use of GPUs. You can push and pull molecules around and work with dynamic models instead of static pictures. YASARA's innovative 3D engine, which is up to 10 times faster than what you usually know from OpenGL, you can load multiple structures at the same time, create publication-quality ray-traced images including labels, and program your own macros and Python plugins.

II. MATERIALS AND METHODS

Tissue Extraction and cDNA Synthesis

The Timber rattlesnake was captured in the wild and placed under controlled conditions of laboratory. The venom was collected by milking the snake three day prior to tissue extraction. The tissue was kept in the TRI reagent and to isolate total RNA the tissue was homogenized using Qiagen (Tokyo, Japan). The integrity of the RNA is checked using agarose gel electrophoresis.

The cDNA first strand was synthesized using SMART^{TM} cDNA Library Construction Kit (BD Biosciences Clontech, San Jose, CA, USA), from 0.5 to 1 μg of mRNA. The primers were the BD Biosciences oligonucleotides SMART™ Oligo VI and CDS-3M containing the SfiI A and SfiI B recognition sequence. Using LD PCR (Qiagen, Tokyo, Japan), first strand of cDNA and 2 μl of polymerase was amplified. Twenty-one cycles (initial denaturation at 95°C for 1 min; 95°C for 15s, 66°C for 20s, and 72°C for 3 min; and final extension at 66°C for 20s and 72°C for 3 min) is performed and cDNA purification was performed using QIA quick PCR purification kit (Qiagen, Tokyo, Japan) to remove primer excess, dNTPs, and salts. And, precipitated in ethanol.

Using duplex specific nuclease (DSN) enzyme and amplified. To perform the second PCR amplification, PCR was carried out as follows: 80 μL of MilliQ water, 10 μL of 10× PCR

buffer, 2 µL of dNTP mix (10 mM), 4 µL of PCR primer-M2, 2 µL of diluted normalized cDNA, and 2 µL of polymerase mix. PCR cycling used the following program: initial denaturation at 95°C for 1 min; 12 PCR cycles at 95°C for 15 s, 64°C for 20 s, and 72°C for 3 min; and a final extension at 64°C for 15 s and 72°C for 3 min. After completing cDNA amplification, 5 µL of the PCR products were analyzed by electrophoresis alongside of 1 kb DNA size markers on a 1.5% (w/v) agarose/EtBr gel run in 1× TAE buffer to check PCR quality.

Normalized cDNA samples digested by SfiI were ligated and packaged into to λTriplEx2 Vector. Three separate ligation reactions were set up with different ratios of cDNA to phage vector. It was then incubated at 16°C overnight. Lastly, MaxPlax™ Lambda Packaging Extracts for λ packaging reaction were used according to the manufacturer's instruction.

To determine the titer of unamplified library, first the overnight culture was prepared using a stock plate of single isolated E. coli, inoculated in 15 ml of LB/MgSO4/maltose broth in a 50 ml test tube. And incubated at 37ºC overnight at 140 rpm on a shake until OD600 of the culture reached 2.0. 1:5, 1:10, 1:15, and 1:20 dilutions of each of the packaging extracts were prepared in 1x lambda dilution buffer (NaCl, MgSO4•7H2O Tris-HCI (pH 7.5)). 1 µl of the diluted phage was added to 200 µl of the XL1-Blue overnight culture, and allowed the phage to adsorb at 37°C for15 min. After that 5 ml of melted LB/MgSO4 top agar was added and poured onto LB/MgSO4 plates and incubated at 37°C for 18 hours.

50µl of 1:20 diluted lysate was added to twenty 10ml tubes containing 200µL of overnight bacterial culture and incubated at 37°C for 15 minutes. Following incubation, we added 5ml of melted LB/MgSO4 soft top agar into the tubes and then poured the content on to LB/MgSO4 agar plates. After waiting 10 minutes for the top agar to harden, the inverted plates

were placed in the incubator at 37°C for 18 hours. Following incubation, 12ml of 1X lambda dilution buffer was added to each plate and then stored at 4°C overnight. The next day, plates were placed on a shaker (50 rpm) for an hour at a room temperature. From that, the pooled λ phage lysate was poured into a sterile 50ml screw-cap tube. To clear phage lysate of cell debris, 10 ml of chloroform was added to the tube, then it was vortexed for 2 minutes, and centrifuged at 7000rpm for 10 minutes. The supernatant was then collected and stored at 4°C.

Collection of raw sequence data of proteins

The target protein VG35 raw sequences data was taken and was quality checked. The quality check and annotation were based on comparing the obtained sequence data to the sanger sequence avi peak file. The changes in the nucleotide and the quality assurance were made as per the sequencer peaks of each base pair.

The resultant quality assured sequence was then used to determine the region similarity between different biological sequences. This was achieved using the BLAST (Basic Local Alignment Search Tool) which is an online tool offered by NCBI and acts as a cloud-based data warehouse of different biological genetic sequences and their relative functions. The BLAST tool compares the target nucleotide sequence to the data of different sequence available and calculate the statistical significance. By the results obtained from the BLAST search gave details of the different species of rattlesnake's venom serine protease has the highest statistical significance to the target VG35 serine protease.

A total of eleven different rattlesnake venom serine protease were selected which has the highest statistical significance to the target VG35 serine protease. Once the highest significance species were determined, UNIPROT was used to extract the raw sequence data of venom serine

protease for the other eleven different rattlesnake species. UNIPROT (The Universal Protein Resource) is a tool which is offered by the collaboration between European Bioinformatics Institute (EMBL-EBI) and Swiss Institute of Bioinformatics (SIB). It is a comprehensive online collection tool of protein sequence and its annotation data.

Once the raw nucleotide sequence of the target VG35 and serine protease other eleven species of rattlesnake was obtained using UNIPROT, it needs to be translated to amino acid sequence which can be used for Homology modelling in YASARA. Expasy, an online tool operated by Swiss Institute of Bioinformatics (SIB) was used to translate the nucleotide sequences to amino acid sequences. Once the translation was completed, each file was downloaded in FASTA format which can then be used to YASARA for homology modeling.

Homology Modeling using YASARA

Homology modeling in YASARA is done in two steps, i.e., homology modeling and refinement. From all the FASTA files of eleven different rattlesnake venom serine protease and VG35 serine protease, homology modeling was commenced in YASARA one at a time and left over night or consecutive day for the program to run. When the program has the run complete the resultant model was save in SCE and PDB file format. This was done with VG35 and eleven different species of rattlesnake proteins. When all the models are obtained and saved refinement of the models was done. The refinement process of program refines the slightest and minute changes to the model structure by determining the energy retention of each amino acid impact on structure. The refinement process was left over night or conservative day for the 500 ps simulation program to run. When completed the refinement process result in a text file which give the list of different possible models and their energy level.

Evolutionary Comparison

To determine the evolutionary comparison between the VG35 serine protease and the eleven serine proteases of different species of rattlesnakes, the secondary structure of each serine protease was found using YASARA. The secondary structure of each serine protease was then complied in a spreadsheet and different color was coded to the different aspect of the secondary structure. Using the different parts of secondary structure of each proteins helps to determine the evolutionary relationship between the different species of rattlesnake serine protease and the VG35 serine protease.

Epitope of VG35 and VG172

When the evolutionary relationship was found of VG35 serine protease, it was then subjected to determine the potential epitope site which can acts as a protein binding site for the anti-venom. The epitopes were found using an online tool called Discotope 2.0, in which the target VG35 serine protease was compared with the epitopes found on VG172 and a randomly selected serine protease from the eleven species given.

III. RESULTS

The sequence data obtained was used to gather the serine protease data of other species. As per the BLAST result the serine protease K8RSA4 of *Crotalus horridus* was taken and used to make the homology ribbon model in YASARA (Figure 1). The second result was of serine protease of *Crotalus tigris* was taken and ribbon model was created using YASARA (Figure 2). This process was then continued to obtain the serine protease of different species, VG172 and VG35 and the resultant homology ribbon model was created using YASARA (Figures 3-13). When all the homology ribbon model and model with least energy retention was obtained, the models were used to identify secondary structure of each protein. Each aspect of secondary structure was color coded for visual understanding of the secondary structure across of each protein (Table 1). If an amino acid has a secondary structure of coil, it is colored as black, white amino acid represents the turn, blue amino acid represents sheet and yellow amino acid represent helix (Table 1).

The obtained homology ribbon model was then superimposed to on each other using YASARA and the correlation of each protein was found to determine the conservation of VG35. The conservation of each amino acid helps to determine impact of insertion, deletion, and other variants of an amino acid at position found in all the serine protease (Table 2 and Figure 14).

The homology ribbon model was used in Discotope 2.0 to determine the possible epitopes found in VG35, VG172 and *Crotalus tigris* serine protease. It was determined that

serine protease of *Crotalus tigris* has 8 sites out of 244 residues which are found to be potential epitopes (Table 3). VG172 was found to have 37 sites out of 194 residues which can be possible epitopes (Table 4). In VG35 it was determined that it has 28 sites out of 201 total residue which can be possible epitopes (Table 5).

IV. DISCUSSION

Snake venom is a unique combination mixture of biologically active substances which has highly specialized mode of action of the complex biological processes. Many types of proteins which were isolated from snake venom has found its application in different aspects of medicine and many are found useful in other research (Stocker 1999). In this research by using the computational methods helps to determine the structure of various serine proteases. This helps to visualize the role of protein structure and determine its stability. When compared to the different serine proteases of different rattlesnake species it was found that the difference in amino acid to set location has different impact on the stability of the protein ribbon model and then leading to the difference in the secondary structure. Thus, this change in structure of protein and its impact on stability may affect the enzymatic activities of the protein. It was found that the change in the amino acids and structure of protein may lead to change in the hydrolytic activity of the serine protease activity and may also lead to complete for partial loss of catalytic action of the protein.

While analyzing the data of secondary structure and variants in amino acid chain (Table 1), it can be predicted that the variants among the sequence of amino acids in each protein has led to difference in the secondary structure of the proteins. These variants can be explained due to the different environmental factors which lead to the evolutionary changes in the protein splicing during post-transcriptional process of the mRNA.

By using the homology modeling of different serine protease helps to understand the function of the protein and thus different potential epitope sites were determined. This site can be used for future studies and can be tested to develop any possible anti-venom to battle the snake venom. A computational recombinant anti-venom can be designed and by using homology modeling in YASARA different antibody site and ligand/receptors sites can be found. This program of YASARA will give the detail function and result of how the designed anti-venom would act on the target protein.

BIBLIOGRAPHY

Baker M, Marti-Renom MA, Chiu W, Sali A. 2006. Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. Mol Biol. 357(5):1655–68.

Bauchot R. 1994. Snakes: a natural history. New York City, NY, USA: Sterling Publishing Co., Inc. 194-209.

Bottrall JL, Madaras F, Biven CD, Venning MG, Mirtschin PJ. 2010. Proteolytic activity of Elapid and Viperid Snake venoms and its implication to digestion. Journal of Venom Research. 1(3):18-28.

Brahmbhatt VVK, Bhatt H, Parmar U. 2013. Therapeutic potential of snake venom in cancer therapy: current perspectives. Asian Pacific Journal of Tropical Biomedicine. 3(2):156-162.

Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J. 1986-5(4):823-6.

Halliday A, Kraig T. 2002. Encyclopedia of reptiles and amphibians. Toronto, Canada: Firefly Books Ltd. 202-203.

Hedstrom L. 2002. Serine protease mechanism and specificity. Chem Rev. 102(12):4501-4502.

Jain D, Kumar S. 2012. Snake venom: a potent anticancer agent. Asian Pacific Journal of Cancer Prevention. 13(10):4855–4860.

John B, Sali A. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res. 31(14):3982-3992.

Kaczanowski S, Zielenkiewicz P. 2010. Why similar protein sequences encode similar three-dimensional structures. Theoretical Chemistry Accounts. 125:643-50.

Lazaridis T, Karplus M. 1999. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J. Mol. Biol. 288(3):477-487.

Levitt, M. 1992. Accurate modeling of protein conformation by automatic segment matching. J Mol Biol. 226(2):507–33.

Maroun RC, Serrano SM. 2004. Identification of the substrate-binding exosites of two snake venom serine proteinases: molecular basis for the partition of two essential functions of thrombin. J. Mol. Recognit. 17, 51–61.

Martin, C.J., and G. Lamb. 1907. Snake-poison and snakebite. A System of Medicine. Vol. ii (part ii). 783–821.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct. 29:291–325.

Mattison, C. 2007. The New Encyclopedia of Snakes. Princeton, New Jersey, USA Princeton University Press.

McCleary JG. 2007. Ziconotide: a review of its pharmacology and use in the treatment of pain. Neuropsychiatric Disease and Treatment. 3(1):69–85.

Page MJ, Di Cera E. 2008. Evolution of peptidase diversity. J Biol Chem. 283(44):30010-30014.

Pothineni NV, Watts TE, Ding Z, Dai Y, Deshmukh AJ. 2016. Eptifibatide-induced thrombocytopenia: when inhibitor turns killer. American Journal of Therapeutics. 23(1): e298–299.

Raufman JP. 1996. Bioactive peptides from lizard venoms. Regulatory Peptides. 61(1):1–18.

Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. 2008. MEROPS: the peptidase database. Nucleic Acids Res. 2008:36.

Redi F, 1664. Osservazioni Intorno alle Vipere. Florence: All'Insegna della Stella 98.

Richardson DC, Richardson JS. 2002. Teaching molecular 3-D literacy. Biochemistry and Molecular Biology Education. 30:21–26.

Richardson JS. 1985. Schematic drawings of protein structures. Methods in Enzymology. 115:359–380.

Saab F, Ionescu C, Schweiger MJ. 2012. Bleeding risk and safety profile related to the use of eptifibatide: a current review. Expert Opinion on Drug Safety. 11(2):315-24.

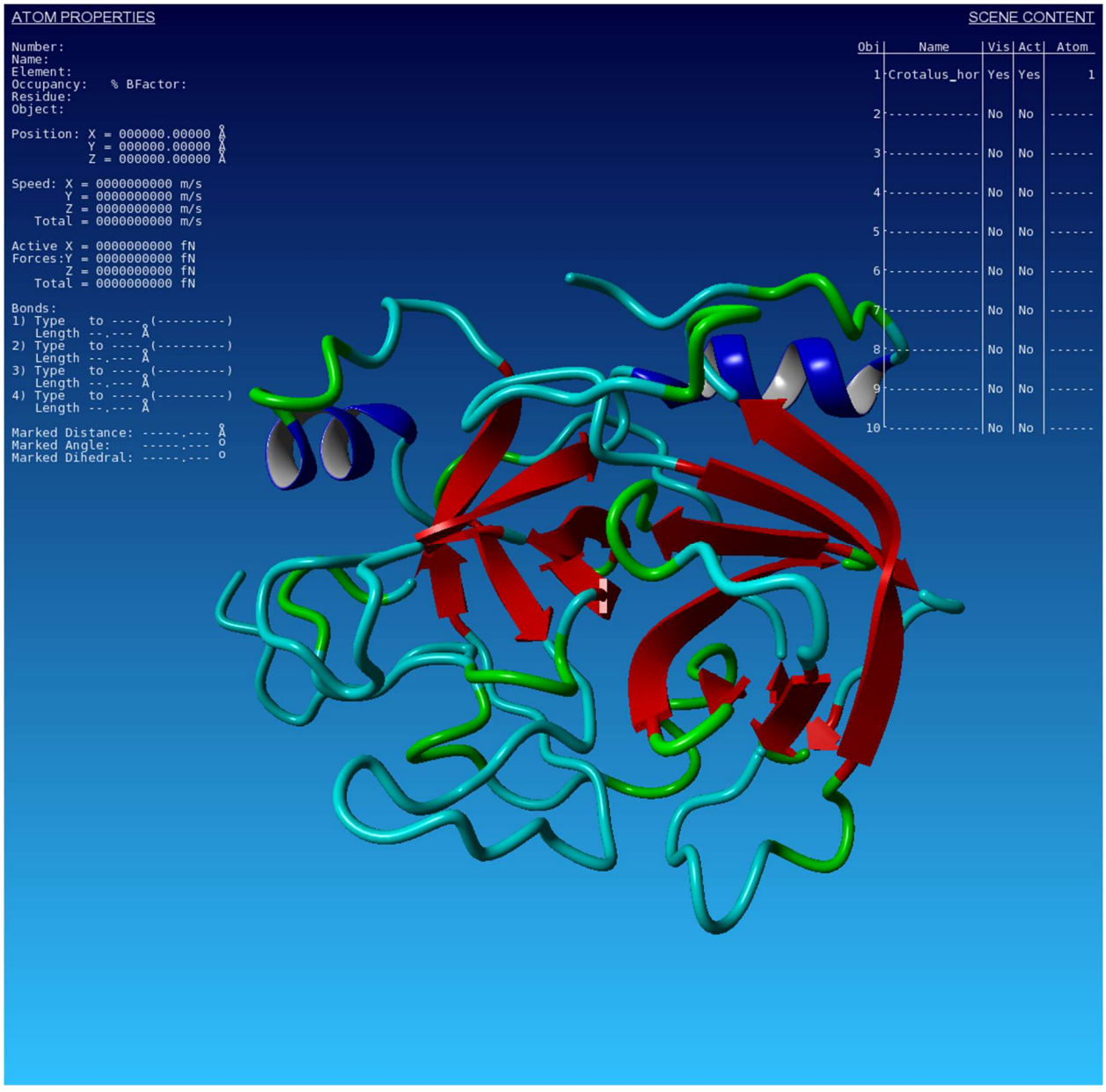
Slagboom J, Kool J, Harrison RA, Casewell NR. 2017. Haemotoxic snake venoms: their functional activity, impact on snakebite victims and pharmaceutical promise. British Journal of Haematology. 177(6):947-959.

Williamson AR. 2000. Creating a structural genomics consortium. Nature Structural Biology 7, 953.

APPENDIX A: FIGURES

Figure 1

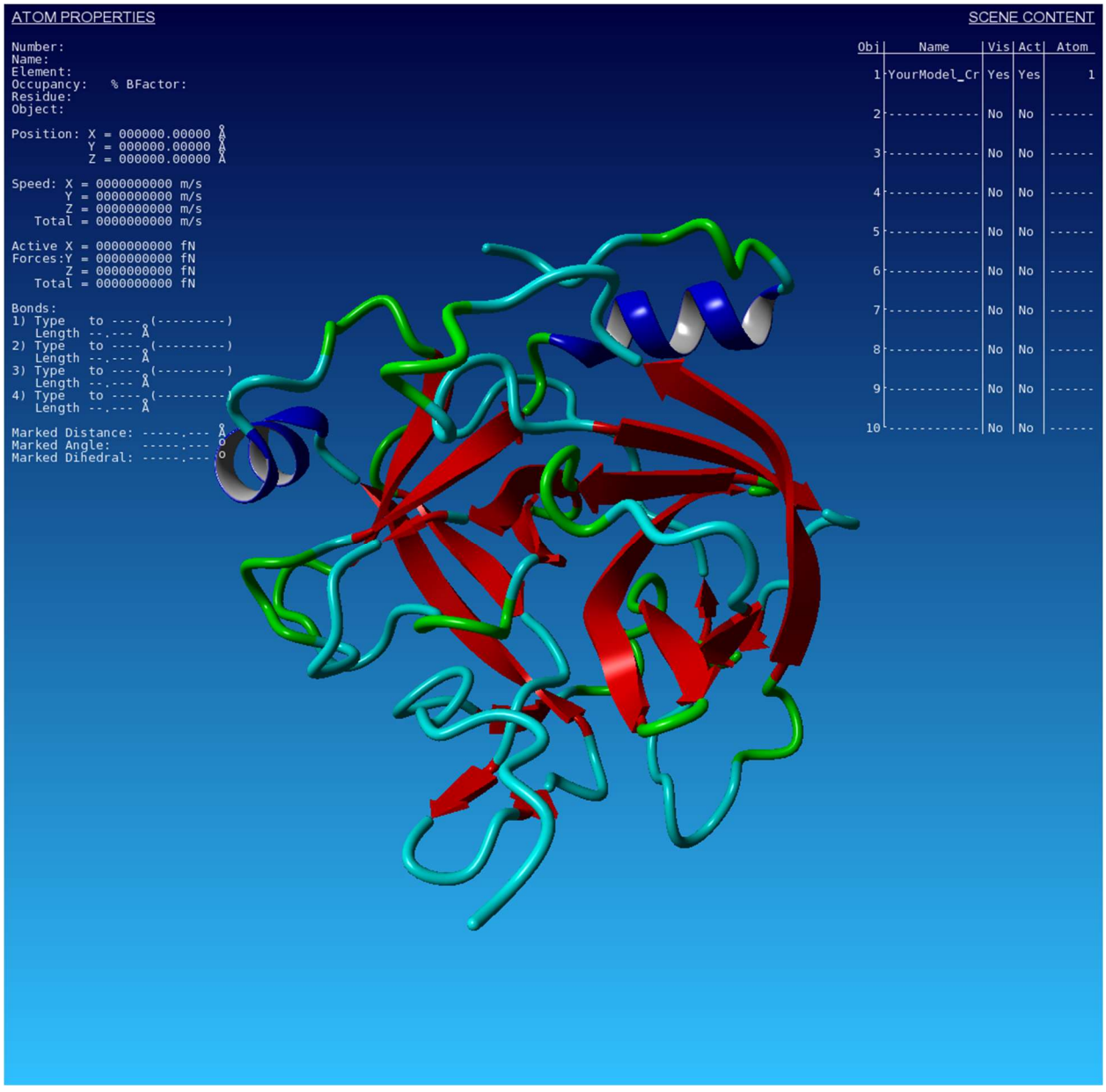
Ribbon Model of *Crotalus horridus* Protein



Note: The ribbon model of *Crotalus horridus* modelled using YASARA showing the secondary structure of the protein.

Figure 2

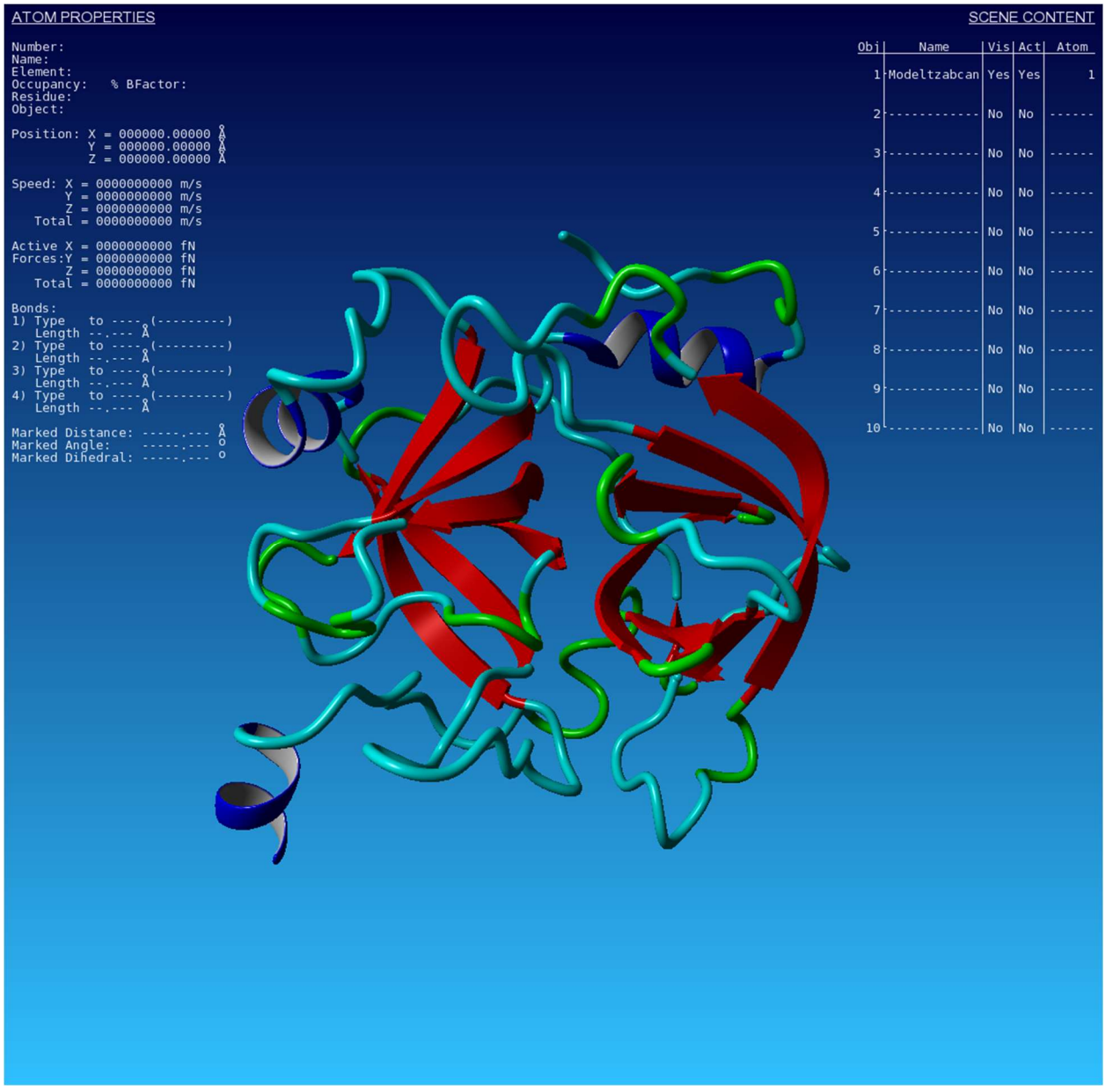
Ribbon Model of *Crotalus tigris* Protein



Note: The ribbon model of *Crotalus tigirs* modelled using YASARA showing the secondary structure of the protein.

Figure 3

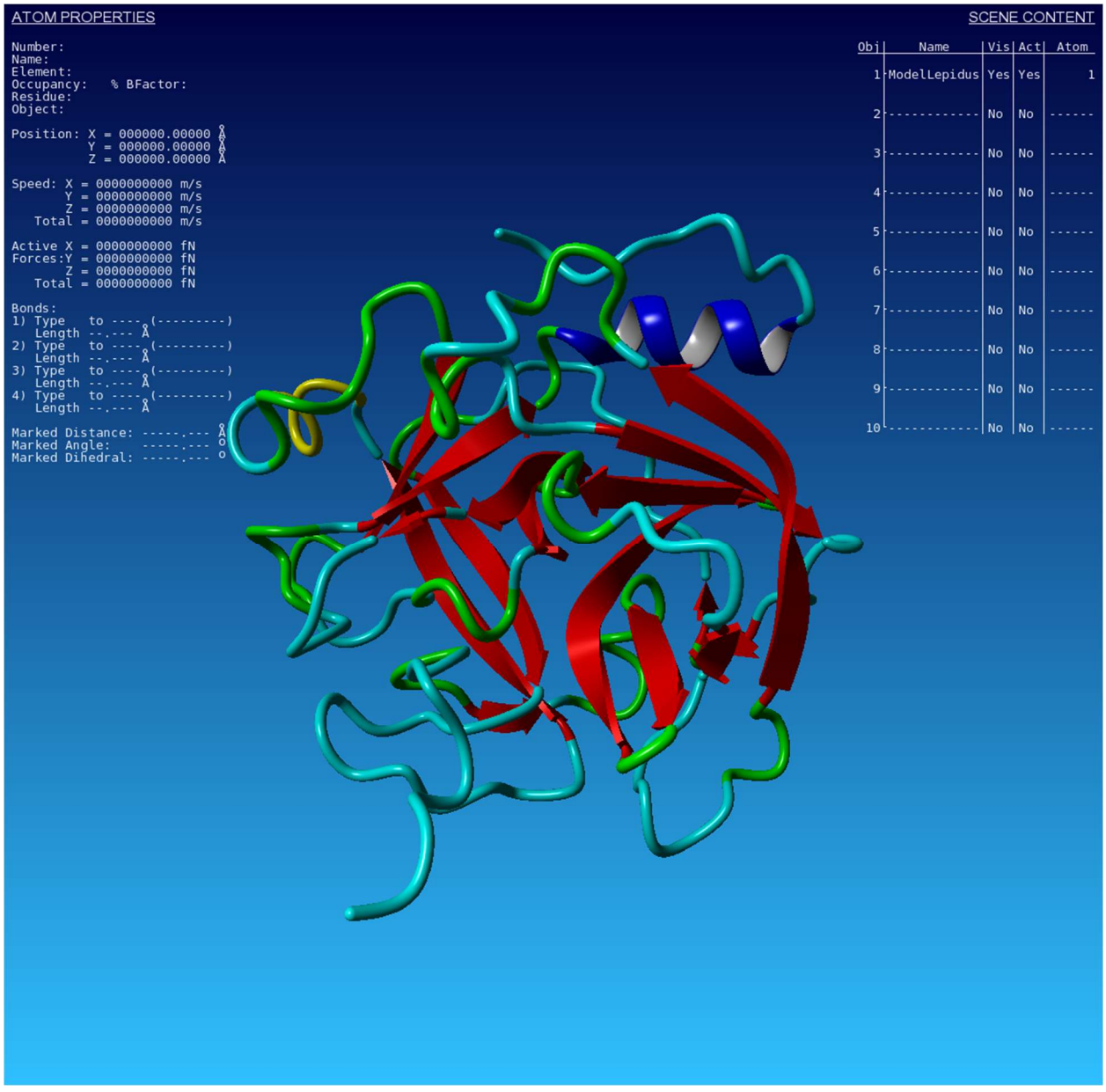
Ribbon Model of *Crotalus tzabcan* Protein



Note: The ribbon model of *Crotalus tzabcan* modelled using YASARA showing the secondary structure of the protein.

Figure 4

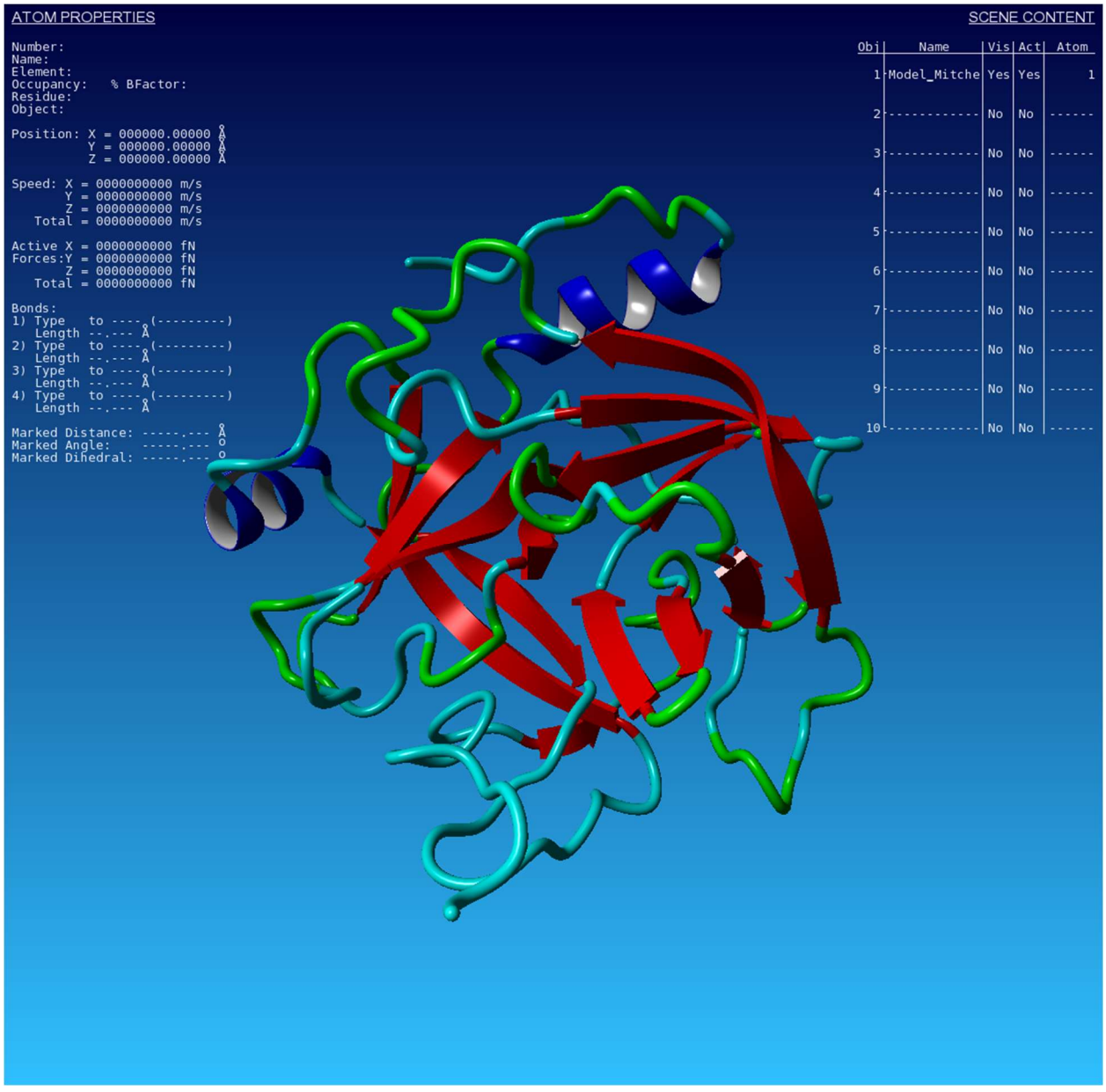
Ribbon Model of *Crotalus lepidus* Protein



Note: The ribbon model of *Crotalus lepidus* modelled using YASARA showing the secondary structure of the protein.

Figure 5

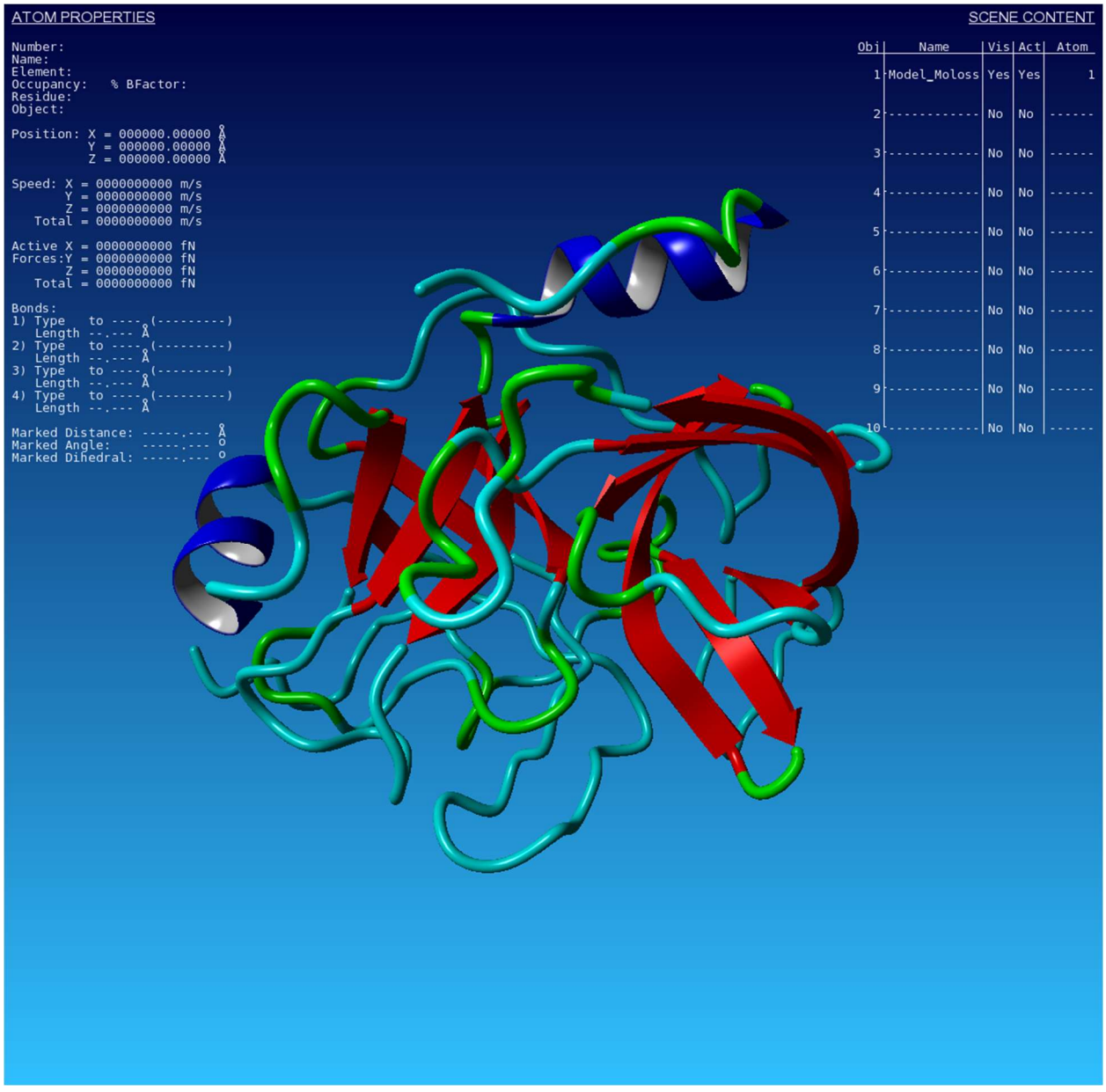
Ribbon Model of *Crotalus mictchellii* Protein



Note: The ribbon model of *Crotalus mictchellii* modelled using YASARA showing the secondary structure of the protein.

Figure 6

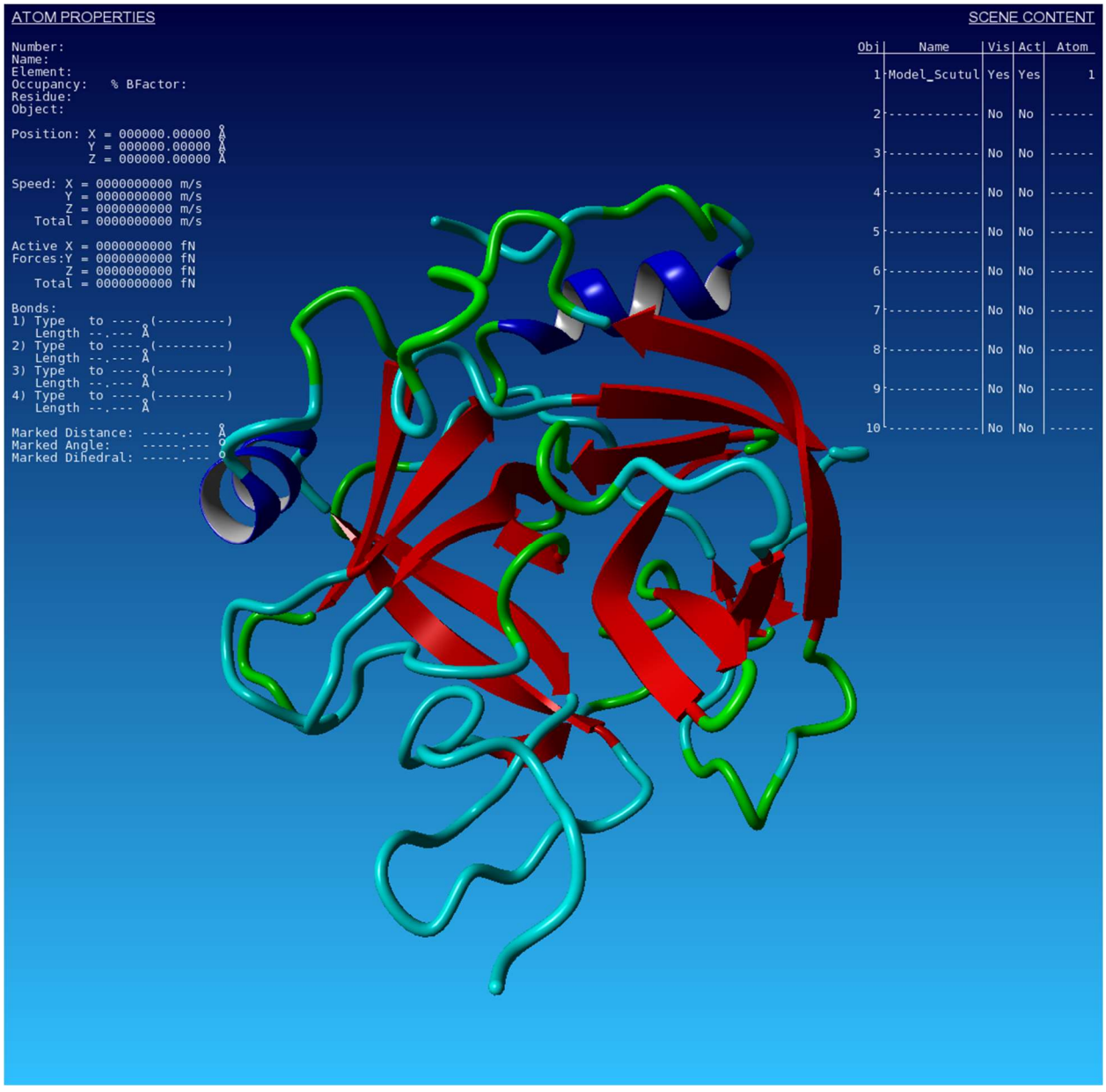
Ribbon Model of *Crotalus molossus* Protein



Note: The ribbon model of *Crotalus molossus* modelled using YASARA showing the secondary structure of the protein.

Figure 7

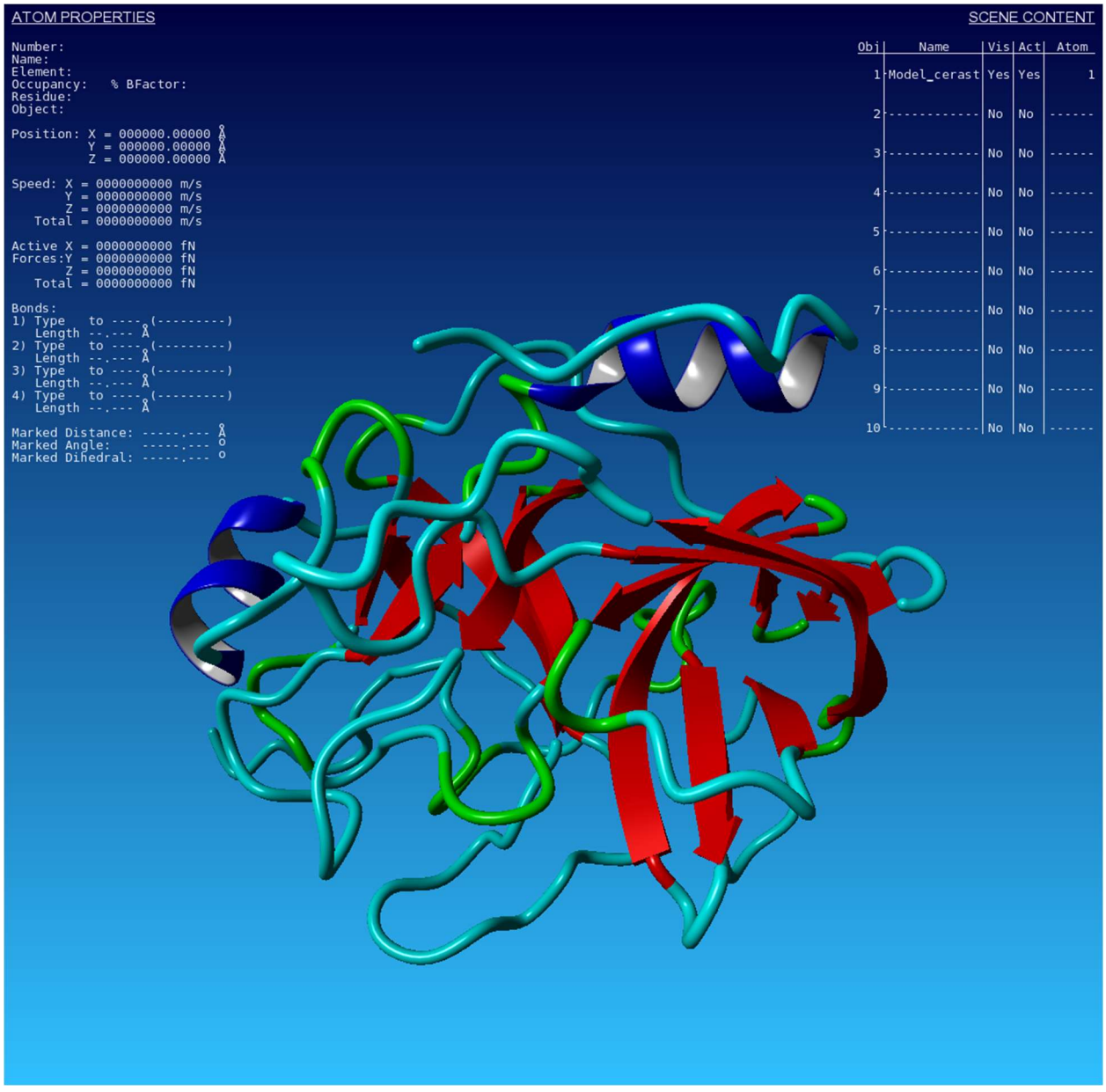
Ribbon Model of *Crotalus scutulatus* Protein



Note: The ribbon model of *Crotalus scutulatus* modelled using YASARA showing the secondary structure of the protein.

Figure 8

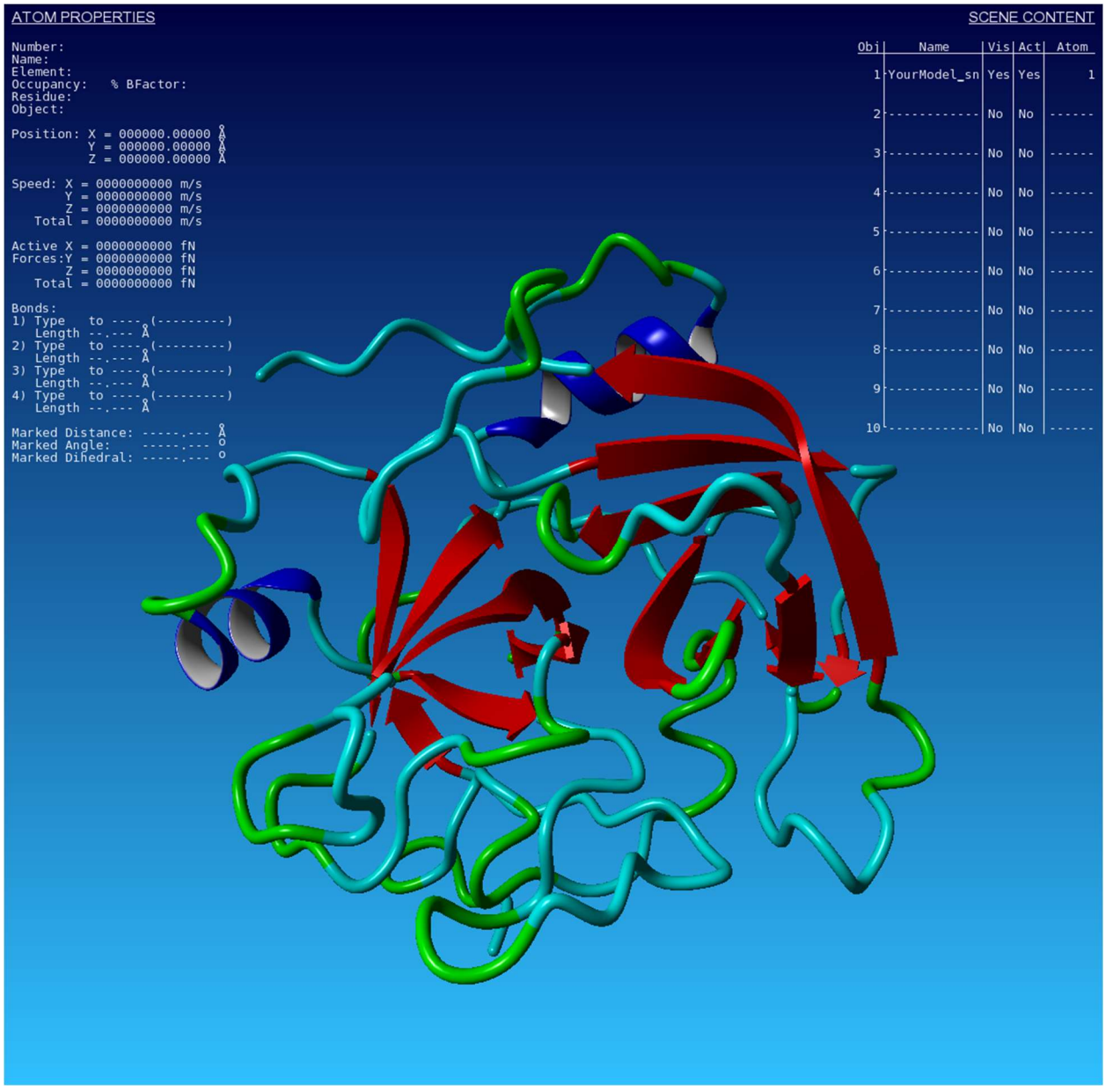
Ribbon Model of *Crotalus cerastes* Protein



Note: The ribbon model of *Crotalus cerastes* modelled using YASARA showing the secondary structure of the protein.

Figure 9

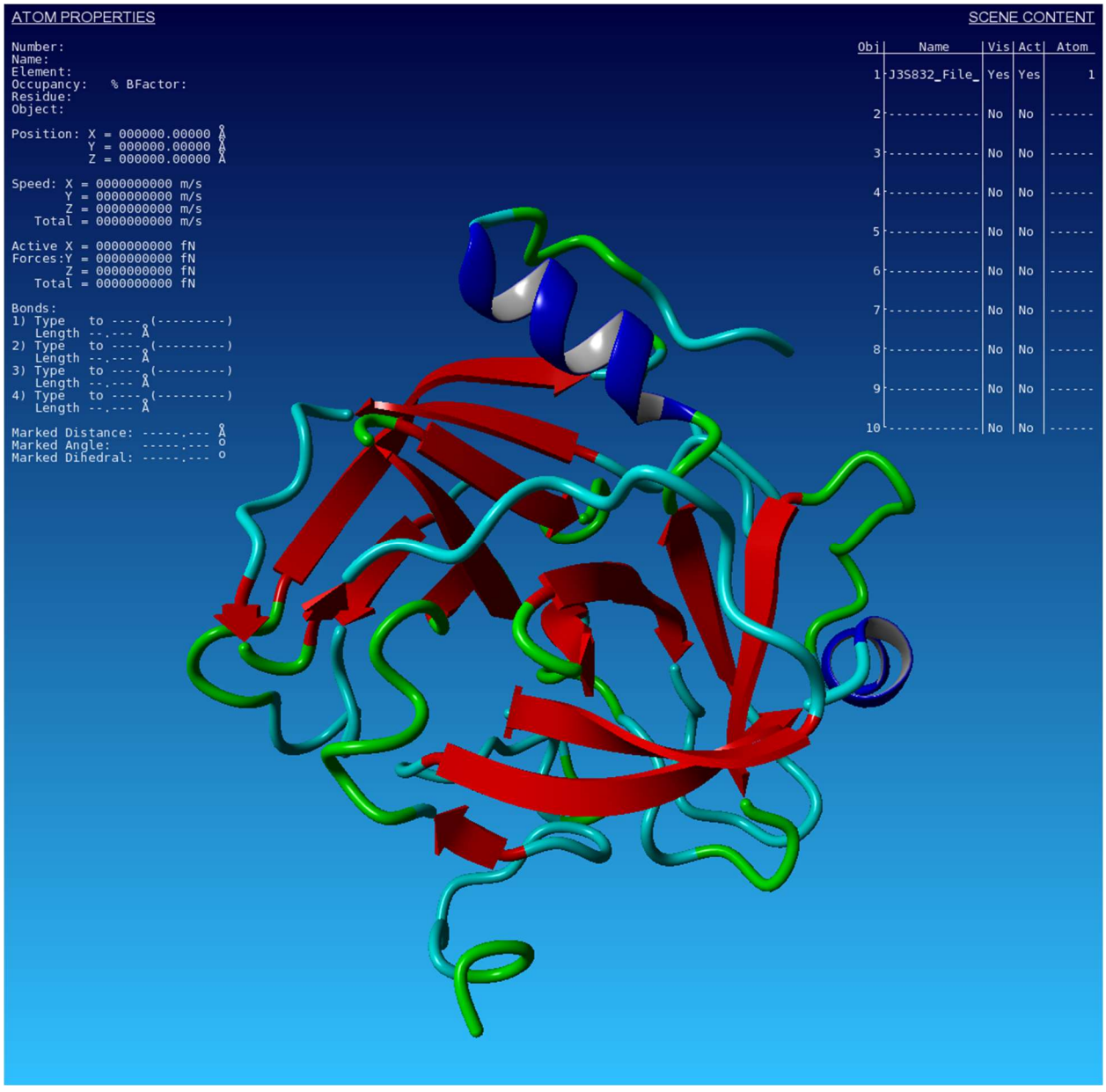
Ribbon Model of *Crotalus atrox* protein



Note: The ribbon model of *Crotalus atrox* modelled using YASARA showing the secondary structure of the protein.

Figure 10

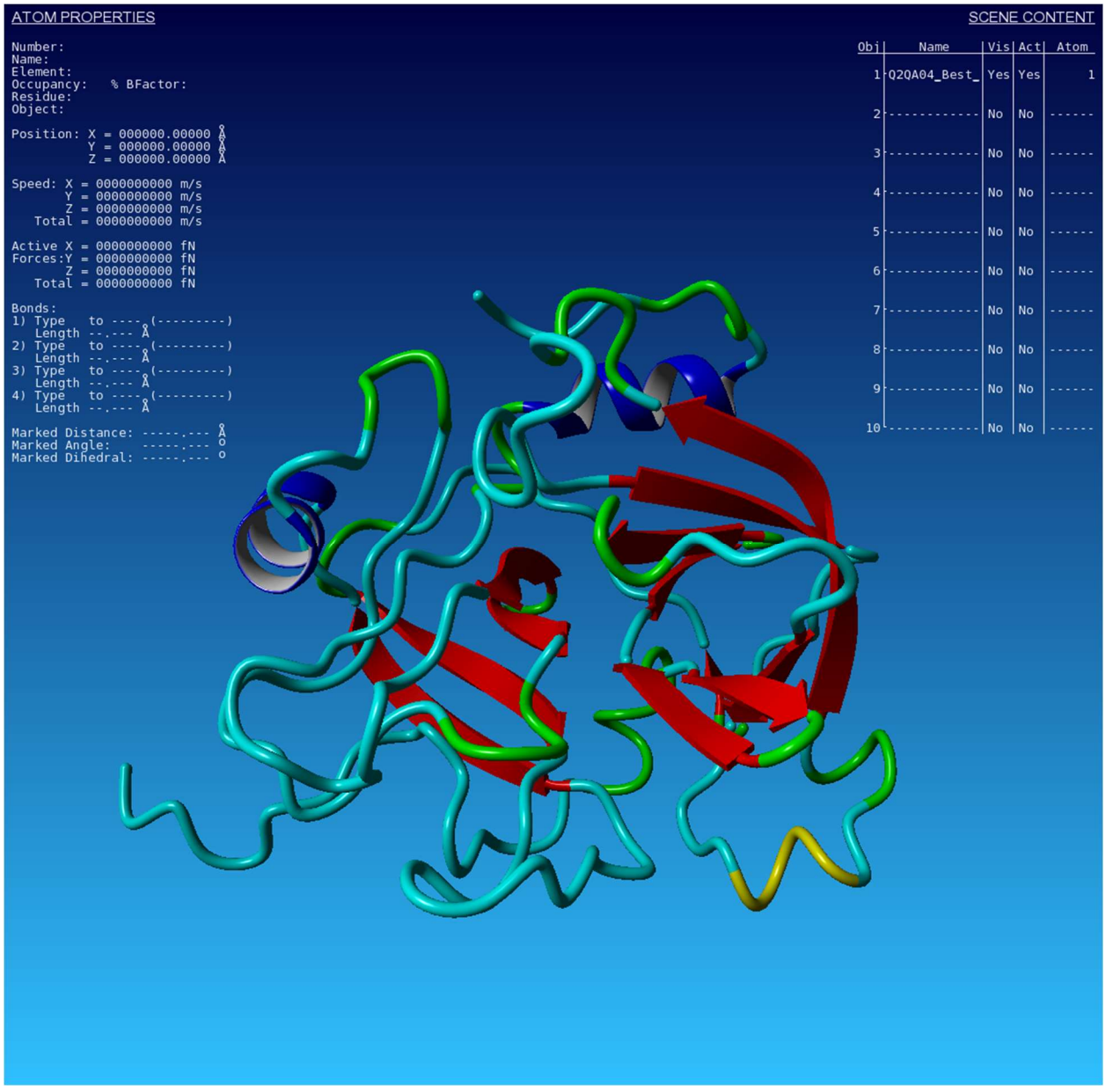
Ribbon Model of *Crotalus adamanteus* Protein



Note: The ribbon model of *Crotalus adamanteus* modelled using YASARA showing the secondary structure of the protein.

Figure 11

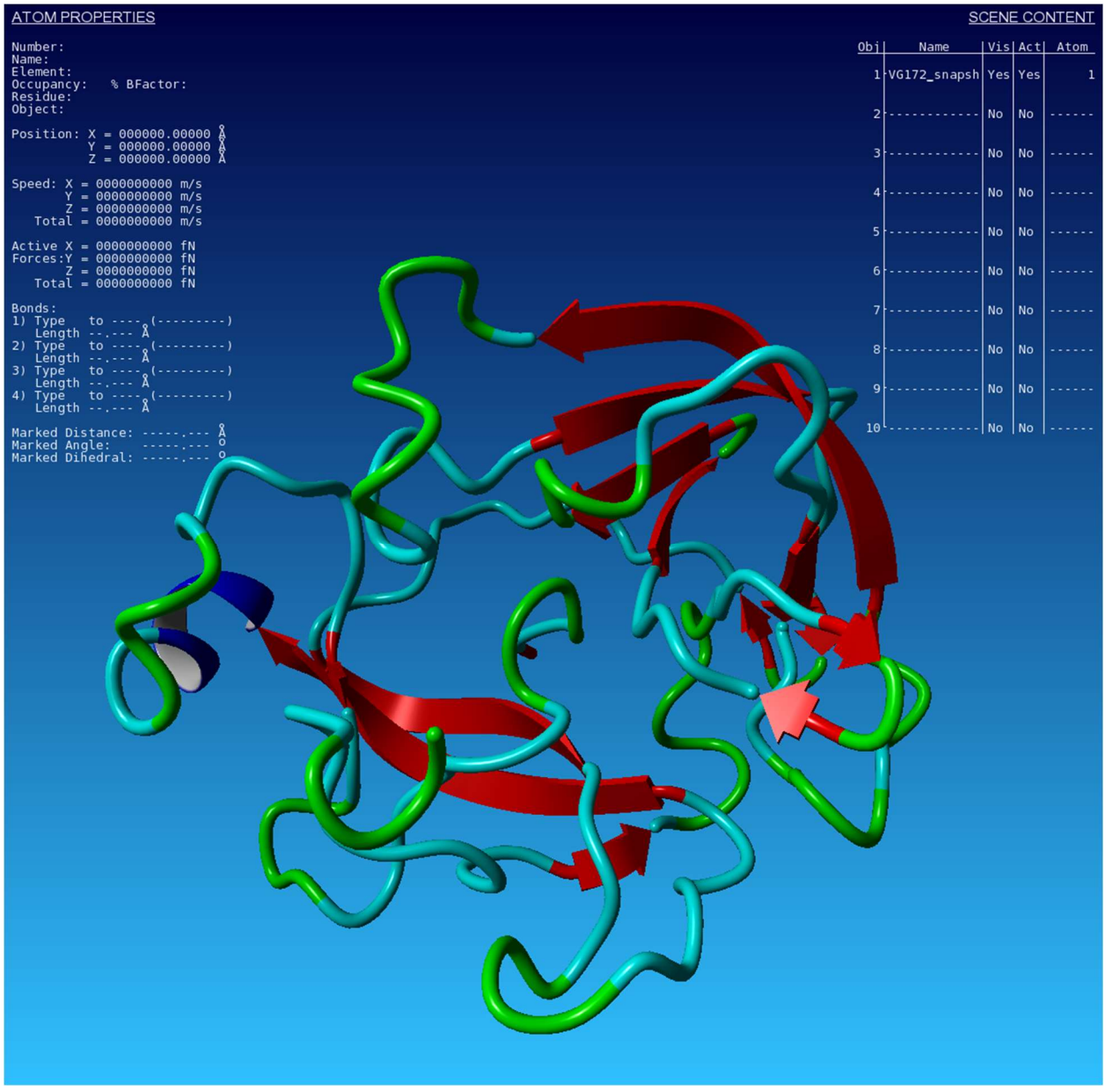
Ribbon Model of *Crotalus durissus* Protein



Note: The ribbon model of *Crotalus durissus* modelled using YASARA showing the secondary structure of the protein.

Figure 12

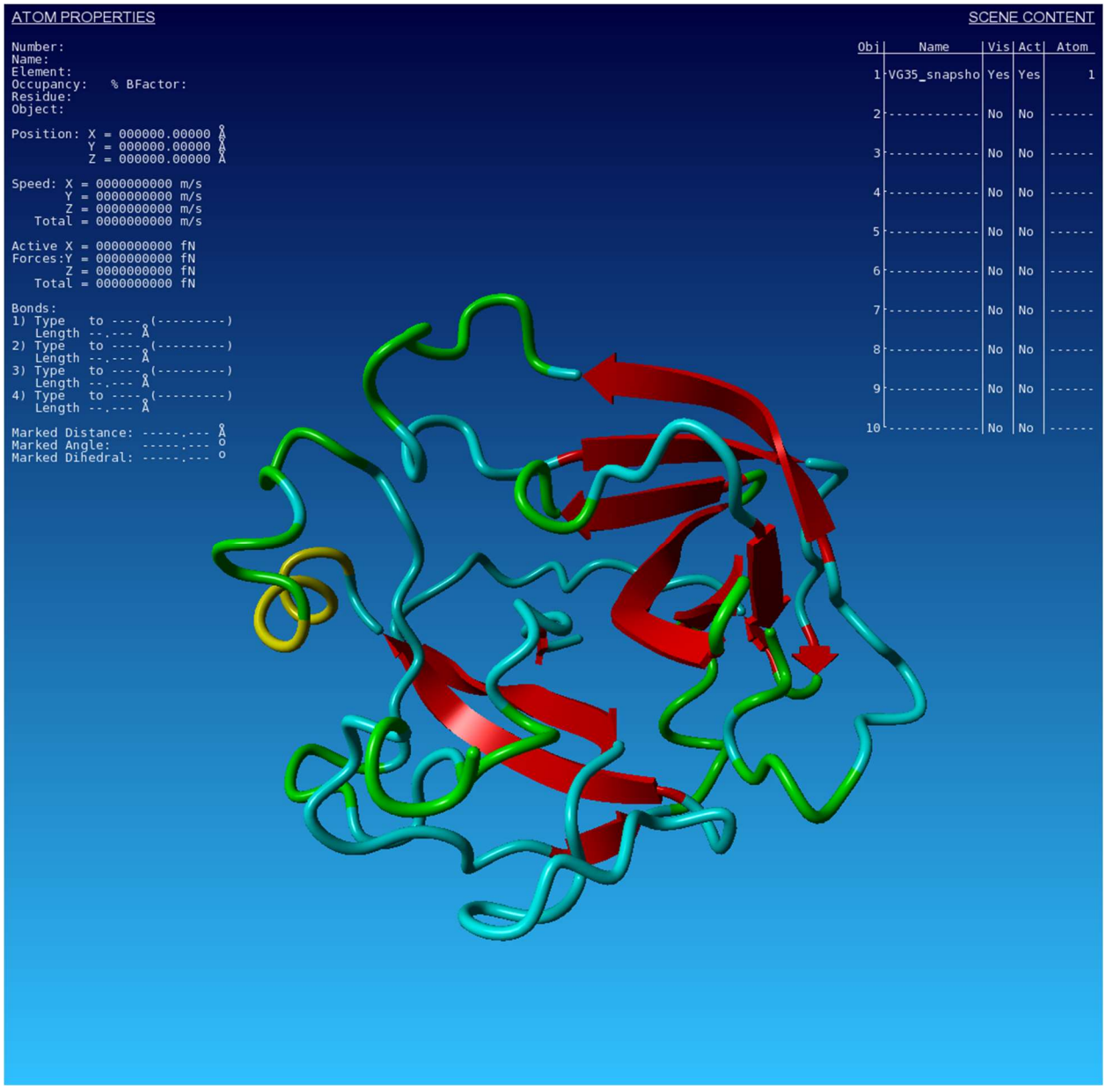
Ribbon Model of VG172 Protein



Note: The ribbon model of VG172 modelled using YASARA showing the secondary structure of the protein.

Figure 13

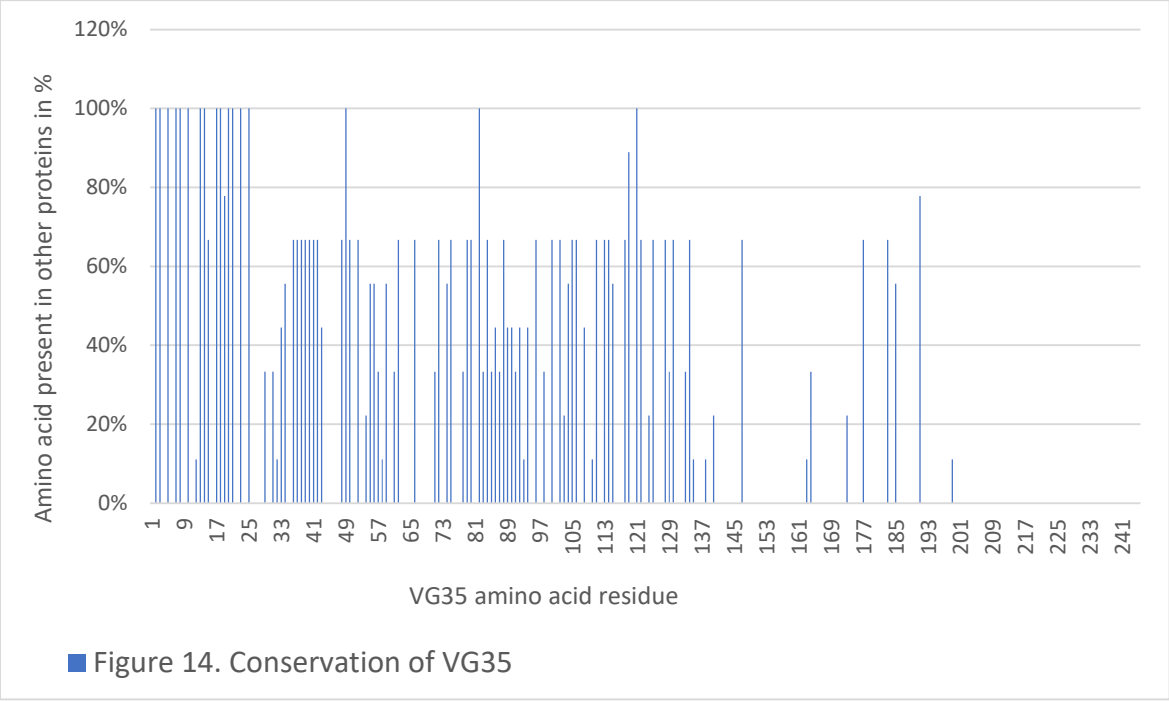
Ribbon Model of VG35 Protein



Note: The ribbon model of VG35 modelled using YASARA showing the secondary structure of the protein.

Figure 14

Conservation of VG35



APPENDIX B: TABLES

Table 1

Evolutionary Comparison

Note: The secondary structure of serine protease (black = coil, white = turn, blue = sheet, yellow = helix).

Residue	VG172	VG35	adamanteus	horridus	Cerastes	mitchellii	lepidus	molossus	scutulatus	tigris	tzabcan
15				ILE	LEU	LEU	LEU	LEU	LEU	LEU	
16	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
17	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR	TYR
18	PRO	PRO	ALA	ALA	GLY	ALA	ALA	ALA	ALA	ALA	ALA
19	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN	GLN
20	ASN	ASN	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS	LYS
21	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
22	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	SER
23	ASP	ASP	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU
24	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
25	ALA	ALA	VAL	ILE	VAL	VAL	VAL	VAL	VAL	VAL	ILE
26	PHE	PHE	VAL	ILE	THR	ILE	VAL	VAL	ILE	VAL	PHE
27	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
28	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
29	ASP	ASP	ASP	ASP	ASP	ASP	ASP	ARG	ASP	ARG	ARG
30	LYS	LYS	GLU	GLU	GLU	GLU	GLU	PRO	GLU	PRO	PRO
31	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
32	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN
33	ILE	ILE	ILE	LYS	ILE	ILE	ILE	ILE	ILE	ILE	ARG
34	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN	ASN
35	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU	GLU

36	LEU	LEU	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS	HIS
37	ARG	ARG	ARG	ARG	ARG	ARG	ARG	ARG	ARG	ARG	ARG
38	PRO	PRO	SER	PHE	SER	SER	SER	SER	SER	SER	PHE
39	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU	LEU
40	ALA	GLU	VAL	ALA	ALA	ALA	VAL	VAL	ALA	VAL	ALA
41	PRO	PRO	LEU	LEU	LEU	LEU	ALA	VAL	LEU	VAL	LEU
42	GLY	PHE	VAL	VAL	VAL	VAL	ILE	LEU	VAL	LEU	VAL
43	TYR	PHE	TYR	TYR	TYR	TYR	PHE	PHE	TYR	PHE	TYR
44	ILE	THR	SER	SER	ILE	ILE	ASN	ASN	ILE	ASN	SER
45	THR	SER	ASP	ASP	THR	THR	SER	SER	THR	SER	ASP
46	LYS	THR	GLY	GLY	ASN	SER	THR	SER	SER	SER	GLY
47	GLY	GLY	ILE	ASN	ASP	GLY	GLU	GLY	GLY	GLY	ASN
48	PHE	PHE	GLN	GLN	GLY	PHE	PHE	PHE	PHE	PHE	GLN
49	PRO	SER	CYS	CYS	PHE	LEU	PHE	LEU	LEU	LEU	CYS
50	CYS	CYS	GLY	GLY	CYS	CYS	CYS	CYS	CYS	CYS	SER
51	PRO	GLY	GLY	GLY	GLY	GLY	SER	ALA	GLY	ALA	GLY
52	GLY	GLY	THR	THR	GLY	GLY	GLY	GLY	GLY	GLY	THR
53	THR	THR	LEU	LEU	THR	THR	THR	THR	THR	THR	LEU
54	LEU	LEU	ILE	ILE	LEU	LEU	LEU	LEU	LEU	LEU	ILE
55	ILE	ILE	ASN	ASN	ILE	ILE	ILE	ILE	ILE	ILE	ASN
56	ASN	ASN	GLN	GLU	ASN	ASN	ASN	ASN	ASN	ASN	LYS
57	LYS	GLN	GLU	GLU	GLN	GLU	GLN	GLN	GLU	GLN	GLU
58	LYS	LYS	TRP	TRP	GLU	GLU	GLU	GLU	GLU	GLU	TRP
59	GLY	GLY	MET	VAL	TRP	TRP	TRP	TRP	TRP	TRP	VAL
60	GLY	GLY	LEU	LEU	VAL	VAL	VAL	VAL	VAL	VAL	LEU
61	LEU	ALA	THR	THR	LEU	LEU	VAL	LEU	LEU	LEU	THR
62	THR	THR	ALA	ALA	THR	THR	THR	THR	THR	THR	ALA
63	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA	ALA
64	ALA	ALA	HIS	HIS	ALA	ALA	ALA	ALA	ALA	ALA	HIS
65	PRO	LEU	CYS	CYS	HIS	HIS	HIS	HIS	HIS	HIS	CYS
66	CYS	CYS	ASP	GLU	CYS	CYS	CYS	CYS	CYS	CYS	GLU
67	ASN	ASN	GLY	ARG	ASP	ASP	ASP	LYS	ASP	ASP	GLY
68	GLY	ARG	LYS	ASN	LYS	ARG	SER	ILE	ARG	ILE	LYS
69	GLU	LYS	LYS	LYS	GLY	GLY	THR	LYS	GLY	LYS	LYS
70	ASN	ASN	MET	MET	ASP	ASN	ASN	ASN	ASN	ASN	MET
71	PHE	PHE	LYS	LYS	MET	MET	PHE	PHE	MET	PHE	LYS
72	PRO	LYS	LEU	ILE	LEU	LEU	LYS	GLN	LEU	GLN	ILE
73	PHE	ILE	GLN	HIS	ILE	ILE	MET	ILE	ILE	ILE	HIS

74	LEU	ASN	PHE	LEU	PHE	PHE	LYS	GLN	PHE	GLN	LEU
75	LEU	LEU	GLY	GLY	LEU	PHE	PHE	LEU	PHE	LEU	GLY
76	GLY	GLY	LEU	VAL	GLY	GLY	GLY	GLY	GLY	GLY	VAL
77	GLY	GLY	HIS	HIS	MET	VAL	MET	VAL	VAL	VAL	HIS
78	LEU	LEU	SER	SER	HIS	HIS	HIS	HIS	HIS	HIS	SER
79	THR	THR	LYS	LYS	SER	ARG	SER	SER	ARG	SER	LYS
80	LEU	LYS	LYS	LYS	LEU	LEU	LYS	LYS	LEU	LYS	LYS
81	ARG	ARG	VAL	VAL	GLU	LYS	LYS	LYS	LYS	LYS	VAL
82	GLU	GLU	PRO	PRO	VAL	GLY	VAL	VAL	GLY	VAL	PRO
83	LEU	SER	ASN	ASN	PRO	LEU	PRO	ARG	LEU	AR G	ASN
84	ILE	ILE	LYS	LYS	ASN	ASN	ASN	ASN	ASN	ASN	LYS
85	LYS	GLU	ASP	ASP	LYS	LYS	GLU	GLU	LYS	GLU	ASP
86	ASP	ASP	LYS	LYS	ASP	ASP	ASP	ASP	ASP	ASP	LYS
87	VAL	ASP	GLN	GLN	VAL	VAL	GLU	GLU	VAL	GLU	GLN
88	GLN	GLN	THR	THR	GLN	GLN	GLN	GLN	ASN	GLN	THR
89	LYS	THR	ARG	ARG	THR	THR	THR	THR	THR	THR	ARG
90	LYS	LYS	VAL	VAL	ARG	ARG	ARG	ARG	ARG	AR G	VAL
91	ALA	THR	PRO	PRO	VAL	VAL	ASN	ASP	VAL	ASP	ALA
92	PRO	PRO	LYS	LYS	ALA	ALA	PRO	PRO	ALA	PR O	LYS
93	LYS	LYS	GLU	GLU	LYS	LYS	LYS	LYS	LYS	LYS	GLU
94	GLU	GLU	LYS	LYS	GLU	GLU	GLU	GLU	GLU	GLU	LYS
95	ASN	ASN	PHE	PHE	LYS	LYS	LYS	LYS	LYS	LYS	PHE
96	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE	PHE
97	PHE	PHE	CYS	CYS	ILE	ILE	PHE	PHE	ILE	PHE	CYS
98	CYS	CYS	LEU	VAL	CYS	CYS	CYS	CYS	CYS	CYS	VAL
99	PRO	PRO	SER	SER	PRO	PRO	PRO	LEU	SER	LEU	SER
100	ILE	ASN	SER	SER	ASN	ASN	ASN	GLY	ASN	GLY	SER
101	ARG	LYS	LYS	GLU	ARG	ARG	LYS	SER	ARG	SER	LYS
102	LYS	LYS	ASN	ASN	LYS	LYS	LYS	LYS	LYS	LYS	ASN
103	LYS	LYS	ASN	TYR	LYS	LYS	LYS	THR	LYS	THR	TYR
104	ILE	ASN	LYS	THR	ASP	ASN	ASP	ASN	ASN	ASN	THR
105	ASP	ASP	GLU	PHE	ASN	ASP	ASP	ASN	ASP	ASN	LYS
106	GLU	GLU	TRP	TRP	LYS	GLU	VAL	GLU	GLU	GLU	TRP
107	LYS	LEU	ASP	ASP	LYS	LYS	LEU	TRP	LYS	TRP	ASN
108	ASP	ASP	LYS	LYS	ASP	ASP	ASP	GLU	ASP	GLU	LYS
109	ARG	ARG	ASP	ASP	LYS	LYS	LYS	LYS	LYS	LYS	ASP
110	ASP	ASP	ILE	ILE	ASP	ASP	ASP	ASP	ASP	ASP	ILE

111	PHE	PHE	MET	MET	ILE	ILE	ILE	ILE	ILE	ILE	MET
112	LEU	LEU	LEU	LEU	MET	MET	MET	MET	MET	MET	LEU
113	PHE	PHE	ILE	ILE	LEU	LEU	LEU	LEU	LEU	LEU	ILE
114	ILE	ILE	ARG	ARG	ILE	ILE	ILE	ILE	ILE	ILE	ARG
115	ARG	SER	LEU	LEU	ARG	ARG	LYS	ARG	ARG	ARG	LEU
116	PHE	LEU	ASN	ASP	LEU	LEU	LEU	LEU	LEU	LEU	ASP
117	ASN	ASN	ARG	ARG	ASP	ASP	ASP	ASN	ASP	ASN	ARG
118	SER	SER	PRO	PRO	SER	SER	SER	ASN	SER	ASN	SER
119	PRO	PRO	VAL	VAL	PRO	PRO	PRO	PRO	PRO	PRO	VAL
120	VAL	VAL	ASN	SER	VAL	VAL	VAL	VAL	VAL	VAL	SER
121	THR	THR	ASN	ASN	ASN	ASN	SER	SER	ASN	SER	ASN
122	SER	ASN	SER	SER	ASN	SER	ASN	ASN	SER	ASN	SER
123	SER	ARG	LYS	GLU	SER	SER	SER	SER	SER	SER	GLU
124	THR	GLU	HIS	HIS	THR	THR	GLU	ALA	THR	ALA	HIS
125	HIS	HIS	ILE	ILE	HIS	HIS	HIS	HIS	HIS	HIS	ILE
126	PHE	PHE	ALA	ALA	ILE	ILE	ILE	ILE	ILE	ILE	ALA
127	ALA	ALA	PRO	PRO	ALA	ALA	ALA	ALA	ALA	ALA	PRO
128	PRO	PRO	LEU	LEU	PRO	PRO	PRO	PRO	PRO	PRO	LEU
129	LEU	LEU	SER	SER	LEU	LEU	LEU	LEU	VAL	LEU	LYS
130	SER	THR	LEU	LEU	SER	SER	SER	SER	SER	SER	LEU
131	TRP	TRP	PRO	PRO	LEU	LEU	LEU	LEU	LEU	LEU	PRO
132	PRO	PRO	SER	SER	PRO	PRO	PRO	PRO	PRO	PRO	SER
133	SER	SER	LYS	SER	SER	SER	SER	SER	SER	SER	SER
134	ASN	THR	PRO	PRO	ASN	ASN	SER	SER	ASN	SER	SER
135	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
136	PRO	PRO	SER	SER	PRO	PRO	PRO	PRO	PRO	PRO	SER
137	ARG	ARG	GLN	VAL	SER	SER	SER	SER	SER	SER	VAL
138	VAL	GLY	ASP	GLY	VAL	VAL	VAL	VAL	VAL	VAL	GLY
139	GLY	GLY	THR	SER	GLY	GLY	GLY	GLY	GLY	GLY	SER
140	PRO	PRO	VAL	VAL	SER	SER	SER	SER	SER	SER	VAL
141	VAL	PHE	CYS	CYS	VAL	VAL	VAL	LEU	VAL	LEU	CYS
142	CYS	CYS	ASN	ARG	CYS	CYS	CYS	CYS	CYS	CYS	ARG
143	ARG	ARG	ILE	ILE	ARG	HIS	HIS	ARG	GLN	ARG	ILE
144	ILE	ILE	MET	MET	ILE	ILE	ILE	ILE	ILE	ILE	MET

145	MET	THR	GLY	GLY	MET	MET	MET	MET	MET	MET	GLY
146	GLU	GLU	TRP	TRP	GLY	GLY	GLY	GLY	GLY	GLY	TRP
147	GLY	GLY	GLY	GLY	TRP	TRP	TRP	TRP	TRP	TRP	GLY
148	GLY	GLY	THR	THR	GLY	GLY	GLY	GLY	GLY	GLY	THR
149	ALA	SER	ILE	ILE	THR	ALA	SER	THR	ALA	THR	ILE
150	ILE	PHE	SER	SER	ILE	ILE	ILE	ILE	ILE	ILE	SER
151	PRO	HIS	PRO	PRO	THR	THR	THR	SER	THR	SER	PRO
152	PRO	PRO	THR	THR	SER	SER	PRO	SER	SER	SER	THR
153	PRO	GLN	GLU	GLU	PRO	PRO	ILE	THR	PRO	THR	GLU
154	LYS	LYS	GLU	VAL	ASN	ASN	GLU	LYS	ASN	LYS	VAL
155	GLY	ASN	THR	ILE	VAL	VAL	LYS	GLU	VAL	GLU	ILE
156	THR	ASP	TYR	LEU	THR	THR	THR	ILE	THR	ILE	LEU
157	LEU	PHE	PRO	PRO	LEU	LEU	LEU	TYR	LEU	TYR	PRO
158	PRO	ALA	ASP	ASP	PRO	PRO	PRO	PRO	PRO	PRO	ASP
159	GLY	GLN	VAL	VAL	ALA	GLY	ASP	ASN	GLY	ASN	VAL
160	ALA	CYS	PRO	LEU	VAL	VAL	VAL	VAL	VAL	VAL	PRO
161	PRO	PRO	HIS	HIS	PRO	PRO	PRO	PRO	PRO	PRO	HIS
162	LEU	LEU	CYS	CYS	HIS	HIS	TYR	ARG	HIS	ARG	CYS
163	LEU	LEU	ALA	VAL	CYS	CYS	CYS	CYS	CYS	CYS	ALA
164	GLY	CYS	ASN	ASN	ALA	ALA	ALA	ALA	ALA	ALA	ASN
165	TYR	LYS	ILE	ILE	ASN	ASN	ASN	ASN	ASN	ASN	ILE
166	PHE	HIS	ASN	ASN	ILE	ILE	ILE	ILE	ILE	ILE	ASN
167	ASN	LYS	ILE	LEU	ASN	ASN	ASN	ASN	ASN	ASN	LEU
168	ILE	PRO	LEU	LEU	ILE	ILE	LEU	ILE	ILE	ILE	LEU
169	LEU	THR	ASP	ASN	VAL	LEU	LEU	LEU	LEU	LEU	ASN
170	ASN	GLN	HIS	TYR	ARG	ASP	ASP	ASP	ASP	ASP	TYR
171	TYR	TRP	ALA	SER	ASN	TYR	ASP	TYR	ASP	TYR	SER
172	LYS	GLY	VAL	VAL	ALA	GLU	ALA	ALA	GLU	ALA	GLU
173	GLY	ARG	CYS	CYS	LEU	VAL	VAL	VAL	VAL	VAL	CYS
174	GLY	GLY	ARG	ARG	CYS	CYS	CYS	CYS	CYS	CYS	ARG
175	GLN	CYS	ALA	ALA	ARG	ARG	GLN	ARG	ARG	ARG	ALA
176	PRO	HIS	ILE	ALA	LYS	ALA	PRO	ALA	ALA	ALA	ALA
177	ALA	PRO	TYR	TYR	ALA	ALA	PRO	ALA	ALA	ALA	TYR
178	TYR	PRO	PRO	PRO	TYR	TYR	TYR	TYR	TYR	TYR	PRO
179	ALA	LEU	GLY	GLN	ALA	ALA	PRO	PRO	ALA	PRO	GLU

180	ARG	SER	LEU	TYR	GLY	GLY	GLU	TRP	GLY	TRP	TYR
181	LEU	GLN	LEU	GLY	LEU	LEU	LEU	TRP	LEU	TRP	GLY
182	PRO	ILE	GLU	LEU	PRO	PRO	PRO	PRO	PRO	PRO	LEU
183	GLN	CYS	LYS	PRO	ALA	ALA	ALA	VAL	ALA	VAL	PRO
184	THR	PRO	SER	ALA	THR	THR	THR	THR	THR	THR	ALA
185	ARG	ALA	ARG	THR	SER	SER	SER	THR	SER	THR	THR
186	LYS	THR	VAL	SER	ARG	ARG	ARG	ARG	ARG	ARG	SER
187	LYS	THR	LEU	ARG	THR	ILE	THR	ILE	ILE	ILE	ARG
188	LEU	LYS	CYS	THR	LEU	LEU	LEU	LEU	LEU	LEU	THR
189	GLY	HIS	ALA	LEU	CYS	CYS	CYS	CYS	CYS	CYS	LEU
190	CYS	ILE	GLY	CYS	ALA	ALA	ALA	ALA	ALA	ALA	CYS
191	ARG	GLY	ILE	ALA	GLY	GLY	GLY	GLY	GLY	GLY	ALA
192	ILE	VAL	LEU	GLY	ILE	ILE	ILE	ILE	ILE	ILE	GLY
193	PRO	GLN	GLU	ILE	LEU	LEU	PRO	LEU	LEU	LEU	ILE
194	LYS	VAL	GLY	LEU	GLN	GLU	GLU	GLU	GLU	GLU	LEU
195	ARG	SER	GLY	GLU	GLY	GLY	GLY	GLY	GLY	GLY	GLU
196	GLY	TRP	LYS	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
197	LYS	LYS	ASP	GLY	LYS	LYS	LYS	LYS	LYS	LYS	GLY
198	SER	GLY	THR	LYS	ASP	SER	ASP	ASP	SER	ASP	LYS
199	SER	SER	CYS	ASP	SER	SER	THR	SER	SER	SER	ASP
200	CYS	ASN	GLY	THR	CYS	CYS	CYS	CYS	CYS	CYS	THR
201	LYS	GLU	GLY	CYS	GLN	ASP	GLY	GLN	ASP	GLN	CYS
202	GLY	TYR	ASP	VAL	GLY	GLY	GLY	GLY	GLY	GLY	VAL
203	ASP	LEU	SER	GLY	ASP	ASP	ASP	ASP	ASP	ASP	GLY
204	SER	TRP	GLY	ASP	SER	SER	SER	SER	SER	SER	ASP
205	GLY	GLY	GLY	SER	GLY	GLY	GLY	GLY	GLY	GLY	SER
206	GLY	TRP	PRO	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY
207	PRO	THR	LEU	GLY	PRO	PRO	PRO	PRO	PRO	PRO	GLY
208	LEU	SER	ILE	PRO	LEU	LEU	LEU	LEU	LEU	LEU	PRO
209	LEU	GLY	CYS	LEU	ILE	ILE	ILE	ILE	ILE	ILE	LEU
210		GLU	ASN	ILE	CYS	CYS	CYS	CYS	CYS	CYS	ILE
211		PRO	GLY	CYS	ASN	ASN	ASN	ASN	ASN	ASN	CYS
212		PHE	GLU	ASP	GLY	GLY	GLY	GLY	GLY	GLY	ASN
213		ILE	ILE	GLY	GLN	GLU	GLN	GLN	GLU	GLU	GLY
214		CYS	GLN	GLU	ILE	ILE	PHE	PHE	ILE	ILE	GLN
215		ASN	GLY	PHE	GLN	GLN	GLN	GLN	GLN	GLN	PHE
216		GLY	LEU	GLN	GLY	GLY	GLY	GLY	GLY	GLY	GLN

217	LEU	GLY	ILE	ILE	ILE	ILE	ILE	ILE	GLY
218	SER	ILE	VAL	VAL	VAL	VAL	VAL	VAL	ILE
219	VAL	ALA	SER	SER	PHE	SER	SER	SER	ALA
220	GLY	SER	GLY	TRP	TYR	TRP	TRP	TRP	SER
221	GLY	TRP	GLY	GLY	GLY	GLY	GLY	GLY	TRP
222	ASP	GLY	VAL	GLY	ALA	ALA	GLY	ALA	GLY
223	PRO	SER	HIS	ASP	HIS	HIS	ASP	HIS	SER
224	CYS	PRO	PRO	ILE	PRO	PRO	ILE	PRO	THR
225	ALA	ASN	CYS	CYS	CYS	CYS	CYS	CYS	LEU
226	GLN	CYS	GLY	ALA	GLY	GLY	ALA	GLY	CYS
227	PRO	GLY	GLN	GLN	GLN	ARG	GLN	ARG	GLY
228	HIS	TYR	ARG	PRO	ALA	ARG	PRO	ARG	TYR
229	VAL	VAL	LEU	HIS	LEU	LEU	HIS	LEU	VAL
230	PRO	GLY	LYS	GLU	LYS	ASN	GLU	ASN	ARG
231	ALA	GLU	PRO	PRO	PRO	PRO	PRO	PRO	GLU
232	LEU	PRO	GLY	GLY	GLY	GLY	GLY	GLY	PRO
233	TYR	ALA	PHE	VAL	VAL	PHE	VAL	PHE	ALA
234	ILE	LEU	TYR	TYR	TYR	TYR	TYR	TYR	LEU
235	LYS	PHE	THR	THR	THR	THR	THR	THR	TYR
236	VAL	THR	LYS	LYS	LYS	LYS	LYS	LYS	THR
237	PHE	LYS	VAL	VAL	VAL	VAL	VAL	VAL	LYS
238	ASP	VAL	PHE	PHE	PHE	PHE	PHE	PHE	VAL
239	TYR	PHE	ASP	ASP	ASP	ASP	ASP	ASP	PHE
240	THR	ASP	TYR	TYR	TYR	TYR	TYR	TYR	ASP
241	GLU	HIS	ILE	THR	ASN	ILE	THR	ILE	HIS
242	TRP	LEU	ASP	ASP	ASP	ASP	ASP	ASP	LEU
243	ILE	ASP	TRP	TRP	TRP	TRP	TRP	TRP	ASP
244	GLN	TRP	ILE	ILE	ILE	ILE	ILE	ILE	TRP
245	SER	ILE	GLN	GLN	GLN	GLN	GLN	GLN	ILE
246	ILE	GLN	SER	ASN	SER	SER	ASN	SER	GLN
247	ILE	SER	ILE	ILE	ILE	ILE	ILE	ILE	SER
248	THR	ILE	VAL	ILE	ILE	ILE	ILE	THR	ILE
249	GLY	ILE	ALA	ALA	ALA	ALA	ALA	ALA	ILE
250	ASN	ALA	GLY	GLY	GLY	GLY	GLY	GLY	ALA
251	THR	GLY	ASN	ASN	ASN	ASN	ASN	ASN	GLY
252	ALA	ASN	THR	THR	THR	THR	THR	THR	ASN

253	ALA	LYS	THR	ASP	ALA	THR	ASP	THR	THR
254	THR	ASP	VAL	ALA	ALA	VAL	SER	VAL	ASP
255	CYS	ALA	THR	THR	THR	THR	THR	THR	ALA
256	PRO	THR	CYS	CYS	CYS	CYS	CYS	CYS	THR
257	PRO	CYS	PRO	PRO	PRO	PRO	PRO	PRO	CYS
258		PRO	GLN	PRO	PRO	GLN	PRO	GLN	PRO
259		THR							LEU

Table 2.

Conservation of VG35 to the different serine proteases.

Residue	Conservation of VG35	Residue	Conservation of VG35	Residue	Conservation of VG35
15	0%	102	67%	189	0%
16	100%	103	44%	190	0%
17	100%	104	44%	191	67%
18	0%	105	33%	192	0%
19	100%	106	44%	193	0%
20	0%	107	11%	194	0%
21	100%	108	44%	195	0%
22	100%	109	0%	196	0%
23	0%	110	67%	197	67%
24	100%	111	0%	198	0%
25	0%	112	33%	199	56%
26	11%	113	0%	200	0%
27	100%	114	67%	201	0%
28	100%	115	0%	202	0%
29	67%	116	67%	203	0%
30	0%	117	22%	204	0%
31	100%	118	56%	205	78%
32	100%	119	67%	206	0%
33	78%	120	67%	207	0%
34	100%	121	0%	208	0%
35	100%	122	44%	209	0%
36	0%	123	0%	210	0%
37	100%	124	11%	211	0%
38	0%	125	67%	212	0%
39	100%	126	0%	213	11%

40	0%	127	67%	214	0%
41	0%	128	67%	215	0%
42	0%	129	56%	216	67%
43	33%	130	0%	217	0%
44	0%	131	0%	218	0%
45	33%	132	67%	219	0%
46	11%	133	89%	220	0%
47	44%	134	0%	221	0%
48	56%	135	100%	222	0%
49	0%	136	67%	223	0%
50	67%	137	0%	224	0%
51	67%	138	22%	225	0%
52	67%	139	67%	226	0%
53	67%	140	0%	227	0%
54	67%	141	0%	228	0%
55	67%	142	67%	229	0%
56	67%	143	33%	230	0%
57	44%	144	67%	231	0%
58	0%	145	0%	232	0%
59	0%	146	0%	233	0%
60	0%	147	33%	234	0%
61	0%	148	67%	235	0%
62	67%	149	11%	236	0%
63	100%	150	0%	237	0%
64	67%	151	0%	238	0%
65	0%	152	11%	239	0%
66	67%	153	0%	240	0%
67	0%	154	22%	241	0%
68	22%	155	0%	242	0%
69	56%	156	0%	243	0%
70	56%	157	0%	244	0%
71	33%	158	0%	245	0%
72	11%	159	0%	246	0%
73	56%	160	0%	247	0%
74	0%	161	67%	248	0%
75	33%	162	0%	249	0%
76	67%	163	0%	250	0%
77	0%	164	0%	251	0%
78	0%	165	0%	252	0%
79	0%	166	0%	253	0%
80	67%	167	0%	254	0%
81	0%	168	0%	255	0%

82	0%	169	0%	256	0%
83	0%	170	0%	257	0%
84	0%	171	0%	258	0%
85	33%	172	0%	259	0%
86	67%	173	0%		
87	0%	174	0%		
88	56%	175	0%		
89	67%	176	0%		
90	0%	177	11%		
91	0%	178	33%		
92	33%	179	0%		
93	67%	180	0%		
94	67%	181	0%		
95	0%	182	0%		
96	100%	183	0%		
97	33%	184	0%		
98	67%	185	0%		
99	33%	186	0%		
100	44%	187	22%		
101	33%	188	0%		

Table 3.

Epitopes of *Crotalus tigris* serine protease.

Residue#	Amino acid	Site
15	LEU	Epitope
81	LYS	Epitope
82	VAL	Epitope
83	ARG	Epitope
84	ASN	Epitope
85	GLU	Epitope
86	ASP	Epitope
124	ALA	Epitope

Table 4.

Epitopes of VG172 serine protease.

Residue#	Amino acid	Site
18	PRO	Epitope
19	GLN	Epitope

69	GLU	Epitope
98	CYS	Epitope
99	PRO	Epitope
100	ILE	Epitope
101	ARG	Epitope
102	LYS	Epitope
103	LYS	Epitope
104	ILE	Epitope
105	ASP	Epitope
108	ASP	Epitope
121	THR	Epitope
137	ARG	Epitope
138	VAL	Epitope
139	GLY	Epitope
151	PRO	Epitope
153	PRO	Epitope
154	LYS	Epitope
170	ASN	Epitope
171	TYR	Epitope
172	LYS	Epitope
173	GLY	Epitope
174	GLY	Epitope
175	GLN	Epitope
176	PRO	Epitope
177	ALA	Epitope
178	TYR	Epitope
179	ALA	Epitope
180	ARG	Epitope
181	LEU	Epitope
182	PRO	Epitope
183	GLN	Epitope
184	THR	Epitope
193	PRO	Epitope
194	LYS	Epitope
195	ARG	Epitope

Table 5.

Epitopes of VG35 serine protease.

Residue#	Amino acid	Site
85	GLU	Epitope
100	ASN	Epitope
104	ASN	Epitope
118	SER	Epitope
119	PRO	Epitope
121	THR	Epitope
122	ASN	Epitope
123	ARG	Epitope
124	GLU	Epitope
132	PRO	Epitope
137	ARG	Epitope
138	GLY	Epitope
139	GLY	Epitope
140	PRO	Epitope
153	GLN	Epitope
171	TRP	Epitope
172	GLY	Epitope
173	ARG	Epitope
177	PRO	Epitope
178	PRO	Epitope
180	SER	Epitope
181	GLN	Epitope
184	PRO	Epitope
185	ALA	Epitope
186	THR	Epitope
195	SER	Epitope
196	TRP	Epitope
197	LYS	Epitope