International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015
ISSN 2091-2730

# SEQUENCE ANALYSIS OF BASIC PHOSPHOLIPASE A2 (NEUROTOXIN) AS A POTENTIAL DRUG TARGET: AN IN SILICO APPROACH

Praveen Kumar K.S[1] and Dr. Mahesh pattabhiramiah[2]
Centre for Applied Genetics, Jnanabharathi, Bangalore University,
Bangalore-560056, India
Email:reply2mahesh@gmail.com
Phone:+91-9916130942

**Abstract**— Snake venom is merely modified saliva, or a combination of many different proteins and enzymes. Hemotoxins along with Basic PLA2 present in venom cause hemolysis, the destruction of red blood cells (erythrocytes), or induce blood coagulation (clotting) as well as damage vascular endothelium. There by causing dysfunctions of the normal body activity. Characterization of viper venom reveals hydropathic amino acid distribution. The aliphatic index computed by Expasy's Protparam infers most of these proteins are stable at wide range of temperatures, ranging from 60-65(Aliphatic index by CLC). Protein homology prediction using (Phyre 2) reveals probability and scoring for a particular template and confidence level and percentage of Identity. The secondary structure analysis showed more number of coils and percentile turns (range:50-60%).Multiple sequence alignment of viper PLA2 proteins revealed highly conserved regions. The positions of amino acids residues of 3G8G as the ideal in the core. The allowed and disallowed regions in Rampage are within the accepted limits for the model 3D structure and most favored regions in R-Plot suggests 3G8H and 1OQS has highly favored regions of about 96% and 98.3% respectively, indicating basic PLA2 of snake venom protein is of good quality. SOPMA and SMART reveals domains, motifs, and e-Values which is 1.23e-65 to 8.02e-66. SOSUI revealed percentile membrane proteins of 64.3%.3D structure analysis provide the importance of these proteins. SVM prot analysis of PLA2 proteins reveals that they are the potential neurotoxin targets and to the receptors of post synaptic membranes and drug targets for the venomous snake bites and its treatment as well.

**Keywords**— PLA2 protein, CLC work bench , Expasy ProtParam , MSA, Phyre2 , SOPMA, SOSUI

## INTRODUCTION

Phospholipases A2 (PLA2s) are enzymes that release fatty acids from the second carbon group of glycerol. This particular phospholipase specifically recognizes the sn-2 acyl bond of phospholipids and catalytically hydrolyzes the bond releasing arachidonic acid and lysophospholipids. Upon downstream modification by cyclooxygenases, arachidonic acid is modified into active compounds called eicosanoids. Eicosanoids include prostaglandins and leukotrienes, which are categorized as anti-inflammatory and inflammatory mediators [6].

PLA2 are commonly found in mammalian tissues as well as insect and snake venom [18].Venom from both snakes and insects is largely composed of melittin, which is a stimulant of PLA2. Due to the increased presence and activity of PLA2 resulting from a snake or insect bite, arachidonic acid is released from the phospholipid membrane disproportionately. Snake venom phospholipase A2 (PLA2) acts as a presynaptic neurotoxin and has been found to bind with high affinity to intracellular proteins. As a result, inflammation and pain occur at the site[1].The structure includes four residues which occur less frequently in PLA2's. His1, Arg6 and Trp70 located at the interfacial recognition site may play an important role in the interaction with aggregated substrates, such that it can be considered as neurotoxic drug targets. While Trp77 contributes to the hydrophobic interactions between the beta-wing and the main body of the molecule [12].

The physicochemical and the structural properties of the proteins are well understood with the use of computational tools by through in-silico analysis. The statistics about a protein sequence such as number of amino acid, frequency is predicted by CLC work bench(http://w.w.w..clcbio.com/index.php?id=28). Sequence length, and the physico-chemical properties of a proteins such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, etc. can be computed by ProtParam, the protein 3D model and its characteristics were predicted by Swiss model server [22].Protein homology modeling [10][11][25] and analogy recognition is made through Phyre2 online server. Ramachandran plot [20] (RAMPAGE) is the way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure. Further Computer-aided techniques for the

efficient identification and optimization of novel molecules with a desired biological activity have become a part of the drug discovery process.

Bioinformatics has revolutionized the field of molecular biology. The raw sequence information of proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools [3]. Prediction of protein function is important application of bioinformatics [19].The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties. Sequence analysis and physicochemical characterization of proteins using biocomputation tools [4][15] [16] have been done by many researchers and reported.

However, physicochemical characterization of Basic phospholipase A2 proteins has not been done so far. The purpose of this study was to perform in-silico analysis to determine the molecular characterization, identity, physicochemical characteristics of phospholipase family so as to pave the way to find out better therapeutic method and to say these are potential targets to treat and control the overwhelming of Snake bites. The Importance of aberrant basic phospholipase A2 protein function in Snake bite and for the Drug discovery, makes it an object of study.

## Materials and methods

**Protein sequence retrieval :**The Protein Sequences of PLA2(14 sequences)were retrieved in FASTA format from Uniprot (Swiss-prot) database(Table1) used for further analysis.

**Amino acid Composition:** The amino acid composition of selected proteins were computed using the tool CLC free workbench ( www.clc.bio.com/..../clc-main-workbench), tabulated in (Table-2).

**Primary structure analysis** - Percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis by SOPMA (Table-3).

Simple Modular Architecture Research Tool (SMART) is a biological database that is used in the identification and analysis of protein domains within protein sequences [21] [14] (Table-6).

**Physio-chemical parameters:** ProtParam (http://www.expasy.org/tools/protparam.html)[24] computes various physicochemical properties that can be deduced from a protein sequence. No additional information is required about the protein under consideration [23].

The physicochemical parameters, theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient ,instability index [8], aliphatic index [7] and grand average hydropathy (GRAVY)[13] were computed using the Expasy's ProtParam server [17],and tabulated in (Table-4).

SVM prot analysis (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) which is a protein function prediction tool, and classification of distantly related proteins can be Analyzed.(Table-7).

**Secondary structure prediction**: The secondary structure was predicted by self-optimized prediction method with alignment by SOPMA server [4] (Table-5).

The system SOSUI for the discrimination of membrane proteins and soluble ones together with the prediction of trans membrane helices. The system SOSUI is available through internet access: http://www.tuat.ac.jp/mitaku/sosui/ (Fig:5).

**Sequence Homology Analysis:** Method employed in alignment of divergent protein sequences, it is used to align divergent sequences in order, locally reduced gap penalties to encourage the opening up of new gaps at these positions. MULTILIN online tool used to do pair wise and multiple sequence alignment (Fig-1).

**Tertiary structure Prediction:** Tertiary structure prediction [8][9].(Fig-2) of PLA2 proteins was performed using bioinformatics tool Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/index.cgi). RASMOL visualization provided 3D structure of selected PDB ids. The modeled 3D structure were evaluated and validated with RAMPAGE (mordred.bioc.cam.ac.uk/…/rampage.php) (Fig-3).

**(Table-1):The selected neurotoxin protein basic PLA2 retrieved from Uniprot .**

| Sl. no | Species | ID | Length | Protein sequence |
|---|---|---|---|---|
| 1 | Vipera ammodytes ammodytes | P11407 | 138 | MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCCFVHDCCYGNLPDCSPKTDRY KYHRENGAIVCGKGTSCENRICECDRAAAICFRKNLKTYNYIYRNYPDILCKEESEKC |
| 2 | Vipera ammodytes ammodytes | P17935 | 138 | MRILWIVAVCLIGVEGSVIEFGKMIQEETDKNPLTSYSFYGCHCGLGNKGKPKDATDRCCFVHSCCYAKLPDCSPKTNRY EYHRENGAIVCGSSTPCKKQICECDRAAAICFRENLKTYNKKYKVYLRFKCKGVSEKC |
| 3 | Daboia siamensis | Q02471 | 138 | MRTLWIVAVCLIGVEGNLFQFARMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCCFVHDCCYGGVKGCNPKLAIY SYSFQRGNIVCGRNNGCLRTICECDRVAANCFHQNKNTYNKEYKFLSSSKCRQRSEQC |
| 4 | Vipera ammodytes ammodytes | P00626 | 138 | MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCCFVHDCCYGNLPDCSPKTDRY KYHRENGAIVCGKGTSCENRICECDRAAAICFRKNLKTYNYIYRNYPDFLCKKESEKC |
| 5 | Vipera aspis aspis | Q8JFGZ | 138 | MRILWIVAVCLIGVEGNLFQFAKMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCCFVHDCCYGRVRGCNPKLAIY SYSFKKGNIVCGKNNGCLRDICECDRVAANCFHQNKNTYNKNYRFLSSSRCRQTSEQC |
| 6 | Vipera ammodytes ammodytes | P14424 | 138 | MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCCFVHDCCYGNLPDCSPKTDRY KYHRENGAIVCGKGTSCENRICECDRAAAICFRKNLKTYNHIYMYYPDFLCKKESEKC |
| 7 | Daboia siamensis | A8CG84 | 137 | MRTLWIVAVCLIGVEGSLLEFGKMILEETGKLAIPSYSSYGCYCGWGGKGTPKDATDRCCFVHDCCYGNLPDCNPKSDRY KYKRVNGAIVCEKGTSCENRICECDKAAAICFRQNLNTYSKKYMLYPDFLCKGELRC |
| 8 | Vipera nikolskii | Q1RP79 | 138 | MRILWIVAVCLIGVEGNLFQFAKMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCCFVHDCCYGRVRGCNPKLAIY AYSFKKGNIVCGKNNGCLRDICECDRVAANCFHQNQNTYNKNYKFLSSSRCRQTSEQC |
| 9 | Sistrurus catenatus tergeminus | Q6EER3 | 138 | MRTFWIVAVLLVGVEGNLLQFNKMIKIMTKKNAIPSYSSYGCYCGWGGRGRPKDATDRCCFVHDCCYEKLTDCSPKTDTY SYSLKSGVIICGGNDPCKKQICECDKAAAVCFGENLSTYKKRYMFYPDFLCTDPSETC |
| 10 | Sistrurus catenatus tergeminus | Q6EER2 | 138 | MRALWIVAVLLVGVEGNLLQFNKMIKFETNKNAIPFYAFYGCYCGWGGRGRPKDATDRCCFVHDCCYGKLPNCDTKWDIY SYSLKSGFITCGKGTWCEEQICECDRVAAECLRRSLSTYKYGYMFYLDSRCKGPSEQC |
| 11 | Protobothrops flavoviridis | Q805A2 | 138 | MRTLWIMAVLLVGVEGNLLQFNKMIKIMTKKNGFPFYTSYGCYCGWGGRGKPKDATDRCCFVHDCCYEKLTDCSPKSDIY SYSWKTGVIICGEGTECEKQICECDRAAAVCFGQNLRTYKKKYMFYPDFLCTDPTEKC |
| 12 | Protobothrops mucrosquamatus | Q90W39 | 138 | MRTLWIVAVLLLGVEGNLLQFNKMIKIMTKKNAIPFYSSYGCYCGWGGQGKPKDATDRCCFVHDCCYGKLTDCSPKSDIY SYSWKTGIIICGEGTECEKKICECDRAAAVCLGHNLRTYKKRYMFYPDFLCTDPSEKC |
| 13 | Deinagkistrodon acutus | Q1ZY03 | 138 | MRTLWIVAVLLVSVEGHLLQFNKMIKIMTRKNAFPFYTSYGCYCGWGGRGWPKDATDSCCFVHDCCYQKLTGCSPKWDIY PYSWKTGVIICGEGTPCEKEICECDRAAAVCLGENLRTYKTKYMFYPDFLCKKPSKQC |
| 14 | Vipera nikolskii | Q1RP78 | 138 | MRILWIVAVCLIGVEGNLFQFAKMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCCFVHDCCYGRVRGCNPKLAIY AYSFKKGNIVCGKNNGCLRDICECDRVAANCFHQNKNTYNKNYRFLSSSRCRQTSEQC |

**(Table-2): Representation of frequency of amino acids.**

| | Amino Acid | A8CG84 | P00626 | P11407 | P14424 | P17935 | Q1RP78 | Q1RP79 | Q1ZY03 | Q6EER2 | Q6EER3 | Q8JFG0 | Q90W39 | Q805A2 | Q02471 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | |
| 1 | Alanine | 0.051 | 0.043 | 0.043 | 0.043 | 0.051 | 0.058 | 0.058 | 0.043 | 0.051 | 0.043 | 0.051 | 0.043 | 0.036 | 0.051 |
| 2 | cysteine | 0.109 | 0.109 | 0.109 | 0.109 | 0.109 | 0.109 | 0.109 | 0.101 | 0.101 | 0.101 | 0.109 | 0.101 | 0.101 | 0.109 |
| 3 | Aspartic Acid | 0.051 | 0.001 | 0.051 | 0.051 | 0.036 | 0.036 | 0.036 | 0.043 | 0.051 | 0.065 | 0.036 | 0.058 | 0.058 | 0.029 |
| 4 | Glutamic Acid | 0.058 | 0.058 | 0.065 | 0.058 | 0.065 | 0.022 | 0.022 | 0.043 | 0.051 | 0.036 | 0.022 | 0.043 | 0.051 | 0.029 |
| 5 | phenylalanine | 0.029 | 0.036 | 0.029 | 0.036 | 0.036 | 0.051 | 0.051 | 0.043 | 0.051 | 0.043 | 0.051 | 0.036 | 0.051 | 0.051 |
| 6 | Glycine | 0.095 | 0.101 | 0.101 | 0.101 | 0.072 | 0.101 | 0.101 | 0.08 | 0.094 | 0.08 | 0.101 | 0.087 | 0.087 | 0.109 |
| 7 | Histidine | 0.007 | 0.014 | 0.014 | 0.022 | 0.022 | 0.014 | 0.014 | 0.014 | 0.007 | 0.007 | 0.014 | 0.014 | 0.007 | 0.014 |
| 8 | Isoleucine | 0.051 | 0.051 | 0.058 | 0.051 | 0.058 | 0.058 | 0.058 | 0.051 | 0.043 | 0.051 | 0.058 | 0.065 | 0.051 | 0.051 |
| 9 | Lysine | 0.088 | 0.08 | 0.072 | 0.08 | 0.116 | 0.065 | 0.065 | 0.094 | 0.072 | 0.094 | 0.065 | 0.101 | 0.101 | 0.058 |
| 10 | leucine | 0.08 | 0.065 | 0.065 | 0.065 | 0.051 | 0.051 | 0.051 | 0.065 | 0.072 | 0.058 | 0.051 | 0.072 | 0.058 | 0.051 |
| 11 | Methionine | 0.022 | 0.022 | 0.022 | 0.029 | 0.014 | 0.014 | 0.014 | 0.029 | 0.022 | 0.029 | 0.014 | 0.029 | 0.036 | 0.014 |
| 12 | Asparagine | 0.044 | 0.051 | 0.051 | 0.043 | 0.043 | 0.087 | 0.087 | 0.022 | 0.036 | 0.036 | 0.087 | 0.029 | 0.029 | 0.08 |
| 13 | Proline | 0.036 | 0.036 | 0.036 | 0.036 | 0.036 | 0.014 | 0.014 | 0.051 | 0.029 | 0.043 | 0.014 | 0.036 | 0.036 | 0.014 |
| 14 | Glutamine | 0.007 | 0 | 0 | 0 | 0.014 | 0.036 | 0.043 | 0.022 | 0.022 | 0.014 | 0.036 | 0.014 | 0.022 | 0.043 |
| 15 | Arginine | 0.051 | 0.058 | 0.058 | 0.051 | 0.051 | 0.065 | 0.058 | 0.036 | 0.058 | 0.036 | 0.065 | 0.036 | 0.036 | 0.065 |
| 16 | Serine | 0.051 | 0.043 | 0.043 | 0.043 | 0.058 | 0.051 | 0.051 | 0.043 | 0.051 | 0.065 | 0.058 | 0.051 | 0.036 | 0.058 |
| 17 | Threonine | 0.044 | 0.058 | 0.058 | 0.058 | 0.043 | 0.029 | 0.029 | 0.065 | 0.043 | 0.065 | 0.029 | 0.058 | 0.072 | 0.036 |
| 18 | valine | 0.044 | 0.043 | 0.043 | 0.043 | 0.058 | 0.058 | 0.058 | 0.051 | 0.043 | 0.051 | 0.058 | 0.036 | 0.043 | 0.058 |
| 19 | Tryptophan | 0.015 | 0.007 | 0.007 | 0.007 | 0.007 | 0.022 | 0.022 | 0.036 | 0.029 | 0.014 | 0.022 | 0.022 | 0.022 | 0.022 |
| 20 | Tyrosine | 0.066 | 0.072 | 0.072 | 0.072 | 0.058 | 0.058 | 0.058 | 0.065 | 0.072 | 0.065 | 0.058 | 0.065 | 0.065 | 0.058 |
| | | | | | | | | | | | | | | | |

**(Table3):Hydrophilic and hydrophobic residues content computed by SOPMA.**

| Accession no. | Percentage of Hydrophilic residues | Percentage of hydrophobic residues |
|---|---|---|
| P11407 | 3.8 | 5.6 |
| P17935 | 1.3 | 2.8 |
| Q02471 | 1.8 | 6.4 |
| P00626 | 1.3 | 9.3 |
| Q8JFGZ | 2.7 | 6.8 |
| P14424 | 5.6 | 7.5 |
| A8CG84 | 4.9 | 5.3 |
| Q1RP79 | 3.1 | 2.5 |
| Q6EER3 | 3.5 | 1.1 |
| Q6EER2 | 3.2 | 15 |
| Q805A2 | 8 | 6.5 |
| Q90W39 | 2.5 | 8.1 |
| Q1ZY03 | 1.1 | 7.7 |
| Q1RP78 | 3.8 | 5.6 |

**(Table-4):Parameters computed by Expasy ProtParam**

| Accession no. | PI | Mol.wt | -R | +R | EC | II | AI | GRAVY |
|---|---|---|---|---|---|---|---|---|
| A8CG84 | 8.32 | 15380.8 | 15 | 19 | 25285 | 23.71 | 69.05 | -0.215 |
| P00626 | 8.33 | 15530.9 | 15 | 19 | 21275 | 17.19 | 62.17 | -0.299 |
| P11407 | 7.86 | 15497.8 | 16 | 18 | 21275 | 21.5 | 65 | -0.284 |
| P14424 | 8.14 | 15528.9 | 15 | 18 | 21420 | 26.5 | 62.17 | -0.251 |
| P17935 | 8.87 | 15636.2 | 14 | 23 | 18295 | 35.54 | 64.28 | -0.364 |
| Q1RP78 | 9 | 15594 | 8 | 18 | 29295 | 25.93 | 65 | -0.205 |
| Q1RP79 | 8.91 | 15565.9 | 8 | 17 | 28420 | 24.9 | 65 | -0.198 |
| Q1ZY03 | 8.62 | 15853.6 | 12 | 18 | 37775 | 39.5 | 64.28 | -0.146 |
| Q6EER2 | 8.36 | 15844.3 | 14 | 18 | 25285 | 50.77 | 62.9 | -0.218 |
| Q6EER3 | 8.35 | 15525 | 14 | 18 | 28420 | 33.63 | 61.45 | -0.213 |
| Q8JFG0 | 9 | 15610 | 8 | 18 | 29295 | 24.22 | 64.28 | -0.224 |
| Q90W39 | 8.5 | 15668.3 | 14 | 19 | 29910 | 41.75 | 68.55 | -0.181 |
| Q805A2 | 8.35 | 15817.4 | 15 | 19 | 30785 | 37.84 | 58.62 | -0.256 |
| Q02471 | 8.92 | 15554.8 | 8 | 17 | 28420 | 26.76 | 61.45 | -0.236 |

**(Table-5):**Representation of helix ,sheet, turns, coils by Garnier peptide Analysis by through online tool SOPM, SOPMA and SSCP (Secondary Structural Content Prediction)

| | | P11407 | P17935 | Q02471 | P00626 | Q8JFG0 | P14424 | A8CG84 | Q1RP79 | Q6EER3 | Q6EER2 | Q805A2 | Q90W39 | Q1ZY03 | Q1RP78 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Residue totals | 37 | 35 | 22 | 39 | 23 | 39 | 38 | 22 | 38 | 38 | 42 | 45 | 47 | 24 |
| helix[H] | | | | | | | | | | | | | | | |
| | Percent% | 30.3 | 28.7 | 18 | 32 | 18.9 | 32 | 31.4 | 18 | 31.1 | 31.1 | 34.4 | 36.9 | 38.5 | 19.7 |
| | Residue totals | 25 | 24 | 30 | 23 | 30 | 25 | 31 | 29 | 26 | 23 | 22 | 23 | 21 | 30 |
| sheet[E] | | | | | | | | | | | | | | | |
| | Percent% | 20.5 | 19.7 | 24.6 | 18.9 | 24.6 | 20.5 | 25.6 | 23.8 | 21.3 | 18.9 | 18 | 18.9 | 17.2 | 24.6 |
| | Residue totals | 62 | 65 | 73 | 62 | 74 | 61 | 61 | 76 | 66 | 69 | 66 | 63 | 62 | 73 |
| turns[T] | | | | | | | | | | | | | | | |
| | Percent% | 50.8 | 53.3 | 59.8 | 50.8 | 60.7 | 50 | 50.4 | 62.3 | 54.1 | 56.6 | 54.1 | 51.6 | 50.8 | 59.8 |
| | Residue totals | 14 | 14 | 13 | 14 | 11 | 13 | 7 | 11 | 8 | 8 | 8 | 7 | 8 | 11 |
| coils[c] | | | | | | | | | | | | | | | |
| | Percent% | 11.5 | 11.5 | 10.7 | 11.5 | 9 | 10.7 | 5.8 | 9 | 6.6 | 6.6 | 6.6 | 5.7 | 6.6 | 9 |

**(Fig:1): Multiple sequence alignment [MSA] by MULTILIN.**

Conserved sequences for hierarchical clustering, primary constructions ,identity percentage strong and weakly similar sequences is predicted.

```
                          10        20        30        40        50        60
                           |         |         |         |         |         |
PA2BC_VIPAA    MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCC
PA2BA_VIPAA    MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCC
PA2BB_VIPAA    MRTLWIVAVCLIGVEGSLLEFGMMILGETGKNPLTSYSFYGCYCGVGGKGTPKDATDRCC
PA2BS_DABSI    MRTLWIVAVCLIGVEGSLLEFGKMILEETGKLAIPSYSSYGCYCGWGGKGTPKDATDRCC
PA2HL_VIPAA    MRILWIVAVCLIGVEGSVIEFGKMIQEETDKNPLTSYSFYGCHCGLGNKGKPKDATDRCC
PA2B1_VIPNI    MRILWIVAVCLIGVEGNLFQFAKMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCC
PA2B4_DABSI    MRTLWIVAVCLIGVEGNLFQFARMINGKLGAFSVWNYISYGCYCGWGGQGTPKDATDRCC
               ** :**:** *:.*** :::*  **         .  *  ***:** *.:* ****** **
Prim.cons.     MRTLWIVAVCLIGVEGNLLQF2KMIKGETGKNA3PSYSSYGCYCGWGG2GTPKDATDRCC
                         70        80        90       100       110       120
|        |        |         |         |         |
PA2BC_VIPAA    FVHDCCYGNLPDCSPKTDRYKYHRENGAIVCGKGTSCENRICECDRAAAICFRKNLKTYN
PA2BA_VIPAA    FVHDCCYGNLPDCSPKTDRYKYHRENGAIVCGKGTSCENRICECDRAAAICFRKNLKTYN
PA2B4_DABSI    FVHDCCYGGVKGCNPKLAIYSYSFQRGNIVCGRNNGCLRTICECDRVAANCFHQNKNTYN
               ***.***  : .*..*   * *    * * *  .  * . *****:.** *: ..  **.
Prim.cons.     FVHDCCYGKLPDCSPKTDIYSYSRKNGAIVCGKGT2CEK2ICECDRAAAICFRQNLNTYN
```

**Alignment data :**

Primary construction. KKYMFYPDFLCKQPSE2C,Alignment length : 138
Identity (*) : Strongly similar (:) : Weakly similar (.) : Different : 57 is 41.30 %

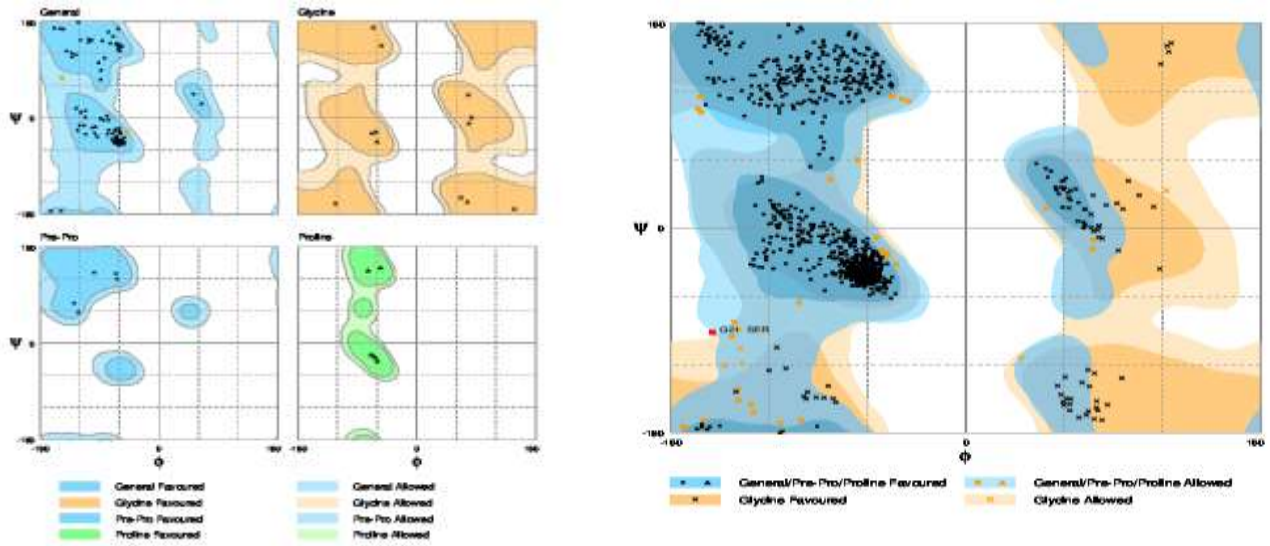
**(Fig-2): Protein homology/analogy recognition by Phyre2.**

| # | Template | Alignment Coverage | 3D Model | Confidence | % i.d. | Template Information |
|---|---|---|---|---|---|---|
| 1 | c3dihA_ | Alignment |  | 100.0 | 73 | **PDB header:**toxin<br>**Chain:** A: **PDB Molecule:**phospholipase a2 homolog, ammodytin l;<br>**PDBTitle:** crystal structure of ammodytin l |
| 2 | d1jlaa_ | Alignment |  | 100.0 | 63 | **Fold:**Phospholipase A2, PLA2<br>**Superfamily:**Phospholipase A2, PLA2<br>**Family:**Vertebrate phospholipase A2 |
| 3 | d1mc2a_ | Alignment |  | 100.0 | 59 | **Fold:**Phospholipase A2, PLA2<br>**Superfamily:**Phospholipase A2, PLA2<br>**Family:**Vertebrate phospholipase A2 |
| 4 | d1oqsb_ | Alignment |  | 100.0 | 52 | **Fold:**Phospholipase A2, PLA2<br>**Superfamily:**Phospholipase A2, PLA2<br>**Family:**Vertebrate phospholipase A2 |
| 5 | c3bjwE_ | Alignment |  | 100.0 | 63 | **PDB header:**hydrolase<br>**Chain:** E: **PDB Molecule:**phospholipase a2;<br>**PDBTitle:** crystal structure of ecarpholin s complexed with suramin |

**(Fig-3):Crystallographic Structures of Basic PLA2 proteins developed.**

P00626, 3G8G.PDB P11407, 3G8H.PDB P17935, 3DIH.PDB

**(Fig-4): Ramchandran plot analysis for tertiary structure:**




### 3G8H.pdb Results

Residue [A 23: PHE] (-147.08, 75.26) in Allowed region
Residue [A 105: TYR] (-48.46, -25.47) in Allowed region
Number of residues in favored region (~98.0% expected): 118 (98.3%)
Number of residues in allowed region (~2.0% expected): 2 (1.7%)
Number of residues in outlier region : 0 (0.0%)

### 1OQS.pdb result

Residue [A 21: SER] (-138.79, -89.44) in Allowed region
Residue [A 32: GLY] (-136.39,-118.72) in Allowed region
Residue [A 39: ASP] (-171.43,-175.23) in Allowed region
Residue [A 73: TYR] (-100.41,-168.66) in Allowed region

Number of residues in favored region (~98.0% expected) : 923 (96.1%)
Number of residues in allowed region (~2.0% expected) : 36 (3.8%)
Number of residues in outlier region : 1 (0.1%)

**(Fig-5): SOSUI Predictions for Trans membrane protein.**

>sp|P11407|PA2BC_VIPAA Basic phospholipase A2 ammodytoxin C
OS=Vipera ammodytes ammodytes PE=1 SV=1

**At the result part**

| | |
|---|---|
| TOTAL PROTEINS TAKEN IS | **14** |
| TOTAL MEMBRANE PROTEINS ARE | **9** |
| PERCENTAGE OF ARE MEMBRANE PROTEINS | **64.3 %** |

**(Table-6): SMART Analysis.**

**Confidently predicted domains, repeats, motifs and features:**

| Name | Start ▲ | End | E-value |
|---|---|---|---|
| PA2c | 17 | 132 | 7.6e-66 |
| PA2c | 167 | 282 | 4.01e-65 |
| PA2c | 316 | 431 | 1.23e-65 |
| PA2c | 1677 | 1792 | 8.02e-65 |
| PA2c | 1828 | 1943 | 2.09e-63 |
| PA2c | 1979 | 2094 | 1.65e-66 |

**(Table-7): Support Vector Machine (SVM) Prot Analysis.**

Predicted results are given in TP (true positive), FN (false negative), TN (true negative), FP (false positive), and Q (overall accuracy). Number of positive or negative samples in testing and independent evaluation sets is TP + FN or TN + FP, respectively.

| Protein family | Training set | | Testing set | | | | Independent evaluation set | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | positive | negative | positive | | negative | | positive | | | negative | | |
| | | | TP | FN | TN | FP | TP | FN | Sensitivity | TN | FP | Specificity |
| EC1.1 Oxidoreductases acting on the CH-OH group of donors | 1164 | 2324 | 1795 | 10 | 7594 | 14 | 494 | 105 | 82.5% | 4760 | 192 | 96.1% |
| EC1.2 Oxidoreductases acting on the aldehyde or Oxo group of donors | 665 | 1920 | 705 | 14 | 8051 | 25 | 259 | 69 | 79.0% | 4908 | 77 | 98.5% |
| EC1.3 Oxidoreductases acting on the CH-CH group of donors | 491 | 1917 | 131 | 3 | 8090 | 17 | 73 | 37 | 66.4% | 4941 | 57 | 98.9% |
| EC2.8 Transferases transferring sulfur-containing groups | 203 | 1549 | 43 | 0 | 8531 | 7 | 20 | 10 | 66.7% | 5021 | 11 | 99.8% |
| EC6.1 Ligases forming carbon-oxygen bonds | 281 | 1115 | 381 | 1 | 1185 | 13 | 286 | 29 | 90.8% | 980 | 27 | 97.3% |
| EC6.2 Ligases forming carbon-sulfur bonds | 149 | 1233 | 154 | 4 | 8858 | 4 | 51 | 13 | 79.7% | 5203 | 13 | 99.8% |
| EC6.3 Ligases forming carbon-nitrogen bonds | 381 | 1133 | 358 | 2 | 1148 | 3 | 294 | 57 | 83.8% | 946 | 45 | 95.5% |
| EC6.4 Ligases forming carbon-carbon bonds | 99 | 1543 | 45 | 0 | 8548 | 8 | 28 | 16 | 63.6% | 5033 | 4 | 99.9% |
| EC6.5 Ligases forming phosphoric ester bonds | 94 | 1679 | 36 | 2 | 8408 | 3 | 22 | 9 | 71.0% | 5027 | 6 | 99.9% |
| Zinc-binding | 2731 | 6416 | 6610 | 569 | 5931 | 360 | 4616 | 1546 | 74.9% | 6289 | 127 | 98.0% |
| Metal-binding | 5013 | 3101 | 11806 | 1015 | 4217 | 522 | 12070 | 3391 | 78.1% | 4529 | 617 | 88.0% |
| Antioxidant | 145 | 3450 | 81 | 4 | 12429 | 5 | 83 | 10 | 89.2% | 7937 | 23 | 99.7% |
| Hormone | 584 | 3371 | 309 | 2 | 12389 | 1 | 285 | 25 | 91.9% | 7840 | 64 | 99.2% |
| Immune response | 447 | 2998 | 268 | 1 | 12561 | 2 | 132 | 20 | 86.8% | 7815 | 4 | 99.9% |
| Inflammatory response | 134 | 3320 | 83 | 15 | 12425 | 7 | 56 | 8 | 87.5% | 7882 | 20 | 99.7% |
| Innate immunity | 193 | 3382 | 69 | 16 | 12276 | 9 | 52 | 6 | 89.7% | 7860 | 8 | 99.9% |
| Motor protein | 212 | 1656 | 198 | 8 | 13630 | 425 | 85 | 8 | 91.4% | 7666 | 217 | 97.2% |
| Virulence | 544 | 2983 | 268 | 2 | 12278 | 17 | 165 | 8 | 95.4% | 7665 | 6 | 99.9% |

## Results and Discussion:-

### Amino acid composition:

The results of Primary sequence analysis of 14 PLA2 proteins analyzed by CLC work bench revealed the sequence length of all amino acids is found to be138 in number, except A8CG84(137), tabulated in (Table 4). Most abundant amino acids were serine, leucine, valine, arginine, Proline and threonine which are tabulated in CLC work bench. Abundant amino acid was found to be Arginine and Lysine in a protein, which promotes the phenomenon of neurotoxicity by hydrolyzing and inactivating components of the Nervous system, one of the major causes for promoting inflammation and neurotoxic activity that can be seen in P00626, P11407,and P17935 (Table 2).

We conclude that Lysine has highest frequency in P17935 and Cysteine present in large frequency in Q1RP79. The aliphatic index computed for particular PLA2 protein and A8CG84 shows highest aliphatic index.

### Primary sequence analysis:

The result of primary structure analysis suggests that most of the PLA2 are hydropathic in nature due to presence of high non-polar residues content (Table-3). Presence of high content of cysteine residues in all indicates the more no. of disulphide bridges except in P17935 and Q6EER3 found to be 2.1 which indicates the absence of sulphide bonds (Table 3).

The start and end point with their E-values is predicted for Confidently predicted domains, repeats, motifs by through SMART. **Expect value** (**E**) a parameter that counts the number of hits one can "expect" to see by chance for a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Here it is in the expected range.

### Physico-chemical parameters:

The average molecular weight of basic PLA2 was found in between 15380.8-15853.61Da,

ProtParam of Expasy computes atomic composition of carbon ,hydrogen ,Nitrogen ,Oxygen ,Sulphur is for a range of (671,1044,184,202 and 18) wavelength, 280nm is favored because proteins absorb strongly there while other substances commonly in proteins solution, do not.

Extinction co-efficient of PLA2 at 280nm is ranging from 20400-52784 $M^{-1}$ $Cm^{-1}$.

ProtParam server predicted that P11407 , P17935, Q02471, P00626 Q8JFG0, P14424 , A8CG84 , Q1RP79 ,Q6EER3 , Q6EER2 ,Q805A2 , Q90W39 , Q1ZY03 , Q1RP78 are having Asp+Glu no. is 8-15 infers ATP-dependent RNA activity part of neurotoxic activity.

Isoelectric point is the pH at which the surface of protein is covered with charge but net charge of protein is zero.pI of PLA2 found to be Basic in nature. This important property, because it is at point that the protein is least soluble. Computed isoelectric point of proteins > 7 soluble in basic buffers. Isoelectric point is predicted ranges from 8.32 – 9.23 (Table 4).Useful for developing buffer system for purification of proteins.

The Aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains:alanine, valine, isoleucine, and Leucine of P11407, Q8JFG0 , and Q90W39 having 69.05, 64.28, and 68.55respectively. Which infers positive factor for thermo-stability[26].

The Grand Average hydropathy (GRAVY values) showed that all proteins are hydrophilic ranging from -0.3 to -0.1, supports the soluble nature of PLA2 proteins. Though it can play a role in substrate recognition. Here the protein sequences showing negative that indicates stability of the protein. In particular, hydrophobic amino acids can be involved in binding/recognition of ligands.

A protein whose instability index is smaller than 40 are predicted as stable, and a value above 40 predicts that the protein may be unstable, here the instability index of all proteins found to be less than 50 [9] (Table 4).

Support vector machines (SVM) method for the classification of proteins with diverse sequence distribution. SVMProt shows a certain degree of capability for the classification of distantly related proteins and homologous proteins of different function and thus may be used as a protein function prediction tool that complements sequence alignment methods. It has been employed in protein studies

including protein–protein interaction prediction , fold recognition , solvent accessibility and structure prediction . The prediction accuracy ranges from 65 to 91.4% in this study. Thus SVM classification of protein functional family ,a potentially developed into a protein function prediction tool to complement methods based on sequence similarity and clustering.

Based on the Classification of proteins of our interest and its values, we predict that , these proteins may be act as drug targets for inflammatory response(Fig-6), Antioxidant property, Metal binding sites, bonding involved ligation and Hormonal action for the Venomous snake bites is concerned.

**Secondary structure prediction:**

SOPMA was employed for calculating the secondary structural features of the selected protein sequences considered in this study. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method [17].

The secondary structure of alpha helix, beta turn, extended strand, random coil ranging from 49-60% predicted. The secondary analysis showed that PLA2 contain more random coils and alpha helices (range: 20-40%) than Beta sheets.

Being hydrophobic, Leucine prefers to be buried in protein hydrophobic cores. It also shows a preference for being within alpha helices more so than in beta strands. The very high coil structural content of PLA2is due to the rich content of more flexible glycine and hydrophobic Proline amino acids. Proline has a special property of creating links in polypeptide chains and disrupting ordered secondary structure.

The consequence in which most of the amino acid side chains of *trans membrane segments* is non-polar (*e.g.* Ala, Val, Leu, Ile, Phe). and the very polar CO-NH groups (peptide bonds) of the polypeptide backbone of trans membrane segments which participates in hydrogen bonding (H-bonds) in order to lower the cost of transferring them into the hydrocarbon interior. This H-bonding is most easily accomplished with alpha-helices for which all peptide bonds are H-bonded internally. On this basis we can say this may act as a neurotoxic drug target.(Table-5)

**Secondary structure analysis**

SOSUI that predicts a part of the secondary structure of proteins from a given amino acid sequence (AAS). The main objective is to determine whether the protein in question is a soluble or a trans membrane protein. Here it is at most are soluble proteins and predicted 64.3% are membrane proteins.

**Sequence homology Analysis:**

Multiple Sequence alignment by MULTILIN online tool. Homology sequences revealed significant conserved (Leucine) and semi conserved regions (Proline, Alanine). Residues conserved for 90 % and above is 59 which is 42.75 % Residues conserved for 50 % and less than 90 %: 48 is 34.78 %.Residues conserved less than 50 %: 29 is 21.01 %, Alignment length : 138 ,Identity (*) : 56 is 40.58 %, Strongly similar (:) : 11 is 7.97 % ,Weakly similar (.) : 14 is 10.14 % ,Different : 57 is 41.30 %.(Fig-1).

**Tertiary structure Prediction:** The Tertiary structure Analysis of 14 PLA2 proteins reveals the ideal structures with PDB ID : 3G8G ,3G8H ,3DIH. 3D structures of PDB IDs were generated through Phyre2 (Fig-2). Which predicts PLA2 ammoditin I, which shows highest % i.d. of 73%, with the use of psi-BLAST found 88% coverage which shows single highest scoring template [2] .Validation of results determined that the distribution of amino acid residues were at the most favorable region in the Ramchandran plot (more than 95%).The Crystallographic structures was developed for the sequence accession number P00626, P11407, P17935. shown in (Fig-3)

Ramchandran plot [20] (Fig-4) is an indication of the stereo chemical quality of the model taken for the structural analysis. Ramchandran plot displays the main chain torsion angles phi, psi (, ψ); (Ramchandran angles) in a protein of known structure. Dihedral angle checks Ramchandran plot shows phi-psi distribution. Each residue is classified according to its region ‘core’, ‘allowed’, ‘generous’, or ‘disallowed’. Residues in the generous and disallowed regions are high-lighted on the plot. A log-odds score shows how normal or unusual the residue's location is on the Ramchandran plot (fig-2) for the given residue type. Results gave us the value of 98.3% residues for 3G8H,96.1% residues for10QS in most favored regions in R-Plot suggests ,the predicted PLA2 proteins is of good Quality. Very useful in molecular medicine for designing a drug or biomedicine.[12].

## Conclusion:

The present analysis entitles members of Basic Phospholipase A2 selected P11407 , P17935, Q02471, P00626 Q8JFG0, P14424 , A8CG84 , Q1RP79 ,Q6EER3 , Q6EER2 ,Q805A2 , Q90W39 , Q1ZY03 , Q1RP78 from Uniprot database showing high conservation, suggests their functional similarity. In our studies we depicted that P00626, P11407 some of the protein families which has metal binding, immune response, inflammatory response, immunity and motar protein by through training, testing, independent evaluation by SVM Analysis prediction can say high sensitivity of PLA2 family proteins of venomous vipers, which induces cell invasion and might cause failure of nerve transmission and blockage may be the reason for the death of an individual due to its neurotoxicity. From the present analysis it can be concluded that selected PLA2 proteins have high degree of homology.

Prediction infers that these are membrane proteins could result in better interaction with water [5]. The number of venom components in venomous snake's ranges from 50-200 toxins [22]. Snake venoms are important tools in toxicology, neuroscience, and pharmacology. The venom components are highly variable and functionally complex and they offer many research opportunities [11]. The main toxins from snake venom that affect the CNS are neurotoxins. Neurotoxins form one of the largest families of proteins with established primary structures.

PLA2 in its neurotoxic effect and its emerging importance can turn into potential targets in venomous disorders like inflammation, failure of nervous system, and some deadly disorders which opens new areas for future research.

In the present study the sequence and structure analysis of Basic PLA2 protein was done by various tools and software's. Based on the findings it could be concluded that further characterization of Venomous viper proteins is novel and will be important for evaluating how the regulation of this proteins is related in the complications connected to neurotoxicity. Although PLA2 proteins were initially identified as key members which participate in regulating the synaptic signaling and their strong association with snake bite and different types of deadly disorders. Because of the network of signals is rather complex and cell-context-dependent, further studies may help to establish the relevance of individual family members as neuronal predictors and therapeutic targets for venomous Snake bites. Moreover, bioinformatics studies will aid in the development of improved molecular tools for the study of Basic PLA2 proteins. Identification of novel PLA2 protein functions and crucial signaling events provide additional targets and new therapeutic approaches. Although significant progress has been made towards elucidating its role in neurotoxic effects and role as a Drug target. Additional work needed to fulfill the regulation of PLA2 protein as if Venomous venom is considered.

Moreover, bioinformatics studies will aid in the development of improved molecular tools for the study of proteins like PLA2. It is becoming clear that PLA2 may have many important functions, and hydropathicity might contribute to its role in signaling and immunogenic responses. Identification of novel PLA2 functions and crucial signaling events provide additional targets and new therapeutic approaches. Further work will be required in order to fully understand how PLA2 is regulated.

## REFERENCES:

1. Argiolas A, Pisano J.J and Pisano. "Facilitation of phospholipase A2 activity by mastoparans, a new class of mast cell degranulating peptides from wasp venom". J. Biol. Chem. 258 (22): 13697/702. 1983
2. Arnold K, Bordoli L and Kopp J.. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics. 22:195-201. 2006
3. Ashokan K and Pillai M..Characterization of biomimetic material, Mussel adhesive protein, using computer tools and servers. The Internet Journal of Bioengineering. *Vol* 4(1). 2008
4. Ashokan K. V, Mundaganur D. S, and Mundaganur Y D.. "Catalase: Phylogenetic characterization to explore protein cluster".Journal of research in Bioinformatics. 1:001-008. 2011
5. Burke J. E, and Dennis E.A. "Phospholipase A2 biochemistry," Cardiovascular Drugs and Therapy. 23(1): 49–59. 2009
6. Dennis E.A.. "Diversity of group types, regulation, and function of phospholipase A2".J. Biol. Chem. 269 (18): 13057/60. 1994
7. Eisenhaber F, Imperiale F, Argos P and Froemmel C.. Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. *Proteins:* Struct. Funct. Design.25:157-168. 1996
8. Gill S.C. and Von Hippel P.H. Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem.182: 319-326. 1989

9. Guruprasad K, Reddy B.V.P and Pandit M.W..Correlation between stability of a 273 protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot.Eng. 4 (2):155-164. 1990
10. Garnier O and Robson J. " Analysis of the accuracy and the implications of simple methods for predicting the secondary structure of globular proteins" . Mol. Biol. 120:97-120. 1978
11. Joyce P.Y.S, Khan A.M, Paul T. J, Tan, Koh J.L.Y, Seah S.H, Koo C.Y, Chai S.C, Armugam A, Brusic V, and Jeyaseelan K.. Systematic analysis of snake neurotoxins functional classification using a data warehousing approach. Bioinformatics. 20:3466 -3480. 2004
12. Koh D. C. I, Armugam A and Jeyaseelan K.. "Snake venom components and their applications in biomedicine". Cellular and Molecular Life Sciences. 63:3030-3041. 2006
13. Kitchen D.B, Decornez H, Furr J.R, and Bajorath J. "Docking and scoring in virtual screening for drug discovery: methods and applications". Nature reviews Drug discovery. 3(11):935-949. 2007
14. Letunic I, Doerks T and Bork P. "SMART 6: recent updates and new developments". *Nucleic Acids Res.* 37: 229/32. 2009
15. Madhu S and Mahesh P. " Sequence analysis of Semaphorin in tumor progression: An Insilico approach" International Journal of Asian Academic Research of Multidisciplinary .1(29): 407-423. 2015
16. Mahesh P, Divya P , Akshatha M , Prathima R and Lava Kumar C." Insilico characterization and phylogenetic analysis of novel probiotic bacteria in honey bees" International Journal of Asian Academic Research of Multidisciplinary .1(32): 337-357.2015
17. Mugilan A , Ajitha M. C , Devi and Thinagar.. "In silico Secondary Structure Prediction Method (Kalasalingam University Structure Prediction Method) using Comparative Analysis". Trends in Bioinformatics. 3(1):11-19. 2010
18. Nicolas J.P , Lin Y, Lambeau G, Ghomashchi F, Lazdunski M, Gelb M.H, Lin, Lambeau, Ghomashchi, Lazdunski, Gelb. "Localization of structural elements of bee venom phospholipase A2 involved in N-type receptor binding and neurotoxicity". J. Biol. Chem. 272 (11): 7173/81. 1997
19. Prashanth V.T, Uddhav S. C, Madura S. M, Vishal P. D, and Renuka R.K. "Secondary Structure Prediction and Phylogenetic Analysis of salt Tolerant Proteins".Global Journal of molecular Sciences. 5(1):30-36. 2010
20. Ramachandran G.N, Ramakrishnan C, and Sasisekharan V."Stereochemistry of polypeptide chain configurations". J. Mol. Biol. 7: 95-9. 1963
21. Schultz J, Milpetz F, Bork P, Ponting CP.. "SMART, a simple modular architecture research tool: identification of signaling domains". Proc. Natl. Acad. Sci. U.S.A. 95 (11): 5857/64. 1998
22. Tsetlin V.I, and Hucho F." Snake and snail toxins acting on nicotinic acetylcholine receptors. Fundamental aspects and sciences", Bioinformatics. 4: 53-62. 2004
23. Tramontano A, Leplae R, and Morea V. "Analysis and assessment of comparative modeling predictions in CASP4". Proteins. 45(5):22-38. 2001
24. Vinita H. "Physiochemical, Functional and Structural Characterization of Wheat Germin Using In silico Methods".Current Research Journal of Biological Sciences. 3(1): 35-41. 2011
25. Warren G.L, Andrews C.W, Capelli A.M, Clarke B, LaLonde J, Lambert M.H, Lindvall M, Nevins N, Semus S.F, Senger S, Tedesco G, Wall I.D, Woolven J.M, Peishoff C.E, and Head M.S.. B-95 critical assessment of docking programs and scoring functions. J Med Chem. 49(20): 5912-31. 2006
26. ZIkai A. Thermo stability and aliphatic index of globular proteins. J. Biochem. 88:1895-1898. 1980