WestminsterResearch

http://www.wmin.ac.uk/westminsterresearch

Animal toxins: what features differentiate pore blockers from gate modifiers.

Saranjit Bhogal[1]
Kenneth Revett[2]

1 School of Biosciences
[2] Harrow School of Computer Science

Animal Toxins: What Features Differentiate Pore Blockers From Gate Modifiers?

Saranjit Bhogal, and Kenneth Revett

Abstract — **A surprisingly large number of animal toxins target voltage sensitive ion channels. Even though there exists toxins for all four major voltage sensitive ion channels, a majority act either on sodium or potassium channels. Given a specific primary sequence, the challenge is to determine in an automated fashion whether a given substance is toxic, and what its site of action might be. Currently, there are signals such as functional dyads that are indicative of a toxin, but are not yet specific enough to allow accurate prediction of the site of action. In this paper, an automated approach for detecting whether a toxin acts on voltage-sensitive sodium versus potassium channels is presented. In addition, our consensus sequence is also able to reliably determine whether the toxin acts as a gate modifier or pore blocker (> 93% accuracy).**

Keywords — **gate blocker, multiple sequence alignment, pore blocker, venom, and voltage-sensitive ion channels**

I. INTRODUCTION

Evolution and the natural world have endowed animal species with the ability to produce a bewildering array of toxic substances – for both protection and predation. The effects of toxins range from being mildly irritating to lethal on the inflicted victim. Toxins can be broadly classified as either venoms or poisons.[1-3]. A venom is a cocktail of many different components (enzymes and peptides) that are synthesized by a particular organism and stored in specialised glands until required. The venom is always delivered into the prey through teeth, stingers or harpoons and work much like a hypodermic needle. A poison is generally a chemical substance that may not even be produced by the organism it is found in. It if often produced by microorganisms and stored in the host organism, such as under the skin in the case of the Poison Arrow Frog. It is not delivered into a host through penetration but through absorption. Animals produce venoms as peptides (proteins) and crude (whole) venom can contain many different types of active substances. Toxins interfere with the normal functioning of specific cell types within the prey organism, although

Saranjit Bhogal is from the Cavendish School of Biosciences, University of Westminster, London, England, UK. (e-mail: g0uk1@hotmail.com).

Kenneth Revett is from the harrow School of Computer Science, University of Westminster, London, England, UK HA1 3TPT. (e-mail: Biomodelling@aol.com).

the mechanism(s) of action of several toxins remain to be elucidated. [4].

In general, the most common sites of action of known toxins can be classed into the following categories [5]:

- Neurotoxins - cause paralysis or interfere with the central nervous system.
- Haemotoxins - these affect the blood, particularly
- clotting abilities.
- Myotoxins - cause damage to muscle.
- Haemorrhagins - damage blood vessels resulting
- in bleeding.
- Haemolysins - cause damage to red blood cells.
- Cardiotoxins - directly affect the heart.
- Nephrotoxins - directly affect the kidneys.
- Necrotoxins - cause the death of tissue.

Considering the wide range of toxicities exhibited by animal toxins, it would be very helpful to be able to have a structure function relationship for a given toxin. This task has not yet been accomplished, as there appears to be a small number of structural features (i.e. protein folds) for the vast array of toxins that produce overlapping mechanisms of action. With respect to toxins acting on ion channels, multiple folds may act on the same ion channel type (e.g. $\alpha\alpha$ or $\beta\alpha\beta\beta$) [6,7]. In addition, there are folds that act across multiple channel types such as the. $\alpha\alpha$ fold which acts on *both* potassium and sodium channels. This lack of specificity has made the automated determination of site of action versus structure very difficult. What is still required is a clear consensus sequence that relates structure to site of action.

We have attempted to address the issue of generating a consensus sequence – focusing solely on potassium and sodium channels. These ion channels are the targets of the majority of the protein folds (16/21) [7,8]. Ion channels are large macromolecular structures that are found in all excitable tissue found in the animal kingdom [8]. They form channels, continuous water filled pathways, that connect the extracellular and intracellular compartments of cells. The opening and closing of the channel is controlled by changes in membrane potential, which causes a "gate' to open and close. Ions are then allowed into the mouth (vestibule) of the channel. Channel types are determined by the specific ions that are allowed to pass through the pore. The specificity of a channel is conferred by a 'selectivity' filter, which determines which type of ion is allowed to pass through (e.g. cation or anion).

In addition, most channels are able to differentiate between a sodium and potassium ion for example, which is critical for proper functioning of excitable cells. There are thus two functional aspects of an ion channel – the gating mechanism that opens and closes based on the perceived membrane potential and the pore element that spans across the membrane and allows ions to pass through (see Figure 1 below). It has been reported that some toxins can selectivity bind to specific regions of these channels – binding either to the gating mechanism or the pore element. The challenge has been to find a direct association between the structure of the toxin and its specific site of action.
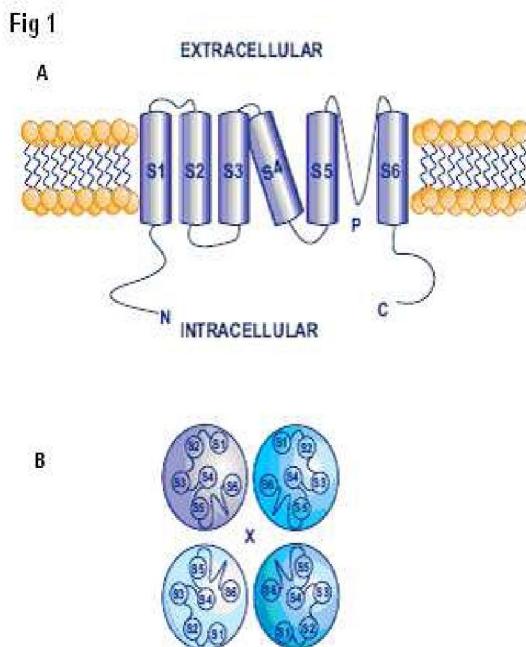

Figure 1. A cartoon of a typical ion channel, containing for subunits. In A) each subunit is composed of 6 transmembrane elements that span the entire lipid bilayer. In B), a superior view where you are looking down into the mouth of the ion channel

Animal toxins from a variety of species ranging from scorpions to mollusks bind all four major ion channels with the following predominance: sodium > potassium > calcium > chloride. This trend is substantiated by the reported number of protein folds for these particular categories: 8, 8, 4, and 1 respectively. It should be noted that there are roughly twice as many sodium channel toxins than potassium [9]. There are several internet-based databases that contain specific information about various toxins. Primary structures of various toxins have been determined and are available through online public databases [10] such as those The National Centre of Biotechnology Information (http://www.ncbi.nlm.nih.gov/) and The ExPASy website (http://ca.expasy.org/). Specialized databases, such as ToxProt (http://www.expasy.org/sprot/tox-prot/), have been established which contain specific peptide sequences and additional relevant biological information. Species specific databases have also been established to further

develop understanding of all known toxins, such as the Scorpion only database, Scorpion (http://research.i2r.a-star.edu.sg:8080/scorpion/).

Studies have revealed certain similarities that appear to be highly conserved across a bewildering array of species [11]. For instance, most toxins contain a functional dyad, consisting of a Lys residue and an aromatic (Tyr or Phe) or aliphatic (Leu) residue separated by approximately 7 Å. The Lys residue is generally positioned in the centre of a ring which is composed on negatively charged residues (carbonyl groups_ of four acidic residues (e.g. Asp) from each of the ion channel subunits. In addition, Mouhat reported the existence of a "triad" functional group, which consists two homologous hydrophobic residues (Tyr, Trp, or Phe) act in unison in place or alongside the Lys residue – otherwise it is virtually identical to the functional triad mentioned above [7]. In addition, another characteristic motif is the "Inhibitor Cystine Knot" (ICK) which consists of a ring of residues formed by 2 pairs of disulfide bridges (C1-c4 and C2-C5), with a 3rd disulfide bridge penetrates (C3-C6) forming a knot-like structure [7,8]. This motif presents an extremely stable and compact structure that can serve as scaffolding upon which various side groups can be placed.

Almost all of the toxins described in this paper can be classified according to a particular fold, based on the relative numbers and positions of their secondary structural elements (α-helices and β-sheets) and a disproportionately large number of disulfide bonds (2-5). Considering the small size of most animal toxins (on the order of 22-74 residues), the large number of disulfide bridges appears to impart extreme stability, enabling them to withstand proteolysis and denaturation. The small number of residues, and the highly conserved nature of the cystine bridges has made structure/function determination quite difficult, if specificity is required. Considering the mentioned difficulties, this is still a very active area of research, as suitable results have far reaching consequences with respect to anti-venom development, ion subtype specificity in disease, and basic neurophysiological research.

In this work, we sought to determine if we could discover one or more consensus sequences for sodium and potassium channels. In addition, we wished to determine if a consensus sequence could be obtained that would provide a classification based on whether they acted on the pore or the gate. Vestibule blockers have also been reported, which bind the opening of the pore from the outside – as opposed to inserting into the pore itself. There are very few instances of this type (vestibule blocker) and we have therefore classed them together with the pore blockers. The results would one to determine directly from sequence databases whether the toxin acted on sodium/potassium channels and also whether they would inactivate the channel either by binding to the gate or by blocking the channel pore.

This paper is organised as follows: the next section present a description of the basic methodology employed, followed by a presentation of the major results, and lastly by a brief conclusion section.

II. Methods

This study entailed the use of several internet based protein structure repositories. The basic outline of our consensus development strategy is as follows:

1. Obtaining the toxins targeting sodium and potassium channels
2. Refining toxins by extracting the active forms
3. Separate toxins based on their site of actions
4. Using various consensus sequence extraction tools sat as PRATT
5. Comparing the resultant consensus sequences thus obtained by doing a search through the PDB looking to see what the sensitivity and specificity of the resultant hit list was
6. PRATT consensus builder was employed
7. Database search using the generated consensus sequences
8. Repeat from 2 as necessary

Below we describe the process in more detail.

Obtaining the toxins targeting Na and K Voltage gate ion channels:

The keywords 'potassium channel inhibitor' were entered into the SRS [7,8] on the ExPASy server returning a list of peptide toxins targeting Kv channels, returning a list of 155 peptides. Sequences were in FASTA [8] format and output saved into a text document named Master_K.txt. FASTA format was chosen as it is recognised by many types of Bioinformatics analysis tools available online [10]. The process was repeated for the Sodium channel toxins using the keywords 'Sodium channel inhibitor', returning 283 toxins.

Refining toxins by extracting active forms:

Toxins in ExPASy can contain entire active sequences but sometimes contain Signal peptides and/or Propeptide sequences. These regions were removed in order to obtain active peptide sequences only, as programs may take those features to generate results. Under the region 'Features' is where details are presented if available, that describes details such as domains and disulphide bonds. e.g. Charybdotoxin b precursor from the scorpion *Leiurus quinquestriatus hebraeus*, SWISSPROT id P59943:

```
MKILSVLLLA    LIICSIVGWS    EAQFTDVSCT
TSKECWSVCQ  RLHNTSIGKC  MNKKCRCYS
Key   From To  Length Description
SIGNAL 1   22   22      by similarity.
```

MKILSVLLLA LIICSIVGWS EA

(leaving the active peptide):

```
QFTDVSCT     TSKECWSVCQ     RLHNTSIGKC
MNKKCRCYS
```

These features in ExPASy are not all experimentally derived and have been assigned 3 types of comments [10].
- Potential
- Probable
- By similarity

'Potential' - there is some logical evidence that given annotation could apply. This non-experimental qualifier is often used to present the results from protein sequence analysis tools if the results make sense in respect to a given protein [10,11].
'Probable' - a stronger indicator than 'Potential', and is based on some experimental evidence, that the given information is expected to be found in the natural environment the protein [12,13]. 'By similarity' - facts proven for the protein or part of it, and then transferred to other protein family members within a certain taxonomic range. Sites within conserved domains to each other include active sites and disulfide bonds [14]. The Master files contain active toxin peptides, were 'saved as' Primary files: Prim_Na.txt and Prim_K.txt files. Master files remained untouched since downloading.

Separating toxins by site of action:

The toxins in the Primary files were separated by site of action, sites where determined through literature reading from associated links of ExPASy. The resulting data was stored on disc for further analysis (i.e. see below).

Prim_K.txt: K_pores.txt & K_gaters.txt
Master_Na.txt: Na_pores.txt & Na_gaters.txt

K_files:

SWISSPROT ID's for each toxin in the Prim_K.txt file was entered into ExPASy. The resulting page has links to related literature (published) that were reviewed to determine the toxin as a pore blocker or a gate modifier. Once determined, the toxin entry was cut and pasted into the respective file.

Na files:

The process for Kv toxins was repeated, however due to the structural differences between the channels, Nav was found to have 6 binding sites, 5 of which are associated with the gate (sites 2-6) and site 1 the pore.

Veratridine, Batrachotoxin and Grayanotoxin are toxins acting on site 2, but are not peptides. They were not present on the list of the ExPASy output for Sodium channel inhibitors. They are poisons that are stored in the organism's body (Batrachotoxin) or in seeds (Veratridine). These toxins are thought to be created by microorganisms, living within the organisms, and stored by the host. Brevetoxin and Ciguatoxin acting on site 5 again are not peptide toxins that were returned on the output for Sodium channel inhibitors, Therefore these toxins will be excluded.

PRATT: Consensus builder

PRATT (http://www.ebi.ac.uk/pratt/) is part of EBI (http://www.ebi.ac.uk) and generates consensus sequence

motifs from unaligned fasta files. ExPASy has many other databases and tools for analysis of proteins and PROSITE will be initially used to search against the generated consensus. PRATT search input parameters can be modified to the user preferences, i.e. consensus must match at least 50% of inputted sequences. The output of PRATT is in PROSITE format and feed directly into the PROSITE database search [14].

I. DATABASE SEARCH

PROSITE is a database that can be searched when a Motif is entered as a parameter (http://www.ExPASy.org/prosite/). The website is able to understand patterns and motifs by using the accepted one letter abbreviations of the amino acids (e.g. G is Glycine), i.e. the usual format of a typical pattern/motif can be something like:

G-[ASP]-V-X(2)-GLA-{SP} (1)

Letters correspond to their amino acids
'-'   separates the next amino acid
X   calls for any amino acid (the number in the brackets determines how many of the preceding letter).
{}   Amino acids within curly braces are not to be included in the pattern/Motif search.

Example Interpretation (1 above):

Starting with a Glycine, followed by Alanine, Serine or Proline, followed by a Valine, then by any 2 amino acids, followed the amino acid order Glycine, Leucine and Alanine and finally ending on any amino acid barring Serine or Proline.

III RESULTS

Potassium - GATE
The gate (voltage sensor) of the K_v channel currently has two known folds targeting it. These folds are: $\beta\beta\beta$, $3_{10}\beta\beta$ and 2 non-categorised folds $\beta\beta$ and $\beta\beta\beta\beta$.

The consensus that was applied to the structures was:
C-X(3)-[WMILF]-X(9,10)-C-X(0,3)-[REKH]-X(1,5)-C-X(3,10)-C   (2)

This consensus finds 19 out of the 20 toxins that have been classed to target the potassium gate. The only toxin that does not contain this consensus is Kappa-conotoxin BtX from the Cone snail (beach Cone).

$\beta\beta\beta$ GsMTx-2 (Chilean rose tarantula)

MTX2_GRASP: 2- 25:   y
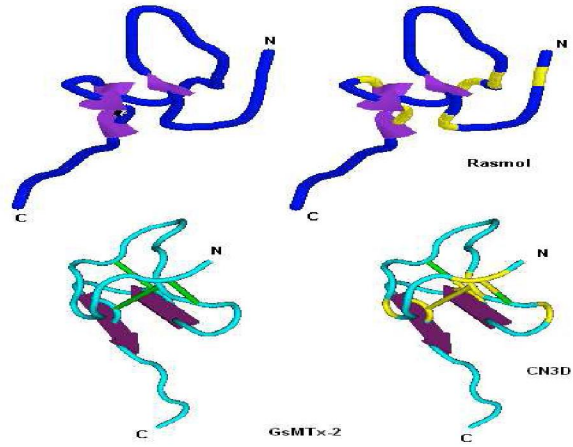
CqkwMwtcdeerk-CcEglv----Crlw-------C kriin

Figure 1. Note: Both structure files only show 2 beta sheets. These Rasmol (top) and CN3D (bottom) images present the structural configurations of 2 gate blocking toxins specific to potassium channels

II.

III. POTASSIUM - PORE

Starting with the potassium channel pore consensus:
C-X(9,12)-C-X(2,5)-C (3)
Folds include: $\beta\beta\beta$, $\alpha\alpha$ (hairpin), $\alpha\alpha$ (helical cross), $3_{10}\alpha\alpha$, $\alpha\beta\beta$, $\beta\alpha\beta\beta$ and $3_{10}\beta\beta\alpha$
$\beta\beta\beta$ PRATT ouput: CXK7A_CONPU*:   8- 26:
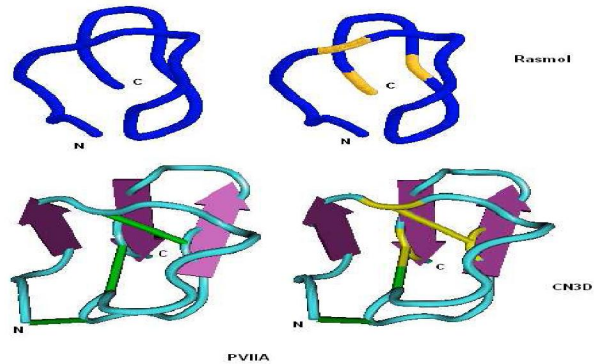Cfqhlddccsrk-CnrfnkC   (purple cone)

Figure 2. Note: Both structure files only show 2 beta sheets. These Rasmol (top) and CN3D (bottom) images present the structural configurations of 2 pore blocking toxins specific to potassium channels

IV. SODIUM - PORE

Site 1 is the pore of the sodium channel. The folds that recognise the Na_v Pore are: $\beta\beta$, $\beta\beta\beta$, and $\beta\alpha\beta\beta$
The Pore consensus will be applied to a structure that represents each fold.
C-x(3,6)-C-x(4,6)-C-[ACGN] (4)

Matches 15 out of 16 known Nav pore blocking toxins.

Example displayed: $\beta\beta$ - Conotoxin GS (Cone snail)
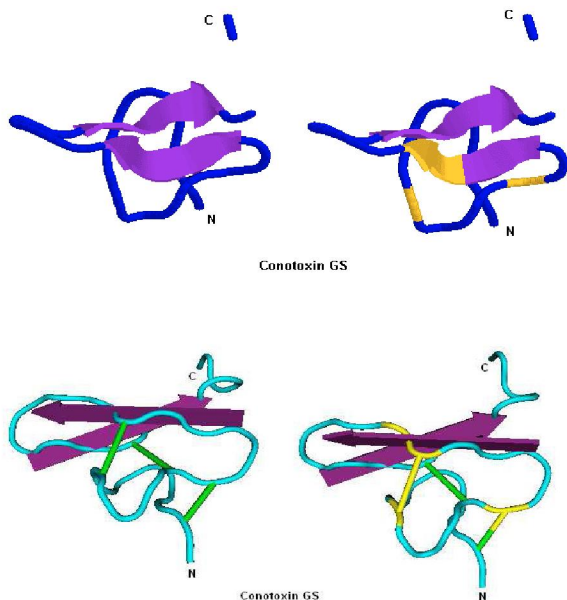CXGS_CONGE: 9-20: grgsr Cppqc--Cmglr--CGrgnpq

Figure 3. Note: Both structure files only show 2 beta sheets. These Rasmol (top) and CN3D (bottom) images present the structural configurations of 2 pore blocking toxins specific to sodium channels

V. SODIUM - GATE

Site 3 best consensus: **C-X(6,9)-C-X(0,6)-C**
This consensus matches some toxin sequences six times, Neurotoxin Tx2-6 from the Brazilian armed spider. As a result alignment B will be chosen as it matches some toxin sequences only 3 times.

$$C\text{-}x(5,7)\text{-}C\text{-}x(10,13)\text{-}C \quad (5)$$

All folds except $\beta\beta\alpha\beta\beta\alpha$ target site 3 of the sodium channel. Each fold type will be shown:
$\beta\beta$ Robustoxin (funnel web spider)

```
  TXDT1_ATRRO: 1-20: Cakkrnw-Cgknedcccpmk--C
iyawy
 TXDT1_ATRRO:14-31: gkned Cccpmk--
Ciyawynqqgs---Cqttit
```

Consensus finds two matches:
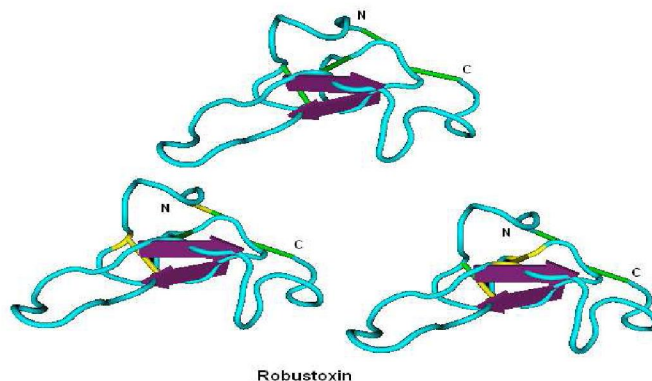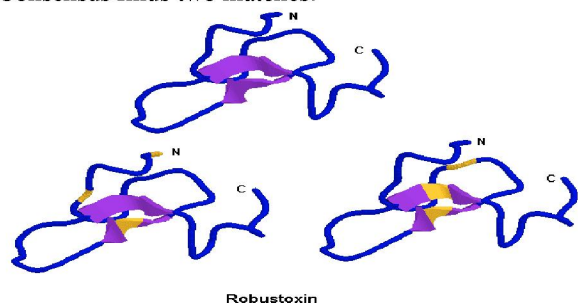

Robustoxin

Robustoxin

Figure 4. Note: Both structure files only show 2 beta sheets. These Rasmol (top) and CN3D (bottom) images present the structural configurations of 2 gate blocking toxins specific to sodium channels

Table 1. The summary statistic of the overall consensus sequences for each of the categories – potassium pore and gate along with sodium pore and gate. The last column indicates the accuracy of the search based on the consensus sequences from the PDB

Site		Summary consensus table	
Toxin Target	**No of toxins**	**Consensus**	**% Found**
K Pore	127	C-X(9,12)-C-X(2,5)-C	93.7
K Gate	20	C-X(3)-[WMILF]-X(9,10)-C-X(0,3)-[REKH]-X(1,5)-C-X(3,10)-C	95
Na Pore (site 1)	16	C-X(3,6)-C-X(4,6)-C-[ACGN]	93.75
Na Gate Site 3	143	C-X(5,7)-C-X(10,13)-C	93
Na Gate Site 4	96	C-x(0,1)-K-x(2,8)-C-x(6,8)-C	93.67
Na Gate Site 6	7	C-X(3,4)-G-X(1,2)-C-X(1,4)-[FIL]-X(2)-[GN]-X(0,1)-C-C-X(3,4)C-X(2,3)-[FIV]-C	100
All Na Gates	247	C-X(3,6)-C-X(9,11)-C	93.9
All K Toxins	145	C-X(11,14)-C-X(2,5)-C	91.72
All Na Toxins	262	C-X(3,6)-C-X(10,13)-C	93.13

Finally a consensus was determined for all potassium and sodium channel toxins. This produced an interesting output, in that the individual consenses seem to be a reverse of each other.

Potassium toxins   **C-X(9,12)-C-X(2,6)-C   (6)**
Sodium toxins   **C-X(3,6)-C-X(5,9)-C   (7)**

The potassium toxin consensus has a long spacer region between the first and second Cystiene and the sodium has a longer spacer region between the second and third Cystiene. The predominant type of channel toxin for the potassium channel seems to be Pore specific and for the sodium channel the gate conversely. The relationships between these observations are not clear but may help t o

understand the differences between the two types of ion channel toxins.

IV. CONCLUSIONS

In this study, we developed a set of consensus sequences that could differentiate potassium from sodium channel acting toxins. The algorithm we employed was able to generate consensus sequences for potassium gate and pore blockers (2 and 3 respectively) as well as sodium channel gate and pore blockers (4 and 5 respectively). In addition, we were able to come up with a consensus sequence that was able to generically differentiate potassium from sodium channel toxins (6 and 7 respectively). To date, no literature report has provided a consensus sequence that could distinguish pore from gate blockers for either sodium or potassium channels. This is the novel result from this study. When the consensus sequences we have generated are entered into a standard protein sequence database such as PDB, we find that the accuracy – in terms of number of hits per known number of toxins acting on that particular site is quite high - on the order of 93% or higher in this study. It should be noted however that consensus generation has been optimal when sequence numbers are low or toxins are from a related family. The toxins that act on the channels in this project are from a wide range of organisms. There is no ideal consensus sequence that can *perfectly* identify a specific type toxin; however the types of folds the toxins adopt shows specificity towards toxins site and also species specificity.

In future work, we will investigate the two remaining voltage gated ion channels, the chloride and calcium channels. It will be interesting to investigate the similarities in protein folds between all four major ion channels to see if there is an overall relationship between them. In addition, it should also be noted that during the course of this study, we discovered 2 new protein folds for potassium gate blockers that have not yet been reported in the literature ($\beta\beta$ and $\beta\beta\beta\beta$). We will pursue these toxin structures in further detail in a follow up study.

REFERENCES

[1] Miller C, Naranjo D, 1996. A strongly interacting pair of residues on the contact surface Charybdotoxin and a shaker channel. *Neuron.* 16:123-130.

[2] Possani, L.D., E. Merino, M. Corona, F. Bolivar, and B. Becerril. 2000. Peptides and genes coding for scorpion toxins that effect ion-channels. *Biochemie.* 82:861-868

[3] Tan P.T., Khan A.M and Brusic V. 2003.Bioinformatics for venom and toxin sciences. Brief Bioinformatics. 4: 53-62.

[4] Jungo F., Bairoch A. Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein Knowledgebase. 2004. Toxicon. 45: 293-301.

[5] Elsworth J. Deadly by Nature, Web Project. http://www.chm.bris.ac.uk/webprojects2002/elsworth/deadly_by_nature.htm

[6] Schonbach C., Kowalski-Saunders and Brusic V. 2000. Data warehousing in molecular biology. Brief Bioinformatics. 1: 190-198.

[7] Mouhat S., Besma J., Mosbah A., De Ward M., and Sabatier JM. 2004. Diversity of folds in animal toxins acting on ion channels. Journal of Biochemistry. 378: 717:726

[8] Mouhat s, Waard M and Sabatier JM. 2005. Contribution of the functional dyad of animal toxins acting on voltage-gated Kv1-type channels. Journal of Peptide Science 11: 65–68.

[9] Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. *ExPASy: the proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Res. 31:3784-3788(2003).

[10] Boeckmann B., Bairoch A., Apweiler R., Blatter M.-C., Estreicher A., Gasteiger E., Martin M.J., Michoud K., O'Donovan C., Phan I., Pilbout S. and Schneider M.: 2003. The Swiss-Prot protein knowledgebase and its supplement TrEMBL in. Nucleic Acids Res. 31:365-370

[11] Jonassen I, Collins JF and Higgins D. Finding flexible patterns in unaligned protein sequences. 1995. Protein Science ;4(8):1587-1595.

[12] Inge Jonassen . 1997. Efficient discovery of conserved patterns using a pattern graph. CABIOS 13, 509-522.

[13] Bucher P., Bairoch A. 1994. A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAIPress, Menlo Park,

[14] Sigrist C.J.A., Cerutti L., Hulo N., Gattiker A., Falquet L., Pagni M., Bairoch A., Bucher P. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 3:265-274.