1-1-2007

Structural comparison of scorpion activator peptides

Regina Sampson

NORTHERN ILLINOIS UNIVERSITY

Structural Comparison of Scorpion Activator Peptides

A Thesis Submitted to the

University Honors Program

In Partial Fulfillment of the

Requirements of the Baccalaureate Degree

With Upper Division Honors

Department of

Biological Sciences

By

Regina Sampson

DeKalb, Illinois

May 2007

University Honors Program

Capstone Approval Page

Capstone Title (print or type):

Structural Comparison of Scorpion Activator Peptides

Student Name (print or type): Regina Sampson

Faculty Supervisor (print or type): RICHARD HAHIN

Faculty Approval Signature: Richard Hahin

Department of (print or type): Biological Sciences

Date of Approval (print or type): 3 May 07

HONORS THESIS ABSTRACT
THESIS SUBMISSION FORM

AUTHOR: Regina Sampson

THESIS TITLE: Structural Comparison of Scorpion Activator Peptides

ADVISOR: Richard Hahin

ADVISOR'S DEPT: Biological Science

DISCIPLINE: Biology

YEAR: 2007

PAGE LENGTH: 19

BIBLIOGRAPHY:

ILLUSTRATED:

PUBLISHED (YES OR NO): NO

LIST PUBLICATION:

COPIES AVAILABLE (HARD COPY, MICROFILM, DISKETTE): Hard Copy

ABSTRACT (100 – 200 WORDS):

Abstract:

Scorpion activator polypeptides specifically bind to sodium channels, however the mechanism of binding is unknown. A hypothesis has been advanced by my project mentor (Dr. R. Hahin) that specifies the part of the molecule that binds and interacts with the channel. Also, certain parts of the activators play a key role in the maintenance of its three dimensional structure. The focus of the project is to compare peptides to further test the hypothesis and obtain important structural information that may well be common to all the molecules. Studies and analysis were achieved by using bioinformatic tools such as NCBI, EMBO, RSCB, SWISS-PROT, and SDMC to locate toxins. Clustal W alignment program was used to perform a multiple sequence alignment of the peptides and Deepview was used to identify key structural regions. New information was gained about how activator peptides bind to and modify sodium channels. Important structural and sequence motifs of the molecules were also identified.

Abstract:

Scorpion activator polypeptides specifically bind to sodium channels, however the mechanism of binding is unknown. A hypothesis has been advanced by my project mentor (Dr. R. Hahin) that specifies the part of the molecule that binds and interacts with the channel. Also, certain parts of the activators play a key role in the maintenance of its three dimensional structure. The focus of the project is to compare peptides to further test the hypothesis and obtain important structural information that may well be common to all the molecules. Studies and analysis were achieved by using bioinformatic tools such as NCBI, EMBO, RSCB, SWISS-PROT, and SDMC to locate toxins. Clustal W alignment program was used to perform a multiple sequence alignment of the peptides and Deepview was used to identify key structural regions. New information was gained about how activator peptides bind to and modify sodium channels. Important structural and sequence motifs of the molecules were also identified.

Theory:

Sodium channels are large proteins that are distributed along nerve, heart and muscle cell membranes in our bodies. The proteins act as a pore or channel in which sodium ions can flow into or out of a cell and produce voltage changes in our cells. The channels can exist in three states: open, closed or inactivated. When stimulated sodium channels typically open, inactivate, and then close to produce a propagated signal called an action potential (AP). AP's act as signals throughout our nervous system, heart and skeletal muscles to cause thoughts, movements, and life processes (heart beats; breathing, blood circulation, etc.).

A number of substances can bind to and alter sodium channels and alter human health. Only a few substances bind specifically to sodium channels. Many chemical agents such as anesthetics bind to and alter sodium channels, but also can alter many other proteins in a human's body. A toxin from the Japanese Puffer Fish, tetrodotoxin is an example of an agent that specifically binds to and alters sodium channels to cause loss of sensation by eliminating AP production. Similarly scorpion polypeptides act to bind to and alter the behavior of sodium channels and alter AP propagation. Alpha scorpion polypeptides are one of the smallest proteins found and act to alter sodium channels so that they open but do not inactivate normally. Alpha scorpion polypeptides act to cause long lasting AP's that are quite harmful to humans because they disrupt the signaling ability of the nervous system, the heart, and muscle. However, the alpha scorpion polypeptides are extremely important molecules because they are one of the few molecules that specifically and tightly bind to sodium channels so that they can be used as a research tool to study the channel's structure and function.

Since scorpion activator polypeptides bind specifically to sodium channels they could potentially be used to modify the sodium channel's behavior without affecting any other biological process. If mutations are made in the activator peptides so that the basic structure of the polypeptide is retained so that it continues to bind to sodium channels but now has reduced potency, it could be a potentially useful therapeutic agent. It could be used to lengthen heartbeats in people with failing hearts, or it could alter skeletal muscle contractions.

Methodology:

This is both a qualitative and quantitative study where both primary and secondary databases were used.

- **Obtain Sequences—**
- **Obtain Multiple Alignment of Sequences—**
- **Identify Key Structural Regions—**
- **Calculate and Specify the Most Probable Dihedral Angles for Tight Turns and Helices—**
- **Compare Results with Similar Results from Other Polypeptides/ Proteins—**

Results:

Amino acid sequences and gene sequences for alpha toxins were the very first pieces of information obtained. Bioinformatics tools available on various data bases such as NCBI, EMBO, RSCB, SWISS-PROT, and SDMC were used to find various toxins and sequences. A set of amino acid sequences and gene sequences were obtained for all available molecules that act to bind to and alter sodium channels. The database that was used most frequently to obtain the sequences was SDMC website database. Sequences and toxins obtained were then compared to previous toxins. A couple of knew toxins were identified. These toxins are still being studied. All other known alpha toxins and their sequences are shown in Figure 1. It was observed that two of the toxins had a prime to go along with the original toxin. A good example of this is the very first toxin in Figure 1, which is Aah1' and the second one is Aah1. The other toxin that had a prime was BmKM4 and BmKM4'. Most of the toxins are not in any type of order in Figure 1,

so the next step was to see how well they relate to each other in certain positions in their sequences.

A multiple alignment of all the peptide sequences was performed on all the found toxins. Using a bioinformatics alignment program called Clustal W, peptide sequences were aligned to produce maximum identity of amino acids. This program found similarities at many analogous positions. The whole multiple sequence alignment is shown in Figure 2. In analyzing the results of the alignment, there seemed to be many pairs of toxins that went as far as having 25 aligned peptides. The significance of both the toxins with primes was identified as having extreme similar sequences. Aah1 and Aah1' expressed all but one identical position. At the 36^{th} position of the alignment Aah1 had a V while Aah1' had an I instead. The rest of the alignment was identical. As for BmKM4 and BmKM4', they posses two unlike positions in their sequences. The first mismatch alignment is in position 15, where BmKM4 expresses a K while BmKM4' expresses a T. The next mismatch alignment is in position 37, where BmKM4 contains a G while BmKM4' contains a T. In observing the rest of the toxins in alignment, it was noticed that some sequences include long strands of dotted lines. These were later identified as precursor sequences that are not counted in the total alignment. They were used in order to start the sequences at the correct positions. For example, AaHIV's sequence started at a much earlier position and it was also longer than LqHVIII, which is the sequence right next to it. My final observation of the multiple sequence alignment was that the toxins with like names where bunched more closely together. An example of this is on page 13 in Figure 2. The toxins CsEV1, CsEV2, CsEV3, CsEV4, and CsEV5 were all in a row

with each other, but they did not go in order. The exact order was 2, 3, 1, 4, and then 5. This just showed that toxins in the same classes are more sequence related.

To get a closer look at the actual structure and to find similarities among the toxins, the key structural regions were identified. Samples of 13 toxins were taken through this process. Polypeptides possess key structural elements such as helices, beta strands, and tight turns. Each was identified in number and in structure. While analyzing each, it was discovered that the samples of toxins all had at least one alpha helix and one tight turn. Beta strands were also not seen in all the toxin samples. An alpha helix is a common structure of proteins, characterized by a single, spiral chain of amino acids stabilized by hydrogen bonds. A tight turn is described as reverse turns where the strand will go one way and quickly, or tightly, change to go the opposite way. The focus of the rest of this research will be geared towards alpha helices and the tight turns of the proteins in toxins.

It is hypothesized that reverse tight turns are a part of the protein structure that interacts with the sodium channel. To discover any unique aspects of the tight turns, the most probable dihedral orientations, or angles, were calculated for each of the 13 protein samples. It was observed that all the proteins included a first tight turn. This turn was large and obvious in its structure. Previous studies have been performed on first tight turns and the dihedral angles were found. In this research, the second, third, fourth, and even fifth tight turn was found in some of the 13 samples of proteins. All of the 13 protein included a second tight turn. Using a program called Deepview, each of the proteins were set up by themselves to be analyzed. The protein had to be seen in ribbon form in order to properly identify the correct fragment. The sections including the tight

turns were then highlighted. The amino acids of the section was observed along with the position number they were in. A Ramachondron Plot was made from the highlighted amino acid section. The values of the plot give the dihedral angles of the tight turns. Figure 3 illustrates all the values and numbers of tight turns each protein expressed. It was surprising to see that 1PE4 has as many as five tight turns. When observing the results of tight turn two, three, four, and five, there did not seem to be any distinct patterns or orders. There was a slight similarity in the omega (ω) angel of all the tight turns. It seemed to range from 170.00 to 179.99, whether the number was negative or positive. The phi (φ) and psi (Ψ) calculations did not seem to have any pattern at all. They each had high, low, positive, and negative numbers all over. When comparing second tight turn to the first tight turn in all 13 of the sample proteins, there seemed to be some type of structure and order to the first tight turns. All of the omega angles were in the high 170's, except for three of them. They were also, either negative or positive. The phi angles were in the 50's to 70's range and they were all negative except for one of them. The psi angles were all negative and in the 20's to 40's range. They clearly show a decrease in angle measurement from the omega to the psi. This observation leads to the explanation that first tight turns are unique in its dihedral measurements.

There have been many experiments and studies done to identify the potential site that scorpion alpha-toxins bind to the sodium channel. There is still not a definitive answer as to exactly where this binding occurs but research is getting closer to discovering the answer. In an article titled, *Identification of Structural Elements of a Scorpion -Neurotoxin Important for Receptor Site Recognition*, the authors try to get down to the bottom of this issue. The experiment they performed used an efficient

bacterial expression system for modifying specific amino acid residues of highly insecticidal alpha-neurotoxin LqhalphaIT from the scorpion *Leiurus quinquestriatus hebraeus*. They modified the toxins at tight turns, the C-terminal region, and other structurally related regions by subjecting them to neuropharmacological and structural analyses. The scientists took AaHII, the most potent anti-mammalian scorpion alpha-toxin and made it compete with LqhalphaIT for a binding site on an insect sodium channel. The authors suggest that there is a multipoint interaction and that binding to the receptor site may involve electrostatic as well as hydrophobic interactions with opposing constituents at the receptor site. Their conclusion is in accordance with their experiment, but the definite answer for all scorpion alpha-toxins was not identified

In doing this study, the reverse tight turns and helices have structures and dihedral angle measurements that suggest their uniqueness. Using data obtained from this experiment as well as others in literature, it is proposed that these two structural elements play a key role in the action of the polypeptides in scorpion alpha-toxins. The final conclusion to this topic is yet to be resolved.

Reference:

Articles from the Scientific Literature:

Bouhaouala-Zahar, B., R. Benkhalifa, N. Srairi, I. Zenouaki, C. Ligny-Lemaire, P. Drevet, F. Sampieri, M. Pelhate, M.E. Ayeb, A. Menez, H. Karoui, and F. Ducancel.
2002. A chimeric scorpion alpha-toxin displays de novo electrophysiological properties similar to those of alpha-like toxins. *Eur. J. Biochem.* 269:2831-2841.

Gilles, N., E. Leipold, H. Chen, S.H. Heinemann, and D. Gordon. 2001. Effect of Deploarization on Binding Kinetics of Scorpion alpha-toxin Highlights Conformational Changes of Rat Brain Sodium Channels. *Biochemistry.* 40:14576-14584.

Guan, R., Y. Xiang, X. He, C. Wang, M. Wang, Y. Zhang, E. Sundberg, and D. Wang. 2004. Structural Mechanism Governing Cis and Trans Isomeric States and an Intramolecular Switch for Cis/Trans Isomerization of a Non-proline Peptide Bond Observed in Crystal Structures of Scorpion Toxins. *J. Mol. Biol.* 341:1189-1204.

Hahin, R., Z. Chen, D. Wang, G. Reddy, and L. Mao. 2003. Scorpion toxins from *Buthus martensii* Karsch all possess a predicted α-tight-turn. *Cell Biochem. & Biophys.* 37:169-185.

Hamon, A., N. Gilles, P. Sautiere, A. Martinage, C. Kopeyan, C. Ulens, J. Tytgat, J. Lancelin, and D. Gordon. 2002. Characterization of scorpion alpha-like toxin group using two new toxins from the scorpion *Leiurus quinquestriatus hebraeus. Eur. J. Biochem.* 269:3920-3933.

He, X., H. Li, Z. Zeng, X. Liu, M. Wang, and D. Wang. 1999. Crystal Structures of Two alpha-like Scorpion Toxins: Non-proline cis Peptide Bonds and Implications for New Binding Site Selectivity on the Sodium Channel. *J. Mol. Biol.* 292:125-135.

Krimm, I., N. Gilles, P. Sautiere, M. Stankiewicz, M. Pelhate, D. Gordon, and J. Lancelin. 1999. NMR structures and activity of a novel alpha-like toxin from the scorpion *Leiurus quinquestriatus hebraeus. J. Mol. Biol.* 285:1749-1763.

Zilberberg, N., O. Froy, E. Loret, S. Cestele, D. Arad, D. Gordon, and M.Gurevitz.1997.Identification of Structural Elements of a Scorpion alpha-Neurotoxin Important for Receptor Site Recognition. *J. Mol. Biol.* 272: 14810-14816

Review articles:

Baltazar, B., M. Corona, C. Garcia, F. Bolivar, and D. Possani. 1995. Cloning of genes encoding scorpion toxins: An interpretative review. *J. Toxicol.—Toxin Reviews* 14339-357.

Possani, L., B. Becerril, M. Delepierre, and J. Tytgat. 1999. Scorpion toxins specific for Na+-channels. *Eur, J. Biochem.* 264:287-300.

Book articles:

Kyte, J. Evolution. *In* Structure in Protein Chemistry. Garland Publishing, Inc., New York. 243-279.

M.S. Thesis:

Yu, N. 2005. Structural analysis of scorpion α and α like toxins that bind to Na+ channels. Biol. Sci. Dept., Northern Illinois University.

Figure 1

>Aah1'
MNYLVMISLALLLMIGVESKRDGYIVYPNNCVYHCIPPCDGLCKKNGGSSGSCSFLVPSGLACWCKDLPDN
VPIKDTSRKCT

>Aah1
MNYLVMISLALLLMIGVESKRDGYIVYPNNCVYHCVPPCDGLCKKNGGSSGSCSFLVPSGLACWCKDLPDN
VPIKDTSRKCT

>AaHIII
MNYLVMISLALLLMTGVESVRDGYIVDSKNCVYHCVPPCDGLCKKNGAKSGSCGFLIPSGLACWCVALPDN
VPIKDPSYKCHSR

>AaHIV
MNYLIMFSLALLLVIGVESGRDGYIVDSKNCVYHCYPPCDGLCKKNGAKSGSCGFLVPSGLACWCNDLPEN
VPIKDPSDDCHKR

>LqHIII
VRDGYIAQPENCVYHCFPGSSGCDTLCKEKGGTSGHCGFKVGHGLACWCNALPDNVGIIVEGEKCHS

>LqHVIII
VRDGYIAKPENCAHHCFPGSSGCDTLCKENGGTGGHCGFKVGHGTACWCNALPDKVGIIVDGVKCH

>LqHVI
VRDGYIAQPENCVYHCIPDCDTLCKDNGGTGGHCGFKLGHGIACWCNALPDNVGIIVDGVKCHK

>LqHVII
VRDGYIAKPENCAHHCFPGSSGCDTLCKENGGTGGHCGFKVGHGTACWCNALPDKVGIIVDGVKCH

>OsIII
GVRDGYIAQPHNCVYHCFPGSGGCDTLCKENGATQGSSCFILGRGTACWCKDLPDRVGVIVDGEKCH

>BomIII
GRDGYIAQPENCVYHCFPGSSGCDTLCKEKGATSGHCGFLPGSGVACWCDNLPNKVPIVVGGEKCH

>BmKM4
MNYLVMISFALLLMKGVESVRDAYIAKPENCVYHCAGNEGCNKLCTDNGAESGYCQWGGRYGNACWCIKLP
DDVPIRVPGKCHR

>BmKM4'
MNYLVMISFALLLMTGVESVRDAYIAKPENCVYHCATNEGCNKLCTDNGAESGYCQWGGKYGNACWCIKLP
DDVPIRVPGKCHR

>BmK10
MNYLVMISFALLLMKGVESVRDAYIAKPENCVYECGITQDCNKLCTENGAESGYCQWGGKYGNACWCIKLP
DSVPIRVPGKCQR

>BmKM1
MNYLVMISFALLLMTGVESVRDAYIAKPHNCVYECARNEYCNDLCTKNGAKSGYCQWVGKYGNGCWCIELP
DNVPIRVPGKCHR

>BmKII
VRDAYIAKPHNCVYECARNEYCNDLCTKDGAKSGYCQWVGKYGNGCWCIELPDNVPIRIPGNCH

>BmKM2
VRDAYIAKPHNCVYECARNEYCNNLCTKNGAKSGYCQWSGKYGNGCWCIELPDNVPIRVPGKCH

>BmKM7
VRDGYIALPHNCAYGCLNNEYCNNLCTKDGAKIGYCNIVGKYGNACWCIQLPDNVPIRVPGRCHPA

>BomIV
GRDAYIAQPENCVYECAKNSYCNDLCTKNGAKSGYCQWLGKYGNACWCEDLPDNVPIRIPGKCHF

>BotI
GRDAYIAQPENCVYECAQNSYCNDLCTKNGATSGYCQWLGKYGNACWCKDLPDNVPIRIPGKCHF

>BotIII
VKDGYIVDDRNCTYFCGRNAYCNEECTKLKGESGYCQWASPYGNACYCYKVPDHVRTKGPGRCN

>BotII
GRDAYIAQPENCVYECAKNSYCNDLCTKNGAKSGYCQWLGRWGNACYCIDLPDKVPIRIEGKCHF

>BotIT1
VRDAYIAQNYNCVYFCMKDDYCNDLCTKNGASSGYCQWAGKYGNACWCYALPDNVPIRIPGKCHS

>BomPI
GRDAYIAQPENCVYECAKSSYCNDLCTKNGAKSGYCQWLGRWGNACYCIDLPDKVPIRIEGKCHFA

>LqqIII
VRDAYIAKNYNCVYECFRDSYCNDLCTKNGASSGYCQWAGKYGNACWCYALPDNVPIRVPGKCH

>BmK8
GRDAYIADSENCTYFCGSNPYCNDVCTENGAKSGYCQWAGRYGNACYCIDLPASERIKEPGKCG

>AaHII
MNYLVMISLALLFVTGVESVKDGYIVDDVNCTYFCGRNAYCNEECTKLKGESGYCQWASPYGNACYCYKLP
DHVRTKGPGRCHGR

>LqHII
IKDGYIVDDVNCTYFCGRNAYCNEECTKLKGESGYCQWASPYGNACYCYKLPDHVRTKGPGRCR

>BotXI
LKDGYIVDDRNCTYFCGTNAYCNEECVKLKGESGYCQWVGRYGNACWCYKLPDHVRTVQAGRCRS

>AmmV
LKDGYIIDDLNCTFFCGRNAYCDDECKKKGGESGYCQWASPYGNACWCYKLPDRVSIKEKGRCN

>LqqV
LKDGYIVDDKNCTFFCGRNAYCNDECKKKGGESGYCQWASPYGNACWCYKLPDRVSIKEKGRCN

>BmKalpha2
MNYMVIISLALLVMTGVESVKDGYIADDRNCPYFCGRNAYCDGECKKNRAESGYCQWASKYGNACWCYKLP
DDARIMKPGRCNGG

>BeM10
VRDGYIADDKDCAYFCGRNAYCDEECKKGAESGKCWYAGQYGNACWCYKLPDWVPIKQKVSGKCN

>BmKalpha1
MNYLVFFSLALLLMTGVGSVRDGYIADDKNCPYFCGRNAYCDDECKKNGAESGYCQWAGVYGNACWCYKLP
DKVPIRVPGKCNGG

>BmKalpha3
MNYLVFFSLALLLMTGVESVRDGYIADDKNCAYFCGRNAYCDDECKKKGAESGYCQWAGVYGNACWCYKLP
DKVPIRVPGKCNGG
>LqHIV
GVRDAYIADDKNCVYTCGANSYCNTECTKNGAESGYCQWFGKYGNACWCIKLPDKVPIRIPGKCR

>LqqIV
GVRDAYIADDKNCVYTCGSNSYCNTECTKNGAESGYCQWLGKYGNACWCIKLPDKVPIRIPGKCR

>AamH2
MNYLITISLALLLMTGVASGVRDGYIADAGNCGYTCVANDYCNTECTKNGAESGYCQWFGRYGNACWCIKL
PDKVPIKVPGKCNGR

>BeM14
ARDAYIADDRNCVYTCALNPYCDSECKKNGADGSYCQWLGRFGNACWCKNLPDDVPIRKIPGEECR

>Tb3
KKDGYPVEADNCAFVCFGYDNAYCDKLCGDKKADSGYCYWVHILCYCYGLPDNEPTKTNGKC

>CsEV
KKDGYPVDSGNCKYECLKDDYCNDLCLERKADKGYCYWGKVSCYCYGLPDNSPTKTSGKCNPA

>AaHIT4
EHGYLLNKYTGCKVWCVINNEECGYLCNKRRGGYYGYCYFWKLACYCQGARKSELWNYKTNKCDL

>CsEV2
MNSLLIITACLFLIGTVWAKEGYLVNKSTGCKYGCLKLGENEGCDKECKAKNQGGSYGYCYAFACWCEGLP
ESTPTYPLPNKSCSRK

>CsEV3
KEGYLVKKSDGCKYGCLKLGENEGCDTECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNKSC

>CsEV1
KEGYLVKKSDGCKYDCFWLGKNEHCNTECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNKSC

>CsEV4
KEGYMVNKSTGCSYSCPKTGESVYCDKECKAKNQGGSYGFCQYSNCWCEGLPESTPTWPLDDKPCD

>CsEV5
KDGYPVDSKGCKLSCVANNYCDNQCKMKKASGGHCYAMSCYCEGLPENAKVSDSATNIC

>Cn12
RDGYPLASNGCKFGCSGLGENNPTCNHVCEKKAGSDYGYCYAWTCYCEHVAEGTVLWGDSGTGPCRS

>BeI2
ADGYVKGKSGCKISCFLDNDLCNADCKYYGGKLNSWCIPDKSGYCWCPNKGWNSIKSETNTC

Figure 2

CLUSTAL W (1.74) multiple sequence alignment

```
Tb3             --------------------KKDGYPVE-ADNCAFVCFGYDNA--YCDKLCGDK-
-KADS
CsEV            --------------------KKDGYPVD-SGNCKYECLKDD----YCNDLCLER-
-KADK
Aah1'           -MNYLVMISLALLLMIGVESKRDGYIVY-PNNCVYHCIP------PCDGLCKKN-
-GGSS
Aah1            -MNYLVMISLALLLMIGVESKRDGYIVY-PNNCVYHCVP------PCDGLCKKN-
-GGSS
AaHIII          -MNYLVMISLALLLMTGVESVRDGYIVD-SKNCVYHCVP------PCDGLCKKN-
-GAKS
AaHIV           -MNYLIMFSLALLLVIGVESGRDGYIVD-SKNCVYHCYP------PCDGLCKKN-
-GAKS
LqHVIII         --------------------VRDGYIAK-PENCAHHCFPGSS---GCDTLCKEN-
-GGTG
LqHVII          --------------------VRDGYIAK-PENCAHHCFPGSS---GCDTLCKEN-
-GGTG
LqHIII          --------------------VRDGYIAQ-PENCVYHCFPGSS---GCDTLCKEK-
-GGTS
LqHVI           --------------------VRDGYIAQ-PENCVYHCIP------DCDTLCKDN-
-GGTG
OsIII           -------------------GVRDGYIAQ-PHNCVYHCFPGSG---GCDTLCKEN-
-GATQ
BomIII          --------------------GRDGYIAQ-PENCVYHCFPGSS---GCDTLCKEK-
-GATS
BomIV           --------------------GRDAYIAQ-PENCVYECAKNS----YCNDLCTKN-
-GAKS
BotI            --------------------GRDAYIAQ-PENCVYECAQNS----YCNDLCTKN-
-GATS
BotII           --------------------GRDAYIAQ-PENCVYECAKNS----YCNDLCTKN-
-GAKS
BomPI           --------------------GRDAYIAQ-PENCVYECAKSS----YCNDLCTKN-
-GAKS
BmKM4           MNYLVMISFALLLMK-GVESVRDAYIAK-PENCVYHCAGNE----GCNKLCTDN-
-GAES
BmKM4'          MNYLVMISFALLLMT-GVESVRDAYIAK-PENCVYHCATNE----GCNKLCTDN-
-GAES
BmK10           MNYLVMISFALLLMK-GVESVRDAYIAK-PENCVYECGITQ----DCNKLCTEN-
-GAES
BmKM1           MNYLVMISFALLLMT-GVESVRDAYIAK-PHNCVYECARNE----YCNDLCTKN-
-GAKS
BmKII           --------------------VRDAYIAK-PHNCVYECARNE----YCNDLCTKD-
-GAKS
BmKM2           --------------------VRDAYIAK-PHNCVYECARNE----YCNNLCTKN-
-GAKS
BmKM7           --------------------VRDGYIAL-PHNCAYGCLNNE----YCNNLCTKD-
-GAKI
BotIT1          --------------------VRDAYIAQ-NYNCVYFCMKDD----YCNDLCTKN-
-GASS
LqqIII          --------------------VRDAYIAK-NYNCVYECFRDS----YCNDLCTKN-
-GASS
```

```
BmK8            --------------------GRDAYIAD-SENCTYFCGSNP----YCNDVCTEN-
-GAKS
LqHIV           --------------------GVRDAYIAD-DKNCVYTCGANS----YCNTECTKN-
-GAES
LqqIV           --------------------GVRDAYIAD-DKNCVYTCGSNS----YCNTECTKN-
-GAES
AamH2           MNYLITISLALLLMTGVASGVRDGYIAD-AGNCGYTCVAND----YCNTECTKN-
-GAES
BeM14           --------------------ARDAYIAD-DRNCVYTCALNP----YCDSECKKN-
-GADG
AaHII           -MNYLVMISLALLFVTGVESVKDGYIVD-DVNCTYFCGRNA----YCNEECTKL-
-KGES
LqHII           --------------------IKDGYIVD-DVNCTYFCGRNA----YCNEECTKL-
-KGES
BotIII          --------------------VKDGYIVD-DRNCTYFCGRNA----YCNEECTKL-
-KGES
BotXI           --------------------LKDGYIVD-DRNCTYFCGTNA----YCNEECVKL-
-KGES
AmmV            --------------------LKDGYIID-DLNCTFFCGRNA----YCDDECKKK-
-GGES
LqqV            --------------------LKDGYIVD-DKNCTFFCGRNA----YCNDECKKK-
-GGES
BmKalpha2       -MNYMVIISLALLVMTGVESVKDGYIAD-DRNCPYFCGRNA----YCDGECKKN-
-RAES
BmKalpha1       -MNYLVFFSLALLLMTGVGSVRDGYIAD-DKNCPYFCGRNA----YCDDECKKN-
-GAES
BmKalpha3       -MNYLVFFSLALLLMTGVESVRDGYIAD-DKNCAYFCGRNA----YCDDECKKK-
-GAES
BeM10           --------------------VRDGYIAD-DKDCAYFCGRNA----YCDEECKK--
-GAES
CsEV2           --MNSLLIITACLFLIGTVWAKEGYLVNKSTGCKYGCLKLGENE-
GCDKECKAKNQGGSY
CsEV3           --------------------KEGYLVKKSDGCKYGCLKLGENE-
GCDTECKAKNQGGSY
CsEV1           --------------------KEGYLVKKSDGCKYDCFWLGKNE-
HCNTECKAKNQGGSY
CsEV4           --------------------KEGYMVNKSTGCSYSCPKTGESV-
YCDKECKAKNQGGSY
CsEV5           --------------------KDGYPVD-SKGCKLSCVANN----YCDNQCKMK-
-KASG
Cn12            --------------------RDGYPLA-SNGCKFGCSGLGENNPTCNHVCEKK-
AGSDY
AaHIT4          --------------------EHGYLLNKYTGCKVWCVINNE---ECGYLCNKR-
RGGYY
BeI2            --------------------ADGYVKG-KSGCKISCFLDN---
DLCNADCKYYG-GKLN
                                      ..*      .*   *         *.  *

Tb3             GYCYWV--H-ILCYCYGLPDNEPTK--TNGK-C---
CsEV            GYCYWG--K-VSCYCYGLPDNSPTK--TSGK-CNPA
Aah1'           GSCSFLVPSGLACWCKDLPDNVPIK--DTSRKCT--
Aah1            GSCSFLVPSGLACWCKDLPDNVPIK--DTSRKCT--
AaHIII          GSCGFLIPSGLACWCVALPDNVPIK--DPSYKCHSR
AaHIV           GSCGFLVPSGLACWCNDLPENVPIK--DPSDDCHKR
LqHVIII         GHCGFKVGHGTACWCNALPDKVGII--VDGVKCH--
LqHVII          GHCGFKVGHGTACWCNALPDKVGII--VDGVKCH--
LqHIII          GHCGFKVGHGLACWCNALPDNVGII--VEGEKCHS-
```

```
LqHVI           GHCGFKLGHGIACWCNALPDNVGII--VDGVKCHK-
OsIII           GSSCFILGRGTACWCKDLPDRVGVI--VDGEKCH--
BomIII          GHCGFLPGSGVACWCDNLPNKVPIV--VGGEKCH--
BomIV           GYCQWLGKYGNACWCEDLPDNVPIR--IPGK-CHF-
BotI            GYCQWLGKYGNACWCKDLPDNVPIR--IPGK-CHF-
BotII           GYCQWLGRWGNACYCIDLPDKVPIR--IEGK-CHF-
BomPI           GYCQWLGRWGNACYCIDLPDKVPIR--IEGK-CHFA
BmKM4           GYCQWGGRYGNACWCIKLPDDVPIR--VPGK-CHR-
BmKM4'          GYCQWGGKYGNACWCIKLPDDVPIR--VPGK-CHR-
BmK10           GYCQWGGKYGNACWCIKLPDSVPIR--VPGK-CQR-
BmKM1           GYCQWVGKYGNGCWCIELPDNVPIR--VPGK-CHR-
BmKII           GYCQWVGKYGNGCWCIELPDNVPIR--IPGN-CH--
BmKM2           GYCQWSGKYGNGCWCIELPDNVPIR--VPGK-CH--
BmKM7           GYCNIVGKYGNACWCIQLPDNVPIR--VPGR-CHPA
BotIT1          GYCQWAGKYGNACWCYALPDNVPIR--IPGK-CHS-
LqqIII          GYCQWAGKYGNACWCYALPDNVPIR--VPGK-CH--
BmK8            GYCQWAGRYGNACYCIDLPASERIK--EPGK-CG--
LqHIV           GYCQWFGKYGNACWCIKLPDKVPIR--IPGK-CR--
LqqIV           GYCQWLGKYGNACWCIKLPDKVPIR--IPGK-CR--
AamH2           GYCQWFGRYGNACWCIKLPDKVPIK--VPGK-CNGR
BeM14           SYCQWLGRFGNACWCKNLPDDVPIRK-IPGEECR--
AaHII           GYCQWASPYGNACYCYKLPDHVRTKG--PGRCHGR-
LqHII           GYCQWASPYGNACYCYKLPDHVRTKG--PGRCR---
BotIII          GYCQWASPYGNACYCYKVPDHVRTKG--PGRCN---
BotXI           GYCQWVGRYGNACWCYKLPDHVRTVQ--AGRCRS--
AmmV            GYCQWASPYGNACWCYKLPDRVSIKE--KGRCN---
LqqV            GYCQWASPYGNACWCYKLPDRVSIKE--KGRCN---
BmKalpha2       GYCQWASKYGNACWCYKLPDDARIMK--PGRCNGG-
BmKalpha1       GYCQWAGVYGNACWCYKLPDKVPIR--VPGKCNGG-
BmKalpha3       GYCQWAGVYGNACWCYKLPDKVPIR--VPGKCNGG-
BeM10           GKCWYAGQYGNACWCYKLPDWVPIKQKVSGKCN---
CsEV2           GYCYAF-----ACWCEGLPESTPTYP-LPNKSCSRK
CsEV3           GYCYAF-----ACWCEGLPESTPTYP-LPNKSC---
CsEV1           GYCYAF-----ACWCEGLPESTPTYP-LPNKSC---
CsEV4           GFCQYS-----NCWCEGLPESTPTWP-LDDKPCD--
CsEV5           GHCYAM-----SCYCEGLPENAKVSD-SATNIC---
Cn12            GYCYAW-----TCYCEHVAEGTVLWGDSGTGPCRS-
AaHIT4          GYCYFWK---LACYCQGARKSELWN--YKTNKCDL-
BeI2            SWCIPD--KSGYCWCPNKGWNSIKS---ETNTC---
                . .          *:*
```

Figure 3

Second Tight Turn

PDB	A.A.		ω	φ	Ψ
1AHO	LYS(28)	H	175.323	-59.202	-37.103
	LEU		-179.805	-86.545	7.962
	LYS		177.289	74.923	17.969
	GLY		170.049	-77.856	170.895
	GLU(32)		-179.863	-72.505	-49.017
1BMR	PHE(17)		166.980	-112.528	122.057
	PRO		-169.493	-56.343	96.466
	GLY		-179.342	162.277	-161.058
	SER		176.207	-84.856	-45.352
	SER(21)	H	-145.300	-39.295	-56.222
1CHZ	LYS(28)		177.766	-64.691	-29.494
	ANS		-177.947	-97.495	18.755
	GLY		177.092	89.096	2.419
	ALA		-179.850	-71.698	166.179
	LYA(32)		175.279	-77.384	-30.931
1JZA	LYS(30)	H	-179.364	-83.600	-14.769
	ALA		179.585	-49.661	139.431
	LYS		-178.236	-57.077	-28.223
	ASN		178.077	-75.253	-18.194
	GLN		178.182	-97.582	-46.614
	GLY		178.974	73.550	25.524
	GLY(36)		-179.570	-83.395	169.611
1LQQ	ALA(39)	S	179.724	-140.580	-75.626
	GLY		179.845	157.663	-55.094
	LYS		179.804	-34.718	-47.977
	TYR		179.537	-125.879	14.509
	GLY	S	-179.773	84.127	-132.976
	ASN(44)	S	-179.922	-102.138	157.645

1NH5	CYS(35)	S	-170.222	-102.701	106.794
	TYR	S	-179.046	-108.864	123.408
	ALA		177.062	42.695	51.935
	MET		175.929	54.638	43.319
	SER	S	168.592	-151.407	136.070
	CYS(40)	S	-169.386	-73.501	95.785
1NRA	GLU(28)	H	179.493	-66.755	-33.596
	ARG	H	-179.935	-75.295	2.149
	LYS		179.990	83.482	10.790
	ALA		-179.637	-90.355	146.620
	ASP(32)		179.469	-68.182	-52.198
1PE4	CYS(15)		-179.615	59.097	109.188
	SER		-179.619	-162.043	111.635
	GLY		179.588	-52.885	-65.058
	LEU		179.298	-78.750	58.114
	GLY		-179.662	-155.709	179.748
	GLU(20)		-179.961	-83.357	-161.010
1SN1	VAL(39)	S	177.373	-138.161	122.359
	GLY		179.754	-51.918	-56.595
	LYS		-179.699	-54.691	-20.017
	TYR		179.181	-130.105	-9.513
	GLY(43)	S	-179.058	82.038	-144.722
1SN4	ASP(28)		177.239	-65.900	-18.084
	ASN		-179.838	-112.218	24.500
	GLY		170.320	94.442	-7.523
	ALA		-179.572	-74.049	166.318
	GLU(32)		178.537	-81.719	-32.087
1SNB	TRP(38)		178.522	-62.179	155.302
	ALA		176.780	41.721	46.006
	GLY		178.830	-106.792	-162.954
	ARG		-175.436	-48.333	-40.295
	TYR		-179.747	-118.32-	30.176
	GLY		-177.460	79.994	-154.709
	ASN(44)	S	176.328	-67.059	116.894

1VNB	ALA(31)	-179.983	-50.602	153.988
	LYS	-179.972	-75.098	-0.491
	ASN	-179.903	-87.501	-8.378
	GLN	-179.933	-122.003	-59.800
	GLY	-179.974	104.988	-3.729
	GLY(36)	-179.976	-67.507	147.924
2SN3	ALA(31)	178.696	-56.591	146.426
	LYS	-177.962	-68.077	-18.820
	ASN	168.133	-63.154	-4.035
	GLN	178.643	-129.967	-52.355
	GLY	172.531	84.935	16.363
	GLY	178.112	-80.463	176.897
	SER(37)	-172.813	-97.011	-39.678

Third Tight Turn

PDB	A.A.		ω	φ	Ψ
1AHO	TRP(38)		175.050	-73.859	143.258
	ALA		174.423	57.028	41.178
	SER		167.668	-71.900	172.959
	PRO		176.082	-61.680	-12.665
	TYR		-178.489	-109.662	14.578
	GLY		-174.073	84.997	-165.415
	ASN(44)		177.405	-65.245	131.030
1BMR	LYS(40)	S	178.293	-118.926	99.253
	VAL		-175.710	-69.345	129.910
	GLY		-168.219	124.747	-41.550
	HIS		176.668	-108.937	-69.852
	GLY(44)	S	-170.525	158.387	-155.491
1CHZ	TRP(38)		179.575	-102.358	-20.559
	SER		-178.146	-139.877	46.451
	GLY		178.943	-94.679	-166.229
	LYS		-177.662	-52.139	-30.146
	TYR		179.995	-116.804	7.329
	GLY		-177.683	86.517	-130.142
	ASN(44)		-177.951	-75.878	133.234

1JZA	CYS(41)	S	-179.481	-82.364	126.817
	TYR	S	-179.639	-138.443	115.233
	ALA		-179.424	65.284	38.710
	PHE		179.884	63.212	35.081
	ALA(45)	S	178.959	-154.218	137.519
1NRA	TRP(38)	S	-179.203	-64.836	-100.867
	GLY		-179.472	40.344	29.098
	LYS		-178.997	67.110	54.600
	VAL	S	-179.203	-138.924	-104.478
	SER(42)	S	-179.180	-175.938	-167.254
1PE4	LEU(31)		179.298	-78.750	58.114
	LYS		-179.753	-103.576	-73.127
	LYS	H	-179.766	-83.645	-25.925
	ALA		178.918	-97.457	-14.061
	GLY		179.216	42.993	29.963
	SER(35)		-179.957	-135.098	153.501
1SN4	TRP(38)		178.601	-79.799	142.604
	GLY		176.739	56.463	34.746
	GLY		179.685	-85.681	-171.724
	ARG		178.854	-66.788	-17.925
	TYR		178.362	-119.788	30.273
	GLY		-179.332	85.589	-152.477
	ASN(44)	S	-179.329	-93.463	113.957
1VNB	CYS(41)	S	179.763	-107.422	131.157
	TYR	S	-179.746	-107.662	83.063
	ALA		179.954	84.952	-57.834
	PHE		179.907	-165.491	18.708
	ALA(45)	S	179.968	-126.970	-179.681
2SN3	CYS(41)	S	-176.060	-98.279	122.855
	TYR	S	175.469	-131.582	120.498
	ALA		-178.372	54.550	39.740
	PHE	S	173.971	72.362	22.701
	ALA	S	169.522	-143.119	156.564
	CYS	S	-175.765	-79.721	122.324
	TRP	S	177.297	-104.829	144.643

Fourth Tight Turn

PDB	A.A.	ω	φ	Ψ
1JZA	PRO	-0.100	-99.313	162.068
	LEU	178.157	-79.572	132.667
	PRO	-179.335	-56.708	-19.254
	ASN	-177.863	-136.830	15.431
	LYS	-179.655	-140.825	93.543
	SER	-177.756	-78.754	150.686
1PE4	CYS(40)	179.379	-125.049	117.045
	TYR	-179.734	-124.148	-57.920
	ALA	179.594	-96.136	-148.123
	TRP	-179.959	-85.611	63.941
	THR	-179.991	-178.614	174.667
	CYS(45)	-179.661	-133.855	141.677

Fifth Tight Turn

PDB	A.A.	ω	φ	Ψ
1PE4	ALA(51)	179.942	-85.186	9.663
	GLU	179.820	-79.472	59.217
	GLY	179.973	93.975	-25.435
	THR	-179.924	-60.752	-175.558
	VAL(55)	-179.913	-98.728	164.260