Computational analysis of evolutionary divergence of scorpion toxins

R K Upadhyay
Laboratory of Human Genetics, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Received 24 September 2002; revised 27 December 2002

Conserved and consensus sequences of several members of scorpion toxin families have been analyzed. Multiple sequence alignment define the highly conserved residues that play important structural role in formation of toxin subgroups. Scorpion toxins have characteristic feature in signature pattern \{(GA)\}-K-C \{(LIVM)\}-X \{(2)\}-K-C-C-X-C\}, where lysine and cysteine residues are well conserved in the polypeptide. The toxins show two major groups, one with conserved active region GKC\(\text{MNKGKC}\) and second with CT\(\text{PK}\). There is variability in position of lysine and arginine in different toxins. Most of the scorpion toxins contain six conserved cysteines involved in disulphide bonds. From the evolutionary analysis it is concluded that \(\textit{Leiurus}\) is a primitive ancestral species from which probably various toxins have evolved during the evolution.

About 600 species of scorpion are known belonging to 70 genera and 6 families. Highly venomous scorpion species are from Buthinae family containing short proteins consisting of 60-80 amino acids. Over 80 secondary structures which constitute a group of small basic toxins e.g. neurotoxic, cytotoxic, kaliotoxic and charybotoxic polypeptides\(^1\)\(^3\) and are cross linked by four disulphide bridges in individual primary sequences have been reported from different species\(^4\)\(^5\). These toxins have been found in all the scorpion species belonging to the genera \(\textit{Leiurus}\), \(\textit{Androctonus}\), \(\textit{Buthus}\), \(\textit{Centruroides}\), \(\textit{Tityus}\) and \(\textit{Mesobuthus}\).

The temperate species of scorpion are highly venomous and the toxins are lethal, having very low LD\(_{50}\) (e.g. \(\textit{Tityus serrulatus}\) toxin has LD\(_{50}\) 0.016 mg/kg). These toxins are further divided into subgroups according to their nature of action and binding inside the tissues. Certain naturally occurring toxins from scorpion venom have high binding affinities for the sodium channel\(^6\) and are also highly selective for channel sub-types\(^7\)\(^8\). The potent toxicity of these toxins against mammals is related to high-affinity binding to their voltage sensitive sodium channels\(^9\). This specific binding results in the impairing of the initial, rapid depolarization phase of the action potential in nerve and muscle causing paralytic effect. Some scorpion toxins selectively block the high conductive calcium activated potassium channels\(^10\).

Another important feature of scorpion toxin family which might be involved in structure-function relationship is the presence of four disulphide bridges which are placed between 12-65, 16-41, 25-46 and 29-48 position in the polypeptide chain. Unlike most of other proteins which have disulphide linkages in their structure, scorpion toxins also have topological isomeric forms of same topological chirality\(^11\) and inhibit the post synaptic currents\(^12\). These toxins have significant homologous amino acid sequences within the families and display various degrees of toxicity towards different animal classes which also have diverse pharmacological properties. These toxins also have beta pleated sheets in their structure connected by long exposed loops and folds. Since the toxicity of venoms from different scorpion species shows great variation, it is interesting to analyze the diverse structural and functional homologues of toxins in different species of scorpion. The question arises as to how these toxins are evolved during evolution to show high degree of toxicity and variability?

**Methodology**

The scorpion toxins were searched under the "stringsearch" programme from the GCG Swissprot database. Extensive search was done for collecting information on scorpion toxins. Searches were also made for current protein identification resources (PIR) from protein Data Bank to compare per cent identity of one protein with another. Multiple sequence alignment was done by using "pileup" program. Pair-wise similarity between two toxin polypeptides was obtained from the dot plots. For consensus and conserved residues “pretty” was also

*Present Address: Department of Zoology, Deen Dayal Upadhyay Gorakhpur University, Gorakhpur, U.P. 273009 India
E-mail rkupad@yahoo.com
used. For hydrophobicity analysis plot structures were prepared based on the Kyte-Doolittle hydrophathy scale. The presence and number of shared “motifs” in sets of aligned and unaligned protein sequences were identified by using motif programme. Blast searches were also made for sequences similar to query sequences. To reconfirm the homology, “fasta” was used to search the similarity between a query sequence and a group of sequences. For getting consensus sequences, “pretty” and “profile” matching programmes were used.

The phylogenetic tree was constructed using pre-aligned sequences from the distance programme and degree of divergence was calculated. The toxins were grouped according to the nature of toxins and consensus pattern followed a particular rule in particular set. The number of representatives of each class are SCX (54), SIX (5), SCK (15) SCA (3).

Results

Comparison and multiple alignment

In the present study six families of scorpion toxins are analyzed. Alignment of sequences of scorpion toxins with the known secondary structures gave a clear picture of active site region and amino acids involved in toxicity. The N to C terminal multiple alignment of toxins is shown in Fig. 1. The alignment begins with residue 1 at N terminal side and ends on the last position of amino acids at C terminal end. Position number indicated in the figure represents the residue numbers in the proteins, but due to homology present among different toxins their position has been changed. The residue RDGYIVD is all conserved. In 80% of the species a hydrophobic/hydrophilic residue in alpha helical region is indicated by V. Residue 36, 49, 69 is a cysteine present in all toxins except agitoxin. The residue Y present at position 24 in the alpha helix A1 is conserved in all the neurotoxins. In other toxins, positions 21-30 is highly conserved sequence KDGYIDKST where Y and C are fully conserved; RDANVY conserved sequences are replaced by KEKNKST. In SEX3_Andau R, at position No. 21 is replaced either by K or E. Cysteine at 36 and 69 are fully conserved in the beta strand region as well as in loop region of 21-100. This constitutes a pocket of amino acids which show 88% homology and represents an evolutionary box for the scorpion toxins. Amino acid residues in Leirus and Titus have a conserved box KDGYYP with slight variations in amino acids in which lysine at position no. 21 is replaced by arginine. While the same box in other species Androctonus mauretanicus, Buthus occitanus, Orthochirus, Androctonus mauretanicus, Leiurus quinguestiatus and Buthus euperus has been changed by RDGYI. There was observed and conserved active site region between 26-33 amino acid positions which is represented by GKCMNK amino acids (Fig. 1). There is also another characteristic box CTPK which is also conserved in some of the toxins. Short stretches of YHC amino acids are also found conserved. In some of them H is replaced by E or T and G. Further towards C-terminal site, YC conserved stretches are also found which show an alternative arrangement ofYW, AW and YC. At residue no. 69 and 70 toxins have GL, AC, SC, GC amino acids in sequential order in different toxin peptides. At some places instead of above amino acids WC, WC, YC short stretches are found. Besides above features a short domain of KCH, KCY, GRCN, GRCR is also present in the end of C-terminal site which has similarity with toxins of other animal groups.

Sequence motifs in toxins are identified by using motifs programme and a sequence pattern of \{(GA)−K-C\}−\{(LIVM)−X\}-(2)−K-C-C-X-C) is derived for characterizing the scorpion toxin family. But these signature sequences do not provide overall homology for most of the toxin clusters as different toxins have different active amino acids in their alpha and beta chain regions. In certain species per cent identity seen was nearly 50% but in some cases very little homology was present in toxins. Agitotoxins have amino acid residues different from kaliotoxins. Scyllatoxins are also far away from the signature sequences matching (Fig. 1). Dotplots showing percent identity for different scorpion toxins are shown in Fig. 2. The most identical matches showing highest homology are SCX_Buteu and SCX_3_Butoc; SCX_Cenno and SCX_Titse respectively. Most of the toxins show homology in middle portion of the polypeptide chain (Fig. 2).

When the signature sequences of scorpion toxin family were used as template for the “fasta” and “best fit” programmes maximum 883 toxins and non-toxins showed homology. When signature sequences were matched with the Leiurus (SCX_LEIQH 4 species), a maximum homology of 100% was obtained while minimum homology was observed in SCX2_TITSE (34%) (Fig. 3a). Percent identity of signature sequences of scorpion toxins show a 100 % homology with leiurotoxin II while with other toxins it shows a 70-80% similarity. Among the non-toxins, disintegrins, defensin and acetylcholinesterases show homology between 40-50% (Fig. 3b) which could be
Fig. 1 — Alignment of consensus and conserved sequences of scorpion toxins. Alignment is made with appropriate gaps and

<table>
<thead>
<tr>
<th>Consensus Sequence</th>
<th>Conserved Regions</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For selected toxins, which show close homology
Fig. 2 — Dotplots showing the percent identity for different scorpion toxins [Ordinates and abscissae are expressed as amino acid residue number of the sequences employed. The C-terminal residue is indicated in the upper right corner of the dotplot]
either in functional region or in the structural region. Interestingly, disintegrins, defensins, granulins and immunophilins also show homology in disulphide bridge region. When consensus sequences of scorpion toxins are compared with other toxins a homology of 50-70% is noted (Fig. 3c) and with non-toxins 25-50% (Fig. 3d). This shows that the amino acids which are partially shared by toxins are also present in non-toxins. Due to environmental impact non toxic amino acids are replaced by basic or toxic amino acids in polypeptides.

Position of active site region in scorpion toxins

Scorpion toxins have specific binding sites for the sodium and potassium channels either in the helix region or in the beta chain region. It has been shown from the motifs and signature sequence pattern that active site region in toxins is always found between 26-34 amino acid residues (Table 1). In neurotoxins amino acids at position 21 to 36 are conserved in very few species and there is a replacement of arginine at position 21 by lysine while aspartic acid is replaced by glutamic acid. There are five triplets which are GYI, GYL, GYA, GYP and GYC representing the different toxins, most of these being neurotoxins. In kaliotoxins residues N, KC, CTPK are well conserved at C terminal end but in insectoxins GCK at 31-33, C at 55 residue and C at 67 are well conserved. A very significant difference between the neurotoxin to kaliotoxins is that the former has the long stretch between the 21-100 amino acids in their secondary structure. The lowest percent identity is observed in the distantly related groups SCX Leiqh chlorotoxin. There is common structural variability in amino acids based on the nature of toxins in different species. The best example of such more visible inter group homology is present in the genus Leiurus.

Presence of beta chain sub-region

Amidation and phosphorylation sites are found in Androctonus AaH II and AAI neurotoxin between 20-83 amino acid residues and alpha toxin of Mesobuthus between 17-79 residues. In SCX2-Butzu, Buthus judiacus (22-82), in Androctonus australis hector neurotoxin (22-82), SCXZ Leiqh, noxiustoxin (20-85), neurotoxin AAH1T2 (SIX_Andau) 19-88,
SCXx_Ttse (21-81), SCK2_Andma kaliotoxin 23-59, SCX4_Cenno (Toxin 4 precursor) 20-85 shows similarity in beta chain region.

**Sodium channel binding domain of scorpion toxin**

In scorpion toxins sodium channel binding domains in the active site region occur at different places. The sequences GKCMNRCHC which are present in the active region have high binding affinity with receptors of sodium channels. In scorpion toxins lysine (K) is a very active amino acid residue which plays an important role in toxicity determination and most frequently occur both in alpha and beta helical regions. The important role of cysteine is to establish the receptor binding and to block the sodium channels.

**Discussion**

Scorpion toxins have multiple functions according to the basic nature and activity of amino acids. But helix and turn region residues are most important which can determine the activity of a toxin polypeptide. Variability in active site region can determine the ionic regulation and hydrophobic activity of the toxins. It seems likely that the majority of the effects of the scorpion toxins on the body are due to their direct actions on sodium and potassium channels. The most toxic venoms of *Leiurus* and *Buthus* scorpion have a slow sodium channel inactivation. Different toxins have different receptor sites associated with sodium channel. Homologous proteins may have same binding site or different binding sites. In this category, alpha toxins from African and Asian scorpion species e.g. *Androctonus australis*, *Leiurus quinquestriatus* and *Buthus occitanus* act upon the receptor binding site III whereas beta toxin from *Centruroides sculpturatus* binds to a different site in the sodium channel site (IV). There is also a second possibility that many ionic channels may share common toxin receptor sites which have high specific binding for some of the amino acids in toxins. If a single amino acid is modified or replaced in the active site region, the affinity of toxin may be decreased or in creased to a specific channel.

In some scorpion toxins, sequence similarity extends to the small domain of defensins of arthropods in both C and N terminal region. Signature sequences from the scorpion family do not predict the structural role of active site region in many toxins because most of the proteins have simple primary structures and a single amino acid difference. However, the rate of replacement of amino acids depends on climatic effects which alter the natural survival and mechanism of action in scorpions. It is also evident that most of the toxins were not capable of binding to the sodium channel, but they could enhance the specific binding of other types of toxins. An evolutionary peculiarity in scorpion toxins is anti-mammalian toxic activity and induction of a paralytic effect on non-mammals. Lqq IT2 Lei and AaH IT4 have no effect on mammals but have high paralytic effect on insects larvae or on the prey.

From an evolutionary point of view, the presence of anti mammal toxic effect next to anti-insect toxicity is a progressive adaptation of the venomous secretion of putative preys of scorpions. On the other hand selective anti-insect toxins are not found in the new world scorpion venoms. Only a few toxins have

<table>
<thead>
<tr>
<th>Active site</th>
<th>Residue</th>
<th>Species</th>
<th>Toxin type</th>
</tr>
</thead>
<tbody>
<tr>
<td>26-33</td>
<td>GKCMGKKC</td>
<td>SCKi_Mesta</td>
<td>Iberiotoxin</td>
</tr>
<tr>
<td>26-34</td>
<td>GAKCMNGK</td>
<td>SCK1_Cenli</td>
<td>Toxin IT</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCMKKKC</td>
<td>SCKC_Leiqh</td>
<td>Charybdotoxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCMNGKC</td>
<td>SCKA_TITSE</td>
<td>Tityustoxin K</td>
</tr>
<tr>
<td>26-33</td>
<td>GAKCMKKC</td>
<td>SCK2_Leiqh</td>
<td>Charybdotoxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCNRKNC</td>
<td>SCK1-Andma</td>
<td>Kaliotoxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCINGKC</td>
<td>SCK1_Cenll</td>
<td>Toxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCNGKC</td>
<td>SCK2_Andma</td>
<td>Toxin</td>
</tr>
<tr>
<td>47-54</td>
<td>GKCNGKC</td>
<td>SCK2_Leiqh</td>
<td>Toxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCQNKQC</td>
<td>SCK3_Leiqh</td>
<td>Agitoxin</td>
</tr>
<tr>
<td>26-33</td>
<td>GKCINGKC</td>
<td>SCA1_Leiqh</td>
<td>Agitoxin</td>
</tr>
</tbody>
</table>
common nature of old and new world toxins which are in the transition phase of the evolution hence show bi-specificity i.e. Ts VII toxin from Brazilian scorpion, *Tityus serrulatus* which shows an activity against enzymes and sodium channel receptors of mammals and insects as well. It is also well known fact that scorpion toxins are ion channel blockers but are diversified according to survival of scorpion species. In the long array of time how old world toxins of alpha type are converted into beta type new world toxins is a great question to answer.

There are a few scorpion venom toxins which impart high toxic effect on both mammals and insects but their toxic effects are very weak in comparison to anti-insect toxins from old world scorpion toxins. We can infer that anti-mamalian toxins (cytotoxins) were present in new world scorpions while anti-insect toxins (paralytic) are less evolved and are absent in old world scorpion species. From the present study two clusters show more conservancy and diversity of alpha and beta types. It was reported that toxins from the new world were mainly beta type toxins whereas toxins from the old world were exclusively alpha type (*Tityus serrulatus* and *Leiurus quinquestriatus*). *Androctonus australis hector* has characters of both type of toxins. It is also known that tityustoxins from old world scorpions selectively act on insects. Iberiotoxin, an alpha type protein from new world scorpion *Mesobuthus tamulus* act upon mammals. Even small scorpion toxins which have both beta and alpha type structures fall into two groups. One with basic amino acids having non-charged polar groups which may be responsible for non-target channel binding or transformed into a less toxic polypeptide. Second type of toxins are high grade killer toxins containing highly charged basic amino acids in their alpha helical region. The most conserved residues which have shown high activity for channel binding are G,K,M,C,R,Q in different species. During evolution these are further reorganized in different species.

This divergence in the evolution generated different interrelated clusters of toxins within the families/genus. Most diverged genus is *Leiurus* which has different toxins with variable active sites. Leiurotoxin is a possible ancestral protein, which has acquired most diverged mechanism of action. It is plausible that the extended family of scorpion toxins began to diverge from a common ancestor that had similar structure and biochemical function at very early evolutionary stage. Subsequently, the signature sequences required for the definition of basic toxic activity in most of the scorpion toxins have a variable region according to the climatic conditions. In a well defined ancestry of the scorpion toxins *Leiurus* toxin seems to be more diversified proteins.

On the basis of evolutionary divergence, it is hard to evaluate the level of toxicity in different scorpions but evolution of diverse defense mechanisms indicates action mechanism on different tissues and cells of different animals. For example LD$_{50}$ of tityustoxin, a beta type toxin, is less toxic than the alpha type. At present toxin bearing scorpion species are of both types. One line is of alpha type while another one is beta type. It is also clear that speciation also led to the toxin types and subtypes. Changes in environment and climatic changes affected the arrangement of amino acids in toxin polypeptides. The European species of colder region have some characters deviated from the temperate ones. However, in speciation from old world to new world, nature of toxicity is not altered (neurotoxins remain neurotoxins), but the level of toxicity may be reduced by shifting of amino acids. This shows that amino acids which are partially shared by toxins are also present in non toxins but due to environmental impact, site specific changes have taken place.

**Acknowledgement**

The author expresses his gratitude to Prof R N K Bamezai, Laboratory of Human Genetics, School of Life Science, New Delhi 110067. He is also thankful to University Grants Commission, New Delhi for financial assistance.

**References**

9 Jover E, Couraud F & Rochat H (1980) Biochem Biophys Commun 45, 1607-1614
12 Southan A P & Robertson B J (1998) Pharmacol 125 (6), 1375-81
13 Koppenhofer E & Schmidt H (1968) Plugers Arch 303 (2), 133-149