Deduction of functional peptide motifs in scorpion toxins

PAUL T. J. TAN,[a,b] **SHOBA RANGANATHAN**[c,b] **and VLADIMIR BRUSIC**[d,a]*

[a] Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
[b] Department of Biochemistry, Faculty of Medicine, National University of Singapore, Medical Drive, Singapore 117597
[c] Department of Chemistry and Biomolecular Sciences and Biotechnology Research Institute, Macquarie University, Sydney NSW 2109, Australia
[d] Australian Centre for Plant Functional Genomics, School of Land and Food Sciences and Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia

Abstract: Scorpion toxins are important physiological probes for characterizing ion channels. Molecular databases have limited functional annotation of scorpion toxins. Their function can be inferred by searching for conserved motifs in sequence signature databases that are derived statistically but are not necessarily biologically relevant. Mutation studies provide biological information on residues and positions important for structure–function relationship but are not normally used for extraction of binding motifs. 3D structure analyses also aid in the extraction of peptide motifs in which non-contiguous residues are clustered spatially. Here we present new, functionally relevant peptide motifs for ion channels, derived from the analyses of scorpion toxin native and mutant peptides. Copyright © 2006 European Peptide Society and John Wiley & Sons, Ltd.

INTRODUCTION

Scorpion toxin peptides are used as highly potent and specific probes for identification and characterization of ion channels [1–3]. They are also utilized in protein drug target engineering [4], vaccines against scorpion envenomation [5,6] and as biopesticides [7–9]. Identification of new toxin sequences and determination of their functional sites and structural properties are of great interest in scientific research as well as in medical and commercial applications.

Newly identified scorpion toxins are deposited in public databases that provide limited functional annotation. To aid experimental characterization of these toxins, function is inferred from identification of similar characterized sequences by pairwise alignment such as BLAST [10] or FASTA [11] programs. However, pairwise alignment does not distinguish critical functional residues (e.g. an active site) from residues with no critical role [12]. Another approach involves searching sequence signature databases such as PROSITE [12] and PRINTS [13] for conserved motifs which usually have structural or functional meaning. These motifs are derived statistically from multiple sequence alignments of peptide sequences that form a family. For example, the scorpion short toxins signature, C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-C, has been deposited in PROSITE database (accession ID: PS01138). However, the motifs that were derived statistically are not necessarily biologically relevant. Mutation studies of scorpion toxins (such as site-directed mutagenesis and chemical modification) have identified critical residues important for both structural and functional properties [e.g. [14,15]]. The mutation data that provide biologically relevant information on critical residue and position are available in the literature and normally are not used for extraction of functional motifs in scorpion toxins. Additionally, analyses of three-dimensional (3D) structures of toxins can aid identification of motifs in which non-contiguous residues in the primary sequence cluster spatially. Here we describe a set of scorpion toxin motifs extracted from analysis of multiple sequence alignment of scorpion native toxin sequences, 3D structures and information from mutation studies.

This is the first report of peptide binding motifs functionally relevant to sodium and potassium ion channels, which can facilitate the determination of specificity of newly identified scorpion toxin peptides to various ion channel subtypes. Also, they can help in the detection of distant relationships between toxin sequences that may be overlooked in pairwise alignment analysis.

MATERIALS AND METHODS

Scorpion Toxin Sequence, Three-Dimensional Data Collection and Enrichment

Scorpion native toxin sequence data were extracted from public databases, GenBank [16] and Swiss-Prot [17] and the literature. Scorpion mutant toxin sequence data were extracted solely from the literature through keyword searches such as 'scorpion toxin' AND 'mutation', 'mutant', 'mutagenesis', 'modification' and 'analog' in the PubMed (www.ncbi.nlm.nih.gov/entrez/query.fcgi). The collected data

* Correspondence to: V. Brusic, Institute for Molecular Bioscience, University of Queensland Brisbane 4072, Australia;
e-mail: v.brusic@uq.edu.au

were enriched with information on ion channel specificity, binding affinity and toxicity extracted from the literature. 3D structures of scorpion toxins were extracted from PDB [18]. Homology models of scorpion native and mutant toxins were built using SDPMOD [19]. All sequence and 3D data were deposited into SCORPION2 database [20].

Classification of Native Scorpion Toxin Sequences

Native scorpion toxin sequences were classified into four broad groups, namely, sodium- (Na^+), potassium- (K^+), calcium- (Ca^{2+}), and chloride- (Cl^-) specific toxins. Within each broad group they were further classified into groups based on primary structure similarity using BLAST [10], multiple sequence alignment and phylogenetic analysis [21]. Within each broad group, a representative toxin sequence was submitted to the blastp program against the non-redundant (nr) database at NCBI (www.ncbi.nlm.nih.gov/BLAST). Each BLAST result obtained after every submission ranked protein sequences from the most similar to the least. By looking at the E value, a cut-off was determined by manual inspection for each group of sequences. The cut-off value would vary depending on the similarity of the top scoring sequences, but usually occurred where there was a large change in E values between two consecutive sequences. Sequences that had E values lower than this cut-off were clustered into a group. The sequences scoring higher than the cut-off were added to a new group.

Once similar sequences were assigned to a putative group, another toxin from this group was used for a second BLAST search to confirm the results. The grouping was finalized if the new cut-off grouped the toxins in the same way as the first BLAST result. We confirmed this grouping by searching the sequence closest to the cut-off value against the nr database. The grouped sequences were removed from the list and the process was repeated with subsequent sequences until the list was exhausted.

Within each group, multiple sequence alignment was performed using CLUSTAL W [22]. Sequences with distinct primary structure patterns were further classified into subgroups. We verified these groupings by phylogenetic analysis using MEGA 3.0 [23]. Phylogenetic trees were reconstructed using the neighbour-joining algorithm [24]. A 1000-fold bootstrap was used to test the phylogenetic tree significance.

Scaling of Binding Affinities to a Common Scale in Mutant Toxin Data

The effects of mutation on binding affinity were scaled to a common scale for comparison [Eqn (1)].

$$\text{Scaled } X = 1 + \frac{\log_{10}\frac{1}{x} - \log_{10}\frac{1}{\text{lowest binding affinity}}}{\log_{10}\frac{1}{\text{highest binding affinity}} - \log_{10}\frac{1}{\text{lowest binding affinity}}} \times 7 \quad (1)$$

where $\log_{10}$ represents the common base 10 logarithm, x is the value of the binding affinity for the experiment to be scaled. The highest binding affinity values were observed for each ion channel from independent binding experiments. The lowest binding affinity for each ion channel was set at 10 000 nM.

Only K^+ and Na^+ binding affinity data were mapped because no binding affinity data was available for Cl^- toxins and only a set of five binding affinity measurements was available for Ca^{2+} toxins. Within K^+ and Na^+ groups, eight binding categories were defined for each group, namely, (i) non-binding, (ii) very low binding, (iii) low binding, (iv) moderately low binding, (v) moderate binding, (vi) moderately high binding, (vii) high binding, and (viii) very high binding. The highest binding affinity data observed for K^+ and Na^+ were 0.08 pM [25] and 4 pM [26], respectively. The lowest limit for both K^+ and Na^+ was set at 10 000 nM because any concentration higher than this would mean the toxin is a non-binder to the ion channel.

Data Analysis

Mutant toxin peptides were first compared with their native toxin sequences. Mutations that affected structural folding, as detected by circular dichroism spectroscopy, were noted. Mutations that affected binding affinity and toxicity by more than 10- and 100-fold as compared to native toxins were termed as 'influential' and 'critical', respectively. These influential and critical residues were then compared to multiple sequence alignment of classified groups of native toxins to verify possible conservation of residues. The spatial organization of the residues important for structure and function was visualized using Molsoft (www.molsoft.com/index.html) and mapped onto the native toxin primary sequences for extraction of motifs. The motifs are represented as used in PROSITE database [12]. To test the relevance of these motifs, each was searched in Swiss-Prot (release 48.0) and TrEMBL (release 31.0) databases [17] using the ScanProsite program [27].

RESULTS AND DISCUSSION

A total of 402 native and 426 mutant toxin peptides were collected and deposited in the SCORPION2 database [20]. Of 630 three-dimensional structures, 82 were extracted from PDB and 548 were homology models developed using SDPMOD. We classified the 402 scorpion native toxin peptides by sequence similarity using BLAST [10], multiple sequence alignment and phylogenetic analysis into 59 groups where 30 K^+, 18 Na^+, four Ca^{2+} and one Cl^- groups were defined (Table 1). Scorpione sequence was assigned to the 'defensin' group, while five scorpion toxin sequences with no annotated molecular target were assigned to the five 'orphan' groups (http://sdmc.i2r.a-star.edu.sg/scorpion/groups0/index1.html).

By scaling binding affinities of scorpion mutant toxins to a common scale, effects of mutation on function can be compared across each group. An example is shown in Figure 1, where Agitoxin 2 targets shaker K^+ channel at moderate range after mapping to the common scale ($K_d = 0.741$ nM) [28]. Mutation study on Agitoxin 2 suggested that K27 and N30 are critical for binding affinity towards this channel, while S11, R24, F25, M29 and T36 are influential in toxin–channel interaction, and T9, G10, R31, K32, H34 and P37

Binding Affinity   Agitoxin 2: Shaker potassium channel expressed in xenopus oocyte

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Moderate	G	V	P	I	N	V	S	C	T	G	S	P	Q	C	I	K	P	C	K	D	A	G	M	R	F	G	K	C	M	N	R	K	C	H	C	T	P	K
								A	V																						A	A		A			A	
Moderate Low											A													A	A				A							A		
Low																											M			A								

Figure 1 Scaling of agitoxin 2 and its mutant sequences for analysis of residues important for interaction with shaker K$^+$ channel. Native agitoxin 2 binds moderately to shaker K$^+$ channel. Mutations at positions 9, 10, 31, 32, 34 and 37 did not affect binding affinity significantly (<7-fold). However, mutations at positions 11, 24, 25, 29 and 36 resulted in moderately low affinities. Further, mutations of lysine and asparagine to methionine and alanine at positions 27 and 30 respectively resulted in low affinities towards the channel. This suggests that lysine and asparagine at these positions are important for binding towards shaker K$^+$ channel.

Table 1 Number of classified toxin groups for 402 scorpion native toxins

Number of toxin groups	Scorpion Version 2
Potassium channel toxin	30
Sodium channel toxin	18
Calcium channel toxin	4
Chloride channel toxin	1
Defensin	1
Orphan toxin	5
Total number of groups	59

can be substituted without significant effect on binding affinity [28].

A total of eight peptide motifs from analyses of native toxin sequences, 3D structures and mutation studies of Na$^+$ (four motifs) and K$^+$ (four motifs) toxins were extracted (Table 2). These motifs are highly specific to scorpion toxins in which all the sequences returned by ScanProsite were of scorpions i.e. no false positives were extracted. Further, these motifs can be used to search uncharacterized scorpion toxins and infer their specificities to different ion channel subtypes. Supplementary materials that provide more details on scaled binding affinities of scorpion mutant toxins, multiple sequence alignments of classified groups of native toxins and descriptions of spatial arrangement of the motifs are available at sdmc.i2r.a-star.edu.sg/scorpion/Motif/motif.html.

Chloride Channel Motif

In the PROSITE document of scorpion short toxins signature (PDOC00875), Cl$^-$ toxins were annotated with the pattern, C-x(3)-C-x(6,9)-[GAS]-K-C-[IMQT]-x(3)-C-x-C. However, the motif that we extracted on the basis of conservation of residues among 18 Cl$^-$ toxins was C-x-P-C-F-T-x(8)-C-x(2)-C-C-x(5,7)-C-x(2,3)-Q-C-[LI]-C. Importantly, the conserved cysteine pattern

Table 2 Motifs of Scorpion toxins extracted for Na$^+$, K$^+$ and Cl$^-$ channels. Kv – voltage-dependent K$^+$ channels; BK$_{Ca}$ – large-conductance Ca2$^+$-activated channel; SK$_{Ca}$ – small-conductance Ca2$^+$-activated channel; ERG – Ether-a-go-go K$^+$ channel. Hits – number of sequences that contain the submitted pattern upon search in Swiss-Prot and TrEMBL databases using ScanProsite. Expt. – experimentally determined function. No info. – functional information unavailable

Na$^+$		Hits	Expt.	No info.
α	R-D-x-Y-I-x(4)-N-C-x-Y-x-C-x(5,7)-C-N-x-x-C-T-x-x-G-A-x(3,4)-Y-C-x(6)-G-N-x-C-x-C-x-x-L-P-x(4)-I-x(4,5)-[KR]-C-[HR]	22	4	15
α-like	R-D-x-Y-I-A-x(3)-N-C-x(3)-C-x(3,6)-C-x-x-L-C-x(3)-G-x(3)-G-x-C-x(6)-G-x-x-C-W-C-x-x-L-P-x-x-V-x-I-x(3)-G-K-C-H	16	4	9
β-excitatory	G-x(3)-D-x-x-G-K-x-x-E-C-x(4,9)-Y-C-x-x-E-C-x-K-V-x-Y-A-x-x-G-Y-C-C-x(3)-C-Y-C-x-G-L-x(16)-C	9	4	5
β-mammal	K-x-G-Y-x-V-x(4)-G-C-x(3)-C-x-x-L-G-x-N-x-x-C-x-x-E-C-x(9)-G-Y-C-Y-x-F-x-C[WY]-C-x-x-L-x(8)-L-x-x-K-x-C	25	9	14

K$^+$		Hits	Expt.	No info.
Kv	C-x-x-[SP]-x(1,2)-C-[YWIDLG]-x-x-C-x(8,10)-K-C-[MI]-N-x-x-C-[KRH]-C	27	19	8
BK$_{Ca}$	C-x-x-[SP]-x(1,2)-C-[YWIDLG]-x-x-C-x(8,10)-K-C-[MI]-[NG]-x-x-C-[KRH]-C	28	7	19
SK$_{Ca}$	C-x-x-[RK]-[RM]-C-x(3)-C-[RK]-x(7)-C-x(4)-C-x-C	5	5	—
ERG	C-x(3)-Y-x-C-x(3)-C-K-x-R-F-x-K-x(3)-R-C-x(4)-C-x-C	2	1	1
Cl$^-$	C-x-P-C-F-T-x(8)-C-x(2)-C-C-x(5,7)-C-x(2,3)-Q-C-[LI]-C	14	—	14

```
              1        10        20        30        40        50        60
              |         |         |         |         |         |         |
D000004     KKNGYAVDSSGKAPEC-----LLSNYCNNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000137     KKNGYAVDSSGKAPEC-----LLSNYCNNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000125     KKNGYAVDSSGKAPEC-----LLSNYCNNQCTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000005     KKNGYAVDSSGKAPEC-----LLSNYCYNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000139     KKDGYAVDSSGKAPEC-----LLSNYCYNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000824     KKNGYAVDSSGKAPEC-----LLSNYCNNECTKVHYADKGYCCLLSCYCFGLSDDKKVLE 55
D000825     KKNGYAVDSSGKAPEC-----LLSNYCNNECTKVHYADKGYCCLLSCYCFGLSDDKKVLD 55
D000006     KKNGYAVDSSGKAPEC-----LLSNYCYNECTKVHYADKGYCCLLSCYCVGLSDDKKVLE 55
D000007     KKNGYAVDSSGKAPEC-----LLSNYCYNECTKVHYAEKGYCCLLSCYCVGLSDDKKVLE 55
D000233     KKNGYAVDSKGKAPEC-----FLSNYCNNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000234     KKNGYAVDSKGKAPEC-----FLSNYCNNECTKVHYADKGYCCLLSCYCFGLNDDKKVLE 55
D000235     KKNGYAVDSKGKAPEC-----FFSNYCNNECTKVHYAEKGYCCLLSCYCVGLNDDKKVME 55
D000236     KKNGFAVDSNGKAPEC-----FFDHYCNSECTKVYYAEKGYCCTLSCYCVGLDDDKKVLD 55
D000165     KKNGYAVDSSGKVAEC-----LFNNYCNNECTKVYYADKGYCCLLKCYCFGLADDKPVLD 55
D000003     KKNGYAVDSSGKVSEC-----LLNNYCNNICTKVYYATSGYCCLLSCYCFGLDDDKAVLK 55
D000136     KKNGYAVDSSGKVSEC-----LLNNYCNNICTKVYYATSGYCCLLSCYCFGLDDDKAVLK 55
D000184     KKNGYAVDSSGKVSEC-----LLNNYCNINCTKVYYATSGYCCLLSCYCFGLDDDKAVLK 55
D000001     KKNGYPLDRNGKTTECSGVNAIAPHYCNSECTKVYYAESGYCCWGACYCFGLEDDKPIGP 60
D000002     KKNGYPLDRNGKTTECSGVNAIAPHYCNSECTKVYYAKSGYCCWGACYCFGLEDDKPIGP 60
            **:*:..:* .**..**        :   :**   ****:** .****   ***.** *** :


              70
              |
D000004     ISDTRKSYCDTTIIN-- 70
D000137     ISDTRKSYCDTTIIN-- 70
D000125     ISDTRKSYCDTTIIN-- 70
D000005     ISDTRKSYCDTPIIN-- 70
D000139     ISDTRKSYCDTPIIN-- 70
D000824     ISDTRKKYCDYTIIN-- 70
D000825     ISDTRKKYCDYTIIN-- 70
D000006     ISDARKKYCDFVTIN-- 70
D000007     ISDARKKYCDFVTIN-- 70
D000233     ISGTTKKYCDFTIIN-- 70
D000234     ISDTTKKYCDFTIIN-- 70
D000235     ISDTRKKICDTTIIN-- 70
D000236     ISDTRKKLCDFTLFN-- 70
D000165     IWDSTKNYCDVQIIDLS 72
D000003     IKDATKSYCDVQIN--- 69
D000136     IKDATKSYCDVQIIG-- 70
D000184     IKDATKSYCDVQIIN-- 70
D000001     MKDITKKYCDVQIIPS- 76
D000002     MKDITKKYCDVQIIPS- 76
            :  .   *.  **
```

Figure 2 The conserved residues of 19 β-excitatory toxins: **KK**xGxxxDxxGKxxECx(4,9)YCxxx**CT**KVxYAxxGYCCxxxCYCxGLx **DDK**x(9)**K**xxC**D**. The first column represents the record accession number found in the SCORPION2 database; second column represents the toxin primary sequences; and third column represents the peptide length. Mutation studies were performed on Bjxtr-IT (D000001). Multiple sequence alignment was generated by clustal W.

in Cl⁻ toxins is different from that reported earlier in the PROSITE.

Sodium Channels – β-Excitatory Motif

Inclusion of mutation studies in the analyses of conserved residues helps to determine their degree of importance for structure and function of peptides. For example, in Na⁺ β-excitatory toxins, the conserved residues are: **KK**xGxxxDxxGKxxECx(4,9)YCxxx**CT**KV xYAxxGYCCxxxCYCxGLx**DDK**x(9)**K**xxC**D** (Figure 2). Mutation studies of Bjxtr-IT [29,30] demonstrated that toxicity to insects and binding affinity to insect Na⁺ channels as compared to that of wild type Bjxtr-IT were mildly affected (<7-fold) for K1, K2, T32, D54, D55,

K56, K66 and D70 (shown in bold). However, mutations at D8, K12, Y26, K33 and V34 strongly affected binding affinity, ranging from 12- to >10 000-fold reduction in binding affinity. Thus, different conserved residues played different roles for function and structure. Further, when negative-charged E15 was mutated to positive-charged arginine, toxicity to insects was severely affected (>10 000-fold) but not binding affinity (<7-fold). This suggests that negative-charged E15 is involved in toxic action. Another residue important for toxicity and binding affinity is glutamate at position 30 in the multiple sequence alignment of β-excitatory toxins (Figure 2). Residues are not conserved at position 30 because glutamine, isoleucine and asparagine were also observed. Mutation of E30 to

glutamine, leucine, arginine and even conservative substitution to aspartate caused reduction of more than eight-fold in toxicity and 43-fold in binding affinity. Without mutant data, this position and residue identity (negative-charged and hydrophilic), which is important for function, would be overlooked. Thus, by incorporation of mutation information, the functional motif of β-excitatory toxins to insect Na^+ channels can be summarized as G-x(3)-D-x-x-G-K-x-x-E•C-x(4,9)-Y-C-x-x-E•-x-K-V-x-Y-A-x-x-G-Y-C-C-x(3)-C-Y-C-x-G-L-x(16)-C, where E is important for toxicity to insects.

Sodium Channels – β-Mammal Motif

In β-mammal specific toxins, the conserved residues in this group are KxGYxVx(4)GC**K**xxCxxLGxNxxCxxECx(9)GYCYxFxCxCxxLx(7)**P**LxxKxC. Mutation studies of Css 4 [14] demonstrated that mutation of K13 and P61 (shown in bold) did not affect binding affinity to mammal sodium channels significantly ($\leq$2-fold). Y4, W47 and K63 are involved in structural integrity – their mutations to alanine disrupted secondary structures. Substitution of N22, Y40, Y42 and F44 to alanine resulted in a 600-fold reduction of binding affinity while L19 to alanine resulted in a 150-fold reduction in binding affinity to mammal sodium channels. This suggests that some positions in the sequence are more important in binding affinity than others. Information such as the physicochemical properties of functional residues can also be obtained from mutation studies. For example, a stronger reduction on binding affinity was obtained upon charge inversion of E28 to arginine (900-fold) than charge neutralizing substitutions to alanine (600-fold), glutamine (400-fold) and leucine (50-fold). The functional motif for this group can be summarized as K-x-G-Y-x-V-x(4)-G-C-x(3)-C-x-x-L-G-x-N-x-x-C-x-x-E-C-x(9)-G-Y-C-Y-x-F-x-C[WY]-C-x-x-L-x(8)-L-x-x-K-x-C.

Sodium Channels – α-Motif

In α-toxins, the motif extracted from 14 experimentally determined α-toxins was R-D-x-Y-I-x(4)-N-C-x-Y-x-C-x(5,7)-C-N-x-x-C-T-x-x-G-A-x(3,4)-Y-C-x(6)-G-N-x-C-x-C-x-x-L-P-x(4)-I-x(4,5)-[KR]-C-[HR]. From mutation studies of Lqh αIT [31,32], residues at positions 15, 25, 28, 56 and 54 (as in Lqh αIT) can be mutated to alanine without significant effect on binding affinity to insect Na^+ channels (<2-fold reduction), which correlated well with the observation that there is no conservation of residues at these positions. Conserved N44 is involved in structural and functional integrity because mutation to alanine disrupted the secondary structures and reduced binding affinity by 31-fold. Mutation of I57 to alanine and threonine caused more than 92-fold reduction in binding affinity. Positive-charged residues at position 62 and 64 were essential to binding affinity, as

charge neutralization to alanine resulted in more than a 54-fold reduction.

Sodium Channels – α-Like Motif

The motif for α-like toxins is R-D-x-Y-I-A-x(3)-N-C-x-x(3)-C-x(3,6)-C-x-x-L-C-x(3)-G-x(3)-G-x-C-x(6)-G-x-x-C-W-C-x-x-L-P-x-x-V-x-I-x(3)-G-K-C-H. Site-directed mutagenesis of Bmk M1 [33,34] highlighted the importance of four residues important for function and structure in α-like toxins. Mutations at Y5 and N11 disrupted the secondary structures as measured by circular dichroism, suggesting their role in maintaining structural integrity. This is corroborated by the crystal structure of BmK M1 in which N11 forms hydrogen bonds with residues 58 and 59 [35]. Positive charge at positions 62 and 64 is important for binding affinity to insect Na^+ channel in which negative-charged residues (D and E) resulted in more than 167-fold reduction. Mutation at P9 did not significantly affect function (two-fold reduction in toxicity and binding affinity).

Potassium Channel Subtype – Ether-a-go-go-related K$^+$ Channel Motif

Scorpion toxins target various K^+ channel subtypes with different specificities. In ether-a-go-go-related gene K^+ (ERG) channel subtype, residues that are important for binding of BeKm-1 to the ERG channel had been identified by mutagenesis [36]. Y11, K18, R20 and K23 were critical to binding while F21 and R27 were involved in structural folding. The functional residues were located on the α-helix and the following loop. The recognition motif for ERG subtype can be expressed as C-x(3)-Y-x-C-x(3)-C-K-x-R-F-x-K-x(3)-R-C-x(4)-C-x-C.

Potassium Channel Subtype – Small-conductance Ca^{2+}-activated K$^+$ Channel Motif

The ERG functional surface is similarly shared by scorpion toxins that target small-conductance Ca^{2+}-activated K^+ channels (SK_{Ca}). Mutagenesis of leiurotoxin and P05 identified two positions, 6 and 7, as important for binding to this subtype [37–40]. However, Ts-κ also targets this subtype and had two functional residues, R6 and R9, located outside the α-helix structure [41]. Though the positions of the functional residues are not spatially conserved, the residue identity required to bind to SK_{Ca} was positive-charged arginine in which mutation to hydrophobic leucine and even conservative substitution to lysine affected binding affinity.
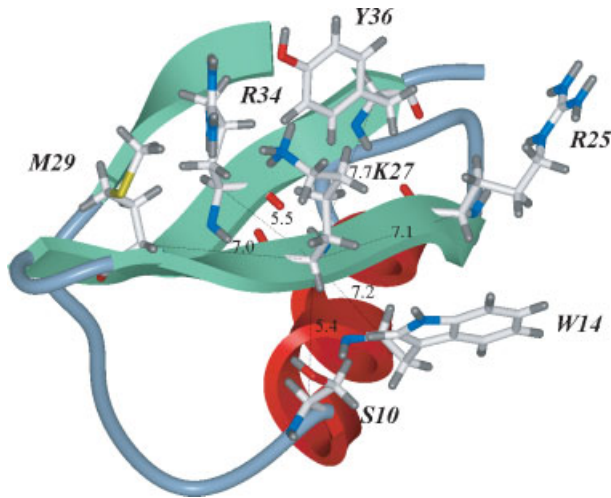
Figure 3 Functional residues of charybdotoxin (PDB: 2CRD) determined by mutagenesis studies, which were important for binding to voltage-dependent K$^+$ and large-conductance Ca^{2+}-activated K$^+$ channels. Most of the functional residues (R25, K27, M29, R34 and Y36), except S10 and W14, were located on the flat surface of the β-sheets. S10 and W14 reside at the α-helix. Spatial distances between Cα atoms of the functional residues with that of critical K27 demonstrated that they are within 6.6 ± 1.0 Å. This may explain their involvement in interaction due to their close proximity to the functional dyad.

Potassium Channel Subtypes – Large-conductance Ca^{2+}-activated K$^+$ Channel and Voltage-dependent K$^+$ Channel Motifs

For scorpion toxins that bind to voltage-dependent K$^+$ channels (K$_v$), Goldstein and Miller [42] had demonstrated that lysine at position 27 in charybdotoxin from *Leiurus quinquestriatus hebraeus* physically occlude the pore of this subtype, thus preventing the flow of K$^+$ ions. This functional residue, lysine, and an aromatic residue (tyrosine or phenylalanine) separated by 6.6 ± 1.0 Å formed the functional dyad that targets the K$_v$ channels [43]. In addition, mutagenesis of charybdotoxin on K$_v$ channels and large-conductance Ca^{2+}-activated channels (BK$_{Ca}$) highlighted several positions that are important for binding [42,44,45]. These include positions 10, 14, 25, 29 and 34 where mutations led to drastic reduction in binding affinity. The functional residues for K$_v$ and BK$_{Ca}$ channels are located on the β-sheets in contrast to that for ERG and SK$_{Ca}$ channels that are located at the α-helix. These residues also fall within the functional dyad radius of 6.6 ± 1.0 Å, which may explain their involvement in toxin–channel interaction due to their close proximity to the functional dyad (Figure 3). Only one residue difference in the functional surface distinguishes K$_v$ channels and BK$_{Ca}$ channels, in which asparagine is specific to K$_v$ channels. This was determined from analyses of multiple sequence alignment of scorpion toxins that target different channel subtypes and mutant data. Iberiotoxin

that is highly specific for BK$_{Ca}$ channel [46] has glycine instead of asparagine at position 30. P05 targets SK$_{Ca}$ channel only and amino acid mutation at positions 22–24 from IGD to MNG resulted in recognition of K$_v$ channels [40]. This is corroborated by Schroeder *et al.* [47], where they mutated glycine to asparagine at position 30 in iberiotoxin, which caused the mutant toxin to target both K$^+$ channel subtypes.

CONCLUSIONS

Given the large diversity of ion channels [48,49], a number of different binding motifs in scorpion toxins can be defined. Dauplais *et al.* [43] reported a conservation of a functional dyad motif in K$^+$ toxins with unrelated structures while none has been reported for Na$^+$, Cl$^-$ and Ca^{2+} toxins. This is the first report of peptide binding motifs for four K$^+$ ion channel subtypes (voltage-dependent K$^+$ channels, large- and small-conductance Ca2$^+$-activated channels and ether-a-go-go channel), four binding site motifs for Na$^+$ channels and a conserved motif for Cl$^-$ channels.

The motifs reported here included information from mutation studies of scorpion toxins and 3D structure analyses, except for Cl$^-$ channel for which mutation study is not available. The motifs reported in this article have biological and functional relevance which complement motifs obtained statistically. Mutation studies help determine whether conserved residues within groups of scorpion toxins are important for the integrity of molecular structure and for function. The mutant data provide information on residues and positions that are critical for structure and function, and those that are not. This is important because even semi-conservative substitution of residues can affect activity in scorpion toxins in which a substitution from arginine to lysine resulted in a 77-fold decreased activity in LqhαIT [31]. Scaling the effects of mutation on binding affinity to a common scale provides a semi-quantitative comparison to differentiate critical residues from those that play no role in binding affinity.

3D structure analyses aid the identification of motifs in which non-contiguous residues are clustered spatially. Spatial proximity to critical functional residues is likely to influence interaction, especially in toxin–channel complexes, which involves a combination of electrostatic, hydrophobic and hydrogen-bonding interactions [50,51]. Information of spatial proximity on residues near critical functional residue facilitates the design of mutation studies.

This systematic approach of including mutant data of scorpion toxins and 3D structure analyses for extraction of motifs can serve as a model for other proteins in which mutation studies and 3D structures are available. This approach complements motifs

that are derived statistically, by including biological information for a more accurate inference of function for newly identified scorpion toxin sequences.

REFERENCES

1. Huang X, Liu H, Cui M, Fu W, Yu K, Chen K, Luo X, Shen J, Jiang H. Simulating the interactions of toxins with K$^+$ channels. *Curr. Pharm. Des.* 2004; **10**: 1057–1067.

2. Mouhat S, Jouirou B, Mosbah A, De Waard M, Sabatier JM. Diversity of folds in animal toxins acting on ion channels. *Biochem. J.* 2004; **378**: 717–726.

3. Arnon T, Potikha T, Sher D, Elazar M, Mao W, Tal T, Bosmans F, Tytgat J, Ben-Arie N, Zlotkin E. BjalphaIT: a novel scorpion alpha-toxin selective for insects-unique pharmacological tool. *Insect Biochem. Mol. Biol.* 2005; **35**: 187–195.

4. Gilquin B, Bourgoin M, Menez R, Le Du MH, Servent D, Zinn-Justin S, Menez A. Motions and structural variability within toxins: implication for their use as scaffolds for protein engineering. *Protein Sci.* 2003; **12**: 266–277.

5. Maria WS, Velarde DT, Alvarenga LM, Nguyen C, Villard S, Granier C, Chavez-Olortegui C. Localization of epitopes in the toxins of *Tityus serrulatus* scorpions and neutralizing potential of therapeutic antivenoms. *Toxicon* 2005; **46**: 210–217.

6. Gazarian KG, Gazarian T, Hernandez R, Possani LD. Immunology of scorpion toxins and perspectives for generation of anti-venom vaccines. *Vaccine* 2005; **23**: 3357–3368.

7. Gilles N, Gurevitz M, Gordon D. Allosteric interactions among pyrethroid, brevetoxin, and scorpion toxin receptors on insect sodium channels raise an alternative approach for insect control. *FEBS Lett.* 2003; **540**: 81–85.

8. Regev A, Rivkin H, Inceoglu B, Gershburg E, Hammock BD, Gurevitz M, Chejanovsky N. Further enhancement of baculovirus insecticidal efficacy with scorpion toxins that interact cooperatively. *FEBS Lett.* 2003; **537**: 106–110.

9. Wang J, Chen Z, Du J, Sun Y, Liang A. Novel insect resistance in *Brassica napus* developed by transformation of chitinase and scorpion toxin genes. *Plant Cell Rep.* 2005; **24**: 549–555.

10. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; **25**: 3389–3402.

11. Pearson WR. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* 2000; **132**: 185–219.

12. Hulo N, Sigrist CJ, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A. Recent improvements to the PROSITE database. *Nucleic Acids Res.* 2004; **32**: Database issue: D134–D137.

13. Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, Mitchell AL, Moulton G, Nordle A, Paine K, Taylor P, Uddin A, Zygouri C. PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.* 2003; **31**: 400–402.

14. Cohen L, Karbat I, Gilles N, Ilan N, Benveniste M, Gordon D, Gurevitz M. Common features in the functional surface of scorpion β-toxins and elements that confer specificity for insect and mammalian voltage-gated sodium channels. *J. Biol. Chem.* 2005; **280**: 5045–5053.

15. Huys I, Tytgat J. Evidence for a function-specific mutation in the neurotoxin, parabutoxin 3. *Eur. J. Neurosci.* 2003; **17**: 1786–1792.

16. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2005; **33**: Database Issue: D34–D38.

17. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-prot: juggling between evolution and stability. *Brief. Bioinform.* 2004; **5**: 39–55.

18. Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM, Bourne PE. The RCSB protein data bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.* 2005; **33**: Database Issue: D233–D237.

19. Kong L, Lee BT, Tong JC, Tan TW, Ranganathan S. SDPMOD: an automated comparative modeling server for small disulfide-bonded proteins. *Nucleic Acids Res.* 2004; **32**: W356–W359.

20. Srinivasan KN, Gopalakrishnakone P, Tan PT, Chew KC, Cheng B, Kini RM, Koh JL, Seah SH, Brusic V. SCORPION, a molecular database of scorpion toxins. *Toxicon* 2002; **40**: 23–31.

21. Tan PT, Srinivasan KN, Seah SH, Koh JL, Tan TW, Ranganathan S, Brusic V. Accurate prediction of scorpion toxin functional properties from primary structures. *J. Mol. Graph. Model.* 2005; **24**: 17–24.

22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; **22**: 4673–4680.

23. Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 2004; **5**: 150–163.

24. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987; **4**: 406–425.

25. Koschak A, Bugianesi RM, Mitterdorfer J, Kaczorowski GJ, Garcia ML, Knaus HG. Subunit composition of brain voltage-gated potassium channels determined by hongotoxin-1, a novel peptide derived from *Centruroides limbatus* venom. *J. Biol. Chem.* 1998; **273**: 2639–2644.

26. Hassani O, Mansuelle P, Cestele S, Bourdeaux M, Rochat H, Sampieri F. Role of lysine and tryptophan residues in the biological activity of toxin VII (Ts gamma) from the scorpion *Tityus serrulatus*. *Eur. J. Biochem.* 1999; **260**: 76–86.

27. Gattiker A, Gasteiger E, Bairoch A. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics* 2002; **1**: 107–108.

28. Ranganathan R, Lewis JH, MacKinnon R. Spatial localization of the K$^+$ channel selectivity filter by mutant cycle-based structure analysis. *Neuron* 1996; **16**: 131–139.

29. Karbat I, Cohen L, Gilles N, Gordon D, Gurevitz M. Conversion of a scorpion toxin agonist into an antagonist highlights an acidic residue involved in voltage sensor trapping during activation of neuronal Na$^+$ channels. *Faseb J.* 2004; **18**: 683–689.

30. Cohen L, Karbat I, Gilles N, Froy O, Corzo G, Angelovici R, Gordon D, Gurevitz M. Dissection of the functional surface of an anti-insect excitatory toxin illuminates a putative "hot spot" common to all scorpion beta-toxins affecting Na$^+$ channels. *J. Biol. Chem.* 2004; **279**: 8206–8211.

31. Karbat I, Frolow F, Froy O, Gilles N, Cohen L, Turkov M, Gordon D, Gurevitz M. Molecular basis of the high insecticidal potency of scorpion alpha-toxins. *J. Biol. Chem.* 2004; **279**: 31 679–31 686.

32. Zilberberg N, Froy O, Loret E, Cestele S, Arad D, Gordon D, Gurevitz M. Identification of structural elements of a scorpion alpha-neurotoxin important for receptor site recognition. *J. Biol. Chem.* 1997; **272**: 14 810–14 816.

33. Wang CG, Gilles N, Hamon A, Le Gall F, Stankiewicz M, Pelhate M, Xiong YM, Wang DC, Chi CW. Exploration of the functional site of a scorpion alpha-like toxin by site-directed mutagenesis. *Biochemistry* 2003; **42**: 4699–4708.

34. Sun YM, Bosmans F, Zhu RH, Goudet C, Xiong YM, Tytgat J, Wang DC. Importance of the conserved aromatic residues in the scorpion alpha-like toxin BmK M1: the hydrophobic surface region revisited. *J. Biol. Chem.* 2003; **278**: 24 125–24 131.

35. He XL, Li HM, Zeng ZH, Liu XQ, Wang M, Wang DC. 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.* 1999; **292**: 125–135.

36. Korolkova YV, Bocharov EV, Angelo K, Maslennikov IV, Grinenko OV, Lipkin AV, Nosyreva ED, Pluzhnikov KA, Olesen SP,

Arseniev AS, Grishin EV. New binding site on common molecular scaffold provides HERG channel specificity of scorpion toxin BeKm-1. *J. Biol. Chem.* 2002; **277**: 43 104–43 109.

37. Sabatier JM, Zerrouk H, Darbon H, Mabrouk K, Benslimane A, Rochat H, Martin-Eauclaire MF, Van Rietschoten J. P05, a new leiurotoxin I-like scorpion toxin: synthesis and structure-activity relationships of the alpha-amidated analog, a ligand of Ca^{2+}-activated K$^+$ channels with increased affinity. *Biochemistry* 1993; **32**: 2763–2770.

38. Sabatier JM, Fremont V, Mabrouk K, Crest M, Darbon H, Rochat H, Van Rietschoten J, Martin-Eauclaire MF. Leiurotoxin I, a scorpion toxin specific for Ca^{2+}-activated K$^+$ channels. Structure-activity analysis using synthetic analogs. *Int. J. Pept. Protein Res.* 1994; **43**: 486–495.

39. Shakkottai VG, Regaya I, Wulff H, Fajloun Z, Tomita H, Fathallah M, Cahalan MD, Gargus JJ, Sabatier JM, Chandy KG. Design and characterization of a highly selective peptide inhibitor of the small conductance calcium-activated K$^+$ channel, SK$_{Ca}$2. *J. Biol. Chem.* 2001; **276**: 43 145–43 151.

40. Wu JJ, He LL, Zhou Z, Chi CW. Gene expression, mutation, and structure-function relationship of scorpion toxin BmP05 active on SK(Ca) channels. *Biochemistry* 2002; **41**: 2844–2849.

41. Lecomte C, Ferrat G, Fajloun Z, Van Rietschoten J, Rochat H, Martin-Eauclaire MF, Darbon H, Sabatier JM. Chemical synthesis and structure-activity relationships of Ts kappa, a novel scorpion toxin acting on apamin-sensitive SK channel. *J. Pept. Res.* 1999; **54**: 369–376.

42. Goldstein SA, Miller C. Mechanism of charybdotoxin block of a voltage-gated K$^+$ channel. *Biophys. J.* 1993; **65**: 1613–1619.

43. Dauplais M, Lecoq A, Song J, Cotton J, Jamin N, Gilquin B, Roumestand C, Vita C, de Medeiros CL, Rowan EG, Harvey AL, Menez A. On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *J. Biol. Chem.* 1997; **272**: 4302–4309.

44. Stampe P, Kolmakova-Partensky L, Miller C. Intimations of K$^+$ channel structure from a complete functional map of the molecular surface of charybdotoxin. *Biochemistry* 1994; **33**: 443–450.

45. Park CS, Miller C. Mapping function to structure in a channel-blocking peptide: electrostatic mutants of charybdotoxin. *Biochemistry* 1992; **31**: 7749–7755.

46. Galvez A, Gimenez-Gallego G, Reuben JP, Roy-Contancin L, Feigenbaum P, Kaczorowski GJ, Garcia ML. Purification and characterization of a unique, potent, peptidyl probe for the high conductance calcium-activated potassium channel from venom of the scorpion Buthus tamulus. *J. Biol. Chem.* 1990; **265**: 083 11–90.

47. Schroeder N, Mullmann TJ, Schmalhofer WA, Gao YD, Garcia ML, Giangiacomo KM. Glycine 30 in iberiotoxin is a critical determinant of its specificity for maxi-K versus K$_V$ channels. *FEBS Lett.* 2002; **527**: 298–302.

48. Zuo XP, Ji YH. Molecular mechanism of scorpion neurotoxins acting on sodium channels: insight into their diverse selectivity. *Mol. Neurobiol.* 2004; **30**: 265–278.

49. Korn SJ, Trapani JG. Potassium channels. *IEEE Trans. Nanobioscience* 2005; **4**: 21–33.

50. Xu CQ, Zhu SY, Chi CW, Tytgat J. Turret and pore block of K$^+$ channels: what is the difference? *Trends Pharmacol. Sci.* 2003; **24**: 446–448; author reply 448–449.

51. Mouhat S, Mosbah A, Visan V, Wulff H, Delepierre M, Darbon H, Grissmer S, De Waard M, Sabatier JM. The 'functional' dyad of scorpion toxin Pi1 is not itself a prerequisite for toxin binding to the voltage-gated Kv1.2 potassium channels. *Biochem. J.* 2004; **377**: 25–36.