Journal of Peptide Science
J. Pept. Sci. 2006; **12**: 679–685
Published online 11 July 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.778

JOURNAL OF PEPTIDE SCIENCE

I-conotoxin superfamily revisited

SUKANTA MONDAL,[a] RAJASEKARAN MOHAN BABU,[b] RAJASEKARAN BHAVNA[b] and SURYANARAYANARAO RAMAKUMAR[a,b]*

[a] Department of Physics, Indian Institute of Science, Bangalore 560012, India
[b] Bioinformatics Centre, Indian Institute of Science, Bangalore 560012, India

Received 16 March 2006; Revised 26 May 2006; Accepted 27 May 2006

Abstract: The I-conotoxin superfamily (I-Ctx) is known to have four disulfide bonds with the cysteine arrangement C-C-CC-CC-C-C, and the members inhibit or modify ion channels of nerve cells. Recently, Olivera and co-workers (*FEBS J.* 2005; 272: 4178–4188) have suggested that the previously described I-Ctx should now be divided into two different gene superfamilies, namely, I_1 and I_2, in view of their having two different types of signal peptides and exhibiting distinct functions. We have revisited the 28 entries presently grouped as I-Ctx in UniProt Swiss-Prot knowledgebase, and on the basis of *in silico* analysis have divided them into I_1 and I_2 superfamilies. The sequence analysis has provided a framework for *in silico* annotation enabling us to carry out computer-based functional characterization of the UniProtKB/TrEMBL entry Q59AA4 from *Conus miles* and to predict it as a member of the I_2 superfamily. Furthermore, we have predicted the mature toxin of this entry and have proposed that it may be an inhibitor of voltage-gated potassium channels. Copyright © 2006 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: signal peptide; mature toxin; gene superfamily; potassium channel inhibitor; functional annotation

INTRODUCTION

Marine cone snails are carnivores that show a high degree of specificity for certain prey types – worms (vermivorous), other molluscs (molluscivorous) or fish (piscivorous). They use venom synthesized from the modified salivary gland and convoluted duct for feeding and defence [1]. Conotoxins are a complex mixture of pharmacologically active and conformationally constrained peptides that target specific ion channels or G protein coupled receptors [2]. They have been classified into several superfamilies on the basis of a number of characteristics such as a highly conserved *N*-terminal precursor sequence, disulphide connectivity and similar mode of action. Each superfamily has been further categorized into different families on the basis of their specific pharmacological targets [1,3,4].

The I-conotoxin superfamily (I-Ctx) is known to have four disulphide bonds with the cysteine arrangement C-C-CC-CC-C-C, and the members inhibit or modify ion channels of nerve cells. I-Ctx were detected both in vermivorous and piscivorous species of *Conus*, suggesting the widespread presence of such toxins beyond evolutionary and ecological groups [5]. The superfamily has been less explored when compared to others in the *Conus* species. Recently, we reported a systematic sequence and structural analysis for I-Ctx based on a theoretical 3D model [6]. We have deposited the sequence pattern for I-Ctx in the PROSITE database [7] with accession number PS60019 under PDOC60004 documentation [6]. This information can be accessed from the InterPro database [8] with accession number IPR013141, which includes 37 entries (28 entries from UniProtKB/Swiss-Prot [9] and nine entries from UniProtKB/TrEMBL [10]). Similar information about the superfamily is present in Pfam [11] with accession number PF08088, and the corresponding InterPro accession number is IPR012624.

The previously characterized I-Ctx have been redefined by Olivera and co-workers [12], who have demonstrated the presence of two different gene superfamilies I_1 and I_2 with different signal peptide sequences, though having similar cysteine arrangements. Earlier, identical cysteine arrangement, although with different cysteine connectivity, has been observed in α- and λ-conotoxin families respectively from A- and T-superfamilies [13,14]. It has been suggested that the I_1 gene superfamily has unique post-translational modifications at the *C*-terminus, while I_2 lacks a propeptide region (Figure 1), differing in this aspect from all other superfamilies. Olivera and co-workers have demonstrated that the peptide r11a, belonging to I-Ctx having 46 amino acids, has a D-Phe residue at position 44 responsible for inducing repetitive activity in the nerve [15]. The I-Ctx peptides r11b and r11c show sequence similarity to r11a to varying extents and were examined with L-Phe44 and L-Phe42, respectively. The experimental results indicate that r11a and r11b induced repetitive activity in the nerve only with D-Phe44, whereas with L-Phe44 both the sequences were inactive. The sequence r11c was equipotent in inducing repetitive action potentials in both motor nerve and skeletal muscle with L-Phe42 as well as D-Phe42. These peptides were identified to undergo

*Correspondence to: S Ramakumar, Department of Physics, Indian Institute of Science, Bangalore 560012, India;
e-mail: ramak@physics.iisc.ernet.in

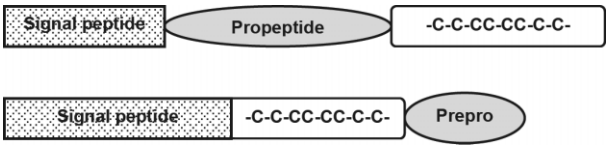

Figure 1 Schematic representation of I_1 and I_2 conotoxin superfamilies, top and bottom, respectively.

post-translational isomerization at the third amino acid from the *C*-terminus and they were put together as group A. Another group of sequences, though sharing high sequence identity with group A but those not having a D–amino acid at the third position from the *C*-terminus, were termed as group B. Group A and group B have the same signal sequence, MKLCVTFLLVLMILPSVTG/EKSSERTLSGALLRGVKRR, and the two groups together have been termed as the *I_1 superfamily*. Another superfamily, group C, which does not undergo post-translational isomerization and which has a different signal sequence MMFRVTSVGCFLLVIVFLNLVVLTDA has been named as the *I_2 superfamily* [12].

In view of the recent proposal that I-Ctx should be divided into I_1 and I_2 superfamilies [12], it is desirable to revisit the database entries that are currently grouped as I-Ctx. We have used a systematic approach to analyse the 28 UniProtKB/Swiss-Prot I-Ctx and have split them into two distinct gene superfamilies using bioinformatics approaches. In this report, we discuss the two groups that apparently share the same cysteine arrangement C-C-CC-CC-C-C, while differing in sequence composition, pattern and therefore their functions. However, experimental 3D structures are not available for these superfamilies, which limits the analysis at the sequence level itself. We have been able to predict the position of signal peptide, prepro-region and mature toxin for the unannotated UniProtKB/TrEMBL entry, namely, Q59AA4 from *Conus miles*, and this approach provides a framework for future annotations.

MATERIALS AND METHODS

Datasets

The amino acid sequences of 28 entries were obtained from the UniProt Swiss-Prot knowledgebase (Release 49.0). The dataset was used for categorization of I_1 and I_2 superfamilies. A non-redundant set was created using the program CD-HIT [16] with sequence matches of <70% sequence identity. This set comprising 13 entries (Table 1) was used for computing the amino acid composition. The data set entries have been referred in this paper with their UniProtKB accession numbers.

Superfamily Grouping

A few sequences, which have been identified by Olivera and co-workers as belonging to either I_1 or I_2 superfamilies have been termed by us as *reference sequences* and have formed the basis of the present analysis. A sequence level comparison was done between the UniProtKB/Swiss-Prot entries and the reference sequences whose superfamily was known as being either I_1 or I_2. For a given superfamily, the signal peptide is well conserved and hence this information, if available, can be exploited for grouping the database sequences. In cases where this information was not present, we compared the mature toxin region alone with the corresponding region of the reference sequences. The percentage identity and similarity were calculated for such cases using the method of global alignment by GCG [17] program 'GAP'.

Multiple Sequence Alignment

A multiple sequence alignment (MSA) separately for I_1 and I_2 superfamilies was done using EBI ClustalW [18] with

Table 1 I_1 and I_2 conotoxins superfamily members: Group B members are in bold and the non-redundant entries are highlighted

Superfamily	Species	UniProtKB/Swiss-Prot entries[a]
I_1	*C. radiatus*	Q7Z0A5 (1–44), Q7Z0A3 (1–44), Q7Z0A1 (1–42), Q7Z098 (1–45), Q7Z091 (1–41), Q7Z0A0 (1–43), Q7Z092 (1–46), Q7Z093 (1–46), Q7Z0A6 (1–42), Q7Z099 (1–45), Q7Z094 (1–46), Q7Z095 (1–46), Q7Z096 (1–43), Q7Z0A4 (1–44), Q7Z0A2 (1–42), Q7Z097 (1–45), **Q7Z090** (1–40), **Q7M4K5** (1–37)
I_2	*C. miles*	P69498 (27–60)
	C. striatus	P69499 (27–61)
	C. vexillum	P69500 (27–60), P69501 (27–60)
	C. betulinum	Q9U3Z3 (27–57)
	C. imperialis	P69495 (27–60), P69496 (27–60), P69497 (27–60)
	C. capitaneus	P69494 (27–61)
	C. virgo	Q7YZS9 (27–60)

[a] Start and end positions of mature toxin are mentioned in parentheses for the corresponding entries.

DOI: 10.1002/psc

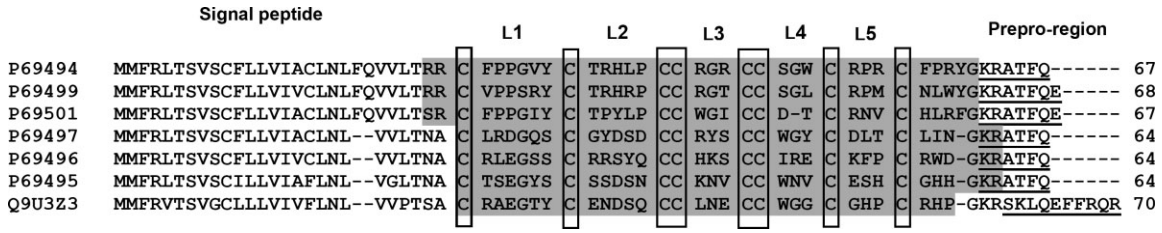

Figure 2 MSA of non-redundant sequences belonging to I_2 conotoxin superfamily: mature toxin regions are highlighted, prepro-regions are underlined, conserved cysteines residues are in boxes and inter-cysteine loops are numbered from L1 to L5. The last column indicates the length of the proteins. Note that the prepro-region for K^+ channel inhibitors is KRATFQ/KRATFQE in contrast with that of the modulator, Q9U3Z3, which is SKLQEFFRQR.

default parameters. Using the MSAs, the type of residues in the inter-cysteine loops as well as conserved positions in the alignment were observed and the sequence pattern for both the superfamilies were defined. The ScanProsite [19] from the ExPASy server (http://www.expasy.org) was used to evaluate the newly created sequence patterns. The MSAs for I_1 and I_2 superfamilies were also used for generating profile Hidden Markov model (HMM) using Dr Sean Eddy's HMMER package [20] available in the GCG Package [17]. We used HmmerBuild to create a position-specific scoring table that represents the primary structure consensus of a given sequence family called *profile HMM*. In order to increase the sensitivity of the database search performed using that profile as a query, we used HmmerCalibrate, which 'calibrates' a profile HMM by generating random sequences and computing a raw score for the comparison between each sequence and the profile. The program then fits the distribution of these scores to an extreme value distribution. The calibrated profile was used for searching the UniProt database using HmmerSearch module. The profiles for I_1 and I_2 can be obtained from the supplementary material. The values for sensitivity, selectivity, specificity and MCC were estimated as defined in [Ref.6] so as to assess the reliability of patterns and profile HMM, when scanned against the UniProtKB/Swiss-Prot.

Amino Acid Composition

The composition of an amino acid i for the mature toxin sequences in a superfamily m (I_1 and I_2) were calculated using [Eqn (1)].

$$p_m^i = \frac{n_m^i}{N_m}, \quad m = I_1, I_2 \tag{1}$$

where n_m^i is the total number of i^{th} amino acid in all sequences of m superfamily and N_m is the total number of all amino acids in all sequences of m superfamily.

RESULTS AND DISCUSSION

Recognition of I_1 and I_2 Members

In order to implement the proposed grouping, 28 UniProtKB/Swiss-Prot sequences (at present referred to as *I-Ctx*) were compared with the reference sequences [12]. A quick observation on the sequences in the UniProtKB/Swiss-Prot indicated the presence of the signal sequence only for ten entries. A consensus was observed in the signal sequence similar to those of the reference sequences of the I_2 superfamily members (Figure 2). Therefore, the ten entries (Table 1) were categorized as belonging to I_2 superfamily.

No information was available about the signal sequences for the remaining 18 UniProtKB/Swiss-Prot entries. These entries were scanned against the mature toxin region of the reference sequences for assessing the similarity and identity. The UniProtKB/Swiss-Prot entry having maximum identity with any one of the reference sequences was taken into account and the grouping was based on the highest value of the identity obtained. The 16 entries showed sequence identities ranging from 73 to 100% with the group A reference sequences. The entry Q7M4K5 was same as the peptide r11e, belonging to the group B of the I_1 superfamily. The peptide r11e was shown to be closely related to the clone R11.3 in Ref. 21. R11.3 is actually the UniProtKB/Swiss-Prot entry Q7Z090, and therefore it is concluded that this entry too belongs to the group B of the I_1 superfamily. Hence, all the 18 entries showed more similarity to the mature toxin of the I_1 superfamily, where 16 entries are from group A and two are from group B. We have grouped the 18 entries as belonging to the I_1 superfamily and the remaining ten entries to the I_2 superfamily (Table 1). For the reference sequences, the consensus sequence pattern (termed as *reference sequence pattern*) as previously reported by Olivera and co-workers [12] for the three groups A, B and C was defined to be C-X(6)-C-X(5)-C-C-X-C-C-X(4)-C-X(8,10)-C, C-X(6)-C-X(5)-C-C-X(3)-C-C-X(4)-C-X(6)-C and C-X(6)-C-X(5)-C-C-X(3)-C-C-X(2,3)-C-X(3)-C, respectively, and these were used for verifying the grouping done by us. The reference sequence patterns for all the three groups were able to pick the same UniProtKB/Swiss-Prot entries (Table 1) that were categorized by us confirming correct grouping.

The I_1 Superfamily

The MSA for the non-redundant set of mature toxins showed eight conserved cysteine positions. Apart from this, conserved residues like His in loop 2, and Gly

DOI: 10.1002/psc

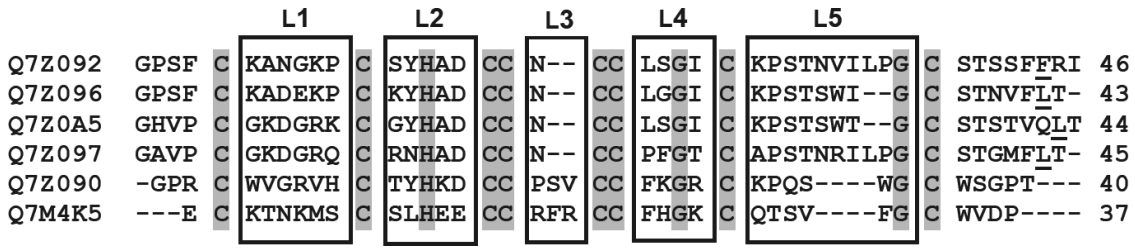

Figure 3 MSA of non-redundant sequences belonging to the I_1 conotoxin superfamily. The first four entries belong to group A and the last two entries to group B. The presence of the D-amino acid (most probable post-translational modification site) at the third position from the *C*-terminus is underlined; arginine cleaved by a carboxypeptidase in Q7Z096, Q7Z0A5, Q7Z097 is not shown. Conserved residues are highlighted and inter-cysteine loops are numbered from L1 to L5. The last column indicates the length of the proteins.

in loop 4 and loop 5, respectively, (Figure 3) were observed in the inter-cysteine loops. The functional and/or structural importance of such conserved positions needs to be ascertained through experimental investigations.

Since only two UniProtKB/Swiss-Prot members form group B, we attempted to create a pattern for all the I_1 superfamily members i.e. for group A and group B together. Using the MSA, we defined the sequence pattern for the I_1 superfamily as C-{C}-{G}-{C}(4)-C-{C}(5)-C-C-{C}(1, 3)-C-C-{C}(4)-C-{C}(6, 10)-C, where the regular expressions are according to the PROSITE format. The pattern was scanned against UniProtKB/Swiss-Prot, release 49.7, using ScanProsite and the results indicated a good-quality cluster since the pattern picked up the 18 true positive hits. The profile HMM created for the I_1 superfamily was also used for searching the same database, which resulted in top 18 hits with an E-value between 10^{-27} and 10^{-12}. The E-value for the remaining hits was greater than 0.6. Thus both the methods performed well yielding only true positives. The sensitivity, selectivity, specificity and correlation coefficient were thus unity. It should be noted that the training set (18 entries used for creating the pattern and profile) is a subset of the independent set (UniProtKB/Swiss-Prot).

The functional characteristics of the entries of this superfamily have been previously described providing glimpses of their molecular complexity [5,6,15,21]. A majority of the members show general excitatory symptoms. But at the same time, it has been shown that two members, Q7Z096 and Q7M4K5, exhibit hyperactivity, circular motion and convulsion. Another member, Q7Z091, has been reported to cause paralysis and death in mice. It has been suggested that this class of conotoxins might be a promising source of pharmacological tools to explore the molecular components that help in axon excitability [21].

The I_2 Superfamily

The reference sequence pattern for group C (I_2 superfamily) was able to pick all the UniProtKB/Swiss-Prot, release 49.7, entries correctly, but when scanned against other databases like UniProtKB/TrEMBL (release 32.7), it picked 9 correct proteins and 15 unrelated proteins (two from *Homo sapiens*, eight hypothetical proteins from *C. elegans* and five hypothetical proteins from *Caenarhabditis briggsae*). This is because the reference sequence patterns with X-positions may pick any residue including cysteine residues, which are not allowed for the I_2 superfamily sequences. Hence, such X-positions were replaced by {C}, which included all residues other than cysteine. We redefined the sequence pattern for the I_2 superfamily as C-{C}(6)-C-{C}(5)-C-C-{C}(3)-C-C-{C}(2, 3)-C-{C}(3)-C. Our pattern was scanned against the UniProtKB/Swiss-Prot database, which picked up ten true positive hits. The profile HMM created for I_2 superfamily was also used for searching the same database, which resulted in top ten hits with an E-value between 10^{-15} and 10^{-12}. The E-value for the remaining hits was greater than 0.8. Thus both the methods performed well yielding only true positives. The sensitivity, selectivity, specificity and correlation coefficient were thus unity. The pattern and profile defined by us was scanned against UniProtKB/TrEMBL, which also gave only the correct hits.

The percentage of positively charged residues, especially Arg, was found to be more for the I_2 superfamily members (Table 2). The mature toxin sequences from *C. imperialis*, (P69495, P69496 and P69497) have positively charged residues Lys and Arg, while in the remaining sequences, the Lys residue was absent. Generally, such mature toxin peptides [22] have been shown to inhibit potassium channels and the positively charged residues play an important role in such interactions. Out of the ten entries of I_2 superfamily conotoxins (Table 1), nine entries inhibit the vertebrate potassium channels Kv1.1 and Kv1.3 but not Kv1.2, [23] and the remaining one entry Q9U3Z3 acts as a modulator of potassium channel [24]. At least two different mechanisms have been proposed for blocking the voltage-gated potassium channels by the inhibitors. One mechanism involves the positively charged lysine and the hydrophobic residue (Tyr or Phe) that acts as

DOI: 10.1002/psc

a dyad, which occludes the potassium channel pore [25,26]. The second mechanism involves clustering of positively charged residues on one surface and it acts as an anchor blocking the pore as a lid [27].

The *in silico* docking studies carried out by our group and reported earlier [6] on a theoretical 3D model of ViTx, Q7YZS9, from *Conus virgo*, reveals the crucial role of the residues Arg 26 and Arg 32 present in the *C*-terminal region as being important for interacting with vertebrate Kv1.1 channels. The entries P69498 and P69501 are isotoxins from *Conus vexillum* and *Conus miles*, which belong to the same clade and differ in only a single amino acid residue from ViTx. Thus their mode of action may be similar to that of ViTx [5]. The entry P69500 having 97% sequence identity with ViTx may also exhibit a similar mode of action. The *in silico* docking studies suggest that ViTx [6] may block by means of the second type of mechanism as observed in κ-M conotoxins RIIIK which complexes with TSha1 K^+ channel [27]. Conus peptides from other superfamilies have also been reported to inhibit the potassium channel, as shown in Table 3. The I_2 superfamily members also exhibit functional similarity with Parabutoxins, (Q6WGI9, P60164, P60165) which belong to a subfamily of acidic α-K^+ toxins from *Parabuthus* scorpion species [28].

Table 2 Amino acid composition for the 20 amino acid residues in the non-redundant set of I_1 and I_2 superfamilies

Category	Residue	Composition (%)	
		I_1 superfamily	I_2 superfamily
Aliphatic	Ala	3.1	0.4
	Gly	10.6	9.7
	Ile	2.7	1.7
	Leu	3.5	4.2
	Pro	7.1	6.8
	Val	3.9	2.1
Aromatic	Phe	4.3	2.1
	Tyr	1.6	5.1
	Trp	2.4	3.0
Negative charged	Asp	3.5	3.4
	Glu	1.6	3.0
Positive charged	Arg	3.9	12.7
	His	3.5	3.8
	Lys	6.7	2.5
Polar	Asn	3.9	3.4
	Gln	1.6	1.3
	Ser	9.4	7.6
	Thr	7.1	3.4
Sulphur containing	Cys	18.8	23.6
	Met	0.8	0.4

A Framework for *in Silico* Annotation of I_1 and I_2 Superfamilies

The reference sequence pattern for group A was scanned against UniProtKB/TrEMBL release 32.7, which picked eight entries from *C. elegans*, four from *C. briggsae* and one entry from *Babesia bovis and H. sapiens*. The reference pattern for group B picked one entry from *H. sapiens*. But the sequence pattern and profile HMM defined by us for I_1 superfamily did not pick any false entry.

The defined sequence pattern and profile HMM (E-value between 10^{-14} and 10^{-5}) for I_2 was able to pick nine UniProtKB/TrEMBL entries – Q59AA2, Q59AA3, Q59AA4, Q59AA5, Q59AA6, Q59AA7, Q59AA8, Q59AA9 and Q514E5. This observation is supported by the entries present in IPR013141. Out of the nine entries, seven entries, Q59AA2, Q59AA3, Q59AA5, Q59AA6, Q59AA7, Q59AA8 and Q59AA9, are same as in UniProtKB/Swiss-Prot, which are P69498, P69501, P69499, P69494, P69495, P69496 and P69497, respectively.

Of the remaining two entries, for one entry, Q514E5, the information about the signal peptide region (1–25), mature toxin region (26–56) is available. The *C*-terminal region of this entry is similar to the UniProtKB/Swiss-Prot entry Q9U3Z3, a modulator of potassium channel, as shown in Figure 4(a). We have predicted the prepro-region (60–69) of this entry to be AKLLEFFRQR. Since the prepro-regions are similar, one may expect that the post-translational modifications in the mature toxins may be similar to that of a potassium channel modulator like Q9U3Z3 [24]. But from Figure 4(a), it is seen that only the cysteine residues are aligned in the mature toxin sequence. Consequently, the mature toxin of Q514E5 alone is checked for similarity using GAP and the best alignment is obtained with P69496

Table 3 Different superfamilies of conotoxins acting on voltage-gated potassium channels

Superfamily (family)	Cys arrangement	Example	Function	Reference
A (αA)	CC-C-C-C-C	SIVA	Binds and inhibits voltage-sensitive K^+ channels	29
M (κM)	CC-C-C-CC	RIIIK	Binds and inhibits reversibly Shaker K^+ channels	30
O (κ)	C-C-CC-C-C	PVIIA	Binds and inhibits voltage-sensitive K^+ channels	31
I_2	C-C-CC-CC-C-C	ViTx	Inhibits the vertebrate K^+ channels Kv1.1 and Kv1.3, but not Kv1.2	23

DOI: 10.1002/psc

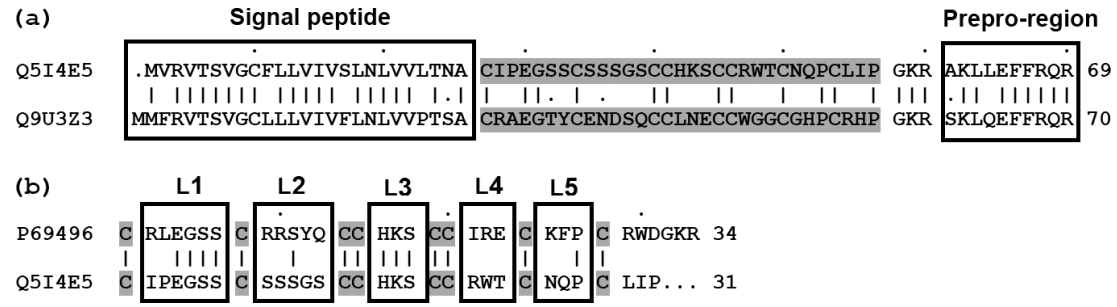

Figure 4 (a) Full-length sequence alignment between Q5I4E5 and Q9U3Z3, a modulator of potassium channel. The signal peptide and prepro-regions are shown in boxes, which are apparently well conserved compared to the highlighted mature toxin region. (b) Alignment between Q5I4E5 and P69496 mature toxin: Mature toxin of Q5I4E5 shows more similarity to P69496, which is an inhibitor of potassium channel.

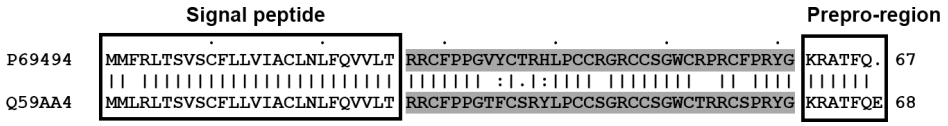

Figure 5 Full-length sequence alignment reflecting the identity between Q59AA4 and P69494. Signal peptide and prepro-region are shown in boxes and the mature toxin region is highlighted. The signal peptide region of Q59AA4 predicted by the SignalP server aligns with that of P69494.

(sequence identity 54.8%) (Figure 4(b)), which is a vertebrate potassium channel inhibitor. In order to clarify the functional mechanism, i.e. whether the mature toxin would act as a modulator or inhibitor, further experimental investigations are required.

Our work provides a framework for annotation of uncharacterized sequences as illustrated below: Firstly, we detected the three important regions in the sequence, namely signal peptide, prepro-region and mature toxin. The signal peptide was used for categorization and the mature toxin region for predicting the probable functional characterization based on similarity. For the purpose of functional annotation of Q59AA4, the SignalP server [32] was used for predicting the signal peptide region. The server gave the same cleavage site by Neural Networks as well as HMM methods, and thus the signal peptide was predicted from 1–26 confirming that the entry belonged to the I_2 superfamily. A pairwise sequence comparison using GAP was done for the entire sequence with I_2 superfamily members. On the basis of the highest identity obtained with the P69494, (86.6% sequence identity) a member of I_2, we were able to predict the mature toxin (27–61) and prepro-regions (62–68) for this entry (Figure 5). In addition to this, on the basis of functional similarity, it is suggested that the entry might inhibit vertebrate potassium channels. It is also noted that the entry does not have a single Lys residue, in contrast with a number of Arg residues distributed along the length of the sequence.

From the results of our analysis, we have been able to show the utility of such an *in silico* functional annotation, which proved to be valuable in predicting the characteristics of one UniProtKB/TrEMBL entry (Q59AA4). The approach can be applied for the functional annotation of new entries and the predictions can be concomitantly evaluated through experimental works.

CONCLUSIONS

Our results reiterate the existence of two distinct gene superfamilies I_1 and I_2, which are presently described as I-conotoxin superfamily in the UniProtKB/Swiss-Prot database. We have defined selective and sensitive sequence patterns and profile HMM for the two gene superfamilies, which would be useful in protein classification and functional annotation. The I_1 superfamily conotoxins could provide valuable information about various molecular factors that are involved in excitable properties of axons. The majority of the I_2 superfamily members are actively involved in inhibiting the therapeutically important vertebrate potassium channels. They also show functional similarity with not only the superfamilies of the *Conus* species but also with some others like scorpion species. There is a need to understand the degree of relatedness within and between species that have same targets. Moreover, investigations are required to explain the cysteine connectivity and elucidate the appropriate 3D structure in order to appreciate the role of specific amino acids that are involved in interacting with their respective targets. The unique features, sequence patterns and functions of the two superfamilies provide scope for further examination in order to explore their pharmacological properties.

DOI: 10.1002/psc

Acknowledgements

Facilities at the Bioinformatics Centre of Excellence funded by the Department of Biotechnology (DBT), India, were used and are gratefully acknowledged. S. Ramakumar thanks International Business Machines (IBM) for a CAS fellowship grant IBMC002. We are grateful for the constructive comments offered during the review process.

REFERENCES

1. Jones RM, Bulaj G. Conotoxins–new vistas for peptide therapeutics. *Curr. Pharm. Des.* 2000; **6**: 1249–1285.
2. Rajendra W, Armugam A, Jeyaseelan K. Toxins in anti-nociception and anti-inflammation. *Toxicon* 2004; **44**: 1–17.
3. McIntosh JM, Jones RM. Cone venom–from accidental stings to deliberate injection. *Toxicon* 2001; **39**: 1447–1451.
4. Terlau H, Olivera BM. Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* 2004; **84**: 41–68.
5. Kauferstein S, Huys I, Kuch U, Melaun C, Tytgat J, Mebs D. Novel conopeptides of the I-superfamily occur in several clades of cone snails. *Toxicon* 2004; **44**: 539–548.
6. Mondal S, Vijayan R, Shichina K, Babu RM, Ramakumar S. I-superfamily conotoxins: sequence and structure analysis. *In Silico Biol.* 2005; **5**: 557–571.
7. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJA. The PROSITE database. *Nucleic Acids Res.* 2006; **34**: D227–D230.
8. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McDowall J, Mitchell A, Nikolskaya AN, Orchard S, Pagni M, Ponting CP, Quevillon E, Selengut J, Sigrist CJ, Silventoinen V, Studholme DJ, Vaughan R, Wu CH. InterPro, progress and status in 2005. *Nucleic Acids Res.* 2005; **33**: D201–D205.
9. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: Juggling between evolution and stability. *Brief. Bioinform.* 2004; **5**: 39–55.
10. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 2006; **34**: D187–D191.
11. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon M, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res.* 2004; **32**: D138–D141.
12. Buczek O, Yoshikami D, Watkins M, Bulaj G, Jimenez EC, Olivera BM. Characterization of D-amino-acid-containing excitatory conotoxins and redefinition of the I-conotoxin superfamily. *FEBS J.* 2005; **272**: 4178–4188.
13. Nicke A, Loughnan ML, Millard EL, Alewood PF, Adams DJ, Daly NL, Craik DJ, Lewis RJ. Isolation, structure, and activity of GID, a novel alpha 4/7-conotoxin with an extended N-terminal sequence. *J. Biol. Chem.* 2003; **278**: 3137–3144.
14. Balaji RA, Ohtake A, Sato K, Gopalakrishnakone P, Kini RM, Seow KT, Bay BH. λ-conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. Isolation and characterization from the venom of Conus marmoreus. *J. Biol. Chem.* 2000; **275**: 39516–39522.
15. Buczek O, Yoshikami D, Bulaj G, Jimenez EC, Olivera BM. Post-translational amino acid isomerization: a functionally important D-amino acid in an excitatory peptide. *J. Biol. Chem.* 2005; **280**: 4247–4253.
16. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 2001; **17**: 282–283.
17. GCG. *Wisconsin Package Version 10.3*. Accelrys Inc: San Diego, CA, 1999.
18. Higgins D, Thompson J, Gibson T, 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.
19. Gattiker A, Gasteiger E, Bairoch A. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics* 2002; **1**: 107–108.
20. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998; **14**: 755–763.
21. Jimenez EC, Shetty RP, Lirazan M, Rivier J, Walker C, Abogadie FC, Yoshikami D, Cruz LJ, Olivera BM. Novel excitatory Conus peptides define a new conotoxin superfamily. *J. Neurochem.* 2003; **85**: 610–621.
22. Huang X, Dong F, Zhou HX. Electrostatic recognition and induced fit in the κ-PVIIA toxin binding to shaker potassium channel. *J. Am. Chem. Soc.* 2005; **127**: 6836–6849.
23. Kauferstein S, Huys I, Lamthanh H, Stocklin R, Sotto F, Menez A, Tytgat J, Mebs D. A novel conotoxin inhibiting vertebrate voltage-sensitive potassium channels. *Toxicon* 2003; **42**: 43–52.
24. Fan CX, Chen XK, Zhang C, Wang LX, Duan KL, He LL, Cao Y, Liu SY, Zhong MN, Ulens C, Tytgat J, Chen JS, Chi CW, Zhou Z. A novel conotoxin from Conus betulinus, kappa-BtX, unique in cysteine pattern and in function as a specific BK channel modulator. *J. Biol. Chem.* 2003; **278**: 12624–12633.
25. 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.
26. Mouhat S, De Waard M, Sabatier JM. Contribution of functional dyad of animal toxins acting on voltage-gated K_v 1-type channels. *J. Pept. Sci.* 2005; **2**: 65–68.
27. Verdier L, Al-Sabi A, Rivier JE, Olivera BM, Terlau H, Carlomagno T. Identification of a novel pharmacophore for peptide toxins interacting with K^+ channels. *J. Biol. Chem.* 2005; **280**: 21246–21255.
28. Huys I, Olamendi-Portugal T, Garcia-Gomez BI, Vandenberghe I, Van Beeumen J, Dyason K, Clynen E, Zhu S, van der Walt J, Possani LD, Tytgat J. A subfamily of acidic alpha-K(+) toxins. *J. Biol. Chem.* 2004; **279**: 2781–2789.
29. Wang CZ, Jiang H, Ou ZL, Chen JS, Chi CW. cDNA cloning of two A-superfamily conotoxins from Conus striatus. *Toxicon* 2003; **42**: 613–619.
30. Ferber M, Sporning A, Jeserich G, DeLaCruz R, Watkins M, Olivera BM, Terlau H. A novel conus peptide ligand for K+ channels. *J. Biol. Chem.* 2003; **278**: 2177–2183.
31. Naranjo D. Inhibition of single Shaker K channels by kappa-conotoxin-PVIIA. *Biophys. J.* 2002; **82**: 3003–3011.
32. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004; **340**: 783–795.