I-superfamily conotoxins: sequence and structure analysis

S. Ramakumar
Department of Physics and Bioinformatics Centre, Indian Institute of Science, Bangalore
Talk plan

• Overview of Cone snails and conotoxins
• Sequence patterns of Conotoxins
• Theoretical structural 3D model of ViTx from *Conus virgo*
• Interaction studies of ViTx with the vertebrate K+ channels
Cone snails

- the cone snails (genus Conus) venomous marine molluscs
- the most successful living marine animals alive today
- can grow up to ~25 cm in length
- about 500 different species of predatory cone snails
Conus in human history

Shells of cone snails and their use in early human cultures

Cone snails

Classification

Class: Gastropoda
Subclass: Prosobranchia
Order: Caenogastropoda
Superfamily: Conacea
Family: Conidae

Major Genera
- Genus: Conus
- Genus: Asprella
- Genus: Chelyconus
- Genus: Floraconus
- Genus: Leptoconus

They are classified according to their prey:

Piscivorous ........ fish eaters
  * C.striatus
  * C.geographus
  * C.magnus

Molluscivorous .... mollusk eaters
  * C.textile
  * C.marmoreus
  * C.pennaceus

Vermivorous ....... worm eaters
  * C.imperialis
  * C.eburneus
  * C.quercinus
Cone snails: hunters using poisoned arrows

- Harpoon
- Tube
- Tongue
- Throat
Conus purpurascens rapidly catching a clown fish
Conus – a sign of danger

- ~30 human deaths have been recorded from cone snail envenomation
- *Conus geographus*, is known colloquially as the "cigarette snail"
- symptoms of a cone snail stings include intense pain, swelling, numbness and tingling
- severe cases involved muscle paralysis, changes in vision and respiratory failure

Most DEADLY species
*Conus geographus*
- the organization involves the precursor region, pro-peptide region and mature region

- **precursor region** contains highly conserved signal sequences

- **pro-peptide region** contains potential anchor binding sites for post-translational modification enzymes

- **mature region** contains functional toxin
Ion channels and receptors

- **voltage-gated ion channels** are membrane-bound proteins activated by change in transmembrane voltage

- they are **multi-subunit** complexes with circular arrangement of identical or different proteins forming a pore region

- it conducts specific species of ions such as $\text{Na}^+$, $\text{K}^+$, $\text{Ca}^{2+}$

- ion channels are linked with various **neurodegenerative disorders**

- conotoxins work on this channels by plugging ion pores and becomes main target for **therapeutically important channels**

- **ligand gated ion channels**: Nicotinic acetylcholine receptors
Block of $K^+$-and prolonged opening of $Na^+$-channels

Conus purpurascens - Toxin

$k$-conotoxin PVI A

$\delta$-conotoxin PVI A

extracellular

membrane

intracellular

potassium channel

sodium channel

$K^+$- efflux: control, with toxin

$Na^+$- influx: control, with toxin
Classification of Conus venoms

CONOPEPTIDES

Non-Disulfide-Rich

No S-S
- Contulakin
- Conantokin
- Conorhamide

Single S-S
- Conopressin
- Contryphan

Disulfide-Rich

Superfamily

O
- VII/VIII
  - C-C-CC-C-C

M
- III
  - CC-C-C-CC

A
- J/II
  - CC-C-C
- IV
  - CC-C-C-C

S
- VIII
  - C-C-C-C-C-C-C

T
- V
  - CC-CC
- X
  - CC-CCPC

P
- IX
  - C-C-C-C-C-C-C
- I
  - C-C-C-C-C-C-C-C

Framework # S-S motif

General Target

Conotoxin Family
- δ
- μO
- κ
- ω
- μ
- ψ
- kM
- α
- αA
- κA
- σ
- ?
- NE
- Transporter
- ?
- χ
- K Channels
Applications of conotoxins

- **Ziconotide** (Elan Pharmaceuticals) - synthetic ω-conotoxin MVIIA, acts by blocking N-type Calcium channels involved in pain pathways

- **α conotoxins** are antagonists of Nicotinic acetylcholine receptors which are involved in Alzheimer's disease, Parkinson's disease

- **ω conotoxins** are widely used in the treatment of chronic pain, spinal chord injury

- **κ conotoxins** are blockers of K+ channels which are linked with hypertension and epilepsy

- **μ conotoxins** are blockers of TTX-R Na+ channels which are involved in pain
Ziconotide, the first novel drug from the sea

- isolated from the cone snail, Conus magus
- ω-conotoxin MVIIA (SNX-111)
- recently approved as a new analgesic (pain killer)
- trade name Prialt®
- mechanism of action is different to that of morphine
A
Spinal chord
Nociceptor

B
Nociceptor
Morphin
Ca^{2+}
ω-Conotoxin

Spinal Chord cell

Skins
Muscle
Organ

Nerve cell
Synapse
Axon
Dendrite
Venominformatics for conotoxins

- therapeutic applications of conotoxins
- sequence and structural analysis
- interaction studies between conotoxins and targets
- output helpful for protein engineering
Swiss-Prot; a **curated protein sequence database** which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, PTMs, variants, etc.), a minimal level of redundancy and high level of integration with other databases
What are patterns?

Patterns are regular expressions matching short sequence motifs usually of biological meaning, generally:

a) Enzyme catalytic sites,
b) Prosthetic group attachments sites (heme, pyridoxal-phosphate, etc.),
c) Amino acids involved in binding a metal ion,
d) Cysteines involved in disulfide bonds,
e) Regions involved in binding a molecule or another protein.

Applications of the pattern:
- **Classification**: discriminating between family members and non-members
- **Functional annotation**: describe biologically important features
The three principal methods for building pattern databases

Single motif methods

- fuzzy regular expression (IDENTIFY)
- exact regular expression (PROSITE)

C-Y-X(2)-[DG]-G-X-[ST] regular expression

C-yedggiS cyeeggit cyhgdggs cyrgdgs

Multiple motifs methods

- identity matrices (PRINTS)
- weight matrices (BLOCKS)

Full domain alignment methods

- profiles (PROSITE LIBRARY)
- HMMs (PFAM)
## Information about a few pattern databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Information</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROSITE</td>
<td>Regular expressions and profiles</td>
<td><a href="http://www.expasy.org/prosite">http://www.expasy.org/prosite</a></td>
</tr>
<tr>
<td>BLOCKS</td>
<td>Blocks (ungapped local alignments)</td>
<td><a href="http://www.blocks.fhrc.org/">http://www.blocks.fhrc.org/</a></td>
</tr>
<tr>
<td>Prints</td>
<td>Blocks</td>
<td><a href="http://umber.sbs.man.ac.uk/dbbrowser/PRINTS/">http://umber.sbs.man.ac.uk/dbbrowser/PRINTS/</a></td>
</tr>
<tr>
<td>Identify</td>
<td>Regular expressions</td>
<td><a href="http://motif.stanford.edu/identify/">http://motif.stanford.edu/identify/</a></td>
</tr>
<tr>
<td>Pfam</td>
<td>HMMs</td>
<td><a href="http://sanger.ac.uk/Pfam">http://sanger.ac.uk/Pfam</a></td>
</tr>
<tr>
<td>Interpro</td>
<td>protein families, domains</td>
<td><a href="http://www.ebi.ac.uk/interpro/">http://www.ebi.ac.uk/interpro/</a></td>
</tr>
</tbody>
</table>
Database of protein families and domains

Release 19.19, of 24-Jan-2006 (contains 1400 documentation entries that describe 1329 patterns, 4 rules and 593 profiles/matrices)
PROSITE syntax

- The standard IUPAC one-letter codes.
- `x` : any amino acid.
- `[ ]` : residues allowed at the position.
- `{ }` : residues forbidden at the position.
- `( )` : repetition of a pattern element are indicated in parenthesis. X(n) or X(n,m) to indicate the number or range of repetition.
- `-` : separates each pattern element.
- `<` : indicated a N-terminal restriction of the pattern.
- `>` : indicated a C-terminal restriction of the pattern.
- `.` : the period ends the pattern.

PS60019 I_CONOTOXIN
C-{C}(6)-C-{C}(5)-C-C-{C}(1,3)-C-C-{C}(2,4)-C-{C}(3,10)-C
How we develop Prosite patterns!

ScanProsite

The ScanProsite tool (Help) allows to scan protein sequences (either from SWISS-PROT or TRAFOS) or provided by the user) for occurrences of patterns, profiles and motif (PROSITE) entries in the PROSITE database, or to derive protein databases for hit by specific motifs (Help). Download pc.exe from the developer's website. The program PCPRO can be used to generate your own patterns. You may either:

- Enter one or more PROSITE accessions (e.g., tradi002) to search the SWISS-PROT/EMBL and/or PDB databases, OR
- Enter one or more sequences (e.g., tradi002 or data file) and/or their PROSITE accession numbers (e.g., tradi002) to be scanned with all patterns, profiles, and motif (PROSITE, OR
- Fill the right fields to find all occurrences of a motif in a sequence.

Prosite patterns entry form (Help)

ScanProsite tool entry form (Help)
Pattern assessment

sensitivity = \frac{TP}{TP+FN}

specificity = \frac{TN}{TN+FP}

accuracy = \frac{TP+TN}{TP+TN+FP+FN}

MCC = \frac{TP \times TN - FP \times FN}{\sqrt{[(TP+FN)(TN+FP)(TP+FP)(TN+FN)]}}
The core of the PROSITE database is composed of two text files: **PROSITE .DAT** and **PROSITE .DOC**.
Cone snail toxins, conotoxins, are small peptides with disulfide connectivity, that target ion-channels or G-protein coupled receptors. Based on the number and pattern of disulfide bonds and biological activities, conotoxins can be classified into several families [1]. Omega, delta and kappa families of conotoxins have a knottin or inhibitor cystine knot scaffold. The knottin scaffold is a very special disulfide through disulfide knot, in which the III-VI disulfide bond crosses the macrocycle formed by two other disulfide bonds (I-IV and II-V) and the interconnecting backbone segments, where I-VI indicates the six cysteine residues starting from the N-terminus [2,El].

-Expert(s) to contact by email:
  Ramakumar S.; ramak@physics.iisc.ernet.in
Online servers for automated pattern detection

**TEIRESIAS**
- Generate patterns from a collection of unaligned protein or DNA sequences; at IBM.

**PRATT**
[http://www.ebi.ac.uk/pratt/](http://www.ebi.ac.uk/pratt/)
- Interactively generates conserved patterns from a series of unaligned proteins.
Contributions in prosite from Prof. S. Ramakumar's group

**Conotoxin families signatures**

**PDOC60004** Conotoxin families signatures

- PS60004; **OMEGA_CONOTOXIN**, Omega-conotoxin family signature.
- PS60005; **DELTA_CONOTOXIN**, Delta-conotoxin family signature.
- PS60013; **MU_CONOTOXIN**, Mu-conotoxin family signature.
- PS60014; **ALPHA_CONOTOXIN**, Alpha-conotoxin family signature.
- PS60019; **I_CONOTOXIN**, I-superfamily conotoxin signature.

**PDOC60025** PS60025; **CONANTOKIN**, Conantokin family signature.

**Spider toxins**

- PS60018; **DELTA_ACTX**
- PS60016; **OMEGA_ACTX_1**
- PS60021; **HWTX_1**
- PS60015; **MU_AGATOXIN**
- PS60026; **ERGTX**

- PS60020; **J_ACTX**
- PS60017; **OMEGA_ACTX_2**
- PS60022; **HWTX_2**
- PS60023; **OMEGA_AGA_II_III**

**Many more on the way ...**
Growth of Biological Databases

is an integrated documentation resource for protein families, domains and sites

dynamic controlled vocabulary of Biology
Our contribution in dynamic controlled vocabulary of Biology

**PDOC60004** **PS60005** **DELTA_CONOTOXIN**

**IPR012322** **Delta-conotoxin**

<table>
<thead>
<tr>
<th>Name</th>
<th>GO ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>process</strong></td>
<td></td>
</tr>
<tr>
<td>pathogenesis</td>
<td>GO:0009405</td>
</tr>
<tr>
<td><strong>function</strong></td>
<td></td>
</tr>
<tr>
<td>sodium channel inhibitor activity</td>
<td>GO:0019871</td>
</tr>
<tr>
<td><strong>component</strong></td>
<td></td>
</tr>
<tr>
<td>extracellular region</td>
<td>GO:0005576</td>
</tr>
</tbody>
</table>
I-superfamily conotoxins

- recently characterized and explored comparatively less

ViTx (Q7YZS9) is isolated from venom of *Conus virgo*

- comprises a single chain of 35 aa linked by four SS bridges

- inhibits the vertebrate potassium channel $K_v1.1$ and $K_v1.3$ and not $K_v1.2$
SDPMOD – Modeling Tool for SDPs

- provides an automated comparative modeling of Small SS-bonded Proteins (SDPs)

SDPMOD offers 3 different modes:

- fully automated
- semi-automated

allows users to select the template from a suggested list from server

- manual

allows for the selection of the desired template

This server was used for modeling the structure of ViTx
ViTx-theoretical 3D model

Template 1DLO (J-AcTx-Hv1c) from Blue Mountains funnel-web spider *Hadronyche versuta*
Comparisons of ViTx with other toxins (acting on Shaker related Potassium channels)
Potassium channels

- $K^+$ channels are the most diverse group of the ion channel family
- the $K_{v1.1}$ subfamily is expressed in the embryonic nervous system, brain
- the $K_v$ family can be divided into 4 subfamilies on the basis of sequence similarity and function:
  Shaker ($K_v1$), Shab ($K_v2$), Shaw ($K_v3$) and Shal ($K_v4$)
- contains four identical subunits forming a central pore
- point mutations in $K_{v1.1}$ result in episodic ataxia type I, a rare autosomal dominant neurological disorder
Modeling of $K_v1.1$ potassium channel

- template KcsA from *Steptomyces lividans* (1BL8)
- the sequence identity between KcsA and $K_v1.1$ is 20%
- the Homology module in InsightII was used for modeling process
- the stereochemistry of model was checked by GROMACS
- the quality of the model was checked by PROCHECK

KcsA-Potassium channel (1BL8)  Modeled $K_v1.1$
Docking of $K_v1.1$ with ViTx

- the interaction between $K_v1.1$ and ViTx was studied using **Affinity module** of InsightII
- ViTx perfectly blocked the selectivity filter and ion passage is restricted
- **electrostatic interactions** play a crucial role in interaction
Energetically favourable 2 different orientations of the vertebrate channel blocker ViTx docked with P-loop region of Kv1.1 from Human

C-terminal region of ViTx in blocking therapeutically important voltage-gated potassium channels
Conclusion

• Sequence patterns useful for Classification & functional annotations

• Interaction studies with pharmacological target helpful for mutagenesis study
I-superfamily conotoxins: sequence and structure analysis

Sukanta Mondal, Ramachandran Vijayan, Kannambath Shichina, Rajasekaran Mohan Babu and Suryanarayananarao Ramakumar

In Silico Biology (2005) 5(4):0050
Contributors

Research Scholar
Sukanta Mondal

Project Assistants
Mohan Babu Rajasekaran
Bhavna Rajasekaran

Project students
Shichina
Vijayan
Vinod
Acknowledgements

The facilities at the Bioinformatics Centre funded by the Department of Biotechnology (DBT), India, were utilized and are gratefully acknowledged.

S. Ramakumar thanks International Business Machines (IBM) for a CAS fellowship grant.
Thank you