Computational Biology and Chemistry 61 (2016) 8–14

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Analysis of molecular structures and mechanisms for toxins derived from venomous animals

L.F.O. Rocha*

Department of Physics and Chemistry, Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo. Av. do café, s/n, 14040-903, Ribeirão Preto, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 9 November 2014
Received in revised form 22 September 2015
Accepted 10 November 2015
Available online 2 December 2015

Keywords:
Ion channel
Molecular sequence data
Secondary structure composition
Sequence-structure-function relationship
Steric and hydrophobic interaction
Toxic protein

ABSTRACT

As predominant component in the venom of many dangerous animal species, toxins have been thoroughly investigated for drug design or as pharmacologic tools. The present study demonstrated the use of size and hydrophobicity of amino acid residues for the purposes of quantifying the valuable sequence–structure relationship and performing further analysis of interactional mechanisms in secondary structure elements (SSEs) for toxin native conformations. First, we showed that the presence of large and hydrophobic residues varying in availability in the primary sequences correspondingly affects the amount of these residues being used in the SSEs in accordance with linear behavioral patterns from empirical assessments of experimentally derived toxins and non-toxins. Subsequent derivation of prediction rules was established with the aim of analyzing molecular structures and mechanisms by means of 114 residue compositions for venom toxins. The obtained results concerning the linear behavioral patterns demonstrated the nature of the information transfer occurring from the primary to secondary structures. A dual action mechanism was established, taking into account steric and hydrophobic interactions. Finally, a new residue composition prediction method for SSEs of toxins was suggested.

1. Introduction

Toxins are proteins that predominantly constitute the venom of dangerous animals, such as snakes, scorpions, spiders, cone snails, sea anemones, insects and fish (Rash and Hodgson, 2002). Venom is strategically used when performing predation and defense, and often results in high morbidity and mortality rates, which have been reported as a serious hazard to public health in many countries (Ushanandini et al., 2006). Therefore, toxins have been actively investigated in basic research (Nagaraju et al., 2006; Whetstone and Hammock, 2007) as pharmacological and diagnostic tools (Harvey, 2002), templates for drug design, and therapeutic agent candidates, thereby transforming toxicity into profit. Animal toxins vary exceedingly in terms of their ability to endamage preys and threats due to diversified biological effects arising from the actions of venom (Cologna et al., 2009; Dokmetjian et al., 2009). These effects can include hemorrhagic, myotoxic, neurotoxic, hemolytic and proteolytic activities. For example, neurotoxic activity drastically influences neural transmission, and significantly alters muscular contraction and relaxation, respiration and cardiac function, which can have a potentially fatal effect on a prey or threat.

Multiple covalent disulfide bridges (Kong et al., 2004) formed by means of sequential distant cysteine residues (Matsunaga et al., 2009) strengthen native-state structures in toxins (or toxic proteins) (Redfern et al., 2008) as well as diminish macromolecule susceptibility to enzymatic digestion (Nayak et al., 1999), so that only a few toxic proteins without such bridges have been isolated and characterized in some breeds, e.g., scorpions (Zeng et al., 2005). In terms of function, toxins work as cell modulators through recognition and selective binding to ion channels (mainly calcium, chloride, potassium, and sodium) and membrane receptors (Restrepo-Angulo et al., 2010). In the case of ion channels (Gabashvili et al., 2007), toxins target these channels by blocking them, enhancing or decreasing their opening, preventing or slowing their inactivation, and modifying gating (electrophysiology). Thus, toxins affect various kinds of cells, alter fundamental physiological processes, and subsequently inflict severe biological effects (Cologna et al., 2009; Dokmetjian et al., 2009).

Functional native configurations of toxins and non-toxins (or other proteins that are not labeled as toxins) have secondary structure elements (SSEs) as underlying constituents, mainly α-helices and β-pleated sheets formed by strands. They are commonly divided into four classes: mostly α; mostly β; mixed α

* Fax: +55 16 3602 4880.
E-mail address: luiz@fcfrp.usp.br (L.F.O. Rocha).

http://dx.doi.org/10.1016/j.compbiolchem.2015.11.004
1476-9271/

and β; and few secondary structures (Orengo et al., 1997). The process for predicting SSEs from the primary sequence isvery challenging. It has been thoroughly investigated using different approaches and specific details regarding the molecular forces stabilizing secondary structures and, significantly improving the process. However, the performance of prediction methods is still dependent on some peculiar features of the water-soluble and transmembrane proteins (Majorek et al., 2009). Another issue is related to secondary structural classes, where better predictive analyses usually occur for α classes (Gromiha and Selvaraj, 2004). Therefore, the greatest challenge should be to improve the prediction accuracy for β and mixed classes (Midic et al., 2007; Best and Mittal, 2010) of the interspersed α/β or segregated α + β forms (Lindström et al., 2009).

Size (steric hindrance) and hydrophobicity (hydrophobic clustering) are considered to be universal and common attributes of residues in toxins and non-toxins. These two attributes have long been recognized as the indispensable and primary determinants of the native folded state (Chothia, 1984; Ramachandran and Sasisekharan, 1968), and both were examined here. Excluded volume or steric interactions comprise the main factor involved in biomolecular structural organization (Richards, 1977), SSE packing (Jiang et al., 2003) and molecular interplays (Halperin et al., 2002). Computational simulations (Rocha et al., 2009), statistical analyses of known macromolecular conformations (Miyazawa and Jernigan, 1996) and structure prediction methods (Zhang et al., 1998) point to the fact that hydrophobic and hydrophilic interactions are very important and represent valuable driving forces required to maintain biological folding and stability.

In the present study, first we demonstrated five linear relations between the availabilities of large and hydrophobic residues by means of primary structures and how these residues can be employed by SSEs in the functional conformations of small toxins and non-toxins. Next, we performed an analysis of molecular structures and mechanisms in toxins through direct comparison of real and predicted residue compositions, which we surmised would reveal complementary (or predominant) contributions of residue size and hydrophobicity in 80% (or 20%) of SSEs. In order to carry out proper analysis a purely empirical approach was used, taking into account sequence and structure data extracted from experimentally determined macromolecules. A mathematical formulation utilizing representative measures of primary, secondary and tertiary structures, as well as computational algorithms allowing automatic processing of our sample choice, empirical approach and mathematical formulation, were employed during the investigation.

2. Materials and methods

2.1. Macromolecules of the benchmark dataset

The native macromolecules in our dataset were obtained from a special set containing toxins and non-toxins. However, many elusive and redundant macromolecules are already deposited in the Protein Data Bank (PDB) (Berman et al., 2000). Thus, we defined the following criteria and restrictions: primary structure consisting solely of natural amino acids (among the 20 genetically encoded types); complete information provided regarding whole residue sequences and secondary structures; and use of unrelated proteins having either different residue sequences (pairwise sequence identity below 25%) or SSEs with at least four different residues.

Several chain extensions were initially investigated, taking into account the above restraint conditions, which resulted in choosing the 35-residue proteins for analysis. This extension allowed maximization of the quantity of exemplars; of 39 proteins, 8 were toxins and 31 were non-toxins (for more details about these toxins, see Table S1a in Supplementary materials). Furthermore, our dataset also included all the toxins with 36–40 residues that are currently available in the PDB archives and under the conditions above, totaling 32 toxins.

In addition to being of remarkably diverse origins, our 71 unrelated toxin and non-toxin chains exhibited diverse biological roles and were present in several subcellular locations, as well as being involved in many biochemical processes. For instance, in humans the 35-residue non-toxins 1PB5, 1ZWD, 2ODB, 2OV2 were classified as signaling, hormone, binding and transferase, respectively.

2.2. Simplified alphabets for residue sizes and hydrophobicities

We selected and used various proteins with diversified sequences and structures that were in compliance with the selection criteria above, since the present research was based on two omnipresent and common attributes of residues, size or volume and hydrophobicity, which operate in various environmental circumstances and sustain specialized biological functions. Twenty natural amino acids were simplified according to their sizes (and hydrophobicities) by means of reduced binary codification, denoted as Large/Small, or LS (Zamyatnin, 1984), and Hydrophobic/Polar was denoted as HP (Nozaki and Tanford, 1971; Fauchere and Pliska, 1983). The amino acids were considered large if above an arbitrary apparent molar volume of 129 Å^3 (Glu), and hydrophobic was defined as a hydrophobicity index above 0.18 kcal/mol (His). From the smallest to largest in terms of size (with HP codes in parentheses), small amino acids were presented as follows: Gly (P), Ser (P), Ala (H), Cys (H), Asp (P), Thr (H), Asn (P), Pro (H); and the large ones were presented in a similar manner: Glu (P), Val (H), Gln (P), His (H), Met (H), Leu (H), Ile (H), Lys (P), Phe (H), Tyr (H), Arg (P), Trp (H).

Table 1
Information about primary (n_i) and secondary (L_h,$t_{i,h}$, $p_{i,h}$) structures for the 35-residue toxin 1ROO, and calculation procedure for fractions $p_{i,h}$, shown in %. 1ROO had only helices encompassing the sequences 9–11, 14–19 and 21–25, consequently L_h was equal to 14 and $p_{i,e}$ were not calculated. All the sequence and structure data employed in the present study are free for the use at the websites of PDB (http://www.pdb.org) and PROMOTIF (http://www.ebi.ac.uk/pdbsum/).

Amino acid residue sequences (20-letter, LS, HP codes) in helices of 1ROO

K S R A F Q C K H M K Y R L Total number of residues, L_h=14
L S L S L L S L L L L L L Total of large residues, $t_{L,h}$=11
P P P H H P H P H H P H P H Total of hydrophobic residues, $t_{H,h}$=7

Toxin	n_L; n_H	L_h	$t_{L,h}$; $t_{H,h}$	$p_{L,h}$; $p_{H,h}$
1ROO	17; 20	14	11; 7	78.6%; 50.0%

Fractions $p_{i,j}$ (Eq. (1)) of L, H residues in helices: $p_{L,h} = (t_{L,h}/L_h)100 = (11/14)100 = 78.6\%$; $p_{H,h} = (t_{H,h}/L_h)100 = (7/14)100 = 50.0\%$