Toxicon 148 (2018) 16–25

Contents lists available at ScienceDirect

Toxicon

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

Accelerated evolution of toxin genes: Exonization and intronization in snake venom disintegrin/metalloprotease genes

R. Manjunatha Kini

Protein Science Laboratory, Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, 117543, Singapore

ARTICLE INFO

Article history:
Received 16 January 2018
Received in revised form
21 March 2018
Accepted 1 April 2018
Available online 7 April 2018

Keywords:
Rapid evolution of proteins
Neofunctionalization
Origin of exons
Origin of intron
Evolution of reptile genes

ABSTRACT

Toxin genes in animals undergo accelerated evolution compared to non-toxin genes to be effective and competitive in prey capture, as well as to enhance their predator defense. Several mechanisms have been proposed to explain this unusual phenomenon. These include (a) frequent mutations in exons compared to introns and nonsynonymous substitutions in exons; (b) high frequency of point mutations are due to the presence of more unstable triplets in exons compared to introns; (c) Accelerated Segment Switch in Exons to alter Targeting (ASSET); (d) Rapid Accumulation of Variations in Exposed Residues (RAVERs); (e) alteration in intron-exon boundary; (f) deletion of exon; and (g) loss/gain of domains through recombination. By systematic analyses of snake venom disintegrin/metalloprotease genes, I describe a new mechanism in the evolution of these genes through exonization and intronization. In the evolution of RTS/KTS disintegrins, a new exon (10a) is formed in intron 10 of the disintegrin/metalloprotease gene. Unlike more than 90% new exons that are from repetitive elements in introns, exon 10a originated from a non-repetitive element. To incorporate exon 10a, part of the exon 11 is intronized to retain the open reading frame. This is the first case of simultaneous exonization and intronization within a single gene. This new mechanism alters the function of toxins through drastic changes to the molecular surface via insertion of new exons and deletion of exons.

1. Introduction

Venoms and toxins in animals have evolved as chemical armamentarium to allow the animals to be effective and competitive in prey capture, as well as to enhance their predator defense. The genes encoding toxins are thought to have been evolved by repeated events of gene duplications followed accelerated evolution and positive Darwinian selection (Nakashima et al., 1993; Ogawa et al., 1995; Deshimaru et al., 1996; Juarez et al., 2008; Moura-da-Silva et al., 2011; Casewell et al., 2011). Point mutations contribute to nonsynonymous mutations in exons of venom phospholipase A2 enzymes (Nakashima et al., 1993; Ogawa et al., 1995; Chuman et al., 2000) and serine proteases (Deshimaru et al., 1996; Nikandrov et al., 2005). We correlated the increased frequency of point mutations to the occurrence of stable and unstable triplets (Kini and Chinnasamy, 2010). The exons of toxin genes encoding venom phospholipase A2 enzymes, neurotoxins and cardiotoxins have higher percentage of unstable triplets compared to their introns. In contrast, non-venom genes have higher percentage of unstable triplets in introns compared to their exons (Kini and Chinnasamy, 2010). Interestingly, accelerated amino acid mutations in phospholipase A2 enzymes appear to target the molecular surface of toxins (Kini and Chan, 1999). We speculated that these mutations assist in changing from one target protein to other target protein (Kini and Chan, 1999). Point mutations alone are insufficient to explain the observed drastic changes in toxin isoforms. Thus, alternative molecular mechanisms, which may explain drastic changes in the molecular surface by altering the loop sizes and/or addition or removal of domains, were proposed. The short-chain neurotoxins diverged early in the course of evolution from the long-chain neurotoxins by the deletion of the fifth disulfide bond in loop 2 (Fujimi et al., 2003). This 'removal' of short segment was accommodated through the alteration in intron-exon boundary. A point mutation, the deletion of 'A', resulted in the loss of AG at the end of intron 2 and the shift of the intron 2/Exon 3 splicing site (Fujimi et al., 2003). The altered structure in loop 2 of short-chain neurotoxins is concomitant with their inability to bind to neuronal α7 nicotinic acetylcholine receptors. In snake venom metalloproteases the loss or gain of domains appears to play critical

E-mail address: dbskinim@nus.edu.sg.

https://doi.org/10.1016/j.toxicon.2018.04.005
0041-0101/

role in the evolution of their structure and function. The changes in the domains occur through recombination between closely related genes (Moura-da-Silva et al., 2011) or through 'acquisition' or 'forfeiture' of domains (Casewell et al., 2011). Thus, a number of hypothetical mechanisms have been proposed to explain the accelerated evolution of toxin genes.

Disintegrins are a family of nonenzymatic proteins from snake venoms that interfere in the function of integrins (Dennis et al., 1990; Gould et al., 1990). They are potent platelet aggregation inhibitors that block the final common step of fibrinogen interaction with platelet receptor glycoprotein IIbIIIa (integrin $\alpha_{IIb}\beta_{III}$) (Chao et al., 1989; Gan et al., 1988; Huang et al., 1989; Shebuski et al., 1989). These disintegrins have RGD/KGD tripeptide in a loop that plays critical functional role. Subsequent studies have shown several structural variants with distinct integrin selectivity and concomitant biological properties (for details, see Calvete et al., 2005; Calvete et al., 2003; Walsh and Marcinkiewicz, 2011). The structural variants include monomeric disintegrins that belonged to small, medium and large classes (Gould et al., 1990) and dimeric disintegrins (Calvete et al., 2003). The vast majority of disintegrins are part of a large precursor, which also encodes propeptide, metalloprotease, disintegrin and Cys-rich domains and the disintegrins are derived by proteolysis of this common precursor (Hite et al., 1992; Kini and Evans, 1992; Paine et al., 1992). Interestingly, the α-subunit of the dimeric disintegrin acostatin from *Agkistrodon contortrix contortrix* venom is derived from a much shorter precursor (Okuda et al., 2002). This precursor has only the disintegrin domain and lacks the propeptide and metalloprotease domains compared to the canonical large precursor of disintegrins. The β-subunit, however, is encoded by a typical large precursor. Similar domain organizations were also found in α- and β-subunits of piscivostatin from *Agkistrodon piscivorus piscivorus* venom (Okuda et al., 2002). These precursors, although shorter, had homologous signal peptide, followed by partial propeptide and medium disintegrin domains. In 2005, Sanz et al. (2005) determined the sequence of the short precursor of short disintegrin jerdostatin. This precursor has a 17-residue segment that has no similarity in the medium and large disintegrins (Fig. 1). The authors described the role of this segment in reducing the number of disulfide bonds in small disintegrins (Sanz et al., 2005). I was intrigued by this unusual, non-homologous segment and was interested in finding the origin of this segment. Such drastic change in the protein sequence would contribute to significant change in the structure and hence, the function. Thus, the understanding of this radical change may help towards the mechanism of accelerated evolution of toxins. By analyzing the draft sequence of the PIII metalloprotease gene from *Echis ocellatus* (Sanz et al., 2012), I have identified that this segment is encoded by a new exon formed from the intron 10 of this gene. Similar exons are also found in *E. ocellatus* PI and PII metalloprotease genes (Sanz and Calvete, 2016). In this paper, I describe the origin of this segment and how exonization and intronization contribute to the evolution of disintegrins. This is the first report on the role of exonization and intronization in evolution of toxin genes.

2. Materials and methods

2.1. Nucleotide and protein sequences

All sequences were obtained from public databases including NCBI. The sequences of disintegrins such as Acostatin α (Q805F7.1; *Agkistrodon contortrix contortrix*), Bitistatin (P17497.1; *Bitis arietans*), Piscivostatin α (Q805F5.1; *Agkistrodon piscivorus piscivorus*), Echistatin β (Q7LZK1.1; *Echis carinatus*), Eristostatin (P0C6S4.1; *Eristicophis macmahoni*), Jerdostatin (AAP20878.1; *Protobothrops jerdonii*), Lebestatin (CAJ34939.1; *Macrovipera lebetina*), Trigramin (P17495.1; *Protobothrops gramineus*), and CV11 α (CAJ34937.1; *Cerastes vipera*), and metalloprotease genes from *Echis ocellatus*, namely PI (EOC00028-like; KX219965), PII (EOC00006-like; KX219964) and PIII (EOC00089-like; KX219963) were analyzed in this study. Metalloprotease genes, namely, PIII (scaffold BCNE02095328; contig_95328), PII (scaffold BCNE02082875; contig_82875) and PI (scaffold BCNE02082869; contig_82869), from *Protobothrops mucrosquamatus* were downloaded from the genome database. *RPTLN* sequences (KU563546-KU563619) are from different organisms. These snake species are all phylogenetically distantly related (Wuster et al., 2008).

2.2. Sequence alignments and analyses

All translated amino acid sequences were obtained using Expasy (http://web.expasy.org/translate/). During the search for potential new exons, which may encode the unusual, non-homologous segment of jerdostatin (Fig. 1), the intron (2–10) sequences of PIII metalloprotease (KX219963) were translated in all three reading frames and the deduced amino acid sequences were searched for this unusual segment. Subsequently, new exons were identified by evaluating deduced amino acid sequences of intron 10. Nucleotide and amino acid sequences were compared by BLAST (Altschul et al., 1990) using default settings. All sequences were aligned and analyzed using default settings in MUSCLE (Edgar, 2004) (http://www.ebi.ac.uk/Tools/msa/muscle/), followed by manual evaluation. The boundaries of exon 10a and exon 11' determined by trial-and-error using the complete protein sequences of disintegrins for guidance.

3. Results and discussion

Toxin genes evolve at an accelerated rate compared to non-toxin genes. Although point mutations contribute to nonsynonymous mutations, they alone are insufficient to explain the observed drastic changes. The molecular surface changes drastically by altering the loop sizes and/or addition or removal of domains. The alteration in intron-exon boundary in three-finger toxin genes

Type	Name	Amino acid sequence
Long	Bitistatin	SPPVCGNKILEQGEDCDCGSPANCQDRCCNAATCKLTPGSQCNYGECCDQCRFKKAGTVCRIARGDWNDDYCTGKSSDCPWNH*
Medium	Trigramin	EAGKDCDCGSPAN PCCDAATCKLLPGAQCGEGPCCDQCSFMKKGTICRRARGDDLDDYCNGRSAGCPRNPFHA*
Medium	Acostatin α	IQPKN PCCDAATCKLTPGSQCAEGLCCDQCKFIKAGKICRRARGDNPDYRCTGQSGDCPRKHFYA*
Medium	Piscivostatin α	GAV QPKN PCCDAATCKLTPGSQCAEGLCCDQCKFIKAGKICRRARGDNPDYRCTGQSGDCPRKHFYA*
Short	Jerdostatin	GAVKQP**EK** **KHEPMKGNTLQKLPL**CTTGPCCRQCKLKPAGTTCW RTSVSSHY CTGRSCECPSYPGNG*
Short	Echistatin β	DCASGPCCRDCKFLKEGTICKRARGDNMDDYCNGKTCDCPRNPHKGEHDP*
Short	Eristostatin	QEEPCATGPCCRRCKFKRAGKVCRVARGDWNDDYCTGKSCDCPKNPWNG*

Fig. 1. Sequence alignment of disintegrins. Only jerdostatin shows the presence of a 17-residue segment (bold and underlined) that has no similarity with the medium and large disintegrins.