Peptides xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Peptides

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

Genome-wide search and comparative genomic analysis of the trypsin inhibitor-like cysteine-rich domain-containing peptides

Xian-Chun Zeng*,[1], Yichen Liu[1], Wanxia Shi, Lei Zhang, Xuesong Luo, Yao Nie, Ye Yang

State Key Laboratory of Biogeology and Environmental Geology & Department of Biological Science and Technology, School of Environmental Studies, China University of Geosciences (Wuhan), Wuhan 430074, People's Republic of China

ARTICLE INFO

Article history:
Received 21 June 2013
Received in revised form 13 August 2013
Accepted 13 August 2013
Available online xxx

Keywords:
TIL domain
Intron gain
Genomic organization
Alternative splicing
Overlapping gene
Scorpion *Mesobuthus martensii* Karsch
Protease inhibitor

ABSTRACT

It was shown that peptides containing trypsin inhibitor-like cysteine-rich (TIL) domain are able to inhibit proteinase activities, and thus play important roles in various biological processes, such as immune response and anticoagulation. However, only a limited number of the TIL peptides have been identified and characterized so far; and little has been known about the evolutionary relationships of the genes encoding the TIL peptides. BmKAPi is a TIL domain-containing peptide that was identified from *Mesobuthus martensii* Karsch. Here, we conducted genome-wide searches for new peptides that are homologous to BmKAPi or possess a cysteine pattern similar to that of BmKAPi. As a result, we identified a total of 80 different TIL peptides from 34 species of arthropods. We found that these peptides can be classified into seven evolutionarily distinct groups. Furthermore, we cloned the genomic sequence of BmKAPi; the genomic sequences of the majority of other TIL peptides were also identified from the GenBank database using bioinformatical approaches. Through phylogenetic and comparative genomic analysis, we found 26 cases of intron gain events occurred in the genes of the TIL peptides; however, no instances of intron loss were observed. Moreover, we found that alternative splicing contributes to the diversification of the TIL peptides. It is interesting to see that four genes of the TIL domain-containing peptides overlap in a DNA region located on the chromosome LG B15 of *Bombus terretris*. These data suggest that the evolution of the TIL peptide genes are dynamic, which was dominated by intron gain.

1. Introduction

Trypsin inhibitor-like cysteine rich (TIL) domain typically contains ten cysteine residues that form five disulfide bridges. The cysteine residues that form the disulfide bridges are C_1–C_7, C_2–C_6, C_3–C_5, C_4–C_{10} and C_8–C_9 [2]. Besides the trypsin inhibitors, it is interesting to see that many extracellular proteins, such as XP_001866937, XP_001197995 and EHJ63159, also contain multiple TIL domains.

Peptides containing the TIL domain generally consist of 56–84 amino acid residues. It was shown that a typical peptide of the TIL family is able to inhibit proteinase activity, and thus play roles in biological processes, such as inhibition of anticoagulation and participation in immune response [4,7,20,31]. Until now, at least five TIL family peptides have been described, including Api m-6 from the honey bee *Apis mellifera* [20], BMSI-7 and Ixodidin from the cattle tick *Boophilus microlpus* [7,31], BmKAPi from the Chinese scorpion *Mesobuthus martensii* Karsch [39], and SjAPI from the scorpion *Scorpiops jendeki* [4]. Api m-6 is a bee venom allergen that binds to IgE. BMSI-7 is an inhibitor of subtilisin. Ixodidin is an antimicrobial peptide that is also able to inhibit the activities of elastase and chymotrypsin. SjAPI is capable of inhibiting the activities of both α-chymotrypsin and elastase. Because TIL domain-containing peptides exhibit some unique features, such as compact rigidity, well-defined structure and small size, they are used as a molecular scaffold for the development of new pharmaceutical drugs. However, only a limited number of the TIL domain-containing peptides have been identified so far, which hinders the structure/function relationship investigation, evolutionary analysis and potential uses of this kind of peptides.

Here, we conducted a genome-wide searches for the homologues of BmKAPi using bioinformatical approaches. As results, we identified a total of 80 different TIL family peptides from 34 species of arthropods. We also cloned the genomic sequence coding for the BmKAPi peptide; the genomic sequences of the majority of other TIL domain-containing peptides were also identified from the GenBank database by bioinformatical approaches. Comparisons for the genomic sequences of the TIL peptides lead to the discovery of intron number and position polymorphisms, as well as extensive

* Corresponding author. Tel.: +86 027 67883481.
E-mail address: xianchun.zeng@gmail.com (X.-C. Zeng).
[1] These authors contributed equally to the work.

0196-9781/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.peptides.2013.08.012

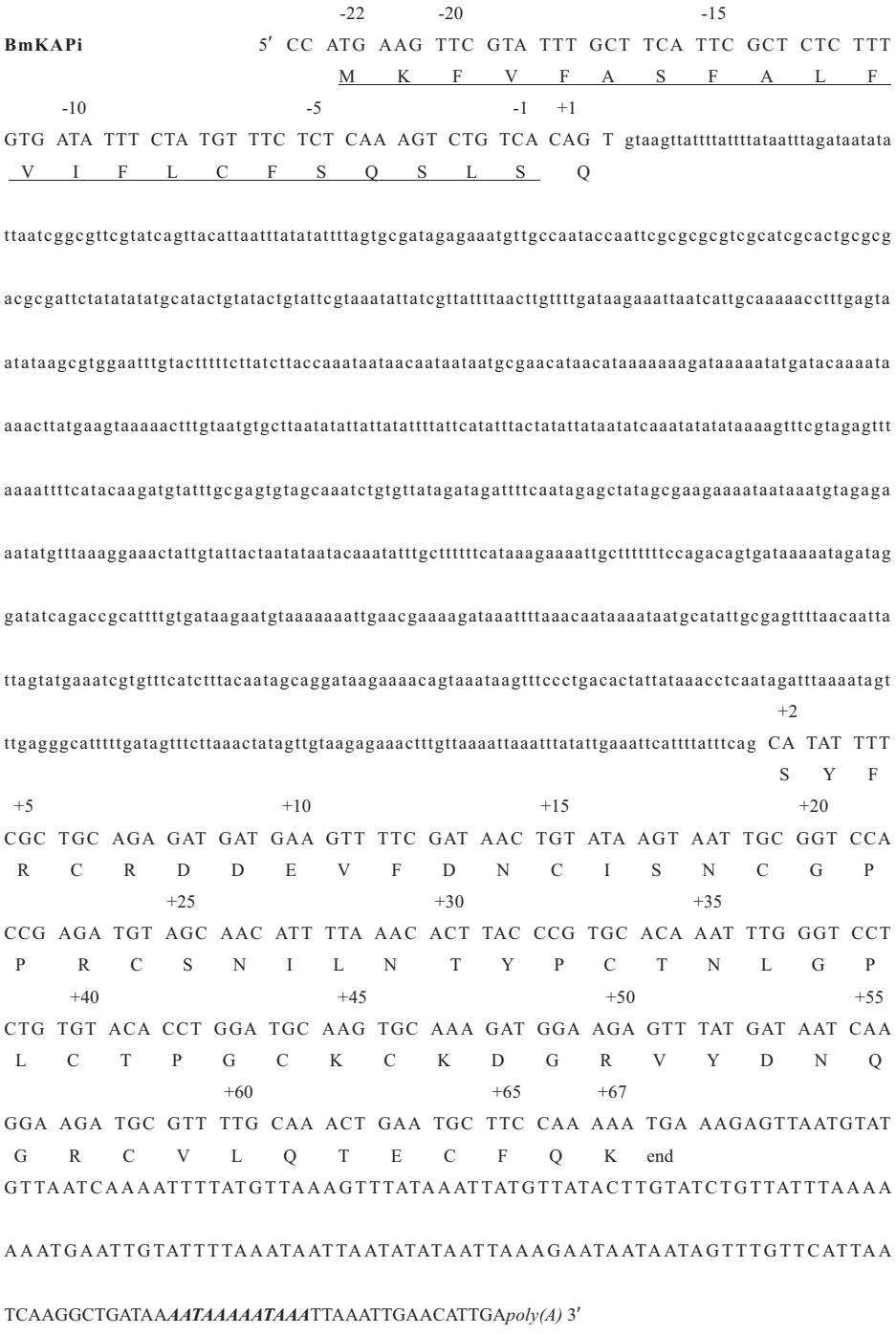

Fig. 1. Genomic organization of BmKAPi. Exons were written in capital letters, whereas intron was in lower-case letters. The genomic sequence of BmKAPi has been deposited in the GenBank database under the accession number of KF146892.

intron gain events. The evolution of the genes of the TIL domain-containing peptides was dominated by intron gain. We also found the alternative splicing is one of the factors responsible for the generation of new TIL peptides. It is interesting to see that the reverse complementary sequence of an intron of a TIL peptide gene codes for a different TIL peptide; and this gene overlaps with another gene that encodes a TIL peptide as well. This suggests that the genes of the TIL peptides could be transposable. Our data expand the knowledge of the TIL domain-containing peptides from arthropods.

2. Materials and methods

2.1. Animals

Specimens of the scorpion *M. martensii* Karsch were captured from Henan province, central China. These specimens were provided with cricket food and a wet sponge for water, and maintained in our laboratory for no more than 10 days.

Please cite this article in press as: Zeng X-C, et al. Genome-wide search and comparative genomic analysis of the trypsin inhibitor-like cysteine-rich domain-containing peptides. Peptides (2013), http://dx.doi.org/10.1016/j.peptides.2013.08.012