Molecular cloning and sequence analysis of serine protease cDNA from the venom of the centipede *Scolopendra subspinipes dehaani*

Objective: Proteolytic enzymes are one of the well-known components of animal venom. The aim of this study was to identify protease from the venom gland of *Scolopendra subspinipes dehaani*.

Methods: After total RNA extraction, a cDNA encoding a serine protease was identified by RT-PCR and was subsequently sequenced. The sequence was analysed by multiple alignment and homology modelling.

Results: A cDNA encoding the precursor of a centipede venom serine protease was identified. The full-length cDNA was 1,014 nucleotides with 780 nucleotides of open reading frame. The precursor nucleotide sequence encoded a signal peptide of 19 residues and a mature protein of 260 residues. It belonged to clan PA of serine proteases in the MEROPS database classification. Catalytic residues were identified by multiple alignments, including an unusual hydrogen bond network in the catalytic site. Secondary structure prediction and homology modelling revealed a chymotrypsin-like fold, which is characteristic of PA clan proteases.

Conclusion: The sequence of the identified serine protease showed unusual features, making this centipede venom serine protease an interesting candidate to investigate. A better understanding of the molecular mechanism of venom components will be useful for further applications and help to improve envenomation treatment.

Key Words: *Scolopendra subspinipes dehaani*, cDNA, serine protease

Conflict of Interest: The authors have no conflict of interest.

ABSTRACT

**Amaç:** Proteolitik enzimler hayvan zehirlerinin en çok bilinen bileşenlerinden biridir. Bu çalıșmanın amacı *Scolopendra subspinipes dehaani* zehir salgı bezinden proteaz enziminin tanımlanmasıdır.

**Metod:** Toplam RNA eldesinin ardından, RT-PCR tekniği kullanılarak serin proteaz proteinini kodlayan gene ait cDNA tanımlanmış ve dizi analizine tabi tutulmuştur. Elde edilen diziler çoklu dizi hizalama ve homoloji modelleri ile analiz edilmiştir.

**Bulgular:** Kırkayak zehirinde bulunan serin proteaz öncülünü kodlayan bir cDNA tanımlandır. cDNA, 780 nükleotit açık okuma çerçevesinde olmak üzere, toplam 1,014 nükleotit uzunluğundadır. Oncül nükleotit dizisi 19 amino asitlik bir sinyal peptidini ve 260 amino asitlik olgun proteini kodlamaktadır. Bu protein MEROPS veritabanına eklenmiştir ve genel proteazdan PA grubuna aitir. Çoklu dizi hizalama kullanılarak alışılamadık bir hidrojen bağlı örgüne sahip katalitik bölge ve bu bölgedeki katalitik aminoasitler tanımlanmıştır. İçindeki yapılı koordinasyonu ve homoloji modellerine sahip PA grubu proteazlara oranla kimyasal benzeri katlanmayı göstermiştir.

**Sonuç:** Tanımlanan serin proteaz dizisi alışılamadık özellikler göstermesi sebebiyle kırkayak zehirindeki serin proteaz enzimini ileri çalışmalar için aday yapmıştır. Venom bileşenlerinin moleüler mekanizmalarının daha iyi anlasılması sonraki çalışmalar için yardımcı olabileceği gibi zehirlenmelerin tedavilerine de yardımcı olacaktır.

**Anahtar Kelimeler:** *Scolopendra subspinipes dehaani*, cDNA, serin proteaz

**Çıkar Çatışması:** Yazaların çıkar çatışması yoktur.
Introduction

Serine proteases are widely distributed in all kingdoms of cellular life [1]. They are well-known proteolytic enzymes that are named due to the role played by a nucleophilic serine residue in the catalysis of peptide bond hydrolysis along with two other amino acids: a basic residue, as an histidine, and an acidic residue, as an aspartate [2,3]. Proteases using this catalytic triad exhibit at least four distinct folds, as illustrated by trypsin, subtilisin, prolyl oligopeptidase and Clp peptidease, and correspond to clan PA, SB, SC and SK, respectively from the MEROPS protease database [4]. The majority of serine proteases uses the canonical catalytic triad and follow the chymotrypsin-like fold of clan PA [5]. Proteolytic enzymes are a major component of venom. They mainly affect the haemostatic system of the prey or victim [6]. Indeed, a number of snake venom serine proteases that affect haemostasis have been reported, including procoagulant, anticoagulant, platelet aggregating or fibrinolytic proteases [7-9].

Although stings and envenomation by centipede are fairly frequent, they result more often in benign and local symptoms with spontaneous healing, and treatment is mainly supportive. Thus, knowledge on centipede venom is limited, because centipedes are regarded as a neglected group of venomous animals due to their limited venom quantity [10]. Nevertheless, centipede venom is a complex mixture and may be a potential source of bioactive molecules. In this study, we report the molecular cloning of a cDNA from the S. subspinipes dehaani venom gland that encodes a serine protease. The sequence was analysed in comparison to the known sequence.

Materials and Methods

Total RNA isolation and cDNA synthesis

The centipedes, S. subspinipes dehaani, were collected from a suburban area of Khon Kaen City, Thailand. Their forcipules were dissected, and the venom glands were then immediately ground in liquid nitrogen. Fifty mg of tissue was homogenised in TRIzol® reagent. Total RNA was isolated using TRIzol® (Invitrogen Life Technologies, USA) according to the manufacturer’s instructions. The quality of the RNA was verified by 1.2% agarose gel electrophoresis. RNA concentration was determined using a nanodrop spectrophotometer (Thermo Scientific, USA). Then, the RNA was directly used for cDNA synthesis. First-strand cDNA was generated using the ReverseAid™ First Strand cDNA Synthesis Kit (Thermo Scientific, USA). The PCR reactions were performed as described above. The PCR products were visualised by 1.5% agarose gel electrophoresis and then cloned and sequenced using the methods described above.

Sequence analyses and multiple alignment

Sequence analysis was performed using BioEdit software. The Basic Local Alignment Search Tools (BLAST) (http://www.ncbi.nlm.nih.gov/) program was used to retrieve similar sequences from Nr/Nt, Nr, SwissProt and the Protein Data Bank (PDB). The blastn algorithm with the Nr/Nt and Nr databases was used for nucleotide se-

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>primer H</td>
<td>GTGCTTACTGCTGCTATTG</td>
</tr>
<tr>
<td>primer S</td>
<td>GGACCACCTGGAATCACCC</td>
</tr>
<tr>
<td>SPaseD</td>
<td>TCCGATTCAGAAGCTGTCG</td>
</tr>
<tr>
<td>dehaF1</td>
<td>ATGATGCCTTTTGAATATTTG</td>
</tr>
<tr>
<td>dehaF2</td>
<td>ATTACGCGGGGAACATTGTCG</td>
</tr>
<tr>
<td>dehaR</td>
<td>GGAATGAGTAGTAATAATGTTATT</td>
</tr>
</tbody>
</table>

Table 1. List and sequence of primers
were then aligned using the “K*Sync” alignment method [17]. From this alignment, a template was generated and the variable regions were modelled in the context of the fixed template using Rosetta fragment assembly. Finally, the side chains of the final model were repacked using a Monte-Carlo algorithm with a backbone dependent side chain rotamer library. Pymol was used to visualise and generate pictures of three-dimensional models (http://pymol.org/).

Results and Discussion

Full-length cDNA determination

A partial sequence of the cDNA encoding a centipede venom serine protease (CVSP) was identified from total RNA (Fig. 1a) by RT-PCR using specific primers (Table 1) that were designed according to the conserved catalytic amino acids of serine protease: histidine (primer H) and serine (primer S) [11]. The expected PCR product (443 bp) showed a single sharp band for the partial CVSP sequence (Fig. 1b). The full-length cDNA was obtained by RACE system using specific primers (primer H and SPas-eD), designed on the basis of the partial sequence and the universal primers provided from the RACE kit. The full-length cDNA consists of 1,014 bp with 780 bp of an open reading frame (ORF), a 5′-untranslated region (5′-UTR) of 101 bp upstream from the first ATG and a 3′-UTR of 133 bp downstream from the stop codon. The ORF encodes a 260 amino acid residue protein with a 19-amino acid signal peptide sequence (Fig. 2a). The precursor (780 bp) and the mature serine protease (723 bp) were cloned using the reverse primer dehaR and the forward primer dehaF1 or dehaF2, respectively (Fig. 1c).

Alignment and amino acid sequence analysis of CVSP

The mature CVSP has a calculated MW of 26940.61 Da and a pl of 5.74. The N-terminal amino acid sequence was then aligned using the “K*Sync” alignment method [17]. From this alignment, a template was generated and the variable regions were modelled in the context of the fixed template using Rosetta fragment assembly. Finally, the side chains of the final model were repacked using a Monte-Carlo algorithm with a backbone dependent side chain rotamer library. Pymol was used to visualise and generate pictures of three-dimensional models (http://pymol.org/).
serine proteases, CVSP can be classified in the S1 family of clan PA [4,18]. Substrate specificity of chymotrypsin-like proteases of the PA clan is mainly determined by the interaction between the S1 pocket and the P1 amino acid side chain of the substrate [3]. In CVSP, the S1 pocket is negatively charged due to the conserved Asp202 (Fig. 3). This suggested a substrate specificity similar to trypsin for peptides with positively charged amino acids in position P1, such as arginine or lysine.

According to PSIPRED, CVSP is composed of 12 beta sheets and 2 alpha helices (Fig. 3). The predicted second-

is similar to other serine proteases with two aliphatic amino acids, isoleucines, followed by two glycines [I(V)I(V)GG] (Fig. 3). The amino acids His61, Asp108 and Ser208 that form the catalytic triad of CVSP are strictly conserved among all analysed sequences. The oxyanion hole from CVSP is formed by Gly206 and Ser208. As the catalytic residue, Ser208 is strictly conserved whereas Gly206 is conserved among centipede and snake serine protease sequences, but not with hymenoptera. According to the comparison of the sequences surrounding the catalytic amino acids of CVSP and evolutionary markers of serine proteases, CVSP can be classified in the S1 family of clan PA [4,18]. Substrate specificity of chymotrypsin-like proteases of the PA clan is mainly determined by the interaction between the S1 pocket and the P1 amino acid side chain of the substrate [3]. In CVSP, the S1 pocket is negatively charged due to the conserved Asp202 (Fig. 3). This suggested a substrate specificity similar to trypsin for peptides with positively charged amino acids in position P1, such as arginine or lysine.

According to PSIPRED, CVSP is composed of 12 beta sheets and 2 alpha helices (Fig. 3). The predicted second-
234 form disulfide bonds 1, 2 and 4, respectively, and are strictly conserved, while cysteines 180 and 195 that form disulfide bond 3 are only missing in the hymenoptera serine protease sequence (Fig. 3). Residue Ile229 of CVSP corresponds to Ser214 in bovine trypsin, which is known to form a hydrogen bond with the catalytic histidine and aspartate by its carbonyl and its OH side chain, respectively. This last hydrogen bond is present in most

Figure 3. Multiple sequence alignment of the *S. subspinipes dehaani* serine protease CVSP, and serine proteases from the venom of centipede (*Scolopendra subspinipes mutilans* [AAD00320]), hymenoptera (*Polistes dominula* [AAP37412], *Vespa magnifica* [ABY78988], *Apis mellifera* [NP_001011584]) and snake (*Gloydius saxatilis* [Q7SZE1], *Bothrops jararacussu* [Q2PQ13], *Gloydius halys* [AFM29142]) and trypsin (*Bos taurus* [NP_001107199]). Conserved catalytic residues are indicated by a black star (★), oxyanion residues by a black circle (●), Ile229 participating in the hydrogen bond network with the catalytic triad by a black triangle (▲), Asp202 participating in the S1 pocket by a black square (■) and predicted disulfide bonds by number. The secondary structure of trypsin (*Bos taurus* [4GUX]) and the PSIPRED predicted secondary structure of CVSP are indicated on the top.
Conflict of Interest

There are no conflicts of interest among the authors.

References


Figure 4. Homology modelling of the S. subspinipes dehaani serine protease. (a) The homology model shows the two β-barrels at the N- and C-terminus, the C-terminal α helix and the loop from amino acids 146–167. The side chains of the catalytic triad residues His61, Asp108, and Ser208 lie in a cleft between the two β-barrels and are shown as sticks, while the amide nitrogen of the backbone for oxyanion residues Gly206 and Ser208 are indicated as blue spheres. Ile229 is represented as a stick, and the methyl group that is usually replaced by the OH of a serine in the homologous serine protease is indicated as a yellow sphere. (b) Homo sapiens hepatocyte growth factor activator, PDB 1YC0, was used as template for homology modelling. The OH of the serine 245 side chain is indicated as a red sphere. (c) Superposition of both structures, with the model in blue and the template in green.

chymotrypsin-like proteases [3,19]. The polar character of position 229 is well conserved in the alignment; however, only CVSP presents a non-polar residue for this position (Fig. 3). The presence of a non-polar isoleucine at position 229 suggests an unusual hydrogen bond network in the catalytic site of CVSP.

Homology modelling of CVSP

The homology model was built using the Robetta server [17] on the structure 1YC0, the human hepatocyte growth factor activator, which has a 33.6% identity with CVSP. The model exhibited the classical fold of clan PA proteases, referred to as a chymotrypsin-like fold and composed of two β-barrels with a classical Greek-key architecture for the β strand topology of each β-barrel (Fig. 4) [5]. The catalytic triad, His61, Asp108 and Ser208, is found in close proximity in the active site cleft between the two β-barrels. Ile229 is part of the amino acids surrounding the catalytic triad that form an extensive hydrogen bond network (Fig. 4) [20]. However, the non-polar properties of the Ile229 side chain suggest an unusual organisation of this hydrogen bond network, making CVSP an interesting candidate to investigate. A better understanding of the molecular mechanism of venom components will be useful for further applications and help to improve envenomation treatment.

Acknowledgements

This research was supported by the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. program (PHD/0282/2549) and the “TRF-CHE Research Grant for Mid-Career University Faculty” and was jointly funded by the TRF and the Office of the Higher Education Commission (OHEC), Ministry of Education, Thailand for fiscal years 2007-2009 and additionally supported by the Khon Kaen University (KKU) Research Fund for fiscal years 2007-2010.
Supplementary Figure 1. Figure 1. Pairwise alignment of CVSP (centipede venom serine protease, Scolopendra subspinipes dehaan [AIJ10719]) and HGFA (hepatocyte growth factor activator, Homo sapiens [1YC0_A]). Conserved catalytic residues are indicated by a black star (★), oxyanion residues by a black circle (●), Ile229 participating in the hydrogen bond network with the catalytic triad by a black triangle (▲), Asp202 participating in the S1 pocket by a black square (■) and predicted disulfide bonds by number. The secondary structure of HGFA (Homo sapiens [1YC0_A]) is indicated on the top.