ELSEVIER

Toxicon 42 (2003) 715–723

TOXICON

www.elsevier.com/locate/toxicon

cDNA sequence analysis of seven peptide toxins from the spider *Selenocosmia huwena*

Jianbo Diao[a], Ying Lin[b], Jianzhou Tang[b], Songping Liang[b,*]

[a]*National laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing 100871, China*
[b]*Department of Biology, College of Life Science, Hunan Normal University, Changsha, Hunan 410081, China*

Received 21 May 2003

Abstract

Seven cDNAs encoding six toxins HWTX-I, HWTX-II, HWTX-IIIa, HWTX-IV, HWTX-V, HWTX-VII and one lectin SHL-I, from the spider *Selenocosmia huwena*, were cloned and sequenced. On the basis of their amino acid sequences, we designed and synthesized 3′ RACE and 5′ RACE primer. By overlapping the two partial cDNA sequences obtained by 3′ and 5′ RACE, their full-length cDNA sequences were obtained. All of the cDNAs of these seven peptides encode a precursor including a potential signal peptide of 21–24 residues, a mature toxin of about 30 residues and an intervening pro region. The prepro regions of HWTX-I, HWTX-IIIa, HWTX-IV, HWTX-V and SHL-I were demonstrated, by the comparison of the cDNA sequences, to have high similarity, which is concert with the similar inhibitor cystine knot motif of HWTX-I, HWTX-IV and SHL-I although their functions are different. It was also demonstrated that, HWTX-II and HWTX-VII share the highly similar prepro region which is different from that of HWTX-I, HWTX-IV and SHL-I. The three dimensional structure of HWTX-II has been determined to exhibit a different motif. This indicates that the seven peptides from *S. huwena* could be classified into two different superfamilies according to the prepro region of cDNA sequences.

Keywords: cDNA sequence; RACE; Huwentoxin; Inhibitor cystine knot motif; Superfamily; *Selenocosmia huwena*

1. Introduction

The spider *Selenocosmia huwena* is distributed in the hilly areas of Guangxi and Yunnan in the south of China. The venom from the spider *S. huwena* contains a mixture of compounds with different types of biological activities. The mixture includes components such as enzymes, lectins, enzyme inhibitors and several classes of neurotoxins, which are of interest as tools for studying neurophysiology and as potential lead structures for insecticidals and pharmaceuticals. In our previous work, by the means of reverse phase and ion-exchange high performance liquid chromatography, some peptides with different activities were purified and characterized. Among them, huwentoxin-I (HWTX-I) has 33 amino acid residues and three disulfide bonds (Liang et al., 1993). Its 3D (three dimensional) structure has been determined by the 2D-^{1}H NMR technique and displays as typical inhibitor cystine knot motif (ICK motif) (Qu et al., 1995). It also has been proved that HWTX-I acted as a presynaptic toxin and can block the N-type high-voltage activated calcium channel (Peng et al., 2001). Huwentoxin-IV (HWTX-IV) is a 35 amino acid peptide with three disulfide bonds, and its 3D structure is also ICK motif according to the 2D-^{1}H NMR experiment. Different from HWTX-I, the physiological activity of HWTX-IV is an inhibitor of tetrodotoxin (TTX) sensitive voltage-gated sodium channel (Peng et al., 2002). Selenocosmia huwenlectin-I (SHL-I) is a 32 amino acid peptide with three disulfide bonds, and also exhibits ICK motif as well as HWTX-I and HWTX-IV

Abbreviations: RACE, rapid amplification of cDNA ends; HWTX-I, huwentoxin-I; HWTX-II, huwentoxin-II; HWTX-IV, huwentoxin-IV; HWTX-VII, huwentoxin-IIa; SHL-I, Selenocosmia huwenlectin-I; HWTX-IIIa, huwentoxin-IIIa; HWTX-V, huwentoxin-V; ICK, inhibitor cystine knot.

* Corresponding author. Tel./fax: +86-731-8861-304.
E-mail address: liangsp@public.cs.hn.cn (S. Liang).

0041-0101/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.toxicon.2003.08.007

(Lu et al., 1999), however, the activity of SHL-I is haemagglutination (Liang and Pan, 1995), and has almost no neurotoxin activity. Interestingly, the pairing model of three disulfide bonds of three peptides is the same model of Cys(I–IV), Cys(II–V), Cys(III–VI) (It is conventional to label the six Cys residues involved in the knot as I–VI, in order from the N- to C-terminus), at the same time, the key structure feature of these three peptides is a three-strand antiparallel β-sheet and a disulfide knot, then they all have the characteristic property of ICK motif (Craik et al., 2001). huwentoxin-II (HWTX-II), huwentoxin-IIa (HWTX-VII), huwentoxin-IIIa (HWTX-IIIa) and huwentoxin-V (HWTX-V) (Zhang et al., 2003) can all cause reversatile paralysis of cockroaches and are insecticidal toxins. Their amino acid residues are separately 37, 35, 33 and 35, and all of them have three disulfide bonds. The disulfide paring of HWTX-II, different from the former three peptides, is in a way of I–III, II–V, IV–VI, thus the steric conformation resolved by NMR spectrometry is a new motif different from ICK (Shu et al., 2002). Different from the former toxins, the disulfide bonds paring and steric structures of HWTX-VII, HWTX-IIIa and HWTX-V have not yet been elucidated.

Although the structures and functions of these seven venom compounds from the spider have been intensively studied, there is not any report on the cDNAs of these peptides, which are significant for the understanding of the evolutionary relationship and the genetic classification of these peptides.

In this study, the full-length cDNA sequences of the seven peptides from the venom of the spider *S. huwena* were elucidated by using the method of 3′ and 5′ rapid amplification of cDNA ends (RACE). We have got the evidence that these seven peptides can be classified into two different superfamilies according to the prepro regions of cDNA sequences.

2. Materials and methods

2.1. Materials

The spiders *S. huwena* were collected in Guangxi province, China. Their venom glands were isolated and immediately frozen in liquid nitrogen. The 3′ and 5′ RACE kits and TRIzol reagent were purchased from Invitrogen Inc., restriction enzymes, Taq DNA polymerase and pGEM-T easy vector system were from Promega. IPTG, X-gal and other chemical reagents were analytical reagent grade.

2.2. Preparation of total RNA

Venomous glands (100 mg) frozen in liquid nitrogen were ground into fine powder. By using TRIzol reagent kit, the total RNA extraction was performed in accordance with the instructions of the supplier.

2.3. 3′ RACE

Five micro grams mRNA was taken to convert mRNA into cDNA by using the 3′ RACE kit supplied with Superscript II reverse transcriptase and universal adapter primer (5′-GGCCACGCGTCGACTAGTAC$(dT)_{17}$-3′). The cDNA was then used as template for PCR amplification using a pair of primers, namely the gene specific primer and the abridged universal adapter primer containing an additional *Hin*dIII restriction site (5′-CGAAGCTTGGCCACGCGTCGACTAGTAC-3′). In this article, when we cloned the partial cDNA sequence of HWTX-II by 5′ RACE, we also got a similar partial sequence which is consistent with the protein sequence of HWTX-VII. Then, on the basis the 5′ partial cDNA sequence of HWTX-VII, the primer 9 was designed and synthesized. Finally, in the former work by Min Li, the partial cDNA sequence that encoded the mature peptide of HWTX-I has been cloned (Li et al., 2001) (GeneBank No. AF157504). In order to get the full sequence of the HWTX-I cDNA, the primer 10 was also designed and synthesized. The sequences of all of the 5′ RACE primers were listed in Table 1. The amplified products were then purified and cloned into the pGEM-T easy vector for sequencing.

2.4. 5′ RACE

On the basis of the partial cDNA sequences of HWTX-IV, SHL-I, HWTX-II, HWTX-IIIa and HWTX-V which were determined by 3′ RACE and also the partial cDNA sequence of HWTX-I, their anti-sense primers were designed and synthesized for 5′ RACE, which had also been shown in the Table 1. With the strategy described by the RACE kit supplier, the 5′-end cDNA species of HWTX-I, HWTX-IV, SHL-I, HWTX-IIIa, HWTX-V and HWTX-II were cloned by using their own gene-specific primers and nested primers, respectively. The amplified products were then precipitated and cloned into the pGEM-T easy vector for sequencing.

2.5. DNA sequencing and computer analysis

DNA sequencing was performed by Bioasia Inc. Nucleic acid sequences were analyzed using the software of DNAclub (by Xiongfong Chen) and DNAman (by Nynnon biosoft).

2.6. Mass spectrometry determination of amidation of HWTX-IV

Mass spectrometry analysis of the HWTX-IV was performed using a Voyager-DE™ STR MALDI-TOF mass spectrometer of ABI Company. The HWTX-I, whose relative molecular mass has been accurately determined, is taken as a relative molecular mass standard. After displaying all the isotopic relative molecular mass, only

Table 1
Structure of oligonucleotide primers

3′ RACE primers		
Toxin	Composition	a.a.sequence
HWTX-IV	P1: 5′-CGGA(A/G)TG(T/C)CT(A/G/C/T)GA (A/G)AT(A/C/T)TT(T/C)AA-3′	ECLEIFK
	P2: 5′-CGTG(T/C)AA(C/T)CC (A/G/C/T)(A/T)(G/C) (A/G/C/T)AA(C/T)GA(C/T)CA-3′	CNPSNDQ
SHL-I	P3: 5′-CGGA(T/C)AA(A/G)TG (T/C)GA(T/C)TA(T/C)AA(T/C)AA-3′	DKCDYNN
HWTX-IIIa	P4: 5′-CGGA(T/C)TG(T/C) GC(A/T/C/G)GG(A/T/C/G)TA(T/C)ATG-3′	DCAGYM
	P5: 5′-CGATG(A/C)G(A/T/C/G)GA(A/G)TG(T/C) AA(G/A)GA(A/G)AA-3′	MRECKEK
HWTX-V	P6: 5′-GG(T/C/A/G)GG(T/C/A/G)TG(T/C) (T/A)(C/G)(T/C/A/G)CA(G/A)GA(T/C)GG-3′	GGCSQDG
	P7: 5′-CGTG(T/C)AA(G/A)CA(C/T)(C/T) T(T/C/A/G)CA(G/A)TG(T/C)C-3′	CKHLQCH
HWTX-II	P8: 5′-CGTG(T/C)GA(G/A)AT(A/C/T)GA(G/A) AA(A/G)GA(G/A)GG-3′	CEIEKEG
HWTX-VII	P9: 5′-GAAAAAAAAGGCGAATCATGC-3′	IEKEGDK
HWTX-I	P10: 5′-TGC ACACCTGGAAAGAATGAG-3′	CTPGKNE
5′ RACE primers		
Toxin	Composition	a.a.sequence and 3′UTR
HWTX-XI	P11: 5′-GGATCTCATTTGCCTATTTGG-3′	3′ UTR
	P12: 5′-TTT ACACCACCTGGTTTTTCG-3′	RKTRWCK
SHL-I	P13: 5′-ACA AAGAATAGAATCTCAAC-3′	3′ UTR
	P14: 5′-CCATTTCCAAGTGGTG AAC-3′	CSRTWKW
HWTX-IIIa	P15: 5′-CTGTCTGGTACAAGGAA TAG-3′	3′UTR
	P16: 5′-TGGCAATACGCACCATTT CCG-3′	RKWCVLP
HWTX-V	P17: 5′-TGTTATAGGGATTGGGA ACG-3′	3′UTR
	P18: 5′-GGAAAATGTTCCATCCCATACGC-3′	CVWDGTFS
HWTX-II	P19: TAATGATTGAAGGACGCATGC-3′	3′UTR
	P20: GAATTTGCATTTCCATCCACC-3′	GGWKCKF
HWTX-I	P21: 5′-GTTTGCACCACTTGTGTTTATC-3′	DKHKWCK

the monoisotopic relative molecular mass was chosen for analysis.

3. Results

3.1. cDNA sequence of HWTX-I, IIIa, IV, V and SHL-I

The cDNA sequences of the five toxins HWTX-I, IIIa, IV, V and SHL-I were completed by overlapping two fragments amplified by 3′ and 5′ RACE. Their open reading frames all encoded a signal peptide of 21–24 residues, a mature peptide of 30–40 residues and an intervening pro region (Fig. 1). The deduced amino acid sequences from five cDNAs were consistent with the determined except that in some toxins additional residues were found at the C terminus of the mature peptide before the stop codon. Among them, HWTX-IIIa and SHL-I had additional Arg-Arg dipeptide while HWTX-V had an additional Lys at the C terminus. All of the above residues were removed during the post-translational processing; different from the above three toxins, HWTX-IV had additional Gly-Lys dipeptide

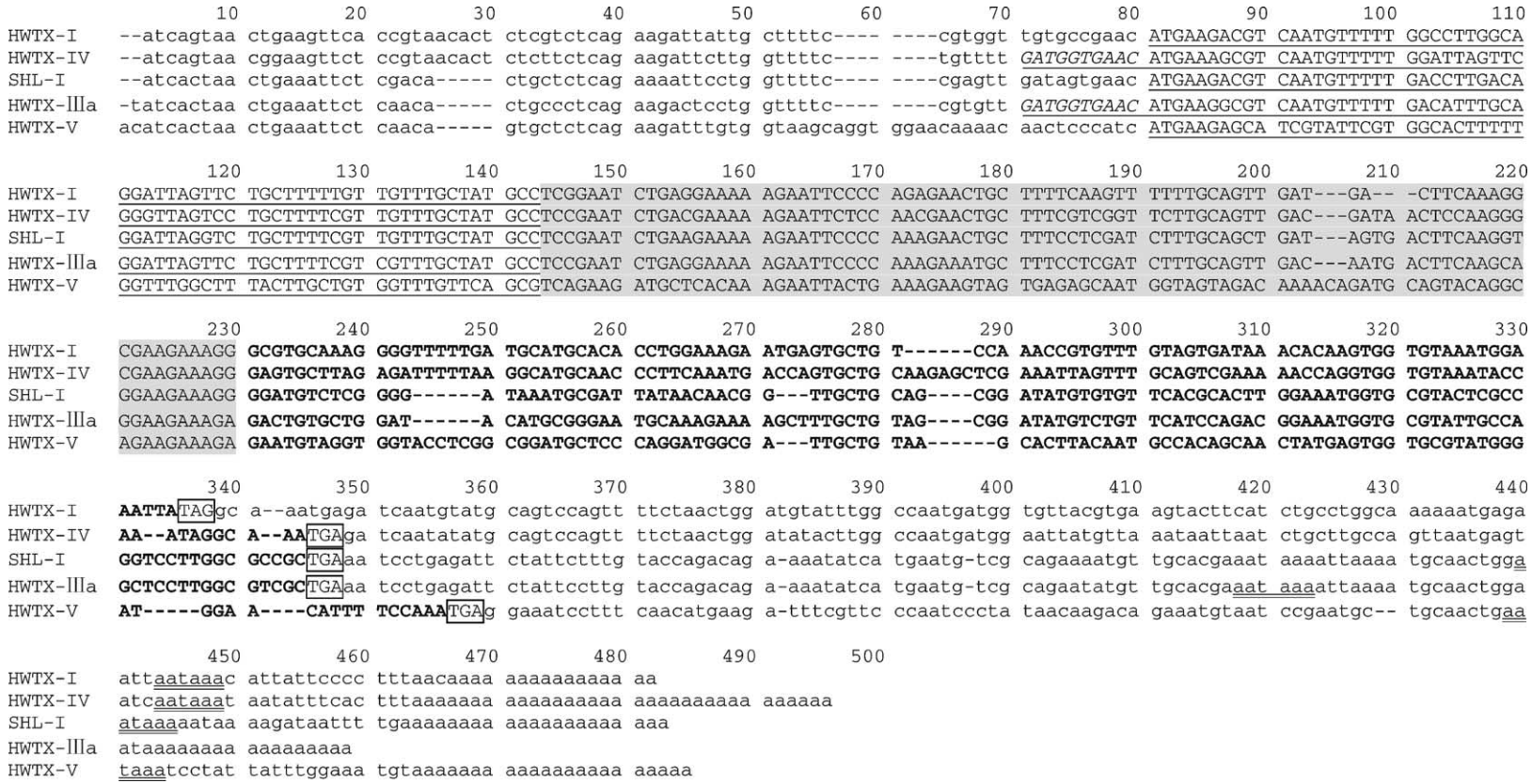

Fig. 1. Comparison of the nucleotide sequences of cDNAs encoding HWTX-I, HWTX-IV, SHL-I, HWTX-IIIa and HWTX-V. The signal sequence is underlined, the propeptide is given in the gray box, the mature toxin in bold and the stop codon in the white box. Gaps(-) were introduced to maximize regions of similarity between sequence. The polyadenylations signal, AATAAA, is double underlined.

```
                10        20        30        40        50        60        70        80        90       100       110
HWTX-II   ----tgtagg tgattgcttt cgtagtagat ctgattccat agttagacaa cttcagctag ttctggtcaa ttgcaggaga atttgaaaga caATGAAGGT GACATTGATT
HWTX-VII  tgagtgcggg tgatttcttt cgtagtagat ctcatttcat agtttcacaa cttcag---- ---------- ------gaga atttgaaaga caATGAAGGT GACATTAATT

HWTX-II   GCCATTCTGA CATGCGCTGC AGTGTTAGTT CTTCACACAA CAGCAGCAGA AGAACTCGAA GCAGAAAGTC AGCTGATGGA AGTTGGTATG CCCGATACAG AATTAGCAGC
HWTX-VII  GCCATTCTGA CATGCGCTGC AGTGTTAGTT CTTCACACAA CAGCAGCAGA AGAACTCGAA ---GAAAGTC AGCTGATGGA AGTTGGTATG CCCGATACAG AATTAGCAGC

HWTX-II   TGTGGATGAA GAAAGACTCT TCGAATGCTC TTTTTCATGC GAAATTGAGA AAGAAGGCGA CAAACCATGC AAAAAGAAGA AATGTAAAGG TGGATGGAAA TGCAAATTCA
HWTX-VII  TGTGGATGAA GAAAGACTCT TCGAATGCTC TATTTCATGC GAAATTGAAA AAAAAGGCGA ---ATCATGC AAACCGAAGA AATGTAAAGG TGGATGGAAA TGCAAATTCA

HWTX-II   ATATGTGTGT GAAGGTTTAA agtgccgtaa tcatcacgac aaaattcgga atgaaagtg gagagatttg tccttggtga cccattggaa agtacgcatg cgtccttcaa
HWTX-VII  ATATGTGTGT GAAGGTTTAA agtgccgaaa tcatcacgac aaaattcgga atgaaatgta tagagatctg tccttggtga cccattggaa agtacgcatg cgtccttcaa

HWTX-II   tcattaaccg caatattgtt tgaatctctt ctgtttctga aacgaattct caataaaatt gttaaaacat ttcaaaaaaa aaaaaaaaaa aaaaa
HWTX-VII  tcattaaccg caatattgtt tgaatctctt ctgtttctga aacgaattct gaataaaatt attaaacct aaaaaaaaaa aaaaaaaaaa aaaaaa
```

Fig. 2. Comparison of the nucleotide sequences of cDNAs encoding HWTX-II and HWTX-VII. The signal sequence is underlined, the propeptide is given in the gray box, the mature toxin in bold and the stop codon in the white box. Gaps (-) were introduced to maximize regions of similarity between sequence. The polyadenylations signal, AATAAA, is double underlined.

which was processed after translation and was absolutely required for the C terminal amidation of the mature toxin. The cDNA partial sequence of HWTX-I was once determined by Min Li (Li et al., 2001), and it was not a complete sequence for shortage of the prepro and untranslated sequence. In this article, we got its whole cDNA sequence by the method of RACE. The determined sequence not only supplied the prepro region but also rectified a mistake of site 32 of mature peptide of the sequence submitted to Genebank by Min Li in that it should be AAA instead of AAG. The latter might occur for the mistake of PCR amplification. The prepro region of the five toxins shared high similarity (79.87%) although their mature peptide just displayed relative low similarity (58.42%). Finally, polyadenylation signal, AATAAA, was found in the 3′ untranslated region at position 15–20 upstream of the poly(A).

3.2. cDNA sequence of HWTX-II and HWTX-VII

The cDNA sequence of HWTX-II and HWTX-VII was completed by overlapping two fragments amplified by 3′ and 5′ RACE, their open reading frames encoded a signal peptide of 21 residues and a mature peptide of 37 or 35 residues and a pro region (Fig. 2). The deduced amino acid sequences were consistent with the determined. The cDNA sequences of HWTX-II and HWTX-VII displayed high similarity not only for their prepro region (97.22%) but also their mature peptide (89.19%). Comparison with the former five toxins, although the organization of cDNAs was in a same way, the prepro region of the later two toxins had relatively low similarity with them. The polyadenylation signal, AATAAA, was also found in the 3′ untranslated region upstream of the poly(A).

3.3. MS determination of C terminal amidation of HWTX-IV

According to the sequence of its primary structure, the monoisotopic relative molecular mass of HWTX-IV is 4111.9470 (calculated with the software in the computer of Voyager-DE™ STR MS spectrometer) if the C terminal is free acid. After deletion of six hydrogen atoms when the three pairs of disulfide bridges are formed, the relative molecular mass is 4105.9001 (4111.9470 − 1.007825 × 6 = 4105.9001). Then if the C terminal is amidated, the mass becomes 4104.8923 (4105.9001 − 1.0078). In our MS spectroscopy experiment, we used the toxin HWTX-I as a relative molecular mass standard, whose monoisotopic mass is 3748.6890. After strictly calibrating with this peptide, we found that the relative molecular mass of HWTX-IV determined by MS is 4104.7398. This is close with the amidated peptide. Then the MS result gave a simple and convenient biochemistry evidence for the C terminal amidation.

4. Discussion

The spider venoms often contain many active peptides such as neurotoxins, lectins, inhibitors to enzyme, etc. These peptides are very important for spiders' hunting and defending. During the long history of spider evolution, the peptides evolved into different structures and functions, which were separately subject to mammalians and insects. In our previous work, from the specie of *S. huwena*, we separated and characterized many peptides, and by using the NMR technique, the 3D structures of these peptides were elucidated. Among them, HWTX-I, HWTX-IV, SHL-I display ICK motif, although their functions are separately high-voltage calcium channel blocker, TTX-sensitive sodium channel blocker and erythrocytes aggregation activity. ICK motif was first reported by Pallaghy in 1994 (Pallaghy et al., 1994), which incorporates a small triple-stranded antiparallel β-sheet with a topology of +2x, −1 and a cystine knot formed by three disulfide bonds with a linkage pattern of I–IV, II–V, III–VI. The former three toxins all accorded these characteristics and exhibited typical ICK motif. HWTX-II is an insecticidal toxin which has a different structure motif, and HWTX-VII, HWTX-IIIa, HWTX-V are also insecticidal toxins though their 3D structures have not yet been resolved. In this article, we cloned and sequenced seven peptides' cDNAs and found that the lengths of their cDNAs are all between 400 and 500 bp, and all the seven cDNAs encoded peptide precursors about 80 amino acid residues. The precursors all included a signal peptide, a mature peptide and an intervening sequence. These cDNA structures were similar with most other spider cDNA sequences searched from the Genebank database (Penaforte et al., 2000; Diniz et al., 1993). We found that cDNAs of scorpion toxins usually just contain a signal peptide and a mature peptide, but without a propeptide region (Dai et al., 2000; Vazquez et al., 1995). However, the cDNA sequences of cone snail toxins contain intervening peptides and the prepro region with high sequence similarity (Olivera and Cruz, 2001). The phenomenon that some toxins from the three different animals above mentioned share similar functions but with different kind of cDNA organizations might give us some important insights to analyze the toxin evolution in these animals.

The signal sequences of the seven peptides are all 21 or 24 amino acid long. They have all the characteristics of a typical signal peptide (Perlman and Halvorson, 1983), including a hydrophobic core rich in valine and leucine, a consensus cleavage point for a signal peptidase. The intervening propeptide regions of these toxins were between 27 and 29 amino acid. This segment is rich in glutamate residues as other spider toxins such as *Phoneutria nigriventer* and *Agelenopsis aperta* (Penaforte et al., 2000; Santos et al., 1992). From the analysis of the acid residues and basic residues of pro region and mature region, we find that the mature peptides are rich in basic residues except that SHL-I, which is a neutral molecular. In exploration of the relationship of structure and function of scorpion toxins, people found that the toxin activity was related with the basic residues such as lysine and arginine. Then the pro region rich in acid residues might have two main potential usages. One is to neutralize the basic residues of mature peptide so as to stabilize the toxin precursor in the cell cytoplasm; Second, the acid residues might form a precursor structure that cover the basic mature functional side, then it could prevent the toxins interact with other molecule before they are transported to venom glands. Except for HWTX-VII, which had an additional leucine, the pro regions of the seven peptides all ended in a sequence of EER, which might be the cleavage signal of propeptide processing enzyme. During purification and sequence of peptides from *S. huwena*, we found many peptides have polymorphism phenomenon such as HWTX-II with HWTX-VIII (Fig. 4), the later is just short of a leucine residue in the N terminal of peptide; SHL-I with SHL-Ia, HWTX-III with HWTX-IIIb (unpublished data), both of their later analogies are short of a tryptophan in the C terminal. So the cDNAs of HWTX-II, SHL also might be the cDNA sequences of HWTX-VIII, SHL-Ia, The reason that brought forth different peptides from the same cDNA sequence might be the ambiguous cleavage site of precursor processing enzyme.

Different from other toxins, the translated protein sequence from HWTX-IV cDNA ended with glycine and lysine just before the stop coden. Which was an amidation signal of C terminal. In order to give further evidence, we did the MS experiment using the ABI Voyager-DE™ STR MALDI-TOF MS-spectrometer. The MS results gave a simple and convenient evidence to prove the amidation toxins. Amidation was often found in toxins from spiders, scorpion and conotoxin. In studying the pro region of ω-conotoxin, Goldenberg found that C terminal additional glycine could improve the refolding efficiency, thus it might involve in the peptide folding (Price-Carter et al., 1996). The additional basic residue lysine or Arg-Arg was also found in C terminal of HWTX-V, HWTX-IIIa and SHL-I, which was also post-translational processing result found in some other toxins. During investigation toxins from the scorpion *Androlonus australis*, Bougis once found toxin precursors that activated on insects were not processed at their C terminal, however, toxins activated on mammalians contained either a single arginine or Gly-Arg dipeptides at the C-terminus which got removed (Leisy et al., 1996). This is obviously not a universal rule in the world of toxins, for that in our toxin precursors, HWTX-IIIa and HWTX-V could cause cockroaches paralysis. Further experiment needs to be done to investigate the exact function of C terminal processing and amidation.

According to the alignment of the protein sequences of the seven peptides' precursors, we can get further interesting information (Figs. 3 and 4). The prepro regions of HWTX-I, HWTX-IV, SHL-I, HWTX-IIIa have high similarity (over 75%), although their coding sequences have almost no similarity and their biological functions are different.

	Signal peptide	Intervening propeptide	Identity%
HWTX-I	---MRASMFLALAGLVLLFVVCYA	SESEEKEFPRELLFKFFAVD-D-FKGEER	100
HWTX-IV	MVN·K·····	·····SN···SSVL···DN-S·····	76.9
SHL-I	---·KT···T·T·····	·····K···SSI··A·S·-··V···	77.2
HWTX-IIIa	MVN·K·····TF·····	·····K·M·SSI····N·-··Q···	75.0
HWTX-V	---·KSIV·V··F··A··A···S·	··DAH··LLK·VVRAMVVDKT·AVQA···	40.0

	Mature peptide	Length
HWTX-I	ACKGVFDACTPGKNECCPNR—VCSDKHKWCKWKL	81
HWTX-IV	E·LEI·K··NPSNDQ··KSSKL···R·TR···YQIG*K*	89
SHL-I	G·L·DKCDYNNGCCSGYVCSRTWKWCVLAGPWR*R*	81
HWTX-IIIa	D·A·YMRE·KEKLCCSGYVCSSRRKWCVLPAPWR*R*	87
HWTX-V	E·RWYLGG·SQDGDC·KHLQ-CHSNYE-··V·DGTFSK*	86

Fig. 3. The alignment of amino acid sequence deduced for the cDNA sequences of HWTX-I, IIIa, IV, V and SHL-I. A point indicates that the amino acid is identical to the one of HWTX-I. When there is no similarity the corresponding change is indicated. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. The previously determined amino acid sequence of HWTX-I, IIIa, IV, V and SHL-I is underlined. The hydrophobic region (signal peptide) is shown by open box. The extra C terminal residue of HWTX-IV, IIIa, V and SHL-I is designated with asterisks. The identity of the prepro region is annotated in the end of prepro region.

HWTX-V also has a certain similarity with the former four toxins (40%), but with low similarity with HWTX-II. It appeared that HWTX-I, HWTX-IV, SHL-I, HWTX-IIIa and HWTX-V evolved from the same ancestor. At the same time, the prepro region of HWTX-II and HWTX-VII have high similarity (over 95.8%), as well as their coding sequences which also share a high similarity (about 81.1%), but the prepro regions of the two toxins have low similarity with that of the former five peptides, then it seemed that HWTX-II and HWTX-VII came from another same progenitor. Interestingly, by the NMR technique, the structures of HWTX-I, II, IV and SHL-I have been elucidated. The 3D structure of HWTX-I, HWTX-IV and SHL-I are all ICK motif, but the molecular of HWTX-II

	Signal peptide	Intervening propeptide	Identity%
HWTX-II	MKVTLIAILTCAAVLVLHTTA	AEELEAESQLMEVGMPDTELAAVDEER-	100
HWTX-VII	·····-·····L		95.8

	Mature peptide	Length
HWTX-II	LFECSFSCEIEKEGDKPCKKKKCKGGWKCKFNMCVKV	85
HWTX-VII	-····I·····K·E-S··P·····	85
HWTX-VIII	-·····	36

Fig. 4. Comparison of amino acid sequences of HWTX-II, VII, VIII. A point indicates that the amino acid is identical to the one of HWTX-II. When there is no similarity the corresponding change is indicated. Gaps (dashes) were introduced to maximize the deduced polypeptide sequence similarities. The previously determined amino acid sequence of HWTX-II, VII and VIII is underlined. The hydrophobic region (signal peptide) the intervening region (pro-peptide) and the mature toxin are show by open boxes. HWTX-VIII is shown only of its mature peptide sequence.

exhibits a different motif. Combined the prepro region similarity and the structure similarity, it indicates that HWTX-I, HWTX-IV and SHL-I, HWTX-IIIa, HWTX-V belong to the same superfamily, however, HWTX-II and HWTX-VII belong to another superfamily (Actually, HWTX-II, HWTX-VII belong to a same family which is included in a superfamily different from the former one). HWTX-V share a relative low similarity with HWTX-I, IV, IIIa and SHL-I, but it still shows homology with the four toxins especially its pro region share 57.1% similarity with HWTX-I. Then we classified HWTX-V into the superfamily of the former four toxins. The classification based on the prepro region similarity was as well as conotoxin, in which the prepro regions of the toxins from the same superfamily had high similarity. In this article, we also analyzed the similarity of different regions of the precursors of the former five toxins, and found that different segments of the peptide precursors evolved at extremely different rates. Signal sequences of peptides within the same superfamily have the highest similarity (about 80%, excluded HWTX-V), also the intervening pro regions have a lower similarity about 70% (exluded HWTX-V), however, the mature toxin regions are hypermutated (<40% similarity). This phenomenon of the spider cDNA precursors is also been found in the atracotoxins from the Australian funnel-web spider by Wang in 2001(Wang et al., 2001), furthermore, we also found that the signal sequence elucidated by Wang for the omega-atracotoxin-2 is very similar to the signal sequence of huwentoxins. It appears that in the long history of evolutionary time, under the pressure of environments, the peptides in the spider venom generated molecular diversity to fit for different environments for defending and hunting by hypermutating the mature toxin region, while conserving the basic structure framework. Moreover, the signal peptide was related to the orientation of peptides, its residues were very important for the sort of molecule, so it became the most conserved region.

Acknowledgements

We thank Professor Chengwu Chi and Dr Chunguang Wang in Shanghai Institute of Biochemistry, Academia Sinica for their help in this work. The work was supported by National Natural Science Foundation of China, under contract No. 30170193 39990600.

References

Craik, D.J., Daly, N.L., Waine, C., 2001. The cystine knot motif in toxins and implications for drug design. Toxicon 39, 43–60.

Dai, L., Wu, J.J., Gu, Y.H., Lan, Z.D., Ling, M.H., Chi, C.W., 2000. Genomic organization of three novel toxins from the scorpion *Buthus martensi* Karsch that are active on potassium channel. Biochem. J. 346, 805–809.

Diniz, M.R.V., Paine, M.J.I., Diniz, C.R., Theakston, R.D.G., Crampton, J.M., 1993. Sequence of the cDNA coding for the lethal neurotoxin Tx1 from the Brazilian armed spider *Phoneutria nigriventer* predicts the synthesis and processing of a preprotoxin. J. Biol. Chem. 268, 15340–15342.

Leisy, D.J., Mattson, J.D., Quistad, G.B., Kramer, S.J., Beek, N.V., Tsai, L.W., Enderlin, F.E., Woodworth, A.R., Digan, M.E., 1996. Molecular cloning and sequencing of cDNAs encoding insecticidal peptides from the primitive hunting spider, *Plectreurys tristis* (Simon). Insect. Biochem. Mol. Biol. 26, 411–417.

Li, M., Zhou, Z., Liang, S.P., 2001. Sequence and cloning of huwentoxin-I cDNA. Chin. J. Biochem. Mol. Biol. 17, 51–55.

Liang, S.P., Pan, X., 1995. A lectin-like peptide isolated from the venom of the Chinese bird spider *Selenocosmia huwena*. Toxicon 33, 875–882.

Liang, S.P., Zhang, D.Y., Pan, X., Chen, Q., Zhou, P.A., 1993. Properties and amino acid sequence of huwentoxin-I, a neurotoxin purified from the venom of the Chinese bird spider *Selenocosmia huwena*. Toxicon 31, 969–978.

Lu, S.Y., Liang, S.P., Gu, X.C., 1999. Three-dimensional Structure of *Selenocosmia huwena* Lectin-I (SHL-I) from the venom of the spider *Selenocosmia huwena* by 2D-NMR. J. Protein Chem. 18, 609–617.

Olivera, B.M., Cruz, L.J., 2001. Conotoxins, in retrospect. Toxicon 39, 7–14.

Pallaghy, P.K., Nielsen, K.J., Craik, D.J., Norton, R.S., 1994. A common structural motif Incorporating a cystine knot and a triple-stranded β-sheet in toxic and inhibitory polypeptides. Protein Sci. 3, 1833–1839.

Penaforte, C.L., Prodo, V.F., Prado, M.A.M., Romano-Silva, M.A., Guimaraes, P.E.M., Marco, L.D., Gomez, M.V., Kalapothakis, E., 2000. Molecular cloning of cDNAs encoding insecticidal neurotoxic peptides from the spider *Phoneutria nigriventer*. Toxicon 38, 1443–1449.

Peng, K., Chen, X.D., Liang, S.P., 2001. The effect of huwentoxin-I on Ca^{2+} channels in differentiated NG108-15 cells, a patch-clamp study. Toxicon 39, 491–498.

Peng, K., Shu, Q., Liang, S.P., 2002. Functional and solution structure of huwentoxin-IV, a potential neuronal TTX-sensitive sodium channel antagnisit from the Chinese bird spider *Selenocosmia huwena*. J. Biol. Chem. 277, 47564–47572.

Perlman, D., Halvorson, H.O., 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. 167, 391–409.

Price-Carter, M., Gray, W.R., Goldenberg, D.P., 1996. Folding of ω-conotoxins. 2. Influence of precursor sequences and protein disulfide isomerase. Biochemistry 35, 15547–15557.

Qu, Y.X., Liang, S.P., Ding, J.Z., Ma, L.B., Zhang, R.J., Gu, X.C., 1995. Proton nuclear magnetic resonance studies on huwentoxin-I from the venom of the spider *Selenocosmia huwena*: 1. Sequence-specific ^{1}H-NMR assignments. J. Protein Chem. 14, 549–557.

Santos, A.D., Imperial, J.S., Chaudhary, T., Beavis, R.C., Chait, B.T., Hunsperger, J.P., Olivera, B.M., Adams, M.E., Hillyard, D.R., 1992. Heterodimeric structure of the spider toxin ω-agatoxin IA revealed by precursor analysis and mass spectrometry. J. Biol. Chem. 267, 20701–20705.

Shu, Q., Lu, S.Y., Gu, X.C., Liang, S.P., 2002. The structure of spider toxin huwentoxin-II with unique disulfide linkage: evidence for structural evolution. Protein Sci. 11, 245–252.

Vazquez, A., Tapia, J.V., Eliason, W.K., Martin, B.M., Lebreton, F., Delepierre, M., Possani, L.D., Becerril, B., 1995. Cloning and characterization of the cDNAs encoding Na^+ channel-specific toxins 1 and 2 of the scorpion *Centruroides noxius* Hoffmann. Toxicon 33, 1161–1170.

Wang, X.H., Connor, M., Wilson, D., Wilson, H.I., Nicholson, G.M., Smith, R., Shaw, D., Machay, J.P., Alewood, P.F., Christie, M.J., King, G.F., 2001. Discovery and structure of a potent and highly specific blocker of insect calcium channels. J. Biol. Chem. 276, 40306–40312.

Zhang, P.F., Chen, P., Hu, W.J., Liang, S.P., 2003. Huwentoxin-V, a novel insecticidal peptide toxin from the spider *Selenocosmia huwena*, and a natural mutant of the toxin: indicates the key amino acid residues related to the biological activity. Toxicon 42, 15–20.