Cloning of a galactose-binding lectin from the venom of *Trimeresurus stejnegeri*

Qiang XU, Xiang-Fu WU, Qi-Chang XIA and Ke-Yi WANG¹

Shanghai Institute of Biochemistry, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China

A galactose-binding lectin isolated from the venom of *Trimeresurus stejnegeri* is a homodimer C-type lectin. The cloned cDNA encoding the monomer of *Trimeresurus stejnegeri* lectin (TSL) was sequenced and found to contain a 5'-end non-coding region, a sequence which encodes 135 amino acids, including a typical 23 amino acid signal peptide followed by the mature protein sequence, a 3'-end non-coding region, a polyadenylation signal, and a poly(A) region. To completely characterize the deduced amino acid sequence, on-line HPLC–MS and tandem MS were used to analyse the intact monomer and its proteolytic peptides. A modified peptide fragment was also putatively identified by HPLC–MS analysis. The deduced amino acid

sequence was found to contain a carbohydrate-recognition domain homologous with those of some known C-type animal lectins. Thus TSL belongs to group VII of the C-type animal lectins as classified by Drickamer [(1993) *Prog. Nucleic Acid Res. Mol. Biol.* **45**, 207–232]. At present, a number of C-type lectins have been purified from snake venom, but most of them have been characterized only at the protein level. To our knowledge, this is the first known cDNA sequence of a true C-type lectin from snake venom.

Key words: carbohydrate-recognition domain, cDNA sequence, mass spectrometry.

INTRODUCTION

Snake venoms contain a variety of proteins, some of which affect vertebrate blood coagulation and the platelet aggregation system. Several of these venom proteins, with molecular masses of approx. 30 kDa, consist of two similar disulphide-linked monomers containing a C-type lectin motif, such as the rattlesnake lectin from *Crotalus atrox* venom [1], the *Lachesis muta stenophrys* lectin [2], Factor IX/X-binding protein [3], albo-aggregin-B [4] and echicetin [5]. However, most of these proteins do not possess saccharide-binding activity and thus are not true C-type lectins. In a previous study, we purified a galactose-binding C-type lectin from the venom of the snake *Trimeresurus stejnegeri* designated as TSL [6]. The results of SDS/PAGE, under non-reducing and reducing conditions, suggested that the major constituent of the lectin was a disulphide-linked homodimer of (approx.) 20 kDa monomers with a pI of 8.5. TSL can agglutinate rabbit red cells through a Ca²⁺-dependent mechanism. The interaction of TSL protein with ¹²⁵I-labelled asialo- α_1 -acid glycoprotein was studied to ascertain the saccharide-binding conditions [7]. The results showed that the saccharide-binding activity of the lectin remained almost unchanged over the pH range 5–10. Among the monosaccharides assayed, galactose was found to be the best inhibitor of binding, with the most important structural feature being the arrangement of the 3- and 4-hydroxy groups.

Proteins that consist entirely of the carbohydrate-recognition domains (CRDs) have been known for a period of time. Only recently have many of these proteins been isolated from snake venoms and characterized at the protein level. Moreover, it has been impossible to be sure that these proteins were not proteolytic

fragments of larger precursors until their cDNA sequences became available [8].

Here we report the cloning, sequencing and deduced amino acid sequence of the cDNA encoding the lectin from *T. stejnegeri* venom. The sequence of the N-terminal 30 amino acids was determined directly by automated Edman degradation of the purified protein. The deduced amino acid sequence, consisting of 135 amino acid residues, was confirmed by mapping of proteolytic peptides that were analysed by on-line HPLC–electrospray ionization MS (HPLC–ESI–MS) and tandem MS. When compared with other C-type CRDs, TSL was shown to be composed of covalently linked CRDs belonging to group VII of the C-type animal lectins as classified by Drickamer [9].

MATERIALS AND METHODS

Purification of TSL and separation of the monomers

TSL was purified from freeze-dried *Trimeresurus stejnegeri* venom (Huangshan, Anhui, China) as described by Liang and Wang [6] with the following modification: acid-treated Sepharose 6B was used in the first step of the purification, and reverse-phase HPLC was used in the second step. Purified TSL was freeze-dried and stored at –20 °C.

Purified TSL (0.1 mg) dissolved in 60 μ l of 0.1 M NH₄HCO₃ (pH 8.3) was reduced by incubation (42 °C for 1 h) with 8 μ l of 100 mM dithiothreitol and then S-carboxamidomethylated with 16 μ l of 100 mM iodoacetamide (37 °C for 35 min in the dark). Finally, the S-carboxamidomethylated-TSL (CAM-TSL) was isolated free of reagents by reverse-phase HPLC with an acetonitrile/water/trifluoroacetic acid system as described below.

Abbreviations used: CAM-, S-carboxamidomethylated; CRD, carbohydrate-recognition domain; ESI, electrospray ionization; LA-PCR, ligation-mediated-anchor-PCR; MBP, mannose-binding protein; TSL, *Trimeresurus stejnegeri* lectin.

¹ To whom correspondence should be addressed (e-mail kyw0717@sunm.shnc.ac.cn).

The cDNA sequence of *Trimeresurus stejnegeri* lectin will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF119097.

Name	5' SEQUENCE 3'
N1	AGCTGTTGTAAGTACTGACAGCCTGCCNATGAA
N2	CGCTCGAGCCTATGAAYGGHATGTGYTAYAAAT
T1	CGCTCGAGT ₁₆
T2	GCTAGAATTGCACTGGCAGAGAA
ANCHOR	pAGCTCGAG GTCGTGACTGGGAAAACCTGGCGC-ddA
AP	CGCCAGGGTTTTCCAGTCACGAC
R1	TTGGACTCGCACCTGATCATTCCA
R2	CCCTGTAAGGGACACAAGCTCAACGCA

Figure 1 Sequences, designation and relative position of the oligonucleotides used in this study

Primer sequences (top panel). Primer N1 is a 'guesser', and primer N2 is degenerate. Primers T2, R1 and R2 are designed according to the results of partial cDNA sequencing. In addition, an *XhoI* restriction site was included in primers N2 and T1 to facilitate the cloning of the PCR product. The 34 nt anchor contains an altered M13 sequence including an *XhoI* restriction site. The primer AP is complementary to a sequence within the anchor. Diagram of oligonucleotide positions relative to both ends of TSL mRNA (bottom panel). The thick line designates coding sequences and the thin line untranslated sequences. Arrows indicate the location of synthetic primers relative to TSL mRNA. Sense and antisense oligonucleotides are represented above and below respectively.

N-terminal sequence analysis of TSL

The HPLC-purified CAM-monomer was sequenced by automated Edman degradation using an Applied Biosystems (Foster City, CA, U.S.A.) model 491A protein sequencer equipped with a 140C microgradient system.

Amplification of TSL cDNA and determination of a partial cDNA sequence

Total RNA was prepared from *T. stejnegeri* venom glands by the method of Chomczynski and Sacchi [10] using TRIzol® reagent (Life Technologies). Single-stranded cDNA was prepared from the mRNA contained in total RNA (5 µg) using the SuperScript Preamplification System (Life Technologies) with the primer T1 (Figure 1). The single-stranded cDNA was then amplified by PCR using N1 (sense) and T1 (antisense) as primers. PCR was carried out with the Expand High Fidelity PCR system (Boehringer Mannheim) and performed in a Perkin-Elmer GeneAmp PCR thermocycler, model 2400, with the following program: cycle 1, 94 °C for 5 min; cycles 2–7, 94 °C for 45 s, 50 °C for 30 s, 72 °C for 1 min; cycles 8–41, 94 °C for 45 s, 52 °C for 30 s, 72 °C for 1 min; cycle 42, 72 °C for 7 min. Primary PCR products were re-amplified using primers T1 and N2 to introduce a *XhoI* restriction site into the 5'- and 3'-ends of the cDNA under the following conditions: cycle 1, 94 °C for 5 min; cycles 2–6, 94 °C for 45 s, 52 °C for 30 s, 72 °C for 1 min; cycles 7–41, each 94 °C for 45 s, 55 °C for 30 s, 72 °C for 1 min; cycle 42, 72 °C for 7 min.

Amplification of 5'-end of TSL mRNA

The sequence of 5'-end of the cDNA was determined by the method of single-strand ligation-mediated-anchor-PCR (LA-PCR) [11,12]. In brief, the oligonucleotide T2 was used to prime the first strand cDNA synthesis. T4 RNA ligase (Boehringer Mannheim) was used to covalently join the 5'-phosphorylated-,

3'-blocked-anchor oligonucleotide to the reverse-transcribed first-strand cDNA. The primers were removed by QIAquick PCR Purification Kit (Qiagen). The products of the single-strand ligation reaction were amplified in two successive reactions, PCR1 and PCR2. PCR1 amplification was carried out with the primers AP and R1 under the following conditions: cycle 1, 94 °C for 5 min, 60 °C for 30 s, 72 °C for 10 min; cycles 2–36, 94 °C for 45 s, 60 °C for 30 s, 72 °C for 1 min; cycle 37, 72 °C for 7 min. One µl of the amplified products was subsequently used in a nested amplification reaction (PCR2) with the primers R2 and AP under the following conditions: cycle 1, 94 °C for 5 min; cycles 2–6, 94 °C for 45 s, 62 °C for 30 s, 72 °C for 1 min; cycles 7–31, 94 °C for 45 s, 60 °C for 30 s, 72 °C for 1 min; cycle 32, 72 °C for 7 min.

DNA sequence analysis

The amplified PCR products were purified by low-melting-point agarose gel (Bio-Rad) electrophoresis and then cloned into a pGEM-T (Promega) or pBluescript II SK⁺ (Stratagene) vector. *Escherichia coli* host cells, ElectroMAX DH12S (Life Technologies) were transformed with the recombinant plasmids by electroporation. DNA sequencing was performed by the dideoxy chain-termination method using the Sequenase (version 2.0) DNA sequencing kit (Amersham Life Science) and a nucleotide sequencer (model 373A, Applied Biosystems).

ESI-MS of intact TSL monomers

A Hewlett-Packard Model 1100 HPLC system equipped with a 2.1 mm × 30 mm Aquapore RP300 C₈ column (Applied Biosystems, Santa Clara, CA, U.S.A.) was used to purify 10 µg of reduced and CAM- monomers. Solvent A was 0.02 % (v/v) trifluoroacetic acid in water, and solvent B was 0.02 % (v/v) trifluoroacetic acid in acetonitrile/water (50:1, v/v). Proteins were eluted with a linear gradient from 0–100 % (v/v) B over 15 min at a flow rate of 0.2 ml/min. Absorbance at 214 nm was monitored. The HPLC eluate was subjected to ESI-MS using a Finnigan Mat (San Jose, CA, U.S.A.) LCQ™ ion-trap mass spectrometer equipped with an ESI source.

Confirmation of the deduced sequence by peptide mass fingerprinting

Purified CAM-TSL (0.1 mg) was digested with the sequencing grade endoproteinase Glu-C (Sigma) at a ratio of 50:1 (w/w) in 80 µl of 30 mM NH₄HCO₃, pH 7.8, at 37 °C for 4 h. The digestion was continued for another 20 h with an addition of the same amount of endoproteinase Glu-C. Purified CAM-TSL was also digested with the endoproteinase Lys-C under the same conditions. In each case a 30 µl aliquot of the resulting digest mixture was analysed by on-line reverse-phase HPLC-ESI-MS and tandem MS.

Reverse-phase HPLC was performed as above except the linear gradient was 0–50 % (v/v) solvent B over 90 min at a flow rate of 0.2 ml/min. The MS conditions were as follows: ionization spray voltage, 4.25 kV; capillary temperature, 200 °C; capillary voltage, 14.64 kV. Tandem MS was performed in the data-dependent mode.

Computer analysis

Multiple sequence alignment was performed using the CLUSTAL W program (version 1.7, June 1997) [13]. As the CLUSTAL W program did not produce exact values of percentage similarity, pair-wise sequence alignments were performed on a SGI workstation using the Wisconsin sequence

analysis package (version 9.1, September 1997, Genetics Computer Group, Madison, WI, U.S.A.). Moreover, some identical alignment parameters such as gap penalties and weight matrix were also reset, in agreement with the CLUSTAL W program.

RESULTS AND DISCUSSION

Cloning and sequence analysis of the cDNA encoding TSL

N-terminal analysis of the HPLC-purified CAM-TSL yielded the following sequence:

SCCTXDSLPMNGMCYKIFDEPKTWEDAEMF

with an unidentified residue Xaa⁵. Generally, phosphorylated or glycosylated amino acid residues are not identified by regular Edman degradation. Therefore we suspected Xaa⁵ to be a threonine residue. Edman degradation of CAM-TSL yielded only a single N-terminal amino acid sequence, which was in agreement with the result of reducing SDS/PAGE, suggesting that TSL is a homodimer.

On the basis of the experimentally determined N-terminal sequence of TSL, we designed a 'guessamer' N1 and degenerate primer N2 (Figure 1), encoding the N-terminal peptides Ser¹-Asn¹¹ and Pro⁹-Ile¹⁷ respectively. For amplification of the partial coding and 3'-end non-coding regions, two successive PCR reactions were performed with the primers N1 and N2 successively paired with the corresponding antisense primer T1. The major PCR product was 567 bp.

After sequencing of this fragment, antisense primers R1, R2 and T2 were designed for LA-PCR. The single-stranded cDNA reverse transcription was first primed using a TSL gene-specific

```

1          CACAAGCTAAAAGCCTGGAGTTGCCCTGAG
32  CAGATTGCTACCTGTGGAGGCCAAAGGACAGTTCTCTCGACGGGAAGGAAGACC
92  ATGGGGCGATTACTCTTCGTGAGCTTCGGCTTGTGGTCTGCTCTCCCTGAGTGGA
-23  M G R F I F V S F G L L V V F L S L S G
152 GCCAAGGCAGTTGTTGTACCAATGATTGCTGCCCATGAATGGGATGTGTACAAGATC
-3   A K G S . . C . . T N D S . . P M N G M C Y K I
212 TTCGATGAACCGAAGACCTGGGAGGATGCAGAGATGTTCTGCAGGAATACAAGCAGGC
18   F D E P K T W E D A E M F C R K Y K P G
272 TGCCACCTTGCTCCTTTCACCGCCTTGCAATCACTTGCATTGCTGAGTACATCTCT
38   C H L A S F H R L A E S L D I A E Y I S
332 GACTATCACAAAGCTCAGGCTGAAGTTGGATTGGACTGTTGGATAGGAAGAAGATTTC
58   D Y H K R Q A E V W I G L L D R K K D F
392 TCCTGGGAGTGGACGACAGATCTTGCACTGACTACCTCACTGGGACAAAATCAACCC
78   S W E W T D R S C T D Y L N W D K N Q P
452 GATCACTACAAGGACAAAGAATTTCGCGTTGAGCTTGTCCTTACAGGGTATCACCGG
98   D H Y K D K E F C V E L V S L T G Y H R
512 TGGAAATGATCAGGTGTGCGAGTCCAAGAATTCTTTCTCTGCCAGTGCAAAATCTAGCGC
118  W N D Q V C E S K N S F L C Q C K F ***
572 ACGTCGCTCTTCTCCCGCTCCGGAAGGAGAGAAGCCCGAGCGATGTCTGGAGAAGCAA
632 CGAAGCTCTGCCACAATCTCTGCTCTGCCCTTTCGCTCAAGGGAGGCTCTCCGTAGCT
692 GCGATCTGGTTTTGCTGCTCTGACGGGCCAGAAGCTCCAAATAAATCTGCCTAACCGGGA
752 AAAAAAAAAA -poly(A)

```

Figure 2 Complete nucleotide and deduced amino acid sequence of TSL monomer

Numbering of the nucleotides is in the 5' to 3' direction. Beneath the nucleotide sequence is the deduced amino acid sequence encoded by the open reading frame between nucleotides 92 and 565. The broken lines indicate the positions of peptides obtained by direct sequencing of TSL monomer. The putative signal sequence is printed in *italics*. The probable polyadenylation signal is printed in **bold** type. The stop codon is starred (*). The underlined nucleotide sequences correspond to primers used in the LA-PCR.

Figure 3 ESI-MS of reduced (a) and CAM-TSL monomers (b)

antisense primer T2. This step permitted the cDNA conversion of specific mRNA, and maximized the potential for complete extension to the 5'-end of the mRNA. Following cDNA synthesis, the purified first-strand products were amplified by PCR using the primers R1 and AP. To increase the specificity of the primary PCR, a nested PCR was carried out with the primers R2 and AP. The sequencing of the nested PCR products revealed the 5'-end non-coding region and a signal-peptide coding sequence of TSL. Combining the results from sequences extending into both the 5'- and 3'-ends allowed us to elucidate the complete cDNA sequence of TSL. The 750 bp sequence contained a 5'-untranslated region of 91 bp, an open reading frame of 474 bp, a stop codon, and a 3'-end non-coding region of 182 bp that included a poly(A) signal site followed by a poly(A) tail (Figure 2).

The deduced amino acid sequence contained a very hydrophobic sequence of 23 residues, corresponding to a signal peptide that is similar to the one described for Factor IX/X-binding protein [14], bothrojaracin [15], and convulxin [16]. In addition, the unidentified residue at position 5 from the N-terminal sequencing by Edman degradation was identified as Asn by sequencing of the cDNA clones. Since the lectin has a potential Asn-X-Thr/Ser sequence, which is known as a sequon for the attachment of carbohydrate to asparagine, the residue Asn⁵ is likely to be linked to a carbohydrate.

Confirmation of amino acid sequences by HPLC-ESI-MS

To better characterize the deduced amino acid sequence, the reduced and CAM-intact monomers were analysed by HPLC-ESI-MS. Mass determination of TSL monomer by ESI-MS was repeated several times, and two typical spectra are presented in Figures 3(a) and 3(b). The molecular mass of reduced TSL (17924.2 ± 2.4 Da) was increased by 571 Da for CAM-TSL

Table 1 HPLC-ESI-MS analysis of proteolytic products of CAM-TSL

All cysteine residues have been converted into CAM-cysteine. Prefixes S and K denote peptides generated by digestion with endoproteinase Glu-C and Lys-C respectively. The products are numbered according to the order of the proteolytic peptide fragments separated by HPLC.

Number	Position	Theoretical mass (Da)	Observed mass (Da)	Retention time (min)
S1	59–65	931.0	930.9	17.01
S2	55–58	496.5	496.1	20.09
S3	81–88	1040.0	1039.9	22.29
S4	43–48	771.9	772.0	25.68
S5	105–108	553.6	553.2	27.00
S6	49–54	646.7	646.3	27.50
S7	29–42	1755.0	1754.7	27.76
S8	89–104	2093.2	2093.7	29.72
S9	81–104	3115.2	3115.3	32.21
S10	125–135	1418.5	1418.6	32.53
S11	109–124	1977.1	1976.9	33.45
S12	7–28	2662.0	2662.6	38.57
S13	7–25	2346.7	2346.6	39.02
S14	66–80	1892.2	1892.0	40.23
K1	95–103	1143.2	1143.7	3.58
K2	95–101	900.3	900.7	3.72
K3	23–33	1472.6	1472.0	29.58
K4	127–135	1203.3	1202.6	30.82
K5	62–75	1712.0	1711.9	36.33
K6	17–33	2202.4	2201.9	38.82
K7	95–126	3983.3	3982.9	42.70
K8	104–126	2857.1	2856.8	44.34
K9	76–94	2524.6	2524.9	45.54
K10	34–61	3321.7	3322.0	48.21

(18495.3 ± 1.5 Da). This result was consistent with the presence of ten half-cystine residues in the derivatized TSL, which corresponds to the ten cysteine residues in the deduced amino acid sequence.

The mass spectrum of the peptide mixture resulting from enzymic digestion of a protein can provide a mass fingerprint, which is specific enough to identify the protein uniquely [17]. Therefore, the endoproteinase Glu-C and Lys-C digests of CAM-TSL were separated by reverse-phase HPLC on-line with an ion-trap mass spectrometer. The experimental mass values were then compared with peptide mass values, calculated by applying the enzyme cleavage rule to the deduced amino acid sequence (Table 1). The assignment of every peak was further confirmed by tandem MS sequencing of the peptides (results not shown). In general, the observed masses matched those deduced from cDNA sequence. However, peaks S4 and S7 showed values in mass inconsistent with the predicted proteolytic fragments. These corresponded to the peptides Phe⁴³-Glu⁴⁸ and Met²⁹-Ser⁴², possibly representing a non-specifically cleaved peptide bond between residues 42 and 43. This result was confirmed by tandem MS. We also found that the deduced molecular masses of the unmodified peptides Ser¹-Asp⁶ and Ser¹-Lys¹⁶ were not compatible with corresponding peaks in the HPLC-ESI-MS profile of the proteolytic products. A possible explanation was that the two peptide fragments were modified. In addition, close examination of the ESI-MS spectrum of reduced monomers showed some heterogeneity in the mass range. These results suggested that there might be heterogeneity due to glycosylation of the residue Asn⁵, supporting the assumption that TSL is a glycoprotein (R. Zeng, Q. Xu, K.-Y. Wang and Q.-C. Xia, unpublished work). Of course it should be possible to locate the precise modification site(s) by sequencing the modified and

TSL	SCCT--NDSL	PMNGMCYKIF	DEPKTWEDAE	MFCRKYPK--	(1-36)
LMSL	NNCP--QDWL	PMNGLCYKIF	DEQKAWEDAE	MFCRKYPK--	(1-36)
RSL	NNCP--LDWL	PMNGLCYKIF	NQLKTWEDAE	MFCRKYPK--	(1-36)
ASGPR	ICCP--INWV	EYEGSCYWF	SSVKPWTEAD	KYQLEN---	(151-185)
HLIT	ISCP--EGTN	AYRSYCYFYN	EDRETWVAD	LYCQNMNS--	(12-47)
SUL	GCCP--TFWT	SFGSNCYRFF	AVSLTWAEGE	QFCQSFSVPS	(1-38)
ABL	--CPGNLDWQ	EYDGHYCWAS	TYQVRWNAQ	LACQTVHP--	(2-37)
MBP	-----	-SGKKFFVTN	HERMFFSKVK	ALCSELR---	(107-132)
				*	
TSL	-G-----	CHL ASFHRLAESL	DIAEYISDYH	KRQ--AEVWI	(37-68)
LMSL	-G-----	CHL ASFHRYGESL	EIAEYISDYH	KGQ--AEVWI	(37-68)
RSL	-G-----	CHL ASFHRYGESL	EIAEYISDYH	KGQ--ENVWI	(37-68)
ASGPR	-----	AHL VVVTSWEEQR	FVQGHMGP--	-----LNTWI	(186-211)
HLIT	-G-----	N-L VSVLTQAEQA	FVASLIKESG	TDD--ENVWI	(48-78)
SUL	RGDIDSIGHL	VSIHSETEQN	FVYHYFETST	KDDTTPMWL	(39-78)
ABL	-G-----	AYL ATIQSQLENA	FISETVSN--	-----NRLWI	(38-64)
MBP	-G-----	T VAIPNAEEN	KAIQEVAK--	-----TSAFL	(133-157)
				*	
TSL	GLLDRKKDFS	WEWTDRS--C	TDYLNWDKNQ	PDHYKD----	(69-102)
LMSL	GLWDKKKDFS	WEWTDRS--C	TDYLTWDKNQ	PDHYEG----	(69-102)
RSL	GLRDKKKDFS	WEWTDRS--C	TDYLTWDKNQ	PDHYQN----	(69-102)
ASGPR	GLTDQ--NGP	WKWVDGTDYE	TGFKNWRPGQ	PDDWYGHGLG	(212-249)
HLIT	GLHDPKKNRR	WHWSSGS--L	VSYKSWGIGA	PSSVN-----	(79-111)
SUL	GFNDRTTEGN	FQWTDGS--P	NDETAWVGSN	PDNYGS----	(79-112)
ABL	GLNDIDLEGH	YVWSNGE--A	TDFTYWSSNN	PNNWE-----	(65-97)
MBP	GITDEVTEGQ	FMVVTGG--R	LTYSNNKKDE	PNDHGS----	(158-191)
				*	
TSL	-KEFCVELVS	LTGYHRWNDQ	VCESKNS-FL	CQCKF	(103-135)
LMSL	-KEFCVELVS	LTGYRLWNDQ	VCESKNA-FL	CQCKF	(103-135) 85.2%
RSL	-KEFCVELVS	LTGYRLWNDQ	VCESKDA-FL	CQCKF	(103-135) 82.2%
ASGPR	GGEDCAHFTT	D-G--HWND	VCRRPYR-WV	CET--	(250-278) 41.1%
HLIT	-PGYCVSLTS	STGFKWKVDV	PCEDEKFS-FV	CKEKN	(112-144) 33.1%
SUL	-GEDTQMVM	GAG-LNWIDL	PCSSTRHYLI	CKL--	(113-143) 31.6%
ABL	-NQDCGVVNY	DTVTGQWDD	DCNKNRN-FL	CKMP-	(98-129) 31.0%
MBP	-GEDCVTIVD	N-G--LWNDI	SCQASHT-AV	CEFPA	(192-221) 25.6%
				*	

Figure 4 Alignment of sequence of TSL with those of the known C-type CRDs

The sequences are: LMSL, C-type lectin from *L. muta stenophrys* venom [2]; RSL, rattlesnake lectin from *C. atrox* venom [1]; ASGPR, rat liver asialoglycoprotein receptor-1 [18]; HLIT, human lithostathine [19]; SUL, sea-urchin lectin [20]; ABL, acorn-barnacle lectin [21]; MBP, rat serum mannose-binding protein [22]. Gaps have been inserted to maximize similarity. Stars (*) indicate the residues that are completely conserved between TSL and other CRDs. Dots (·) indicate the residues that are conserved in more than five proteins. Ligands for Ca²⁺ number 1 in the MBP are underlined, and ligands for Ca²⁺ number 2 in the MBP are shown in **bold**. Percentage identity between TSL and each of the other proteins (within the sequence of the CRD) is shown on the right.

unmodified peptides. Nevertheless, the HPLC-ESI-MS and tandem MS results have verified the deduced amino acid sequence.

Sequence similarity between TSL and other C-type CRDs

To further characterize the structure of TSL, we compared its deduced amino acid sequence with those of other known C-type CRDs (Figure 4). There was an obvious high degree of similarity between TSL and rattlesnake lectin and *L. muta stenophrys* lectin (82.2% and 85.2% identity respectively). The above three lectins contain 135 residues per monomer and the positions of the nine cysteine residues are the same. However, TSL possesses an additional cysteine residue at position 2. In addition, TSL showed the following percentages of identity with other heterodimeric snake venom C-type-lectin-like proteins: the α and β subunits of botrocetin [23], 39.4% and 38.2% respectively; the β and α subunits of echicetin [5], 37.8% and 36.3% respectively; the B and A chains of habu IX/X-binding protein [3], 33.6% and 33.3% respectively. From the known patterns of disulphide bridges in three of the C-type-lectin-like proteins from snake venoms (e.g. factor IX/X-binding protein from *Trimeresurus flavoviridis* [24], botrocetin from *Bothrops jararaca* [23], and the homodimeric rattlesnake lectin from *C. atrox* [1]), the location of the disulphide bridges in the TSL protein could be deduced. Four intrachain disulphide bridges linked Cys³ to Cys¹⁴, Cys³¹ to

Cys¹³¹, Cys³⁸ to Cys¹³³, and Cys¹⁰⁶ to Cys¹²³. Two half-cystine residues, Cys² and Cys⁸⁶, could have participated in the formation of the interchain disulphide bridges. Furthermore, the additional cysteine (Cys²) may account for a minor component (homooligomer) observed by SDS/PAGE under reducing conditions.

The amino acid sequence of TSL is also homologous with those of two venom and some other known non-venom CRDs, with eleven amino acid residues (Cys³¹, Glu⁴⁸, Gly⁶⁹, Asp⁷², Trp⁹², Pro⁹⁷, Cys¹⁰⁶, Trp¹¹⁸, Asp¹²⁰, Cys¹²³ and Cys¹³¹) completely conserved, and fourteen residues well conserved. It is also striking that TSL shows 41.1% amino acid sequence identity with rat liver asialoglycoprotein receptor, a well-characterized galactose-binding C-type animal lectin.

In comparison with the rat mannose-binding protein (MBP), four residues (Asp¹⁶¹, Glu¹⁶⁵, Asp¹⁸⁸, and Asp¹⁹⁴), which co-ordinate Ca²⁺ number 1, are not well conserved in TSL [25]. This fact suggests that differences in the position of Ca²⁺ site 1 in TSL, as observed in selectins [26], are likely lead to differences in the details of Ca²⁺-dependent local conformational changes. In addition, we have also noted that the TSL molecule conserves three of the five Ca²⁺-binding residues in the Ca²⁺ number 2 site found in MBP, with the tripeptide sequence Glu¹⁸⁵-Pro¹⁸⁶-Asn¹⁸⁷ in MBP replaced by the corresponding sequence Gln⁹⁶-Pro⁹⁷-Asp⁹⁸ in TSL. This interesting phenomenon may be due to the different sugar-binding specificity in TSL compared to MBP. In fact, experiment has shown that the double substitution (Glu¹⁸⁵ → Gln and Asn¹⁸⁷ → Asp) MBP mutant binds to galactose in preference to mannose by a factor of 3 [27]. Taken together, these data imply that TSL possesses a C-type CRD folding mode and a Ca²⁺-binding site 2 similar to MBP, in spite of its low overall sequence similarity.

Although many of the CRD-containing proteins have been sequenced, most of them are mosaic proteins composed of CRDs and other domains. From the evolutionary point of view, C-type lectins from snake venoms seem to lie between the simple proteins that consist of just a C-type CRD and the complex, multidomain proteins of other vertebrates. Moreover, these C-type lectins exhibit little or no homology in domains other than the CRD [28]. However, TSL, as a galactose-binding protein in group VII of the C-type animal lectins as classified by Drickamer [9], consists simply of one isolated free CRD. Thus, further studies on its binding sites and three-dimensional structure will be helpful to a better understanding of the structure-function relationships of CRDs. We believe that TSL should also be an excellent model system to study the interaction between CRD and carbohydrate ligands.

We thank Dr. Rong Zeng for her excellent technical assistance and advice. We also thank Dr. Hua Pan for fruitful discussions. This work was supported by Grant 39570174, National Natural Science Grand Council, China.

REFERENCES

- Hirabayashi, J., Kusunoki, T. and Kasai, K. (1991) *J. Biol. Chem.* **266**, 2320–2326
- Aragon-Ortiz, F., Mentele, R. and Averswald, E. A. (1996) *Toxicon* **34**, 763–769
- Atoda, H., Hyuga, M. and Morita, T. (1991) *J. Biol. Chem.* **266**, 14903–14911
- Usami, Y., Suzuki, M., Yoshida, E., Sakurai, Y., Hirano, K., Kawasaki, T., Fujimura, Y. and Titani, K. (1996) *Biochem. Biophys. Res. Commun.* **219**, 727–733
- Polgar, J., Magnenat, E. M., Peitsch, M. C., Wells, T. N., Saqi, M. S. and Clemetson, K. J. (1997) *Biochem. J.* **323**, 533–537
- Liang, X.-L. and Wang, K.-Y. (1993) *Acta Biochim. Biophys. Sin.* **25**, 515–521
- Liang, X.-L. and Wang, K.-Y. (1994) *Acta Biochim. Biophys. Sin.* **26**, 137–143
- Kini, R. M. (1996) *Toxicon* **34**, 1287–1294
- Drickamer, K. (1993) *Prog. Nucleic Acid Res. Mol. Biol.* **45**, 207–232
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Garrity, P. A., Wold, B. and Mueller, P. R. (1995) in *PCR 2: a Practical Approach* (McPherson, M. J., Hames, B. D. and Taylor, G. R., eds.), pp. 329–322, Oxford University Press, New York
- Troutt, A. B., McHeyzer-Williams, M. G., Pulendran, B. and Nossal, G. J. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9823–9825
- Higgins D. G., Thompson J. D. and Gibson, T. J. (1996) *Methods Enzymol.* **266**, 383–402
- Matsuzaki, R., Yoshiara, E., Yamada, M., Shima, K., Atoda, H. and Morita, T. (1996) *Biochem. Biophys. Res. Commun.* **220**, 382–387
- Arocas, V., Castro, H. C., Zingali, R. B., Guillin, M. C., Jandrot-Perrus, M., Bon, C. and Wisner, A. (1997) *Eur. J. Biochem.* **248**, 550–557
- Leduc, M. and Bon, C. (1998) *Biochem. J.* **333**, 389–393
- Pappin, D. J. C., Rahman, D., Hansen, H. F., Jeffery, W. and Sutton, C. W. (1995) in *Methods in Protein Structure Analysis* (Atassi, M. Z. and Appella, E., eds.), pp. 161–173, Plenum Press, New York
- Leung, J. O., Holland, E. C. and Drickamer, K. (1985) *J. Biol. Chem.* **260**, 12523–12527
- Watanabe, T., Yonekura, H., Terazono, K., Yamamoto, H. and Okamoto, H. (1990) *J. Biol. Chem.* **265**, 7432–7439
- Giga, Y., Ikai, A. and Takahashi, K. (1987) *J. Biol. Chem.* **262**, 6197–6203
- Muramoto, K. and Kamiya, H. (1986) *Biochim. Biophys. Acta* **874**, 285–295
- Weis, W. I., Crichton, G. V., Murthy, H. M., Hendrickson, W. A. and Drickamer, K. (1991) *J. Biol. Chem.* **266**, 20678–20686
- Usami, Y., Fujimura, Y., Suzuki, M., Ozeki, Y., Nishio, K., Fukui, H. and Titani, K. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 928–932
- Atoda, H. and Morita, T. (1993) *J. Biochem. (Tokyo)* **113**, 159–163
- Weis, W. I., Drickamer, K. and Hendrickson, W. A. (1992) *Nature (London)* **360**, 127–134
- Graves, B. J., Crowther, R. L., Chandran, C., Rumberger, J. M., Li, S., Huang, K. S., Presky, D. H., Familletti, P. C., Wolitzky, B. A. and Burns, D. K. (1994) *Nature (London)* **367**, 532–538
- Drickamer, K. (1992) *Nature (London)* **360**, 183–186
- Vasta, G. R., Ahmed, H., Fink, N. E., Elola, M. T., Marsh, A. G., Snowden, A. and Odom, E. W. (1994) *Ann. N.Y. Acad. Sci.* **712**, 55–73