A unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif

Jennifer J. Smith[1], Justine M. Hill[2], Michelle J. Little[3], Graham M. Nicholson[3], Glenn F. King[1,*], and Paul F. Alewood[1,*]

[1]Institute for Molecular Bioscience and [2]School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD 4072, Australia;

[3]Neurotoxin Research Group, School of Medical and Molecular Biosciences, University of Technology, Sydney, NSW 2007, Australia

***Address for correspondence:** Prof. Glenn F. King or Prof. Paul Alewood, Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia. Email: glenn.king@imb.uq.edu.au and p.alewood@imb.uq.edu.au; Phone: +61 7 3346-2025

ABBREVIATIONS: CPP, cell penetrating peptide; CSα/β, cysteine-stabilised α/β; DDH, disulfide-directed β-hairpin; DQF-COSY, double-quantum filtered correlation spectroscopy; ICK, inhibitor cystine knot; MALDI-ToF, matrix-assisted laser desorption ionization-time of flight; NOESY, nuclear Overhauser effect spectroscopy; rpHPLC, reversed-phase high performance liquid chromatography; TCEP, tris(2-carboxyethyl)phosphine; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

KEYWORDS: scorpion toxin, protein structure, molecular evolution, inhibitor cystine knot, disulfide-directed β-hairpin

Abstract

The three-disulfide inhibitor cystine knot (ICK) motif is a fold common to venom peptides from spiders, scorpions and aquatic cone snails. Over a decade ago it was proposed that the ICK motif is an elaboration of an ancestral two-disulfide fold coined the disulfide-directed β-hairpin (DDH). Here we report the isolation, characterization, and structure of a novel toxin (U_1-LITX-Lw1a) from the venom of the scorpion *Liocheles waigiensis* that is the first example of a native peptide that adopts the DDH fold. U_1-LITX-Lw1a not only represents the discovery of a missing link in venom protein evolution, it is the first member of a fourth structural fold to be adopted by scorpion venom peptides. Additionally, we show that U_1-LITX-Lw1a has potent insecticidal activity across a broad range of insect pest species, thereby providing a novel structural scaffold for bioinsecticide development.

Introduction

Scorpions are one of the most ancient venomous animals, with the oldest fossil scorpion dating to the Silurian Period around 430 million years ago (1). Scorpion venoms are consequently the product of millions of years of evolutionary fine tuning, resulting in the chemically and pharmacologically complex mixture present in extant venoms (2). The predominant components of scorpion venoms are bioactive peptides, many of which have potential application in the pharmaceutical (3) and agrochemical (4) industries.

Over the past 60 years, ~730 scorpion venom peptides have been isolated from 56 species (5). The vast majority of scorpion toxins characterized to date contain a common core topology comprising one or two short α-helices connected to a triple-stranded antiparallel β-sheet stabilized by three or four disulfide bonds (6). This fold, known as the cystine-stabilized α/β (CSα/β) motif, is one of only three structural scaffolds that have been found in disulfide-containing scorpion venom peptides. The topologically unrelated cystine stabilized α-helix-loop-helix (CSα/α) fold consists of two short α-helices connected by a β-turn; only seven scorpion peptides have been discovered that adopt this fold (7). The third fold found in scorpion-venom peptides is the inhibitor cystine knot (ICK) motif (8) that dominates the venom peptidome of spiders (9) and is also present in peptides from evolutionarily unrelated organisms (10).

The CSα/β scorpion neurotoxins evolved by recruitment into the venom and subsequent neofunctionalisation of CSα/β defensins (11). The CSα/β defensins are antimicrobial peptides that are widely distributed in plants, fungi, nematodes, and the arthropod classes Insecta and Arachnida. The conservation of gene structure between scorpion CSα/β defensins, which are

found in the haemolymph, and the CSα/β venom neurotoxins, supports a paralogous relationship between these two classes of peptides (12). In contrast, little is known about the evolutionary origins of the CSα/α and ICK folds. It was previously suggested that the three-disulfide ICK fold is an elaboration of a simpler, ancestral two-disulfide fold coined the disulfide-directed β-hairpin (DDH) (13).

Here we describe the discovery of a scorpion venom peptide that represents the first example of a novel fourth scorpion toxin fold. Moreover, this structure is the first example of a native peptide that adopts the previously hypothetical DDH fold. Thus, the current work provides support for the hypothesis that the DDH fold is the evolutionary precursor of the ICK motif.

Results

Mass profile of the Liocheles waigiensis *venom peptidome*

Research on scorpion venoms has focused primarily on members of the medically important Buthidae family. There have been far fewer studies on non-buthids, and very little research has been performed on Australian scorpions despite their evolutionary significance. We therefore decided to examine the venom of the non-buthid Australian scorpion *L. waigiensis* (Scorpiones: Scorpionoidea: Liochelidae) as this species had not been studied previously. MS analysis of *L. waigiensis* venom revealed 200 distinct masses ranging from 1 kDa to >10 kDa. The majority (88%) of observable masses were smaller than 5 kDa (Fig. 1*B*), indicating that the venom is dominated by small peptides. Mass profiling of the venom using both nano-electrospray and MALDI-ToF instruments proved to be complementary, with only ~30% of the total masses (57 of 192) observed using both methods (Fig. 1*B*).

Isolation and sequencing of a novel peptide

A novel peptide of mass 4171.91 Da was purified from *L. waigiensis* venom using reversed-phase (rp) HPLC (Fig. 1*A*). Since its molecular target remains to be determined, the peptide was named U_1-liotoxin-Lw1a (U_1-LITX-Lw1a) based on the recently introduced nomenclature for venom peptides (14). Reduction of the purified peptide followed by alkylation with iodoacetamide led to a mass increase of 232 Da, indicating the presence of two disulfide bonds. The reduced and alkylated peptide was subjected to N-terminal sequencing which, combined with tandem mass spectrometry (MS/MS) analysis, revealed the primary structure of the 36-residue peptide as DFPLSKEYESCVRPRKCKPPLKCNKAQICVDPNKGW.

U_1-LITX-Lw1a is expressed as a prepropeptide

The DNA sequences of clones from 5' RACE analysis of a venom-gland cDNA library revealed the presence of a unique transcript upstream of the region encoding the mature U_1-LITX-Lw1a toxin. Analysis of the transcript using SignalP 3.0 (15) indicated that it is comprised of a 5' UTR, signal peptide, propeptide, mature toxin encoding region, and a 3' UTR (Fig. S3 and Fig. 5), with a polyadenylation signal (AATAAA) 41 nucleotides downstream of the stop codon. The signal peptide consists of 25 residues, 16 of which are hydrophobic. The propeptide comprises 14 residues and half of these are acidic. While acidic propeptide regions are a common feature of spider-venom toxin transcripts (16), of those scorpion toxin precursors known to have a propeptide region, only a few members of the calcine family have a high number of acidic residues in the propeptide region.

Synthesis of U1-LITX-Lw1a and determination of disulfide bond connectivity

Synthetic U_1-LITX-Lw1a was produced by solid-phase peptide synthesis. Oxidation of the synthetic peptide resulted in one major peak that was shown to co-elute with the native peptide on rpHPLC (Fig. S1). Since the synthetic peptide co-eluted with the native peptide, and there was only a limited amount of native material, the disulfide-bond connectivity was determined by tryptic digest of the synthetic toxin. There are three connectivities possible for a peptide with two disulfide bonds (Fig. S2*A*). However, as there is a tryptic cleavage site between each of the cysteine residues in U_1-LITX-Lw1a, digestion of the fully oxidized peptide with trypsin should give a unique mass fingerprint for each of the three possible disulfide bond connectivities. Indeed, the mass spectra obtained for the oxidized peptide following tryptic digestion (Fig. S2*B*) allowed unambiguous determination that the cysteine residues in U_1-LITX-Lw1a are arranged in a 1–3, 2–4 connectivity (top panel in Fig. S2*A*).

Structure of U1-LITX-Lw1a

The three-dimensional structure of U_1-LITX-Lw1a was determined using homonuclear NMR methods. Statistics highlighting the high precision and stereochemical quality of the ensemble of 20 U_1-LITX-Lw1a structures are shown in Table S1. The highest-ranked member of the ensemble has a MolProbity score of 2.54, placing it in the 45th percentile relative to all other structures ranked by MolProbity.

The structure of U_1-LITX-Lw1a is remarkable for a scorpion venom peptide. It does not contain the CSα/β, CSα/α or ICK motif common to other disulfide-rich scorpion toxins but instead contains a novel two-disulfide scaffold (Fig. 3). The only elements of secondary structure are two short but well-defined two-stranded β sheets. The N-terminal β sheet comprises β strands 1 and 2 (residues 4–5 and 16–17, respectively), while the C-terminal β sheet is composed of β strands 3

and 4 (residues 22–23 and 29–30, respectively) (Fig. 3).

Remarkably, a search for structural homologs of U_1-LITX-Lw1a using DALI (17) yielded a total of 29 unique matches with a statistically significant Z score ≥ 2. Of these structural homologs, 28 contain an ICK motif, including 23 spider toxins and one cone snail toxin. Notably, however, all of these structurally homologous toxins contain at least one additional disulfide bond.

ICK toxins have proliferated in spider venoms to the point where they now dominate most spider-venom peptidomes (18). The marked insensitivity of this structural scaffold to changes in intercystine resides has enabled spiders to develop diverse pharmacologies using the same disulfide framework. Fig. 4 shows an overlay of U_1-LITX-Lw1a on the structure of ICK spider toxins with different pharmacologies, including toxins with high affinity for P2X3 receptors, Na_V channels, K_V channels, and calcium-activated potassium (K_{Ca}) channels. The inherent functional diversity of ICK toxins means that these structural homologs provide few clues as to the likely target of U_1-LITX-Lw1a. In contrast, as discussed below, U_1-LITX-Lw1a might provide an indication of the evolutionary origin of ICK venom peptides.

U_1-LITX-Lw1a is a potent insecticidal toxin

U_1-LITX-Lw1a was lethal to all three insect pest species tested (Fig. 2). In crickets, intermittent twitching of appendages was initially observed and by 2 h post-injection the degree of paralysis ranged from slight at low doses to complete paralysis at high doses. Blowfly larvae also exhibited a dose-dependent contractile paralysis. The LD_{50} ranged from 0.78 nmol/g in mealworms to 5 nmol/g in adult blowflies (Fig. 2).

We attempted to determine the molecular target of the toxin by analyzing its effect on ion

channel currents in cockroach dorsal unpaired median (DUM) neurons. However, no significant effects were observed on the amplitude, kinetics, or the voltage-dependence of activation of voltage-gated sodium (Na_V), potassium (K_V), and calcium (Ca_V) currents when up to 3 μM U_1-LITX-Lw1a was applied to DUM neurons.

Discussion

Evolution of the inhibitor cystine knot

Despite its abundance in invertebrate venoms such as spiders, scorpions, and cone snails, the evolutionary origin of the ICK motif is unclear as there are no invertebrate body proteins known to contain this motif. The ICK motif comprises a "ring" formed by two disulfides and the intervening sections of polypeptide backbone, with a third disulfide piercing the ring to create a pseudo-knot (Fig. 4*F*). The compact hydrophobic core of the ICK motif consists largely of the two central disulfide bridges that emanate from the two β-strands that characterize the ICK fold (19). In contrast, the N-terminal disulfide bridge contributes very little to the hydrophobic core and it has been demonstrated that the tertiary structure and thermal stability of the ICK-containing trypsin inhibitor EETI II is largely unperturbed by removal of the N-terminal disulfide bridge (20, 21). Thus, we previously proposed that the ICK fold is a minor elaboration of a simpler ancestral fold that we referred to as the disulfide-directed β-hairpin (DDH) (13).

The DDH fold is shown schematically in Fig. 4*F* and its amino acid consensus sequence was determined to be $CX_{5–19}CX_2$[G or P]$X_2CX_{6–19}C$, where X is any amino acid (13). The DDH fold differs from the ICK fold in that there are only two mandatory disulfide bridges that form the bulk of the hydrophobic core, so that loop 1 is no longer necessarily bounded by an N-terminal

Cys residue, and loop 3 is generally five residues in length with a central Gly or Pro to ensure a tight turn prior to the first β-strand. The residue following the Gly/Pro in loop 3 is generally a hydrophobe as this residue, along with the two buried disulfide bridges, constitute the mini-hydrophobic core of this domain. Based on their studies on EETI II, Heitz and coworkers (20) also proposed the existence of an elementary motif which they named the cystine-stabilized β-sheet (CSB). The CSB motif is in fact a more stringently defined DDH, containing smaller first and third loops and a triple-stranded β-sheet. Because the DDH motif is less rigidly defined, it should be considered the ancestral fold as it includes many proteins that are excluded from the CSB fold.

It is clear from Fig. 4 that the two disulfides in U_1-LITX-Lw1a overlay well with the two central disulfides in the ICK motif; in contrast, the "outer" solvent-exposed disulfide that closes the ring of the ICK motif is missing in U_1-LITX-Lw1a. Thus, the structure of U_1-LITX-Lw1a corresponds exactly to the previously hypothesized DDH fold. Although the Liochelidae family to which *L. waigiensis* belongs is not believed to be an evolutionarily ancient scorpion (22), U_1-LITX-Lw1a adopts the heretofore-missing native ancestral DDH fold that may be the evolutionary precursor to the ICK motif. Thus, U_1-LITX-Lw1a lends supports to the hypothesis that the ICK toxins, at least in arachnids, may have originated from a DDH-encoding gene that was recruited into the venom and subsequently elaborated to encode an additional disulfide bond.

The phylogeny of scorpions is highly controversial, with recent estimates of the number of extant families ranging from 13 to 18 (23-25). Nevertheless, it is clear that Liochelidae is not a basal group (22). Thus, it is possible that U_1-LITX-Lw1a is not the precursor of an ICK peptide but

rather derived from it. However, this would leave unsolved the question of how the ICK fold was recruited into the venom peptidome. In contrast, as pointed out previously (13), the DDH fold is present in numerous eukaryotic body proteins and hence it seems likely that the highly stable ICK fold was derived by simple modification of an ancestrally recruited DDH motif.

A novel bioinsecticide scaffold

The primary structure of U_1-LITX-Lw1a shows homology to three scorpion peptides: LaIT1 (91% identity) from *Liocheles australasiae* (family Liochelidae) (26), OcyC10 (68% identity) from *Opisthacanthus cayaporum* (Liochelidae) (27), and SmpIT2 (80% identity) from *Scorpio maurus palmatus* (Scorpionidae) (28) (Fig. 5). LaIT1 and SmpIT2, like U_1-LITX-Lw1a, are insecticidal, which is not surprising given that insects are the primary prey of most scorpions and hence this is the function of most venom toxins. LaIT1 induced limb spasms when injected into crickets at a dose of ~4.8 nmol/g (i.e., ~4-fold higher than the LD_{50} value for U_1-LITX-Lw1a) (26), whereas injection of a venom fraction containing SmpIT2 into *Sarcophaga falculata* blowfly larvae caused flaccid paralysis (28). This contrasts with the contractile paralysis caused by U_1-LITX-Lw1a in *Lucilia cuprina* blowfly larvae; thus, SmpIT2 and U_1-LITX-Lw1a possibly have different effects on the same molecular target or act on different receptors altogether.

The similar LD_{50} values obtained when U_1-LITX-Lw1a was tested across a range of invertebrate orders implies that the molecular target of the toxin is evolutionarily conserved. Moreover, this molecular target may be novel for arachnid toxins since high doses of U_1-LITX-Lw1a (3 μM) did not affect Na_V, K_V, or Ca_V currents in cockroach DUM neurons, ruling out most of the common targets of scorpion and spider toxins. U_1-LITX-Lw1a did not alter Na_V or Ca_V currents in rat

dorsal root ganglion neurons at a concentration of 1 μM, and no toxic effects were reported when ~240 pmol LaIT1 (which shares 91% sequence identity with U_1-LITX-Lw1a) was applied intracerebroventricularly to mice (26). Thus, the combined data suggests that U_1-LITX-Lw1a is an insect-selective toxin that represents a new molecular scaffold for bioinsecticide development.

Structure of the U_1-LITX-Lw1a transcript

It is becoming evident that venom from non-buthid species possess fewer high molecular weight long-chain Na_V channel toxins than found in Buthidae scorpion venoms (2). The venom profile of *L. waigiensis* continues this trend as most of the observed peptide masses were less than 5 kDa. However, while the venom peptidome of *L. waigiensis* is consistent with that of other non-buthid venoms, the transcript encoding U_1-LITX-Lw1a is atypical for scorpion-venom peptides.

The mRNA transcript encoding U_1-LITX-Lw1a comprises a 25-residue signal sequence and a 14-residue propeptide preceding the 36-residue mature toxin sequence. Propeptide sequences are ubiquitous in transcripts encoding peptide toxins from cone snails and spiders, but they are uncommon in scorpion toxin transcripts (18). Of the 250 scorpion toxin transcripts that have been sequenced, only ~20 have a pro-region following the signal peptide (5); these correspond to members of the long-chain β and short-chain α potassium channel toxins, non-disulfide-bonded antimicrobial peptides, and the calcine family (29-31).

Ten of the pro-region containing peptides belong to the calcine family. The calcines are cell-penetrating peptides (CPPs) that target ryanodine receptors (32-34). They contain 33–35 residues, adopt the ICK fold (35) and, like most CPPs, have a high net positive charge (+7) (36, 37). An

alanine scan of maurocalcine and imperatoxin-A revealed a critical basic region containing an essential Arg residue at position 24 that is essential for activity at the ryanodine receptor (37, 38). Although U_1-LITX-Lw1a and the calcines both have numerous basic residues and are similar in size, U_1-LITX-Lw1a has a net charge of only +4 and it does not appear to have the attributes necessary for ryanodine receptor activation (Fig. S4). Nevertheless, the conservation of transcript architecture and the similarity in size and structure of the mature toxins suggests that the calcines and U_1-LITX-Lw1a might be evolutionarily related.

In conclusion, the solution structure of U_1-LITX-Lw1a represents a new scorpion toxin fold. This new fold corresponds to the previously proposed but heretofore hypothetical DDH motif, and it provides support for the hypothesis that the DDH motif is the ancestral precursor of scorpion-venom ICK toxins such as the calcines. Further research into the venoms of scorpions and other invertebrates might reveal the presence of additional peptides with the DDH fold, and analysis of their structure and function should provide a greater understanding of the evolutionary relationship between DDH and ICK toxins.

Materials and Methods

Venom extraction

Liocheles waigiensis were purchased from The Green Scorpion (Port Macquarie, NSW, Australia). Venom was extracted manually by inducing the scorpion to sting a parafilm-covered Eppendorf tube. Venom was lyophilized before storage at –20°C. Each milking occurred at least two weeks after the previous milking.

Venom fractionation

Lyophilized venom was reconstituted in 0.05% trifluoroacetic acid (TFA) in water (solvent A) before HPLC fractionation using a narrow bore reversed-phase C_8 column (Zorbax C8, 2.1 × 150 mm, 300 Å) on a Shimadzu VP system. Venom components were eluted at a flow rate of 0.25 mL/min using a linear gradient of 5–60% solvent B (90% acetonitrile, 0.043% TFA) over 115 min preceded by 5% solvent B for 5 min. Absorbance was monitored at 214 nm and 280 nm using a SPD-10A_{VP} UV detector. Fractions were manually collected and lyophilized.

Mass spectrometric analysis

Lyophilized HPLC fractions were dissolved in solvent A and 0.5 μL was spotted onto a MALDI-ToF plate with 0.5 μL α-cyano-4-hydroxycinnamic acid matrix (10 mg/mL in 60% acetonitrile). Spots were analyzed using an ABI 4700 Voyager mass spectrometer (MS) in positive reflector mode. For nanospray MS analysis, crude venom was dissolved in a minimal amount of a 50:50 mixture of solvent A and B and then diluted to 0.5 mg/mL in solvent A. 0.75 μL of dissolved venom was fractionated by rpHPLC using a Vydac C_{18} column (300 Å, 5 μm, 150 × 0.3 mm). Venom components were eluted at a flow rate of 4 μL/min using a gradient of 2% buffer B (90% acetonitrile, 0.1% formic acid) for 10 min, followed by 10–40% buffer B over 110 min, then 40–80% buffer B over 10 min. Online nano-electrospray analysis was performed using a QSTAR Elite mass spectrometer.

Peptide sequencing

U_1-LITX-Lw1a was isolated from crude *L. waigiensis* venom. Disulfides were reduced with 100 mM tris(2-carboxyethyl)phosphine (TCEP) in 100 mM ammonium bicarbonate (pH 6.5) at 50°C for 1 h. Cysteines were alkylated by incubation with 200 mM iodoacetamide in 100 mM

ammonium bicarbonate (pH 9.0) for a further hour at 50°C and the peptide desalted using rpHPLC. Reduced and alkylated U_1-LITX-Lw1a was sequenced using an Applied Biosystems Procise HT Protein Sequencer at the Biomolecular Research Facility (University of Newcastle, Australia). For *de novo* sequencing using MS/MS, reduced and alkylated U_1-LITX-Lw1a was digested by incubation with a ~1:50 molar ratio of trypsin (Proteomics Grade, Sigma) at pH 9 for 14 h at 37°C, and the resultant cleavage products were separated using rpPHLC. A 4700 Voyager mass spectrometer was used for analysis of tryptic fragments; MS ions were selected for MS/MS followed by manual analysis of spectra using GPS DeNovo Explorer (version 3.6) and MS-Product (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct).

Chemical synthesis of U_1-LITX-Lw1a

U_1-LITX-Lw1a was chemically synthesized using the stepwise *in situ* neutralization protocol for Boc chemistry followed by HF cleavage (39). Boc-protected amino acids were purchased from Novabiochem and the following side-chain protected amino acids were used: C(MeBzl), D(Chxl), E(Chxl), K(ClZ), N(Xan), Q(Xan), R(Tos), S(Bzl), Y(BrZ). All other amino acids were unprotected. The cleaved peptide was purified using rpHPLC, then oxidized/folded in 30% isopropanol/0.1 M ammonium bicarbonate, pH 8.0 for 3 days. The folded peptide was purified to homogeneity using rpHPLC.

Sequencing of U_1-LITX-Lw1a transcript

A QuickPrep Micro mRNA Purification Kit (Illustra, GE Healthcare) was used to extract mRNA from the telson of one *L. waigiensis* specimen three days after milking. The FirstChoice RLM-RACE Kit (Ambion) was used to ligate a 5' adapter sequence to the mRNA. Single stranded cDNA was synthesized from the ligated mRNA using Superscript III reverse transcriptase

(Invitrogen) and a poly dT primer. The 5' region of the U_1-LITX-Lw1a transcript was established by amplifying the region of cDNA between the 5' adaptor and the start of the mature toxin. The 5' adaptor primer used was as per the Ambion kit and a degenerate primer was designed based on the N-terminal amino acid sequence of U_1-LITX-Lw1a determined from Edman sequencing. The complete U_1-LITX-Lw1a prepropeptide transcript was found by PCR using the 5' adaptor primer and a poly dT primer. Amplified fragments were subsequently cloned and sequenced.

Insect bioassays

U_1-LITX-Lw1a solutions in insect saline (40) were injected into *Acheta domestica* (house crickets, mass 60–100 mg), *Tenebrio molitor* larvae (mass 190–210 mg), and *Lucilia cuprina* larvae and adults (mass 38–42 mg and 18–22 mg, respectively) (41). The median lethal dose (LD_{50}) was determined as described previously (42). Whole-cell recordings of ionic currents in DUM neurons isolated from the American cockroach *Periplaneta americana* were made as described previously (43), with minor modifications as described in SI text.

NMR structure determination

Lyophilized U_1-LITX-Lw1a was resuspended at a final concentration of 300 μM in 90% H_2O/10% D_2O, pH 3.1. The sample was filtered using a low-protein-binding Ultrafree-MC centrifugal filter (0.22 μm; Millipore, MA, USA), then 550 μL was added to a 5 mm outer-diameter susceptibility-matched microtube (Shigemi, Japan). NMR spectra were acquired at 283 K using a cryoprobe-equipped 900 MHz Avance spectrometer (Bruker BioSpin, Germany). Two-dimensional TOCSY (spin-lock time of 80 ms), NOESY (mixing time of 250 ms), and DQF-COSY spectra were acquired.

NMR spectra were analyzed using the program XEASY (44). A complete set of sequence-specific resonance assignments were obtained and chemical shifts were deposited in BioMagResBank (accession number 16963). The NOESY spectrum was manually peak picked and integrated, then the peaklists were assigned and an ensemble of structures calculated automatically using the CANDID module of the torsion angle dynamics package CYANA (45, 46). The tolerances used for CANDID were 0.025 ppm in both ^{1}H dimensions. The Phe2–Pro3 and Lys18–Pro19 peptide bonds were determined to be in the *trans* conformation on the basis of characteristic H_{α}–H_{δ} NOEs, whereas the Arg13–Pro14 and Pro19–Pro20 peptide bonds were both clearly identified as *cis* based on characteristic H_{α}–H_{α} NOEs (47). Initial structure calculations in combination with analysis of NOE networks allowed assignment of χ_1 restraints for all four Cys residues.

The disulfide bond configuration of U_1-LITX-Lw1a was unambiguously determined via chemical methods (see Results), thus allowing disulfide-bond restraints of $2.0 \leq d \leq 2.1$ Å for S_{γ}(i)–S_{γ}(j), and $3.0 \leq d \leq 3.1$ Å for both C_{β}(i)–S_{γ}(j) and S_{γ}(i)–C_{β}(j) to be used in the structure calculations. Six hydrogen bonds were clearly identified in preliminary rounds of structure calculation. Hydrogen-bond restraints of 1.7–2.2 Å and 2.7–3.2 Å were employed for the H_N–O and N–O distances, respectively, in subsequent structure calculations (48). In the final round of structure calculations, CYANA was used to calculate 100 structures from random starting conformations, then the 20 conformers with highest stereochemical quality as judged by MolProbity (49) were selected to represent the solution structure of U_1-LITX-Lw1a. During the automated NOESY assignment/structure calculation process the CANDID module of CYANA assigned 93.6% of all NOESY crosspeaks (1699 out of 1815), a significantly higher assignment rate than in most

previous reports (46, 50). Coordinates for the final ensemble of structures are available from the Protein Data Bank (accession number 2KYJ).

Acknowledgements

We acknowledge financial support from the Australian Research Council (Discovery Grants DP0774245 and DP1095728 to GFK), the National Health and Medical Research Council (NHMRC) of Australia (Program Grant 351446), and The University of Queensland (UQ). JMH is an NHMRC RD Wright Fellow (401748). We thank Dr Geoffrey W. Brown (Agriscience Queensland) for supply of sheep blowfly larvae, and Dr Brid Callaghan (Queensland Brain Institute, UQ) for DRG assays.

References

1. Dunlop, J. A. & Selden, P. A. (2009) *Exp. Appl. Acarol.* **48,** 183–197.
2. Rodriguez de la Vega, R. C., Schwartz, E. F. & Possani, L. D. (2010) *Toxicon* **56,** 1155–1161.
3. Mamelak, A. N. & Jacoby, D. B. (2007) *Expert Opin. Drug Deliv.* **4,** 175–186.
4. Zlotkin, E., Fishman, Y. & Elazar, M. (2000) *Biochimie* **82,** 869–881.
5. He, Q. Y., He, Q. Z., Deng, X. C., Yao, L., Meng, E., Liu, Z. H. & Liang, S. P. (2008) *Nucleic Acids Res.* **36,** D293–D297.
6. Rodriguez de la Vega, R. C. & Possani, L. D. (2004) *Toxicon* **43,** 865–875.
7. Chagot, B., Pimentel, C., Dai, L., Pil, J., Tytgat, J., Nakajima, T., Corzo, G., Darbon, H. & Ferrat, G. (2005) *Biochem. J.* **388,** 263–271.
8. Pallaghy, P. K., Nielsen, K. J., Craik, D. J. & Norton, R. S. (1994) *Protein Sci.* **3,** 1833–1839.
9. King, G. F., Tedford, H. W. & Maggio, F. (2002) *J. Toxicol.-Toxin Rev.* **21,** 359–389.
10. Craik, D. J., Daly, N. L. & Waine, C. (2001) *Toxicon* **39,** 43–60.
11. Fry, B. G., Roelants, K., Champagne, D. E., Scheib, H., Tyndall, J. D., King, G. F., Nevalainen, T. J., Norman, J. A., Lewis, R. J., Norton, R. S., Renjifo, C. & de la Vega, R. C. (2009) *Annu. Rev. Genomics Hum. Genet.* **10,** 483–511.
12. Zhu, S., Li, W., Jiang, D. & Zeng, X. (2000) *IUBMB Life* **50,** 57–61.
13. Wang, X., Connor, M., Smith, R., Maciejewski, M. W., Howden, M. E., Nicholson, G. M., Christie, M. J. & King, G. F. (2000) *Nat. Struct. Biol.* **7,** 505–513.
14. King, G. F., Gentz, M. C., Escoubas, P. & Nicholson, G. M. (2008) *Toxicon* **52,** 264–276.
15. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004) *J. Mol. Biol.* **340,** 783–795.
16. Tedford, H. W., Sollod, B. L., Maggio, F. & King, G. F. (2004) *Toxicon* **43,** 601–618.
17. Holm, L. & Rosenstrom, P. (2010) *Nucl. Acids Res.* **38,** W545–W549.
18. Sollod, B. L., Wilson, D., Zhaxybayeva, O., Gogarten, J. P., Drinkwater, R. & King, G. F. (2005) *Peptides* **26,** 131–139.

19. Fletcher, J. I., Smith, R., O'Donoghue, S. I., Nilges, M., Connor, M., Howden, M. E. H., Christie, M. J. & King, G. F. (1997) *Nat. Struct. Biol.* **4,** 559–566.

20. Heitz, A., Le-Nguyen, D. & Chiche, L. (1999) *Biochemistry* **38,** 10615–10625.

21. Le-Nguyen, D., Heitz, A., Chiche, L., El Hajji, M. & Castro, B. (1993) *Protein Sci.* **2,** 165–174.

22. Prendini, L. (2000) *Cladistics* **16,** 1–78.

23. Prendini, L. & Wheeler, W. C. (2005) *Cladistics* **21,** 446–494.

24. Fet, V. & Soleglad, M. E. (2005) *Euscorpius* **31,** 1–13.

25. Volschenk, E. S., Mattoni, C. I. & Prendini, L. (2008) *Zool. J. Linn. Soc.* **154**.

26. Matsushita, N., Miyashita, M., Sakai, A., Nakagawa, Y. & Miyagawa, H. (2007) *Toxicon* **50,** 861–867.

27. Silva, E. C., Camargos, T. S., Maranhao, A. Q., Silva-Pereira, I., Silva, L. P., Possani, L. D. & Schwartz, E. F. (2009) *Toxicon* **54,** 252–261.

28. Lazarovici, P., Yanai, P., Pelhate, M. & Zlotkin, E. (1982) *J. Biol. Chem.* **257,** 8397–8404.

29. Yuan, W., Cao, L., Ma, Y., Mao, P., Wang, W., Zhao, R., Wu, Y., Cao, Z. & Li, W. (2010) *Peptides* **31,** 22–26.

30. Zeng, X. C., Li, W. X., Zhu, S. Y., Peng, F., Zhu, Z. H., Wu, K. L. & Yiang, F. H. (2000) *Toxicon* **38,** 1009–1014.

31. Zhu, S., Li, W., Zeng, X., Jiang, D., Mao, X. & Liu, H. (1999) *FEBS Lett.* **457,** 509–514.

32. Nabhani, T., Zhu, X., Simeoni, I., Sorrentino, V., Valdivia, H. H. & Garcia, J. (2002) *Biophys. J.* **82,** 1319–1328.

33. Fajloun, Z., Kharrat, R., Chen, L., Lecomte, C., Di Luccio, E., Bichet, D., El Ayeb, M., Rochat, H., Allen, P. D., Pessah, I. N., De Waard, M. & Sabatier, J. M. (2000) *FEBS Lett.* **469,** 179–185.

34. Shahbazzadeh, D., Srairi-Abid, N., Feng, W., Ram, N., Borchani, L., Ronjat, M., Akbari, A., Pessah, I. N., De Waard, M. & El Ayeb, M. (2007) *Biochem. J.* **404,** 89–96.

35. Mosbah, A., Kharrat, R., Fajloun, Z., Renisio, J. G., Blanc, E., Sabatier, J. M., El Ayeb, M. & Darbon, H. (2000) *Proteins* **40,** 436–442.

36. Boisseau, S., Mabrouk, K., Ram, N., Garmy, N., Collin, V., Tadmouri, A., Mikati, M., Sabatier, J. M., Ronjat, M., Fantini, J. & De Waard, M. (2006) *Biochim. Biophys. Acta* **1758,** 308–319.

37. Mabrouk, K., Ram, N., Boisseau, S., Strappazzon, F., Rehaim, A., Sadoul, R., Darbon, H., Ronjat, M. & De Waard, M. (2007) *Biochim. Biophys. Acta* **1768,** 2528–2540.

38. Lee, C. W., Lee, E. H., Takeuchi, K., Takahashi, H., Shimada, I., Sato, K., Shin, S. Y., Kim, D. H. & Kim, J. I. (2004) *Biochem. J.* **377,** 385–394.

39. Schnolzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. (1992) *Int. J. Pept. Protein Res.* **40,** 180–193.

40. Eitan, M., Fowler, E., Herrmann, R., Duval, A., Pelhate, M. & Zlotkin, E. (1990) *Biochemistry* **29,** 5941–5947.

41. Wang, X.-H., Smith, R., Fletcher, J. I., Wilson, H., Wood, C. J., Howden, M. E. H. & King, G. F. (1999) *Eur. J. Biochem.* **264,** 488–494.

42. Tedford, H. W., Fletcher, J. I. & King, G. F. (2001) *J. Biol. Chem.* **276,** 26568–26576.

43. Gunning, S. J., Maggio, F., Windley, M. J., Valenzuela, S. M., King, G. F. & Nicholson, G. M. (2008) *FEBS J.* **275,** 4045–4059.

44. Bartels, C., Xia, T. H., Billeter, M., Güntert, P. & Wüthrich, K. (1995) *J. Biomol. NMR* **6,** 1–10.

45. Güntert, P. (2004) *Methods Mol. Biol.* **278,** 353–378.

46. Herrmann, T., Güntert, P. & Wüthrich, K. (2002) *J. Mol. Biol.* **319,** 209–227.

47. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (John Wiley & Sons, Inc.

48. Fletcher, J. I., Chapman, B. E., Mackay, J. P., Howden, M. E. H. & King, G. F. (1997) *Structure* **5,** 1525–1535.

49. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007) *Nucl. Acids Res.* **35,** W375–W383.

50. Robson, S. A. & King, G. F. (2006) *Proc. Natl. Acad. Sci. USA* **103,** 6700–6705.

Figure Legends

Figure 1. Analysis of the venom peptidome of *L. waigiensis*. **(A)** rpHPLC chromatogram of crude venom from the scorpion *L. waigiensis*. The acetonitrile gradient is shown in magenta and the peak marked with an asterisk corresponds to the U_1-LITX-Lw1a peptide. (**B)** Left: Histogram showing the abundance of peptide toxins in the venom of *L. waigiensis*, sorted into 1 kDa molecular mass bins. The overlaid curve shows the cumulative total number of peptide masses identified from MS analyses. Right: Venn diagram showing overlap of peptide masses determined using electrospray and MALDI-ToF mass spectrometry.

Figure 2. Dose-response curves resulting from injection of U_1-LITX-Lw1a into various insects. The LD_{50} values are indicated. Data points are the mean ± SE of three experiments.

Figure 3. Stereo view of the structure of U_1-LITX-Lw1a. The 20 members of the structural ensemble are overlaid for best fit over the backbone atoms of residues 2–36. β-strands 1-4 are shown in cyan, and the sidechains of the two disulfide bonds are highlighted in red. The N- and C-termini are labeled. The hydrophobic core of the protein is composed of the two disulfide bonds and the sidechain of Leu21, which is shown in magenta.

Figure 4. Overlay of U_1-LITX-Lw1a and structurally homologous ICK toxins. The NMR structure of U_1-LITX-Lw1a (red) is overlaid on the structures of different structurally homologous spider ICK toxins (grey), including **(A)** guangxitoxin (PDB code 2WH9), **(B)** GsMTX-4 (PDB code 1TYK), **(C)** hainantoxin-I (PDB code 1NIX), **(D)** purotoxin (PDB code 2KGU), and **(E)** κ-hexatoxin-Hv1c (PDB code 1DL0). The two central disulfides of the ICK

toxins, shown as grey tubes, overlap with the two disulfides of U_1-LITX-Lw1a (shown as gold tubes). Additional disulfide bonds in the ICK toxins are shown as blue tubes. The molecular target of each spider toxin is indicated. (**F)** Graphical representation of the DDH and ICK motifs, with the disulfides and β sheet of the DDH motif shown in orange and green, respectively. The third disulfide necessary for the formation of the ICK motif is shown in blue.

Figure 5. Alignment of the amino acid sequences of U_1-LITX-Lw1a, OcyC10, LaIT1 and SmpIT2. The amino acid translation of the predicted signal and predicted propeptide regions of the U_1-LITX-Lw1a and OcyC10 transcripts are shown. Only the mature toxin regions of LaIT1 and SmpIT2 are available. Signal, propeptide and mature toxin regions are indicated with arrows.