MONASH University

Peptide toxins: structure, stability and engineering

Stephen Bernard Drane
BSc(Hons)

A thesis submitted for the degree of Master of Pharmaceutical Science at
Monash University in 2017
Monash Institute of Pharmaceutical Sciences

Table of contents

Abstract

Peptide toxins from animal venom are a rich source of novel biochemistry, with many toxins having potential utility as therapeutic leads and biochemical tool compounds. There is a current push to investigate newly described toxins, with only a very small proportion of the estimated toxin diversity having been characterised at the present time. This thesis documents an investigation of the scorpion toxin HsTX1, as well as the conotoxins H_Vc7.1, H_Vc7.2 and contryphan-Vc2 from the marine cone snail *Conus victoriae*.

The stability and efficient folding of the scorpion toxin HsTX1 suggested possible utility as a scaffold for molecular grafting applications. To test this possibility, the integrin receptor binding motif RGD and the protein-protein interaction recognition motif DINNN were chosen as grafts to be displayed on this scaffold. Molecular dynamics computer simulations were used to test the stability of proposed construct designs, and the best performers were produced by solid-phase peptide synthesis. Despite the high folding propensity of the native HsTX1 structure, the selected constructs did not fold to a stable three-dimensional conformation. This suggests that modifying the sequence at the chosen insertion point disrupts the formation of the disulfide bond network. This was not an outcome considered in the computer simulations, which did not report on the folding pathway. Given this failure, HsTX1 did not seem to be suited for use as a scaffold.

The H-superfamily conotoxins H_Vc7.1 and H_Vc7.2 had previously been identified in the venom duct transcriptome of *C. victoriae* and were produced by recombinant expression for testing. Unfortunately the complexity of folding these multiply-disulfide-bonded peptides proved intractable, and it was not possible to produce a sample of the required purity for structural and functional analyses.

Contryphan-Vc2 was synthesised in sufficient quantity and purity to determine its three-dimensional structure by nuclear magnetic resonance spectroscopy. This structure was found to be similar to other contryphans, but displayed unexpected dynamic behaviour in computer simulations. Activity assays in mice revealed a novel depressive behavioural phenotype, in contrast to the hyperactive phenotype previously reported for peptides in this class. Further experiments and simulations suggested an interaction with the lipid membrane, which may also be a feature for other members of the contryphan family.

While some aspects of the work did not yield positive results, the initial characterisation of contryphan-Vc2 lays the basis for future work on this peptide and the contryphan family more broadly, and may aid in elucidating the biochemical targets of this class of molecules.

Publications during enrolment

Drane, S. B., Robinson, S. D., MacRaild, C. A., Chhabra, S., Chittoor, B., Morales, R. A. V., Leung, E. W. W., Belgi, A., Espino, S. S., Olivera, B. M., Robinson, A. J., Chalmers, D. K., and Norton, R. S. (2017) Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue, *Toxicon 129,* 113-122.

Declaration (Thesis including published works)

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one (1) original paper published in a peer reviewed journal and no unpublished publications. The core theme of the thesis is investigating the potential utility of peptide toxins. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Medicinal Chemistry in the Faculty of Pharmacy and Pharmaceutical Sciences under the supervision of Professor Raymond Norton.

(The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.)

In the case of Chapter 3 my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student Y/N
3	*Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue*	*Published*	*57%. Concept, sample preparation, data collection, data analysis and writing manuscript.*	*1) Samuel Robinson: concept, mouse assay, input into manuscript 9%* *2) Christopher MacRaild: NMR data analysis, SPR assays, 4%* *3) Sandeep Chhabra: NMR data analysis 4%* *4) Balasubramanyam Chittoor: proteolysis assay 2%* *5) Rodrigo Morales: sample preparation assistance 2%*	*Yes* *No* *No* *Yes* *No*

				6) Eleanor Leung: SPR assays 2% *7) Alessia Belgi: sample preparation 2%* *8) Samuel Espino: mouse assay 2%* *9) Baldomero Olivera: mouse assay, input in manuscript 2%* *10) Andrea Robinson: sample preparation 2%* *11) David Chalmers: MD simulation assistance 2%* *12) Raymond Norton: concept, input in manuscript 10%*	*No* *No* *No* *No* *No* *No* *No*

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: [redacted] **Date:** 15 December, 2017

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: [redacted] **Date:** 7 December, 2017

Acknowledgements

This document is the product of a long journey; a journey in which many people have assisted and aided, both in and out of the laboratory.

Gratitude is due first and foremost to my supervisor Prof Ray Norton, for welcoming me into his laboratory group and letting me loose on an intensely interesting project. His suggestions and guidance have always had the effect of making the project better, and I am thankful for all the time and attention he has spent on it and me.

Thanks must also be given to Dr Rodrigo Morales, for taking me under his wing when I was an uncertain student in the early days of the project, and again for later taking on supervisory duties. The lessons he taught in how to approach a research question and plan out your experimental attack laid a solid foundation that stood me in good stead as I came to take more complete control and ownership of the project.

A great deal is also owed to the other postdocs in the Norton laboratory, who have assisted in many and varied ways. Whenever I had a question or a difficulty, they were unfailingly helpful and generous with their time, so many thanks to Indu Chandrashekaran, Sandeep Chhabra, Eleanor Leung, Christopher MacRaild, Dorothy Wai and Hiromasa Yagi – all have left their mark on me, and I am a better scientist for their input.

The senior students in the lab, although now moved on to bigger and better things, were also instrumental in setting my foundation. Particular thanks to Samuel Robinson, whose work formed the basis of a large part of my own, and who assisted greatly in my first year.

My fellow students at the Monash Institute of Pharmaceutical Sciences have also been of great help, both academically and as companions. Whether troubleshooting code or shooting the breeze, having a sympathetic ear to complain to or an interested party to which I could brag about my latest achievement has been wonderful. I can't name you all, but if we've played cards together or caught up out-of-hours to watch a movie or just relax, I'm talking about you!

I have been financially supported through this journey by the Faculty of Pharmacy and Pharmaceutical Sciences of Monash University, which has allowed me to focus on the work without worrying where my next meal was coming from. Many members of the faculty have also enabled my activities, such as Jason Dang, Kade Roberts and Stephen Headey (and many more). Thanks also to Karen McConalogue for helping me navigate the processes and procedures of the bureaucracy associated with the university.

Moving beyond the Institute, I am also thankful for the members of the Society for Creative Anachronism, the RMIT Occasional Choral Society and my wider friendship group for providing much-needed social outlets and a balancing force on the work-life scales. Having learnt from my experiences writing my Honours thesis, I appreciate how important it is to keep up with recreational activities to avoid falling in a non-productive heap. Specific thanks to those who read and commented on my work, such as Yael Bornstein and Justina Ashman.

Finally, thanks to my family for always standing behind me and providing a safety net 'just in case'. I may not always need you, but it is very helpful to know that you're there anyway, ready to swoop in if I do.

Abbreviations

6xHis	polyhistidine
A_{214}	absorbance at 214 nm
ACE	angiotensin converting enzyme
Acm	acetamidomethyl
AgTx	agatoxin
AU	absorbance units
BmBKTx1	*Buthus martensi* high-conductance potassium channel toxin 1
Boc	*tert*-butyloxycarbonyl
Ca_V	voltage-gated calcium channel
CCK	cyclic cystine knot
CCL	cyclic cysteine ladder
CDR2	complementarity determining region 2
CD4	cluster of differentiation 4
cDNA	complementary deoxyribonucleic acid
CPU	central processing unit
CSαα	cysteine-stabilised α-helix-loop-helix
CSαβ	cysteine-stabilised α-helical and β-sheet
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DNA	deoxyribonucleic acid
DQF-COSY	double-quantum-filtered correlation spectroscopy
DTT	dithiothreitol
EETI-II	*Ecballium elaterium* trypsin inhibitor II
ERG	ether-a-go-go-related gene
FDA	Food and Drug Administration
Fmoc	9-(fluorenyl)methyloxycarbonyl
FPLC	fast protein liquid chromatography
GPCR	G protein-coupled receptor
GSH	reduced glutathione
GSSG	oxidised glutathione
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	hydrofluoric acid
HsTX1	*Heterometrus spinifer* toxin 1

ICK	inhibitory cystine knot
Ig	immunoglobulin
IPTG	isopropyl-β-D-1-thiogalactopyranoside
iNOS	inducible nitric oxide synthase
KTx	potassium toxin
K_V	voltage-gated potassium channel
LB	Luria-Bertani
LC	liquid chromatography
MD	molecular dynamics
MeBzl	4-methyl-benzyl
MS	mass spectroscopy
MS/MS	tandem mass spectroscopy
nAChR	nicotinic acetylcholine receptor
NaScTx	sodium channel scorpion toxin
Na_V	voltage-gated sodium channel
NDBP	non-disulfide-bridged peptides
NET	noradrenaline (norepinephrine) transporter
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NTA	nitrilotriacetic acid
OD_{600}	optical density at 600 nm
ON	overnight
PCR	polymerase chain reaction
PDB	protein data bank
ppm	parts per million
R_g	radius of gyration
RMSD	root-mean-square deviation
RNA	ribonucleic acid
ROESY	rotating-frame nuclear Overhauser effect spectroscopy
RP-HPLC	reverse-phase high performance liquid chromatography
rpm	revolutions per minute
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
ShK	*Stichodactyla helianthus* potassium channel toxin

SPC	single point charge
SPPS	solid-phase peptide synthesis
SPSB2	SPRY-containing SOCS box protein 2
tBu	*tert*-butyl
TCEP	tris(2-carboxyethyl)phosphine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TIP3P	transferable intermolecular potential with 3 points
TMS	tetramethylsilane
TOCSY	total correlation spectroscopy
Tris	tris(hydroxymethyl)aminomethane
Trt	trityl
Trx	thioredoxin
TSP	3-(trimethylsilyl)-propionate
TTX	tetrodotoxin
TTX-R	tetrodotoxin-resistant
TTX-S	tetrodotoxin-sensitive
UV	ultraviolet
v	version

Abbreviations for the common amino acids (L-isomers unless indicated otherwise) are in accordance with the recommendations of the IUPAC-IUB Joint Commission on Biochemical Nomenclature (Eur. J. Biochem. 1984, 138:9-37).

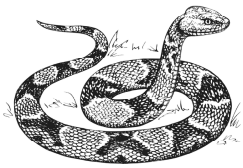
Introduction

1 Overview

A widespread evolutionary strategy in the animal kingdom is the use of venom for predation and defence. Venom-producing species can be found among many animal groups, including arachnids, reptiles, amphibians, insects, fish, molluscs and even some mammals.[1] The venom of each species is a distinct mixture of toxic components, which vary in their composition and effect. Some toxins are large proteins (especially common in snake venom), and some are small molecules, but the venoms of many species are rich in peptides (≤100 amino acid residues) which are commonly reticulated by disulfide bonds.[2] These venom peptides have been subject to selective pressure over millions of years of evolution, and many have shown both significant potency and exquisite selectivity for their target receptors. These properties make such molecules very valuable as tool compounds for further research, and even as leads for developing new therapeutics.[3]

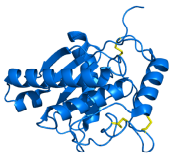
The estimated diversity of toxin sequences is vast. For example, each of the 700 known species of marine cone snail have been shown to produce 100–200 toxins in their venom,[4] and more recent reports using high-resolution mass spectroscopy (MS) techniques have raised that figure to 1000 toxins per venom.[5, 6] Relatively few sequences are found in multiple species, giving an estimate of over 500,000 distinct toxin sequences in this single genus.[5] The total repertoire of spider toxins is estimated to be 12 million sequences or more,[7] while in scorpions 100,000 toxins are predicted to exist across 1500 species.[8] Animal venom is therefore a very rich source of biodiversity with the potential to yield interesting compounds. Despite the efforts of several research groups across the globe, relatively few sequences have been examined in any depth – some estimates state that fewer than 0.1% of cone snail toxins have been investigated thus far.[9] Therefore, an opportunity exists to explore this diversity and harness these molecules for scientific and medical benefit.

This thesis describes an investigation of peptide toxins, with the ultimate objective of expanding the number of toxin sequences with known properties which may be of use in future applications. One avenue pursued was the potential to use toxin structures in peptide engineering applications for the display of functional epitopes. A separate effort was devoted to understanding the structure and function of novel toxin sequences, to determine if they may possess novel activity to go with their novel sequence.

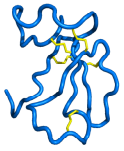


Snakes

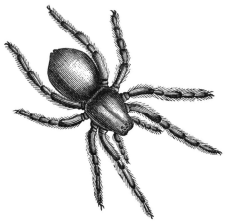
2200 venomous species
~90,000 toxins
Systems affected:
neuronal
cardiovascular
muscular



BaP1
metalloprotease enzyme

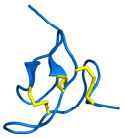


α-bungarotoxin
nAChR antagonist



Spiders

42,000 species
~12,000,000 toxins
Wide range of activities:
neurotoxic analgesic
haemolytic enzyme inhibition
ICK folds dominate



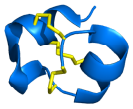
phrixotoxin 1
K_V channel antagonist

psmalmotoxin 1
ASIC1a antagonist

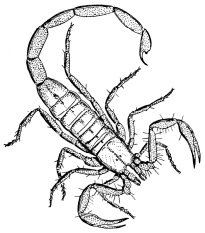
Cone Snails

700 species
~500,000 toxins
Target: ion channels
neurotransmitter receptors

α-ImI
nAChR antagonist



μ-BuIIIB
Na_V channel antagonist



Scorpions

1,500 species
~100,000 toxins
Primarily target ion channels:
Na^+ Ca^{2+}
K^+ Cl^-
CSαβ folds dominate

charybdotoxin
K_V channel antagonist

chlorotoxin
Cl^- channel antagonist

Centipedes

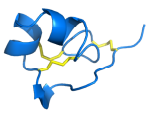
3,500 species
~300,000 toxins
Largely unstudied to date
Evidence of: enzymatic activity
pore-forming toxins
Cys-rich peptides

μ-SLPTX-Ssm6a
Na_V channel antagonist

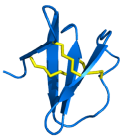
SsTx
KCNQ K^+ channel
antagonist

Sea anenomes

1,100 species
~500,000 toxins
Few toxins studied
Evidence of: cytolytic activity
neurotoxicity



ShK
K_V1.3 channel antagonist



APETx1
hERG channel antagonist

Figure 1. Summary of diversity of toxins found in the venom of various animal groups.

2 Toxins as research tools

As already stated, many toxins are both highly potent and highly selective for their biochemical targets. This can make them valuable tools for studying the macromolecules they affect. The proper functioning of cells requires the coordinated action of a vast array of biomolecular machinery, including ligand receptors, ion channels and more. Even within a certain class of molecule, such as voltage-gated potassium ion channels, there are often several different subtypes that are expressed in different tissues and make a different contribution to the overall health and functioning of the organism.[10] Researchers often need to isolate or block a certain subtype of receptor (rather than the whole population) in order to fully understand the system under investigation. Many molecules derived from animal venom have become valuable tools to facilitate this kind of research.[10]

One widely known example is tetrodotoxin (TTX), a small molecule toxin found in the tissues of pufferfish and other marine species that blocks voltage-gated sodium channels.[11] Even before the existence of different subtypes of this channel was recognised, it was known that some excitable cells and neuronal populations were sensitive to the toxin while others were resistant.[12] TTX was used extensively in research to separate the components of membrane depolarisation events: those of the transient current (due to sodium) and the steady-state current (due to potassium).[13-16] This afforded insights into the mechanisms of membrane depolarisation and the internal workings of excitable cells. Now that the various subtypes of sodium channel have been identified and enumerated ($Na_V1.1$–$Na_V1.9$), they can be classified into being either tetrodotoxin resistant (TTX-R, subtypes $Na_V1.8$ and $Na_V1.9$) or tetrodotoxin sensitive (TTX-S, subtypes $Na_V1.1$–$Na_V1.7$), which has allowed finer probing of the particular channels involved in various processes.

Other toxins have also shown their value in this type of investigation. The μ-conotoxins are a class of venom peptide from marine cone snails that have potent effects at voltage-gated sodium channels. In one study, a panel of 11 μ-conotoxins was tested on isolated preparations of channel subtypes $Na_V1.1$ through $Na_V1.8$, expressed in *Xenopus laevis* oocytes.[17] The researchers found that the overlapping selectivity profiles of the different toxins allowed discrimination between all subtypes except $Na_V1.8$ (which none of the toxins blocked and was also the sole TTX-R channel tested). Based on these data, μ-conotoxins have been used to determine the Na_V subtypes responsible for sodium currents in a variety of neuronal populations.[17-20]

Research on potassium channels has also benefitted from the availability of toxin tool compounds, with some studies showing that toxin probes can reveal information on structure as well as function. For example, the scorpion toxin charybdotoxin was used to determine the subunit stoichiometry of the Shaker potassium channel.[21] In 1991, it was known that the Shaker channel was a multimeric membrane protein, but the number of subunits required to form a functional channel was unknown. Charybdotoxin was known to block the passage of K^+ ions through the channel, but a point mutation in the channel subunit raised the inhibition constant by a factor of 250, resulting in a much less sensitive channel. Co-expressing the sensitive wild-type subunit and the resistant mutant subunit in *Xenopus* oocytes gave rise to a channel population with mixed subunits that still displayed marked toxin sensitivity. The observed K^+ current was the aggregate of the unblocked current of all channels in the population – channels that could have any number of sensitive subunits, from zero to the total number of subunits in a channel. By measuring the unblocked current with different ratios of sensitive and resistant subunits expressed, it was possible to calculate the number of subunits in a channel, yielding valuable information on the structure of the complex.[21] The channel was predicted to be a tetramer, and this has been borne out by determination of the structure of a potassium channel.[22]

Toxins have also been used to probe finer details of potassium channel structure. In addition to charybdotoxin, the Shaker channel is also known to interact with the scorpion toxin agatoxin 2 (AgTx2) via electrostatic interactions. By mutating charged residues on both the channel and the toxin in a manner that closed a thermodynamic cycle (a technique termed 'mutant cycle analysis'), it was possible to discover pairs of toxin/channel residues that interacted with one another.[23] This allowed the toxin binding site to be mapped on the pore region of the channel, from which a structural constraint was derived that showed the pore region extended much farther from the channel axis than had been thought previously.[23] Again, this prediction was validated when a crystal structure of a related channel was solved some years later.[22, 24]

Thus, it is clear that toxin molecules can be very useful as probes and tools for molecular biology. With the vast majority of animal venom components still uncharacterised, there is the potential for many more useful compounds to be discovered and put to use.

3 Venom-derived therapeutics

The use of venom in a medicinal capacity has been part of traditional practice in many different cultures.[25] In recent decades, venom components have also been used as the basis for modern pharmaceuticals. The first drug derived from a venom component to be approved by the FDA was the angiotensin converting enzyme inhibitor (ACE inhibitor) captopril, developed from the venom of the pit viper *Bothrops jararaca* and placed on the market in 1981.[25, 26] The ACE inhibitors are a class of drug used in the treatment of hypertension that work by blocking the cleavage of angiotensin I and thus the production of angiotensin II.[27] Angiotensin II leads to vasoconstriction by acting on its receptor, and also simulates aldosterone secretion.[28] Aldosterone is a hormone that increases sodium reabsorption in the kidneys and also raises blood pressure.[29] Therefore, inhibiting the action of ACE and reducing the levels of angiotensin II in the blood has a two-fold effect in lowering blood pressure. Captopril was the first orally available drug in this class and has proved a phenomenal success, bringing in over a billion dollars for Squibb, the company that developed it.[30, 31]

The success of captopril proved that venom-derived drugs could compete in the marketplace. In 2011, the list of FDA-approved therapeutics mimicking or derived from venom peptides included six molecules.[25]

One of those is ziconotide (trade name PRIALT®), an alternative to morphine in the treatment of chronic pain. Clinical development of this molecule began in 1995, and FDA approval was granted in 2004.[32] Ziconotide is a synthetic version of the ω-conotoxin MVIIA from the cone snail *Conus magus* (see Section 5 for details on conotoxin naming practices), and is a 25-residue peptide with a hydrophilic character.[33, 34] Ziconotide exerts its effects by blocking N-type voltage-gated calcium channels in the presynaptic terminal of the dorsal horn of the spinal column.[35] Calcium ion influx through these channels is required for the release of neurotransmitters to propagate nociceptive signals across the synapse. By inhibiting the flow of ions, the pain signal is prevented from continuing to the brain, thus providing analgesia.[33] Ziconotide is >1000-fold more selective for N-type channels than any other type,[36] which is a key factor in its tolerability and clinical efficacy. However, the peptide is not orally bioavailable, and is also rapidly degraded by proteases in the vasculature, necessitating intrathecal administration (that is, an infusion of the molecule directly into the cerebrospinal fluid).[37] Many peptide drugs suffer from similar

issues in administration. Patients are less likely to favour such invasive therapies, but where the clinical need is great enough – as is the case with refractory chronic pain and ziconotide – a less convenient treatment is preferable to no treatment. A venom-derived therapy for headaches or bacterial infection is unlikely to gain traction in the marketplace, but in more poorly serviced clinical indications the potency and selectivity of venom-derived therapeutics can provide the competitive advantage needed to succeed. Continued research on the toxic peptides found in animal venom may uncover molecules that bear the novel activity required to meet these needs.

4 Scorpion toxins

One of the peptide toxins investigated for this thesis was a scorpion toxin from the species *Heterometrus spinifer*, named HsTX1. A detailed background on this molecule is presented in Section 4.4, but first some background on scorpion toxins in general is required.

Scorpions are among the most ancient of animal groups, having changed little over 400 million years of evolution.[38] Scorpion venom consists of a mixture of components,[39] perhaps the best studied of which are neurotoxic peptides that act at Na^+, K^+, Ca^{2+} and Cl^- ion channels.[40] There is no standardised nomenclature for scorpion toxins, with most molecules being assigned an ad-hoc name which may or may not relate to some property of the toxin. Some names and abbreviations relate back to the species from which the toxin was isolated — thus 'maurotoxin' (from *Scorpio maurus*)[41] and 'TsIV-5' (from *Tityus serrulatus*).[42] 'Chlorotoxin' was thought to act on chloride ion channels,[43] while 'kurtoxin' was isolated from *Parabuthus transvaalicus* and acts on T-type voltage gated calcium ion channels.[44]

Despite the confusion in naming, some attempts have been made to classify scorpion toxins in logical groupings. These attempts have largely focused on the neurotoxic peptides that act on ion channels, as these are the molecules that have been studied most intently. However, more recent investigations have uncovered a plethora of other molecules found in scorpion venom, which have been taken into account in some of the systems described below.

4.1 Classification by chain length

One widely used distinction is drawn between 'long-chain' scorpion toxins (53–78 residues) and 'short-chain' scorpion toxins (23–42 residues).[45] Some sources in the literature state that long-chain toxins are active on Na^+ channels, while short-chain toxins affect K^+ channels.[46, 47] As research has continued, however, it is clear that although this can be generalised as being 'usually true' the distinction is not so clear-cut, as there is also a subset of long-chain K^+ channel toxins.[48]

4.2 Classification by structural motif

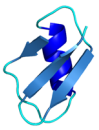
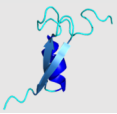
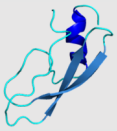
Classifications have also been made based on the three-dimensional structure of toxin molecules. Many scorpion toxins adopt a structure characterised by an α-helix connected to an antiparallel β-sheet by three or four disulfide bonds, termed the 'cysteine-stabilised α-helical and β-sheet' (or CSαβ) fold. Early reports noted that this structure was universal for known scorpion toxins,[40] but ongoing research has uncovered toxins bearing different structural motifs.

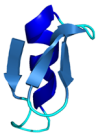
One of these motifs consists of two α-helices connected by two disulfide bonds, called the 'cysteine-stabilised α-helix-loop-helix' (CSαα) fold,[49] found in certain K^+ channel toxins. Scorpion toxins have also been discovered that conform to the 'inhibitory cystine knot' (ICK) motif, which is also widely found in toxins from other animal groups, such as snakes and cone snails. (For a more detailed discussion of this motif, see Section 6.3.) There is also a class of scorpion venom peptides that eschews the concept of structural motifs: the non-disulfide-bridged peptides (NDBPs). These largely adopt an α-helical conformation, and display a variety of properties such as bradykinin potentiating activity, antimicrobial and haemolytic action and immune modulating effects.[50, 51] These peptides have traditionally been neglected in favour of the neurotoxic peptides, but are now becoming an active field of study in their own right.

4.3 Classification by ion channel family

Another system of classification that has received much attention is based primarily upon classes determined by the ion channel affected by a toxin. Within each class, families are created that group toxins with similar structures or mechanisms of action, which may again be further subdivided. **Table 1** outlines the families that have been defined for the major classes of toxin. This system is most robust for toxins affecting Na^+ and K^+ channels, as relatively few molecules have been reported that act on Ca^{2+} and Cl^- channels.

Table 1. Classification and characteristics of scorpion toxin families that affect different ion channels.

Family	Characteristics		Refs
	Na^+ toxins (NaScTx)		
α	Voltage-dependent binding to receptor site 3 on channel Slow or block channel inactivation mechanism Long-chain, CSαβ fold		*8, 47* 1LQH
β	Voltage-independent binding to receptor site 4 on channel Affect channel activation (shift activation potential) Long-chain, CSαβ fold		*8, 47* 1B3C
	K^+ toxins (KTx)		
α	Two interaction modes: 'pore-plugging' (K-Y/F dyad) and 'intermediate' (charge interactions) Inhibit K^+ current Short-chain, CSαβ fold		*52* 2CRD
β	Inhibit K^+ current Long-chain, CSαβ fold		*52, 53* 5IPO
γ	Primarily target ERG channels Bind on turret region rather than pore Short-chain, CSαβ fold		*52, 54* 1J5J
δ	Pore-plugging (K-Y/F dyad) action Inhibit K^+ current, also inhibit serine proteases Kunitz domain fold		*55, 56* 2M01
ε	(Proposed, two members currently described) Weak inhibition of K^+ current ICK fold, lacking regular secondary structure		*56* 2MSF
κ	Pore-plugging (K-Y/F dyad) action Inhibit K^+ current Short-chain, CSαα fold		*52* 1HP9
λ	Pore-plugging (K-Y/F dyad) action Inhibit K^+ current ICK fold		*54* 1C6W

Family	Characteristics		Refs
	Ca^{2+} toxins		
Kurtoxin-like	Bind T-type calcium channels Inhibit channel by modifying voltage-dependant gating mechanism Also active on Na^{+} channels Long-chain, CSαβ fold		*44, 57, 58* 1T1T
Calcins	Affect ryanodine-sensitive calcium channels Varying mechanisms of action ICK fold		*52, 54* 1IE6
	Cl^{-} toxins		
None defined	Too few toxins have been discovered to support distinction and classification into meaningful groups Discovered toxins vary in mechanism		*43, 59-61* 1CHL

4.4 Project target: Scorpion toxin HsTX1

Now our discussion on the organisation of scorpion toxins is concluded, we can describe the scorpion toxin that was the focus of the first part of this project. HsTX1 is a member of the α-KTx family, isolated from the venom of the scorpion *Heterometrus spinifer* (the Malaysian black scorpion).[62] This peptide was identified by its potent antagonism of the rat $K_V1.3$ channel and found to be a 34-residue chain constrained by four disulfide bonds.[63] The molecule bears the CSαβ motif, with an α-helix packed against an antiparallel two-strand β-sheet (**Figure 2**).

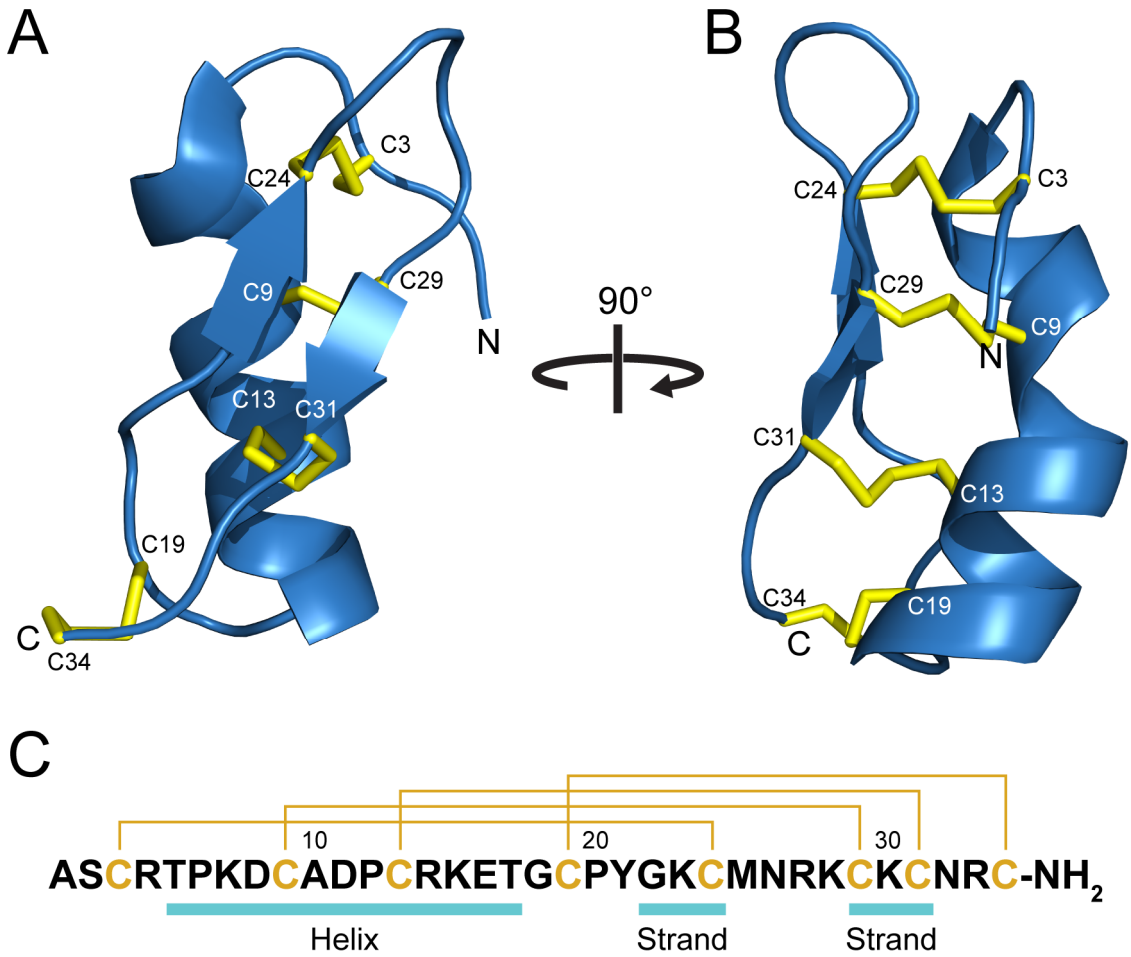


Figure 2. Structure and sequence of wildtype HsTX1. (A) Front view of peptide structure. Disulfide bonds are *yellow*. (B) Side view of peptide structure. (C) Peptide sequence. Disulfide connectivity is indicated by gold staples above Cys residues, secondary structure elements are indicated by underscores below relevant residues.

The potassium channel $K_V1.3$ is of interest as a target in the treatment of autoimmune diseases, as it is overexpressed in activated effector memory T cells.[64] The initial impetus for studying HsTX1 in our laboratory was to evaluate the molecule for potential use as a therapeutic in such disease states, but during this investigation it was found to be remarkably stable against proteolytic degradation and thermal denaturation. Moreover, despite the complexity inherent in a structure featuring four disulfide bonds it folded readily into its active conformation. This combination of stability and folding efficiency suggested that it may be suitable as a scaffold molecule for peptide engineering. The work undertaken in investigating this possibility is described in the following chapter, *Peptide grafting design using molecular dynamics*.

5 Conotoxins and conopeptides

The novel toxin sequences investigated as part of this thesis were discovered in the venom duct transcriptome of the cone snail *Conus victoriae*. Details of the peptides investigated are provided in Sections 6.4 and 6.6, following some background material on the organisation and classifications of cone snail toxins.

Marine cone snails are venomous predators that have evolved to produce venom rich in small peptide toxins, referred to as 'conotoxins' or 'conopeptides'. Historically the term 'conotoxin' was reserved for peptides bearing multiple disulfide bonds, but this distinction has fallen out of use and the two terms are now used interchangeably.[65] The existence of peptide toxins in cone snail venom was determined in the 1970s,[66] and the first complete amino acid sequence described was for the 13-residue α-conotoxin GI, from the venom of *Conus geographus*.[67] Since then, over 2000 conotoxins have been described and the rate of discovery has increased by the application of transcriptomic technology.[68]

Conotoxins are produced in the venom duct as large precursor proteins (~80–100 residues)[69] consisting of three distinct regions,[70] shown in **Figure 3**. At the N-terminus is a signal presequence, which is generally hydrophobic and is highly conserved among related conotoxins.[65] The intermediate pro-region is typically highly charged and thought to assist with folding.[71] The mature toxin region is found at the C-terminus, usually as a single copy, and is released by proteolytic cleavage as part of post-translational processing.[72] Many mature toxins undergo further post-translational modification, incorporating unusual amino acids such as 4-hydroxyproline and γ-carboxyglutamate, or modifying the C-terminus with amidation.[69]

Figure 3. Schematic diagram of conotoxin precursor protein. The precursor is cleaved at the position indicated by the triangles to generate the mature toxin.

The great diversity of conotoxins has led to the development of several complementary classification systems, associating peptides by evolutionary relationships, by structure and by function. The three classifications are made based on gene superfamily, cysteine framework, and pharmacological family.

5.1 Classification by gene superfamily

The definition of conotoxin 'gene superfamilies' is based on the sequence identity of the precursor protein signalling sequence, which is highly conserved between related conotoxins.[65] These superfamilies group conotoxins that are related by evolution. The distinctive presequences are highly conserved within a gene superfamily, but share little sequence identity with presequences of other gene superfamilies. For an analysis of the sequence identity within and between superfamilies from a single species, see **Table 2**.

Table 2. Sequence identity within and between gene superfamily presequences for peptides discovered in the venom duct transcriptome of *Conus marmoreus*. (Reproduced from [73])

A	B	E	F	H	I1	I2	M1	M2	MC	N	O1	O2	S	T	
95	25	33	24	24	26	14	35	22	12	33	33	39	35	32	A
	100	25	25	37	19	12	12	25	12	25	19	50	19	25	B
		100	17	22	17	28	29	22	12	17	39	17	41	22	E
			100	27	26	23	47	33	47	23	26	39	18	37	F
				91	26	32	35	39	24	41	18	28	29	26	H
					79	26	29	28	35	32	26	28	24	37	I1
						100	24	28	12	20	23	22	29	26	I2
							82	65	76	41	35	35	59	29	M1
								89	71	56	17	39	59	33	M2
									100	53	18	35	41	29	MC
										96	27	33	35	26	N
											86	33	24	26	O1
												72	24	33	O2
													100	29	S
														79	T

Numbers indicate % sequence identity. Dark blue background = Min. identity within gene superfamily. Light background = Max. identity between gene superfamilies.
MC = conomarphin (part of the M superfamily).

One strategy useful in the search for new conotoxin sequences is to screen cDNA libraries for known superfamily presequences.[74] Of course, this approach will only identify new members of existing superfamilies and will not be able to uncover novel superfamilies. The search for novel superfamilies often relies on non-genomic sources of information. For example, peptides belonging to the H-superfamily were first identified in MS/MS data, which was then linked back to transcriptomic data and identified the presequence for this new superfamily.[73]

5.2 Classification by cysteine framework

In comparison to toxins from other invertebrate sources such as scorpions, spiders and centipedes, mature cone snail toxins are unusually short. Approximately 80% of those studied are between 12–33 residues in length,[75] compared to (for example) the short-chain scorpion toxins that are typically 23–42 residues.[45] Despite their short length, many conotoxins contain multiple disulfide bonds in distinct cysteine frameworks (**Figure 4**). These frameworks are thought to stabilise the specific three-dimensional structure of the toxin, allowing the amino acid sidechains from residues in the inter-cysteine loops to be displayed in the proper orientation for binding to the molecular target.[76, 77]

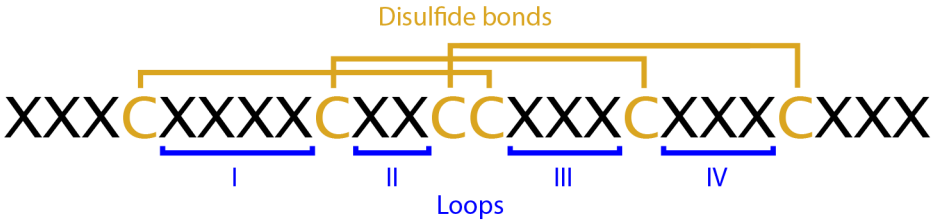


Figure 4. Schematic diagram of cysteine framework VI/VII, showing typical disulfide bond connectivity I–IV, II–V, III–VI.

There are currently 27 recognised frameworks (shown in **Table 3**), each defined by the number and arrangement of cysteine residues in the primary sequence. The rapid pace of transcriptomic discovery is adding new frameworks frequently. The frameworks do not define how many residues lie between non-vicinal cysteine residues, merely whether a loop is present or not. This means that loops can vary in length as well as composition, which in turn allows conotoxins that share the same framework enough sequence diversity to target different macromolecular entities.

Table 3. Recognised cysteine frameworks used in the classification of conotoxins. (Adapted from [78])

Name	# Cysteines	Pattern
I	4	CC—C—C
II	6	CCC—C—C—C
III	6	CC—C—C—CC
IV	6	CC—C—C—C—C
V	4	CC—CC
VI/VII	6	C—C—CC—C—C
VIII	10	C—C—C—C—C—C—C—C—C—C
IX	6	C—C—C—C—C—C
X	4	CC—C—C
XI	8	C—C—CC—CC—C—C
XII	8	C—C—C—C—CC—C—C
XIII	8	C—C—C—CC—C—C—C
XIV	4	C—C—C—C
XV	8	C—C—CC—C—C—C—C
XVI	4	C—C—CC
XVII	8	C—C—CC—C—CC—C
XVIII	6	C—C—CC—CC
XIX	10	C—C—C—CCC—C—C—C—C
XX	10	C—CC—C—CC—C—C—C—C
XXI	10	CC—C—C—C—CC—C—C—C
XXII	8	C—C—C—C—C—C—C—C
XXIII	6	C—C—C—CC—C
XXIV	4	C—CC—C
XXV	6	C—C—C—C—CC
XXVI	8	C—C—C—C—CC—CC
XXVII	8	C—C— C—CCC—C—C
Novel (from[79])	6	C—CC—C—C—C

5.3 Classification by pharmacological family

Conotoxins with the same biological activity are grouped together into pharmacological families, each designated by a Greek letter (**Table 4**). For example, κ-conotoxins antagonise voltage-gated K^+ channels, while μ-conotoxins antagonise voltage-gated Na^+ channels. Voltage-gated Na^+ channels are also targeted by δ-conotoxins, but these peptides delay the fast inactivation of such channels and are thus separate from μ-conotoxins.[9] Classifying conotoxins in this way allows peptides with the same activity to be compared against each other or analysed as a group, giving insights into the pharmacology of their target and clues to the mechanism of action.[80]

Table 4. Conotoxin pharmacological families.

Name	Target	Effect
α	Nicotinic acetylcholine receptor (nAChR)	Antagonist
γ	Voltage-gated pacemaker channel	Agonist
δ	Voltage-gated sodium channel (Na_V)	Delays fast inactivation
ε	Presynaptic Ca^{2+} channel or GPCRs	Antagonist
ι	Voltage-gated sodium channel (Na_V)	Agonist
κ	Voltage-gated potassium channel (K_V)	Antagonist
μ	Voltage-gated sodium channel (Na_V)	Antagonist
μO	Voltage-gated sodium channel (Na_V)	Gating modifier
μO§	Voltage-gated sodium channel (Na_V)	Antagonist
ρ	α1-adrenoceptor (GPCR)	Antagonist
σ	Serotonin-gated ion channel	Antagonist
τ	Somatostatine sst3 receptor (GPCR)	Antagonist
χ	Noradrenaline transporter (NET)	Antagonist
ψ	Nicotinic acetylcholine receptor (nAChR)	Allosteric antagonist
ω	N-type voltage-gated calcium channel (Ca_V)	Antagonist

5.4 Conotoxin standard nomenclature

The standard nomenclature for describing conotoxins includes structural and functional information in the toxin name, based on the set of classification schemes previously described (see **Figure 5**).[81, 82] The pharmacological family of the toxin is indicated by the corresponding Greek letter (**Table 4**). Species of origin is abbreviated to one or two letters, followed by a Roman numeral denoting the cysteine framework (**Table 3**). Order of

discovery is communicated by an uppercase letter, which serves as a unique identifier between toxins that would otherwise share the same name.

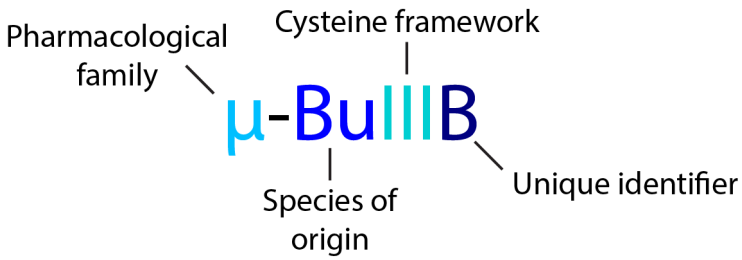


Figure 5. Identification of component parts of standard conotoxin nomenclature. This toxin blocks voltage-gated sodium channels (it is a μ-conotoxin), was discovered in the venom of *Conus bullatus* (abbreviated 'Bu'), displays a CC—C—C—CC cysteine framework (type III), and was the second such toxin identified (denoted by 'B').

6 *Conus victoriae* venom duct transcriptome

Conus victoriae is a molluscivorous cone snail that is endemic to the north-western coast of Australia.[83] Like all known cone snail species it is venomous, and a recent study in the Norton laboratory aimed to document the diversity of sequences produced in the venom duct. This study made use of the next generation paradigm that has been termed 'venomics', mining the transcriptome of the venom duct and matching the sequence data obtained with experimental results from MS/MS matching of the crude venom. Prior to this study, 24 conotoxin sequences were known from *C. victoriae*, spread across four superfamilies (A, O1, O2 and T). The venomics approach discovered 113 sequences, spread across 20 gene superfamilies.[84] Seven of the previously known sequences were not detected. (Robinson, S.D.; personal communication) A summary of the sequences discovered is presented in **Table 5**.

Table 5. Summary of sequences discovered from the venom duct transcriptome of *C. victoriae*. (Adapted from [84]) N.D. = not determined.

Superfamily	Cysteine framework	Number of sequences identified
A	I	2
	XXII	1
Conantokin (B)	Cysteine-free	1
B2	Cysteine-free	1
E	N.D.	1
F	N.D.	1
H	VI/VII	2
	Cysteine-free	1
I1	XI	6
I2	XI	4
I4	XII	1
J	XIV	4
M1	III	4
M2	III	6
Conomarphin (M)	Cysteine-free	2
M	Single disulfide	1
O1	VI/VII	20
	Cysteine-free	1
O2	VI/VII	18
Contryphan (O2)	Single disulfide	2
O3	Cysteine-free	1
P	IX	2
	XIV	1
S	VIII	1
T	V	24
	XIII	1
	X	1
U	VI/VII	2
Con-ikot-ikot	XXI	1
Conodipine		3

Several of the sequences identified from the transcriptome were tempting targets for further research, but the sheer number made it impractical to fully explore the newly recognised diversity of conotoxins in *C. victoriae* in a single project. The investigation documented in this thesis confined itself to three peptides – the two disulfide-bonded members of the H-superfamily, named H_Vc7.1 and H_Vc7.2, as well as contryphan-Vc2 (part of the O2-superfamily).

6.1 H-superfamily toxins

The H-superfamily was first described by Dutertre et al. in 2013, with seven sequences being discovered in a deep venomics investigation of *Conus marmoreus* venom.[73] The presequence for this superfamily exhibited <50% sequence identity with other known superfamily presequences. Being a novel superfamily it was not detected in the initial homology search of the transcriptomic database generated in this study and only became apparent after analysis of MS/MS data. Six of the seven sequences discovered conformed to the type VI/VII cysteine framework.[73] The two disulfide-bonded H-superfamily toxins discovered in the venom duct transcriptome of *C. victoriae* also share the type VI/VII cysteine framework.

6.2 Type VI/VII framework

The type VI/VII cysteine framework was first described in the conotoxin ω-GVIA,[85] and has since been found in almost 700 sequences, making it the most common framework known at the present time. The framework is composed of six cysteines separated by four loops, with the disulfide bonds connected in a I–IV, II–V, III–VI pattern. Conotoxins bearing this framework have been found to act at several different targets, including voltage-gated sodium channels (δ- and μ-conotoxins), voltage-gated calcium channels (ω-conotoxins), voltage-gated potassium channels (κ-conotoxins) and voltage-gated pacemaker channels (γ-conotoxins). Hence, knowledge of the cysteine framework does not provide a good indicator of the pharmacological target of a given toxin and thus provides no clues as to the activity of the H-superfamily toxins. Many conotoxins bearing this framework do share some structural similarity, folding into the defined toxin scaffold of the ICK motif.

6.3 Inhibitory cystine knot (ICK) motif

The ICK motif is a common structural scaffold, found in toxins from the venom of scorpions, cone snails and spiders as well as peptides from plants and fungi.[86] Peptides that adopt this motif usually bear an antiparallel β-sheet structural element, and always possess three disulfide bonds that form the 'knot' structure.

As can be seen in **Figure 6**, a ring is formed by two of the disulfide bonds (along with the intervening backbone residues of the peptide). The third disulfide bond is threaded through this ring, completing the 'knot'.[87]

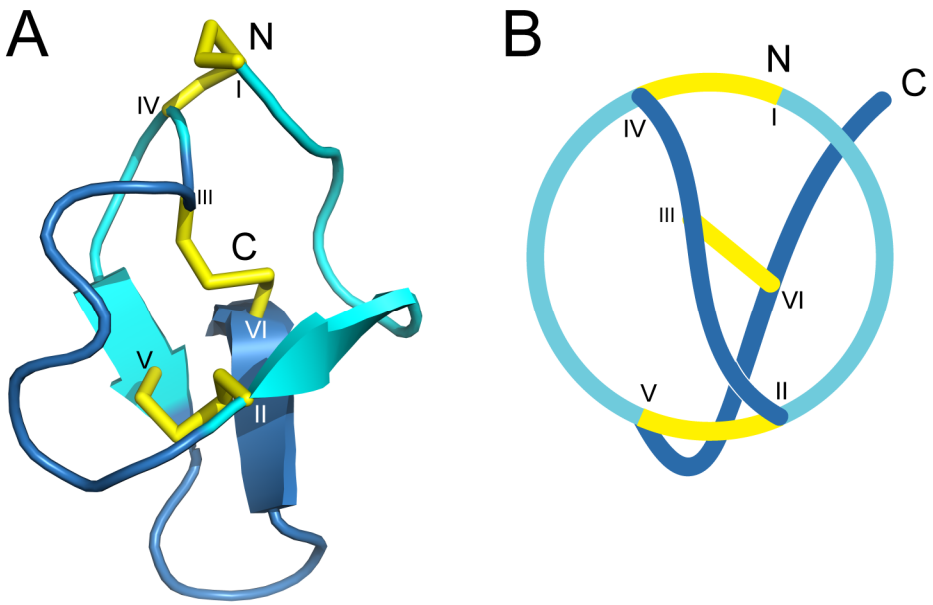


Figure 6. Representation of the ICK fold. The ring is *cyan*; disulfide bonds are *yellow*. (A) Conotoxin ω-GVIA, one of the first molecules in which the ICK motif was identified.[87] (B) Schematic diagram, illustrating the disulfide-bonded ring through which the third disulfide bond passes.

The first few molecules discovered with this motif inhibited ion channels, which led to the term 'inhibitory cystine knot'. As more molecules have been discovered that bear this motif, the collection have been found to include proteases, hormones, anti-microbial peptides and peptides with anti-HIV action and other effects.[88] This confirms that the structure is not a good indicator of eventual function. The ICK motif is noted for being exceptionally stable,[89] and has also been investigated as a possible scaffold for peptide engineering.[90]

6.4 Project targets: H_Vc7.1 and H_Vc7.2

Analysis of the transcriptomic data from the *C. victoriae* venom duct identified two sequences from the newly described H-superfamily which bore the type VI/VII cysteine framework, being named H_Vc7.1 and H_Vc7.2.[84] These peptides were also identified in MS data collected on crude venom samples, providing evidence at the protein level. The sequences of the peptides are given in **Figure 7**.

Figure 7. Sequences of H-superfamily toxins from venom duct transcriptomes of *C. victoriae* and *C. marmoreus*. Disulfide connectivity is assumed by homology with other framework VI/VII and ICK molecules.

The H-superfamily toxins from the *C. victoriae* venom duct transcriptome were chosen for further investigation for a number of reasons. Firstly, it was an opportunity to characterise peptides from a novel superfamily, perhaps establishing a precedent to guide the approaches of future researchers investigating other H-superfamily toxins. Secondly, the MS data collected indicated that the peptides were devoid of post-translational modifications, opening the possibility of producing them by recombinant expression. Thirdly, the novelty of the sequence gave rise to the hope that the toxins may have novel modes of action, expanding the range of molecules that may find use as research tools or therapeutic lead compounds.

6.5 Contryphans

The other peptide to be investigated from the *C. victoriae* venom duct transcriptome was contryphan-Vc2, a member of the contryphan family of conotoxins which was first identified in *Conus radiatus* in 1996.[91] The contryphans share a signal presequence with the O2 gene superfamily and contain only a single disulfide bond (thereby avoiding the concept of cysteine frameworks).[92] No convincing target has been established for the class, although individual contryphans have been reported to have activity at certain ion

channels.[93-95] Therefore contryphans have not been classified into a pharmacological family.

6.6 Project target: Contryphan-Vc2

Contryphan-Vc2 is an unusual contryphan, deviating from the consensus sequence that most members of the class conform to. This novelty recommended the molecule for further study, which is documented in the published work *Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue*, presented as Chapter 3 of this thesis.

7 Research question

From the literature, it is clear that peptide toxins can be useful both as tools and as therapeutic compounds. However, this utility is not universal, and each toxin must undergo a biochemical evaluation to establish its potential. The 113 sequences of the *C. victoriae* venom duct transcriptome are ripe for such investigations, although to fit within the scope of the current project only a few can be investigated. By contrast, the scorpion toxin HsTX1 is already under investigation for its potential therapeutic properties, but may also be useful as a tool in peptide engineering. This too is a possibility worth investigating. The research question for this thesis is therefore: What utility do the peptide toxins under investigation possess?

8 Aims

The specific aims of this project are to:

1. Evaluate the potential for scorpion toxin HsTX1 to be used as a scaffold in peptide engineering, using the RGD and DINNN motifs as example graft sequences. Design of hybrid molecules will be informed by molecular dynamics computer modelling, before production and testing of selected constructs.
2. Investigate the structure and function of novel conotoxin sequences discovered in the venom duct transcriptome of *C. victoriae*. Samples will be produced by recombinant expression or solid-phase peptide synthesis, and tested via nuclear magnetic resonance spectroscopy and mouse bioassay.

9 References

[1] Fry, B. G., Roelants, K., Champagne, D. E., Scheib, H., Tyndall, J. D. A., King, G. F., Nevalainen, T. J., Norman, J. A., Lewis, R. J., Norton, R. S., Renjifo, C., and Rodríguez de la Vega, R. C. (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms, In *Annu. Rev. Genomics Hum. Genet.*, pp 483-511, Annual Reviews, Palo Alto.

[2] Kordiš, D., and Gubenšek, F. (2000) Adaptive evolution of animal toxin multigene families, *Gene 261*, 43-52.

[3] Lewis, R. J., and Garcia, M. L. (2003) Therapeutic potential of venom peptides, *Nat. Rev. Drug Discov. 2*, 790-802.

[4] Olivera, B. M. (2006) *Conus* peptides: biodiversity-based discovery and exogenomics, *J. Biol. Chem. 281*, 31173-31177.

[5] Davis, J., Jones, A., and Lewis, R. J. (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS, *Peptides 30*, 1222-1227.

[6] Biass, D., Dutertre, S., Gerbault, A., Menou, J. L., Offord, R., Favreau, P., and Stocklin, R. (2009) Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*, *J. Proteomics 72*, 210-218.

[7] Saez, N. J., Senff, S., Jensen, J. E., Er, S. Y., Herzig, V., Rash, L. D., and King, G. F. (2010) Spider-venom peptides as therapeutics, *Toxins 2*, 2851-2871.

[8] Possani, L. D., Becerril, B., Delepierre, M., and Tytgat, J. (1999) Scorpion toxins specific for Na^+-channels, *Eur. J. Biochem. 264*, 287-300.

[9] Lewis, R. J., Dutertre, S., Vetter, I., and Christie, M. J. (2012) *Conus* venom peptide pharmacology, *Pharmacol. Rev. 64*, 259-298.

[10] Dutertre, S., and Lewis, R. J. (2010) Use of venom peptides to probe ion channel structure and function, *J. Biol. Chem. 285*, 13315-13320.

[11] Bane, V., Lehane, M., Dikshit, M., O'Riordan, A., and Furey, A. (2014) Tetrodotoxin: chemistry, toxicity, source, distribution and detection, *Toxins 6*, 693-755.

[12] Yoshida, S., Matsuda, Y., and Samejima, A. (1978) Tetrodotoxin-resistant sodium and calcium components of action potentials in dorsal root ganglion cells of the adult mouse, *J. Neurophysiol. 41*, 1096-1106.

[13] Narahashi, T. (1974) Chemicals as tools in the study of excitable membranes, *Physiol. Rev. 54*, 813-889.

[14] Hille, B. (1968) Pharmacological modifications of the sodium channels of frog nerve, *J. Gen. Physiol. 51*, 199-219.

[15] Narahashi, T., Shapiro, B., Deguchi, T., Scuka, M., and Wang, C. M. (1972) Effects of scorpion venom on squid axon membranes, *Am. J. Physiol. 222*, 850-857.

[16] Nakamura, Y., Nakajima, S., and Grundfest, H. (1965) The action of tetrodotoxin on electrogenic components of squid giant axons, *J. Gen. Physiol. 48*, 985-996.

[17] Wilson, M. J., Yoshikami, D., Azam, L., Gajewiak, J., Olivera, B. M., Bulaj, G., and Zhang, M.-M. (2011) μ-conotoxins that differentially block sodium channels $Na_V1.1$ through 1.8 identify those responsible for action potentials in sciatic nerve, *Proc. Natl. Acad. Sci. U. S. A. 108*, 10302-10307.

[18] Zhang, M.-M., Wilson, M. J., Gajewiak, J., Rivier, J. E., Bulaj, G., Olivera, B. M., and Yoshikami, D. (2013) Pharmacological fractionation of tetrodotoxin-sensitive sodium currents in rat dorsal root ganglion neurons by μ-conotoxins, *Br. J. Pharmacol. 169*, 102-114.

[19] Feng, B., Zhu, Y., La, J. H., Wills, Z. P., and Gebhart, G. F. (2015) Experimental and computational evidence for an essential role of $Na_V1.6$ in spike initiation at stretch-sensitive colorectal afferent endings, *J. Neurophysiol. 113*, 2618-2634.

[20] Wilson, M. J., Zhang, M.-M., Gajewiak, J., Azam, L., Rivier, J. E., Olivera, B. M., and Yoshikami, D. (2015) α- and β-subunit composition of voltage-gated sodium channels investigated with μ-conotoxins and the recently discovered μO§-conotoxin GVIIJ, *J. Neurophysiol. 113*, 2289-2301.
[21] MacKinnon, R. (1991) Determination of the subunit stoichiometry of a voltage-activated potassium channel, *Nature 350*, 232-235.
[22] Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity, *Science 280*, 69-77.
[23] Hidalgo, P., and MacKinnon, R. (1995) Revealing the architecture of a K^+ channel pore through mutant cycles with a peptide inhibitor, *Science 268*, 307-310.
[24] Imredy, J. P., and MacKinnon, R. (2000) Energetic and structural interactions between δ-dendrotoxin and a voltage-gated potassium channel, *J. Mol. Biol. 296*, 1283-1294.
[25] King, G. F. (2011) Venoms as a platform for human drugs: translating toxins into therapeutics, *Expert Opin. Biol. Ther. 11*, 1469-1484.
[26] Cushman, D. W., and Ondetti, M. A. (1991) History of the design of captopril and related inhibitors of angiotensin converting enzyme, *Hypertension 17*, 589-592.
[27] Brown, N. J., and Vaughan, D. E. (1998) Angiotensin-converting enzyme inhibitors, *Circulation 97*, 1411-1420.
[28] Griendling, K. K., Murphy, T. J., and Alexander, R. W. (1993) Molecular biology of the renin-angiotensin system, *Circulation 87*, 1816-1828.
[29] Lyubarova, R., and Gosmanova, E. O. (2017) Mineralocorticoid receptor blockade in end-stage renal disease, *Curr. Hypertens. Rep. 19*, 40.
[30] Hodsman, G. P., and Robertson, J. I. S. (1983) Captopril – 5 years on, *Br. Med. J. 287*, 851-852.
[31] Antunes, A. M. D., Guerrante, R. D., Avila, J. D. C., Mendes, F. M. L., and Fierro, I. M. (2016) Case study of patents related to captopril, Squibb's first blockbuster, *Expert Opin. Ther. Pat. 26*, 1449-1457.
[32] Klotz, U. (2006) Ziconotide — a novel neuron-specific calcium channel blocker for the intrathecal treatment of severe chronic pain – a short review, *Int. J. Clin. Pharm. Ther. 44*, 478-483.
[33] Miljanich, G. P. (2004) Ziconotide: neuronal calcium channel blocker for treating severe chronic pain, *Curr. Med. Chem. 11*, 3029-3040.
[34] Pope, J. E., and Deer, T. R. (2013) Ziconotide: a clinical update and pharmacologic review, *Expert Opin. Pharmacother. 14*, 957-966.
[35] Schmidtko, A., Lötsch, J., Freynhagen, R., and Geisslinger, G. (2010) Ziconotide for treatment of severe chronic pain, *Lancet 375*, 1569-1577.
[36] McGivern, J. G. (2007) Ziconotide: a review of its pharmacology and use in the treatment of pain, *Neuropsychiatr. Dis. Treat. 3*, 69-85.
[37] Kress, H. G., Simpson, K. H., Marchettini, P., Ver Donck, A., and Varrassi, G. (2009) Intrathecal therapy: what has changed with the introduction of ziconotide, *Pain Pract. 9*, 338-347.
[38] Ma, Y., He, Y., Zhao, R., Wu, Y., Li, W., and Cao, Z. (2012) Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: implication for proteome evolution of scorpion venom arsenal, *J. Proteomics 75*, 1563-1576.
[39] Rodríguez de la Vega, R. C., Schwartz, E. F., and Possani, L. D. (2010) Mining on scorpion venom biodiversity, *Toxicon 56*, 1155-1161.

[40] Possani, L. D., Merino, E., Corona, M., Bolivar, F., and Becerril, B. (2000) Peptides and genes coding for scorpion toxins that affect ion-channels, *Biochimie 82*, 861-868.

[41] Kharrat, R., Mabrouk, K., Crest, M., Darbon, H., Oughideni, R., Martin-Eauclaire, M.-F., Jacquet, G., El Ayeb, M., Van Rietschoten, J., Rochat, H., and Sabatier, J.-M. (1996) Chemical synthesis and characterization of maurotoxin, a short scorpion toxin with four disulfide bridges that acts on K^+ channels, *Eur. J. Biochem. 242*, 491-498.

[42] Kirsch, G. E., Skattebøl, A., Possani, L. D., and Brown, A. M. (1989) Modification of Na channel gating by an α scorpion toxin from *Tityus serrulatus*, *J. Gen. Physiol. 93*, 67-83.

[43] DeBin, J. A., Maggio, J. E., and Strichartz, G. R. (1993) Purification and characterization of chlorotoxin, a chloride channel ligand from the venom of the scorpion, *Am. J. Physiol. Cell Physiol. 264*, C361-C369.

[44] Chuang, R. S., Jaffe, H., Cribbs, L., Perez-Reyes, E., and Swartz, K. J. (1998) Inhibition of T-type voltage-gated calcium channels by a new scorpion toxin, *Nat. Neurosci. 1*, 668-674.

[45] Diego-García, E., Schwartz, E. F., D'Suze, G., González, S. A. R., Batista, C. V. F., García, B. I., Rodríguez de la Vega, R. C., and Possani, L. D. (2007) Wide phylogenetic distribution of Scorpine and long-chain β-KTx-like peptides in scorpion venoms: identification of "orphan" components, *Peptides 28*, 31-37.

[46] Céard, B., Martin-Eauclaire, M.-F., and Bougis, P. E. (2001) Evidence for a position-specific deletion as an evolutionary link between long- and short-chain scorpion toxins, *FEBS Lett. 494*, 246-248.

[47] Rodríguez de la Vega, R. C., and Possani, L. D. (2005) Overview of scorpion toxins specific for Na^+ channels and related peptides: biodiversity, structure–function relationships and evolution, *Toxicon 46*, 831-844.

[48] Zhu, S., Li, W., Zeng, X., Jiang, D., Mao, X., and Liu, H. (1999) Molecular cloning and sequencing of two 'short chain' and two 'long chain' K^+ channel-blocking peptides from the Chinese scorpion *Buthus martensii* Karsch, *FEBS Lett. 457*, 509-514.

[49] Chagot, B., Pimentel, C., Dai, L., Pil, J., Tytgat, J., Nakajima, T., Corzo, G., Darbon, H., and Ferrat, G. (2005) An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*, *Biochem. J. 388*, 263-271.

[50] Zeng, X.-C., Corzo, G., and Hahin, R. (2005) Scorpion venom peptides without disulfide bridges, *IUBMB Life 57*, 13-21.

[51] Sunagar, K., Undheim, E. A. B., Chan, A. H. C., Koludarov, I., Munoz-Gomez, S. A., Antunes, A., and Fry, B. G. (2013) Evolution stings: the origin and diversification of scorpion toxin peptide scaffolds, *Toxins 5*, 2456-2487.

[52] Quintero-Hernández, V., Jiménez-Vargas, J. M., Gurrola, G. B., Valdivia, H. H., and Possani, L. D. (2013) Scorpion venom components that affect ion-channels function, *Toxicon 76*, 328-342.

[53] Rogowski, R. S., Krueger, B. K., Collins, J. H., and Blaustein, M. P. (1994) Tityustoxin Kα blocks voltage-gated noninactivating K^+ channels and unblocks inactivating K^+ channels blocked by α-dendrotoxin in synaptosomes, *Proc. Natl. Acad. Sci. U. S. A. 91*, 1475-1479.

[54] Gao, B., Harvey, P. J., Craik, D. J., Ronjat, M., De Waard, M., and Zhu, S. (2013) Functional evolution of scorpion venom peptides with an inhibitor cystine knot fold, *Biosci. Rep. 33*, 513-527.

[55] Chen, Z. Y., Hu, Y. T., Yang, W. S., He, Y. W., Feng, J., Wang, B., Zhao, R. M., Ding, J. P., Cao, Z. J., Li, W. X., and Wu, Y. L. (2012) Hg1, novel peptide inhibitor specific for $K_V1.3$ channels from first scorpion Kunitz-type potassium channel toxin family, *J. Biol. Chem. 287*, 13813-13821.

[56] Cremonez, C. M., Maiti, M., Peigneur, S., Cassoli, J. S., Dutra, A. A. A., Waelkens, E., Lescrinier, E., Herdewijn, P., De Lima, M. E., Pimenta, A. M. C., Arantes, E. C., and Tytgat, J. (2016) Structural and functional elucidation of peptide Ts11 shows evidence of a novel subfamily of scorpion venom toxins, *Toxins 8*, 288.

[57] Olamendi-Portugal, T., Inés García, B., López-González, I., Van Der Walt, J., Dyason, K., Ulens, C., Tytgat, J., Felix, R., Darszon, A., and Possani, L. D. (2002) Two new scorpion toxins that target voltage-gated Ca^{2+} and Na^{+} channels, *Biochem. Biophys. Res. Commun. 299*, 562-568.

[58] Lee, C. W., Bae, C., Lee, J., Ryu, J. H., Kim, H. H., Kohno, T., Swartz, K. J., and Il Kim, J. (2012) Solution structure of kurtoxin: a gating modifier selective for Ca_V3 voltage-gated Ca^{2+} channels, *Biochemistry 51*, 1862-1873.

[59] Fuller, M. D., Thompson, C. H., Zhang, Z.-R., Freeman, C. S., Schay, E., Szakács, G., Bakos, É., Sarkadi, B., McMaster, D., French, R. J., Pohl, J., Kubanek, J., and McCarty, N. A. (2007) State-dependent inhibition of cystic fibrosis transmembrane conductance regulator chloride channels by a novel peptide toxin, *J. Biol. Chem. 282*, 37545-37555.

[60] Thompson, C. H., Olivetti, P. R., Fuller, M. D., Freeman, C. S., McMaster, D., French, R. J., Pohl, J., Kubanek, J., and McCarty, N. A. (2009) Isolation and characterization of a high affinity peptide inhibitor of ClC-2 chloride channels, *J. Biol. Chem. 284*, 26051-26062.

[61] Ali, S. A., Alam, M., Abbasi, A., Undheim, E. A. B., Fry, B. G., Kalbacher, H., and Voelter, W. (2016) Structure-activity relationship of chlorotoxin-like peptides, *Toxins 8*, 36.

[62] Nirthanan, S., Joseph, J. S., Gopalakrishnakone, P., Khoo, H.-E., Cheah, L.-S., and Gwee, M. C. E. (2002) Biochemical and pharmacological characterization of the venom of the black scorpion *Heterometrus spinifer*, *Biochem. Pharmacol. 63*, 49-55.

[63] Lebrun, B., Romi-Lebrun, R., Martin-Eauclaire, M. F., Yasuda, A., Ishiguro, M., Oyama, Y., Pongs, O., and Nakajima, T. (1997) A four-disulphide-bridged toxin, with high affinity towards voltage-gated K^{+} channels, isolated from *Heterometrus spinnifer* (Scorpionidae) venom, *Biochem. J. 328*, 321-327.

[64] Beeton, C., Wulff, H., Standifer, N. E., Azam, P., Mullen, K. M., Pennington, M. W., Kolski-Andreaco, A., Wei, E., Grino, A., Counts, D. R., Wang, P. H., LeeHealey, C. J., Andrews, B. S., Sankaranarayanan, A., Homerick, D., Roeck, W. W., Tehranzadeh, J., Stanhope, K. L., Zimin, P., Havel, P. J., Griffey, S., Knaus, H.-G., Nepom, G. T., Gutman, G. A., Calabresi, P. A., and Chandy, K. G. (2006) Kv1.3 channels are a therapeutic target for T cell-mediated autoimmune diseases, *Proc. Natl. Acad. Sci. U. S. A. 103*, 17414-17419.

[65] Puillandre, N., Koua, D., Favreau, P., Olivera, B. M., and Stocklin, R. (2012) Molecular phylogeny, classification and evolution of conopeptides, *J. Mol. Evol. 74*, 297-309.

[66] Cruz, L. J., Corpuz, G., and Olivera, B. M. (1976) A preliminary study of *Conus* venom protein, *Veliger 18*, 302-308.

[67] Cruz, L. J., Gray, W. R., and Olivera, B. M. (1978) Purification and properties of a myotoxin from *Conus geographus* venom, *Arch. Biochem. Biophys. 190*, 539-548.

[68] Prashanth, J. R., Lewis, R. J., and Dutertre, S. (2012) Towards an integrated venomics approach for accelerated conopeptide discovery, *Toxicon 60*, 470-477.
[69] Buczek, O., Bulaj, G., and Olivera, B. M. (2005) Conotoxins and the posttranslational modification of secreted gene products, *Cell. Mol. Life Sci. 62*, 3067-3079.
[70] Colledge, C. J., Hunsperger, J. P., Imperial, J. S., and Hillyard, D. R. (1992) Precursor structure of ω-conotoxin GVIA determined from a cDNA clone, *Toxicon 30*, 1111-1116.
[71] Buczek, O., Olivera, B. M., and Bulaj, G. (2004) Propeptide does not act as an intramolecular chaperone but facilitates protein disulfide isomerase-assisted folding of a conotoxin precursor, *Biochemistry 43*, 1093-1101.
[72] Terlau, H., and Olivera, B. M. (2004) *Conus* venoms: a rich source of novel ion channel-targeted peptides, *Physiol. Rev. 84*, 41-68.
[73] Dutertre, S., Jin, A.-H., Kaas, Q., Jones, A., Alewood, P. F., and Lewis, R. J. (2013) Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom, *Mol. Cell. Proteomics 12*, 312-329.
[74] Kauferstein, S., Melaun, C., and Mebs, D. (2005) Direct cDNA cloning of novel conopeptide precursors of the O-superfamily, *Peptides 26*, 361-367.
[75] Halai, R., and Craik, D. J. (2009) Conotoxins: natural product drug leads, *Nat. Prod. Rep. 26*, 526-536.
[76] Bulaj, G., and Olivera, B. M. (2008) Folding of conotoxins: formation of the native disulfide bridges during chemical synthesis and biosynthesis of *Conus* peptides, *Antioxid. Redox Signal. 10*, 141-155.
[77] Khoo, K. K., and Norton, R. S. (2011) Role of disulfide bonds in peptide and protein conformation, In *Amino Acids, Peptides and Proteins in Organic Chemistry*, pp 395-417, Wiley-VCH Verlag GmbH & Co. KGaA.
[78] Kaas, Q., Westermann, J. C., and Craik, D. J. (2010) Conopeptide characterization and classifications: an analysis using ConoServer, *Toxicon 55*, 1491-1509.
[79] Kancherla, A. K., Meesala, S., Jorwal, P., Palanisamy, R., Sikdar, S. K., and Sarma, S. P. (2015) A disulfide stabilized β-sandwich defines the structure of a new cysteine framework M-superfamily conotoxin, *ACS Chem. Biol. 10*, 1847-1860.
[80] Zhang, M.-M., Fiedler, B., Green, B. R., Catlin, P., Watkins, M., Garrett, J. E., Smith, B. J., Yoshikami, D., Olivera, B. M., and Bulaj, G. (2006) Structural and functional diversities among μ-conotoxins targeting TTX-resistant sodium channels, *Biochemistry 45*, 3723-3732.
[81] Cruz, L. J., Gray, W. R., Yoshikami, D., and Olivera, B. M. (1985) *Conus* venoms – a rich source of neuroactive peptides, *J. Toxicol. Toxin Rev. 4*, 107-132.
[82] Gray, W. R., Olivera, B. M., and Cruz, L. J. (1988) Peptide toxins from venomous *Conus* snails, *Annu. Rev. Biochem. 57*, 665-700.
[83] Nishi, M., and Kohn, A. J. (1999) Radular teeth of Indo-Pacific molluscivorous species of *Conus*: a comparative analysis, *J. Molluscan Stud. 65*, 483-497.
[84] Robinson, S. D., Safavi-Hemami, H., McIntosh, L. D., Purcell, A. W., Norton, R. S., and Papenfuss, A. T. (2014) Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*, *PLoS One 9*, e87648.
[85] Olivera, B. M., McIntosh, J. M., Cruz, L. J., Luque, F. A., and Gray, W. R. (1984) Purification and sequence of a presynaptic peptide toxin from *Conus geographus* venom, *Biochemistry 23*, 5087-5090.
[86] Norton, R. S., and Pallaghy, P. K. (1998) The cystine knot structure of ion channel toxins and related polypeptides, *Toxicon 36*, 1573-1583.

[87] Pallaghy, P. K., Nielsen, K. J., Craik, D. J., and 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.

[88] Gracy, J., Le-Nguyen, D., Gelly, J.-C., Kaas, Q., Heitz, A., and Chiche, L. (2008) KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007, *Nucleic Acids Res. 36*, D314-D319.

[89] Daly, N. L., and Craik, D. J. (2011) Bioactive cystine knot proteins, *Curr. Opin. Chem. Biol. 15*, 362-368.

[90] Craik, D. J., Daly, N. L., and Waine, C. (2001) The cystine knot motif in toxins and implications for drug design, *Toxicon 39*, 43-60.

[91] Jimenez, E. C., Olivera, B. M., Gray, W. R., and Cruz, L. J. (1996) Contryphan is a D-tryptophan-containing *Conus* peptide, *J. Biol. Chem. 271*, 28002-28005.

[92] Robinson, S. D., and Norton, R. S. (2014) Conotoxin gene superfamilies, *Mar. Drugs 12*, 6058-6101.

[93] Massilia, G. R., Eliseo, T., Grolleau, F., Lapied, B., Barbier, J., Bournaud, R., Molgo, J., Cicero, D. O., Paci, M., Schinina, M. E., Ascenzi, P., and Polticelli, F. (2003) Contryphan-Vn: a modulator of Ca^{2+}-dependent K^{+} channels, *Biochem. Biophys. Res. Commun. 303*, 238-246.

[94] Hansson, K., Ma, X., Eliasson, L., Czerwiec, E., Furie, B., Furie, B. C., Rorsman, P., and Stenflo, J. (2004) The first γ-carboxyglutamic acid-containing contryphan: a selective L-type calcium ion channel blocker isolated from the venom of *Conus marmoreus*, *J. Biol. Chem. 279*, 32453-32463.

[95] Sabareesh, V., Gowd, K. H., Ramasamy, P., Sudarslal, S., Krishnan, K. S., Sikdar, S. K., and Balaram, P. (2006) Characterization of contryphans from *Conus loroisii* and *Conus amadis* that target calcium channels, *Peptides 27*, 2647-2654.

Chapter 1
Peptide grafting design using molecular dynamics

1 Introduction

Producing peptides for experimental analysis is not always straightforward. In some cases, the procedure is further complicated by the nature and properties of the peptide to be produced, as it may be unstable, toxic, or otherwise challenging to produce in its normal state. Such challenges may make production of the native peptide unfeasible, and if it cannot be produced it cannot be tested. However, if the active section of the peptide is known, it may still be possible to study its function by engaging in peptide engineering, redesigning the peptide to be more amenable to production and testing.

This chapter documents such an exercise in peptide engineering – an attempt to validate the scorpion toxin HsTX1 as a scaffold for molecular grafting.

1.1 Peptide engineering

In this context, peptide engineering refers to the rational re-design of peptides of interest, modifying the sequence or structure of an existing molecule to reduce undesired traits and promote beneficial ones. This thesis chapter is concerned with redesigning peptides from a functional standpoint, although the term 'peptide engineering' can also cover engineering peptides for structural purposes – such engineered peptides often self-assemble into nanoscale structures for applications in nanotechnology.[1, 2] These applications are not considered further in the current work.

There are many techniques applicable to the peptide engineering field from a functional standpoint. One popular technique is combinatorial peptide engineering (also called 'directed evolution'),[3] in which extra sequences are introduced to a protein scaffold and randomly mutated before being screened for activity. The best performers are mutated and screened again through successive iterations, until the final molecule is produced. Other techniques include forced cyclisation of molecules,[4] ligation of peptide fragments to create larger molecules,[5] selective mutation of sequences with natural and unnatural amino acids[6] and molecular grafting.

1.1.1 Molecular grafting

'Molecular grafting' is a technique in which sections of two source peptides are melded together to form a new hybrid molecule that shares in the properties of both component pieces. Each component has a specific function in the hybrid. First is the 'scaffold', which makes up the bulk of the hybrid and is chosen for its structural properties, such as thermal

stability, protease resistance and folding efficiency. By using a stable molecule as the basis for the hybrid, it is hoped the hybrid will be likewise stable. The other source molecule is chosen for its function, with the intent that the hybrid molecule will exhibit the same functionality. This is done by selecting the stretch of sequence that is responsible for the desired activity, whether it be the active site, a recognition motif or a set of catalytic residues. This active sequence, commonly referred to as the 'motif', is then grafted onto the scaffold molecule at an appropriate location. A hybrid is thus formed which displays the grafted active sequence in a molecule with much higher stability, circumventing the issues inherent in studying the source molecule on its own.

Several labs have been pursuing peptide engineering by molecular grafting, producing molecules with applications in areas such as tumour imaging,[7] tumour suppression,[8] treatment of autoimmune disease,[9] foot-and-mouth disease[10] and chronic pain.[11]

1.1.2 Molecular grafting: scaffolds

Selecting a scaffold for a molecular grafting project is a key element in the ultimate success of the endeavour. There are two broad ways in which scaffolds are selected. The first is by homology matching – that is, by searching a structural database for a molecule that has a structural element similar to the sequence to be grafted, and using that molecule as the scaffold. This approach is often used when engineering enzymatic activity or mimicking protein-protein interactions, as the structural similarity being exploited allows the inclusion of multiple point mutations at disparate locations, rather than a single stretch of sequence.[12, 13] However, a new search must be undertaken for each new application, as the geometry of each grafted sequence is different from the last. A generalised solution is not possible.

The second approach is to select a molecule that has the potential to present the grafted sequence in an active conformation. There is a collection of molecules that have already been shown to possess this attribute, which forms a toolkit for peptide engineers from which the scaffold most suited for their purposes can be selected without an exhaustive search. The graft to be introduced is often limited to a single insert, as opposed to the more far-reaching modifications sometimes attempted using the homology matching approach. Some versatile scaffold molecules have been used to display multiple distinct peptide grafts in the reported literature (**Table 1**).

There are two qualities of a scaffold molecule that are of primary importance: stability, and the capacity to tolerate a grafted sequence. The first scaffolds tested in early molecular grafting applications were sections of the immunoglobulin molecular fold, as the structure was stable and the variable-region loops were, by definition, tolerant of a wide diversity of sequences. More recent efforts have focused on validating smaller, more synthetically tractable molecules as scaffolds, including many multiply-disulfide-bonded peptides that fold into fairly rigid three-dimensional structures. These molecules are sometimes termed 'miniproteins', as their short sequences nevertheless form structures that include a solvent-excluded core reminiscent of that of larger proteins. In the case of miniproteins this core is made up primarily of the cystines of the disulfide bonds, rather than buried hydrophobic sidechains.

An overview of the molecules that have been used as scaffolds for molecular grafting applications is presented in **Table 1**.

Table 1. Summary of molecules used as scaffolds in molecular grafting peptide engineering studies.

Molecule name	Abbreviation	Structural motif	No. of disulfide bonds	Insert lengths (# res.)	Ref
Immunoglobulin and portions thereof (V_L, REI V_L)	Ig	Ig	≥1	20, 16, 15, 13, 12, 9	*14-16*
Lysozyme		Globular	4	4, 5, 6, 7, 8, 10, 12	*17, 18*
Interleukin-1β	IL-1β	Up-and-down β barrel	0	3, 10	*16, 19*
ω-agatoxins	AgTx	ICK	4	9	*20*
Agouti-related protein	AgRP	ICK	5	7, 11, 14, 16, 18, 20	*21, 22*
Ecballium elaterium trypsin inhibitor	EETI-II	ICK	3	7, 12, 14, 18	*21, 22*
Momordica cochinchinensis trypsin inhibitor	MCoTI-I/-II	CCK	3	3, 6, 7, 9, 14, 15, 16,	*8, 10, 23-25*
Kalata B1		CCK	3	4, 5, 6, 7, 8, 9, 10, 12, 13	*9, 11, 26, 27*
Charybdotoxin	ChTx	CSαβ	3	8 (NC), 9 (NC)	*28-30*
Scyllatoxin	ScyTx	CSαβ	3	11 (NC), 12 (NC)	*31*
Buthus martensi high-conductance potassium channel toxin 1	BmBKTx1	CSαβ	3	4 (NC)	*32*
θ-defensin RTD-1	RTD-1	CCL	3	3	*33*
Obtustatin		Disintegrin	4	11	*21*
Microcin J25	MccJ25	Lasso	0	3	*34*
Sunflower trypsin inhibitor	SFTI-I	β-hairpin	1	6, 7, 9	*24*

ICK = inhibitory cystine knot, CCK = cyclic cystine knot, CSαβ = cysteine-stabilised α/β, CCL = cyclic cystine ladder, NC = non-continuous insert

As can be seen from **Table 1**, there are three general disulfide-bonded motifs that have been extensively investigated as scaffolds, with multiple molecules bearing the motif used in grafting experiments.

One of these is the inhibitory cystine knot (ICK) motif, which was detailed in the Introduction to this thesis. This motif is attractive as a scaffold, as most of the stabilising interactions are found in the cystine knot – as long as the disulfide bonds remain intact, the structure remains stable. The non-cysteine residues have little impact on stability, meaning that they can be mutated and foreign sequences can be grafted in without perturbing the structure. EETI-II in particular has been used in several molecular grafting applications (references in **Table 1**).

The cyclic cystine knot (CCK) motif is very similar to ICK, save that the backbone of the peptide is cyclic. This motif is the uniting feature of the cyclotides, a class of plant defence peptides with diverse activities.[35] As with the ICK motif, the structural stability resides in the disulfide bonds, leaving the rest of the peptide open for grafting. Kalata B1 is an example of this class that has been intensively studied. Gunasekera et al.[26] evaluated each of the six intercystine loops of kalata B1 as insertion points for grafted sequences, excluding loops 1 and 4 (as they form the ring of the cystine knot and opening the backbone at these positions causes a failure to fold[36]). The antiangiogenic RRKRRR 'poly-R' graft tested was most active when displayed in loop 3.[26] In other reports loop 6 has been favoured as an insertion site, as it is the largest loop in the structure and is not in close proximity to the bulk of the molecule, reducing the risk of steric clash between the grafted hybrid and the target molecule.[11, 27] Grafting in this site has resulted in a hybrid capable of selectively blocking the bradykinin B1 receptor (with applications in easing chronic inflammatory pain)[11] as well as a molecule selective for the melanocortin 4 receptor (of potential use for the treatment of obesity).[27]

The final well-studied scaffold is the CSαβ fold, commonly found in scorpion toxins – the same scaffold borne by the proposed scaffold of HsTX1. The key difference between the previous reports and the current work is that each of the previous studies that made use of the CSαβ scaffold chose it by homology matching, rather than utilising it as a general scaffold. Drakopoulou et al. replaced two of the β-strands of charybdotoxin with β-hairpins from both a curaremimetic snake neurotoxin[28] and the CDR2-like region from the immunological glycoprotein CD4.[29] The short N-terminal β-strand of the scaffold was truncated in each case to improve the accessibility of the grafted hairpin and avoid clashes with the target molecule. Vita et al. followed up by grafting the CD4 hairpin into the β-sheet

of scyllatoxin, which lacks the N-terminal strand and was therefore posited to be a superior scaffold.[31] Matching the α-helix also led Li et al. to engineer a tumour-suppressing p53 sequence into the CSαβ scorpion toxin BmBKTx1.[32] None of these examples involved grafting a motif into a generic insertion site on the CSαβ structure, as is proposed in this work.

1.1.3 Potential graft: RGD integrin binding motif

To evaluate the potential for HsTX1 to act as a scaffold, an appropriate graft sequence needed to be selected for testing. A classic sequence used to test the feasibility of scaffolds is the tripeptide RGD, the integrin binding motif. Integrins are a class of membrane-bound cell-surface receptor that integrate the extracellular matrix with the cytoskeleton of cells.[37] Each receptor exists as a noncovalent heterodimer of α and β subunits. Both types of subunit possess a large extracellular domain, a single transmembrane helix and a short cytoplasmic domain. Despite these similarities in architecture, the subunits are not homologous, and generally α subunits are larger than their β counterparts. To date 18 α subunits and 8 β subunits have been characterised in vertebrates, with 24 distinct heterodimers documented.[38] Integrin receptors bind an array of ligands which are abundant in the extracellular matrix, including collagen, fibronectin, fibrinogen, thrombospondin, vitronectin and von Willebrand factor.[39] Upon binding, integrin receptors form clusters and in many cases recruit kinases that go on to phosphorylate other signalling molecules.[37] The minimal recognition motif for an integrin ligand is the tripeptide Arg-Gly-Asp, known as the RGD motif.[40-42] Each heterodimeric integrin receptor interacts with a different subset of extracellular RGD-containing ligands, and each ligand can interact with multiple integrin receptors. The differing recognition profiles of integrin receptors are thought to be due to differing physical conformations of the tripeptide motif.[43]

Since it was first identified in 1984,[40] the RGD motif has seen frequent use as a promoter of cell adhesion. Biopolymers modified with RGD peptides have elicited heightened cell attachment and cell spreading in tissue engineering applications,[44] and many studies have made use of the motif as a graft for peptide engineering.[15-18, 20, 21, 33, 34] These successes make the RGD motif an ideal candidate for validating the utility of a new peptide scaffold.

1.1.4 Potential graft: DINNN protein-protein interaction recognition motif

In addition to the RGD motif, a more focused practical application of HsTX1 as a scaffold was planned by using a second graft sequence. Inducible nitric oxide synthase

(iNOS) was a molecule under study in our laboratory that is a key part of the innate immune response to invading pathogens, as it is upregulated in instances of pathogenic attack. iNOS catalyses the conversion of L-arginine into L-citrulline and nitric oxide, a reactive free radical which is used as a chemical attack to kill the invading pathogen.[45] The production of nitric oxide is closely regulated, as excess nitric oxide is also harmful to native cells, so iNOS is recognised by the cytoplasmic protein SPRY-containing SOCS box protein 2 (SPSB2). Binding between iNOS and SPSB2 leads to ubiquitination and degradation of iNOS, which ends the production of nitric oxide.[46] The recognition motif for this protein-protein interaction is the pentapeptide Asp-Ile-Asn-Asn-Asn (DINNN).[46] An appropriately designed DINNN analogue could occupy the binding site on SPSB2, competing for binding with iNOS and thus delaying iNOS degradation and prolonging the production of nitric oxide. This would enhance the pathogen killing effect, and may be an avenue for the creation of novel anti-infective agents.[47] A grafted HsTX1 molecule containing the DINNN motif was hypothesised to be capable of producing this effect, which would be a good test of HsTX1 as a scaffold.

1.1.5 Hybrid construct design

The structure of HsTX1 includes a loop between the two strands of the β-sheet (residues 25–28), which was chosen as the site of insertion for the graft sequence (illustrated in **Figure 1**). One important consideration was the length of the linker between graft and scaffold – too short and the bulk of the scaffold may interfere with the binding event between graft and target, while a too long linker would confer too much flexibility and negatively impact the stability of the construct. To probe the stability of constructs with varying linker lengths, molecular dynamics simulations were performed on a series of proposed designs.

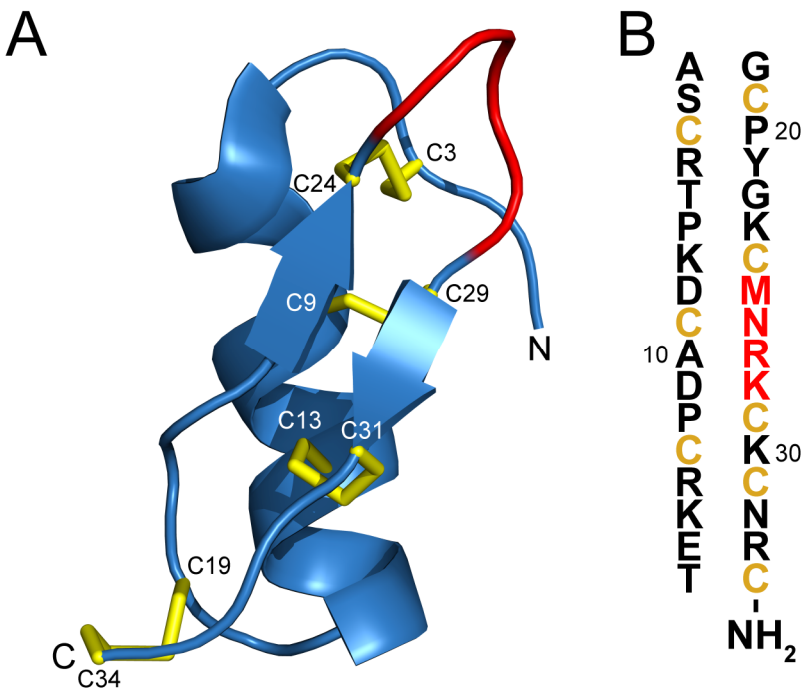


Figure 1. Three-dimensional structure of scorpion toxin HsTX1. Disulfide bonds are *yellow*, graft insertion loop is highlighted in *red*.

1.2 Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations are a computational method for investigating molecular systems, which can provide information on the specific interactions made by individual atoms in the system.

1.2.1 Technical aspects of MD simulations

In a MD simulation, the molecular system to be analysed is represented as a collection of particles. The positions and velocities of the particles evolve in time, with the total length of the simulation made up of many small time steps. At each time step, the forces acting on each particle are calculated and both position and velocity are updated accordingly. The process is very computationally expensive and, despite a range of techniques available for optimising the process, long simulations of complex systems can take many weeks of CPU time. Some of these streamlining techniques are discussed in the following sections, as are other important aspects of the MD framework.

1.2.2 Force fields

The forces applied to each atom in a MD simulation are determined by the molecular mechanics force field, which consists of a potential energy function and a series of

parameters used in that function, such as (for example) bond lengths and atomic charges. The potential energy function is made up of several terms, each of which describes the energetic contribution of a certain type of interaction. A common form of the potential energy function is shown in **Figure 2**. Different interactions between bonded molecules are accounted for by a series of separate terms. The energy associated with deviations from the ideal bond length is represented by a harmonic potential. Bond angles are similarly treated with a harmonic potential, while torsional angles are represented by a periodic sinusoidal term. The other terms in the potential energy function are for long-range interactions between non-bonded atoms. Dispersion forces are represented by a Lennard-Jones '12–6' potential, with a repulsive component acting at short distances and an attractive component that dominates at larger distances. Finally, electrostatic interactions are captured by a coulombic potential.[48] Each of these terms is evaluated for each particle at each time step, with the calculated force being used to update the position and velocity of the particle. Thus, the potential energy function forms the core of the molecular dynamics simulation.

$$E(R) = \frac{1}{2}\sum_{bonds} K_b(b - b_0)^2 + \frac{1}{2}\sum_{\substack{bond\\angles}} K_\theta(\theta - \theta_0)^2 + \frac{1}{2}\sum_{torsional} K_\varphi[1 + \cos(n\varphi - \delta)] + \sum_{non-bonded}\left(\frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr}\right)$$

Figure 2. General form of potential energy function used in molecular dynamics force fields. Taken from [48].

The potential energy function is only part of the force field, however. The function contains several constants that determine the strength of each interaction, and these constants may be different for different atom types. Force fields must therefore be parameterised to ensure these constants and other terms yield a model that is in agreement with experiment. There are several force fields that can be employed in a MD simulation, with a diversity even within the sub-field of biomolecular MD. Many are continually undergoing a process of development and refinement. Each force field may base its parameterisation on a different property of the atoms involved. The GROMOS force field bases its parameterisation on the free enthalpy of solvation,[49, 50] while the CHARMM force field is optimised for condensed phase calculations by reproducing experimental pure liquid, solution and crystal data with a balance between microscopic and macroscopic properties.[51]

1.2.3 Time step length

The length of the time step used in a simulation is informed by two factors. It is more efficient to run simulations with as large a time step as possible, since fewer calculations need to be performed for a simulation of a given duration. The computing power thus saved can then be re-invested to run longer simulations and access motions at larger timescales. However, the time step must sample the system often enough to capture the motions of the system – if the time step is too long, rapid motions can be missed, leading to discontinuities and compromising the simulation. The time step is therefore limited by the fastest dynamic that is observed in the system under analysis. For a peptide in water, the fastest dynamic is technically bond-stretching vibrations, which occur on a timescale of ~10 fs.[52] However, these motions are governed by quantum mechanical effects and are not explicitly considered in classical MD, being treated by a mathematical constraint instead. The fastest dynamic considered by a MD system is therefore bond-angle vibrations (specifically, those that include hydrogen atoms), which have a periodicity of 13 fs.[52] To ensure this motion is sampled sufficiently, MD simulations of peptides are generally limited to a time step of 2 fs.

1.2.4 All-atom, united-atom and coarse-grained force fields

In addition to the length of the time step, another contributor to the computational cost of a simulation is the number of particles to be simulated. The more particles there are in the system, the more calculations are required at each time step – and since long-range interactions in principle involve calculating the force on every particle due to every other particle, the computational expense compounds on itself. To help mitigate this, some force fields combine multiple atoms into larger particles, saving on the number of particles to be simulated. In an all-atom force field, such as AMBER or CHARMM22, each individual atom is represented by its own particle. Another option is a united-atom force field, such as GROMOS or CHARMM19, in which some atoms are considered together in a single particle. For example, the GROMOS united-atom force field merges non-polar hydrogen atoms with their bonded heavy atoms. Extending this concept further are coarse-grained force fields, such as MARTINI, in which several atoms are grouped together into a single particle or ‘bead’. Combining multiple atoms into these coarse-grained beads reduces the number of particles that need to be simulated, and allows the simulation of larger systems. In addition, larger time steps can be used, as the dynamics between beads are slower than those between atoms. It has been estimated that the coarse-grained approach can cut the

computational cost of a simulation by seven orders of magnitude, compared to an atomistic simulation of the same system.[53] This saving comes at the cost of reduced fidelity, so coarse-grained simulations are generally only run for large systems that are beyond the capabilities of modern hardware running more atomistic simulations. The differences between all-atom and united-atom simulations are less severe, although the incorporation of hydrogen atoms into the simulated particle requires a distinct and specialised parameter set to ensure physical behaviour.

1.2.5 Accounting for solvent effects

Hardware constraints on the complexity of systems that can be simulated are a constant reality in the MD field, although the situation is always improving. In the early years of MD, one strategy to mitigate this reality was to perform a biomolecular simulation in a vacuum, without solvent. Indeed the first simulation of a protein, performed in 1977 on the globular bovine pancreatic trypsin inhibitor, took this approach.[54] The justification supplied for neglecting the solvent was that the internal motions being analysed would be little affected by the lack of solvent, even though the surface residues would be subjected to unphysical conditions. With the advent of more powerful computer hardware, simulations incorporating the effects of solvent have become the norm, which allows the analysis of surface effects – particularly important in the simulation of peptides.

One early solution to the solvation problem was implicit solvation, in which the effects of solvent on the molecule of interest were calculated via a series of mathematical approximations.[55] This was mostly useful for bulk solvent and was unable to account for the effects of ordered water molecules at the solvent-solute interface. The implicit solvent model could also break down in areas such as deep binding pockets, where the behaviour of isolated water molecules may vary from that of bulk solvent.

The modern solution to the solvation problem is explicit solvation, which involves including a host of water molecules in the simulation and calculating the interactions of these molecules with the species of interest and with each other. Several water models have been developed for this purpose in an attempt to faithfully replicate experimentally determined properties of liquid water, such as density and intermolecular energy. Models vary in the number of particles used to compose a single water atom. While models with higher particle counts are generally more faithful to the experimental values, they are also more computationally expensive.[56] The choice of which water model to use is often informed by the force field being utilised; when force fields are being parameterised, the

values chosen are usually selected on the assumption that a certain water model will be used. For example, the GROMOS force field is optimised for use with the single point charge (SPC) model, while AMBER and CHARMM make use of TIP3P.[57] Using a different water model may involve re-parameterisation of the force field to ensure results are accurate.

1.2.6 Simulation cells and periodic boundary conditions

If a simulation is run on a system of a finite size, unphysical effects may occur at the boundary, where the particles have no neighbours to interact with. At the same time, it is impractical to simulate a system of infinite size. The solution is to confine the simulation to a finite 'cell' and impose a boundary condition. Most biomolecular simulations make use of so-called periodic boundary conditions, in which the cell is replicated as a periodic image on all boundaries. When forces are calculated, each particle interacts with whichever version of the partner particle is closest – either the one in the cell or one in a periodic image. In effect this creates a simulation of infinite volume and eliminates boundary effects that would otherwise influence the simulation. An important consideration is the size of the cell; although there is no problem if the molecule of interest interacts with the periodic image of a solvent molecule, a situation should be avoided where the molecule is interacting with a periodic image of itself. Therefore the cell must be of a sufficient size to ensure this does not occur. However, a larger cell contains more particles and therefore requires more computational power. To alleviate this, some simulations make use of non-cubic boxes. A rhombic dodecahedron, for example, can maintain the same separation between a molecule and its images while saving 40% of the volume, and therefore the solvent molecule count, over a cubic box.[58] This is due to the more spherical geometry of the rhombic dodecahedron, which cuts corners off a similar-sized cube but still tessellates efficiently.

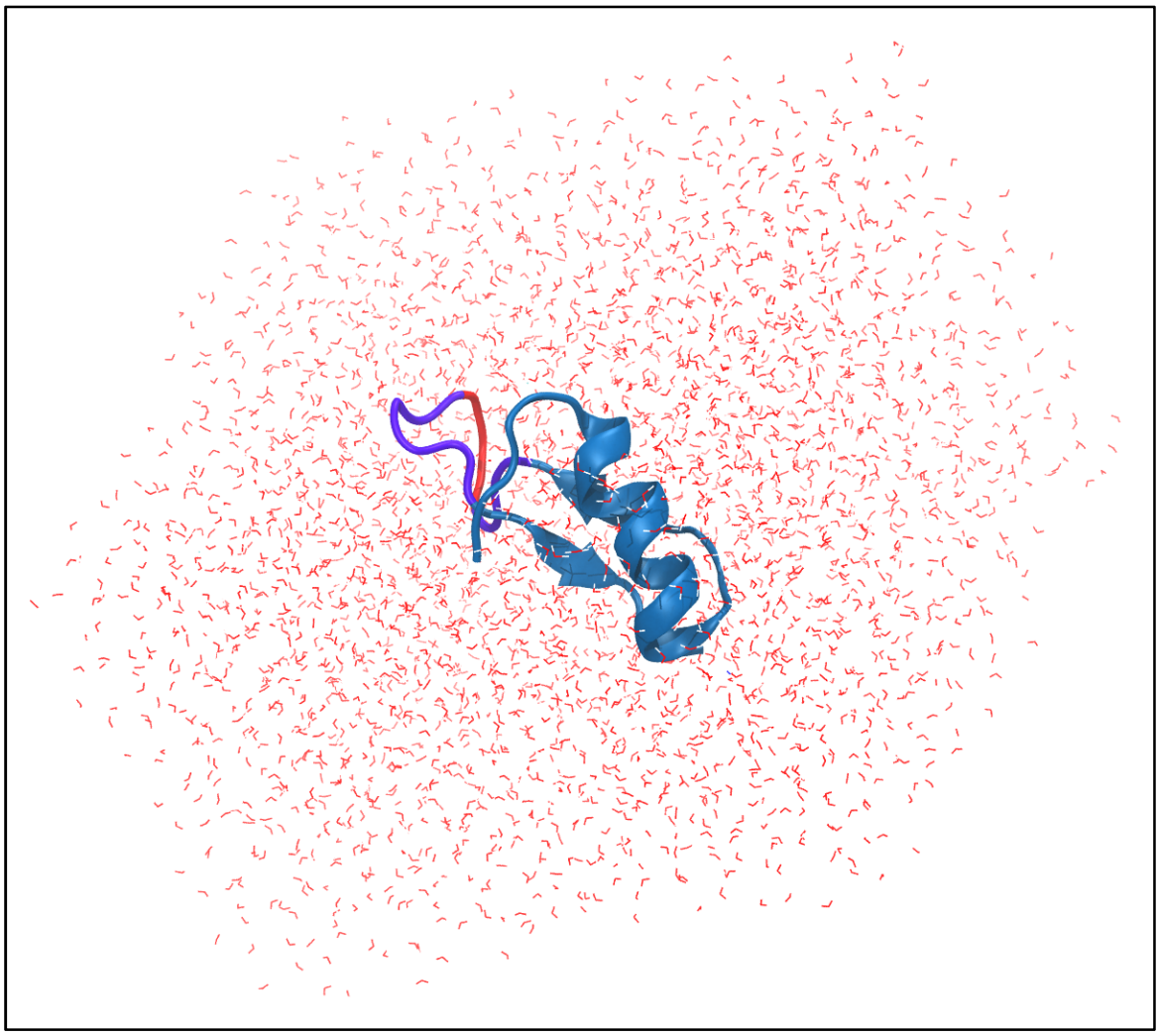


Figure 3. Grafted HsTX1 construct in a simulation cell packed with SPC water models.

The goal of a MD simulation is to investigate the motions of a system at the atomic level, collecting data and allowing inferences to be drawn. The capabilities of the hardware running the simulation determine which systems are tractable and how long the simulation will take to complete. Through careful selection of time step, force field, solvent treatment, cell dimensions and boundary conditions, the computational expense of achieving that goal can be minimised. The computational power thus saved can then be invested in running a longer simulation to collect more data, or in testing more conditions and determining the effect of different variables, or in a simulation of a more complex system to gain new knowledge in a new arena. Eventually the progress of technology makes such simulations viable even without optimising such parameters, but making use of the techniques discussed pays dividends in the meantime.

2 Research rationale

In this study, MD simulations were used to model the interactions of designed constructs of grafted HsTX1, determining the effects of the graft and linker residues on the stability of the molecule. The primary focus was on the length of the peptide linker. Too short and the motif would be unable to interact with the target protein due to steric hindrance by the bulk of HsTX1. Too long and the linker would be too flexible, negatively impacting stability and possibly complicating binding. The site of insertion for the grafts was the loop between the β-strands, residues 25–28 in the native sequence. For each graft (RGD and DINNN), a series of constructs was designed *in silico*, differing by the number of Gly residues in the linker between graft and scaffold. The linkers were symmetric, with an equal number of Gly residues on each side of the graft (0–5, resulting in a total linker length of 0–10 residues).

In addition to these poly-glycine linkers, two constructs were designed based on work published by Kimura et al.,[59] which described RGD-containing peptides developed by directed evolution. The inserts of their constructs 2.5D and 2.5F were transplanted to the HsTX1 scaffold to form constructs JC-D and JC-F. The sequences of the designed constructs (and the final constructs produced) are presented in **Table 2**.

Table 2. Sequences of proposed HsTX1 grafted constructs. Binding motifs are *red*, linker residues are *purple*.

Construct designation	Sequence
Native	ASCRTPKDCADPCRKETGCPYGKC-----MNRK------CKCNRC*
Rx	ASCRTPKDCADPCRKETGCPYGKC-(Gx)RGD(Gx)---CKCNRC*
JC-D	ASCRTPKDCADPCRKETGCPYGKC--PQGRGDWAPTS--CKCNRC*
JC-F	ASCRTPKDCADPCRKETGCPYGKC--PRPRGDNPPLT--CKCNRC*
R-C	ASCRTPKDCADPCRKETGCPYGKC----GRGDWAE----CKCNRC*
Dx	ASCRTPKDCADPCRKETGCPYGKC-(Gx)DINNN(Gx)-CKCNRC*
D10-GS	ASCRTPKDCADPCRKETGCPYGKCGSGSGDINNNGSGSGCKCNRC*

To test the stability of these constructs, MD simulations were carried out with a time step of 2 fs using the GROMOS united-atom force field in the GROMACS software package, with the explicit SPC water model and a rhombic dodecahedral box.

3 Methods

3.1 Molecular dynamics (MD) simulations

Molecular dynamics simulations were carried out using GROMACS version 5.0.4,[60] utilising the GROMOS 54a7 united-atom force field[50] and a 2 fs time step. Temperature coupling made use of the velocity rescale algorithm with a reference temperature of 298 K. Pressure coupling used the Parinello-Rahman algorithm,[61] with reference pressure of 1 bar and compressibility of 4.5×10^{-5} bar^{-1}.

Initial structures were generated using the Maestro software package (v 10.3.014)[62] by *in silico* mutation of the published structure of HsTX1 (PDB code: 1QUZ).[63] These structures were solvated with SPC water and subjected to a steepest-descent minimisation of 2000 steps to remove bad van der Waals contacts between atoms. Temperature equilibration (without pressure coupling) was run for 10,000 steps. Isotropic pressure coupling was then applied for 500,000 steps. Following equilibration, the simulation production runs were executed for 100 ns each. The resultant trajectories were visualised with the Visual Molecular Dynamics (VMD) software package (v 1.9.2)[64] and analysed with VMD and GROMACS built-in tools.

3.2 Trajectory analysis

Simulation trajectories for each construct were visualised with VMD. Images of the constructs were also prepared with this software. Root-mean-square deviation (RMSD) violin plots were generated with a Python script using the matplotlib library,[65] based on data extracted using the *rms* command of the GROMACS software. Radius of gyration was calculated using the *gyrate* command of the GROMACS software and plotted using the gnuplot software package (v 5.0). Regional average RMSD values across the simulation were calculated with the VMD RMSD Trajectory tool.

3.3 Peptide synthesis

Samples of selected constructs were ordered from Peptides International Inc. (KY, USA), where they were produced by solid-phase peptide synthesis (SPPS). Analyses of the samples produced are shown in Section 4.7. A discussion of the SPPS technique is presented in the following chapter, *Accessing peptide toxin samples*.

4 Results

4.1 Initial structure creation

The initial structures for each proposed construct were prepared by *in silico* mutation of the PDB entry for HsTX1 (PDB ID: 1QUZ). Residues 25–28 in the native structure were replaced by the combined graft and linker sequence (**Table 2**). The structures were also energy minimised within the Maestro program before coordinates were output in the PDB file format, ready for conversion to the .gro format.

4.2 Trajectory analysis

To ascertain the stability of the proposed constructs, global RMSD and radius of gyration (R_g) were calculated across the 100 ns simulation time of each trajectory. RMSD values indicated the extent to which the structure was changing, while radius of gyration allowed the compactness of the structure to be monitored. If the structure of the grafted construct were to unravel during the simulation, this would be indicated by an increase in R_g. The results of these analyses are presented in **Figures 4–7**.

4.3 RMSD plots

The RMSD of each construct from the initial pose was calculated for the length of the trajectory. These values were visualised as violin plots, giving an indication of the overall flexibility of each construct and allowing easy comparison to the other constructs in the series. The higher the RMSD of a construct, the more the structure changed over the course of the simulation. Lower RMSD values indicate lower variability and therefore more stable conformations. The plots are presented as **Figure 4** and **Figure 5**.

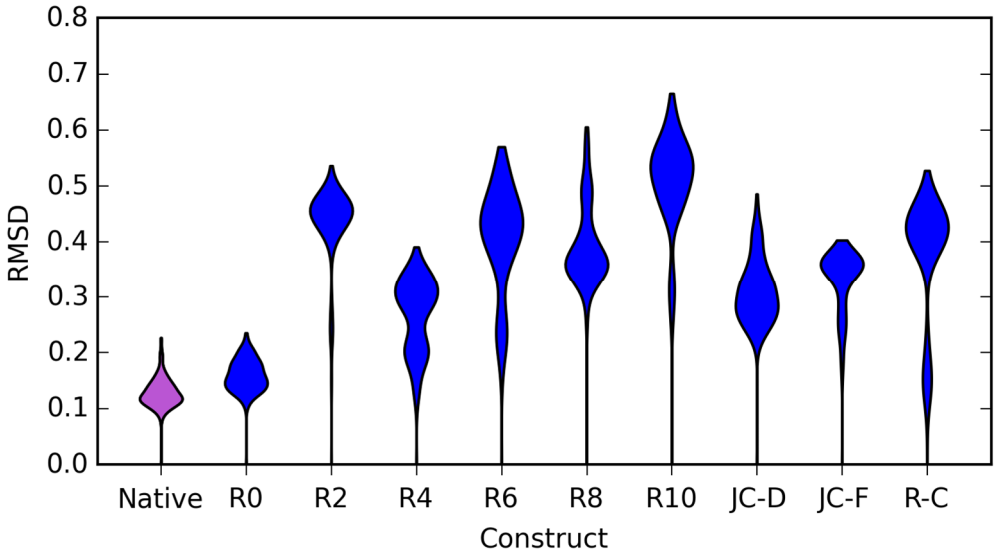


Figure 4. Violin plots of RMSD values for R-series *in silico* designed HsTX1 grafted constructs. Higher values indicate greater deviation from the initial pose.

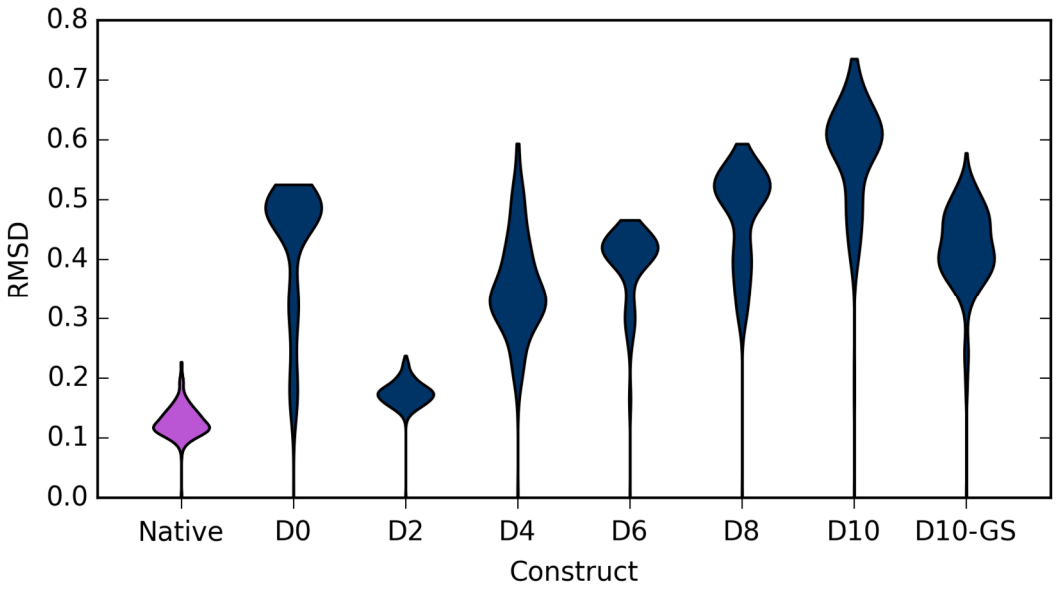


Figure 5. Violin plots of RMSD values for D-series *in silico* designed HsTX1 grafted constructs. Higher values indicate greater deviation from the initial pose.

In general, the longer the linker the higher the RMSD observed. Constructs R2 and D0 are exceptions to this general trend, with greatly elevated RMSD values compared to their short linker length. Construct R-C also defied the trend, as the graft is shorter than that of R4, JC-D and JC-F and yet the RMSD is higher. Also notable is construct D10-GS, which differs by identity from construct D10 by only four residues and by length not at all; regardless, the RMSD value is appreciably reduced from that of construct D10.

4.4 Radius of gyration (R_g) plots

Mathematically, the radius of gyration (R_g) of a molecule is the perpendicular distance from the axis of rotation to the location of a point mass that gives the same inertia as the whole molecule. The less compact the molecule – for example, the less folded the protein – the higher the value of R_g.

R_g plots of the simulation trajectories of HsTX1 constructs are presented in **Figure 6** and **Figure 7** and generally maintain a consistent value, indicating that the structure remained stable and did not unfold. Construct R8 is an exception, with a somewhat heightened R_g value at the end of the simulation. Constructs R6 and D0 also displayed short periods of heightened R_g, but then recovered and maintained a low value for the rest of the simulation. Without running the simulation of construct R8 beyond 100 ns it is not possible to determine if this construct would likewise recover or unfold completely. Construct D4 also deviates from the steady value towards the end of the simulation, but in this case there is a reduction in R_g, reflecting the fact that the grafted loop folds over the side of the molecule in a more compact arrangement.

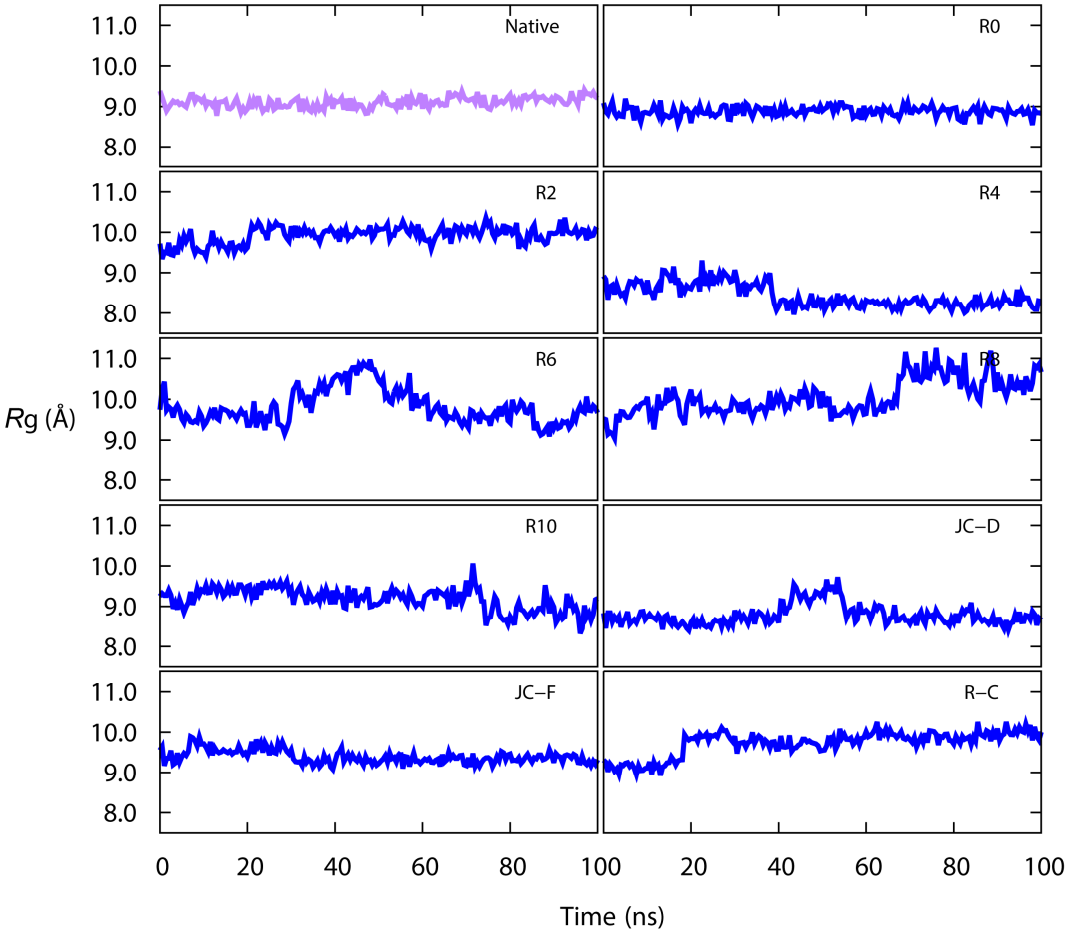


Figure 6. Radius of gyration (R_g) as a function of time across length of simulation for R-series *in silico* designed HsTX1 grafted constructs.

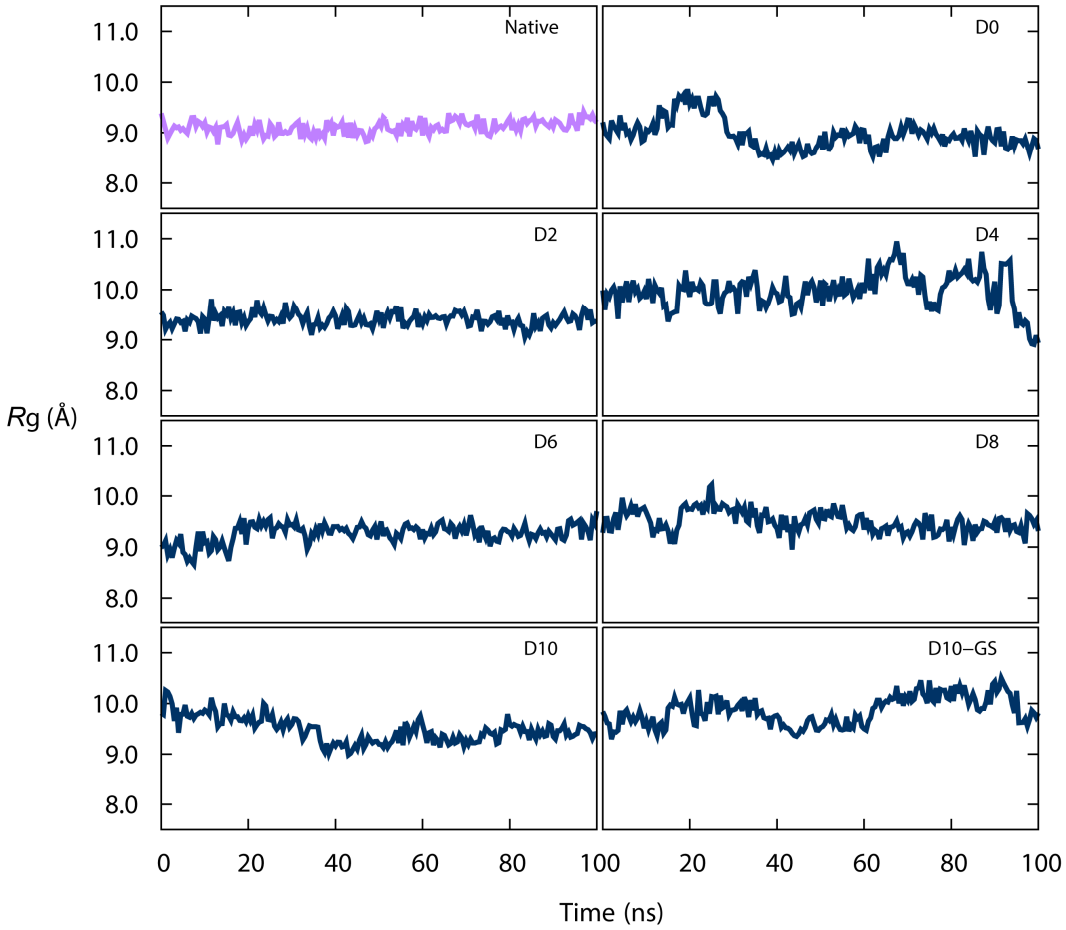


Figure 7. Radius of gyration (R_g) as a function of time across length of simulation for D-series *in silico* designed HsTX1 grafted constructs.

4.5 Regional RMSD calculations

The overall RMSD values for each construct (as shown in **Figure 4** and **Figure 5**) present an incomplete picture, since each construct was composed of two sections – the residues composing the flexible graft and linker as well as the residues composing the core of the scaffold. In terms of ascertaining stability of the molecule, the RMSD values of the core scaffold were far more important, as the linker was expected to be flexible and able to change its orientation throughout the simulation. Any change in the orientation of the linker would alter the overall RMSD, but perhaps not appreciably impact the actual stability of the construct. RMSD values were therefore calculated for each section separately, to investigate the relative contributions of each part of the construct to the overall RMSD value. The results of RMSD calculations for all constructs are presented in **Table 3**.

Construct	RMSD (Å)		
	Global	Core	Grafted loop
R0	1.40	1.36	1.64
R2	2.39	2.27	2.88
R4	1.99	1.57	3.29
R6	3.09	2.62	4.47
R8	2.22	1.90	1.96
R10	3.00	1.36	3.28
JC-D	2.85	2.07	2.55
JC-F	2.48	1.63	4.25
R-C	2.93	2.74	2.00
D0	5.17	5.10	5.42
D2	1.90	1.38	3.26
D4	2.74	2.41	3.72
D6	2.43	2.24	2.77
D8	2.54	2.56	1.66
D10	3.87	1.98	3.45
D10-GS	2.93	2.30	2.46

Table 3. Regional RMSD values for global fold, core residues and grafted loop over simulated trajectory for all *in silico* designed HsTX1 grafted constructs. Values were calculated for heavy atoms with respect to the average structure.

4.6 Selection of constructs

In addition to the stability of the construct, it was important that the linker be long enough to prevent the steric bulk of the scaffold from interfering with the binding between graft and target. The constructs selected for evaluation therefore had a minimum linker length of four residues. The other criterion considered when selecting constructs to be synthesised was the core RMSD value from **Table 3**, with lower values being more favourable. The constructs thus selected were D10 for the D-series and JC-F for the R-series. In consultation with our collaborator Mike Pennington of Peptides International, the linker of D10 was modified from –GGGGG– to –GSGSG–. A simulation of this revised construct (D10-GS) was also run and analysed in the same manner as the previous designs.

4.7 Production of selected constructs

Production of the selected constructs was carried out by solid-phase peptide synthesis by Peptides International, who have considerable experience in producing HsTX1 and associated analogues. Synthesis was successful, and oxidative folding attempted via air oxidation at room temperature for 36 hr in the presence of 0.75 mM oxidised glutathione and 1.5 mM reduced glutathione. Unfortunately, the folding resulted in a multitude of peaks on the HPLC trace rather than the single peak of a well-folded product (**Figure 8**). Previous investigations on native HsTX1 have showed that the unmodified peptide folds to a single peak in almost 100% yield (Figure 4A in Rashid et al.[66]).

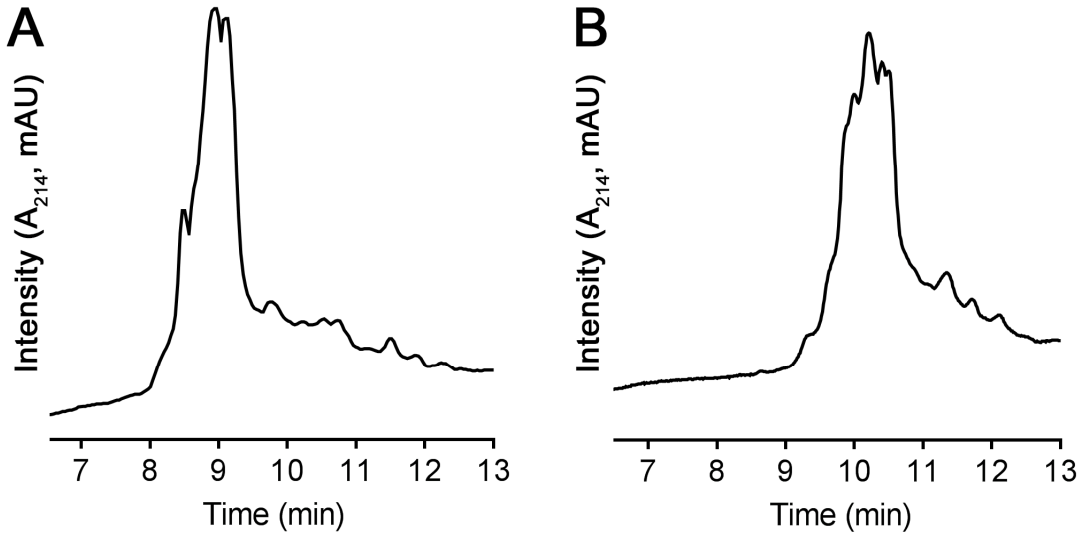


Figure 8. RP-HPLC traces of refolding mixtures of final HsTX1 constructs following overnight incubation. Buffer A: 0.05% TFA in MilliQ H_2O; Buffer B: 0.05% TFA in acetonitrile, gradient: 0–67.5% B over 20 min. (A) Construct D10-GS. (B) Construct JC-F.

4.8 Final construct design

Given the failure of construct JC-F to fold, a shorter RGD graft from a publication by Richards et al.[67] was proposed, in the hope that the shorter graft would have a smaller impact on folding. A simulation of this proposed construct R-C was also run and analysed (**Figure 4**, **Figure 6**, **Table 3**). As with the other constructs produced, the refolding reaction resulted in a multitude of peaks, frustrating attempts to isolate a pure sample for analysis.

5 Discussion

The design of the grafted construct is a key consideration in the pursuit of a molecular grafting study, and impacts on every downstream aspect of the process. Both graft and scaffold must be carefully chosen, and then integrated effectively.

For this project, the explicit aim was to validate HsTX1 as a general scaffold, which set the choice of which scaffold to use. To effectively ascertain the generalised utility of the scaffold, two grafts were chosen to test. The RGD motif has been widely used in proof-of-concept and validation studies for other scaffolds, which made it a logical choice for this new scaffold. The DINNN motif was chosen as a practical example of the ability of the scaffold to exhibit a desired bioactivity. With both scaffold and graft set, the remaining question was how to combine these elements in a manner that resulted in the best grafted construct.

The approach chosen in this work defined 'best' as 'most stable', and to investigate how to maximise this parameter MD simulations were undertaken. The primary strength of this technique is that it furnishes a wealth of atomic-level detail on the system being studied. A more challenging aspect is deciding which questions to ask of that data set, sifting through and extracting the information most relevant to your research objectives.

5.1 Selection of constructs

Two criteria were used to select constructs for production. The first was that the linker be long enough to separate the grafted motif from the bulk of the scaffold. This emerged as a preference for linkers with at least two flanking residues on each side. The second criterion was the RMSD value of the core region of the scaffold, with lower values taken to indicate greater stability. The radius of gyration data was used as a filter; if any constructs underwent an increase in R_g over the course of the simulation, it would indicate an unfolding event that would disqualify them from consideration.

For the R-series, constructs R0 and R10 had the lowest core RMSD values. However, construct R0 had no separation between the graft and the linker and would therefore be unlikely to allow the RGD motif to interact with the integrin receptor unhindered. By contrast, construct R10 had too much linker; RGD motif binding depends on a certain rigidity,[43] and five intervening Gly residues would confer excessive flexibility. The identity of residues flanking the motif has also been shown to affect RGD-integrin binding.[68] For these reasons, construct JC-F was eventually chosen for synthesis, selected over construct

JC-D by virtue of its lower core RMSD value. The PRPRGDNPPLT graft of construct JC-F has been reported to bind to $\alpha_V\beta_3$, $\alpha_V\beta_5$ and $\alpha_5\beta_1$ integrin receptors,[59] and it was hoped this graft could confer similar binding properties on the HsTX1 scaffold. Once construct JC-F failed to fold, a shorter RGD consensus motif was taken from the work of Richards et al.[67] to make construct R-C, in the hope that a shorter graft would lead to more efficient folding. This did not eventuate, as construct R-C also failed to fold effectively.

The lowest core RMSD in the D-series belonged to construct D2, but again concerns about linker length led to disqualification. The next lowest value belonged to construct D10. There are no published data on the effects of rigidity of the DINNN motif on SPSB2 binding, so the enhanced flexibility of the graft was not considered to be as much of an issue as it was for the R-series. Thus construct D10 was chosen as the DINNN graft for this study – although for final production the linker was modified from poly-Gly to mixed Gly/Ser.

The final constructs produced by solid-phase peptide synthesis were:

Construct JC-F

ASCRTPKDCADPCRKETGCPYGKCPRPRGDNPPLTCKCNRC*

Construct R-C

ASCRTPKDCADPCRKETGCPYGKCGRGDWAECKCNRC*

Construct D10-GS

ASCRTPKDCADPCRKETGCPYGKCGSGSGDINNNGSGSGCKCNRC*

5.2 Validity of assumptions

In this study, the initial models for each simulation run were constructed by *in silico* mutation of the published HsTX1 structure from the PDB, inserting the motif and linker into the loop between the β-strands (residues 25–28 in the native sequence). This was the starting point from which the rest of the simulation flowed – it was assumed that the peptide would adopt this conformation, in which the structure was well folded and the disulfide bond network was correctly arranged. This assumption was not without support, given the highly efficient folding of native HsTX1.[66] However, when the time came to produce the selected constructs in the laboratory, it was found that the insertion of the graft between the β-strands disrupted the efficient folding of HsTX1 and prevented the adoption of the folded conformation.

This was not a possibility which was considered in the simulation work. The simulations performed had been designed to assess the stability of the folded conformation, rather than the examine the specifics of the folding pathway. Indeed, a different tool would need to be used to probe the folding pathway, as formation of the disulfide bond network is a critical step and classical MD simulations are incapable of forming or breaking bonds between atoms without including quantum calculations in the paradigm.[48] The 'refolding problem' remains one of the greatest challenges in peptide biochemistry, as the factors that determine and influence the transition from linear primary sequence to three-dimensional tertiary structure are still being investigated and no general schema is known.

5.3 Directed folding

Finding conditions that promote the proper folding of proteins and peptides, especially those with multiple disulfide bonds, is a challenging proposition. Buffers and cofactors all play a role, but although there are general guidelines and principles there are no firm 'rules'.

Rather than relying on conformational dynamics to ensure that the correct Cys residues bond together, another approach is so-called 'directed folding'. In this paradigm, corresponding pairs of Cys sidechains are protected with orthogonal protecting groups, such that the protecting groups can be removed one at a time. This results in only a single pair of Cys sidechains being exposed at any given time and ensures that each pair will bond with each other, rather than a different, incorrect Cys sidechain. This reduces the combinatorial problem of protein folding into a series of discrete steps; instead of the 105 connection patterns possible with 4 disulfide pairs, a single pair is reacted at each stage, resulting in the single, properly formed product. One such orthogonal scheme for 4 disulfide bonds has been reported using the protecting groups trityl (Trt), acetamidomethyl (Acm), *tert*-butyl (tBu) and 4-methyl-benzyl (MeBzl).[69]

Although such a scheme could allow the production of properly folded HsTX1 grafted hybrids, it nullifies a large part of the appeal of HsTX1 as a scaffold – that being the ease of folding observed in the native peptide. If getting such a grafted molecule to fold becomes a complex process, the simplest solution is to use a different scaffold.

6 Conclusion

In the end, the effort to validate the scorpion toxin HsTX1 as a general scaffold for molecular grafting was unsuccessful. Computer modelling, although useful, was limited in scope and considerations outside the scope of the simulation lead to a failure in the folding of the constructs when they were produced physically. There remain a selection of molecules that have been successfully used as scaffolds for arbitrary short motifs in molecular grafting, but the CSαβ fold is not one of them, having only been successfully grafted when chosen on the basis of homology matching with the source structure.

7 References

[1] Gazit, E. (2007) Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization, *Chem. Soc. Rev. 36*, 1263-1269.

[2] Kobayashi, N., and Arai, R. (2017) Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks, *Curr. Opin. Biotechnol. 46*, 57-65.

[3] Jäckel, C., Kast, P., and Hilvert, D. (2008) Protein design by directed evolution, *Annu. Rev. Biophys. 37*, 153-173.

[4] Clark, R. J., Fischer, H., Dempster, L., Daly, N. L., Rosengren, K. J., Nevin, S. T., Meunier, F. A., Adams, D. J., and Craik, D. J. (2005) Engineering stable peptide toxins by means of backbone cyclization: stabilization of the α-conotoxin MII, *Proc. Natl. Acad. Sci. U. S. A. 102*, 13767-13772.

[5] Hofmann, R. M., and Muir, T. W. (2002) Recent advances in the application of expressed protein ligation to protein engineering, *Curr. Opin. Biotechnol. 13*, 297-303.

[6] Link, A. J., Mock, M. L., and Tirrell, D. A. (2003) Non-canonical amino acids in protein engineering, *Curr. Opin. Biotechnol. 14*, 603-609.

[7] Moore, S. J., Hayden Gephart, M. G., Bergen, J. M., Su, Y. S., Rayburn, H., Scott, M. P., and Cochran, J. R. (2013) Engineered knottin peptide enables noninvasive optical imaging of intracranial medulloblastoma, *Proc. Natl. Acad. Sci. U. S. A. 110*, 14598-14603.

[8] Ji, Y., Majumder, S., Millard, M., Borra, R., Bi, T., Elnagar, A. Y., Neamati, N., Shekhtman, A., and Camarero, J. A. (2013) *In vivo* activation of the p53 tumor suppressor pathway by an engineered cyclotide, *J. Am. Chem. Soc. 135*, 11623-11633.

[9] Wang, C. K., Gruber, C. W., Cemazar, M., Siatskas, C., Tagore, P., Payne, N., Sun, G. Z., Wang, S. H., Bernard, C. C., and Craik, D. J. (2014) Molecular grafting onto a stable framework yields novel cyclic peptides for the treatment of multiple sclerosis, *ACS Chem. Biol. 9*, 156-163.

[10] Thongyoo, P., Roque-Rosell, N., Leatherbarrow, R. J., and Tate, E. W. (2008) Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides, *Org. Biomol. Chem. 6*, 1462-1470.

[11] Wong, C. T. T., Rowlands, D. K., Wong, C.-H., Lo, T. W. C., Nguyen, G. K. T., Li, H.-Y., and Tam, J. P. (2012) Orally active peptidic bradykinin B1 receptor antagonists

engineered from a cyclotide scaffold for inflammatory pain treatment, *Angew. Chem. 124*, 5718-5722.

[12] Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J., and Kim, H. S. (2006) Design and evolution of new catalytic activity with an existing protein scaffold, *Science 311*, 535-538.

[13] Liu, S., Liu, S., Zhu, X., Liang, H., Cao, A., Chang, Z., and Lai, L. (2007) Nonnatural protein-protein interaction-pair design by key residues grafting, *Proc. Natl. Acad. Sci. U. S. A. 104*, 5330-5335.

[14] Sollazzo, M., Billetta, R., and Zanetti, M. (1990) Expression of an exogenous peptide epitope genetically engineered in the variable domain of an immunoglobulin: implications for antibody and peptide folding, *Protein Eng. 4*, 215-220.

[15] Zanetti, M., Filaci, G., Lee, R. H., del Guercio, P., Rossi, F., Bacchetta, R., Stevenson, F., Barnaba, V., and Billetta, R. (1993) Expression of conformationally constrained adhesion peptide in an antibody CDR loop and inhibition of natural killer cell cytotoxic activity by an antibody antigenized with the RGD motif, *EMBO J. 12*, 4375-4384.

[16] Lee, G., Chan, W., Hurle, M. R., DesJarlais, R. L., Watson, F., Sathe, G. M., and Wetzel, R. (1993) Strong inhibition of fibrinogen binding to platelet receptor αIIbβ3 by RGD sequences installed into a presentation scaffold, *Protein Eng., Des. Sel. 6*, 745-754.

[17] Yamada, T., Matsushima, M., Inaka, K., Ohkubo, T., Uyeda, A., Maeda, T., Titani, K., Sekiguchi, K., and Kikuchi, M. (1993) Structural and functional analyses of the Arg-Gly-Asp sequence introduced into human lysozyme, *J. Biol. Chem. 268*, 10588-10592.

[18] Yamada, T., Uyeda, A., Kidera, A., and Kikuchi, M. (1994) Functional analysis and modeling of a conformationally constrained Arg-Gly-Asp sequence inserted into human lysozyme, *Biochemistry 33*, 11678-11683.

[19] Wolfson, A. J., Kanaoka, M., Lau, F. T. K., and Ringe, D. (1991) Insertion of an elastase-binding loop into interleukin-1β, *Protein Eng., Des. Sel. 4*, 313-317.

[20] Moore, S. J., Leung, C. L., Norton, H. K., and Cochran, J. R. (2013) Engineering agatoxin, a cystine-knot peptide from spider venom, as a molecular probe for in vivo tumor imaging, *PLoS One 8*, e60498.

[21] Reiss, S., Sieber, M., Oberle, V., Wentzel, A., Spangenberg, P., Claus, R., Kolmar, H., and Lösche, W. (2006) Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins, *Platelets 17*, 153-157.

[22] Krause, S., Schmoldt, H.-U., Wentzel, A., Ballmaier, M., Friedrich, K., and Kolmar, H. (2007) Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists, *FEBS J. 274*, 86-95.

[23] Sommerhoff, C. P., Avrutina, O., Schmoldt, H.-U., Gabrijelcic-Geiger, D., Diederichsen, U., and Kolmar, H. (2010) Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β, *J. Mol. Biol. 395*, 167-175.

[24] Chan, L. Y., Gunasekera, S., Henriques, S. T., Worth, N. F., Le, S. J., Clark, R. J., Campbell, J. H., Craik, D. J., and Daly, N. L. (2011) Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds, *Blood 118*, 6709-6717.

[25] Aboye, T. L., Ha, H., Majumder, S., Christ, F., Debyser, Z., Shekhtman, A., Neamati, N., and Camarero, J. A. (2012) Design of a novel cyclotide-based CXCR4 antagonist with anti-human immunodeficiency virus (HIV)-1 activity, *J. Med. Chem. 55*, 10729-10734.

[26] Gunasekera, S., Foley, F. M., Clark, R. J., Sando, L., Fabri, L. J., Craik, D. J., and Daly, N. L. (2008) Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides, *J. Med. Chem. 51*, 7697-7704.

[27] Eliasen, R., Daly, N. L., Wulff, B. S., Andresen, T. L., Conde-Frieboes, K. W., and Craik, D. J. (2012) Design, synthesis, structural and functional characterization of novel melanocortin agonists based on the cyclotide kalata B1, *J. Biol. Chem. 287*, 40493-40501.

[28] Drakopoulou, E., Zinn-Justin, S., Guenneugues, M., Gilquin, B., Ménez, A., and Vita, C. (1996) Changing the structural context of a functional β-hairpin, *J. Biol. Chem. 271*, 11979-11987.

[29] Drakopoulou, E., Vizzavona, J., and Vita, C. (1998) Engineering a CD4 mimetic inhibiting the binding of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120 to human lymphocyte CD4 by the transfer of a CD4 functional site to a small natural scaffold, *Lett. Pept. Sci. 5*, 241-245.

[30] Vita, C., Roumestand, C., Toma, F., and Ménez, A. (1995) Scorpion toxins as natural scaffolds for protein engineering, *Proc. Natl. Acad. Sci. U. S. A. 92*, 6404-6408.

[31] Vita, C., Drakopoulou, E., Vizzavona, J., Rochette, S., Martin, L., Ménez, A., Roumestand, C., Yang, Y.-S., Ylisastigui, L., Benjouad, A., and Gluckman, J. C. (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein, *Proc. Natl. Acad. Sci. U. S. A. 96*, 13091-13096.

[32] Li, C., Liu, M., Monbo, J., Zou, G., Li, C., Yuan, W., Zella, D., Lu, W.-Y., and Lu, W. (2008) Turning a scorpion toxin into an antitumor miniprotein, *J. Am. Chem. Soc. 130*, 13546-13548.

[33] Conibear, A. C., Bochen, A., Rosengren, K. J., Stupar, P., Wang, C., Kessler, H., and Craik, D. J. (2014) The cyclic cystine ladder of θ-defensins as a stable, bifunctional scaffold: a proof-of-concept study using the integrin-binding RGD motif, *ChemBioChem 15*, 451-459.

[34] Knappe, T. A., Manzenrieder, F., Mas-Moruno, C., Linne, U., Sasse, F., Kessler, H., Xie, X., and Marahiel, M. A. (2011) Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist, *Angew. Chem., Int. Ed. 50*, 8714-8717.

[35] Craik, D. J., Daly, N. L., Bond, T., and Waine, C. (1999) Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif, *J. Mol. Biol. 294*, 1327-1336.

[36] Daly, N. L., and Craik, D. J. (2000) Acyclic permutants of naturally occurring cyclic proteins: characterization of cystine knot and β-sheet formation in the macrocyclic polypeptide kalata B1, *J. Biol. Chem. 275*, 19068-19075.

[37] Giancotti, F. G., and Ruoslahti, E. (1999) Integrin signaling, *Science 285*, 1028-1033.

[38] Luo, B.-H., Carman, C. V., and Springer, T. A. (2007) Structural basis of integrin regulation and signaling, *Annu. Rev. Immunol. 25*, 619-647.

[39] Plow, E. F., Haas, T. A., Zhang, L., Loftus, J., and Smith, J. W. (2000) Ligand binding to integrins, *J. Biol. Chem. 275*, 21785-21788.

[40] Pierschbacher, M. D., and Ruoslahti, E. (1984) Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule, *Nature 309*, 30-33.

[41] Pierschbacher, M. D., and Ruoslahti, E. (1984) Variants of the cell recognition site of fibronectin that retain attachment-promoting activity, *Proc. Natl. Acad. Sci. U. S. A. 81*, 5985-5988.

[42] Ruoslahti, E., and Pierschbacher, M. D. (1986) Arg-Gly-Asp: a versatile cell recognition signal, *Cell 44*, 517-518.

[43] Kostidis, S., Stavrakoudis, A., Biris, N., Tsoukatos, D., Sakarellos, C., and Tsikaris, V. (2004) The relative orientation of the Arg and Asp side chains defined by a pseudodihedral angle as a key criterion for evaluating the structure-activity relationship of RGD peptides, *J. Pept. Sci. 10*, 494-509.

[44] Hersel, U., Dahmen, C., and Kessler, H. (2003) RGD modified polymers: biomaterials for stimulated cell adhesion and beyond, *Biomaterials 24*, 4385-4415.

[45] Mayer, B., and Hemmens, B. (1997) Biosynthesis and action of nitric oxide in mammalian cells, *Trends Biochem. Sci. 22*, 477-481.

[46] Kuang, Z., Lewis, R. S., Curtis, J. M., Zhan, Y., Saunders, B. M., Babon, J. J., Kolesnik, T. B., Low, A., Masters, S. L., Willson, T. A., Kedzierski, L., Yao, S., Handman, E., Norton, R. S., and Nicholson, S. E. (2010) The SPRY domain-containing SOCS box protein SPSB2 targets iNOS for proteasomal degradation, *J. Cell Biol. 190*, 129-141.

[47] Yap, B. K., Leung, E. W. W., Yagi, H., Galea, C. A., Chhabra, S., Chalmers, D. K., Nicholson, S. E., Thompson, P. E., and Norton, R. S. (2014) A potent cyclic peptide targeting SPSB2 protein as a potential anti-infective agent, *J. Med. Chem. 57*, 7006-7015.

[48] Karplus, M., and Petsko, G. A. (1990) Molecular dynamics simulations in biology, *Nature 347*, 631-639.

[49] Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, *J. Comput. Chem. 25*, 1656-1676.

[50] Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and van Gunsteren, W. F. (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7, *Eur. Biophys. J. 40*, 843-856.

[51] MacKerell, A. D., Brooks, B., Brooks, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (2002) CHARMM: the energy function and its parameterization, In *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd.

[52] Feenstra, K. A., Hess, B., and Berendsen, H. J. (1999) Improving efficiency of large timescale molecular dynamics simulations of hydrogen-rich systems, *J. Comput. Chem. 20*, 786-798.

[53] Takada, S. (2012) Coarse-grained molecular simulations of large biomolecules, *Curr. Opin. Struct. Biol. 22*, 130-137.

[54] McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins, *Nature 267*, 585-590.

[55] Onufriev, A. (2008) Implicit solvent models in molecular dynamics simulations: a brief overview, In *Annual Reports in Computational Chemistry* (Wheeler, R. A., and Spellmeyer, D. C., Eds.), pp 125-137, Elsevier.

[56] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

[57] MacKerell, A. D. (2004) Empirical force fields for biological macromolecules: overview and issues, *J. Comput. Chem. 25*, 1584-1604.

[58] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX 1–2*, 19-25.

[59] Kimura, R. H., Levin, A. M., Cochran, F. V., and Cochran, J. R. (2009) Engineered cystine knot peptides that bind αvβ3, αvβ5, and α5β1 integrins with low-nanomolar affinity, *Proteins: Struct., Funct., Bioinf. 77*, 359-369.

[60] Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun. 91*, 43-56.

[61] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling, *J. Chem. Phys. 126*, 014101.

[62] Maestro, version 10.3 (2015). Schrödinger, LLC, New York

[63] Savarin, P., Romi-Lebrun, R., Zinn-Justin, S., Lebrun, B., Nakajima, T., Gilquin, B., and Menez, A. (1999) Structural and functional consequences of the presence of a fourth disulfide bridge in the scorpion short toxins: solution structure of the potassium channel inhibitor HsTX1, *Protein Sci. 8*, 2672-2685.

[64] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graphics 14*, 33-38.

[65] Hunter, J. D. (2007) Matplotlib: a 2D Graphics Environment, *Comput. Sci. Eng. 9*, 90-95.

[66] Rashid, M. H., Huq, R., Tanner, M. R., Chhabra, S., Khoo, K. K., Estrada, R., Dhawan, V., Chauhan, S., Pennington, M. W., Beeton, C., Kuyucak, S., and Norton, R. S. (2014) A potent and Kv1.3-selective analogue of the scorpion toxin HsTX1 as a potential therapeutic for autoimmune diseases, *Sci. Rep. 4*, 4509.

[67] Richards, J., Miller, M., Abend, J., Koide, A., Koide, S., and Dewhurst, S. (2003) Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human αvβ3 integrin, *J. Mol. Biol. 326*, 1475-1488.

[68] Craig, W. S., Cheng, S., Mullen, D. G., Blevitt, J., and Pierschbacher, M. D. (1995) Concept and progress in the development of RGD-containing peptide pharmaceuticals, *Biopolymers 37*, 157-175.

[69] Cuthbertson, A., and Indrevoll, B. (2003) Regioselective formation, using orthogonal cysteine protection, of an α-conotoxin dimer peptide containing four disulfide bonds, *Org. Lett. 5*, 2955-2957.

Chapter 2
Accessing peptide toxin samples

1 Introduction

In this research project, one of the aims was to study newly discovered conotoxins from the venom duct transcriptome of the cone snail *Conus victoriae*. One planned method of analysis was solution nuclear magnetic resonance (NMR) spectroscopy, to determine the three-dimensional structure. In order to facilitate this investigation, samples of toxins were required in multi-milligram quantities at >95% purity. There are several different methods that can be utilised to obtain such samples, such as direct isolation from the venom, solid-phase peptide synthesis (SPPS) and recombinant expression. Each method comes with certain strengths and challenges, which may make it ideal for some investigations and unsuited for others. In the context of the newly discovered H-superfamily conotoxins H_Vc7.1 and H_Vc7.2, the crude venom was a precious resource in limited supply. Moreover, the toxins were not abundant in the venom. The peptides were of a length tractable by SPPS (being less than 50 residues), which would enable the production of native material and analogues. Recombinant production in *E. coli* would additionally allow the prospect of double-labelling the sample with heavy isotopes, which would be prohibitively expensive by SPPS. Double-labelling would allow heavy atom NMR spectra to be collected in a reasonable timeframe, which would assist in assigning resonances and calculating the solution structure. In this chapter, a brief background to the technical approaches to peptide toxin production is provided, followed by details of attempts to obtain samples of newly discovered conotoxins from *C. victoriae* by recombinant expression.

Conceptually, each of the three methods mentioned for obtaining peptide samples follows the same sequence of steps (**Figure 1**). The first is to form the polypeptide chain, with amino acid residues arranged in the correct order to conform to the primary sequence of the peptide of interest. Next, the peptide must be folded into the correct three-dimensional structure with secondary structural elements such as α-helices and β-sheets arranged in the correct orientation. For cysteine-rich peptides, including many toxins, this also requires the disulfide bond network to be properly connected. The final step is purification, to ensure the experiments and assays to be carried out are not compromised by contaminating species.

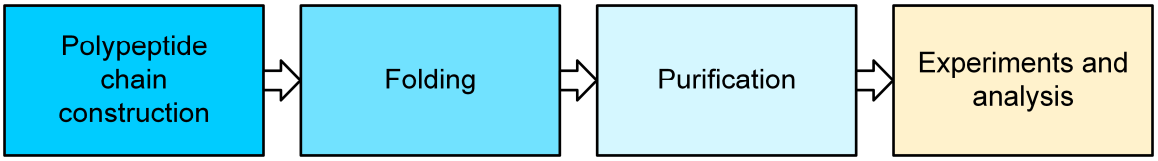


Figure 1. Flowchart outlining the steps necessary to produce a peptide sample for analysis.

Each of the three methods accomplishes these steps in its own manner. When isolating a toxin from the venom, the endogenous protein translation machinery of the source organism accomplishes the first two steps; the challenge for the researcher is to isolate the peptide of interest from the rest of the venom components. In recombinant expression, the formation of the polypeptide chain and (in some cases) the folding are performed by the machinery of the chosen host cell. In other cases the cellular machinery is unable to correctly fold the sample and a separate refolding procedure is required. SPPS accomplishes the first step through synthetic chemistry, sequentially adding amino acids to the growing polypeptide chain. The folding step is also performed *in vitro* and may require testing and optimisation. Purification is again the final step. As the three methods differ in how they produce the target peptide, they also differ in other respects, such as the length of time required, cost, suitability for particular peptides (especially in the area of post-translational modifications) and the complexity of the final product. The benefits and challenges of each method are discussed in more detail in the sections that follow.

1.1 Isolation from venom

Isolating samples of a peptide of interest from the natural source – for example, isolating conotoxins from cone snail venom – is an important technique, as it ensures the material recovered is complete and correct. Any post-translational modifications present can be identified and the structure can be accepted as the canonical native fold. This technique can therefore be used to confirm the identity and accuracy of peptides produced by the other two methods to be discussed. Some early studies on conotoxins generated all of their toxin material in this manner,[1-3] although most recent studies rely on one of the alternative approaches.

In most cases, the abundance of a toxin of interest in the venom is low. Collecting a sufficient stock of venom from which to extract the required amounts of toxin is therefore the first hurdle. This is not always straightforward, as venom composition can vary from individual to individual within the same species. A study comparing the peptide complement of the cone snail species *Conus textile* revealed that fewer than 7% of the

peptides detected were shared by the three individuals studied.[4] The method of venom collection also affects the composition of the sample. Dissecting the venom apparatus from a specimen and extracting the contained venom allows access to all peptides present in the duct, but also commonly introduces non-venom material from the surrounding tissue. In particular, these contaminants can include proteases, which may degrade the peptides of interest before they can be isolated. In addition, this approach limits the amount of venom that can be obtained from each specimen and can have a high cost in animal sacrifice. Milking venom from captive specimens provides an alternative. The venom thus obtained is typically less complex than material extracted from the duct, but many milkings are generally required over an extended period of time to amass sufficient quantities of venom. The state of the animal upon milking can also affect the sample composition, as distressing a cone snail prior to milking elicits the production and expression of a 'defensive' venom, distinct from the predatory venom used on prey.[5] This variability requires researchers to carefully consider how to acquire their stock of venom.

Once a stock of venom is obtained, a method for separating the individual venom components is required. This is typically a chromatographic method that divides the sample into several fractions. Fractions may need to be sub-fractionated to completely purify the components. The final requirement for successfully isolating a peptide from the venom is a technique for identifying which fraction contains the molecule of interest. Historically, venom fractions were tested in assays for biological activity. Fractions that showed an interesting effect in the assay were further investigated and characterised to reveal the identity of the isolated peptide. More recent investigations make use of LC-MS and LC-MS/MS to identify the peptides present in fractions.

To illustrate the process of isolating peptides from venom, the discovery of contryphan-R may be used as an example.[6] This peptide (and a deletion mutant) were identified by isolating the toxins from the venom of the cone snail *Conus radiatus*. The venom sample was obtained by dissecting the venom duct and extracting the toxins with acetonitrile. Initial RP-HPLC of the extracted venom resulted in a multitude of peaks, while bioassay of the fractions by intracranial injection in mice revealed a biological activity in two peaks in particular. The bioactive fractions were then re-chromatographed with a shallower gradient, revealing multiple peaks that were in superposition in the initial run. Another round of bioassays was used to identify the active peaks from these chromatograms, which were then purified to homogeneity. The two peptides thus isolated were investigated and named contryphan-R and des-[Gly1]-contryphan-R.

In the context of this research project, the crude venom was a precious resource in scarce supply and the toxins to be investigated were of very low abundance. In addition, much work had been done at the transcriptomic level that led to a complete knowledge of the peptide sequence, without the need to isolate the toxins from a venom sample. Therefore, alternative methods were used to produce samples for testing and isolation from venom was not attempted in this investigation.

1.2 Recombinant expression

In recombinant expression systems, the transcriptional and translational machinery of a host cell is harnessed to express the protein of interest. The host cell is transformed with an expression vector containing DNA encoding the protein of interest and various genetic markers to promote the translation of the recombinant sequence. The cells are then maintained in optimal growth conditions to allow the protein to be expressed. Several different types of cell are available as recombinant hosts, including mammalian cells, insect cells, yeast, plant cells and bacteria. Choosing the best host for a particular protein is a complex decision, based on the relative importance of several different factors.

1.2.1 Mammalian host cells

The pharmaceutical industry is making increasing use of so-called ‘biologics’, which are proteinaceous therapeutics expressed recombinantly. The proteins thus produced are usually of human origin to maximise biocompatibility with the human end-users. When producing a protein of human origin via recombinant expression, there is an advantage in using a mammalian cell line, such as Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cells. A mammalian cell producing a human (mammalian) protein possesses all of the biochemical machinery required for proper folding and post-translational modification of the expressed sequences,[7] which could be lacking in other potential host strains. However, mammalian cells are also challenging and expensive to maintain and so are rarely used when this advantage is not required.[8] As such, although widely used in the pharmaceutical industry, mammalian cells have not been utilised to any great extent in the production of peptide toxins for research purposes, as the effort and expense do not translate to a substantially better outcome.

1.2.2 Insect host cells

Insect cells can also be used as recombinant hosts, such as the Sf9 strain from *Spodoptera frugiperda*. In the most common system, insect cells are infected with a recombinant baculovirus coding for the protein of interest, under the control of the viral polyhedrin promoter.[9] Once the protein is produced the cells lyse, releasing the product into the media where it can be harvested. The eukaryotic nature of insect cells provides them with much of the post-translational modification machinery that would be lacking in prokaryotic cells, which makes them attractive for the production of modified peptides.[9] Insect cells are also easier to cultivate than mammalian cells and the baculovirus system can deliver higher expression levels of the protein of interest.[10] However, insect cells still incur higher maintenance overheads than other potential host cell systems (discussed below). There are reports in the literature of peptide toxins being produced by this system, such as the spider toxin huwentoxin-1, although insect cells were only used in this study after production by yeast and bacteria yielded unsatisfactory results, rather than as a first preference.[11] Toxins from the venom of the black widow spider have also been produced by Sf21 cells.[12] Although useful, this system is seldom the first choice for the production of toxin peptides.

1.2.3 Yeast host cells

Another widely used kind of host cell are yeasts, most notably *Saccharomyces cerevisiae* and *Pichia pastoris*. As with the other eukaryotic systems, yeast cells are capable of post-translational modifications, although not the full range available to mammalian cells. Yeast cells also boast secretion pathways that can direct the produced protein to the culture medium where it can be easily harvested. The genome of *S. cerevisiae* in particular has been extensively studied, aiding attempts at genetic manipulation. In addition to being eukaryotic, yeast cells can be grown rapidly in inexpensive media to high cell density.[13] These advantages make yeast cells an attractive alternative to other eukaryotic systems and several studies have made use of this expression system. Toxins have been expressed from the venom of a variety of species, including scorpions, spiders, snakes and cone snails.[14-18] A scorpion toxin was expressed in *P. pastoris* and found to have insecticidal activity,[15] while *S. cerevisiae* was used to express the spider peptide jingzhaotoxin-34 for functional studies.[16] The snake neurotoxin α-bungarotoxin has been produced recombinantly by *P. pastoris*,[18] as has the conotoxin TxVIA from *C. textile*.[17]

1.2.4 Plant host cells

Recombinant expression can also make use of plant cells, either in culture or as an entire transgenic plant. Like the other systems discussed so far, plant cells are eukaryotic and thus suitable for producing proteins and peptides with post-translational modifications.[19] There is also less risk of harmful contaminants being introduced to the sample with this system, as plant cells are not afflicted by human pathogens and (unlike bacteria) the cells do not produce endotoxins.[20] The decision to express a protein in whole transgenic plants necessitates consideration of additional requirements, such as which type of plant to use and which plant component to accumulate the recombinant product in. Tobacco leaves, cereal seeds and potato tubers have all been used, the selection of each system influenced by factors such as production time, level of protein accumulation and ease of or need for downstream processing.[21]

Producing recombinant proteins in plant cells in culture is a newer technology, developed to address the lack of precise control inherent in whole transgenic plants. Cultured cells allow greater batch-to-batch consistency, greater containment of the transgenic material and easier compliance with good manufacturing process standards.[19] However, plant cells grow much more slowly than microbial cells and typically deliver relatively low protein yields.[19] Several kinds of recombinant proteins and antibodies have been produced by this host system,[22] but there are few examples of venom peptides being produced in plants for further research; one report details the production of the analgesic-antitumor peptide from the scorpion *Buthus martensi* in this fashion.[23] There are also examples in the literature of venom peptides (commonly from spider venom) being expressed in transgenic plants for their insecticidal properties, proving the viability of plants as recombinant host for the production of toxins and the efficacy of the folding apparatus.[24-26]

1.2.5 Bacterial host cells

Perhaps the most commonly utilised recombinant expression host system is the Gram-negative bacterium *Escherichia coli.* The genetics of *E. coli* have been extensively studied and as a microorganism this host system provides several advantages, such as ease of growth, the existence of plasmids as a convenient mode of genetic material transfer and the high culture densities attainable.[27] There are however some limitations: as prokaryotes, bacteria lack the machinery required for eukaryotic post-translational modifications, making them unsuitable for the production of modified peptides. Additionally, the

cytoplasm is a naturally reducing environment, which inhibits the folding of disulfide-rich peptides and proteins. To overcome this, a signal sequence can be included on the protein construct to direct the expressed sequence to the periplasm, which is an oxidising environment that can facilitate folding.[28] Another option is to make use of a specially engineered strain of bacteria such as *E. coli* Origami™ or SHuffle®, which have an altered cellular redox potential and an oxidising cytoplasm.[29] These strains are optimised for the production of disulfide-bonded sequences, and may not perform as well for other sequences due to the increased oxidative stress experienced by the cell.[30]

An advantage of using recombinant bacteria to produce peptide samples is that the resulting products can be uniformly labelled with heavy isotopes of carbon and nitrogen. Unlike the abundant ^{12}C and ^{14}N, the heavy isotopes ^{13}C and ^{15}N are magnetically active and can be detected by NMR spectroscopy. Minimal growth media supplemented with ^{13}C-glucose and ^{15}N-ammonium ions can be used to produce double-labelled peptides that can be readily analysed by NMR, simplifying the process of calculating a solution structure. This capability has been exploited in previous studies on recombinant toxins.[31, 32] Even without this advantage, the speed and ease of cultivating *E. coli* cells has resulted in this system being used to produce many toxins from several venomous species and it is often the system of choice for unmodified peptides.[33-35]

1.3 Fusion construct design

In bacterial recombinant expression systems, peptides of interest are rarely produced on their own in the host cell; it is far more common to express the peptide of interest as part of a fusion construct, with another protein as a fusion partner.[36] The fusion partner is often selected to enhance the properties of the peptide of interest. Many partners are chosen to increase solubility, although others can direct the peptide into insoluble inclusion bodies.[36] Some fusion partners also act as affinity tags and aid in purification, although others lack this functionality and require the inclusion of specific affinity tags for this purpose. A final component of most fusion constructs is a protease recognition sequence, strategically placed to liberate the desired protein from the fusion so that the protein of interest may be studied in isolation.

Each of the recombinant expression systems therefore have their own mixture of properties that makes them suitable for different applications (**Table 1**). For peptide toxins the two systems most commonly employed are yeast and bacteria, although some studies have made use of others, as noted.

Table 1. Comparison of features of recombinant expression host systems.

Host system	Expense	Timeframe	PTMs possible
Mammalian cells	High	Long	Yes
Insect cells	High	Long	Yes
Transgenic plants	Moderate	Long	Yes (limited)
Plant cells	Low	Moderate	Yes (limited)
Yeast	Low	Short	Yes (restricted)
Bacteria (*E. coli*)	Low	Short	No

1.4 Solid-phase peptide synthesis

Solid-phase peptide synthesis (SPPS) is a technique involving a repeating series of chemical reactions that sequentially couple amino acids together on a solid support to synthesise a peptide.[37] The 'solid phase' of the technique is an inert resin on which the growing peptide chain is anchored. The resin allows the reaction vessel to be drained of reagents at the completion of each step, while retaining the growing peptide chain. The steps required to construct the polypeptide are outlined in **Figure 2**.

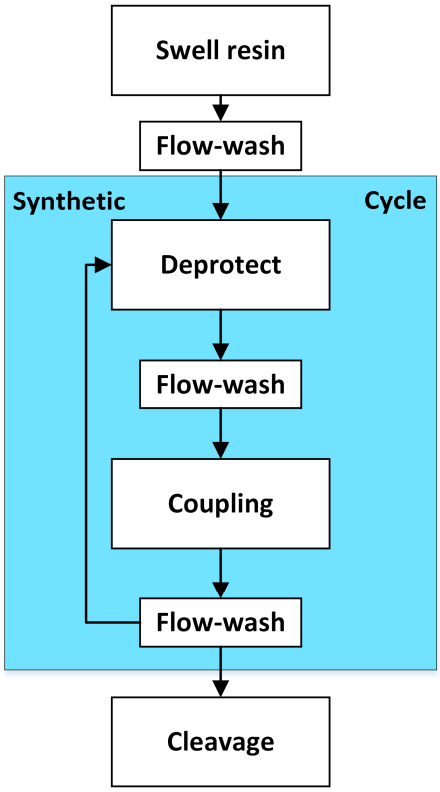


Figure 2. Schematic diagram outlining the synthetic cycle followed in solid-phase peptide synthesis.

During the synthesis, reactive amino acid side chains are chemically protected to prevent side-reactions that would result in misformed peptides. The α-amino group is so protected, so this protecting group must be removed in the 'deprotection' step of the synthetic cycle. The deprotected amino group is then available for coupling with the carboxy group of the next amino acid, which forms the peptide bond and extends the polypeptide chain. As a consequence, the peptide is constructed from the C-terminus towards the N-terminus. There are two common protecting groups for the α-amino group, which differ in the chemistry required to remove them from the chain and therefore necessitate the adoption of different reaction schemes. These α-amino protecting groups are *tert*-Butyloxycarbonyl (Boc) and 9-(Fluorenyl)methyloxycarbonyl (Fmoc), illustrated in **Figure 3**.

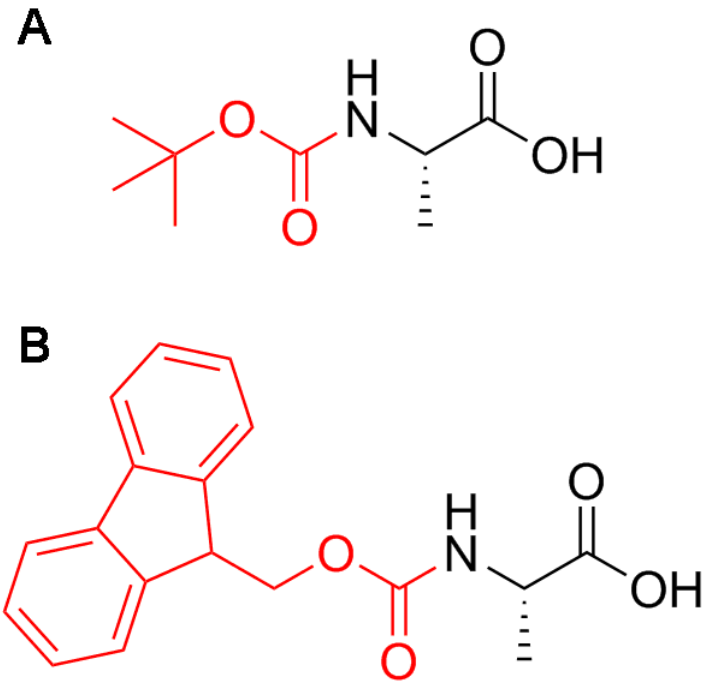


Figure 3. Structures of solid-phase peptide synthesis amino protecting groups.
(A) Boc-(L-Alanine); (B) Fmoc-(L-Alanine).

The Boc protecting group was the first to be adopted and in the early days of SPPS had the advantages of low costs, high purity and high solubility.[37] The utility of this protecting group is somewhat limited by the fact that it is acid-labile, as is the typical linkage to the polymer resin. Having both types of cleavable bond susceptible to the same chemical conditions necessitates a reduction in selectivity: if the deprotection step is protracted, loss of peptide from the resin can occur. When using the Boc protection scheme, selectivity is modulated by the strength of each acid treatment; deprotection is carried out by trifluoroacetic acid (TFA), while final cleavage is performed with the much harsher

hydrofluoric acid (HF). The use of HF imposes a further difficulty, since specialised equipment is required to handle this dangerous chemical safely. The Boc protecting group is therefore less commonly used now that Fmoc has matured as an alternative.

The Fmoc protecting group allows an 'orthogonal' protection scheme, in which each bond can be removed without affecting the others, improving the final yield of the product. The Fmoc group itself is labile under basic conditions and is removed with solutions of piperidine. Cleavage from the resin is performed under acidic conditions using TFA. There is no chance that the attempted cleavage of one bond can affect the other, which, coupled with the safety benefit, gives rise to the preference for Fmoc over Boc in modern SPPS.

The α-amino moiety is not the only chemical entity that requires a protecting group. The sidechains of several amino acids can also be reactive and require protection to ensure that they do not form undesired products. These protecting groups must remain in place throughout the synthesis and are only removed in the final step, when the polypeptide chain is released from the resin by cleaving the covalent bond. Side-chain protecting strategies have therefore been developed to complement both the Boc- and Fmoc-based peptide synthesis protocols.[37]

Peptide synthesis gives exquisite control over the composition of a peptide sample. Additionally, even crude synthetic preparations contain far fewer chemical species than crude venom or recombinant cell lysates, simplifying purification. The primary limitation is the size of the peptide to be synthesised, as long peptides (>~50 residues) face an increased risk of mis-couplings and other errors. The cost of the reagents required is also a consideration and can be a limiting factor. This cost is magnified greatly if heavy isotope labelling is desired, as labelled amino acids must be purchased for use in the synthesis. For this reason recombinant expression is the preferred production method for most labelled peptides. Nonetheless, chemical synthesis has been used by many groups over many years to access peptide toxins for analysis.[38]

1.5 Peptide refolding

Once the peptide chain has been constructed, the peptide must be folded into the native conformation, with disulfide bonds formed correctly and secondary structural elements in place. In most cases this conformation is a thermodynamically favourable arrangement, occupying a minimum on the potential energy landscape for the peptide.[39] The potential complexity of the disulfide bond network is a combinatorial problem; as the number of cysteine residues in a peptide sequence increases, the number of ways they can be paired

into disulfide bonds rises exponentially. A single pair of cysteines have a single way to connect. Two pairs can connect in three different configurations. With three pairs, fifteen possibilities exist – and so on. Of all the possible configurations, generally only one is the desired native form (although some rare peptides do exist in nature with multiple disulfide isoforms, for example the IgG2 subclass of antibodies[40]). If a molecule is produced that connects the cysteines in an alternate manner, the misfolded isomer becomes an impurity in the final sample. Such isomers may be worthy research subjects in their own right, with interesting activities distinct from the native peptide, but if the goal of the study is to investigate the native species then their presence causes a difficulty. Under certain conditions, incorrectly connected disulfide bonds can be 'shuffled', which allows the folding process to seek the energy minimum once more and makes it more likely that the native conformation will be achieved. If a refolding step is necessary in a peptide production workflow, conditions must be found that drive the peptide towards the native state and minimise the formation of misfolded isomers.

Formation of a disulfide bond is an oxidative process, in which a reactive thiolate ion of the Cys side chain mounts a nucleophilic attack on an oxidant. The partner thiolate then attacks this mixed species, resulting in the disulfide bond.[41] To facilitate this process, refolding buffers should contain an oxidant or redox-shuffling reagent and be of basic pH.[42] The presence of a redox-active compound provides the target for the nucleophilic attack, while high pH maintains the thiolates in the active form.[43] The exact composition of a refolding buffer is something that must be determined empirically; a buffer that works for one peptide will not necessarily provide the best results for another.

2 Research rationale

The H-superfamily conotoxins H_Vc7.1 and H_Vc7.2 were to be studied by NMR spectroscopy. Mass spectroscopy data revealed that these peptides lacked post-translational modifications, which lead to the decision to produce the samples recombinantly in *E. coli*. This would allow for double-labelling with heavy isotopes by growing the cells on supplemented minimal media. Having selected this method of production, it was necessary to design a fusion construct for the expression.

A recombinant construct was designed with thioredoxin as a fusion protein partner (**Figure 4**). Thioredoxin enhances the solubility of fusion constructs and has also been

reported to assist in the proper folding of recombinant peptides and antibodies.[44] A polyhistidine (6xHis) tag was included in the construct, in the linker between thioredoxin and the toxin sequence, to assist with purification by nickel affinity chromatography. Finally, a TEV protease cleavage site (ENLFYQ'X) was included immediately adjacent to the toxin sequence, to allow the toxin to be released from the construct. The TEV protease contains a catalytic Cys residue, which requires a reducing agent in order to be maintained in the active state. This protease is also very selective and only cleaves peptides at the canonical cleavage site. Moreover, the P1' position of the recognition site can tolerate a variety of amino acid residues to varying degrees. In the designed fusion constructs, the P1' position of the cleavage site was also the N-terminal residue of the toxin, Ala for H_Vc7.1 and Asn for H_Vc7.2. These residues are well tolerated, having been found to result in cleavage efficiencies of over 95%.[45]

Figure 4. Schematic diagram of H_Vc7.1/H_Vc7.2 recombinant expression fusion construct. Elements are not to scale.

The expected molecular masses of the fusion constructs were 20.6 kDa for Trx-H_Vc7.1 and 20.1 kDa for Trx-H_Vc7.2. The processing of the expressed fusion protein to the final toxin sample involved a number of steps, each of which required optimisation. Details and results of the various trials undertaken are presented in the following sections.

3 Methods

3.1 Production of expression plasmids for H-superfamily toxins

Synthetic genes were ordered from Life Technologies Corporation (CA, USA), containing the sequence for H_Vc7.1 or H_Vc7.2 flanked by NcoI and KpnI restriction sites. Sequences were optimised to use bacterial codons. Pfx Platinum PCR kit (Life Technologies Corporation, USA) was used to amplify the DNA using specific primers (5'–3') GGTAGCGGTAGCGGTACC (forward) and CTACCACCTTCCATGGTTA (reverse) (GeneWorks Pty Ltd Custom Oligo Service, SA, Australia). Reaction products were run on 1% agarose gel to separate amplified product from impurities. QIAquick gel extraction kit (QIAGEN Sample & Assay Technologies, Germany) was used to extract DNA from excised bands of the gel. KpnI-HF and NcoI-HF (New England Biolabs, MA USA) were

used to cleave the DNA ahead of ligation into pET-32a expression vector with ExpressLink T4 DNA ligase (Life Technologies Corporation, USA). Ligated plasmid was used to transform DH5α competent cells, which were grown and harvested. The resultant DNA was purified via Miniprep (AccuPrep Plasmid Mini Extraction Kit, Bioneer, Republic of Korea) and sequenced by the Australian Genome Research Facility.

3.2 Expression of Trx-Vc7.1 and Trx-Vc7.2 fusion constructs

10 mL LB media (100 μg/mL ampicillin) starter cultures were inoculated with 1 mL previously prepared BL21(DE3) glycerol stock and grown overnight (16 h) in an orbital shaker (180–200 rpm) at 37 °C. 500 mL LB media (100 μg/mL ampicillin) expression cultures were inoculated with 1 mL starter culture and incubated at 150 rpm in an orbital shaker at 37 °C until OD_{600} = 0.7. Protein expression was induced by addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), induced cultures were incubated in an orbital shaker at room temperature for a minimum of 4 h.

3.3 Cell harvesting and lysis

Cells were pelleted by centrifugation at 9800 *g* for 10 min at 4 ºC. Pellets were weighed and either processed immediately or stored at -80 ºC. Cell pellets were resuspended in 50 mM phosphate, pH 7.0, with 250 mM NaCl and lysed by sonication, on ice, with 6 x 30 s bursts. Cellular debris was pelleted by centrifugation at 20,000 rpm for 20 min at 4 °C. The supernatant was decanted into fresh centrifuge tubes and debris re-pelleted under the same conditions. Supernatant containing soluble species (including thioredoxin-conotoxin fusion construct) was filtered through a sterile 0.2 μm membrane and stored on ice prior to nickel affinity chromatography.

3.4 Fusion construct purification

The thioredoxin-conotoxin fusion construct was separated from other soluble species using a HisTrap 5 mL Chelating HP column (GE Healthcare, Life Sciences) on a Fast Protein Liquid Chromatography (FPLC) system (ÄKTApurifier, GE Healthcare, USA). Unbound species were washed out with 5 column volumes of buffer (50 mM phosphate buffer, pH 7.0, with 250 mM NaCl). A continuous gradient was run from 0–50% elution buffer (50 mM phosphate buffer, pH 7.0, with 250 mL NaCl and 1 M imidazole) over 6 column volumes. UV absorbance at 214 nm was monitored to detect eluting proteins and

outflow collected in 2.5 mL fractions. These fractions were analysed by SDS-PAGE; fractions found to contain species of the correct molecular weight for the fusion construct were pooled and dialysed against 50 mM phosphate buffer prior to cleavage/refolding.

3.5 Fusion construct cleavage

Cleavage was carried out with a 1:50 ratio of TEV protease:substrate in 50 mM phosphate buffer including oxidised and reduced glutathione to provide reducing power. A pilot cleavage trial tested levels ranging from 0.1 to 1.0 mM oxidised glutathione (GSSG) and 1 to 10 mM reduced glutathione (GSH). Subsequent routine cleavage made use of 0.1 mM GSSG/1 mM GSH.

3.6 Buffer refolding

H_Vc7.1 was incubated with 100 mM TCEP for 2 h to reduce disulfide bonds and ensure a homogeneous starting material for refolding trials. TCEP was removed by RP-HPLC and the peptide lyophilised. The peptide was resuspended in a small amount of MilliQ H_2O and placed in buffer solution (100 mM ammonium bicarbonate, pH 7.0, with 0.5 mM GSSG and 5.0 mM GSH or 100 mM HEPES, pH 7.0, with 0.5 mM GSSG and 5.0 mM GSH). Refolding reactions were stirred for 16 h at room temperature, with timepoints taken at 0, 2, 4, 8 and 16 h. Timepoint fractions were quenched with formic acid (10% volume).

3.7 On-column refolding

Refolding of the peptide while still incorporated in the fusion construct was trialled, based on the method of Chang et al.[32] Cell pellets were thawed and resuspended in Tris buffer (50 mM Tris-HCl, pH 8.0, with 250 mM NaCl and 10 mM imidazole). Cells were lysed by sonication on ice (10 x 30 s bursts) and centrifuged for 30 min at 18,000 rpm prior to being filtered through a sterile 0.2 μm membrane. An equal volume of solubilisation buffer (20 mM Tris-HCl, pH 8.0, with 250 mM NaCl, 10 mM imidazole, 7 M guanidine-HCl and 2 mM DTT) was added. Activated Ni-NTA resin was added to the mixture and left to mix on a roller for 2 h. Slurry was added to a wide-bore gravity-driven purification column and drained. A series of buffers with decreasing amounts of denaturants was applied to the column in sequence (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 10 mM imidazole, with 6 M urea/1 mM DTT; 5 M urea/0.8 mM DTT; 4 M urea/0.6 mM DTT; 3 M urea/0.4 mM DTT; 2 M urea/0.2 mM DTT; 1 M urea/0.1 mM DTT; 0 M urea/0 mM DTT).

Flowthrough was collected. Resin was eluted with 50 mM Tris-HCl, pH 8.0, with 250 mM NaCl and 1 M imidazole.

3.8 Purification

Samples were routinely purified by RP-HPLC on an Altima 5 μm C8 (250 x 22 mm) column with a linear or segmented gradient reaching 100% Buffer B over a variable time period (Buffer A: 0.1% TFA in MilliQ H_2O; Buffer B: 0.1% TFA in 80% acetonitrile).

4 Results

4.1 Fusion construct expression

Initial expression trials of BL21(DE3) cells transformed with the fusion construct using 1 L LB cultures resulted in ~2.5 g cell pellet yield per litre of culture. Halving the volume of culture to 500 mL increased the yield to ~5 g/L. This improvement was attributed to improved aeration of the media. To further increase the yield, the use of 4% glucose in expression media was trialled (media was exchanged after initial growth; induction occurred in the absence of glucose). In one case, use of 4% glucose resulted in a further improvement of yield to 14 g/L, but in other preparations this had little effect. The routine growth of expression culture was carried out in volumes of 500 mL without added glucose.

4.2 Cell lysis and purification of fusion construct

A polyhistidine tag was included in the fusion construct for use in Ni^{2+} affinity chromatography in the post-lysis purification step. Early attempts using gravity-driven Ni-NTA affinity columns were inconsistent with regard to the concentration of imidazole required to elute protein from the resin. This was improved by using an automated FPLC system with a HisTrap column. Typical yields of pure fusion protein were 47 mg per gram of cell pellet for H_Vc7.1 and 1.2 mg per gram of cell pellet for H_Vc7.2.

4.3 Peptide sequence verification

A digested sample of H_Vc7.1 was analysed by LC-MS to determine if the recombinant product conformed to the expected sequence. Results are presented in **Figure 5**; both of the two closely eluting peaks in the chromatogram (indicated by arrows in **Figure 5A**) were found to contain ions diagnostic for the presence of full-length oxidised H_Vc7.1. This confirmed that the construct encoded the correct peptide, while simultaneously indicating

that multiple disulfide isomers had been formed. A separate refolding step was therefore necessary in the workflow and trials were set up to establish the conditions required.

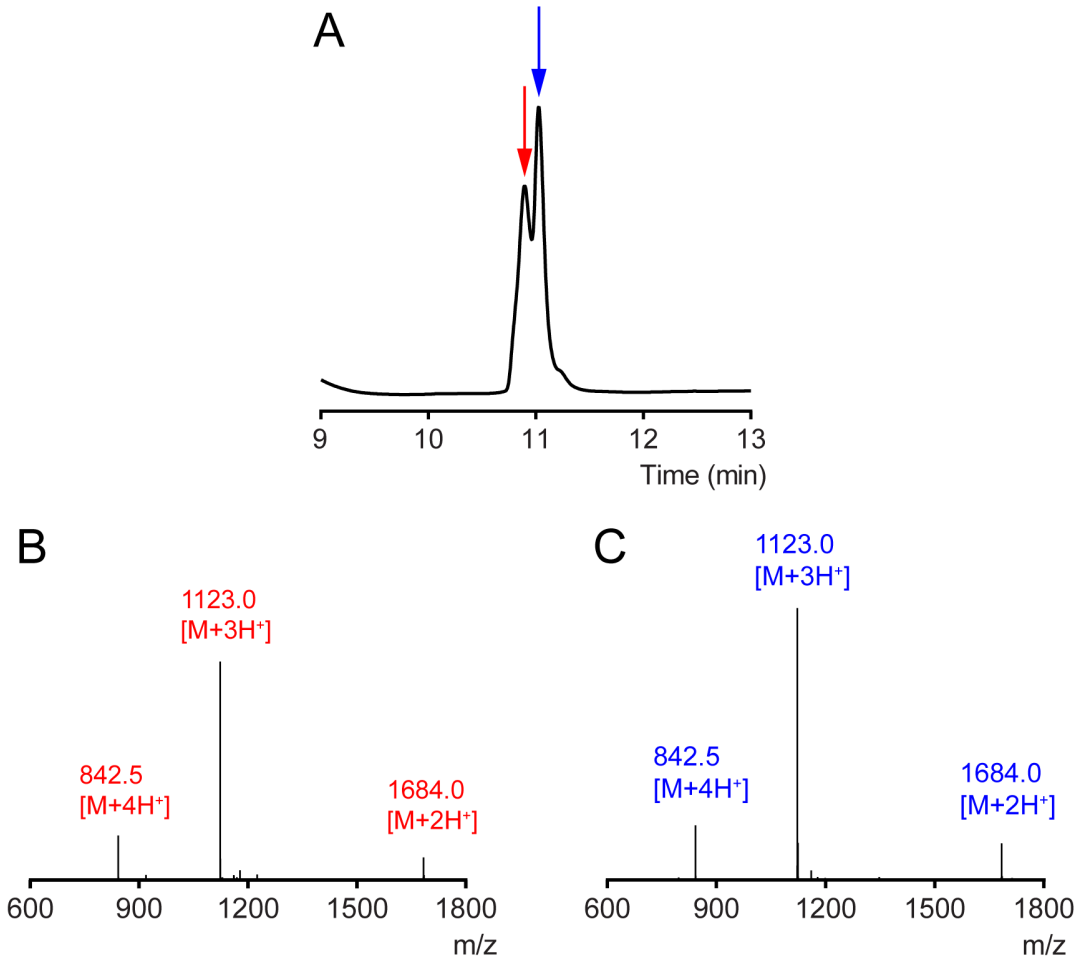


Figure 5. Analysis of cleaved and oxidised H_Vc7.1 by LC-MS. (A) Chromatogram trace; Buffer A: 0.05% TFA in MilliQ; Buffer B: 0.05% TFA in acetonitrile. Sample was run on a C8(2) column with a linear gradient of 0–60% B over 10 min. Peaks containing H_Vc7.1 are indicated by arrows. (B) Ions present in first peak (*red* arrow); diagnostic ions for H_Vc7.1 are labelled. (C) Ions present in second peak (*blue* arrow).

4.4 Detection of multiple folding isomers

To establish a baseline for the upcoming refolding trials, small samples of Trx-H_Vc7.1 and Trx-H_Vc7.2 were digested with TEV protease and run on analytical HPLC using a C18 column with a segmented linear gradient. In the absence of a dedicated refolding procedure a number of isomers were detected, presumed to be caused by various permutations of connections in the three disulfide bridges (**Figure 6**). Two dominant peaks were observed, along with multiple smaller contaminating peaks. The goal of the refolding step was to drive the majority of the product into a single peak, corresponding to the native fold.

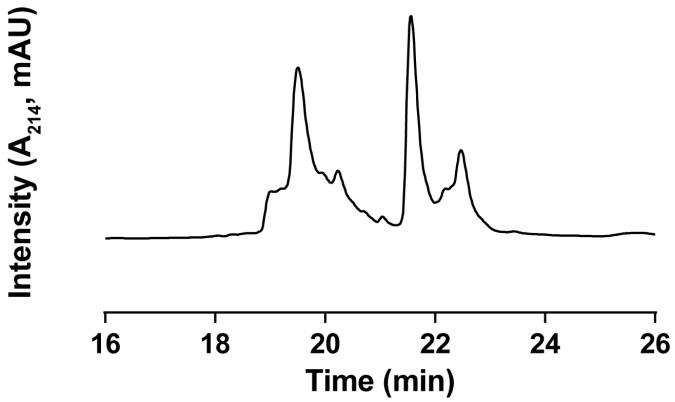


Figure 6. UV absorbance at 214 nm from analytical RP-HPLC of cleaved H_Vc7.1, showing multiple disulfide isomers present in a minimally processed sample. Sample was run on a C18 column at 5 mL/min. Buffer A: 0.1% TFA in MilliQ; Buffer B: 0.1% TFA in 80% acetonitrile, gradient: 5–60% B over 20 min.

4.5 Glutathione cleavage/refolding trials

A pilot cleavage trial was set up using mixtures of oxidised and reduced glutathione to determine the amount of reducing power required for TEV protease to cleave the fusion construct, with DTT as a positive control. The intent was to accomplish the liberation of the toxin from the fusion construct and the folding of the toxin simultaneously. Results for H_Vc7.1 are presented in **Figure 7**. The mass of H_Vc7.1 was 3.4 kDa, too small to be directly visualised on a SDS-PAGE gel. However, comparison of the bands for the fusion and the cleavage mixture in **Figure 7A** was consistent with a mass drop from 20 kDa to 17 kDa, as would be expected when liberating the toxin from the fusion. Samples subjected to overnight incubation showed evidence of degradation products; by contrast the results for 4 h incubation were clear and cleavage appeared to be complete. At high concentrations of glutathione analytical RP-HPLC revealed a number of peaks eluting close together, suggestive of a number of folding isomers. The trace for the lowest glutathione concentration tested showed a single peak dominating, albeit with minor peaks close by. A similar trial was conducted on H_Vc7.2; results in **Figure 8**. Cleavage of H_Vc7.2 was not complete, even after overnight incubation. The analytical RP-HPLC traces were cleaner than those from the trials on H_Vc7.1 and showed fewer peaks, although the best results were still obtained with the lowest concentration of glutathione.

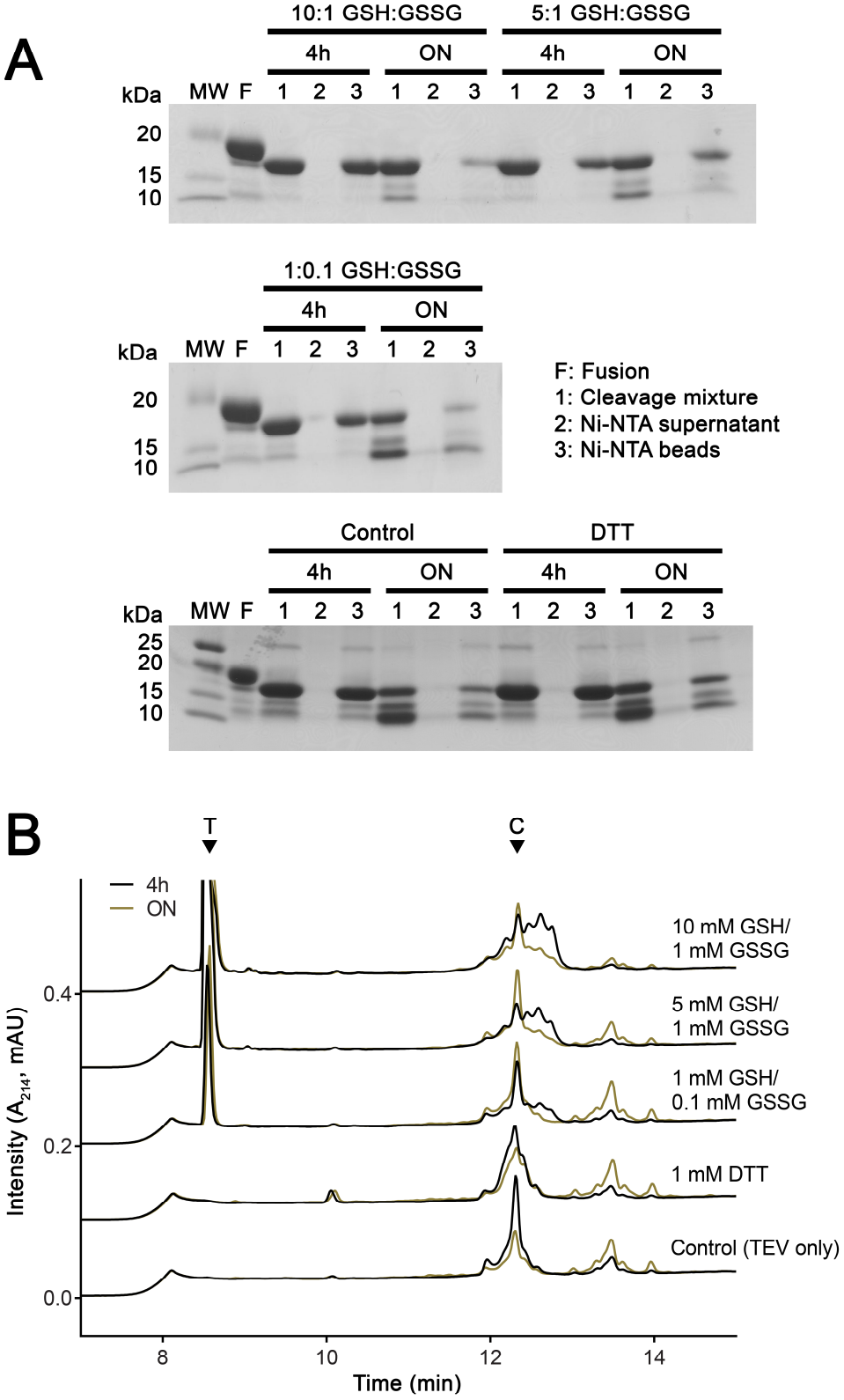


Figure 7. (A) SDS-PAGE analysis of cleavage efficiency for Trx-H_Vc7.1. (B) Analytical RP-HPLC of reaction products. Buffer A: 0.1% TFA in MilliQ H_2O; Buffer B: 0.1% TFA in 80% acetonitrile, gradient: 0–100% B over 20 min. T: thioredoxin, C: conotoxin. ON samples incubated 16 h.

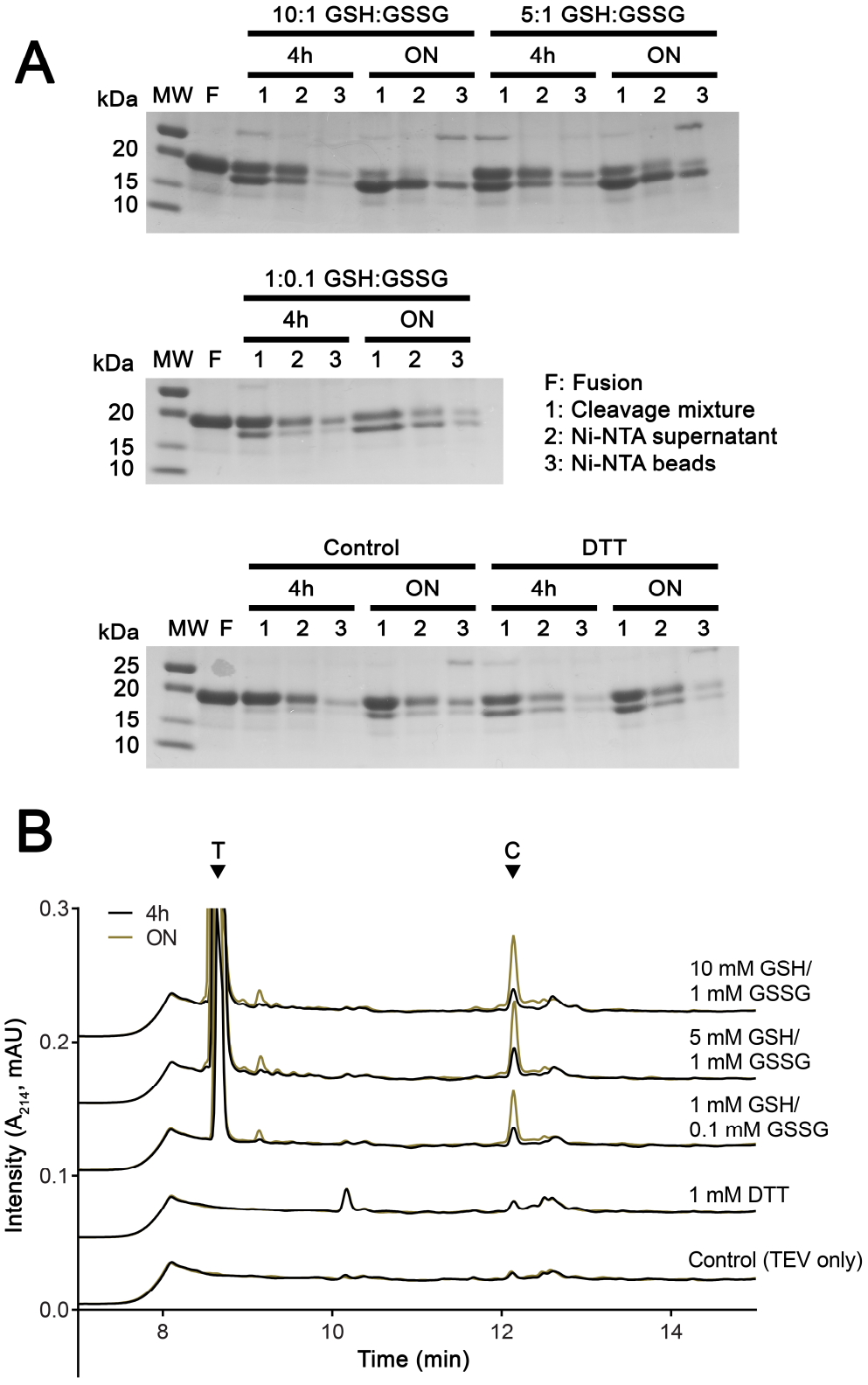


Figure 8. (A) SDS-PAGE analysis of cleavage efficiency for Trx-H_Vc7.2. (B) Analytical RP-HPLC of reaction products. Buffer A: 0.1% TFA in MilliQ H_2O; Buffer B: 0.1% TFA in 80% acetonitrile, gradient: 0–100% B over 20 min. T: thioredoxin, C: conotoxin. ON samples incubated 16 h.

4.6 Cleavage comparison trial

From the glutathione cleavage trial data, it appeared that Trx-H_Vc7.2 was cleaved with lower efficiency that Trx-H_Vc7.1. To ascertain whether this was indeed the case or an artefact of the previous experimental setup, a trial was performed in which both constructs were simultaneously cleaved under identical conditions (1:50 TEV protease:substrate, 1 mM reduced/0.1 mM oxidised glutathione, 4 h incubation at 37 °C).

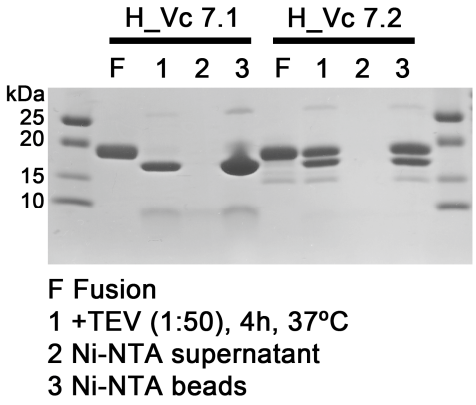


Figure 9. SDS-PAGE analysis of cleavage efficiency comparison for thioredoxin fusions of H-superfamily conotoxins H_Vc7.1 and H_Vc7.2.

As shown in **Figure 9**, the bands corresponding to H_Vc7.1 indicate that the fusion construct is completely digested and that the thioredoxin fusion partner is retained on the affinity resin by the polyhistidine tag. Meanwhile, the results for H_Vc7.2 clearly demonstrate that cleavage is incomplete and approximately 50% of the fusion construct remains resistant to the effect of the protease after 4 h. Thus, it is apparent that the thioredoxin fusion of H_Vc7.1 is cleaved more efficiently than that of H_Vc7.2.

4.7 Purification by RP-HPLC

Trial cleavage/refolding had identified conditions that drove the peptides towards a major conformation, although misfolded peaks were also present. The refolding mixtures were purified by RP-HPLC. However, fractions from these runs did not yield samples of sufficient purity (>95%) for NMR study (**Figure 10**). Gradients were modified to be shallower over the elution point of the peptide, but despite taking a very narrow cut of the peak the required purity was not achieved. The closely eluting impurities were disulfide shuffled analogues, misfolded isomers. Such similar species exhibit a very similar hydrophobicity profile and hence elute from the RP-HPLC column under almost the same conditions, resulting in very little separation from the desired product.

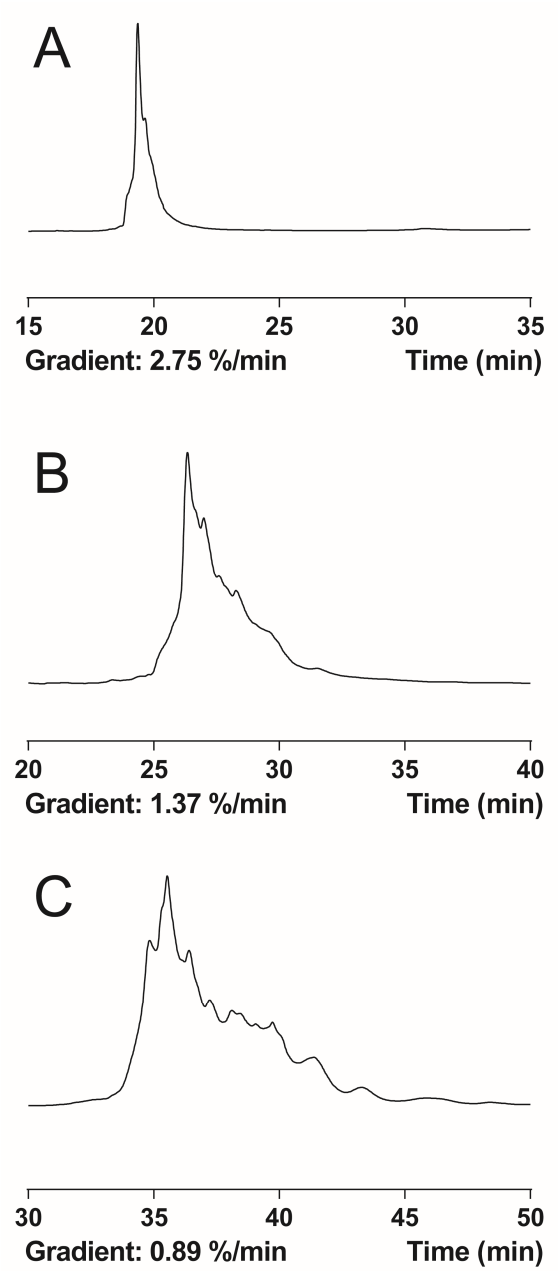


Figure 10. RP-HPLC traces of cleaved H_Vc7.1 samples run with differing gradients. Buffer A: 0.1% TFA in MilliQ H_2O; Buffer B: 0.1% TFA in 80% acetonitrile. Shallower gradients reveal the underlying complexity of the sample, but do not aid separation sufficiently to collect pure samples.

4.8 Refolding strategies

A common refolding strategy is to incubate the linear form of a peptide in a buffer solution with a certain amount of oxidant, such as glutathione. For the H-superfamily toxins, ammonium bicarbonate and HEPES buffer were trialled. As a preparatory step, cleaved peptide was treated with 100 mM TCEP to reduce any disulfide bonds that may have formed and ensure a homogenous starting material for refolding trials. Solutions of

50 mM ammonium bicarbonate and 50 mM HEPES were tested, both at pH 7.0 with 5 mM reduced / 0.5 mM oxidised glutathione. Peptides were incubated for 16 h (**Figure 11**).

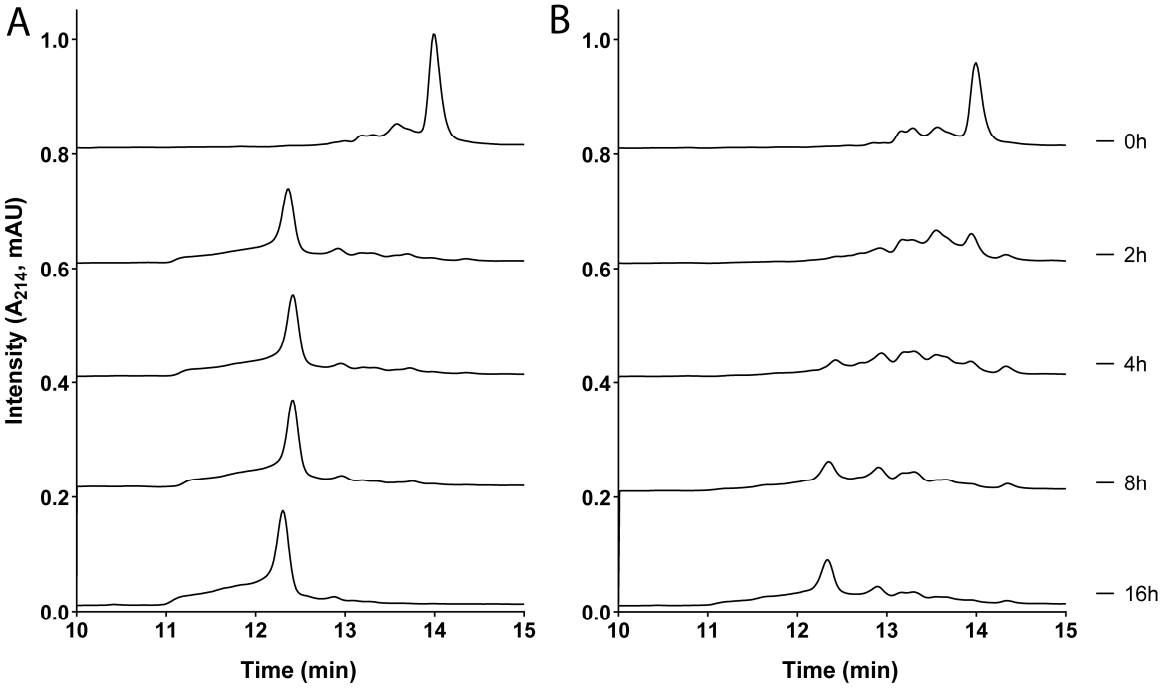


Figure 11. Results of refolding trial of H_Vc7.1. (A) Incubation in 50 mM ammonium bicarbonate, pH 7.0. (B) Incubation in 50 mM HEPES, pH 7.0.

Refolding was most efficient in ammonium bicarbonate, with the oxidised product first appearing after 2 h incubation and persisting for the length of the experiment. HEPES did not perform as well as ammonium bicarbonate, taking the full 16 h to begin forming a major oxidised product and retaining significant contaminating peaks. A change in retention time was observed, with oxidised product eluting from the column earlier than linear material. This is expected, as the folding process should bury hydrophobic moieties and shield them from solvent, reducing the hydrophobicity of the molecule and therefore its tendency to associate with the stationary phase.

4.9 On-column refolding

In another trial, the refolding strategy established for the peptide toxin ShK[32] was tested, to determine if it could be of use in the study of the H-superfamily conotoxins. In this approach, the fusion protein recovered from bacterial lysis is bound to nickel affinity resin in a gravity column and washed with a series of buffers containing diminishing amounts of reductant, starting at 6 M urea/1 mM DTT and working down. This gradually reduces the

denaturing power of the buffer and allows the peptide to adopt ever more energetically favourable conformations, slowly causing it to fold. The folded fusion is eluted from the resin with 1 M imidazole and dialysed into TEV cleavage buffer for liberation of the toxin. In the first instance, this procedure was followed using desalted TEV protease. However, this resulted in a lack of cleavage, so a second trial added 0.6 mM reduced glutathione/0.4 mM oxidised glutathione to the cleavage mixture. This did result in cleavage, but the fractions taken from RP-HPLC were no more pure than those generated by the ammonium bicarbonate refolding (**Figure 12**).

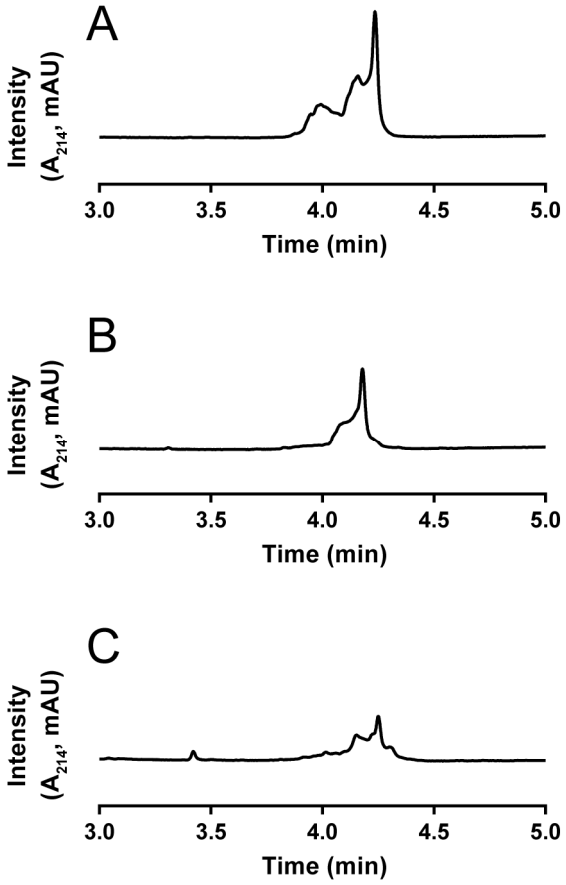


Figure 12. Comparison of fractions taken from refolding experiments on H_Vc7.1. Analytical HPLC traces run on C18 column at 0.4 mL/min. Buffer A: 0.1% TFA in MilliQ H_2O; Buffer B: 0.1% TFA in acetonitrile, gradient: 5–100% B over 9 min. (A) On-column gradient refolding with glutathione. (B) Ammonium bicarbonate buffer refolding. (C) HEPES buffer refolding.

In the end, no conditions could be found that adequately directed the recombinant toxins to the correct folded state. The complexity of the molecules, each with three disulfide bonds, frustrated attempts to gain a sample of sufficient purity for NMR analysis.

5 Discussion

5.1 Production of H-superfamily toxins

Researchers wishing to study peptide toxins have many decisions to make before they can start work. The primary decision considered in this chapter is which process to use to access the peptide sample required, but this decision is informed by many others. In the case of this project, it had already been decided that the H-superfamily conotoxins H_Vc7.1 and H_Vc7.2 would be studied by NMR spectroscopy, with the goal of determining their solution structures. This selection of analysis technique influenced the selection of production technique, as heavy isotope labelling was desired to facilitate acquisition of high-quality carbon and nitrogen chemical shift data in minimal time. Such labelling would be impossible when isolating toxins from the venom and expensive for samples produced by SPPS, but relatively straightforward for recombinant expression in *E. coli*. Regarding the selection of host cell, bacterial cells have the advantage of being easy to maintain. Their primary drawback – the lack of post-translational modification machinery – was not applicable, as the target sequences were devoid of such modifications. Thus recombinant expression in *E. coli* was selected as the method of production for the H-superfamily toxins.

Although cheaper by recombinant expression than by any other method, double-labelling of peptides still requires expensive reagents, namely the ^{13}C-glucose and ^{15}N-ammonium chloride used to supplement the minimal bacterial growth media. To avoid wasting resources, processes were trialled using unlabelled material to establish a production pipeline in the first instance. The recombinant bacterial host cells consistently produced the correct polypeptide chain, but not in a form that was correctly folded. The diversity of folding isomers present in each sample (caused by the incorrect connection of the six Cys residues in the peptide) presented a challenging purification problem. A series of attempts were made to reduce the complexity of the sample by adding a refolding step to the production pipeline, in the hope of driving the majority of the peptide to the native folded state.

5.2 Refolding of H-superfamily toxins

Several methods were tested in the quest for pure fractions of the H-superfamily toxins. These included refolding procedures performed before cleavage, concurrently with cleavage and after cleavage from the fusion construct. In each case, misfolded isomers were present as contaminating species in even the narrowest cut of the final RP-HPLC peak. The

initial trial involved cleaving the construct and folding the peptide at the same time. If successful, this would have been a substantial improvement in efficiency, eliminating an extra step from the production pipeline. However, the competing needs for an oxidising environment (to encourage the formation of disulfide bonds in the folding) and a reducing agent (to maintain the activity of TEV protease for the cleavage) made this impractical. The on-column gradient refolding method, in which the construct is kept intact until after folding, has been used to great effect in our laboratory for the sea anemone toxin ShK, but proved ineffective in a trial with H_Vc7.1. It is possible that there are kinetic traps on the energy landscape of the folding for the H-superfamily peptide; if there are conformations with incorrect disulfide connectivity that also possess energies near the minimum value of the native fold, there would be little opportunity to shuffle those incorrect bonds and arrive at the true minimum energy conformation with the correct disulfides. The buffer refolding in ammonium bicarbonate, performed after cleavage from the construct, was perhaps the most promising regime tested. Certainly the results were better than the equivalent procedure in HEPES buffer (**Figures 11, 12**). A more extensive buffer screen may have found an even better formulation that could drive more of the material into the native fold, reducing the appearance of disulfide isoforms and enabling purification to the level required for NMR studies (>95%).

As it is, each refolding condition trialled resulted in formation of a major peak, but with contaminating species that differed only in the connections between cysteine residues. Such disulfide isomers were very similar to the desired product and the changes in hydrophobicity that followed from the different disulfide connectivity were too small to sufficiently alter the elution time of the peptides in RP-HPLC, resulting in a lack of separation between the species and leading ultimately to impure fractions. While there are other chromatographic techniques that separate molecules based on other properties (such as size, charge and affinity binding), none of these properties was likely to be sufficiently different between the native fold and the disulfide isomers to effectively separate the species present in the refolding reactions.

5.3 Periplasmic expression

Rather than attempting to separate the correctly folded isoform from a mixture, it would be advantageous to reduce or eliminate the appearance of incorrect disulfide isomers in the first place, yielding the correct product from the outset. One strategy that has been of use in the production of disulfide-bonded peptides in *E. coli* cells is localisation of the

expressed fusion proteins to the cellular compartment known as the periplasm. The cytosol of the bacterial cell, where ribosomal translation takes place, is a reducing environment due to the action of the enzymes thioredoxin reductase and glutathione oxido-reductase.[46] These enzymes inhibit the formation of disulfide bonds, which is not ideal for the proper folding of a disulfide-bonded peptide. By contrast, the periplasm is an oxidising environment and is the location of the endogenous disulfide bond formation machinery of the Dsb family.[28]

DsbA is a strong thiol oxidant and is maintained in its oxidised form by the integral membrane protein DsbB. This pair of proteins work in tandem to form disulfide bonds in a peptide substrate. However, in multiply-disulfide-bonded peptides incorrect bonds may be formed by this system. A proofreading and repair function is provided by the other two proteins in the Dsb family, with the isomerase DsbC shuffling incorrect bonds while the integral membrane protein DsbD maintains its partner in the active conformation.[47] Recombinant proteins directed to the periplasm can take advantage of this oxidation machinery and potentially form the native disulfide bond network prior to cell lysis, removing the necessity for a separate refolding step.[48] To accomplish the targeting of a peptide to the periplasm, an N-terminal periplasmic signal sequence is required on the peptide, which is cleaved off after transport. Therefore, to take advantage of this system in the production of the H-superfamily toxins H_Vc7.1 and H_Vc7.2, it would have been necessary to re-engineer the expression plasmids used to transform the bacterial cells.

6 Conclusion

Difficulties with folding the triple-disulfide-bonded H-superfamily toxins resulted in samples of insufficient purity for the proposed biochemical investigation. Thus the structure was unable to be determined and no clues as to the function were gained. Focus turned to the single-disulfide toxin contryphan-Vc2, in the hopes that the simpler disulfide network may make this molecule more amenable to production. Work on this peptide is reported in the next chapter.

7 References

[1] McIntosh, M., Cruz, L. J., Hunkapiller, M. W., Gray, W. R., and Olivera, B. M. (1982) Isolation and structure of a peptide toxin from the marine snail *Conus magus*, *Arch. Biochem. Biophys. 218*, 329-334.

[2] Clark, C., Olivera, B. M., and Cruz, L. J. (1981) A toxin from the venom of the marine snail *Conus geographus* which acts on the vertebrate central nervous system, *Toxicon 19*, 691-699.

[3] Yanagawa, Y., Abe, T., Satake, M., Odani, S., Suzuki, J., and Ishikawa, K. (1988) A novel sodium channel inhibitor from *Conus geographus*: purification, structure, and pharmacological properties, *Biochemistry 27*, 6256-6262.

[4] Davis, J., Jones, A., and Lewis, R. J. (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS, *Peptides 30*, 1222-1227.

[5] Dutertre, S., Jin, A.-H., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V., Dutertre, V., Fry, B. G., Antunes, A., Venter, D. J., Alewood, P. F., and Lewis, R. J. (2014) Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails, *Nat. Commun. 5*, 3521.

[6] Jimenez, E. C., Olivera, B. M., Gray, W. R., and Cruz, L. J. (1996) Contryphan is a D-tryptophan-containing *Conus* peptide, *J. Biol. Chem. 271*, 28002-28005.

[7] Wurm, F. M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells, *Nat. Biotechnol. 22*, 1393-1398.

[8] Schmidt, F. R. (2004) Recombinant expression systems in the pharmaceutical industry, *Appl. Microbiol. Biotechnol. 65*, 363-372.

[9] Luckow, V. A., and Summers, M. D. (1988) Trends in the development of baculovirus expression vectors, *Nat. Biotechnol. 6*, 47-55.

[10] Verma, R., Boleti, E., and George, A. J. T. (1998) Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems, *J. Immunol. Methods 216*, 165-181.

[11] Ji, W., Zhang, X., Hu, H., Chen, J., Gao, Y., Liang, S., and An, C. (2005) Expression and purification of Huwentoxin-I in baculovirus system, *Protein Expr. Purif. 41*, 454-458.

[12] Kiyatkin, N. I., Kulikovskaya, I. M., Grishin, E. V., Beadle, D. J., and King, L. A. (1995) Functional characterization of black widow spider neurotoxins synthesised in insect cells, *Eur. J. Biochem. 230*, 854-859.

[13] Romanos, M. A., Scorer, C. A., and Clare, J. J. (1992) Foreign gene expression in yeast: a review, *Yeast 8*, 423-488.

[14] Pang, S.-Z., Oberhaus, S. M., Rasmussen, J. L., Knipple, D. C., Bloomquist, J. R., Dean, D. H., Bowman, K. D., and Sanford, J. C. (1992) Expression of a gene encoding a scorpion insectotoxin peptide in yeast, bacteria and plants, *Gene 116*, 165-172.

[15] Fitches, E. C., Bell, H. A., Powell, M. E., Back, E., Sargiotti, C., Weaver, R. J., and Gatehouse, J. A. (2010) Insecticidal activity of scorpion toxin (ButaIT) and snowdrop lectin (GNA) containing fusion proteins towards pest species of different orders, *Pest Manag. Sci. 66*, 74-83.

[16] Chen, J., Zhang, Y., Rong, M., Zhao, L., Jiang, L., Zhang, D., Wang, M., Xiao, Y., and Liang, S. (2009) Expression and characterization of jingzhaotoxin-34, a novel neurotoxin from the venom of the tarantula *Chilobrachys jingzhao*, *Peptides 30*, 1042-1048.

[17] Bruce, C., Fitches, E. C., Chougule, N., Bell, H. A., and Gatehouse, J. A. (2011) Recombinant conotoxin, TxVIA, produced in yeast has insecticidal activity, *Toxicon 58*, 93-100.

[18] Levandoski, M. M., Caffery, P. M., Rogowski, R. S., Lin, Y., Shi, Q.-L., and Hawrot, E. (2000) Recombinant expression of α-bungarotoxin in *Pichia pastoris* facilitates identification of mutant toxins engineered to recognize neuronal nicotinic acetylcholine receptors, *J. Neurochem. 74*, 1279-1289.

[19] Hellwig, S., Drossard, J., Twyman, R. M., and Fischer, R. (2004) Plant cell cultures for the production of recombinant proteins, *Nat. Biotechnol. 22*, 1415-1422.

[20] Commandeur, U., Twyman, R. M., and Fischer, R. (2003) The biosafety of molecular farming in plants, *AgBiotechNet 5*, 1-9.

[21] Twyman, R. M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R. (2003) Molecular farming in plants: host systems and expression technology, *Trends Biotechnol. 21*, 570-578.

[22] Ma, J. K. C., Drake, P. M. W., and Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants, *Nat. Rev. Genet. 4*, 794-805.

[23] Lai, L., Huang, T., Wang, Y., Liu, Y., Zhang, J., and Song, Y. (2009) The expression of analgesic-antitumor peptide (AGAP) from Chinese *Buthus martensii* Karsch in transgenic tobacco and tomato, *Mol. Biol. Rep. 36*, 1033-1039.

[24] Hernández-Campuzano, B., Suárez, R., Lina, L., Hernández, V., Villegas, E., Corzo, G., and Iturriaga, G. (2009) Expression of a spider venom peptide in transgenic tobacco confers insect resistance, *Toxicon 53*, 122-128.

[25] Khan, S. A., Zafar, Y., Briddon, R. W., Malik, K. A., and Mukhtar, Z. (2006) Spider venom toxin protects plants from insect attack, *Transgenic Res. 15*, 349-357.

[26] Yao, B., Fan, Y., Zeng, Q., and Zhao, R. (1996) Insect-resistant tobacco plants expressing insect-specific neurotoxin AaIT, *Chin. J. Biotechnol. 12*, 67-72.

[27] Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges, *Front. Microbiol. 5*, 172.

[28] Choi, J. H., and Lee, S. Y. (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*, *Appl. Microbiol. Biotechnol. 64*, 625-635.

[29] Berkmen, M. (2012) Production of disulfide-bonded proteins in *Escherichia coli*, *Protein Expr. Purif. 82*, 240-251.

[30] Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012) SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm, *Microb. Cell Fact. 11*, 753.

[31] Drevet, P., Lemaire, C., Gasparini, S., Zinn-Justin, S., Lajeunesse, E., Ducancel, F., Pinkasfeld, S., Courçon, M., Trémeau, O., Boulain, J. C., and Ménez, A. (1997) High-level production and isotope labeling of snake neurotoxins, disulfide-rich proteins, *Protein Expr. Purif. 10*, 293-300.

[32] Chang, S. C., Galea, C. A., Leung, E. W. W., Tajhya, R. B., Beeton, C., Pennington, M. W., and Norton, R. S. (2012) Expression and isotopic labelling of the potassium channel blocker ShK toxin as a thioredoxin fusion protein in bacteria, *Toxicon 60*, 840-850.

[33] Johnson, T. M., Quick, M. W., Sakai, T. T., and Krishna, N. R. (2000) Expression of functional recombinant scorpion β-neurotoxin Css II in *E. coli*, *Peptides 21*, 767-772.

[34] Che, N. Y., Wang, L. L., Gao, Y., and An, C. C. (2009) Soluble expression and one-step purification of a neurotoxin Huwentoxin-I in *Escherichia coli*, *Protein Expr. Purif. 65*, 154-159.

[35] Hernandez-Cuebas, L. M., and White, M. M. (2012) Expression of a biologically-active conotoxin PrIIIE in *Escherichia coli*, *Protein Expr. Purif. 82*, 6-10.
[36] Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnol. J. 7*, 620-634.
[37] Kent, S. B. (1988) Chemical synthesis of peptides and proteins, *Annu. Rev. Biochem. 57*, 957-989.
[38] Akondi, K. B., Muttenthaler, M., Dutertre, S., Kaas, Q., Craik, D. J., Lewis, R. J., and Alewood, P. F. (2014) Discovery, synthesis, and structure-activity relationships of conotoxins, *Chem. Rev. 114*, 5815-5847.
[39] Govindarajan, S., and Goldstein, R. A. (1998) On the thermodynamic hypothesis of protein folding, *Proc. Natl. Acad. Sci. U. S. A. 95*, 5545-5549.
[40] Wypych, J., Li, M., Guo, A., Zhang, Z., Martinez, T., Allen, M. J., Fodor, S., Kelner, D. N., Flynn, G. C., Liu, Y. D., Bondarenko, P. V., Ricci, M. S., Dillon, T. M., and Balland, A. (2008) Human IgG2 antibodies display disulfide-mediated structural isoforms, *J. Biol. Chem. 283*, 16194-16205.
[41] Bulaj, G. (2005) Formation of disulfide bonds in proteins and peptides, *Biotechnol. Adv. 23*, 87-92.
[42] de Marco, A. (2009) Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*, *Microb. Cell Fact. 8*, 26.
[43] Mamathambika, B. S., and Bardwell, J. C. (2008) Disulfide-linked protein folding pathways, *Annu. Rev. Cell Dev. Biol. 24*, 211-235.
[44] Jurado, P., de Lorenzo, V., and Fernández, L. A. (2006) Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of *Escherichia coli*: evidence that chaperone activity is the prime effect of thioredoxin, *J. Mol. Biol. 357*, 49-61.
[45] Kapust, R. B., Tozser, J., Copeland, T. D., and Waugh, D. S. (2002) The P1' specificity of tobacco etch virus protease, *Biochem. Biophys. Res. Commun. 294*, 949-955.
[46] Wegmuller, S., and Schmid, S. (2014) Recombinant peptide production in microbial cells, *Curr. Org. Chem. 18*, 1005-1019.
[47] Heras, B., Shouldice, S. R., Totsika, M., Scanlon, M. J., Schembri, M. A., and Martin, J. L. (2009) DSB proteins and bacterial pathogenicity, *Nat. Rev. Microbiol. 7*, 215-225.
[48] Klint, J. K., Senff, S., Saez, N. J., Seshadri, R., Lau, H. Y., Bende, N. S., Undheim, E. A. B., Rash, L. D., Mobli, M., and King, G. F. (2013) Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the periplasm of *E. coli*, *PLoS One 8*, e63865.

Chapter 3
Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue

1 Introduction

Production of the H-superfamily peptides from the *Conus victoriae* venom duct transcriptome proved intractable, with both peptides failing to fold to the native conformation following recombinant expression and purification (*Chapter 2: Accessing peptide toxin samples*). Focus therefore turned to another peptide selected from the venom duct transcriptome, contryphan-Vc2. This peptide was successfully produced by solid-phase peptide synthesis and subjected to biochemical characterisation, which included determination of the three-dimensional structure by nuclear magnetic resonance (NMR) spectroscopy. The results of this analysis are described in the following publication, *Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue*. The production of the peptide is detailed below, following a brief discussion of the underpinnings of NMR spectroscopy and its use in the determination of structures of biological macromolecules.

1.1 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is a technique for probing the chemical environment of a magnetically active nucleus, using the phenomenon of nuclear magnetic resonance. Atomic nuclei possess both the quantum property known as 'spin' and an electrical charge, which together result in the generation of a magnetic moment.[1] When an atomic nucleus of non-zero spin is placed in an external magnetic field, the spin aligns itself with the applied field in a parallel or anti-parallel fashion, with the parallel arrangement being a lower-energy state.[1] The spin does not completely align with the applied field; rather it displays precessional motion around the axis of the applied field, with the frequency of precession being termed the Larmor frequency.[2] When a precessing nucleus is excited by a photon with a frequency equal to the Larmor frequency of that nucleus (which is typically in the radio-wave portion of the electromagnetic spectrum), the nucleus absorbs that energy and flips to the other state.[3] This resonance absorption can be measured and forms the basis of NMR spectroscopy.

Many details of the magnetically active nucleus can be determined by NMR spectroscopy, but all results report on the overall average of the sample. Care must therefore be taken when multiple conformations are present in a sample.

1.1.1 Resonant frequency and electronic shielding

The frequency at which a nucleus resonates depends on a few factors, such as the composition of the nucleus (that is, the element and isotope) and the strength of the external magnetic field. If these were the only factors that affected the resonant frequency, the phenomenon would be of limited analytical use – all ^{1}H nuclei would resonate at the same frequency in a given instrument. Fortunately this is not the case, as the applied external magnetic field is not identical to the magnetic field experienced by the nucleus. Electrons around a nucleus also generate magnetic fields, which can 'shield' the nucleus from the full strength of the applied external magnetic field.[4] Chemical phenomena such as bonds and proximal electron-withdrawing groups affect the electron cloud and the degree of shielding afforded to each particular nucleus. This changes the strength of the magnetic field experienced by that nucleus and consequently alters the resonant frequency.[5] Thus NMR spectroscopy is sensitive to the chemical environment surrounding each resonating nucleus.

1.1.2 Chemical shift and referencing

NMR spectrometers come in a variety of different field strengths. As resonant frequency is dependent on field strength, measuring the same sample at different field strengths will yield different values for the resonant frequency. To address this issue and allow comparison of data between instruments, NMR results are reported as normalised resonant frequencies referred to as chemical shifts.

The chemical shift of a resonance is calculated by comparing the frequency of that resonance to the resonant frequency of a reference compound. This removes the field strength dependence of the measurement, converting the frequency (measured in Hz) to a chemical shift (usually presented in ppm).[3] The standard reference compound for ^{1}H nuclei is tetramethylsilane (TMS),[6] but this compound is insoluble in aqueous solution and therefore cannot be used directly for measuring chemical shifts in samples of proteins and peptides. Such samples make use of alternative referencing compounds such as 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), 3-(trimethylsilyl)-propionate (TSP), or 1,4-dioxane for measurement purposes.

1.1.3 Scalar coupling and dihedral angle restraints

Further magnetic resonance phenomena can be detected by NMR spectroscopy. Scalar coupling (or *J*-coupling) is a through-bond interaction that can be observed between two nuclei of non-zero spin. The spin of one nucleus perturbs the electrons involved in the

intervening bonds, which in turn perturbs the other coupled nucleus. This manifests as a 'splitting' of the resonance in the NMR spectrum. For two spin ½ nuclei, the peaks split into doublets, with the spacing between the doublet lines corresponding to the coupling constant.[1]

Scalar coupling values may be calculated between nuclei separated by varying numbers of bonds. The nomenclature is ${}^{n}J_{AB}$, where n is the number of bonds separating nuclei A and B. In protein NMR, vicinal ${}^{3}J_{HNH\alpha}$ coupling constants can be valuable restraints for structure calculation, as the magnitude of the coupling constant is related to the φ dihedral angle of the peptide bond. This relationship between ${}^{3}J$ coupling values and dihedral angle was described mathematically by Karplus based on experiments on ethylene compounds,[7] and has been validated for peptides many times since. Constraining the dihedral angle in this manner is very valuable for structure determination and can greatly increase the quality of the final structural ensemble.

1.1.4 Nuclear Overhauser effect (NOE) peaks and distance restraints

As well as revealing information on bonded nuclei, NMR can also report data on nuclei close in space but separated in sequence. The nuclear Overhauser effect (NOE) arises as a result of dipolar coupling between nuclei,[1] and results in a cross-peak on the recorded NOESY or ROESY spectra correlated to the chemical shifts of the two proximal nuclei. The effect lessens in intensity according to r^{-6}, where r is the distance between the interacting nuclei. Therefore the nuclear Overhauser effect is strictly a short-range interaction and can yield valuable information on which nuclei are spatially proximate in a structure. The strength of an NOE interaction is measured by the volume of the relevant cross-peak; signals are generally classified as strong, medium and weak. Each classification imposes an upper bound on the distance between the two interacting nuclei, with stronger peaks enforcing closer proximity.[8] As previously stated, NMR signals report on the average properties of the sample; it is assumed that a weak NOE peak results from a large distance between two nuclei in a single conformation at full occupancy, rather than a small distance in a minority conformation with low occupancy.

1.1.5 Structure calculation

To determine the structure of a biological macromolecule, such as a peptide or protein, NMR data such as the list of observed NOE peaks and ${}^{3}J_{HNH\alpha}$ coupling constants are converted into structural restraints. Structure calculation software is then used to generate

a series of structures that fulfil the given restraints, from which the 'best' structures are selected to form the structural ensemble. The more restraints that can be extracted from the data, the closer to reality the eventual model will be. The biggest danger is improperly assigned restraints, as these will distort the structure to accommodate an interaction that is actually not present. For this reason, assignments should be made conservatively, to preserve the integrity of the final structural ensemble.

2 Research rationale

2.1 Production of contryphan-Vc2 from *Conus victoriae*

Contryphan-Vc2 diverges from the consensus sequence that is a hallmark of the contryphan family, opening the possibility that the novel sequence may bestow novel structure or function (see the following publication). A high degree of post-translational modifications precluded production by recombinant expression and the toxin was not abundant in the venom, precluding direct isolation. The decision was therefore made to produce contryphan-Vc2 by solid-phase peptide synthesis (SPPS).

To study the effects of the Trp3 D-amino acid on the properties of the peptide, three analogues were prepared for testing. [D-Trp3]-contryphan-Vc2 was the native form, with Trp3 in the D-handed conformation. [L-Trp3]-contryphan-Vc2 was an all-L variant with no residues in the D-handed conformation. [W3A]-contryphan-Vc2 removed Trp3 entirely, replacing it with L-Ala.

3 Methods

3.1 Solid-phase peptide synthesis of contryphan-Vc2

Samples of contryphan-Vc2 containing D-Trp, L-Trp or L-Ala in position 3 were prepared by conventional 9-(fluorenyl)methoxycarbonyl (Fmoc) chemistry on Rink amide resin, as detailed in Section 2.1 of the following publication.

4 Results

4.1 Solid-phase peptide synthesis

Synthesis of [D-Trp3]-contryphan-Vc2 was carried out at 0.1 mmol scale; the peptide was lyophilised following cleavage from the resin. LC-MS of the crude peptide confirmed the presence of full-length reduced peptide ($[M+H^+]$ = 863.6 Da, $[M+2H^+]$ = 432.6 Da).

Ion masses were also detected corresponding to a +44 Da contaminating species, thought to be carboxylation of Trp3 (**Figure 1**). This contaminant was removed by RP-HPLC and the desired fractions from this purification re-lyophilised (**Figure 2**). This lyophilised material had a mass of 42.3 mg, representing 46% of the theoretical yield.

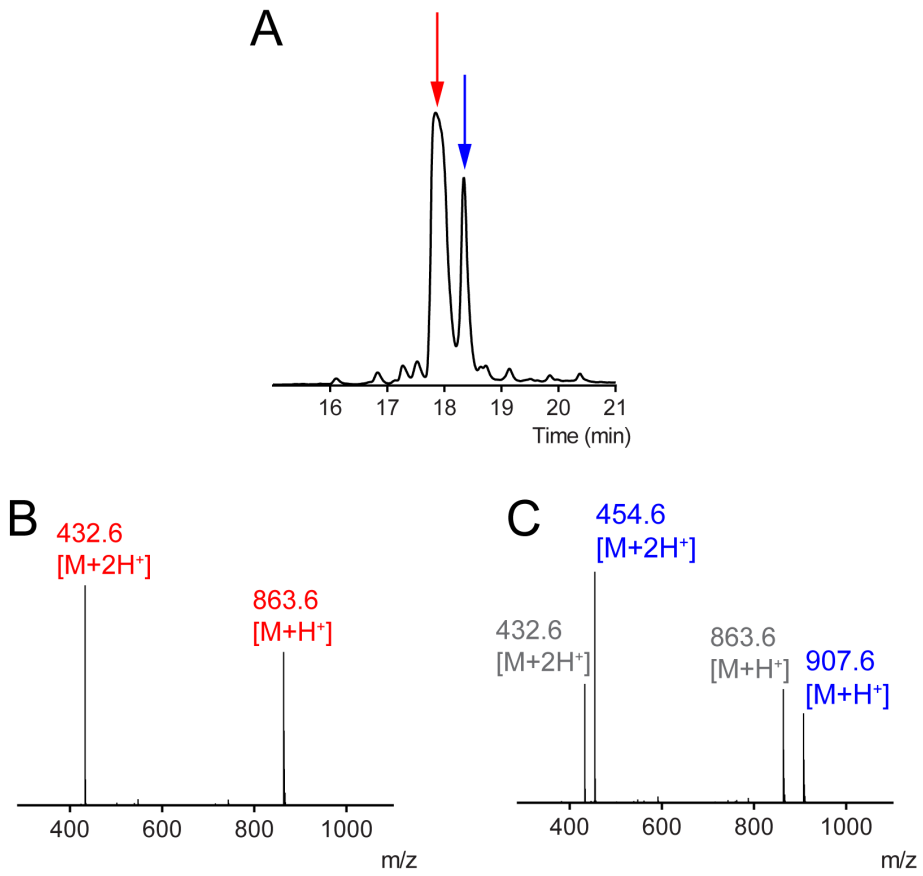


Figure 1. LC-MS analysis of crude synthetic [D-Trp3]-contryphan-Vc2 prior to purification and oxidation. (A) Chromatogram trace; Buffer A: 0.05% TFA in MilliQ H_2O; Buffer B: 0.05% TFA in acetonitrile. Sample was run on a C8(2) column with a linear gradient of 0–60% B over 10 min. Peaks of interest are indicated by arrows. (B) Ions present in first peak (*red* arrow); diagnostic ions for full-length reduced [D-Trp3]-contryphan-Vc2 are labelled. (C) Ions present in second peak (*blue* arrow); diagnostic ions for carboxylated [D-Trp3]-contryphan-Vc2 are highlighted.

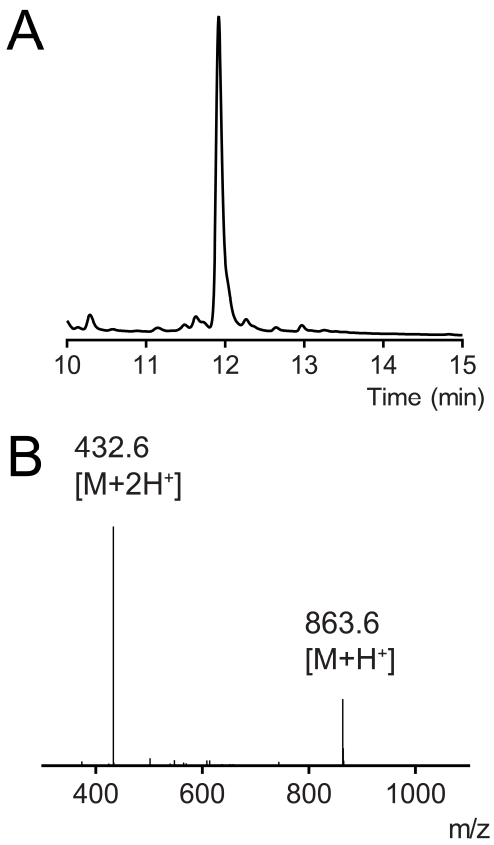


Figure 2. LC-MS analysis of crude synthetic [D-Trp3]-contryphan-Vc2 prior to oxidation. (A) Chromatogram trace; Buffer A: 0.05% TFA in MilliQ; Buffer B: 0.05% TFA in acetonitrile. Sample was run on a C8(2) column with a linear gradient of 0–60% B over 10 min. (B) Ions present in dominant peak; diagnostic ions for full-length reduced [D-Trp3]-contryphan-Vc2 are labelled.

4.2 Oxidative folding

The linear peptide was oxidised according to a standard protocol, via overnight incubation in 50 mM ammonium bicarbonate buffer, pH 8.0. The oxidant was atmospheric oxygen, with the solution left stirring open to the air. LC-MS analysis of the refolding mixture after 17 h incubation confirmed the formation of the disulfide bond through an ion mass loss of 2 Da ($[M+H^+]$ = 861.6 Da, $[M+2H^+]$ = 431.6 Da) (**Figure 3**).

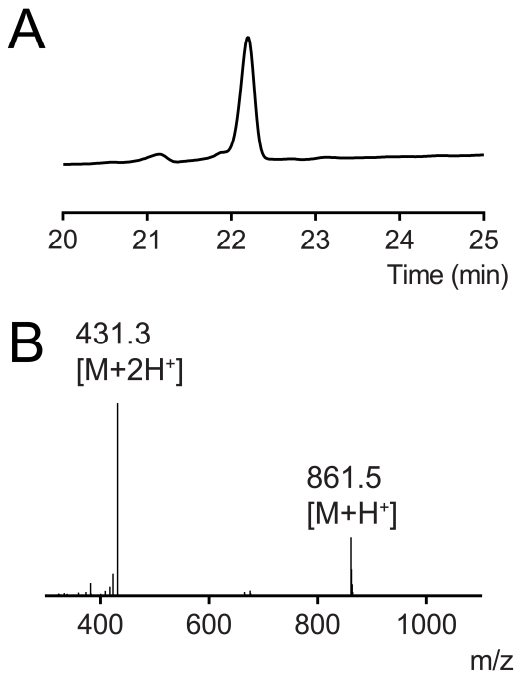


Figure 3. LC-MS analysis of oxidised synthetic [D-Trp3]-contryphan-Vc2. (A) Chromatogram trace; Buffer A: 0.05% TFA in MilliQ; Buffer B: 0.05% TFA in acetonitrile. Sample was run on a C8(2) column with a linear gradient of 0–60% B over 10 min. (B) Ions present in dominant peak; diagnostic ions for full-length oxidised [D-Trp3]-contryphan-Vc2 are labelled.

Preparative RP-HPLC using a C18 column was used to purify the refolding mixture. A single dominant peak was observed, along with a slight shoulder, the ion masses of which did not match any potential deletion product. Careful fractionation of this peak excluded the species creating the shoulder and resulted in a sample of sufficient purity for analysis. The results of the analyses undertaken are described in the following publication.

4.3 Creation of analogues

A sample of [L-Trp3]-contryphan-Vc2 was also prepared at 0.1 mmol scale, using the same techniques as for [D-Trp3]-contryphan-Vc2. LC-MS analysis of the crude cleavage mixture confirmed the presence of full-length reduced peptide ([M+H$^+$] = 863.6 Da, [M+2H$^+$] = 432.6 Da) and no major contaminating species (**Figure 4; A** and **B**). Crude yield before oxidation was 51.5 mg, 55% of theoretical. Oxidation was performed in the same manner as for [D-Trp3]-contryphan-Vc2 and produced the expected 2 Da ion mass loss (**Figure 4; C** and **D**). The oxidised material was then purified by RP-HPLC to >95% purity for analysis (**Figure 4; E** and **F**).

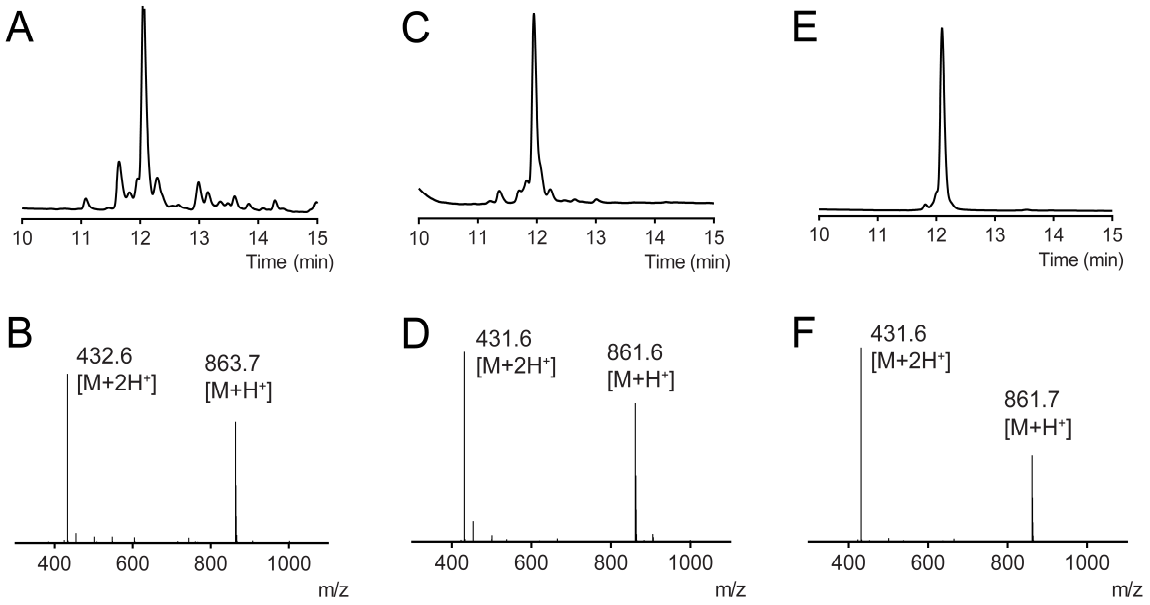


Figure 4. LC-MS analysis of synthetic [L-Trp3]-contryphan-Vc2 throughout production and purification. Chromatogram traces run with Buffer A: 0.05% TFA in MilliQ; Buffer B: 0.05% TFA in acetonitrile. Samples were run on a C8(2) column with a linear gradient of 0–60% B over 10 min. (A) Chromatogram trace of crude peptide, prior to oxidation. (B) Ions present in dominant peak of A; diagnostic ions for full-length reduced [L-Trp3]-contryphan-Vc2 are labelled. (C) Chromatogram trace of oxidation mixture. (D) Ions present in dominant peak of C; diagnostic ions for full-length oxidised [L-Trp3]-contryphan-Vc2 are labelled. (E) Chromatogram trace of purified sample. (F) Ions present in dominant peak of E; diagnostic ions for full-length oxidised [L-Trp3]-contryphan-Vc2 are labelled.

The final analogue prepared was [W3A]-contryphan-Vc2, an Ala mutant that replaced the enantiomeric Trp. The same synthetic procedure and purification regime was followed as before, resulting in a sample suitable for testing. Relevant LC-MS traces are presented as **Figure 5**.

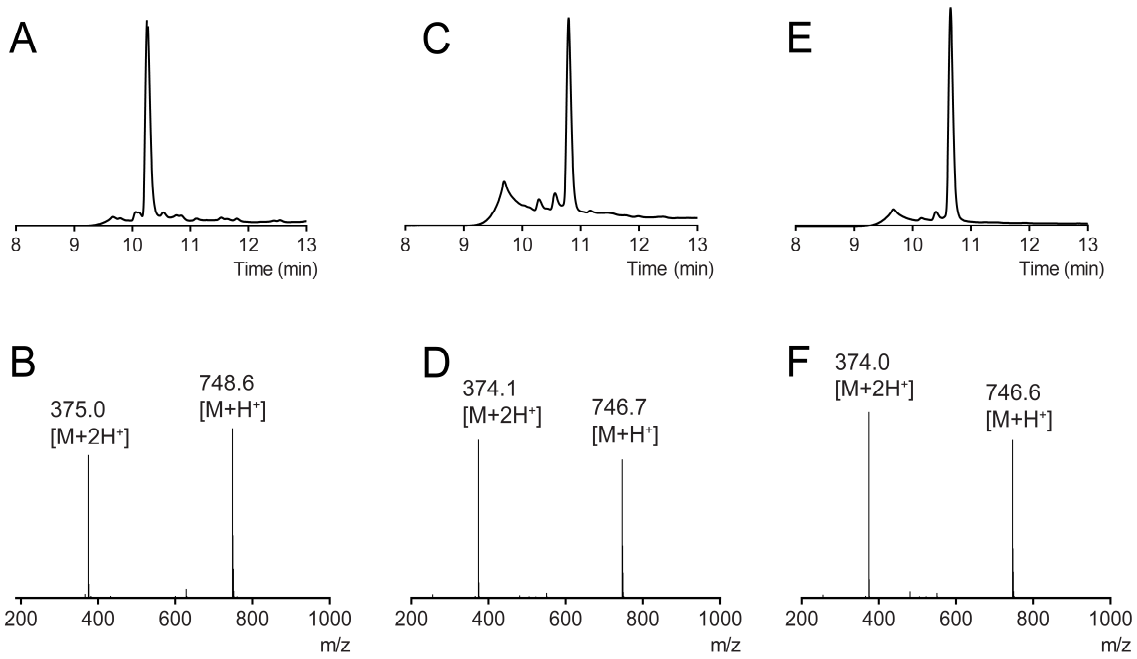


Figure 5. LC-MS analysis of synthetic [W3A]-contryphan-Vc2 throughout production and purification. Chromatogram traces run with Buffer A: 0.05% TFA in MilliQ; Buffer B: 0.05% TFA in acetonitrile. Samples were run on a C8(2) column with a linear gradient of 0–60% B over 10 min. (A) Chromatogram trace of crude peptide, prior to oxidation. (B) Ions present in dominant peak of A; diagnostic ions for full-length reduced [W3A]-contryphan-Vc2 are labelled. (C) Chromatogram trace of oxidation mixture. (D) Ions present in dominant peak of C; diagnostic ions for full-length oxidised [W3A]-contryphan-Vc2 are labelled. (E) Chromatogram trace of purified sample. (F) Ions present in dominant peak of E; diagnostic ions for full-length oxidised [W3A]-contryphan-Vc2 are labelled.

5 Discussion

The SPPS approach was successful in the production of contryphan-Vc2, allowing the easy production of analogues by substituting a different Fmoc-protected amino acid at the relevant coupling step. This allowed production of [D-Trp3]-, [L-Trp3]- and [W3A]-contryphan-Vc2 by varying the amino acid coupled in position 3.

All three analogues were synthesised, oxidised and purified by the same protocol, which gave consistent results. The crude product for [D-Trp3]-contryphan-Vc2 contained a +44 Da impurity, which was identified as carboxylation of Trp3. This modification can arise as a result of the stepwise removal of the protecting group on the indole sidechain.[9] The first step in this process involves removal of a *tert*-butyl component by TFA and proceeds rapidly. This leaves a carboxyl moiety, which is normally removed by the action of weak acids at a slower rate. In the case of [D-Trp3]-contryphan-Vc2 this second step did not proceed to completion, resulting in the impurity. Fractionation by RP-HPLC prior to

oxidation excluded this impurity and allowed the correct product to be isolated. The synthesis of [L-Trp3]- and [W3A]-contryphan-Vc2 proceeded without the need for this extra purification step. Oxidation was carried out in ammonium bicarbonate buffer on a dilute solution of peptide (0.5 mg/mL) to favour formation of the intramolecular disulfide bond. If the concentration of peptide was too high in the buffer, the chance of two separate molecules interacting and dimerising would have increased. Oxidation was complete after 17 h incubation and LC-MS analysis indicated that the correctly cyclised species were indeed formed (as opposed to dimers). The presence of a single cysteine pair eliminated the potential for misfolded species – either the correct bond was formed or the peptide remained linear.

6 Conclusion

The successful production of samples of contryphan-Vc2 was the necessary first step in a biochemical investigation of the structure and function of this peptide. The investigation undertaken is documented in the following publication, *Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue*.

7 References

[1] Evans, J. N. (1995) *Biomolecular NMR spectroscopy*, Oxford University Press, New York.
[2] Dixon, R. L., and Ekstrand, K. E. (1982) The physics of proton NMR, *Med. Phys. 9*, 807-818.
[3] Keeler, J. (2011) *Understanding NMR spectroscopy*, John Wiley & Sons.
[4] Ramsey, N. F. (1950) Magnetic shielding of nuclei in molecules, *Phys. Rev. 78*, 699-703.
[5] Marion, D. (2013) An introduction to biological NMR spectroscopy, *Mol. Cell. Proteomics 12*, 3006-3025.
[6] Harris, R. K., Becker, E. D., de Menezes, S. M. C., Goodfellow, R., and Granger, P. (2001) NMR nomenclature. Nuclear spin properties and conventions for chemical shifts (IUPAC Recommendations 2001), *Pure Appl. Chem. 73*, 1795-1818.
[7] Karplus, M. (1963) Vicinal proton coupling in nuclear magnetic resonance, *J. Am. Chem. Soc. 85*, 2870-2871.
[8] Clore, G. M., and Gronenborn, A. M. (1987) Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy, *Protein Eng., Des. Sel. 1*, 275-288.
[9] Yang, Y. (2016) Chapter 3 - Peptide global deprotection/scavenger-induced side reactions, In *Side Reactions in Peptide Synthesis*, pp 43-75, Academic Press, Oxford.

Toxicon 129 (2017) 113–122

Contents lists available at ScienceDirect

Toxicon

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

Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue

Stephen B. Drane [a], Samuel D. Robinson [a], Christopher A. MacRaild [a], Sandeep Chhabra [a], Balasubramanyam Chittoor [a], Rodrigo A.V. Morales [a], Eleanor W.W. Leung [a], Alessia Belgi [b], Samuel S. Espino [c], Baldomero M. Olivera [c], Andrea J. Robinson [b], David K. Chalmers [a], Raymond S. Norton [a, *]

[a] Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Victoria, Australia
[b] School of Chemistry, Monash University, Clayton 3800, Victoria, Australia
[c] Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

ARTICLE INFO

Article history:
Received 13 October 2016
Received in revised form
11 January 2017
Accepted 16 February 2017
Available online 17 February 2017

Keywords:
Conotoxin
Contryphan-Vc2
Nuclear magnetic resonance
Solution structure
Molecular dynamics
D-amino acid

ABSTRACT

In natural proteins and peptides, amino acids exist almost invariably as L-isomers. There are, however, several examples of naturally-occurring peptides containing D-amino acids. In this study we investigated the role of a naturally-occurring D-amino acid in a small peptide identified in the transcriptome of a marine cone snail. This peptide belongs to a family of peptides known as contryphans, all of which contain a single D-amino acid residue. The solution structure of this peptide was solved by NMR, but further investigations with molecular dynamics simulations suggest that its solution behaviour may be more dynamic than suggested by the NMR ensemble. Functional tests in mice uncovered a novel bioactivity, a depressive phenotype that contrasts with the hyperactive phenotypes typically induced by contryphans. Trp3 is important for bioactivity, but this role is independent of the chirality at this position. The D-chirality of Trp3 in this peptide was found to be protective against enzymatic degradation. Analysis by NMR and molecular dynamics simulations indicated an interaction of Trp3 with lipid membranes, suggesting the possibility of a membrane-mediated mechanism of action for this peptide.

1. Introduction

The contryphans are a family of small disulfide-cyclised peptides found in the venoms of marine cone snails of the genus *Conus*. All contryphans investigated to date produce strong behavioural effects when administered by intracranial injection in mice (Jacobsen et al., 1999; Jimenez et al., 1997, 1996, 2001). The molecular basis for this activity is unknown, although target receptors for four contryphans have been proposed: voltage-gated Ca^{2+} channels (contryphan-Lo and contryphan-Am) (Sabareesh et al., 2006), L-type Ca^{2+} channels (glacontryphan-M) (Hansson et al., 2004) and Ca^{2+}-dependent K^{+} channels (contryphan-Vn) (Massilia et al., 2003). Contryphans range in length from 7 to 11 residues and in most cases share the consensus sequence CO(^{D}W)XPWC. A sequence alignment of the known contryphans that have been investigated at the protein level is presented in Table 1. Consensus features include two Pro residues at positions 2 and 5 of the intercystine loop, with the first proline usually modified to hydroxyproline. A C-terminal Pro-Trp-Cys tripeptide is also present in all family members characterised to date, except contryphan-Tx (in which Trp is replaced by Tyr) (Jimenez et al., 2001). C-terminal amidation is almost ubiquitous, being absent only in Leu-contryphan-P (Jacobsen et al., 1999). Several members also have

Abbreviations: BMRB, Biological Magnetic Resonance Data Bank; BSA, bovine serum albumin; DIPEA, *N,N*-diisopropylethylamine; DMF, *N,N*-dimethylformamide; DPC, dodecylphosphocholine; Hyp, hydroxyproline; LC-MS, liquid chromatography/mass spectroscopy; MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PDB, Protein Data Bank; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPG, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol; ROESY, rotating-frame Overhauser effect spectroscopy; RP-HPLC, reversed-phase high performance liquid chromatography; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy. The abbreviations for the common amino acids (L-isomers unless indicated otherwise) are in accordance with the recommendations of the IUPAC-IUB Joint Commission on Biochemical Nomenclature (*Eur. J. Biochem.* 1984, 138:9-37)..

* Corresponding author. Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville, Victoria 3052, Australia.
E-mail address: Ray.Norton@monash.edu (R.S. Norton).

http://dx.doi.org/10.1016/j.toxicon.2017.02.012
0041-0101/

Table 1
Sequence alignment of members of the contryphan family. W = D-tryptophan, L = D-leucine, W = L-6-bromo-tryptophan, O = hydroxyproline, γ = gamma-carboxyglutamic acid, * = C-terminal amidation. Adapted from (Thakur and Balaram, 2007).

Designation	Sequence	Ref
Contryphan-Vc2	----CRWTPVC*	(Robinson et al., 2014)
Leu-contryphan-P	---GCVLLPWC*	(Jacobsen et al., 1998)
Leu-contryphan-Tx	----CVLYPWC*	(Jimenez et al., 2001)
Contryphan-In (In936)	---GCVLYPWC*	(Thakur and Balaram, 2007)
Glacontryphan-M	NγSγCPWHPWC*	(Hansson et al., 2004)
Contryphan-Vn	--GDCPWKPWC*	(Massilia et al., 2001)
Contryphan-Lo (Lo959)	---GCPWDPWC*	(Sabareesh et al., 2006)
Contryphan-Tx	---GCOWQPYC*	(Jimenez et al., 2001)
Contryphan-Sm	---GCOWQPWC*	(Jacobsen et al., 1998)
Contryphan-P	---GCOWDPWC*	(Jacobsen et al., 1998)
Contryphan-R	---GCOWEPWC*	(Jimenez et al., 1996)
Des(Gly1) contryphan-R	----COWEPWC*	(Jimenez et al., 1996)
Bromocontryphan-R	---GCOWEPWC*	(Jimenez et al., 1997)
Contryphan-Am (Am975)	---GCOWDPWC*	(Sabareesh et al., 2006)
Contryphan-fib	---GCOWMPWC*[a]	(Rajesh, 2015)
Unnamed[b]	-VVGCOWQPWC*	(Thakur and Balaram, 2007)

[a] Detected by mass spectrometry; isomerism of residue 3 not investigated.
[b] Detected in venom of *C. zeylanicus*, *C. betulinus* and *C. figulinus*.

N-terminal extensions to the consensus motif.

One noteworthy feature of contryphans is the presence of a D-amino acid at position 3 of the intercystine loop, which is either Trp or (rarely) Leu (Hansson et al., 2004; Jacobsen et al., 1998, 1999; Jimenez et al., 1997, 1996, 2001; Massilia et al., 2003; Sabareesh et al., 2006). The role of this D-amino acid in contryphans has not been explored in depth. The inclusion of a D-amino acid residue in a natural peptide is uncommon, but not unknown; in 2009, just over 30 examples of D-amino acid-containing peptides in animals were reported, of which nine were contryphans (Bai et al., 2009). In other peptides, D-amino acids have a range of effects. The first D-amino acid-containing peptide to be discovered in vertebrates was dermorphin, which was isolated from the skin of the frog *Phyllomedusa sauvagei* and contains D-Ala in position 2 (Montecucchi et al., 1981). The D-residue was found to be crucial for the potent opiate-like activity, which was lacking in a synthetic all-L analogue (Montecucchi et al., 1981). In the P-type Ca^{2+} channel antagonist ω-agatoxin IVB, found in spider venom, D-Ser46 conferred protection against the proteolytic effect of carboxypeptidase P (Heck et al., 1994). The excitotoxic conopeptide ι-RXIA contains D-Phe, and the analogue with L-Phe, ι-RXIA[L-Phe44], had a two-fold lower affinity and two-fold faster off rate than ι-RXIA on $Na_V1.6$ channels. In addition, the L-analogue was inactive at $Na_V1.2$ channels (Fiedler et al., 2008).

In this study, we investigate a newly-discovered contryphan, contryphan-Vc2, identified in the transcriptome of *Conus victoriae* (Robinson et al., 2014). Mouse bioassays were used to define a novel bioactivity, which is distinct from that seen previously for other contryphans. We assessed the effect of epimerising Trp3 on the activity, solution structure, proteolytic stability and lipid binding of this peptide. Homonuclear 2D ^{1}H NMR was used to calculate a solution structure, and molecular dynamics (MD) simulations were used to reveal further detail of the behaviour of the peptide in solution. Both NMR and MD indicated an interaction with lipid membranes.

2. Materials and methods

2.1. Chemical synthesis

Contryphan-Vc2 was identified in the venom gland transcriptome of *C. victoriae* as reported previously (Robinson et al., 2014). Contryphan-Vc2 peptides containing D-Trp, L-Trp or L-Ala in position 3 were prepared by conventional *N*-(9-fluorenyl)methyloxycarbonyl (Fmoc) chemistry on Rink amide resin at 0.1 mmol scale. Briefly, deprotection was performed in 20% piperidine (in DMF), followed by activation and elongation with 70 mL/L DIPEA (in DMF) and 3 equivalents of HCTU with Fmoc-protected amino acid for 50 min. Cleavage from the resin was performed over 2 h with a mixture of trifluoroacetic acid, triisopropylsilane, 1,3-dimethoxybenzene and 3,6-dioxa-1,8-octanedithiol (TFA:-TIPS:DMB:DODT, 92.5:2.5:2.5:2.5 by volume). The cleavage mixture was evaporated and precipitated with ice-cold diethyl ether. The crude product was lyophilised and stored at −20 °C until further purification.

Disulfide formation was achieved by stirring ~0.5 mg/mL crude peptide in 0.1 M ammonium bicarbonate (pH 8.0) for 17 h at room temperature. The cyclised peptides were purified on a Vydac 10 μm C18 (250 × 10 mm) column using a gradient of 40–70% buffer B over 30 min (buffer A: 0.1% TFA in MilliQ water; buffer B: 0.1% TFA in 80% acetonitrile). Sample purity was assessed by LC-MS to be greater than 95%. Further samples of synthetic [D-Trp3]- and [L-Trp3]-contryphan-Vc2 were purchased from Purar Chemicals (Victoria, Australia) and used for the proteolysis and DPC assays.

2.2. Proteolysis assays

Peptide resistance to proteolysis by trypsin, α-chymotrypsin and pepsin was measured by incubating a 250:1 peptide:enzyme mixture at 37 °C for 4 h. BSA was used as a positive control of protease activity. Stock solutions of trypsin and α-chymotrypsin were prepared in 1 mM HCl/2 mM $CaCl_2$ and reactions were run in 50 mM Tris, 100 mM NaCl (pH 7.4). Pepsin was prepared in 10 mM HCl and reactions run in 10 mM acetic acid/10 mM HCl (pH 2.0). Trypsin and α-chymotrypsin reactions were halted using 2.5% volume of acetic acid solution (25% v/v), and pepsin reactions were halted using 2.5% volume of 200 mM glycine-NaOH buffer (pH 11.4). The extent of digestion was analysed by LC-MS using a Jupiter 5 μm C4 300 Å column (50 × 2.0 mm) (buffer A: 0.1% formic acid in MilliQ water; buffer B: 0.1% formic acid in acetonitrile). Samples were eluted with a gradient of 0–60% B over 10 min.

2.3. Behavioural assay

Swiss Webster mice (15–21 days old; 6.6–10.1 g) were injected intracranially with different doses of synthetic peptides dissolved in 10 μL 0.9% NaCl, as described previously (McIntosh et al., 1994). Control mice were injected with 10 μL 0.9% NaCl solution. Following intracranial injection, mouse behaviour was observed for 2 h to determine differences between treated and control animals. All experiments involving the use of animals were approved by the Institutional Animal Care and Use Committee of the University of Utah.

2.4. Nuclear magnetic resonance spectroscopy

NMR spectra were acquired using a Bruker Avance III 600 MHz instrument. Lyophilised [D-Trp3]-contryphan-Vc2 was dissolved in 93% H_2O/7% ^2H_2O at pH 4.0 to a concentration of 2 mM and 300 μL samples were placed in Shigemi tubes. TOCSY and ROESY experiments were recorded at 5 °C using mixing times of 80 and 350 ms, respectively. An additional ROESY experiment was recorded with a mixing time of 50 ms to assist in χ_1 angle determination. 1D ^{1}H experiments were recorded at temperatures ranging from 10 to 30 °C in 5 °C steps to calculate amide proton temperature coefficients, and at 37 °C to test stability under physiological conditions. Hydrogen-deuterium exchange rates were measured by

dissolving lyophilised [D-Trp3]-contryphan-Vc2 in 100% ^2H_2O and recording 1D ^1H spectra at 5 min intervals at 10 °C. Spectra were referenced using dioxane at 3.75 ppm. ^1H-^{13}C HSQC and ^1H-^{15}N HMQC spectra were collected at natural abundance for assignment of heavy atom chemical shifts, and were referenced indirectly using resonance frequency ratios (Wishart et al., 1995). [L-Trp3]-Contryphan-Vc2 was dissolved in 93% H_2O/7% ^2H_2O at pH 4.1 to a concentration of 3 mM, and TOCSY and ROESY spectra were collected on a 550 μL sample in a standard 5 mm NMR tube at 5 °C. Titration with DPC micelles was carried out by dissolving lyophilised [D-Trp3]- and [L-Trp3]-contryphan-Vc2 separately in 100 mM deuterated acetate buffer, pH 4.4 to a concentration of 0.5 mM. 1D ^1H NMR spectra were acquired at 5 °C both before and after addition of deuterated dodecylphosphocholine (DPC, Sigma-Aldrich), in concentrations of 1, 5, 10, 20 and 40 mM. Peaks were monitored for shifting and broadening. TopSpin (v3.2, Bruker, USA) was used for spectral processing and analysis.

2.5. Structure calculation

Distance constraints for structure determination were derived from intensities of NOE cross-peaks in the ROESY spectrum of [D-Trp3]-contryphan-Vc2 recorded at 5 °C and pH 4.0. Dihedral angle restraints were obtained by measuring the $^3J_{HNHA}$ scalar coupling constants at 5 °C, noting the magnitude of the splitting between doublets in amide resonances. Values below 6.0 Hz were taken to indicate α-helical conformation ($-90° < \varphi < -40°$), while values above 8.0 Hz indicated β-strand conformation ($-160° < \varphi < -80°$). Amide proton temperature coefficients were calculated by recording 1D ^1H spectra over the range 10–30 °C, and plotting the chemical shift values for each residue. Residues with coefficients more positive than −4.5 ppb/K were flagged as potential hydrogen bond donors (Baxter and Williamson, 1997). χ_1 angles were investigated by matching $^3J\alpha\beta$ coupling constants and NOE intensities from a short mixing time (50 ms) NOESY spectrum to the patterns identified by Wagner et al. (1987), but no restraints were generated. The possibility of chemical shift-based restraints was investigated by submitting data to the TALOS-N web server (available at http://spin.niddk.nih.gov/bax/nmrserver/talosn/), but no useful constraints were returned. NMR solution structures were calculated using Cyana v3.0 and refined with XPLOR-NIH v2.40 based on distance restraints (from NOE intensities) and φ angle restraints (from $^3J_{HNHA}$ coupling constants).

2.6. Molecular dynamics (MD) simulations

Molecular dynamics simulations were carried out using GROMACS version 5.0.4 with the GROMOS 54a7 united-atom forcefield (modified to allow inclusion of D-Trp). Simulations used a 2 fs time step. Temperature coupling made use of the velocity rescale algorithm with a reference temperature of 298 K. Pressure coupling used the Berendsen or Parinello-Rahman algorithms, with reference pressure of 1 bar and compressibility of 4.5×10^{-5} bar^{-1}.

Starting models of [D-Trp3]-, [L-Trp3]- and [W3A]-contryphan-Vc2 were built from the NMR-derived solution structure of [D-Trp3]-contryphan-Vc2 with the proline in the *trans* conformation. For membrane simulations, the model system was constructed by placing an appropriate number of lipid molecules in the simulation box and solvating with SPC water. A steepest-descent minimisation of 2000 steps was used to remove bad van der Waals contacts between atoms, followed by a temperature equilibration (without pressure coupling) for 10,000 steps. Semiisotropic pressure coupling was applied using the Berendsen barostat for 500,000 steps. The simulation was then run for 200 ns using the Parinello-Rahman barostat to allow the lipid molecules to relax. Following this equilibration procedure, the peptide was introduced to the system either in the aqueous phase or buried in the bilayer, with the disulfide bond in the $+z$ direction. The minimisation and equilibration steps were then repeated, with the peptide coordinates restrained. The simulation production runs for [D-Trp3]-, [L-Trp3]- and [W3A]-contryphan-Vc2 were executed for 2 μs. [D-Trp3]-Contryphan-Vc2 simulations were extended for an additional 1 μs (total 3 μs). A simulation of [D-Trp3]-contryphan-Vc2 in water used isotropic Parrinello-Rahman pressure coupling, and peptide coordinates were not restrained during equilibration; otherwise the same equilibration procedure was used and the production run was executed for 100 ns. Simulation trajectories were visualised with VMD (v1.9.2) and analysed using VMD and GROMACS built-in tools.

3. Results

3.1. Synthesis and purification of contryphan-Vc2

Solid-phase peptide synthesis using Rink amide resin was used to prepare samples of contryphan-Vc2 containing either D-Trp, L-Trp or L-Ala at position 3. Analysis of each of the linear peptides by LC-MS confirmed that the synthetic Trp-containing peptides had masses of 863.6 Da, closely matching the expected molecular mass for reduced peptide (MH^+ m/z = 863.0 Da). Oxidation in ammonium bicarbonate buffer produced a mass loss of 2 Da to 861.6 Da, consistent with formation of the single disulfide bond (theoretical MH^+ m/z = 861.0 Da). The oxidised Ala-containing peptide contained ions of 746.7 Da, matching the expected LC-MS profile (theoretical MH^+ m/z = 746.9 Da). Purification of peptides by RP-HPLC resulted in samples of >96% purity (Fig. S1, Supporting Information).

3.2. D-isomerism changes peptide elution times

Samples of [D-Trp3]- and [L-Trp3]-contryphan-Vc2 (both singly and as a mixture) were injected on a LC-MS system fitted with a Luna 3u C8(2) 100 Å column and eluted with a gradient of 0–60% acetonitrile over 90 min. [D-Trp3]-Contryphan-Vc2 eluted at 37.0 min, one minute before [L-Trp3]-contryphan-Vc2, which eluted at 38.1 min. The sample containing a mixture of both peptides showed two distinct peaks, with elution times consistent with the single injections. (Fig. S2, Supporting Information). This shows that changing the chirality at Trp3 alters the properties of the peptide sufficiently to affect the chromatographic behaviour.

3.3. Trp3 is important for bioactivity, regardless of chirality

Bioactivity of other contryphans has been demonstrated by intracranial injection in mice (Jacobsen et al., 1999; Jimenez et al., 1997, 1996, 2001). In this study, we employed the same assay to screen for broad biological activity of contryphan-Vc2 and its analogues. Control mice injected with 10 μL normal saline solution showed active exploratory behaviour over the 2 h period of observation, with short periods of rest and grooming. In contrast, [D-Trp3]-contryphan-Vc2 administered by intracranial injection in mice at a dose of 10 nmol produced a strong and reproducible phenotype characterised by limited movement, splaying of hind limbs, dragging of hind limbs and flattening of the lower body when moving (Table 2). At the lowest dose tested (1 nmol) the same behaviour was observed but with reduced severity and duration. At the highest dose tested (20 nmol), the same behaviour was again observed and was increased in both severity and duration. [L-Trp3]-contryphan-Vc2 also produced a strong phenotype which, within the resolution of the assay, was indistinguishable from that elicited

Table 2
Summary of results from intracranial mouse injections. Mice were between 15 and 21 days old and weighed between 6.6 and 10.1 g.

Dose (nmol)	Observed Behaviour (time = approximate post injection)
	[D-Trp3]-contryphan-Vc2
20 ($n = 1$)	0 min – 1 h: Both hind and front legs splayed and body flat. 1 h–2 h: Limited movement. Hind legs splayed, dragging hind legs and lower body flat when moving. Mouse had not recovered by the end of observation at 2 h.
10 ($n = 5$)	From 0 min: Limited movement. Hind legs splayed when still, hind legs dragging and lower body flat when moving, intermittent episodes of body tremor. Recovery was observed at 15 min and 45 min for three of the mice, while the remaining two mice had not recovered by the end of observation at 2 h.
1 ($n = 2$)	5-20 min: Moving around cage, but hind legs dragging when moving. 20 min–2 h: no difference from control.
	[L-Trp3]-contryphan-Vc2
20 ($n = 1$)	0-10 min: Limited movement. Both hind and front legs splayed and body flat. 10–20 min: Limited movement. Recovered by 20 min.
10 ($n = 3$)	0-15 min: Limited movement. Hind legs splayed and body flat. 15 min–1 h: Increased movement with intermittent episodes of body tremor. The peptide produced a stronger response in one mouse with hind legs splayed, dragging hind legs and lower body flat when moving for the entire observation period of 2 h (similar to what was observed for the 20 nmol dose of [D-Trp3]-contryphan-Vc2).
	[W3A]-contryphan-Vc2
10 ($n = 3$)	Active exploratory behaviour with short periods of rest and grooming.
	Control
($n = 5$)	Active exploratory behaviour with short periods of rest and grooming.

by [D-Trp3]-contryphan-Vc2 (Table 2). However, the behaviour of mice injected with [W3A]-contryphan-Vc2 at a dose of 10 nmol could not be distinguished from that of control mice injected with normal saline solution. Together, these data demonstrate that contryphan-Vc2 is bioactive, and that the Trp sidechain in position 3 is critical for this bioactivity. Somewhat surprisingly, however, the chirality of Trp3 has no differential effect.

3.4. NMR spectroscopy

NMR spectra of [D-Trp3]-contryphan-Vc2 showed two sets of resonances, a major and minor species. The resonances of the major species showed good chemical shift dispersion, implying that the peptide adopts a well-defined conformation in solution (Fig. 1A).

Chemical shift assignments were made for ^{1}H and heavy atoms in the major conformer and ^{1}H in the minor conformer (Tables S1, S2 and S3, respectively, of the Supporting Information). Chemical shift data for the major conformer have been deposited in the BMRB (Ulrich et al., 2008), ID: 30152. The conformers differed in the geometry of the Thr4-Pro5 peptide bond, which was in the *trans* orientation in the major form, as shown by the strong NOE cross-peak between the Thr4 Hα and Pro5 Hδ resonances (Fig. S3, Supporting Information). Further corroborating this finding, the ^{13}C chemical shift difference between the Cβ and Cγ resonances of Pro5, 4.61 ppm, conformed to the published value typical for the *trans* conformation (4.51 ppm) (Schubert et al., 2002). In the minor conformer, the Thr4-Pro5 bond was in the *cis* orientation, as shown by the NOE cross-peak between the Hα resonances of Thr4 and Pro5 (Fig. S3, Supporting Information). The signals in the ^{13}C spectrum were too weak to determine the Δβγ value, which for *cis* bonds is typically 9.64 ppm (Schubert et al., 2002). The *cis*:*trans* ratio was determined to be 1:5.

3.5. Structure determination

An ensemble of 20 structures was calculated for the major conformer of [D-Trp3]-contryphan-Vc2. The experimental constraints used for structure calculation and the structural statistics for the ensemble generated are summarised in Table 3.

Dihedral angle constraints were obtained for Arg2 and Val6, both of which possess $^{3}J_{HNHA}$ above the 8.0 Hz threshold for β-

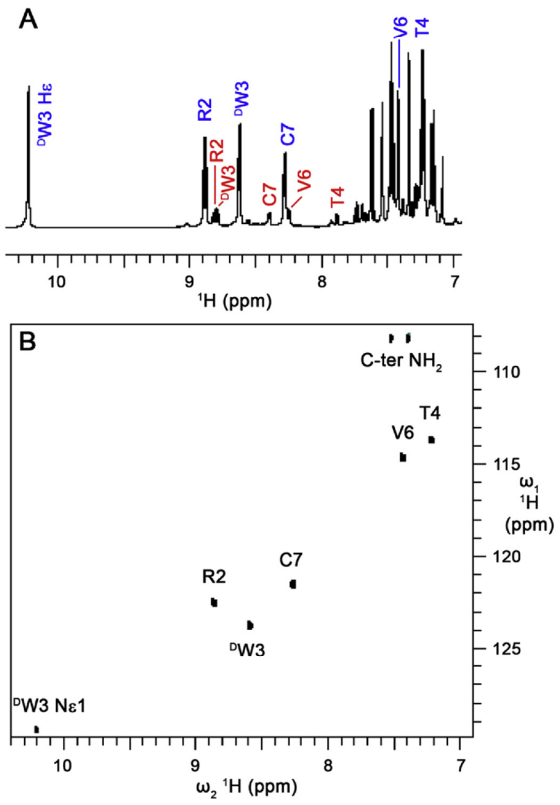


Fig. 1. 600 MHz NMR spectra of [D-Trp3]-contryphan-Vc2. (A) 1D ^{1}H spectrum, collected at 5 °C on a 2 mM sample at pH 4.0. Peaks are labelled from the major (blue) and minor (red) conformers. (B) 2D heteronuclear ^{1}H-^{15}N-HMQC spectrum, collected at 10 °C on a 500 μM sample at pH 4.1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Structural statistics for 20 lowest-energy structures of contryphan-Vc2.

Distance restraints	
Intra-residue (|i–j| = 0)	30
Sequential (|i–j| = 1)	33
Medium-range (2 ≤|i–j| ≤ 4)	10
Long-range (|i–j| > 4)	0
Total	73
Dihedral restraints	
Backbone (φ angle)	2
Sidechain (χ_1 angle)	0
RMSD over 20 structures (all residues)	
Backbone (Å) (N, Cα, C, O)	0.54 ± 0.29
All heavy atoms (Å)	0.80 ± 0.29
RMSD over 20 structures (residues 2–6)	
Backbone (Å) (N, Cα, C, O)	0.11 ± 0.06
All heavy atoms (Å)	0.39 ± 0.20
Ramachandran analysis	
Residues in most favoured regions (%)	13.8
Residues in additionally allowed regions (%)	70.0
Residues in generously allowed regions (%)	16.2
Residues in disallowed regions (%)	0.0
Energies (XPLOR energy units)	
E_{NOE}	2.8 ± 0.1
$E_{bond} + E_{angle} + E_{improper}$	10.2 ± 0.1
RMSDs from idealised geometry	
Bonds (Å)	0.0026 ± 0.0004
Angles (°)	0.515 ± 0.006
Impropers (°)	0.25 ± 0.01

strand conformations (−160° < φ < −80°). Analysis of amide temperature coefficients (Fig. S4, Supporting Information) suggested that the amide proton of Thr4 may be the donor of a hydrogen bond ($\Delta\delta/\Delta T = -2.0$ ppb/K, above the threshold of −4.5 ppb/K) (Baxter and Williamson, 1997), although this value could also be due to ring current effects from the adjacent aromatic D-Trp3 (Cierpicki and Otlewski, 2001). This amide (like all other amides) exchanged rapidly with deuterium when the peptide was dissolved in ^2H_2O, making a hydrogen bond unlikely. The calculated structures showed no consistent acceptor for this hypothetical bond and so it was not used as a constraint. Fig. 2A presents a stereo view of an overlay of the ensemble of final structures calculated for the major conformer of [D-Trp3]-contryphan-Vc2, while Fig. 2B shows the closest-to-average structure. The structural ensemble has been deposited with the PDB (Berman et al., 2003), ID: 5L34.

The NMR-derived structure of [D-Trp3]-contryphan-Vc2 in water reveals that the peptide adopts a well-defined structure consisting of two turn-like regions (formed by residues 1–4 and residues 4–7). The Cα atoms of Cys1 and Thr4 are separated by less than 7 Å, as are the Cα atoms of Thr4 and Cys7. However, the dihedral angles of the internal residues do not match the definitions of the formal types of β-turn (Richardson, 1981; Wilmot and Thornton, 1988). Both turns share Thr4 as a common residue, and are joined by a disulfide bond at the termini. The disulfide bond is the least constrained region of the structure, and makes hydrophobic contact with Thr4. The RMSD values calculated across the structure ensemble show that all structures are similar, especially when the terminal residues are excluded. Ramachandran values are favourable, with most residues falling in the additionally allowed regions of the Ramachandran plot and none present in the disallowed regions (Table 3). This structure is maintained at

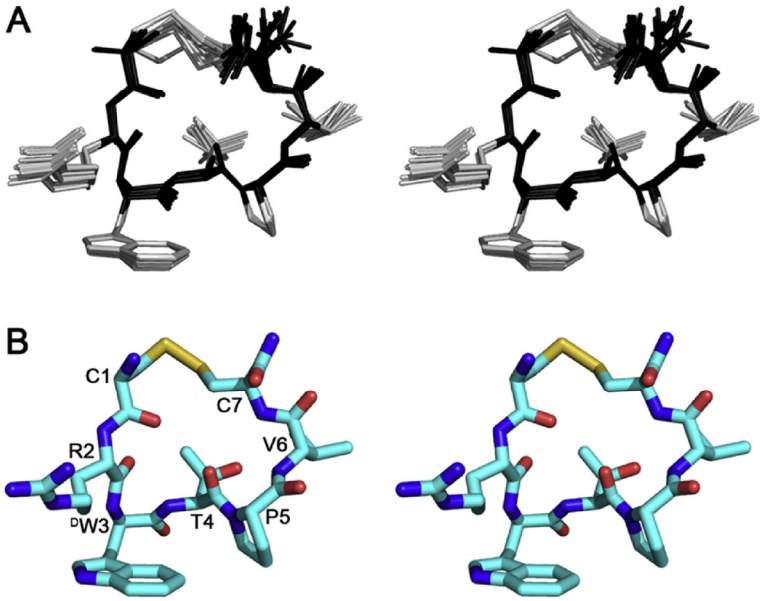


Fig. 2. Stereo images of the calculated solution structure of the major [D-Trp3]-contryphan-Vc2 conformer. (A) Ensemble of 20 structures of [D-Trp3]-contryphan-Vc2 calculated from distance and dihedral angle restraints. Structures have been superimposed over the backbone atoms of residues 2–6. Backbone is coloured *black*, residue side chains are *grey*. (B) Closest-to-average structure from the ensemble. The disulfide bond is *yellow*, nitrogen atoms are *blue*, oxygen atoms are *red*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

physiological temperature (Fig. S5, Supporting Information).

In order to probe the dynamics of [D-Trp3]-contryphan-Vc2 in solution, MD simulations were undertaken. These simulations revealed dynamic behaviour in the conformation associated with rotation of Trp3 around the χ_1 bond (Fig. 3; see also Fig. S6, Supporting Information). Fig. 3A plots the orientation of the Trp3 sidechain throughout the length of the simulation trajectory, revealing the adoption of two primary positions. Fig. 3B illustrates the major conformer (Trp3 $\chi_1 = +60°$), in which the Trp3 and Arg2 sidechains are in close proximity to each other, projecting from a turn-like structure. In the minor conformer (Trp3 $\chi_1 = -60°$, Fig. 3C), the sidechains of Trp3 and Pro5 are in close proximity. Neither of these conformers matches the sidechain positions of the calculated NMR structure, although both maintain the Thr4-Pro5 bond in the *trans* conformation and are similar in terms of the backbone conformation, with the turns intact and the disulfide bond still making hydrophobic contact with Thr4.

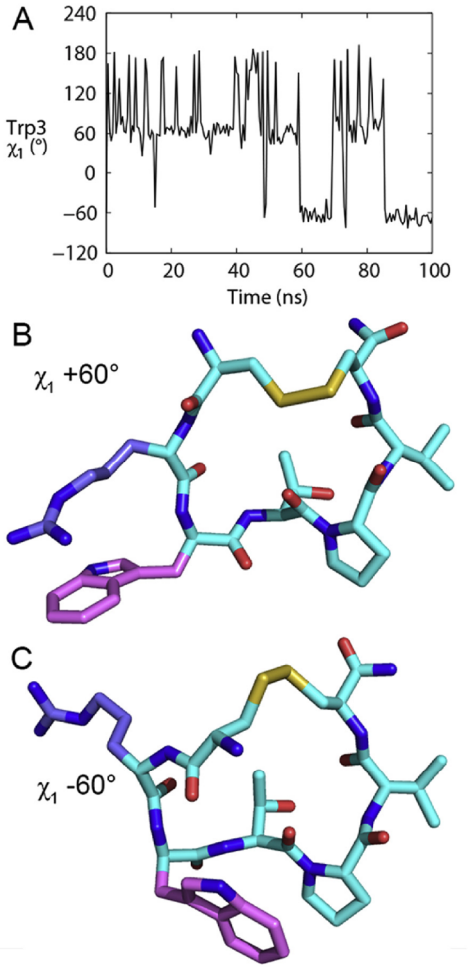


Fig. 3. Conformational data arising from MD simulations of [D-Trp3]-contryphan-Vc2 in water. (A) Plot of Trp3 χ_1 angle vs time, showing presence of two distinct states. (B) Representative structure of high occupancy state (Trp3 $\chi_1 = +60°$), in which Arg2 and Trp3 sidechains are in close proximity. (C) Representative structure of low occupancy state (Trp3 $\chi_1 = -60°$), in which Trp3 and Pro5 sidechains are in close proximity.

3.6. Chemical shift comparison between [D-Trp3]- and [L-Trp3]-containing peptides suggests structural similarity

The spectra of [L-Trp3]-contryphan-Vc2 were well-dispersed, and chemical shift assignments were made for all ^{1}H resonances of the major species (Table S4 of the Supporting Information). Comparison of the ^{1}H resonance chemical shifts of [D-Trp3]-contryphan-Vc2 and [L-Trp3]-contryphan-Vc2 revealed minor differences. The amide proton of Trp3 is shifted upfield by over 1 ppm in [L-Trp3]-contryphan-Vc2, with a smaller downfield shift for the amide proton of Thr4 of around 0.5 ppm. There is a comparable 0.5 ppm downfield shift for the Hα resonance of Pro5, and minor shifts in both directions for other resonances, both Hα and HN (Fig. S7, Supporting Information). These results suggest that the structure of [L-Trp3]-contryphan-Vc2, while slightly different, is largely intact and similar to that of [D-Trp3]-contryphan-Vc2.

3.7. D-isomerism protects against proteolytic degradation

One possible reason for the presence of D-amino acids in peptides is that the atypical chirality protects the molecule from proteolytic degradation. To test this notion in the context of contryphan-Vc2, the digestive enzymes trypsin, α-chymotrypsin and pepsin were used in proteolytic assays. Trypsin degraded [L-Trp3]-contryphan-Vc2 within 4 h, while having no effect on [D-Trp3]-contryphan-Vc2. Neither [D-Trp3]- nor [L-Trp3]-contryphan-Vc2 was susceptible to cleavage by pepsin or α-chymotrypsin over the 4 h duration of the assay. LC-MS traces are presented in Fig. S8 of the Supporting Information. These results suggest that the D-chirality of Trp3 is protective against proteolytic degradation.

3.8. Contryphan-Vc2 interacts with a model lipid system

The insensitivity of the bioactivity assay to the chirality of Trp3 suggested that the peptide may not interact with a protein receptor in a traditional manner. Titration of [D-Trp3]-contryphan-Vc2 with DPC micelles showed clear signs of an interaction between peptide and lipid, with amide peaks broadening and shifting as the concentration of lipid was increased. A similar effect was observed for titration of [L-Trp3]-contryphan-Vc2 into DPC micelles. Plots showing changes in chemical shift with DPC concentration are presented in Fig. 4 (Panels A–C), and spectra of [D-Trp3]- and [L-Trp3]-contryphan-Vc2 at 0 and 40 mM DPC are shown in Fig. S9 of the Supporting Information. These results demonstrate that both [D-Trp3]-contryphan-Vc2 and [L-Trp3]-contryphan-Vc2 interact with this model lipid system. When tested at a peptide concentration of 0.5 mM these effects stabilised upon addition of 20–40 mM DPC and no further alterations to the spectra were apparent in samples on addition of 100 mM DPC (data not shown). Hence saturation was reached around a ratio of 80:1 lipid:peptide. MD simulations were run to gain insight into specific molecular interactions between peptide and lipid bilayers, as described in the Supporting Information, and a representative structure of [D-Trp3]-contryphan-Vc2 in a lipid bilayer is shown in Fig. 4D.

4. Discussion

Contryphan-Vc2 differs from the consensus contryphan sequence (Sonti et al., 2013) at positions 2 and 6 (replacing Pro/Hyp2 with Arg and Trp6 with Val). The behaviour in RP-HPLC is also different from that typically seen for contryphans in containing a single sharp peak and lacking the later-eluting peak observed in elution profiles of other contryphans (Jacobsen et al., 1999). The sequence changes in contryphan-Vc2 are evidently sufficient to abolish the appearance of a second species on the chromatographic

timescale. However, the NMR spectra of [D-Trp3]-contryphan-Vc2 clearly show two sets of peaks, indicating that this peptide does exist in two distinct conformations in solution, as has been observed previously for other members of the contryphan family (Eliseo et al., 2004; Pallaghy et al., 1999, 2000). The difference between the two conformations of contryphan-Vc2 was found to be *cis*/*trans* isomerism about the Thr4-Pro5 bond, with a *cis*:*trans* ratio of 1:5. All but one of the contryphans characterised to date display *cis*/*trans* isomerism (the exception being glacontryphan-M) (Grant et al., 2004). Of the two Pro residues in the consensus sequence, isomerism is usually observed in the conformation of the Cys-Pro bond at the start of the intercystine loop (where Pro can also be modified to Hyp), but in contryphan-Vc2, Pro2 is replaced by Arg and *cis*/*trans* isomerism occurs at the second Pro residue. This is unusual, as in all but one of the previously published solution structures of contryphans this second Xaa-Pro bond has been exclusively in the *trans* conformation (Eliseo et al., 2004; Grant et al., 2004; Pallaghy et al., 1999, 2000). The sole exception is contryphan-In (In936), which a recent study identified as also possessing isomerism at the second Xaa-Pro bond and which, like contryphan-Vc2, lacks the first Pro residue (containing Val instead) (Sonti et al., 2013) (Table 1). The *cis* bond in contryphan-In may be promoted by the Tyr residue preceding the second Pro, as it has been noted that preceding aromatic residues promote the formation of *cis* peptide bonds to Pro through an interaction between the partially positive Pro ring face and the negatively charged aromatic π-face (Zondlo, 2012). The residue preceding Pro5 in contryphan-Vc2 is Thr4, so this type of interaction is not present in this peptide.

Unlike the structures of contryphans solved previously, there is no *i*/*i*+*4* electrostatic interaction between the N-terminal ammonium group and an acidic sidechain, an interaction that has been hypothesised to stabilise *cis* Xaa-Pro bonds in contryphans (Pallaghy et al., 2000). Despite these differences, the orientation of the backbone in contryphan-Vc2 is broadly similar to other members of the contryphan family, as can be seen in an overlay of selected structures with the calculated NMR structure of [D-Trp3]-contryphan-Vc2 (Fig. 5).

Based on the NMR data, a model of the solution structure for [D-Trp3]-contryphan-Vc2 was calculated that satisfied the observed NOE and dihedral angle restraints. However, MD simulations of the peptide in aqueous solvent cast doubt on the adequacy of this single NMR model as a representation of the solution conformation of [D-Trp3]-contryphan-Vc2. The NMR-derived structure positions the sidechains of Arg2, Trp3 and Pro5 in close proximity to each other, satisfying the restraints generated by the observed NOE cross-peaks between Trp3 and the other two residues. However, the MD data suggest that the peptide actually adopts two distinct conformations, distinguished by rotation about the Trp3 χ_1 bond. In one of these, the sidechain of Trp3 is in close proximity to that of Arg2, while in the other it is proximal to Pro5 (as illustrated in Fig. 3). If the peptide does adopt these two distinct conformations in solution, then the assumption made in calculating the solution structure, that a single conformation was present, is compromised. The atomic positions implied by NOE results are assumed to be at full occupancy, with weaker cross-peaks interpreted as arising from an interaction between more distant atoms, rather than being a

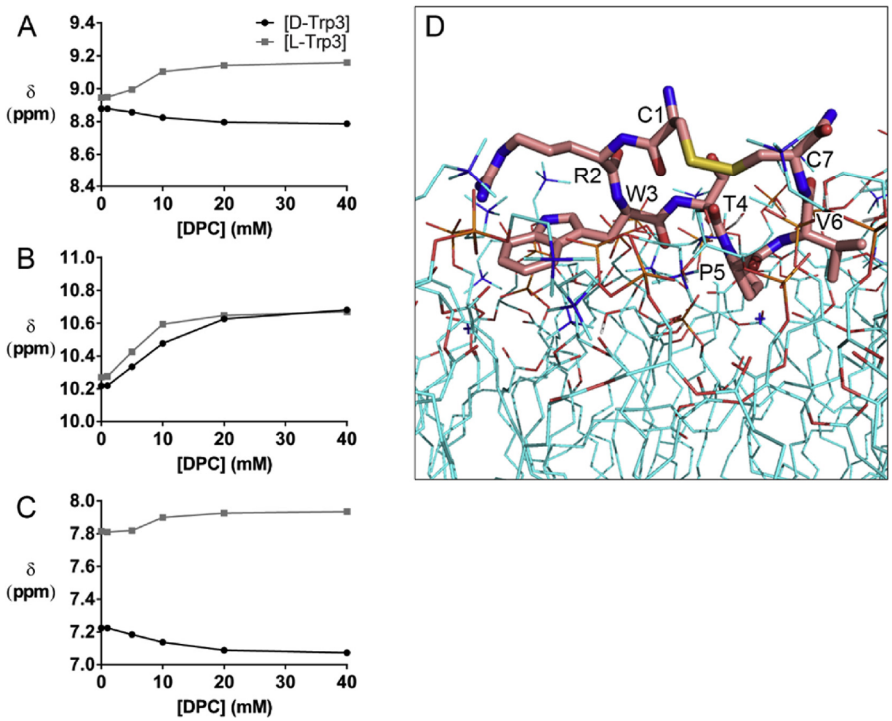


Fig. 4. Interactions of contryphan-Vc2 with lipid. (A) Chemical shift of Arg2 HN resonance in presence of increasing concentrations of DPC. (B) DPC-induced chemical shift changes for Trp3 Hε1 resonance. (C) DPC-induced chemical shift changes for Thr4 HN resonance. (D) Representative frame of [D-Trp3]-contryphan-Vc2 from MD simulation (see also Supporting Information, Figs. S10 and S11), showing position and orientation of the peptide in the polar head group region of the lipid bilayer.

consequence of the peptide sampling multiple conformations. In the latter case, the calculated NMR structure would be the result of averaging the interactions observed in all sampled conformations, and would not necessarily correspond to a pose that the peptide ever actually adopts. Such appears to be the case for contryphan-Vc2, as the MD simulations never show the sidechains of Arg2 and Pro5 in close proximity as in the structures calculated from the NMR data. This is an intriguing finding, as the NMR spectra of this peptide in aqueous solution are sharp and well-dispersed, properties that would typically be interpreted as indicating a single major conformation, or ensemble of very similar conformations. Spectra at substantially lower temperatures or in the solid state could potentially resolve the resonances from these two major conformations detected by the MD simulations.

An effective method for initially assessing the bioactivity of peptide toxins has been administration by intracranial injection in mice. Members of the contryphan family that have been assayed in this way have been reported to induce a 'stiff-tail' syndrome at low doses, while higher doses led to hyperactive behaviours (such as circular motion and barrel rolling), and eventually seizures and death (Jacobsen et al., 1998, 1999; Jimenez et al., 1997, 1996, 2001). By contrast, contryphan-Vc2 appears to induce a depressed phenotype with minimal movement; the mice adopt a splayed posture and drag their hind limbs when moving. This may indicate that the sequence differences in contryphan-Vc2 are sufficient to direct it to a different biological target, as has been demonstrated for several other toxin folds (Froy et al., 1999; Norton and Pallaghy, 1998).

It is intriguing that both [D-Trp3]-contryphan-Vc2 and [L-Trp3]-contryphan-Vc2 elicit indistinguishable biological effects. The lack of activity observed with [W3A]-contryphan-Vc2 suggests that the side chain of Trp3 is directly involved in bioactivity, and presumably receptor binding. However, it would be expected that a change in chirality of this crucial sidechain would disrupt the interaction between peptide and receptor, leading to a loss of or reduction in activity (Dawson et al., 1999; Kreil et al., 1989).

While the similar activities of both D- and L-isomers may point to a mechanism of action mediated by a receptor other than a protein, there are examples of protein receptors that possess a more permissive binding site, for example CXCR4, a chemokine receptor, that was found to interact with all-L, all-D and mixed L/D analogues of a peptide derived from the N-terminal region of viral macrophage inflammatory protein II (vMIP-II) (Zhou et al., 2002).

An alternative to direct receptor binding could be that contryphan-Vc2 interacts with membranes, and then influences a biological target in a membrane-mediated manner. This phenomenon has been reported before; for example, in the interaction between the spider toxin GsMTx4 and gramicidin A channels (Suchyna et al., 2004). The gramicidin A channel is formed when two monomers in opposing leaflets of a lipid bilayer associate together, an event dependent on thinning of the bilayer in a localised area. GsMTx4 alters the lipid packing adjacent to the channel and induces this thinning, and has been observed to enhance the appearance of dimeric channels by a factor of 10–25-fold and extend the lifetime of an open channel by a factor of two. An all-D analogue of GsMTx4 was tested in the same study and exhibited identical activity to the native toxin, mirroring the situation with contryphan-Vc2. It is therefore conceivable that the effects of contryphan-Vc2 are due to a channel which is similarly sensitive to the state of the bilayer.

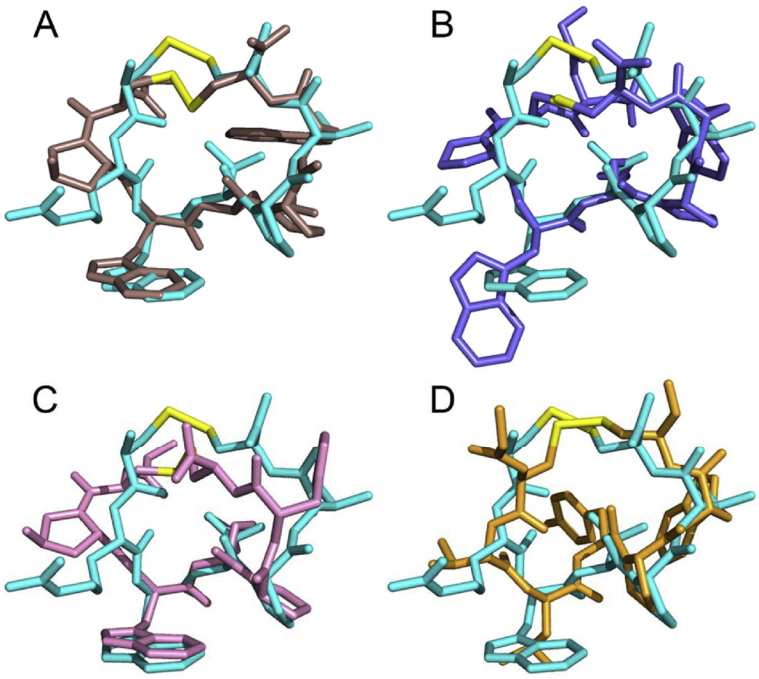


Fig. 5. Superimposed structures of [D-Trp3]-contryphan-Vc2 (*cyan*) with other contryphan family members. (A) Contryphan-R, PDB 1QFB (Pallaghy et al., 1999). (B) Contryphan-Vn, PDB 1NXN (Eliseo et al., 2004). (C) Contryphan-Sm, PDB 1DFY (Pallaghy et al., 2000). (D) Contryphan-In, PDB 2M6D (Sonti et al., 2013). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Titration of [D-Trp3]- or [L-Trp3]-contryphan-Vc2 with DPC micelles led to peak shifting and broadening in NMR spectra, so to gain insight into possible interactions with lipid bilayers, a series of MD simulations was performed with [D-Trp3]-, [L-Trp3]- and [W3A]-contryphan-Vc2 in the presence of mixed POPC/POPG bilayers. Two simulations were run for each peptide, one with the peptide starting in the aqueous phase and the other with the peptide already embedded in the membrane. In all simulations, the peptide migrated from its starting point to the headgroup region of the bilayer (Fig. S10, Supporting Information), although the final equilibrium position of each peptide was influenced by the starting point and did not converge for any pair of simulations. This suggests that there is a significant barrier to transfer of these peptides from aqueous solution to complete burial within the lipid phases. There were no apparent differences between the simulations of [D-Trp3]- and [L-Trp3]-contryphan-Vc2; both migrate to a region near the head group of the lipid and remain in that position for the duration of the simulation.

5. Conclusions

Contryphan-Vc2 is distinct from the rest of the contryphan peptide family in having a novel bioactivity in mice. The chirality of Trp3 did not affect the activity of the peptide in mice, with both D- and L-isoforms eliciting the atypical depressive phenotype, although D-Trp did confer protection against proteolytic degradation. The solution structure determined on the basis of NMR data was well defined, but MD simulations cast doubt on whether the calculated structure is an adequate representation of the ensemble of conformations accessible to this peptide in aqueous solution. Although the backbone is well defined, MD suggested much greater flexibility around the Trp3 Cα-Cβ bond.

NMR studies show that contryphan-Vc2 interacts with the membrane, a finding supported by MD simulations. Whether membrane interactions are crucial to the bioactivity or if the presumed receptor has a permissive binding site capable of accepting either D- or L-Trp3 remains an open question. A role for lipid in contryphan action may also account for the difficulty to date in identifying a molecular target for this class of peptides. If the molecular entities responsible for the activity of contryphan-Vc2 could be identified, it would be interesting to determine the precise structure-activity relationship of the interaction, especially with respect to the isomerism of Trp3 and the apparent inability of other contryphans to elicit a similar effect in mice.

Funding

R.S.N. acknowledges fellowship support from the NHMRC.

Notes

The authors declare no competing financial interest.

Acknowledgements/Author contributions

AB and AJR synthesised the initial sample, further samples were synthesised by SBD. SDR and SSE performed the behavioural assays. SBD collected NMR data and assigned the NMR spectra. SBD, SC and CAM calculated the solution structure. BC performed proteolysis assays. DKC advised on MD simulations. SBD, SDR and RSN wrote the manuscript, with input from all authors. RSN conceived of the study. RSN acknowledges fellowship support from the Australian National Health and Medical Research Council. We thank the Multimodal Australian Sciences Imaging and Visualisation Environment (MASSIVE) and Victorian Life Sciences Computation Initiative (VLSCI) for provision of CPU time and technical support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.toxicon.2017.02.012.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.toxicon.2017.02.012.

References

Bai, L., Sheeley, S., Sweedler, J.V., 2009. Analysis of endogenous D-amino acid-containing peptides in Metazoa. Bioanal. Rev. 1, 7–24.
Baxter, N.J., Williamson, M.P., 1997. Temperature dependence of ^{1}H chemical shifts in proteins. J. Biomol. NMR 9, 359–369.
Berman, H., Henrick, K., Nakamura, H., 2003. Announcing the worldwide protein data bank. Nat. Struct. Mol. Biol. 10, 980–980.
Cierpicki, T., Otlewski, J., 2001. Amide proton temperature coefficients as hydrogen bond indicators in proteins. J. Biomol. NMR 21, 249–261.
Dawson, D.W., Volpert, O.V., Pearce, S.F.A., Schneider, A.J., Silverstein, R.L., Henkin, J., Bouck, N.P., 1999. Three distinct D-amino acid substitutions confer potent antiangiogenic activity on an inactive peptide derived from a thrombospondin-1 type 1 repeat. Mol. Pharmacol. 55, 332–338.
Eliseo, T., Cicero, D.O., Romeo, C., Schinina, M.E., Massilia, G.R., Polticelli, F., Ascenzi, P., Paci, M., 2004. Solution structure of the cyclic peptide contryphan-Vn, a Ca^{2+}-dependent K^{+} channel modulator. Biopolymers 74, 189–198.
Fiedler, B., Zhang, M.-M., Buczek, O., Azam, L., Bulaj, G., Norton, R.S., Olivera, B.M., Yoshikami, D., 2008. Specificity, affinity and efficacy of iota-conotoxin RXIA, an agonist of voltage-gated sodium channels Na_V1.2, 1.6 and 1.7. Biochem. Pharmacol. 75, 2334–2344.
Froy, O., Sagiv, T., Poreh, M., Urbach, D., Zilberberg, N., Gurevitz, M., 1999. Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels. J. Mol. Evol. 48, 187–196.
Grant, M.A., Hansson, K., Furie, B.C., Furie, B., Stenflo, J., Rigby, A.C., 2004. The metal-free and calcium-bound structures of a γ-carboxyglutamic acid-containing contryphan from *Conus marmoreus*, glacontryphan-M. J. Biol. Chem. 279, 32464–32473.
Hansson, K., Ma, X., Eliasson, L., Czerwiec, E., Furie, B., Furie, B.C., Rorsman, P., Stenflo, J., 2004. The first γ-carboxyglutamic acid-containing contryphan: a selective L-type calcium ion channel blocker isolated from the venom of *Conus marmoreus*. J. Biol. Chem. 279, 32453–32463.
Heck, S.D., Kelbaugh, P.R., Kelly, M.E., Thadeio, P.F., Saccomano, N.A., Stroh, J.G., Volkmann, R.A., 1994. Disulfide bond assignment of ω-agatoxins IVB and IVC: discovery of a D-Serine residue in ω-agatoxin IVB. J. Am. Chem. Soc. 116, 10426–10436.
Jacobsen, R., Jimenez, E.C., Grilley, M., Watkins, M., Hillyard, D., Cruz, L.J., Olivera, B.M., 1998. The contryphans, a D-tryptophan-containing family of Conus peptides: interconversion between conformers. J. Pept. Res. 51, 173–179.
Jacobsen, R.B., Jimenez, E.C., De la Cruz, R.G.C., Gray, W.R., Cruz, L.J., Olivera, B.M., 1999. A novel D-leucine-containing Conus peptide: diverse conformational dynamics in the contryphan family. J. Pept. Res. 54, 93–99.
Jimenez, E.C., Craig, A.G., Watkins, M., Hillyard, D.R., Gray, W.R., Gulyas, J., Rivier, J.E., Cruz, L.J., Olivera, B.M., 1997. Bromocontryphan: Post-translational bromination of tryptophan. Biochemistry 36, 989–994.
Jimenez, E.C., Olivera, B.M., Gray, W.R., Cruz, L.J., 1996. Contryphan is a D-tryptophan-containing *Conus* peptide. J. Biol. Chem. 271, 28002–28005.
Jimenez, E.C., Watkins, M., Juszczak, L.J., Cruz, L.J., Olivera, B.M., 2001. Contryphans from *Conus textile* venom ducts. Toxicon 39, 803–808.
Kreil, G., Barra, D., Simmaco, M., Erspamer, V., Falconieri Erspamer, G., Negri, L., Severini, C., Corsi, R., Melchiorri, P., 1989. Deltorphin, a novel amphibian skin peptide with high selectivity and affinity for δ opioid receptors. Eur. J. Pharmacol. 162, 123–128.
Massilia, G.R., Eliseo, T., Grolleau, F., Lapied, B., Barbier, J., Bournaud, R., Molgo, J., Cicero, D.O., Paci, M., Schinina, M.E., Ascenzi, P., Polticelli, F., 2003. Contryphan-Vn: a modulator of Ca^{2+}-dependent K^{+} channels. Biochem. Biophys. Res. Commun. 303, 238–246.
Massilia, G.R., Schinina, M.E., Ascenzi, P., Polticelli, F., 2001. Contryphan-Vn: a novel peptide from the venom of the Mediterranean snail *Conus ventricosus*. Biochem. Biophys. Res. Commun. 288, 908–913.
McIntosh, J.M., Yoshikami, D., Mahe, E., Nielsen, D.B., Rivier, J.E., Gray, W.R., Olivera, B.M., 1994. A nicotinic acetylcholine receptor ligand of unique specificity, α-conotoxin ImI. J. Biol. Chem. 269, 16733–16739.
Montecucchi, P.C., De Castiglione, R., Piani, S., Gozzini, L., Erspamer, V., 1981. Amino acid composition and sequence of dermorphin, a novel opiate-like peptide from the skin of *Phyllomedusa sauvagei*. Int. J. Pept. Protein Res. 17, 275–283.
Norton, R.S., Pallaghy, P.K., 1998. The cystine knot structure of ion channel toxins and related polypeptides. Toxicon 36, 1573–1583.

Pallaghy, P.K., He, W.L., Jimenez, E.C., Olivera, B.M., Norton, R.S., 2000. Structures of the contryphan family of cyclic peptides. Role of electrostatic interactions in *cis-trans* isomerism. Biochemistry 39, 12845–12852.

Pallaghy, P.K., Melnikova, A.P., Jimenez, E.C., Olivera, B.M., Norton, R.S., 1999. Solution structure of contryphan-R, a naturally occurring disulfide-bridged octapeptide containing D-tryptophan: comparison with protein loops. Biochemistry 38, 11553–11559.

Rajesh, R.P., 2015. Novel M-Superfamily and T-Superfamily conotoxins and contryphans from the vermivorous snail *Conus figulinus*. J. Pept. Sci. 21, 29–39.

Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34, 167–339.

Robinson, S.D., Safavi-Hemami, H., McIntosh, L.D., Purcell, A.W., Norton, R.S., Papenfuss, A.T., 2014. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. PLoS One 9, e87648.

Sabareesh, V., Gowd, K.H., Ramasamy, P., Sudarslal, S., Krishnan, K.S., Sikdar, S.K., Balaram, P., 2006. Characterization of contryphans from *Conus loroisii* and *Conus amadis* that target calcium channels. Peptides 27, 2647–2654.

Schubert, M., Labudde, D., Oschkinat, H., Schmieder, P., 2002. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. J. Biomol. NMR 24, 149–154.

Sonti, R., Gowd, K.H., Rao, K.N.S., Ragothama, S., Rodriguez, A., Perez, J.J., Balaram, P., 2013. Conformational diversity in contryphans from *Conus* venom: *cis-trans* isomerisation and aromatic/proline interactions in the 23-membered ring of a 7-residue peptide disulfide loop. Chem. Eur. J. 19, 15175–15189.

Suchyna, T.M., Tape, S.E., Koeppe, R.E., Andersen, O.S., Sachs, F., Gottlieb, P.A., 2004. Bilayer-dependent inhibition of mechanosensitive channels by neuroactive peptide enantiomers. Nature 430, 235–240.

Thakur, S.S., Balaram, P., 2007. Rapid mass spectral identification of contryphans. Detection of characteristic peptide ions by fragmentation of intact disulfide-bonded peptides in crude venom. Rapid Commun. Mass Spectrom. 21, 3420–3426.

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Kent Wenger, R., Yao, H., Markley, J.L., 2008. BioMagResBank. Nucleic Acids Res. 36, D402–D408.

Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Gō, N., Wüthrich, K., 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. J. Mol. Biol. 196, 611–639.

Wilmot, C.M., Thornton, J.M., 1988. Analysis and prediction of the different types of β-turn in proteins. J. Mol. Biol. 203, 221–232.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., Sykes, B.D., 1995. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR 6, 135–140.

Zhou, N., Luo, Z., Luo, J., Fan, X., Cayabyab, M., Hiraoka, M., Liu, D., Han, X., Pesavento, J., Dong, C.-Z., Wang, Y., An, J., Kaji, H., Sodroski, J.G., Huang, Z., 2002. Exploring the stereochemistry of CXCR4-peptide recognition and inhibiting HIV-1 entry with D-peptides derived from chemokines. J. Biol. Chem. 277, 17476–17485.

Zondlo, N.J., 2012. Aromatic–proline interactions: electronically tunable CH/π interactions. Acc. Chem. Res. 46, 1039–1049.

Supporting Information

Structure and activity of contryphan-Vc2: Importance of the D-amino acid residue

Stephen B. Drane[a], Samuel D. Robinson[a], Christopher A. MacRaild[a], Sandeep Chhabra[a], Balasubramanyam Chittoor[a], Rodrigo A. V. Morales[a], Eleanor W. W. Leung[a], Alessia Belgi[b], Samuel S. Espino[c], Baldomero M. Olivera[c], Andrea J. Robinson[b], David K. Chalmers[a], Raymond S. Norton[a*]

[a]Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Victoria, Australia
[b]School of Chemistry, Monash University, Clayton 3800, Victoria, Australia
[c]Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

* Corresponding author: Raymond S. Norton, Monash Institute of Pharmaceutical Sciences, 381 Royal Parade, Parkville, VIC 3052, Australia, [redacted]

Table S1. ^{1}H resonance assignments for [D-Trp3]-contryphan-Vc2 major conformer (*trans* Thr4-Pro5 peptide bond).

Residue	Proton chemical shift (ppm)		
	H^N	H^α	H (other)
Cys1	-	4.33	Hβ1 3.22; Hβ2 3.27
Arg2	8.88	4.52	Hβ1 1.34; Hβ2 1.42; Hγ1 0.99; Hδ1 2.79; Hδ2 2.88; Hε 6.90
Trp3	8.61	4.82	Hβ1 3.21; Hβ2 3.41; Hδ1 7.33; Hε1 10.21; Hε3 7.61; Hζ2 7.47; Hζ3 7.14; Hη2 7.23
Thr4	7.22	4.67	Hβ 4.40; Hγ2* 1.13
Pro5		4.23	Hβ1 1.98; Hβ2 2.38; Hγ1 2.06; Hδ1 3.72; Hδ2 3.85
Val6	7.45	4.18	Hβ 2.07; Hγ1* 0.88; Hγ2* 0.89
Cys7	8.27	4.57	Hβ1 2.96; Hβ2 3.16
Nh2	7.41,7.53		

Table S2. ^{13}C/^{15}N resonance assignments for [D-Trp3]-contryphan-Vc2 major conformer (*trans* Thr4-Pro5 peptide bond).

Residue	Heavy atom chemical shift (ppm)			
	N^H	N (other)	C^α	C (other)
Cys1	-		54.8	Cβ 40.9
Arg2	122.4		55.5	Cβ 31.5; Cγ 26.0; Cδ 43.3
Trp3	123.7	Nε 129.4	57.1	Cβ 29.2; Cδ1 127.4; Cε3 120.9; Cζ2 114.5; Cζ3 121.9; Cη2 124.7
Thr4	113.6		58.7	Cβ 69.5; Cγ2 21.3
Pro5			64.7	Cβ 32.0; Cγ 27.4; Cδ 51.1
Val6	114.6		61.8	Cβ 32.7; Cγ1 20.2; Cγ2 21.3
Cys7	121.5		55.6	Cβ 40.0
Nh2	108.1			

Table S3. ^{1}H resonance assignments for [D-Trp3]-contryphan-Vc2 minor conformer (*cis* Thr4-Pro5 peptide bond).

	Proton chemical shift (ppm)		
Residue	**H^N**	**H^α**	**H (other)**
Cys1	-	4.30	Hβ1 3.37
Arg2	8.81	4.27	Hβ1 1.44; Hγ1 0.73; Hγ2 1.00; Hδ1 2.80; Hδ2 2.85; Hε 6.82
Trp3	8.78	4.90	Hβ1 3.18; Hβ2 3.40; Hδ1 7.28; Hε1 10.25; Hε3 7.73; Hζ2 7.47; Hζ3 7.16
Thr4	7.88	4.35	Hβ 3.91; Hγ2* 1.07
Pro5		4.91	Hβ1 2.25; Hβ2 2.32; Hγ1 1.82; Hγ2 1.99; Hδ1 3.56; Hδ2 3.66
Val6	8.24	4.13	Hβ 2.19; Hγ1* 1.00
Cys7	8.39	4.57	Hβ1 3.09; Hβ2 3.22
Nh2	7.38,7.68		

Table S4. ^{1}H resonance assignments for [L-Trp3]-contryphan-Vc2.

	Proton chemical shift (ppm)		
Residue	**H^N**	**H^α**	**H (other)**
Cys1	-	-	-
Arg2	8.84	4.15	Hβ1 1.31; Hγ1 0.96; Hδ1 2.79; Hε 6.85
Trp3	7.28	4.78	Hβ1 3.30; Hβ2 3.43; Hδ1 7.28; Hε1 10.27; Hε3 7.68; Hζ2 7.47; Hζ3 7.17; Hη2 7.22
Thr4	7.75	4.42	Hβ 4.43; Hγ2* 1.21
Pro5	-	4.87	Hβ1 2.30; Hγ1 1.82; Hγ2 2.07; Hδ1 3.60; Hδ2 3.69
Val6	7.57	4.30	Hβ 2.12; Hγ1* 0.94
Cys7	8.37	4.73	Hβ1 2.96; Hβ2 3.29
Nh2	-		

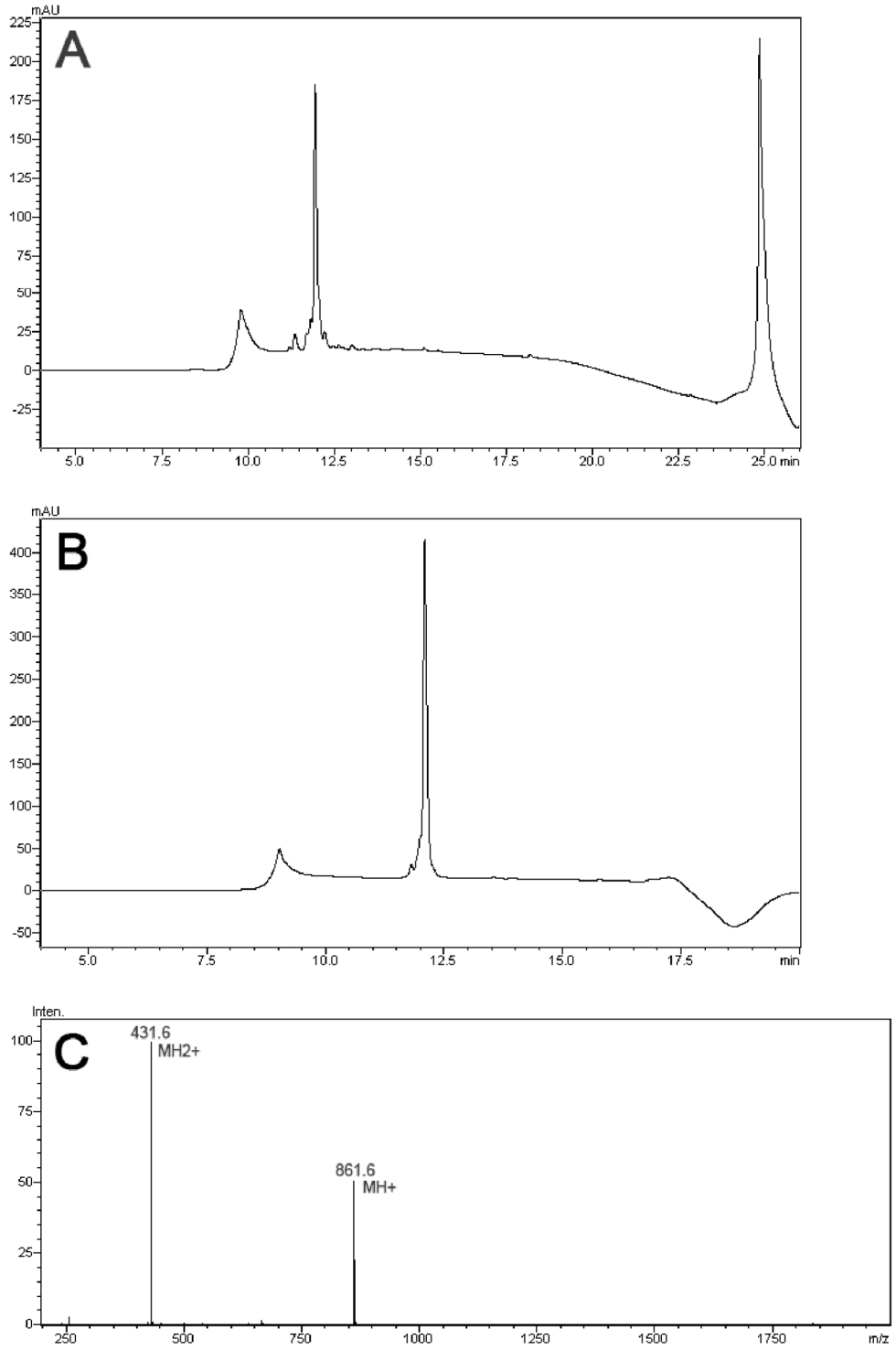


Figure S1. Purification of [L-Trp3]-contryphan-Vc2. **(A)** Oxidative folding mixture, run on LC-MS system with 3 µm C8(2) 100 Å column, flowrate 0.2 mg/mL, gradient 0-100% B over 15 min; Buffer A: 0.05% TFA in H_2O, Buffer B: 0.05% TFA in acetonitrile. **(B)** Final purification, run on LC-MS system with 3 µm C8(2) 100 Å column, gradient 0-60% B over 10 min; Buffer A: 0.05% TFA in H_2O, Buffer B: 0.05% TFA in acetonitrile. **(C)** Ions present in MS of pure peak.

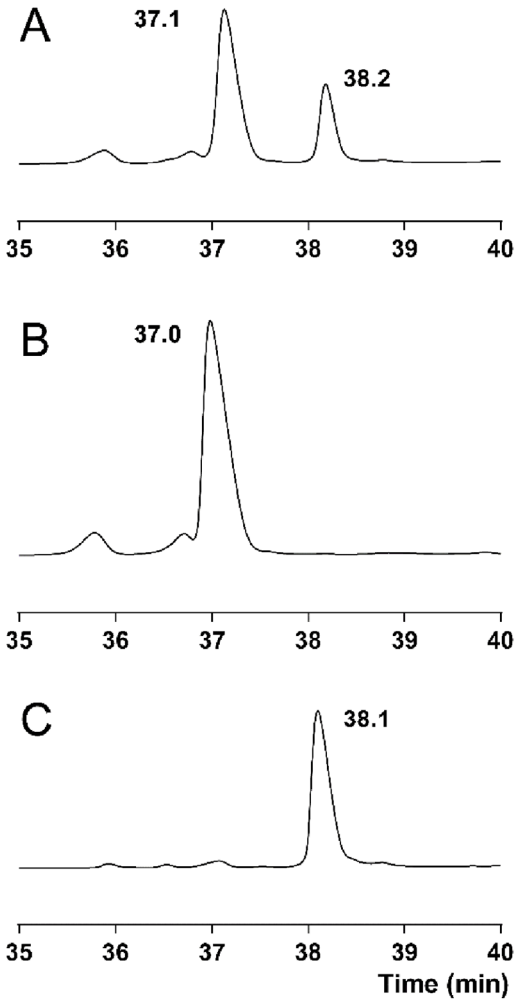


Figure S2. Co-elution of [D-Trp3]-contryphan-Vc2 and [L-Trp3]-contryphan-Vc2, run on LC-MS with a 3 µm C8(2) 100Å column. Gradient 0-60% B over 90 min; Buffer A: 0.05% TFA in H_2O, Buffer B: 0.05% TFA in acetonitrile. (A) 50:50 mixture of [D-Trp3]-contryphan-Vc2 and [L-Trp3]-contryphan-Vc2. (B) [D-Trp3]-contryphan-Vc2. (C) [L-Trp3]-contryphan-Vc2.

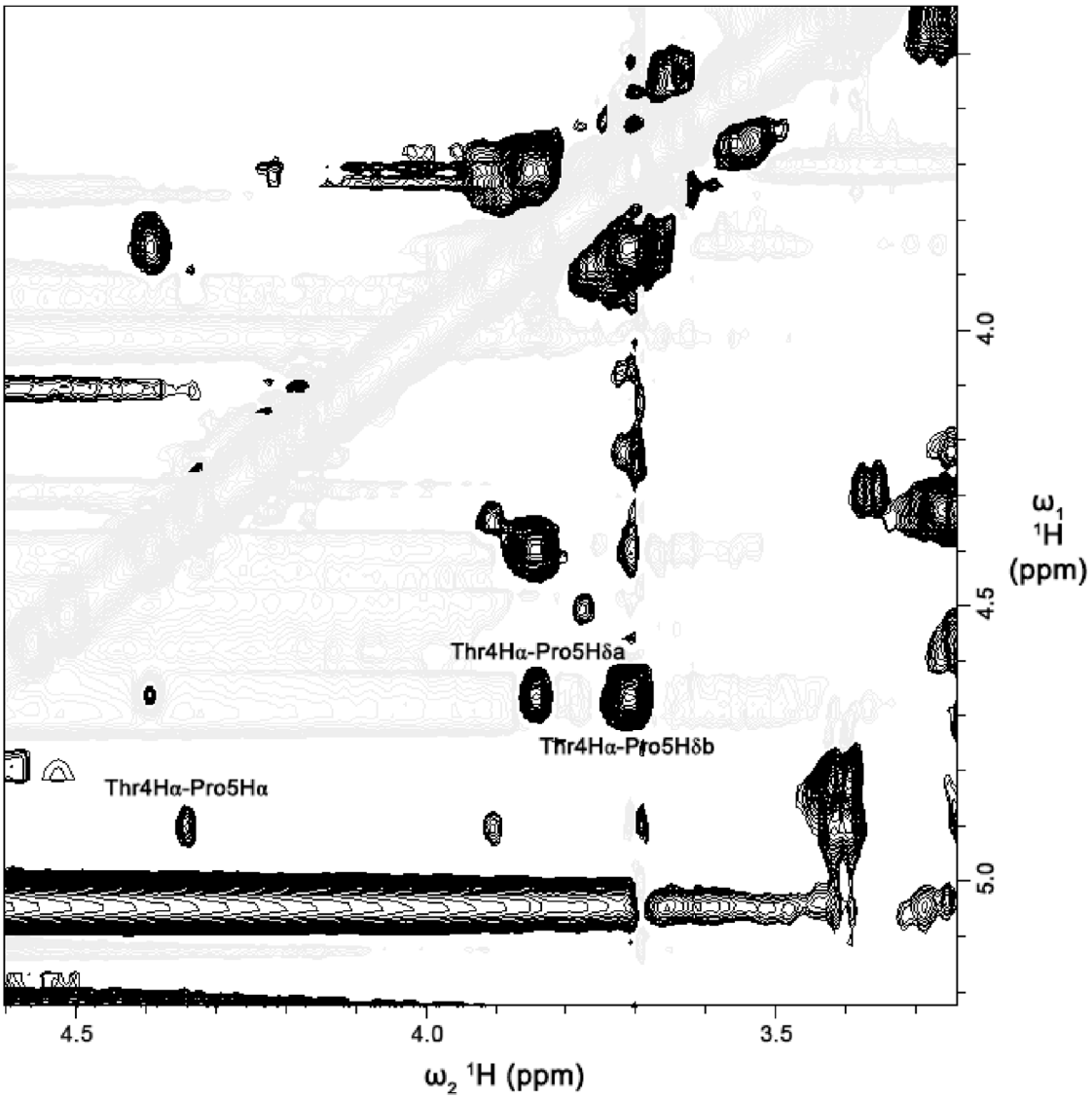


Figure S3. Section of 2D homonuclear ^{1}H ROESY spectrum, showing diagnostic NOE cross peaks indicating *cis*/*trans* isomerism of Thr4-Pro5 peptide bond. The strong Thr4Hα-Pro5Hδ cross peak indicates this bond is *trans* in the major form. The weaker Thr4Hα-Pro5Hα cross peak confirms a *cis* bond is present in the minor conformer.

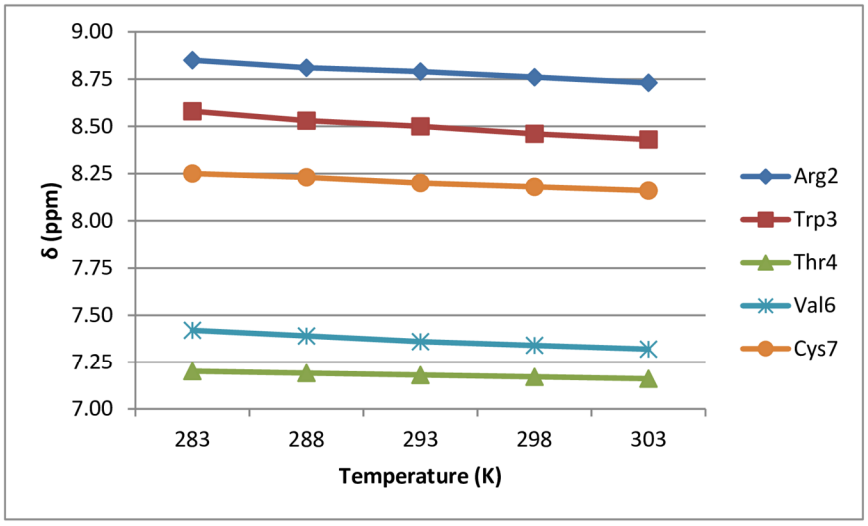


Figure S4. Plot of backbone amide chemical shifts against temperature. The gradients of the lines correspond to amide temperature coefficients.

Contryphan-Vc2 is stable at physiological temperatures

Spectra recorded on [D-Trp3]-contryphan-Vc2 at 37 °C (**Figure S5**) showed amide chemical shifts consistent with those expected from the linear temperature coefficients previously determined over the range 10-30 °C (**Figure S4**). A disruption in structure would disrupt the linearity of these relationships. Additionally, the aliphatic peaks of the spectra exhibited only minor shifts when compared at 10 °C and 37 °C (**Figure S5**), as would be expected for a stable structure. These results confirm that [D-Trp3]-contryphan-Vc2 maintains it structure at physiological temperatures.

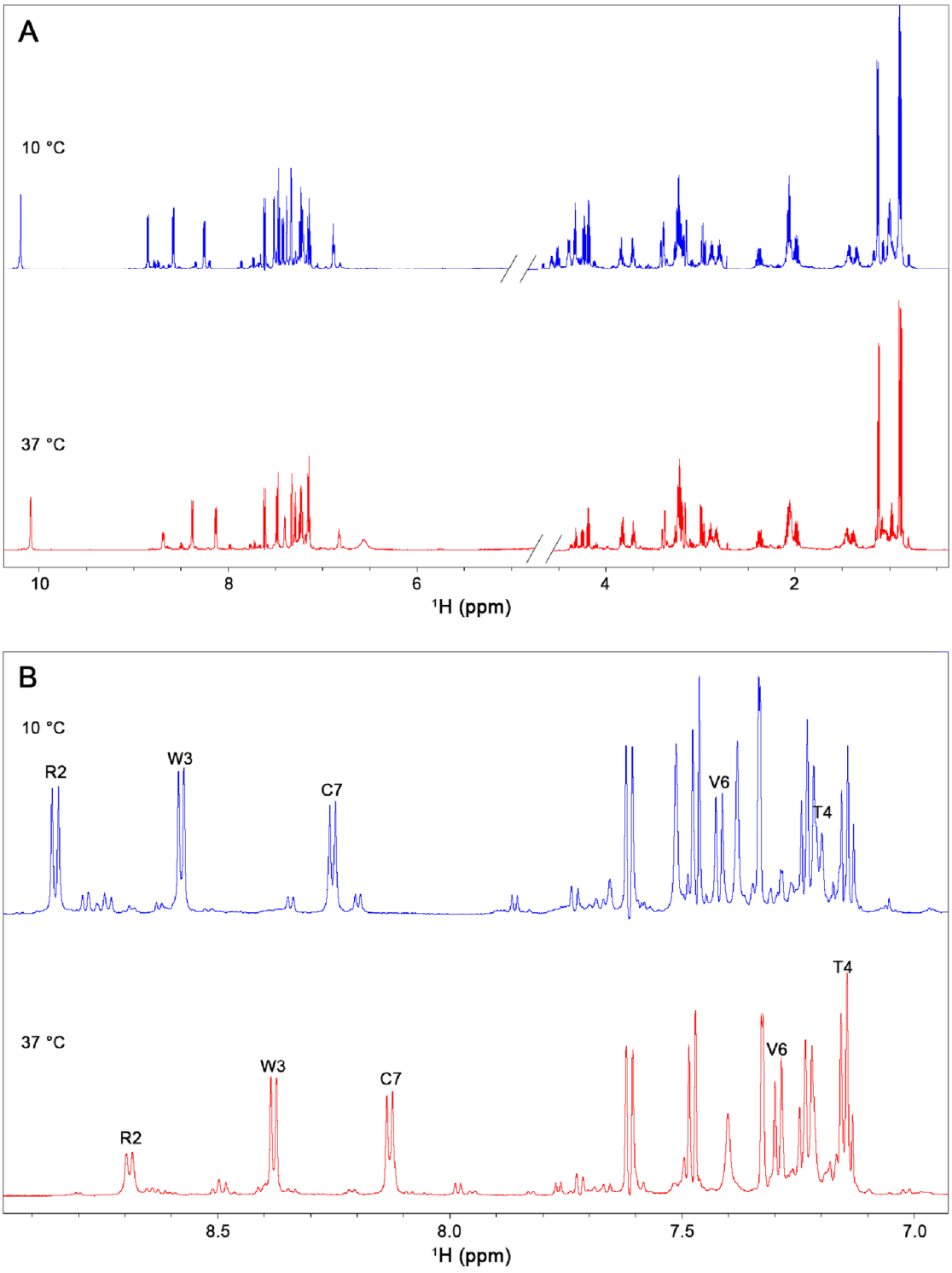


Figure S5. Comparison of NMR spectra of [D-Trp3]-contryphan-Vc2 recorded at 10 °C and 37 °C. (A) Full spectra, showing similarity in aliphatic region. The region of the spectrum in the vicinity of the water resonance has been truncated. (B) Amide region, showing shifts in amide peaks that are consistent with the linear temperature coefficient values measured over the range 10 to 30 ºC (see main text and **Figure S4**), indicating that there is no conformational change over this temperature range.

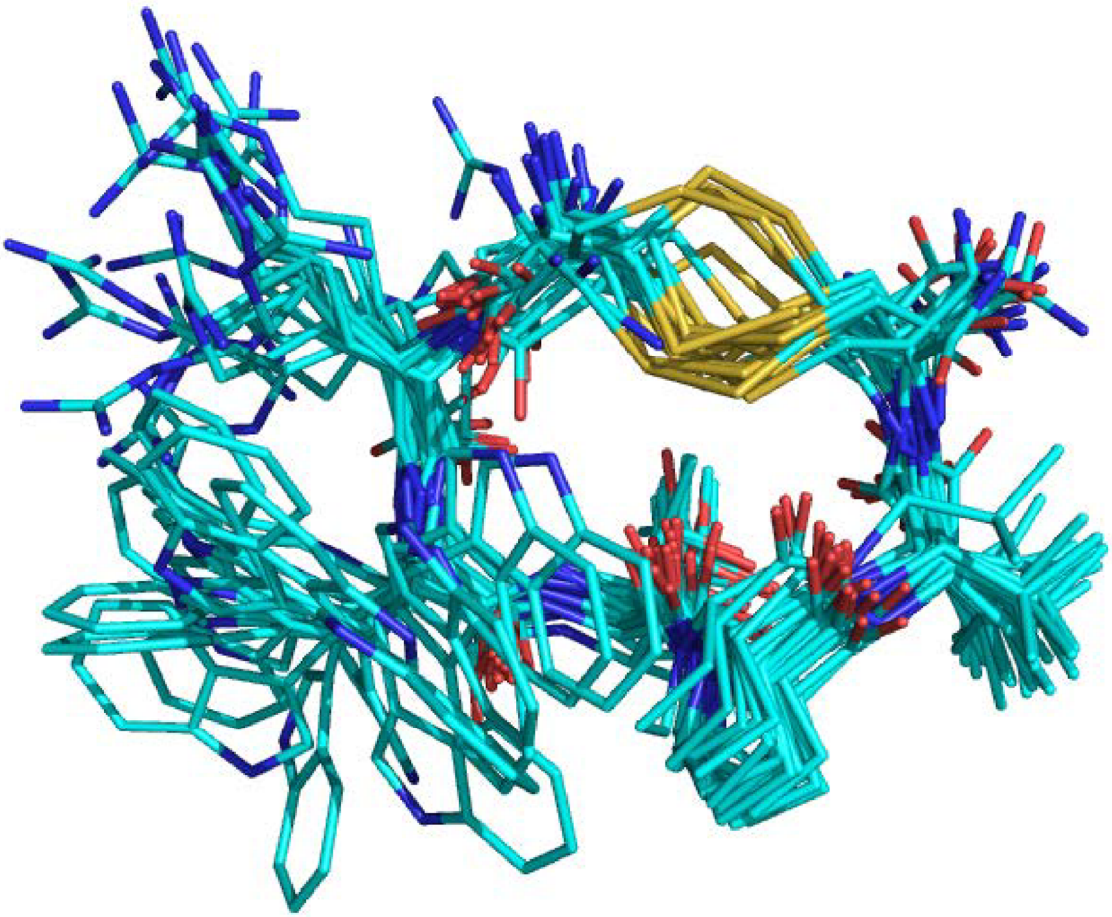


Figure S6. Aligned frames of MD simulation of [D-Trp3]-contryphan-Vc2 in water. Frames taken along length of trajectory at 5 ns intervals. Backbone is largely static, with some deviation at the disulfide bond and notable flexibility in sidechains of Arg2 and Trp3.

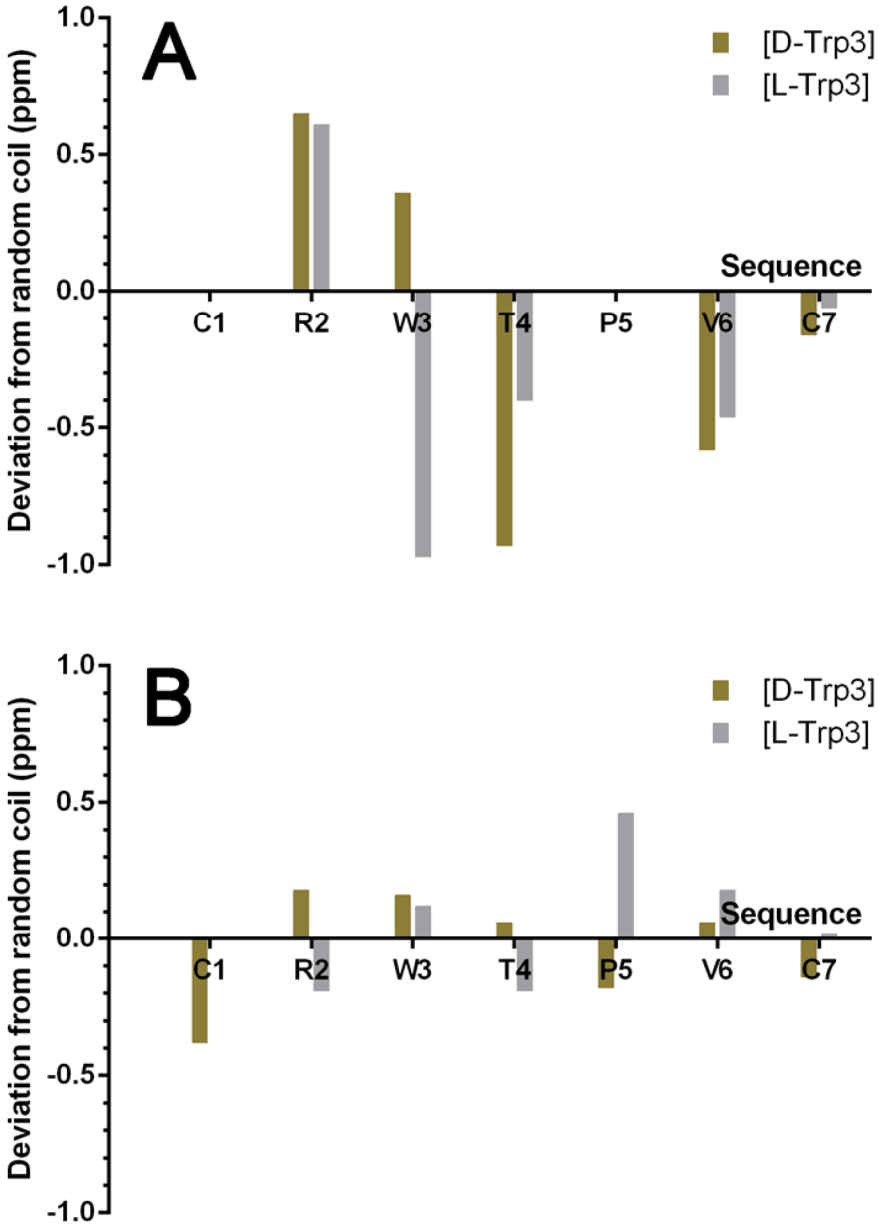


Figure S7. Deviation from random coil chemical shift plots for [D-Trp3]- and [L-Trp3]-contryphan-Vc2. Positive values indicate downfield shifts. (A) Amide proton resonances. (B) Hα proton resonances.

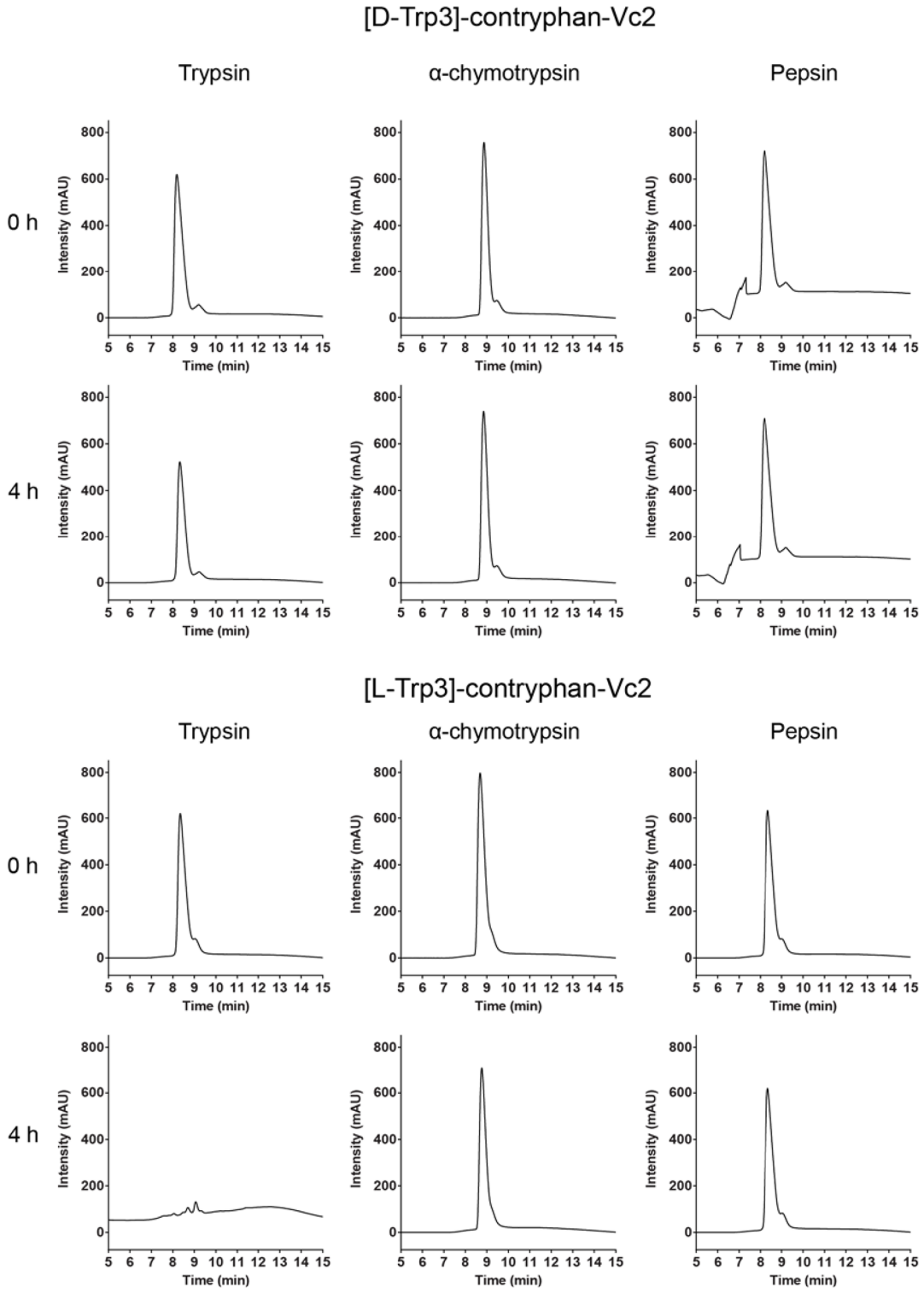


Figure S8. LC-MS traces of contryphan-Vc2 samples showing level of susceptibility to digestive enzymes after 4 h incubation at 37 °C.

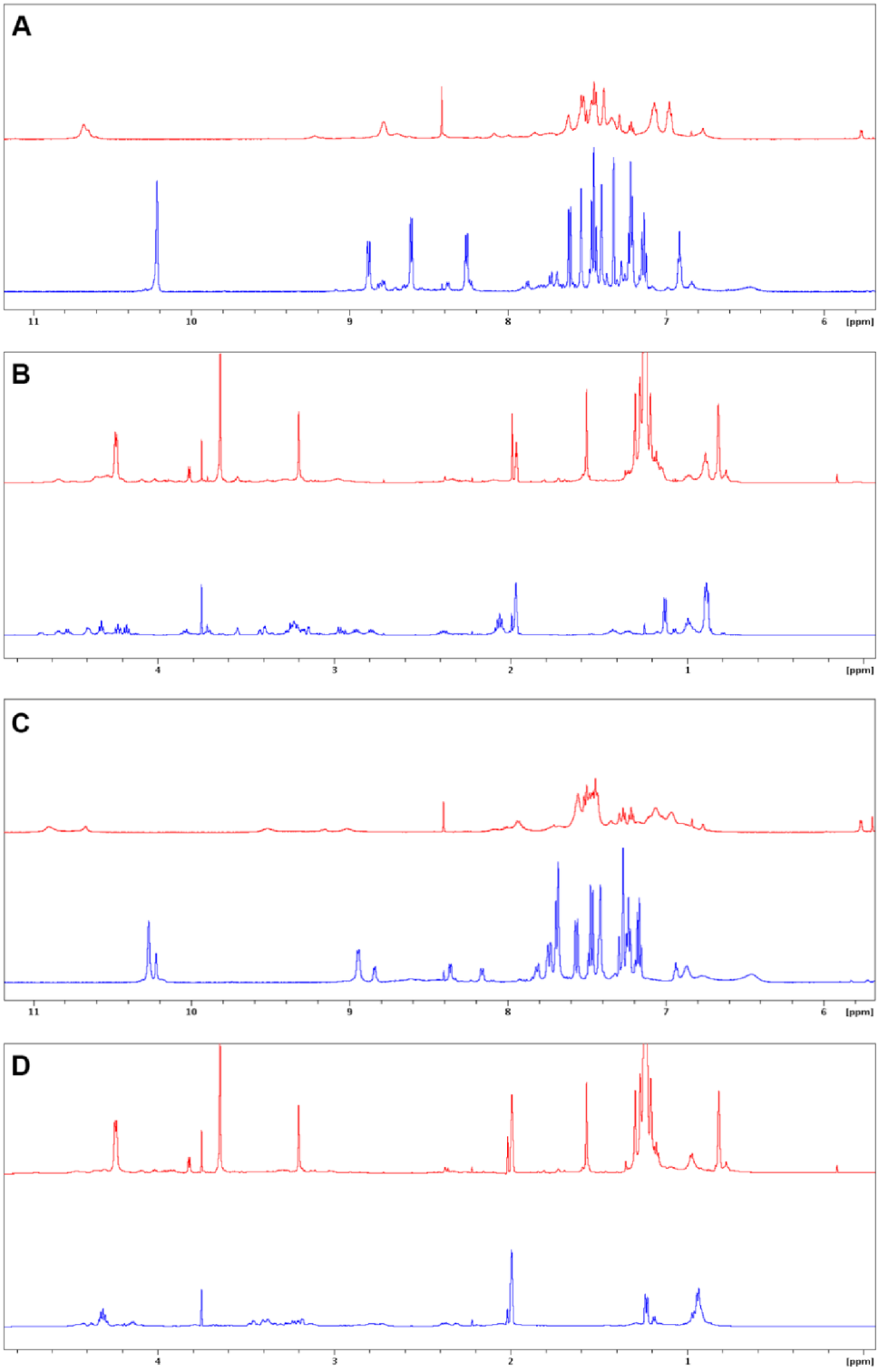


Figure S9. ^{1}H NMR spectra of [D-Trp3]- and [L-Trp3]-contryphan-Vc2 in absence (blue) and presence (red) of 40 mM DPC. (A) [D-Trp3]-contryphan-Vc2 amide region (B) [D-Trp3]-contryphan-Vc2 aliphatic region (C) [L-Trp3]-contryphan-Vc2 amide region (D) [L-Trp3]-contryphan-Vc2 aliphatic region.

Molecular dynamics simulations

MD simulations were run to gain insight into specific molecular interactions between peptides and lipid bilayers. Simulations of [L-Trp3]- and [W3A]-contryphan-Vc2 in the presence of mixed POPC/POPG bilayers were run for 2 μs, while simulations of [D-Trp3]-contryphan-Vc2 were run for 3 μs. Pairs of simulations were set up, varying in the initial position of the peptide. 'Aqueous' simulations started with the peptide in the water phase, while 'buried' simulations began with the peptide within the bilayer. **Figure S10** plots the depth of insertion of the peptide into the bilayer, comparing this value between the pairs of simulations. None of the pairs of simulations had converged after 2 μs, and [D-Trp3]-contryphan-Vc2 had not converged within the extended 3 μs timeframe either. The first 1 μs of each MD trajectory was designated as equilibration time and discounted from further analysis; comparisons and inferences were drawn from the remaining portion of the trajectory.

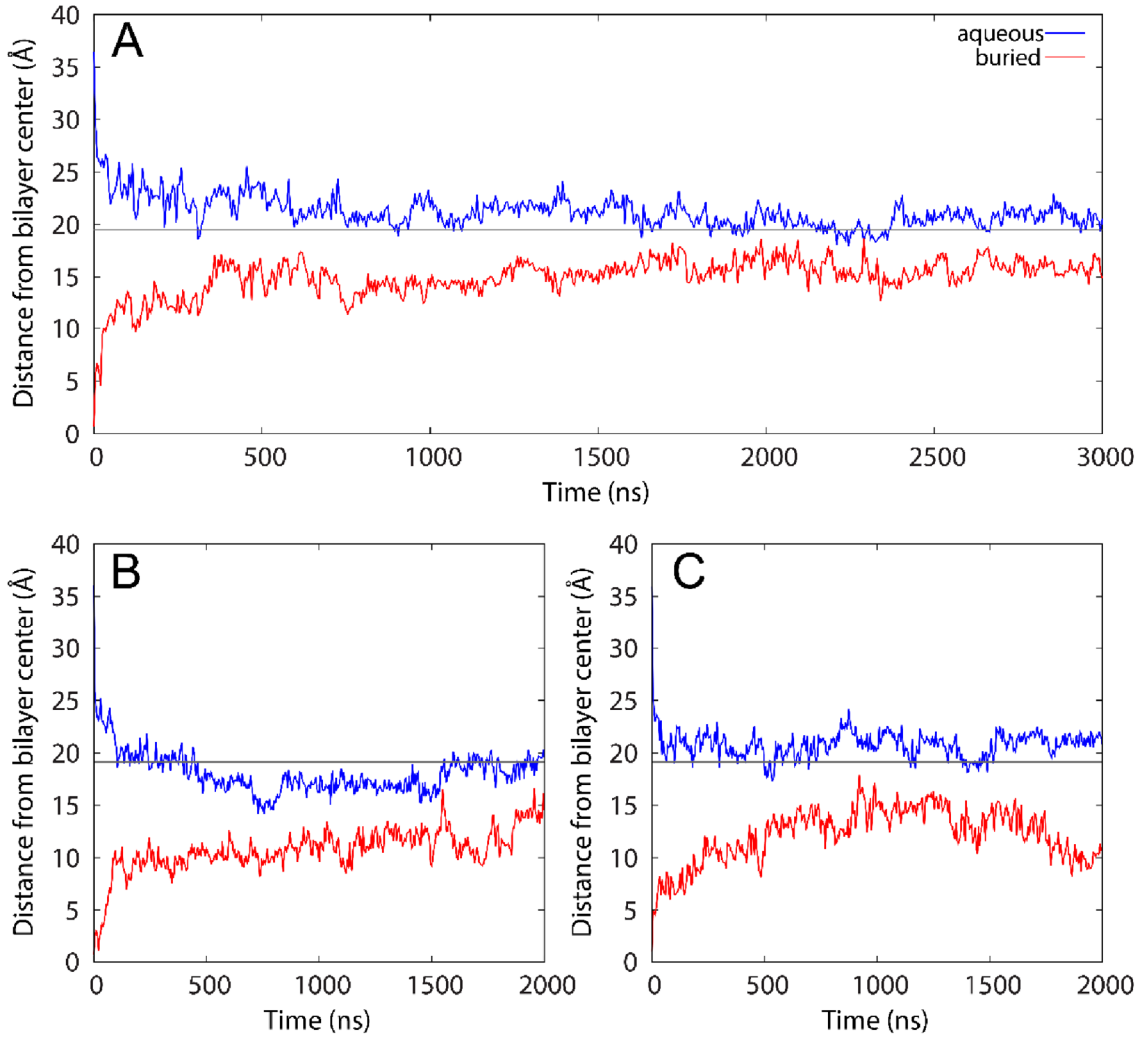


Figure S10. Insertion depth of contryphan-Vc2 centre of mass in lipid bilayer, calculated from trajectories for aqueous (*blue*) and buried (*red*) molecular dynamics simulations. *Grey* line indicates position of head group phosphate. (A) [D-Trp3]-contryphan-Vc2. (B) [L-Trp3]-contryphan-Vc2. (C) [W3A]-contryphan-Vc2.

The solvent-accessible surface area of each residue was calculated as a measure of the extent of interaction with the lipid. Plots for [D-Trp3]-contryphan-Vc2 are presented in **Figure S11**, while average values for all simulations are shown in **Table S5**.

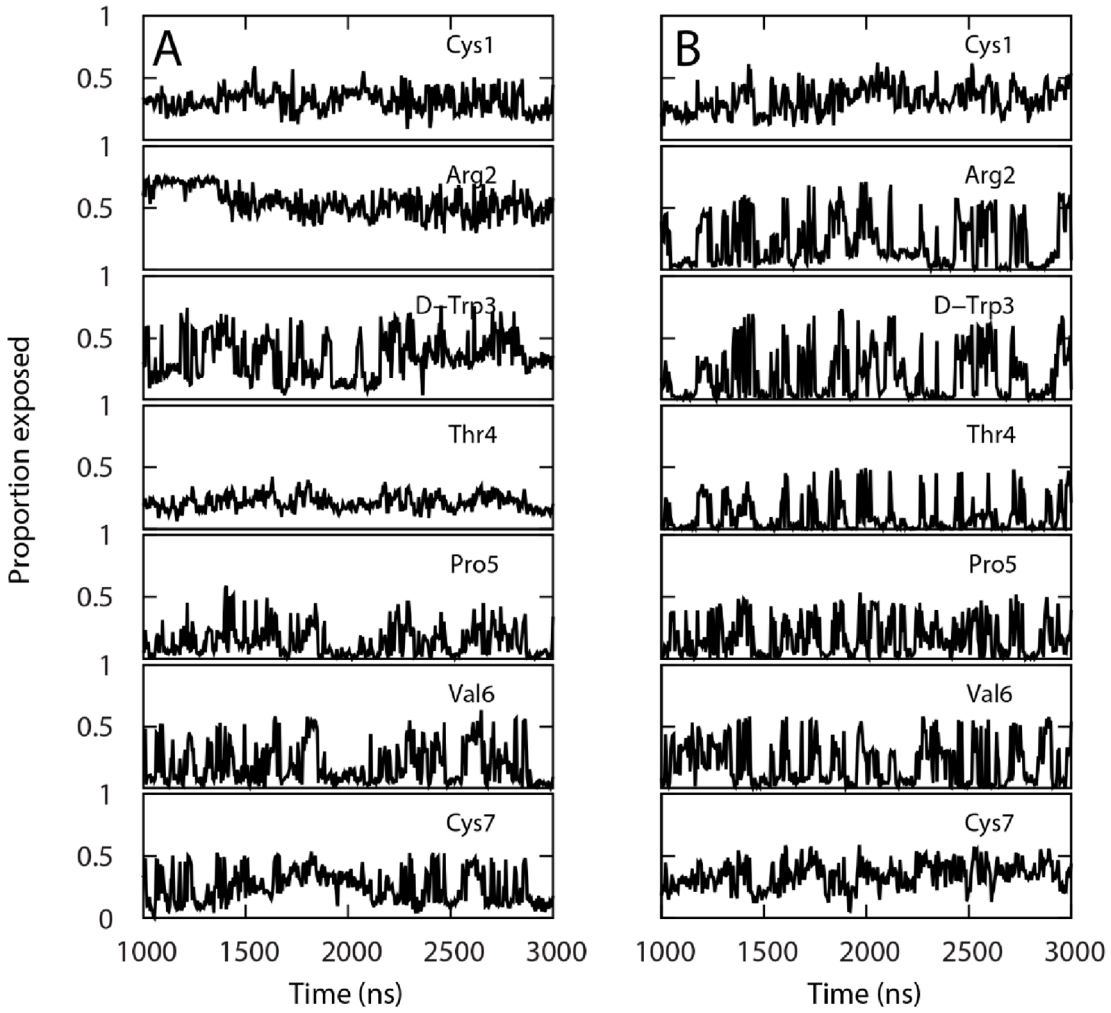


Figure S11. Solvent-accessible surface area of [D-Trp3]-contryphan-Vc2, by residue, calculated from molecular dynamics simulation trajectories. (A) Data from simulation with peptide starting position in aqueous phase. (B) Data from simulation with peptide starting position buried in bilayer.

Table S5. Average values of solvent-accessible surface area (percentage of total area) for each residue, calculated across length of MD simulation.

	Simulation					
	[D-Trp3]		[L-Trp3]		[W3A]	
Residue	aqueous	buried	aqueous	buried	aqueous	buried
Cys1	31	33	14	18	19	21
Arg2	55	22	24	30	13	38
Xaa3	35	20	28	18	8	15
Thr4	22	10	8	13	6	8
Pro5	14	17	42	25	36	17
Val6	19	20	38	22	36	15
Cys7	25	34	19	18	37	22

Thr4 is the most buried residue in all but one simulation (that being 'aqueous' [D-Trp3]-contryphan-Vc2). There is also no clear distinction between the relative solvent exposure of [D-Trp3]-, [L-Trp3]- and [W3A]-contryphan-Vc2, although Ala3 in [W3A]-contryphan-Vc2 is less accessible than D/L-Trp3 in the other peptides.

Thesis outcomes
and future directions

1 Overview

Peptide toxins can be harnessed for a variety of useful purposes, from therapeutic lead compounds to tools and probes for analysing ion channels to components for protein engineering. To generalise, these applications are facilitated by the high potency and exquisite selectivity of many peptide toxins for their target receptors and further aided by their remarkable stability in solution. Determining which applications a particular toxin may be useful for requires a detailed study of that toxin, to establish properties such as activity, structure and molecular target. This thesis investigated the scorpion toxin HsTX1 as well as novel toxins identified from the venom duct transcriptome of the cone snail *Conus victoriae*, to determine what utility these peptide toxins may possess.

2 Scorpion toxin HsTX1

The scorpion toxin HsTX1 has utility as a therapeutic lead for autoimmune diseases owing to its potent inhibition of the $K_v1.3$ voltage-gated potassium channel.[1] In addition, the molecule demonstrates remarkable stability towards thermal and proteolytic degradation. This project evaluated HsTX1 on the basis of this stable structure rather than its potassium channel blocking function, assessing its potential utility as a scaffold for peptide engineering. In addition to the stability of the toxin, previous work on the structure showed that it folded readily into the correct three-dimensional conformation following synthesis.[1] These properties are both hallmarks of good scaffold molecules; the other requirement is the potential to accept a graft, a property of HsTX1 that was heretofore untested. Validating scaffolds for molecular grafting widens the selection of molecules available to peptide engineers and increases the likelihood that a suitable scaffold can be found for their application. The cysteine-stabilised α-helical and β-sheet (CSαβ) fold found in HsTX1 has not yet been validated in the same way as the inhibitory cystine knot (ICK), cyclic cystine knot (CCK) and cyclic cystine ladder (CCL) folds have.

Molecular dynamics (MD) simulations were used to design grafted constructs based on HsTX1 as a scaffold, using the integrin binding motif RGD and the iNOS-SPSB2 interaction recognition sequence DINNN as grafts. The goal was to identify constructs that retained the stability of the toxin. Selected constructs were produced by solid-phase peptide synthesis, but failed to fold to the stable conformation. All of the MD simulations were based on the assumption that the molecules would attain the folded state and gave no insight on the folding pathway. The failure to fold was attributed to the increased length of the loop

chosen as the insertion site; even a seven-residue graft in place of the four-residue native loop led to a failed folding reaction. Although methods exist that could ensure the correct fold is adopted, such measures entail extra levels of complication in the production of the hybrid and make HsTX1 less attractive as a scaffold. In addition, there are no other readily apparent insertion points for arbitrary graft sequences, although the α-helix and β-sheet elements are probably amenable to homology grafting, as has been demonstrated with other CSαβ toxins. Despite the impressive stability of the native toxin, complexities in folding modified forms of HsTX1 make it of marginal utility as a general scaffold.

2.1 Future directions

This investigation suggests that the precise length of the insertion loop is critical for the stability of HsTX1, perhaps facilitating the bonding of the flanking Cys residues with their correct partners and proper formation of the disulfide bond network. If more quantitative data were desired on optimal loop length, a future experiment could be run with several loop analogues – shortened (by deleting residues), or lengthened (by adding Ala residues). Refolding each mutant under the conditions optimised for native HsTX1 would then give an indication of the constraints on loop length consistent with proper folding.

A more extensive investigation could analyse the behaviour of several disulfide-deleted analogues, in the same manner as Price-Carter et al. investigated the folding of ω-conotoxins.[2-5] Probing the propensity for native and non-native disulfide bonds to form can shed light on the folding pathway of a molecule. For the conotoxin ω-MVIIA-Gly, which folds at about 80% efficiency, it was found that non-native bonds formed just as readily as native bonds in the one- and two-disulfide stage.[5] However, species with any two native disulfide bonds displayed great cooperativity in forming the third. It would be interesting to determine the cooperativity of bond formation in HsTX1. Such data would potentially aid a future attempt to engineer a peptide based on this toxin.

3 H-superfamily toxins

The recent elucidation of the venom duct transcriptome of *C. victoriae* afforded an opportunity to study peptide toxins with novel sequence, which carried with it the potential to uncover novel structures and activities. This knowledge gap was initially addressed by investigating two members of the H-superfamily identified in the transcriptome.

Recombinant expression of the toxins H_Vc7.1 and H_Vc7.2 as thioredoxin fusions was successful and cleavage of the thioredoxin fusion partner allowed the recovery of the linear peptides. Unsurprisingly, these triple-disulfide-bonded toxins were not correctly folded in the reducing environment of the recombinant bacterial cytoplasm. Unfortunately, *in vitro* refolding trials also failed to find conditions that favoured the formation of a single major product. As such, no sample could be generated of sufficient purity for investigation. The structure and activity of the toxins, as well as their potential utility, remain unknown.

3.1 Future directions

An avenue that might prove fruitful for future investigations of these toxins is expression in the oxidative environment of the bacterial periplasm, which has proved successful for the production of other multiply-disulfide-bonded peptides in the literature.[6] This would involve re-engineering the expression plasmid used to transform the bacteria, adding a periplasmic export sequence such as the *MalE* signal sequence.[7] After induction, disrupting the bacterial outer membrane would allow recovery of the peptides without completely lysing the cells, which would also minimise the exposure of the peptides to proteases.

If a pure, folded sample could be produced, it would be important to validate that the conformation produced in the periplasm matches that found in the source venom. One established method for achieving this is co-elution of the folded sample and crude venom on RP-HPLC. If the disulfide network of the folded sample differed from that of the venom-derived peptide, this would manifest as an altered retention time on the reverse-phase column. Such a result would be an indication that the folded conformations differed, but unfortunately not give any information on what differences caused the discrepancy.

An alternative to periplasmic expression would be to produce the H-superfamily toxins by solid-phase peptide synthesis (SPPS). The three pairs of Cys sidechains could be protected with orthogonal protecting groups, allowing the disulfide bond network to be constructed by directed folding. This technique requires knowledge of which Cys residues participate in which disulfide bond, which in this case could probably be inferred from homology with other framework VI/VII toxins. It would be very important to validate the conformation by co-elution in this case to confirm the assumed disulfide bonds were indeed formed and correct.

If a pure sample could be obtained, by these or other methods, the next step would be to characterise the peptides at the structural and functional level. This would then reveal if the

peptides possess interesting or novel bioactivity and establish the potential utility of these molecules.

4 Contryphan-Vc2

Another novel peptide arising from the venom duct transcriptome of *C. victoriae* was contryphan-Vc2, which was investigated following production by SPPS. NMR spectroscopy generated data that revealed the presence of two conformers, differing by *cis*/*trans* isomerism around the Thr4-Pro5 bond. Analysis of nuclear Overhauser effect (NOE) crosspeaks enabled a three-dimensional structure of the major conformer to be determined. Interestingly, when this single NMR structure was used as the basis for MD simulations, the peptide alternated between two different conformations, neither of which was that determined by NMR. The NMR data identified the Trp3 sidechain as being in close proximity to both the Arg2 and Pro5 sidechains (based on NOE crosspeaks), a feature reflected in the structure calculated from these data. Collectively, the poses revealed by MD also satisfied this requirement. In one pose, the Trp3 sidechain was in close proximity to the Arg2 sidechain but distant from Pro5. In the other pose, a rotation around the χ_1 angle of Trp3 brought the sidechain into close proximity with the Pro5 sidechain, but distant from Arg2. In both cases the *trans* conformation of the Thr4-Pro5 bond was maintained. Thus the constraints derived from the NMR data were fulfilled, but not by a single conformation of the peptide. This example highlights that NMR spectroscopy can identify some evidence of multiple conformations for a peptide in solution (such as *cis*/*trans* isomerism), while complementary techniques or careful analysis may be required to expose others. Sometimes anomalous NMR signals can give a clue that more than one conformation is present, as in an investigation of a model helical peptide conducted some years ago which detected the expected NOEs consistent with a helical structure, but also some non-helical crosspeaks. The sum total of NOE signals recorded could not be satisfied by a single, static model, indicating the presence of an ensemble of structures in solution.[8] In other cases, new information can shed light on initially misinterpreted data. A study on the amyloid $A\beta_{21–30}$ fragment cast doubt on a previously suggested structure for this peptide by virtue of recording data at a higher field strength (900 MHz vs 500 MHz). Rather than the previously proposed single ‘bent’ structure, which depended on a sole long-range NOE cross-peak that turned out to be incorrectly assigned, this study asserted instead that the peptide was largely disordered and sampled a range of conformations. The authors also made the point that

calculating structure based on multiple weak medium-range NOE peaks may be misleading, as the weakness of the peaks could be due to the occupancy rate of the interaction, rather than the distance between interacting protons.[9] NMR spectra recorded for single conformations (or very similar ensembles) are generally sharp and well-dispersed, while varied ensembles give rise to broad peaks and overlapped signals, which can be a warning sign not to attempt to calculate a single structure. The spectra of contryphan-Vc2 fit in the first category, making the discovery of multiple conformations surprising.

Bioactivity of contryphan-Vc2 was assessed by intracranial injection in mice and revealed a depressive phenotype. This result was intriguing, as other members of the contryphan family tested in this manner have induced a hyperactive phenotype. Furthermore, the depressive activity of contryphan-Vc2 was evident in analogues with Trp3 in both the D-handed and L-handed forms, while a W3A Ala mutant showed no activity. This suggests that Trp3 is important for activity, but the chirality of this residue is not. In turn, this may suggest that the activity of contryphan-Vc2 is not due to an interaction with a protein receptor. Further experiments and MD simulations provided evidence that the peptide may rather interact with the membrane and produce its effects in a membrane-mediated manner. If this is the case for other contryphans, it may help explain why the identity of the molecular target of this class of peptides has remained elusive.

Whether contryphan-Vc2 has any utility as a tool or lead compound would depend on the precise effects it exerts on the membrane. The current study merely demonstrated that it does have an effect; further work is needed to more completely answer this question.

4.1 Future directions

Studies focused on the action of contryphan peptides on membranes, rather than protein receptors, may elucidate the precise effect these molecules exert on the lipid bilayer. Studying the secondary effects of this action could then explain how the peptides generate the distinctive phenotype seen in mice – and perhaps also explain why contryphan-Vc2 differs from other contryphans in this regard.

The dynamic behaviour in the peptide revealed by MD simulations was unexpected and may suggest that other members of the contryphan family have the ability to adopt multiple conformations in solution as well. MD simulations based on previously published structures would go some way to answering the question of whether other contryphan molecules are similarly dynamic in solution.

5 Conclusion

Novelty is a prized quality among research subjects, as anything that is currently not known has the potential to be different from everything that currently is known, opening the way for new discoveries and new applications. This potential is not always fulfilled – many novel compounds are eventually determined to be minor variations on already established themes – but the only way to find out is to perform the investigation.

This thesis attempted to validate a novel scaffold for peptide grafting and also investigate some of the novel sequences previously discovered in the *C. victoriae* venom duct transcriptome. While difficulties with peptide folding frustrated some of these efforts, the conclusions from this body of work have gone some way to addressing the identified knowledge gaps, especially with regard to contryphan-Vc2.

6 References

[1] Rashid, M. H., Huq, R., Tanner, M. R., Chhabra, S., Khoo, K. K., Estrada, R., Dhawan, V., Chauhan, S., Pennington, M. W., Beeton, C., Kuyucak, S., and Norton, R. S. (2014) A potent and Kv1.3-selective analogue of the scorpion toxin HsTX1 as a potential therapeutic for autoimmune diseases, *Sci. Rep. 4*, 4509.

[2] Price-Carter, M., Gray, W. R., and Goldenberg, D. P. (1996) Folding of ω-conotoxins. 1. Efficient disulfide-coupled folding of mature sequences in vitro, *Biochemistry 35*, 15537-15546.

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

[4] Price-Carter, M., Hull, M. S., and Goldenberg, D. P. (1998) Roles of individual disulfide bonds in the stability and folding of an ω-conotoxin, *Biochemistry 37*, 9851-9861.

[5] Price-Carter, M., Bulaj, G., and Goldenberg, D. P. (2002) Initial disulfide formation steps in the folding of an ω-conotoxin, *Biochemistry 41*, 3507-3519.

[6] Klint, J. K., Senff, S., Saez, N. J., Seshadri, R., Lau, H. Y., Bende, N. S., Undheim, E. A. B., Rash, L. D., Mobli, M., and King, G. F. (2013) Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the periplasm of *E. coli*, *PLoS One 8*, e63865.

[7] Bassford, P. J. (1990) Export of the periplasmic maltose-binding protein of *Escherichia coli*, *J. Bioenerg. Biomembr. 22*, 401-439.

[8] Merutka, G., Morikis, D., Bruschweiler, R., and Wright, P. E. (1993) NMR evidence for multiple conformations in a highly helical model peptide, *Biochemistry 32*, 13089-13097.

[9] Fawzi, N. L., Phillips, A. H., Ruscio, J. Z., Doucleff, M., Wemmer, D. E., and Head-Gordon, T. (2008) Structure and dynamics of the $A\beta_{21–30}$ peptide from the interplay of NMR experiments and molecular simulations, *J. Am. Chem. Soc. 130*, 6145-6158.