PORTLAND PRESS

Research Article

Structural characterisation of a cysteine-rich conotoxin, sigma(σ)S-GVIIIA, extracted from the defensive venom of the marine cone snail *Conus geographus*

Yoshimi Peck[1], David T. Wilson[1], Danica Lennox-Bulow[1], Julien Giribaldi[2], Jamie Seymour[1], Sebastien Dutertre[2], K. Johan Rosengren[3], Michael J. Liddell[4] and Norelle L. Daly[1] ID

[1]Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Queensland, Australia; [2]IBMM, University of Montpellier, CNRS, ENSCM, Montpellier, France; [3]School of Biomedical Sciences, The University of Queensland, Brisbane, Queensland, Australia; [4]College of Science and Engineering, James Cook University, Cairns, Queensland, Australia

Correspondence: Norelle L. Daly (norelle.daly@jcu.edu.au)

OPEN ACCESS

The activity of the serotonin type 3 (5-HT$_3$) receptor is associated with neurodegenerative, inflammatory and metabolic diseases, neuropsychiatric disorders and cancer. Structural analysis of modulators of this receptor is likely to aid in future medicinal chemistry studies aimed at developing lead molecules targeting this receptor. Here, we report the structure of a cone snail venom peptide that was purified from the crude venom of *Conus geographus* and shown to be an antagonist of the 5-HT$_3$ receptor more than 25 years ago, sigma(σ)S-GVIIIA. This lag in structural characterisation studies is likely due to challenges in isolating the native peptide and difficulties in producing synthetic peptide due to the presence of ten cysteine residues involved in five disulfide bonds. Using NMR spectroscopy, we show that σS-GVIIIA adopts a growth factor cystine knot (GFCK) fold. This is the first example of a cone snail venom peptide experimentally determined to contain the GFCK structural motif and the first example of a 5-HT$_3$ receptor antagonist containing this motif. Our study also highlights complexities in the use of artificial intelligence (AI)-based structure prediction models. Peptide structure predictions using AlphaFold 3 were consistent with our NMR structure when the input sequence contained the well-conserved precursor sequence but inconsistent when the precursor sequence was excluded. AI-based structure prediction of proteins is a rapidly advancing field, but this inconsistency emphasises the need for more experimental structural training data when novel structures are involved, as was the case here for a cysteine-rich peptide.

Introduction

Cone snails (*Conus* spp.) are predatory marine gastropods found predominantly in tropical or sub-tropical regions [1]. Despite their slow movement, they can immobilise prey and predators quickly by injecting a potent venom [2]. The venom has been found to be a cocktail of pharmacologically active molecules dominated by cysteine-rich peptides referred to as conotoxins [3].

Conotoxins are predominantly neurotoxins that act on ion channels and membrane receptors in the nervous system [4]. They generally have well-defined structures braced by disulfide bonds [3], which confer enhanced chemical and enzymatic stability [5–7]. Conotoxins are also generally potent and specific ligands to their target receptor [8], making them useful pharmacological tools and attractive drug leads for a range of applications including the treatment of pain, neuropsychiatric conditions, cardiac diseases and diabetes [9,10].

Although many conotoxins have been structurally and functionally characterised, they still represent only a small proportion of the more than 80,000 conotoxins estimated to be present in the more than 700 known *Conus* species [11–13]. Characterisation of novel conotoxins can be challenging for several reasons including structural complexity and low abundance in venom samples. Solid-phase peptide synthesis can be used to generate sufficient quantities of low-abundance peptides; however, for cysteine-rich conotoxins, the formation of the correct structural fold *in vitro* can be problematic without prior knowledge of the native disulfide pairing. For example, a conotoxin with ten cysteine residues can have 945 possible disulfide

connectivities [14,15]. Thus, knowledge of the disulfide bond connectivities of cysteine-rich conotoxins is often required to allow the directed formation of the native disulfide bonds during *in vitro* synthesis.

σS-GVIIIA is one of the many conotoxins that has not been well characterised, particularly in terms of structure. It has only been found in the defensive venom of *Conus geographus* and consists of 41 amino acids and contains multiple post-translational modifications, including a bromotryptophan, hydroxy-proline, C-terminal amidation and five-disulfide bonds [16]. It is classified into the S-gene superfamily, has a type VIII cysteine framework and its competitive antagonism at the serotonin type 3 ($5HT_3$) receptor makes it the first, and to our knowledge, currently the only member of the σ pharmacological family.

The 5-HT_3 receptor has been recognized as a clinically important target because its activity has been implicated in substance addiction, clinical depression, cancer, diabetes and inflammatory and neurodegenerative diseases [17–20]. The low abundance of σS-GVIIIA in the venom and highly cysteine-rich nature has made it challenging to work with and has contributed to the limited follow-up studies after the initial identification of the pharmacological activity of this peptide over two decades ago [16]. Consequently, there have been no structural studies, and the disulfide connectivity has not been determined.

In the current study, we have isolated native σS-GVIIIA from milked defensive venom to identify the native three-dimensional (3D) fold and cysteine connectivities using nuclear magnetic resonance (NMR) spectroscopy. We found that σS-GVIIIA is the first conotoxin to contain the growth factor cystine knot (GFCK) motif. The GFCK is a structural motif initially identified in multiple growth factor proteins approximately three decades ago [21]. No structural studies for invertebrate venom peptides containing the GFCK motif have been reported to our knowledge, in contrast with the common inhibitor cystine knot (ICK) [22,23]. σS-GVIIIA is also the first example of a 5-HT_3 receptor antagonist containing the GFCK motif.

To explore the reliability of modern artificial intelligence (AI)-based structure prediction models, we compared our experimental structure of σS-GVIIIA with the predicted structure from AlphaFold 3 and also predicted the structure of σS-GVIIIA in complex with the 5-HT_3 receptor. The aim of these analyses was to provide insight into the strengths and limitations of AI predictions and potential insight into pharmacophore development for the 5-HT_3 receptor.

Results

The molecular profile of *C. geographus* venom in response to offensive and defensive stimuli

Crude *C. geographus* venom was collected from multiple live specimens through defence and predation-provoked stimuli. The liquid chromatography–mass spectrometry (LC–MS) profile of each milking was obtained immediately following each venom collection, and significant variation was observed. The relative abundance of σS-GVIIIA in the crude venom fluctuated between individuals and between subsequent milkings of the same individual. To illustrate this, a comparison of the LC–MS chromatograms of crude defensive venom from two individual *C. geographus* specimens (A and B) milked on the same day is shown in Supplementary Figure S1A and B. These profiles show considerable differences, and σS-GVIIIA was generally not detected in venom samples from specimen A but was detected in the venom of specimen B. An uncharacterized 8916 Da peptide and I_3-GXIA were the most abundant components in the venom of specimen A, while μ-GIIIA conotoxin was the major component in the venom of specimen B. The relative intensity of σS-GVIIIA in specimen B also varied in subsequent milkings taken on different days (Supplementary Figure S1B and C).

Variation in venom composition between offensive and defensive venom was also observed for the venom from *C. geographus* specimen C, and σS-GVIIIA was not detected in the offensive venom but only detected in the defensive venom samples (Supplementary Figure S1D). Analysis of LC–MS profiles showed the only common mass found in both offensive and defensive venom was a 3036 Da peptide, which is predicted to be ω-GVIA.

Purification of the σS-GVIIIA conotoxin from crude venom

To collect σS-GVIIIA for NMR experiments (~0.6 mg), milkings were carried out with defensive stimuli over 36 months. The collected venom was pooled and fractionated using reversed-phase high-performance

liquid chromatography/mass spectrometry (RP-HPLC/MS) (Supplementary Figure S2). The isolated peptide was identified as σS-GVIIIA (sequence ID P58924) based on the mass of 4188/4192 Da (Supplementary Figure S3) and NMR spectroscopy analysis. The presence of two molecular masses is due to the presence of the bromotryptophan (residue 34), which has two naturally occurring isotopes, Br^{79} and Br^{81}, at a 51:49 ratio.

Disulfide bond prediction based on the preliminary NMR structure of σS-GVIIIA

Purified native σS-GVIIIA was solubilised in 90% H_2O:10% D_2O (v/v) at a final concentration of approximately 0.26 mM. Two-dimensional spectra, including ^1H–^1H TOCSY (total correlation spectroscopy), ^1H–^1H NOESY (nuclear Overhauser effect spectroscopy), ^1H–^1H DQF-COSY (double-quantum filtered correlation spectroscopy) and ^1H–^{13}C HSQC (heteronuclear single quantum coherence spectroscopy), were collected for subsequent sequence-specific assignments and structure calculations [24,25]. The signal-to-noise ratio of the ^1H–^{15}N HSQC spectrum was low due to the low concentration of the sample, and the data were not useful in the structure calculations.

Initial NMR structures of σS-GVIIIA were calculated using CYANA (combined assignment and dynamics algorithm for NMR applications) [26,27] without any disulfide bond restraints. An ensemble of 20 structures corresponding to the lowest target function is shown in Figure 1A. These structures were relatively well-defined despite the lack of disulfide bond restraints (backbone root-mean-square deviation [RMSD] of 0.84 ± 0.40 Å for the well-defined region, residues 3–5, 13–15, 25–28 and 35–38, and 1.33 ± 0.67 Å for residues 1–41) and therefore could be used to predict the most likely disulfide bonds. The prediction was based on the distance between sulfur atoms in pairs of cysteine residues. The mean, standard deviation, minimum and maximum distance of each cysteine pair are summarised in Supplementary Table S1, and the specific distances in the individual structures are shown in Supplementary Figure S4. The five cysteine pairs with the lowest mean distance corresponded to Cys^2-Cys^{17}, Cys^6-Cys^{25}, Cys^{11}-Cys^{36}, Cys^{15}-Cys^{38} and Cys^{23}-Cys^{40}, indicating that this is the most likely connectivity. Analysis of the cysteine residues that are not located in close proximity can also be informative for predicting the disulfide connectivity, and in this case, the most definitive finding from the analysis indicated that Cys^2 is only in close proximity to Cys^{17} and does not have any other option to pair based on the individual distances (Supplementary Figure S4).

Analysis of the sulfur–sulfur distances has been done in previous studies of disulfide-rich peptides using cut-off distances [28,29]. We used a cut-off distance of 3.0 Å between the sulfur atoms based on our initial analysis of the average distances. This analysis indicated that the most likely connectivity for σS-GVIIIA is Cys^2-Cys^{17}, Cys^6-Cys^{25}, Cys^{11}-Cys^{36}, Cys^{15}-Cys^{38} and Cys^{23}-Cys^{40} (Figure 1B).

Description of the overall structure and discovery of the GFCK motif in σS-GVIIIA

The final structures for σS-GVIIIA (the PDB and BMRB codes are 9EBE and 31213, respectively) were calculated with disulfide bond restraints based on the most likely connectivities predicted by the inter-sulfur distance analysis. The NMR-derived structure with the lowest target function is shown in Figure 1C, and the structure ensemble containing the 20 structures with the lowest target function is shown in Supplementary Figure S5. The structural statistics are listed in Table 1. The main elements of the secondary structure in σS-GVIIIA are four anti-parallel β-strands, forming two β-hairpins. All four β-strands, referred to as β1–β4, are not found in the same plane, but the β-hairpins are facing each other (Figure 1C and D).

Analysis of the topology of the disulfide bonds in σS-GVIIIA indicated the presence of a cystine knot motif at the core (Cys^6-Cys^{25}, Cys^{11}-Cys^{36} and Cys^{15}-Cys^{38} with two additional disulfide bonds (Cys^2-Cys^{17} and Cys^{23}-Cys^{40} located away from the core of the peptide; Figure 1C and D). For comparative purposes, we will refer to the disulfide bonds directly involved with the cystine knot as $Cys^{I'}$-$Cys^{IV'}$, $Cys^{II'}$-$Cys^{V'}$ and $Cys^{III'}$-$Cys^{VI'}$, respectively (Figure 1E). In this Roman numeral numbering system, the cysteine residues that do not contribute directly to the knot motif have been excluded. The additional disulfide bonds will still be described as Cys^2-Cys^{17} and Cys^{23}-Cys^{40}, without using Roman numerals.

The σS-GVIIIA cystine knot was compared with the ICK motif [22], which is commonly found in conotoxins and has the same disulfide connectivity Cys^I-Cys^{IV}, Cys^{II}-Cys^V and Cys^{III}-Cys^{VI} in the knot (Figure 2). The knot motif of σS-GVIIIA was not superimposable with an ICK motif containing
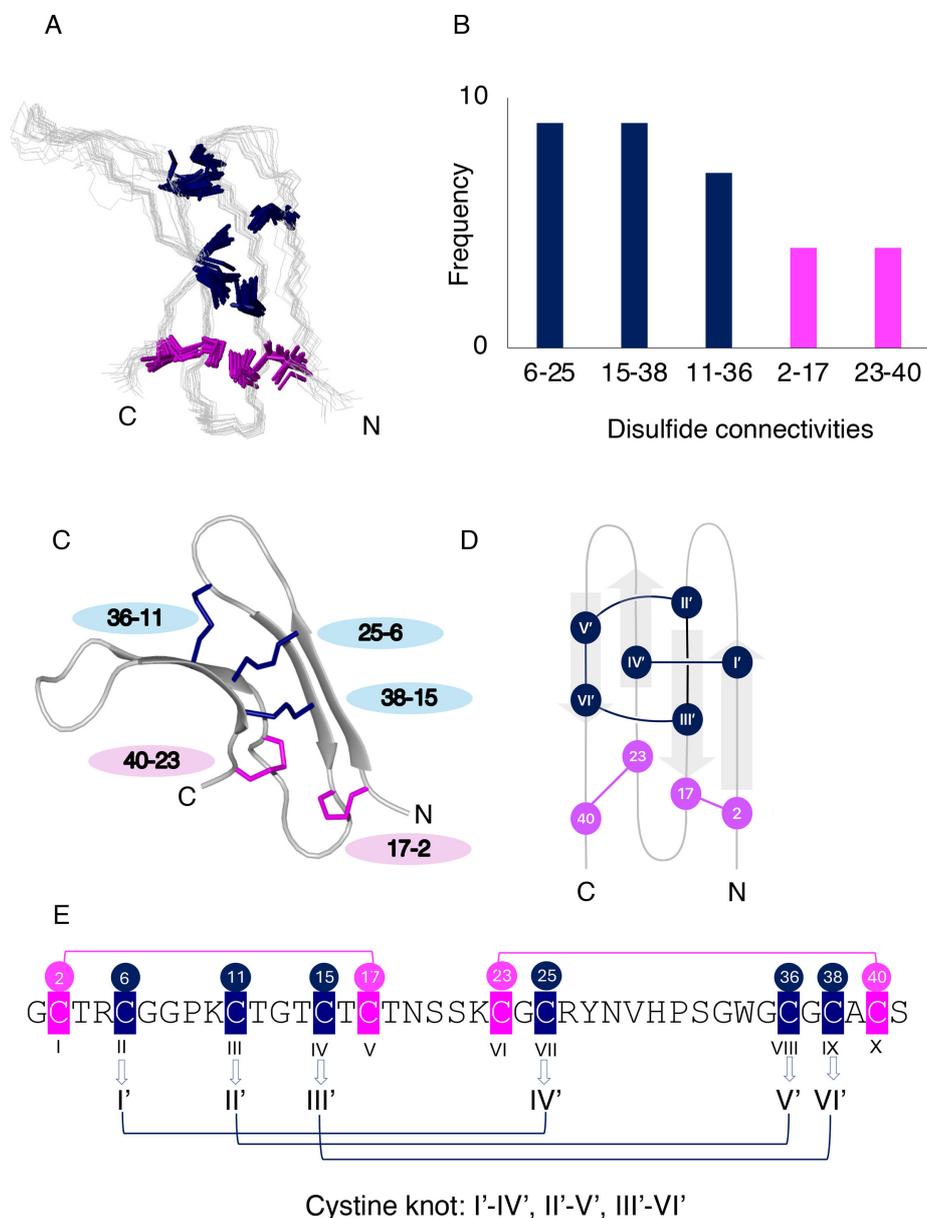
Figure 1: Determination of the structure of σS-GVIIIA.

(**A**) The initial NMR structures were calculated without disulfide bond restraints. Cysteine side chains are highlighted in navy or pink. The ensemble of 20 structures with the lowest target functions is shown. (**B**) The inter-sulfur distance analysis of the preliminary structures of σS-GVIIIA. The number of structures, out of the 20 lowest energy structures, which have particular sulfur–sulfur distances less than 3.0 Å, are plotted. (**C**) The final structure of σS-GVIIIA (PDB:9EBE). The structure with the lowest target function is shown. (**D**) The schematic figure of σS-GVIIIA. The knot-forming cysteine residues are shown in navy, whereas cysteine residues not involved in the knot are shown in magenta. (**E**) The sequence of σS-GVIIIA with re-numbered cysteine residues involved in the cystine knot. All ten cysteine residues were numbered as I–X, and then knot-forming six cysteine residues were re-numbered as I'–VI' and highlighted in navy. The cysteine residues, not involved in the knot, are highlighted in magenta.

peptide, namely, conotoxin ω-GVIA (PDB: 2CCO), also originally isolated from *C. geographus* (Figure 2). Schematic figures of ω-GVIA and σS-GVIIIA highlight the major differences, where the Cys [I']-Cys[IV'] bond threads through the ring of σS-GVIIIA, but Cys [III]-Cys[VI] threads through the ring in ω-GVIA (Figure 2). In contrast, the 3D topology of the knot motif was found to be superimposable with the GFCK motif of the human glycoprotein hormone α subunit (GPH-α) (PDB: 1HRP, Figure 3).

The signature sequence patterns of the GFCK motif have been established previously as CXGXC and CXC, where X refers to any single residue, and collectively these eight residues, including two disulfide bonds, form a ring as a part of the GFCK motif [34]. σS-GVIIIA contains the GFCK motif signature

Table 1: Statistics and analysis of the σS-GVIIIA conotoxin structure.

Experimental restraints	
Interproton distance restraints[1]	
Intraresidue	170
Sequential	98
Medium range (i-j < 5)	13
Long range (i-j ≥ 5)	65
Total	346
Dihedral-angle restraints[2]	50
Restraint statistics	
Average number of violations per structure[1]	
NOE restraints > 0.2 Å	0
Dihedral restraints > 2°	0
Restraint violations	
Distance restraints (Å)[1]	
RMS deviations	0.0123 ± 0.0009
Dihedral-angle restraints (°)[1]	
RMS deviations	0.069 ± 0.050
Structural quality	
RMSD from average structure (Residues 1–41) (Å)[3]	
Backbone atoms	0.68 ± 0.138
All heavy atoms	1.15 ± 0.30
RMSD from average structure (selected residues)[5] (Å)[3]	
Backbone atoms	0.46 ± 0.28
All heavy atoms	1.10 ± 0.25
Ramachandran statistics (%)[4]	
Most favoured regions	73.0
Additionally allowed regions	27.0
Generously allowed regions	0
Disallowed regions	0
Global quality scores (raw/Z scores)[4]	
Verify3D	0.16/−4.82
Prosall (−ve)	0.66/0.04
PROCHECK (all)	−0.89/−5.26
MolProbity clashscore	8.87 / 0.00

[1]Derived from CYANA [26].
[2]Predicted from TALOS-N [30,31].
[3]Calculated using MOLMOL [32].
[4]Evaluated by PSVS [33].
[5]Residues 3-6, 13-16, 25-28 and 35-38.

sequence and the eight-residue ring (Figure 3), including the presence of glycine in the middle, which is likely to prevent the steric clashes with the Cys$^{\text{I'}}$-Cys$^{\text{IV'}}$ disulfide bond that penetrates through the ring [34,35].

The two additional intra-molecular disulfide bonds, not present in the knot motif of σS-GVIIIA, appear to be uncommon based on a search of GFCK-containing proteins in the Protein Data Bank. The most similar arrangement of disulfide bonds to σS-GVIIIA was found in the vertebrate specific hormone GPH-α (PDB:1HRP) [36] (Supplementary Figure S6).
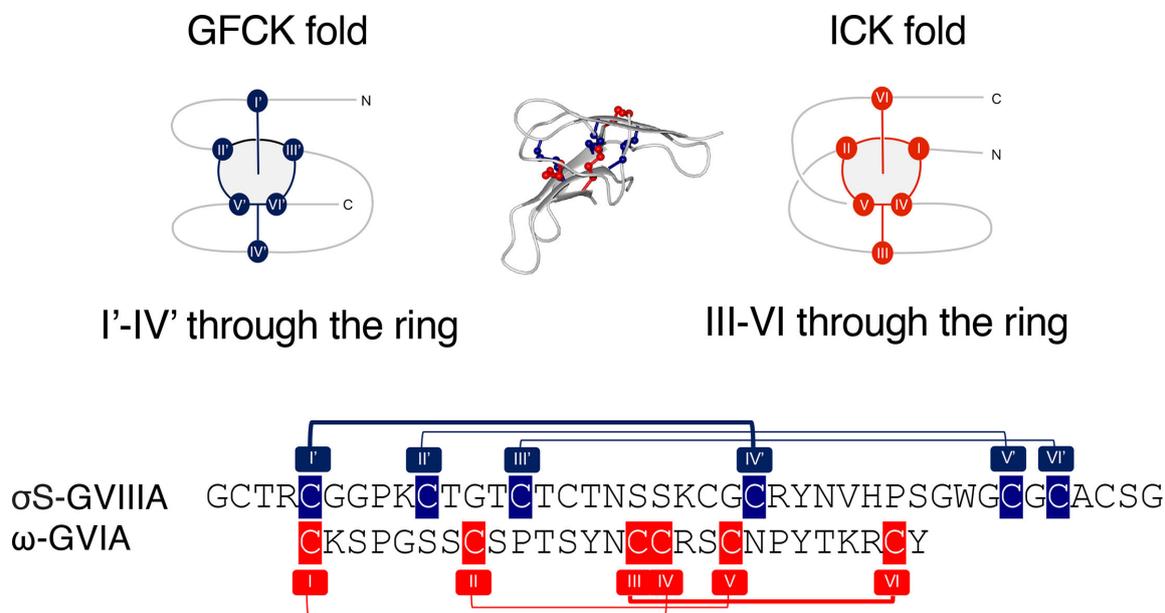
Figure 2: Comparison of the GFCK and ICK motifs.

The GFCK motif is shown in navy blue in σS-GVIIIA (PDB:9EBE), and the ICK motif is shown in red in ω-GVIA (PDB:2CCO) for both the schematic representations of the knot structures and the sequences. Cysteine residues involved in the knot are numbered using Roman numerals, I–VI for ICK and I′–VI′ for GFCK. Cysteine residues, which are not involved in the knot, have not been highlighted. GFCK, growth factor cystine knot.
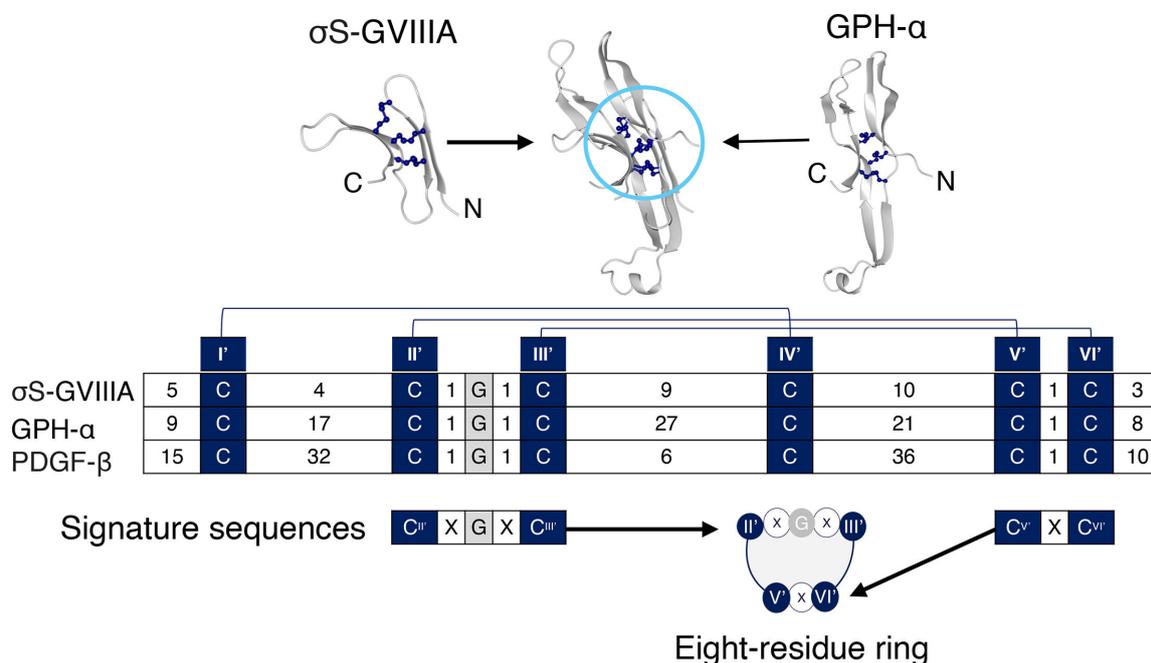

Figure 3: Comparison of the GFCK motif present in σS-GVIIIA and GPH-α.

The disulfide bonds involved in the GFCK motif in the structures of GPH-α (PDB:1HPR) and σS-GVIIIA (PDB:9EBE) are shown in navy blue and superimposed in the middle of the diagram. The signature sequences of the eight-residue ring of the GFCK motif, along with the number of residues in the inter-cysteine loops for σS-GVIIIA, GPH-α and PDGF-β, and the disulfide bonds are shown in the lower half of the diagram. Cysteine residues, which are not participating in the knot formation, have not been highlighted. GFCK, growth factor cystine knot; GPH-α, glycoprotein hormone α-subunit; PDGF, platelet-derived growth factor β-subunit.

Potential GFCK-containing peptides in *Conus* venom

The ConoServer database [37] of *Conus* venom peptides was searched for peptides containing the signature sequences of the GFCK motif (CXGXC and CXC). This sequence motif was found in the majority of peptides with the type VIII cysteine framework. Alignment of all peptides containing the type VIII cysteine framework showed 31 out of 36 sequences contained the GFCK motif signature sequences found in σS-GVIIIA. Cysteine residues not involved in the knot motif were also highly conserved (selected examples are shown in Figure 4). Sequence variability is present at inter-cysteine loops 1, 2 ,5 and 7; in addition, αS-RVIIIA (*C. radiatus*) showed variability in loop 3, where the sequence is CGXXC instead of CXGXC.

AlphaFold prediction of the 3D structure of σS-GVIIIA in isolation

Prediction of the structure of σS-GVIIIA with AlphaFold 3 [38] showed agreement with our NMR structure when the precursor sequence (signal peptide and proregion) was included in the input sequence (Figure 5). The backbone RMSD between the lowest energy NMR structure and the most likely prediction from AlphaFold 3 was 0.791 Å for the well-defined regions (residues 3–5, 15–17, 25–28 and 35–38) and 1.598 Å for residues 1–41.

This agreement between our NMR structure and the AlphaFold 3 predictions only occurred when the precursor sequence was included in the input sequence. When the precursor sequence was not included, the predicted structures were inconsistent with the NMR structure as shown in Figure 5. In the predictions without the precursor sequence, the position of the β3 and β4 strands was altered. Consequently, the knot motif was not topologically formed despite containing the Cys^6-Cys^{15}, Cys^{11}-Cys^{36} and Cys^{15}-Cys^{38} disulfide bonds. These disulfide bonds were also predicted when the precursor sequence was included, but the additional two disulfide bonds were predicted differently, with Cys^2-Cys^{17} and Cys^{23}-Cys^{40} predicted when the precursor sequence was included, and Cys^2-Cys^{23} and Cys^{17}-Cys^{40} predicted when the precursor sequence was excluded (Figure 5). The structure predicted with the precursor region did not show any interaction between the mature peptide and the precursor.

To further explore the differences in the predicted disulfide connectivity and experimentally derived connectivity of σS-GVIIIA, we calculated structures with the three possible disulfide connectivities for Cys^2, Cys^{23}, Cys^{17} and Cys^{40} (Supplementary Figure S7A–C: Set 1, Set 2 and Set 3) with the connectivities, Cys^6-Cys^{25}, Cys^{11}-Cys^{36} and Cys^{15}-Cys^{38} included as disulfide bond restraints. The comparison of the target function and restraint violations among the three sets of structures (Supplementary Figure S7A–C) indicated that Set 1 (i.e. Cys^2-Cys^{17} and Cys^{23}-Cys^{40}) provided the most likely disulfide connectivities for these four cysteine residues, consistent with the prediction based on our initial analysis.

One of the five highest ranked predicted structures was not fully oxidised when the structures were predicted without the precursor sequence, whereas all five predictions that included the precursor sequence were consistent with a fully oxidised form (data not shown). All ten cysteine residues are involved in disulfide bonds based on the mass spectrometry results (Supplementary Figure S3) and a previous study [16]; thus, the partially oxidised structure is inconsistent with the experimental data.

The predicted structure of σS-GVIIIA is also available online at the AlphaFold Database [39]. The precursor region is included for this prediction, and all five disulfide connectivities are consistent with the NMR structure [40].

Overall, the inclusion of precursor residues in the input sequence influenced the structure prediction of σS-GVIIIA, and the resulting structure is consistent with the structure derived from the NMR studies.

AlphaFold 3 prediction of a complex of σS-GVIIIA with the 5-HT$_3$ receptor

To predict the site of the interaction between σS-GVIIIA and mouse 5-HT$_3$ receptor, AlphaFold 3 [38] was used. The predicted complex between σS-GVIIIA and the 5-HT$_3$ receptor is shown in Figure 6. A comparison of this complex with the granisetron–receptor complex [41], a small molecule 5-HT$_3$ receptor antagonist (Supplementary Figure S8), indicates similarity between the binding site of the small molecule antagonist and the peptide antagonist. The critical residues in the orthosteric binding site on the 5-HT$_3$ receptor, established in previous studies with other ligands [41–45], are shown in silver stick format in Figure 6 and Supplementary Figure S8.
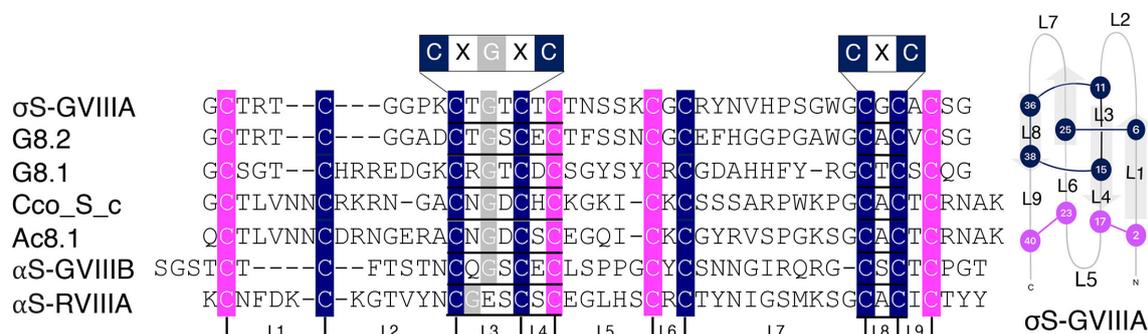
Figure 4: Sequence alignment of selected class VIII cysteine framework peptides.

The GFCK signature sequences are indicated at the top of the sequence alignments. The cysteines involved in the knot formation are shown in navy while cysteine residues not participating in the knot formation are shown in magenta. The loop regions between cysteine residues are indicated as L1-L9 under the sequence as well as in the schematic figure of the σS-GVIIIA structure. GFCK, growth factor cystine knot.
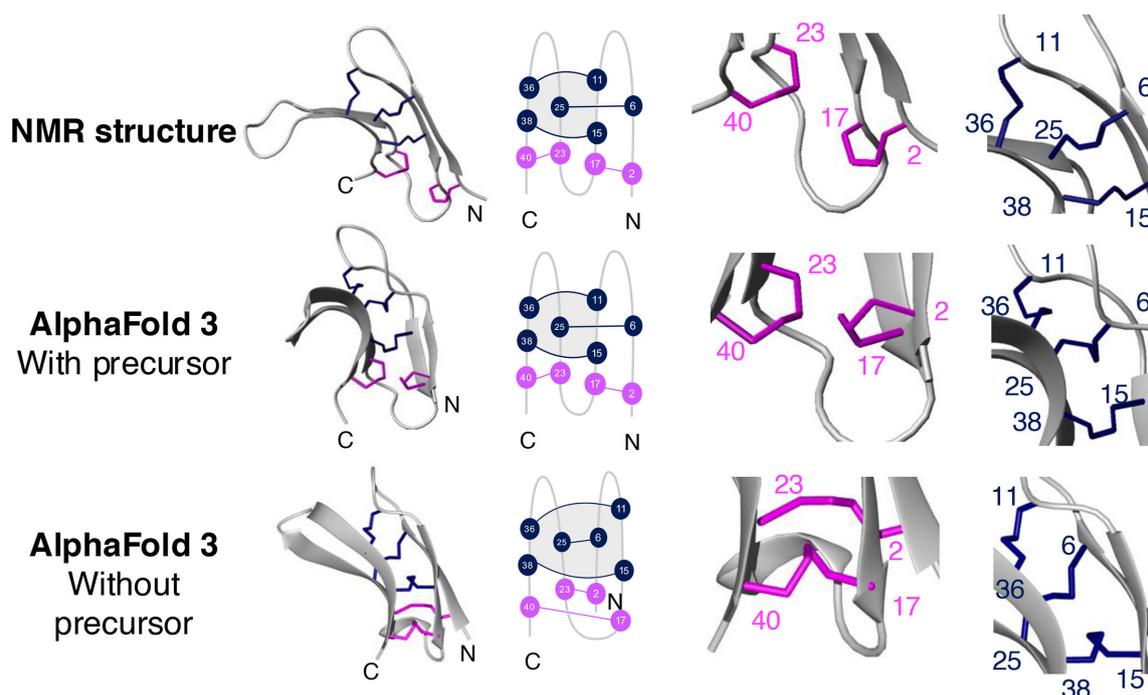

Figure 5: Comparison of the NMR-derived structure of σS-GVIIIA with AlphaFold 3 predictions with or without precursors.

A ribbon and schematic representation of the three-dimensional structures is shown. The precursor region (for the middle structure) is not shown. The NMR-derived structure of σS-GVIIIA represents the structure with the lowest target function. The zoomed-in regions highlight two bonds not involved in the cystine knot (shown in magenta) and the three bonds involved in the knot (shown in blue). The cysteine residues are numbered with Arabic numerals. NMR, nuclear magnetic resonance.

Analysis of the σS-GVIIIA and 5-HT$_3$ receptor complex indicates that several peptide residues are within 4 Å of 5-HT$_3$ receptor atoms (Figure 6). Inter-cysteine loop 7 of σS-GVIIIA (Figure 4), which includes the bromotryptophan at residue 34 (BTR34), is the main region predicted to interact with the active site of the 5-HT$_3$ receptor. Tryptophan is a precursor to the endogenous agonist serotonin (5-hydroxytryptamine, 5-HT) and is thus structurally similar to serotonin. There are also additional regions of σS-GVIIIA, involving the β1 and β4 strands, which are predicted to interact outside of the active site. Several hydrogen bonds between the β-strands of the receptor and the β4 strand of σS-GVIIIA are predicted for this additional site of interaction. Residues 5–26 of σS-GVIIIA are not predicted to interact with the 5-HT$_3$ receptor (Figure 6). In contrast with the mature GVIIIA sequence, the presence of the precursor sequence did not have a significant influence on the prediction results.
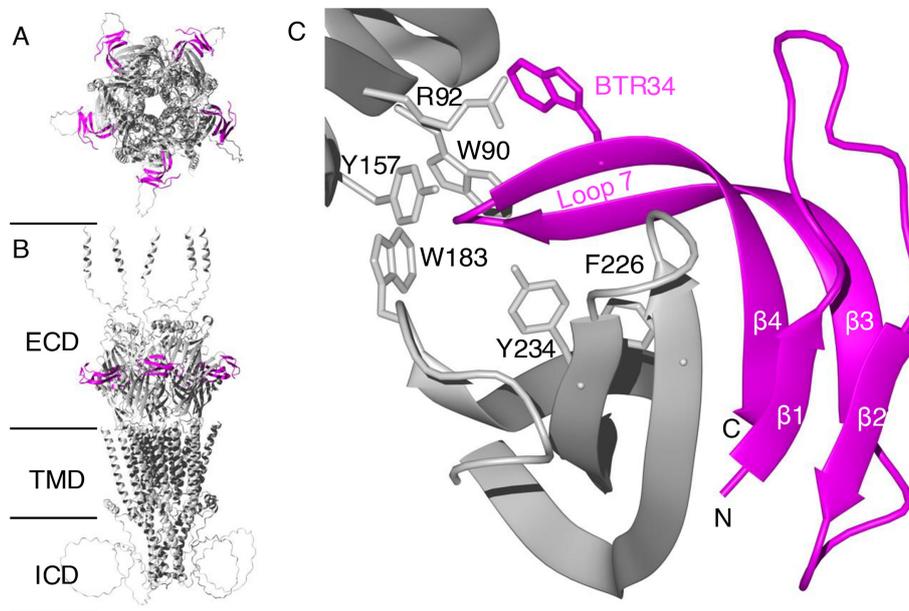
Figure 6: The ligand-receptor complex between σS-GVIIIA and the 5-HT₃ receptor.

(**A**) Side view of the σS-GVIIIA-5-HT$_3$ receptor complex. (**B**) Top view of the σS-GVIIIA-5-HT$_3$ receptor complex. The σS-GVIIIA interacting at the ECD of each subunit is shown in magenta. (**C**) The binding site of the σS-GVIIIA-5-HT$_3$ receptor complex. The key residues on the orthosteric binding site on the 5-HT$_3$ receptor are shown in silver stick format. The σS-GVIIIA conotoxin is shown in magenta. Bromotryptophan at residue 34 (BTR34) of σS-GVIIIA is shown in magenta stick format. The σS-GVIIIA consists of four β-strands, which are labelled as β1 to β4. ECD, extracellular domain; TMD, transmembrane domain; ICD, intracellular domain.

Discussion

In this study, we have shown that σS-GVIIIA contains a GFCK structural motif. This represents the first example of a conotoxin, or indeed an invertebrate venom peptide, experimentally shown to have this structural motif. In conjunction with the earlier study on σS-GVIIIA [16], this is also the first example of a GFCK motif peptide being associated with 5-HT$_3$ receptor antagonist activity.

Structural analysis of venom peptides is often carried out on synthetic/expressed versions of the peptides because of the limited availability of native material. Although the quantity of σS-GVIIIA purified from *C. geographus* venom was limited (0.6 mg), through the use of a 900 MHz NMR spectrometer with a cryogenically cooled probe, we were able to record high-quality NMR data, which allowed the determination of the 3D structure. Recording data on the native material ensured the native fold and native disulfide connectivity were present, and analysis of the sulfur–sulfur distances indicated that the disulfide connectivity corresponded to a GFCK motif (Figure 1, Supplementary Figure S4 and Table S1).

Additional disulfide bonds are often seen in GFCK-containing proteins (Supplementary Figure S6), including two additional inter-molecular disulfide bonds that facilitate functional dimer formation in the platelet-derived growth factor (PDGF) (PDB: 1PDG) [21,34]. σS-GVIIIA lacks such inter-molecular disulfide bonds, but forms two intra-molecular disulfide bonds, located in the region that is critical for receptor interaction based on our predicted ligand-receptor complex (Figure 6). Furthermore, the disulfide connectivities and topological locations of the two additional disulfide bonds present in σS-GVIIIA are generally not present in various signalling proteins and hormones [46,47] (Supplementary Figure S6). This difference in number of disulfide bonds may be critical for the receptor interaction [48]; thus, further study is warranted.

The prediction of protein/peptide structures by AI is a rapidly advancing area [49], and accurate structure prediction is critical to understanding the interactions and functions of proteins and peptides. We have shown that the presence/absence of the precursor sequence of σS-GVIIIA strongly influences the ability of AlphaFold 3 to correctly predict the mature σS-GVIIIA structure. The precursor sequence includes more conserved regions (e.g. signal sequence) than the mature sequence alone [37,50]; thus, this conservation may have influenced the AI structure prediction. We have recently identified similar limitations in AI-based prediction for the cysteine-rich peptide, conotoxin TxVIIB [51]. For new or rare

classes of peptide/protein, reliable training data in AlphaFold 3 is currently lacking, and the outcome can be incorrect predictions of disulfide connectivities [51].

Based on the AlphaFold 3 prediction of the σS-GVIIIA-5–HT$_3$ receptor complex, the solvent-exposed loop 7 region (residues 25–36) of σS-GVIIIA (Figure 4) was predicted to play a role in receptor binding (Figure 6). There is significant sequence variation among type VIII cysteine framework conotoxins in loop 7, along with both the loop 2 and 5 regions (Figure 4). Previous studies have indicated that σS-GVIIIA and σS-GVIIIB exhibit receptor specificity for the 5-HT$_3$ receptor and the α9α10 subtype of nACh receptor, respectively [15,16]. The sequence variation in these loops is likely responsible for the differences in receptor selectivity observed for σS-GVIIIA and σS-GVIIIB against these therapeutic targets.

In the original study of σS-GVIIIA, bromotryptophan at position 34 was suggested to be a critical residue for the 5-HT$_3$ receptor interaction based on the chemical similarity to the endogenous agonist serotonin [16]. In our study, the use of AI-based AlphaFold 3 structural prediction has supported this suggestion that bromotryptophan is located in a critical position to interact with the orthosteric binding site (Figure 5 and Supplementary Figure S8). Further experimental studies to confirm the detailed interaction of σS-GVIIIA with the 5-HT$_3$ receptor are clearly warranted.

The 5-HT$_3$ receptor has multiple subtypes [52], and the distribution of subtype expression appears to be associated with a range of clinical conditions and patient response to therapeutic drugs [19,53–55]. Analysis of the complex relationship in terms of binding of partial agonists, competitive, non-competitive antagonists or other allosteric modulators to this important therapeutic receptor has been revived in recent years [19,56,57]. The determination of the structure of the novel peptide antagonist σS-GVIIIA will further our understanding of the potential modes of interaction of peptides with 5-HT$_3$ receptor subtypes, which in turn may facilitate the design of drug leads targeting this receptor.

Our observation of individual variability in the crude venom composition and the switch of the venom profile between defense- and predation-provoked stimuli was consistent with previous studies [58,59]. However, factors influencing the crude venom profile on each milking occasion, which had a major influence on the relative abundance of σS-GVIIIA in the defensive venom, warrant further investigation.

In conclusion, working with native venom material is challenging but offered us the opportunity to determine the native fold of σS-GVIIIA. Our study highlighted the variabilities in the crude venom composition of *C. geographus*, including differences between defensive and offensive venom. The NMR structure of σS-GVIIIA provided the first experimental evidence of the GFCK motif in conotoxins. Based on sequence similarity, it appears likely that several more known conotoxins also contain this GFCK motif. More experimental studies on conotoxins with the GFCK motif are likely to improve AI prediction of 3D structures of members of this structural family. As the prediction of σS-GVIIIA/5-HT$_3$ receptor complex provided insight into further studies, the AI-based structural prediction of ligand–target interaction appears to be a promising approach to assist in designing/evaluating drug leads prior to synthesis. Our determination of the novel disulfide connectivity in the σS-GVIIIA conotoxin is also likely to facilitate the *in vitro* synthesis of this molecule through selective oxidation. Once this has been achieved, this will allow more detailed structure/function relationships to be explored experimentally and subsequent drug design studies for the therapeutically important 5-HT$_3$ receptor.

Materials and methods

Venom sample milking

Cone snails (*C. geographus*) were collected from the Great Barrier Reef, Queensland, Australia, by a permitted local commercial supplier (Cairns Marine, Cairns, Australia) and kept alive in an aquarium laboratory facility at James Cook University (J. Seymour, eduQuarium). The cone snails were not killed during this project, but some died naturally. The venom was collected from individual animals by milking with defensive and offensive stimuli [58]. Briefly, a 1.5-ml tube covered with parafilm was prepared and the individual animal enticed via a threat (defensive) or prey item (offensive) to inject the venom into the tube through the parafilm. Milking was conducted over a 36-month period on approximately a once weekly basis. All venom samples were stored at −30°C until purification, pooled samples for each individual were generated by pooling all samples for each individual. To create the venom pools, each individual sample tube was centrifuged for 10 min at maximum speed in a benchtop centrifuge (MICRO 200 R, Hettichi, U.S.A., 15,000 rpm) and the supernatants for each specimen were pooled into one tube per specimen. The remaining sample precipitate was re-suspended with 0.5 ml of type 1 water and the same procedure

repeated to collect further supernatant. The pooled venom sample was centrifuged one final time to ensure the removal of any larger particulates. Samples were not filtered to minimise sample loss. More than 20 fractions (~1.5 ml) were collected for each crude venom purification cycle. In total, two rounds of purification were conducted for each of the five pooled samples.

Venom profile analysis

LC–MS analysis, on a Shimadzu LC-MS2020 single quadrupole mass spectrometer (Shimadzu, Kyoto, Japan) equipped with a Shimadzu Prominence high-performance liquid chromatography system, was used to identify the masses of each peptide present in the crude venom and generate a venom profile for each milking. The UV absorbance at 214 and 280 nm was used to monitor the relative abundance of σS-GVIIIA in each milking, which was observed to fluctuate during the pilot studies. The sample (2 μl in 8 μl LC-MS Buffer A) was loaded onto a RP-HPLC column (Phenomenex Aeris PEPTIDE, 150 mm × 2.1 mm, 3.6 μm, 100 Å; Phenomenex, Torrance, CA, USA) and eluted using a linear 1% gradient of LC–MS solvent B (90% ACN with 0.09% formic acid [FA] in type 1 water) in LC–MS solvent A (0.1% FA in type 1 water) at a flow rate of 250 μl/min. Electrospray ionisation mass spectrometry (ESI-MS) data were collected in positive ionisation mode over a scan range of 250–2000 m/z. Molecular weights were reconstructed from ion series using the 'multi-charged ion analysis' module of the Shimadzu LabSolutions software (Version 5.96) (Shimadzu, Kyoto, Japan).

Isolation of σS-GVIIIA conotoxin from the crude venom with RP-HPLC

The pooled sample was diluted with HPLC solvent A (0.05% trifluoroacetic acid [TFA] in type 1 water) and purified using RP-HPLC (Agilent Infinity 1260: Agilent, Santa Clara, CA, U.S.A.). A total of 500 μl of the pooled sample was loaded onto a C18 semi-prep column (Phenomenex Aeris PEPTIDE XB-C18, 250 mm × 10 mm, 5 μm, 100 Å; Phenomenex, Torrance, CA, U.S.A.). Elution used a linear 1% gradient (0–60% solvent B over 60 min) of HPLC solvent B (90% acetonitrile (ACN) with 0.045% TFA in type 1 water) at a flow rate of 3 ml/min. Absorbance was monitored at 214 and 280 nm, and fractions were collected automatically into a 96-well, 2-ml deep well plate (Axygen, Union City, CA, U.S.A.) at 0.33 minute intervals.

An impure fraction containing σS-GVIIIA was re-purified using an analytical column (Phenomenex Aeris PEPTIDE XB-C18, 150 mm × 4.6 mm, 3.5 μm, 100 Å; Phenomenex, Torrance, CA, U.S.A.) with a linear gradient of 0.25% (30–50% HPLC solvent B over 80 min) at a flow rate of 1 ml/min and absorbance monitored at 214 nm and 280 nm. The purified σS-GVIIIA sample was lyophilised and stored at −30°C.

Mass spectrometry

The mass of the peptide in the purified fraction (subsequently confirmed by NMR spectroscopy to be σS-GVIIIA conotoxin) was determined by mass spectrometry using a SCIEX TOF/TOF™ 5800 MALDI mass spectrometer (SCIEX, Framingham, MA, U.S.A.), mixing 0.75 μl of sample with 0.75 μl α-cyano-4-hydroxycinnamic acid (CHCA; Sigma-Aldrich, St. Louis, MO, U.S.A.) matrix at 7.5 mg/ml in 50% ACN/0.1% TFA onto an Opti-TOF 384-well plate. Data were collected in reflector positive ion mode with a scan range of 2000–5000 m/z and averaged over 2000 laser shots. An LC-MS2020 single quadrupole mass spectrometer (Shimadzu, Kyoto, Japan) was also used to confirm the mass and purity of the σS-GVIIIA sample. Sample (30 μl) was loaded onto a RP-HPLC column (Phenomenex Aeris PEPTIDE, 150 mm × 2.1 mm, 3.6 μm, 100 Å; Phenomenex, Torrance, CA, U.S.A.) and eluted using a linear 1% gradient of LC-MS Buffer B in LC-MS solvent A at a flow rate of 250 μL/min. ESI-MS data were collected in positive ionisation mode over a scan range of 250–2000 m/z. Molecular weights were reconstructed from ion series using the 'multi-charged ion analysis' module of the Shimadzu LabSolutions software (Version 5.96) (Shimadzu, Kyoto, Japan).

NMR spectroscopy

Lyophilised sample was dissolved in 90% H_2O:10% D_2O at a concentration of approximately 0.26 mM. All NMR spectra were acquired on either a Bruker 600 MHz AVANCE III NMR spectrometer (Bruker, Karlsruhe, Germany) or Bruker Avance Neo 900 MHz spectrometer; both were equipped with cryogenically cooled probes. Two-dimensional spectra, including ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{1}H

DQF-COSY and ^{1}H-^{13}C HSQC, were collected at 298 K and recorded with a 1 s interscan delay using standard Bruker pulse sequences with an excitation sculpting scheme for solvent suppression. Homonuclear TOCSY and NOESY spectra were acquired with a mixing time of 80 ms and 200 ms, respectively. All spectra were processed using Bruker TopSpin (Version 4.3.0) and assigned using CcpNMR 2.4 (Collaborative Computing Project for NMR 2.4) analysis [60] based on the approach described in Wüthrich et al. [25,61].

Structure calculations

The solution NMR structure of σS-GVIIIA was calculated using the CYANA program [26,27]. The experimental restraints, including distance, backbone torsion-angle and chi1 (χ1) angle, were incorporated into the calculation for the preliminary structure. The distance and angle restraints were derived from the NOESY spectra. Torsion-angle restraints were predicted using TALOS-N [30,31]. The side-chain χ1 angle for cysteine residues was determined based on the information from the intensities of interproton HN-Hβ and Hα–Hβ cross-peaks derived from NOESY spectra [62] and DISH (https://dish-s-407402.web.app/). DISH is a prediction algorithm for side-chain dihedral angles, such as χ1 and χ2 angles, based on the chemical shifts [63]. The χ1 angles from both sources agreed with each other and were used for the structure calculations. Hydrogen bond restraints were subsequently included based on experimental data from the temperature coefficients of the amide protons [64,65] and the preliminary structures. The most likely disulfide connectivities were predicted based on the distances between the sulfur atoms in the preliminary structures, measured using MOLMOL [32]. A set of 100 final structures was calculated with the most likely disulfide connectivity, and the 20 structures with the lowest target function chosen to present the final ensemble. Structures were visualized and the RMSD values were assessed using MOLMOL [32].

Disulfide bonds involving four cysteines (Cys^2, Cys^{17}, Cys^{23} and Cys^{40}) were found to need further clarification based on the discrepancy between the AlphaFold prediction and the NMR structure. We calculated a set of structures for each of the three possible disulfide connectivity patterns possible with these four cysteine residues (Set 1, Set 2 and Set 3: Supplementary Figure S7). The coordinates and chemical shifts were deposited in the PDB (ID code 9EBE) and BMRB (ID code 31213), respectively.

AlphaFold prediction of the 3D structure of σS-GVIIIA and σS-GVIIIA/5-HT$_3$ complex

The neural network based AlphaFold is a novel computational approach for the prediction of peptide/protein structures, developed by Google Deep Mind. In this AI system, multi-sequence alignments, physical, biological and evolutionary knowledge are integrated into a sophisticated machine learning algorithm [66]. AlphaFold 3 became available in May 2024 on the AlphaFold Server [38].

We predicted the structure of σS-GVIIIA using AlphaFold 3 with or without a precursor sequence. This was based on a previous finding which indicated that the precursor sequences can significantly influence the success of AI structure predictions [67]. In the input sequence for AlphaFold 3, the hydroxyproline at residue 9 was able to be distinguished from the proline without a PTM at residue 31; however, there was no option in AlphaFold 3 to specify bromotryptophan at residue 34 or C-terminal amidation. The predicted results were visualised using MOLMOL [32].

Mouse 5-HT$_3$ receptor was used for our AlphaFold prediction of the ligand-receptor complex between σS-GVIIIA and mouse 5-HT$_3$ receptor [38], as multiple binding analyses of various antagonists have been studied with the mouse receptor, including the original σS-GVIIIA study [16]. The two known functional subtypes of the serotonin receptor are 5-HT$_{3A}$ and 5-HT$_{3A/3B}$; with the most studied functional subtype, 5-HT$_{3A}$ being used for our analysis. The sequences of σS-GVIIIA with or without precursor were used for the prediction of the complex.

Multiple sequence alignment of σS-GVIIIA with potentially structurally related proteins and peptides

The signature sequences of the GFCK motif (CXGXC and CXC) were manually searched against *Conus* venom peptides available in the ConoServer database [68] with the type-VIII framework to identify

potential/unrecognised conotoxins containing the GFCK motif. Sequences were aligned based on the signature sequences and the distribution pattern of the other cysteine residues.

Data Availability

The coordinates and chemical shifts were deposited in the PDB (rcsb.org: ID code 9EBE) and BMRB (bmrb.io: ID code 31213), respectively. All other data are contained within the manuscript. All data supporting the findings of this study are available from the corresponding author upon request.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

CRediT Author Contribution

Y.P.: Conceptualisation, Sample collection, Data acquisition, Investigation, Formal analysis, Writing and editing. D.T.W.: Conceptualisation, Data acquisition, Investigation, Review and editing, Supervising. D.L-B.: Sample collection, Resources, Review and editing. J.G.: Data acquisition, Analysis, Review and editing. J.S.: Resources, Review and editing. S.D.: Conceptualisation, Sample collection, Data acquisition, Analysis, Review and editing. K.J.R.: Data acquisition, Analysis, Review and editing. M.J.L.: Conceptualisation, Analysis, Supervising, Review and editing. N.L.D.: Conceptualisation, Data acquisition, Investigation, Formal analysis, Writing and editing, Supervising.

Acknowledgments

Abbreviations

GFCK, growth factor cystine knot; GPH-α, glycoprotein hormone α subunit; 5-HT$_3$ receptor, serotonin type 3 receptor; ICK, inhibitor cystine knot; LC–MS, liquid chromatography–mass spectrometry; NMR, nuclear magnetic resonance; RMSD, root-mean-square deviation; RP-HPLC/MS, reversed-phase high-performance liquid chromatography/mass spectrometry; σS-GVIIIA, sigma-S-GVIIIA.

References

1   Kumar, P.S., Kumar, D.S. and Umamaheswari, S. (2015) A perspective on toxicology of conus venom peptides. *Asian Pac. J. Trop. Med.* **8**, 337–351 https://doi.org/10.1016/S1995-7645(14)60342-4

2   Akondi, K.B., Muttenthaler, M., Dutertre, S., Kaas, Q., Craik, D.J., Lewis, R.J. et al. (2014) Discovery, synthesis, and structure-activity relationships of conotoxins. *Chem. Rev.* **114**, 5815–5847 https://doi.org/10.1021/cr400401e

3   Jin, A.H., Muttenthaler, M., Dutertre, S., Himaya, S.W.A., Kaas, Q., Craik, D.J. et al. (2019) Conotoxins: chemistry and biology. *Chem. Rev.* **119**, 11510–11549 https://doi.org/10.1021/acs.chemrev.9b00207

4   Terlau, H. and Olivera, B.M. (2004) Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **84**, 41–68 https://doi.org/10.1152/physrev.00020.2003

5   Dombkowski, A.A., Sultana, K.Z. and Craig, D.B. (2014) Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 https://doi.org/10.1016/j.febslet.2013.11.024

6   Karimi, M., Ignasiak, M.T., Chan, B., Croft, A.K., Radom, L., Schiesser, C.H. et al. (2016) Reactivity of disulfide bonds is markedly affected by structure and environment: implications for protein modification and stability. *Sci. Rep.* **6**, 38572 https://doi.org/10.1038/srep38572

7   Góngora-Benítez, M., Tulla-Puche, J. and Albericio, F. (2014) Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem. Rev.* **114**, 901–926 https://doi.org/10.1021/cr400031z

8   Mir, R., Karim, S., Amjad Kamal, M., M. Wilson, C. and Mirza, Z. (2016) Conotoxins: structure, therapeutic potential and pharmacological applications. *Curr. Pharm. Des.* **22**, 582–589 https://doi.org/10.2174/1381612822666151124234715

9   Bjørn-Yoshimoto, W.E., Ramiro, I.B.L., Yandell, M., McIntosh, J.M., Olivera, B.M., Ellgaard, L. et al. (2020) Curses or cures: a review of the numerous benefits versus the biosecurity concerns of conotoxin research. *Biomedicines* **8**, 235 https://doi.org/10.3390/biomedicines8080235

10 Han, T.S., Teichert, R.W., Olivera, B.M. and Bulaj, G. (2008) Conus venoms - a rich source of peptide-based therapeutics. *Curr. Pharm. Des.* **14**, 2462–2479 https://doi.org/10.2174/138161208785777469

11 Dutertre, S., Jin, A., 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 https://doi.org/10.1074/mcp.M112.021469

12 Yao, G., Peng, C., Zhu, Y., Fan, C., Jiang, H., Chen, J. et al. (2019) High-throughput identification and analysis of novel conotoxins from three vermivorous cone snails by transcriptome sequencing. *Mar. Drugs* **17**, 193 https://doi.org/10.3390/md17030193

13 Gao, B., Peng, C., Yang, J., Yi, Y., Zhang, J. and Shi, Q. (2017) Cone snails: a big store of conotoxins for novel drug discovery. *Toxins (Basel)* **9**, 397 https://doi.org/10.3390/toxins9120397

14 Zhao, E., Liu, H.L., Tsai, C.H., Tsai, H.K., Chan, C.H. and Kao, C.Y. (2005) Cysteine separations profiles on protein sequences infer disulfide connectivity. *Bioinformatics* **21**, 1415–1420 https://doi.org/10.1093/bioinformatics/bti179

15 Christensen, S.B., Bandyopadhyay, P.K., Olivera, B.M. and McIntosh, J.M. (2015) αS-conotoxin GVIIIB potently and selectively blocks α9α10 nicotinic acetylcholine receptors. *Biochem. Pharmacol.* **96**, 349–356 https://doi.org/10.1016/j.bcp.2015.06.007

16 England, L.J., Imperial, J., Jacobsen, R., Craig, A.G., Gulyas, J., Akhtar, M. et al. (1998) Inactivation of a serotonin-gated ion channel by a polypeptide toxin from marine snails. *Science* **281**, 575–578 https://doi.org/10.1126/science.281.5376.575

17 Lee, J.S., Park, S.Y., Kim, N.Y., Kim, D.W., Oh, J.E., Heo, E. et al. (2019) Anti-tumor potential of a 5-HT3 receptor antagonist as a novel autophagy inducer in lung cancer: a retrospective clinical study with *in vitro* confirmation. *J. Clin. Med.* **8**, 1380 https://doi.org/10.3390/jcm8091380

18 Tone, M., Tahara, S., Nojima, S., Motooka, D., Okuzaki, D. and Morii, E. (2020) HTR3A is correlated with unfavorable histology and promotes proliferation through ERK phosphorylation in lung adenocarcinoma. *Cancer Sci.* **111**, 3953–3961 https://doi.org/10.1111/cas.14592

19 Irving, H., Turek, I., Kettle, C. and Yaakob, N. (2021) Tapping into 5-HT3 receptors to modify metabolic and immune responses. *Int. J. Mol. Sci.* **22**, 11910 https://doi.org/10.3390/ijms222111910

20 Walstab, J., Rappold, G. and Niesler, B. (2010) 5-HT(3) receptors: role in disease and target of drugs. *Pharmacol. Ther.* **128**, 146–169 https://doi.org/10.1016/j.pharmthera.2010.07.001

21 Murray-Rust, J., McDonald, N.Q., Blundell, T.L., Hosang, M., Oefner, C., Winkler, F. et al. (1993) Topological similarities in TGF-beta 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure* **1**, 153–159 https://doi.org/10.1016/0969-2126(93)90029-g

22 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 https://doi.org/10.1016/s0041-0101(00)00160-4

23 Zhu, S., Darbon, H., Dyason, K., Verdonck, F. and Tytgat, J. (2003) Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* **17**, 1765–1767 https://doi.org/10.1096/fj.02-1044fje

24 Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M. et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 https://doi.org/10.1002/prot.20449

25 Schroeder, C.I. and Rosengren, K.J. (2020) Three-dimensional structure determination of peptides using solution nuclear magnetic resonance spectroscopy. *Methods Mol. Biol.* **2068**, 129–162 https://doi.org/10.1007/978-1-4939-9845-6_7

26 Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 https://doi.org/10.1385/1-59259-809-9:353

27 Kuschert, S., Stroet, M., Chin, Y.K.Y., Conibear, A.C., Jia, X., Lee, T. et al. (2022) Facilitating the structural characterisation of non-canonical amino acids in biomolecular NMR. 1–22 Magn. Reson. Discuss. https://doi.org/10.5194/mr-2022-22

28 Chaney, M.O. and Steinrauf, L.K. (1974) The crystal and molecular structure of tetragonal L-cystine. *Acta Crystallogr. B.* **30**, 711–716 https://doi.org/10.1107/S0567740874003566

29 Sun, M.A., Wang, Y., Zhang, Q., Xia, Y., Ge, W. and Guo, D. (2017) Prediction of reversible disulfide based on features from local structural signatures. *BMC Genomics* **18**, 279 https://doi.org/10.1186/s12864-017-3668-8

30 Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 https://doi.org/10.1007/s10858-009-9333-z

31 Shen, Y. and Bax, A. (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 https://doi.org/10.1007/s10858-013-9741-y

32 Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, https://doi.org/10.1016/0263-7855(96)00009-4

33 Bhattacharya, A., Tejero, R. and Montelione, G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778–795 https://doi.org/10.1002/prot.21165

34 Sun, P.D. and Davies, D.R. (1995) The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269–291 https://doi.org/10.1146/annurev.bb.24.060195.001413

35 McDonald, N.Q. and Hendrickson, W.A. (1993) A structural superfamily of growth factors containing a cystine knot motif. *Cell* **73**, 421–424 https://doi.org/10.1016/0092-8674(93)90127-c

36 Lapthorn, A.J., Harris, D.C., Littlejohn, A., Lustbader, J.W., Canfield, R.E., Machin, K.J. et al. (1994) Crystal structure of human chorionic gonadotropin. *Nature* **369**, 455–461 https://doi.org/10.1038/369455a0

37 Kaas, Q., Westermann, J.C., Halai, R., Wang, C.K.L. and Craik, D.J. (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **24**, 445–446 https://doi.org/10.1093/bioinformatics/btm596

38 Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A. et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 https://doi.org/10.1038/s41586-024-07487-w

39 Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G. et al. (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 https://doi.org/10.1093/nar/gkab1061

40 Ho, T.N.T., Tran, T.H., Le, H.S. and Lewis, R.J. (2025) Advances in the synthesis and engineering of conotoxins. *Eur. J. Med. Chem.* **282**, 117038 https://doi.org/10.1016/j.ejmech.2024.117038

41 Basak, S., Gicheru, Y., Kapoor, A., Mayer, M.L., Filizola, M. and Chakrapani, S. (2019) Molecular mechanism of setron-mediated inhibition of full-length 5-HT$_{3A}$ receptor. *Nat. Commun.* **10**, 3225 https://doi.org/10.1038/s41467-019-11142-8

42 Kesters, D., Thompson, A.J., Brams, M., Elk, R., Spurny, R., Geitmann, M. et al. (2013) Structural basis of ligand recognition in 5-HT3 receptors. *EMBO Rep.* **14**, 49–56 https://doi.org/10.1038/embor.2012.189

43 Thompson, A.J. and Lummis, S.C.R. (2006) 5-HT3 receptors. *Curr. Pharm. Des.* **12**, 3615–3630 https://doi.org/10.2174/138161206778522029

44 Basak, S., Gicheru, Y., Rao, S., Sansom, M.S.P. and Chakrapani, S. (2018) Cryo-EM reveals two distinct serotonin-bound conformations of full-length 5-HT$_{3A}$ receptor. *Nature* **563**, 270–274 https://doi.org/10.1038/s41586-018-0660-7

45 Zarkadas, E., Zhang, H., Cai, W., Effantin, G., Perot, J., Neyton, J. et al. (2020) The binding of palonosetron and other antiemetic drugs to the serotonin 5-HT3 receptor. *Structure* **28**, 1131–1140 https://doi.org/10.1016/j.str.2020.07.004

46 Gong, Z., Wang, W., El Omari, K., Lebedev, A.A., Clarke, O.B. and Hendrickson, W.A. (2023) Crystal structure of LGR ligand α2/β5 from *Caenorhabditis elegans* with implications for the evolution of glycoprotein hormones. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218630120 https://doi.org/10.1073/pnas.2218630120

47 Vitt, U.A., Hsu, S.Y. and Hsueh, A.J.W. (2001) Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules. *Mol. Endocrinol.* **15**, 681–694 https://doi.org/10.1210/mend.15.5.0639

48 Hsu, S.Y., Nakabayashi, K. and Bhalla, A. (2002) Evolution of glycoprotein hormone subunit genes in bilateral metazoa: identification of two novel human glycoprotein hormone subunit family genes, GPA2 and GPB5. *Mol. Endocrinol.* **16**, 1538–1551 https://doi.org/10.1210/mend.16.7.0871

49 Yin, R. and Pierce, B.G. (2024) Evaluation of AlphaFold antibody-antigen modeling with implications for improving predictive accuracy. *Protein Sci.* **33**, e4865 https://doi.org/10.1002/pro.4865

50 Robinson, S.D. and Norton, R.S. (2014) Conotoxin gene superfamilies. *Mar. Drugs* **12**, 6058–6101 https://doi.org/10.3390/md12126058

51 Raffaelli, T., Wilson, D.T., Dutertre, S., Giribaldi, J., Vetter, I., Robinson, S.D. et al. (2024) Structural analysis of a U-superfamily conotoxin containing a mini-granulin fold: insights into key features that distinguish between the ICK and granulin folds. *J. Biol. Chem.* **300**, 107203 https://doi.org/10.1016/j.jbc.2024.107203

52 Thompson, A.J. and Lummis, S.C.R. (2013) Discriminating between 5-HT$_3$A and 5-HT$_3$AB receptors. *Br. J. Pharmacol.* **169**, 736–747 https://doi.org/10.1111/bph.12166

53 Fakhfouri, G., Rahimian, R., Dyhrfjeld-Johnsen, J., Zirak, M.R. and Beaulieu, J.M. (2019) 5-HT$_3$ receptor antagonists in neurologic and neuropsychiatric disorders: the iceberg still lies beneath the surface. *Pharmacol. Rev.* **71**, 383–412 https://doi.org/10.1124/pr.118.015487

54 Abad, I.P.L., Fam, R.L., Nguyen, D.T., Nowell, C.J., Trinh, P.N.H., Manallack, D.T. et al. (2020) Visualising functional 5-HT$_3$ receptors containing A and C subunits at or near the cell surface. *Biomed. Pharmacother.* **132**, 110860 https://doi.org/10.1016/j.biopha.2020.110860

55 Yaakob, N., Malone, D.T., Exintaris, B. and Irving, H.R. (2011) Heterogeneity amongst 5-HT$_3$ receptor subunits: is this significant? *Curr. Mol. Med.* **11**, 57–68 https://doi.org/10.2174/156652411794474392

56 Felt, K., Stauffer, M., Salas-Estrada, L., Guzzo, P.R., Xie, D., Huang, J. et al. (2024) Structural basis for partial agonism in 5-HT3A receptors. *Nat. Struct. Mol. Biol.* **31**, 598–609 https://doi.org/10.1038/s41594-023-01140-2

57 López-Sánchez, U., Munro, J., Ladefoged, L., Pedersen, A., Nielsen, C., Lyngby, S.M. et al. (2024) Structural determinants for activity of the antidepressant vortioxetine at human and rodent 5-HT3 receptors. *bioRxiv* https://doi.org/10.1101/2024.02.23.581731

58 Dutertre, S., Jin, A.H., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V. et al. (2014) Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat. Commun.* **5**, 3521 https://doi.org/10.1038/ncomms4521

59 Dutertre, S., Jin, A.H., Alewood, P.F. and Lewis, R.J. (2014) Intraspecific variations in conus geographus defence-evoked venom and estimation of the human lethal dose. *Toxicon* **91**, 135–144 https://doi.org/10.1016/j.toxicon.2014.09.011

60 Skinner, S.P., Fogh, R.H., Boucher, W., Ragan, T.J., Mureddu, L.G. and Vuister, G.W. (2016) CcpNmr analysisassign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 https://doi.org/10.1007/s10858-016-0060-y

61 Wüthrich, K. (1983) Sequential individual resonance assignments in the 1H-NMR spectra of polypeptides and proteins. *Biopolymers* **22**, 131–138 https://doi.org/10.1002/bip.360220121

62 Barsukov, I.L. and Lian, L.Y. (1993) Structure determination from NMR data I. Analysis of NMR data. In NMR of Macromolecules: A Practical Approach, pp. 315–357, Oxford University Press

63 Armstrong, D.A., Kaas, Q. and Rosengren, K.J. (2018) Prediction of disulfide dihedral angles using chemical shifts. *Chem. Sci.* **9**, 6548–6556 https://doi.org/10.1039/c8sc01423j

64 Cierpicki, T. and Otlewski, J. (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* **21**, 249–261 https://doi.org/10.1023/a:1012911329730

65 Rosengren, K.J., Wilson, D., Daly, N.L., Alewood, P.F. and Craik, D.J. (2002) Solution structures of the cis- and trans-Pro30 isomers of a novel 38-residue toxin from the venom of *Hadronyche Infensa sp.* that contains a cystine-knot motif within its four disulfide bonds. *Biochemistry* **41**, 3294–3301 https://doi.org/10.1021/bi011932y

66 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 https://doi.org/10.1038/s41586-021-03819-2

67 Mironov, P.A. and Shenkarev, Z.O. (2023) Patterns of a disulfide bond formation in toxins of spiders from the ctenidae family: a comparison with predictions of the AlphaFold 2.0 neural network. *Moscow Univ. Biol.Sci. Bull.* **78**, S9–S16 https://doi.org/10.3103/S0096392523700141

68 Kaas, Q., Yu, R., Jin, A.H., Dutertre, S. and Craik, D.J. (2012) ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **40**, D325–D330 https://doi.org/10.1093/nar/gkr886