δ-Conotoxin Structure Prediction and Analysis through Large-Scale Comparative and Deep Learning Modeling Approaches

*Stephen McCarthy and Shane Gonen**

The δ-conotoxins, a class of peptides produced in the venom of cone snails, are of interest due to their ability to inhibit the inactivation of voltage-gated sodium channels causing paralysis and other neurological responses, but difficulties in their isolation and synthesis have made structural characterization challenging. Taking advantage of recent breakthroughs in computational algorithms for structure prediction that have made modeling especially useful when experimental data is sparse, this work uses both the deep-learning-based algorithm AlphaFold and comparative modeling method RosettaCM to model and analyze 18 previously uncharacterized δ-conotoxins derived from piscivorous, vermivorous, and molluscivorous cone snails. The models provide useful insights into the structural aspects of these peptides and suggest features likely to be significant in influencing their binding and different pharmacological activities against their targets, with implications for drug development. Additionally, the described protocol provides a roadmap for the modeling of similar disulfide-rich peptides by these complementary methods.

1. Introduction

Conopeptides are peptides that have been evolved by marine cone snails to protect against predators and to capture prey. Each species of cone snail may produce over 1000 different peptides[1] and estimates for the total number of conopeptides range from 50000 to millions,[2,3] but structural characterization remains sparse. The targets for these conopeptides are varied and include several protein classes of significance in human disease such as ion channels and G protein-coupled receptors (GPCRs). Although conopeptides rarely comprise more than 40 amino acids, individual conopeptides can show high levels of target specificity

and potency, including, in some cases, the ability to distinguish between receptor subtypes.[4] These properties have made them the subject of considerable research, both as tools in neurological research and as leads for drugs targeting channelopathies such as chronic pain, epilepsy, and migraines.

Conotoxins are conopeptides that exert a toxic effect on the envenomed organism and are classified according to their target and pharmacological effects. δ-conotoxins (**Figure 1A**) interact with voltage-gated sodium channels (Na$_v$s), where they inhibit the fast inactivation phase of channel gating and prolong an open channel conformation, similar to scorpion α-toxins.[5–7] While the target subtype selectivity of most δ-conotoxins remains undetermined, intriguing differences have been noted between δ-conotoxins derived from molluscivorous (mollusk-eating) cone snails compared with those from piscivorous (fish-eating) cone snails (Figure 1B): while δ-conotoxins from piscivorous cone snails show activity against vertebrate Na$_v$s, peptides from molluscivorous snails (with the exception of Am2766[8]) do not show activity on vertebrate neurons, while retaining their characteristic activity against mollusk Na$_v$s.[4,5,9,10] Radioligand binding studies on the molluscivorous cone snail peptide TxVIA nonetheless suggest that it retains the ability to bind to vertebrate Na$_v$s, but without having any pharmacological effect on the channel.[5] A structural basis for this "silent binding" has not yet been determined, and experimental structures of δ-conotoxins to date are limited to two peptides from molluscivorous cone snails, TxVIA and Am2766, and one from a piscivorous cone snail, EVIA, which has been the subject of several structural studies.[11–14] The peptide EVIA is an outlier among conopeptides from piscivorous snails due to its longer loop 2 (see Figure 1A for nomenclature), and, consequently little structural information exists for most members of this class.

Computational structure prediction is especially valuable when determining protein structures by experimental methods is challenging. Traditional methods for structure prediction include modeling by homology (in which known structures of evolutionarily similar (homologous) proteins are used as templates)[15,16] and ab initio calculations (which aim to mimic protein folding in nature by subjecting an unfolded protein chain to an energy function, without reference to external templates).[17] More recently, protein structure prediction

S. McCarthy, S. Gonen
Department of Molecular Biology and Biochemistry
University of California
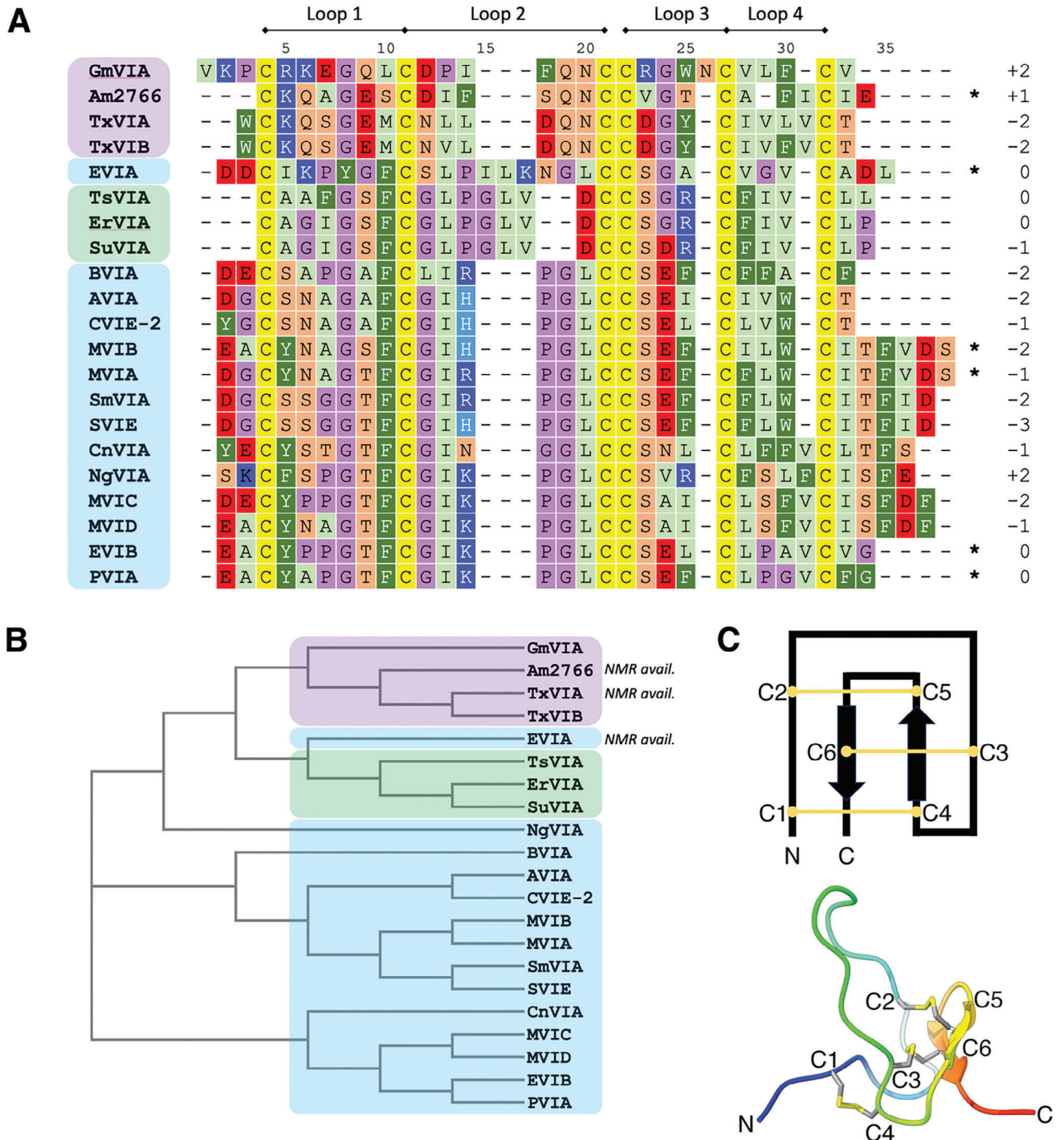Irvine, CA 92697, USA
E-mail: gonens@uci.edu

Figure 1. A) Sequence alignment of the δ-conotoxins modeled in this study, together with the overall net charge and loop nomenclature. Peptides from piscivorous, vermivorous, and molluscivorous cone snails are highlighted in blue, green, and purple, respectively. C-terminal amides are indicated with *. B) Phylogenetic analysis of the sequences shown in A indicating peptides with available experimental structures and colored according to the color scheme in A. C) (top) Features typical of an inhibitor cystine knot (ICK) peptide, showing the connectivity of the disulfide bonds (numbering refers to the order in which cysteines appear, not residue numbering) and (bottom) One conformation of the NMR-derived structure of δ-EVIA (PDB ID: 1G1P)[13] showing the classic ICK fold.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SCIENCE**
Open Access

www.advancedscience.com

algorithms based on machine learning techniques have come to the fore.[18] One such approach is "deep learning" which uses multilayered neural networks trained on large sequence and structure datasets to predict protein structures from large multiple sequence alignments (MSAs)[19,20] and is exemplified by AlphaFold[21] and RoseTTAFold.[22,23] Other methods, such as ESMFold[24] and OmegaFold,[25] take advantage of language models, which treat protein sequences similarly to natural languages and predict structural features by learning the "grammar" of the training sequence data;[26] these methods can predict structures without the need to generate MSAs for the target protein.

Here we take advantage of these complementary computational approaches to determine 3D models of all δ-conotoxins without an experimental structure using AlphaFold and RosettaCM. The deep learning-based algorithm AlphaFold generates models using MSAs and structural data as inputs,[21] whereas RosettaCM is a threading modeling method requiring one or more template structures.[27] Computational modeling is especially appropriate for studies of δ-conotoxins, as the complex inhibitor cystine knot (ICK) disulfide bonding pattern (Figure 1C), low in vitro oxidative folding yields, and high hydrophobicity have hampered the synthetic efforts necessary for structural studies. Analyzing multiple models generated by both AlphaFold and RosettaCM provides insights into key structural features of the δ-conotoxins and the bases for their pharmacological properties, assists in developing modeling strategies for disulfide-rich peptides, and aids in the characterization of δ-conotoxins yet to be discovered.

2. Results

Using AlphaFold, 5 decoys per target δ-conotoxin were generated, and all 5 models were included in the analysis. For the RosettaCM models the lowest-energy representatives of the top 5 clusters, generated from 5000 total models, were used for analysis. Representatives for each peptide, together with corresponding quality metrics, are highlighted in **Figure 2**, and our analysis and observations from these models are described below.

2.1. Peptides Show the Expected ICK Geometry

Peptide models were initially evaluated by checking to see that they had formed the expected ICK fold with the correct disulfide connectivity (C1-C4, C2-C5, C3-C6) (Figure 1C); in the case of the RosettaCM models this geometry was enforced and so all models met this criterion. For AlphaFold models, however, no disulfide bonding pattern was stipulated in the inputs; nonetheless, AlphaFold produced highly accurate predictions, with only a single model (out of a total of 90) displaying incorrect disulfide connectivity (Figure S1, Supporting Information). Notably, an attempt to perform modeling in AlphaFold using only MSA data and without the addition of template structures as inputs resulted in considerably poorer results, with approximately half the target peptides showing no models with the correct ICK topology (data not shown). Templates provide additional residue pair and amino acid torsion angle information that, together with evolutionary information derived from the multiple sequence alignment, contribute to the pair representation used by AlphaFold to calculate

atom positions; for the δ-conotoxins this will necessarily include disulfide bond pair information which will impose restraints on the structure calculation and lead to a higher success rate than for MSA data alone.

2.2. Peptide Termini Are Generally Flexible

For our purposes, the N- and C-terminal regions are defined as the residues prior to the first cysteine residue, and after the last cysteine residue, respectively. Consistent with experimental evidence showing that these regions are largely unstructured in ICK peptides,[13] we observe decreased predicted local distance difference test (pLDDT) scores in the AlphaFold models, which is ascribed to regions that are either intrinsically disordered or ordered only in complex (Figure S2, Supporting Information).[21,28,29] We also observe a high degree of divergence in the RosettaCM models of the termini, and correspondingly higher per-residue root-mean-square deviation (RMSD) scores.

2.3. Secondary Structure Is Limited

The ICK fold classically contains a short β-turn near the C-terminus that incorporates C5 and C6; very occasionally a short third strand is observed around C2.[30] Our models largely conform to this pattern, with limited secondary structure aside from this C-terminal β-turn. However, one RosettaCM model contains a formal helical half-turn just prior to C3, and this turn is also present in 5/18 of the top-ranked AlphaFold models, all of which are derived from piscivorous cone snails (Figure 2). This element is unusual in experimental structures of ICK peptides; an example featuring a helix in this position is the insecticidal funnel-web spider toxin U21-hexatoxin-Hi1a, however this peptide contains a number of additional structural features beyond the ICK scaffold core and the helix in this section of the peptide might be more properly considered to be an insertion into the otherwise unstructured loop.[31] Since this region of the δ-conotoxins is relatively flexible (see below) it is possible that the observed secondary structure in this region is transitory and does not represent a significant population in solution, but it is also possible that conotoxins with a longer loop 2 permit formation of short secondary structure elements that would not otherwise be observed. Additionally, AlphaFold has been shown to slightly over-predict the formation of secondary structural elements (especially helices) in flexible loops,[32,33] which could explain the increased occurrence of this helical turn in the AlphaFold models relative to the RosettaCM models.

2.4. Peptide Flexibility Is Greatest in Loop 2

Of the four disulfide-constrained loops, loop 2 consistently shows the lowest mean pLDDT score for the AlphaFold models and the highest RMSD between RosettaCM models (Figure S2, Supporting Information), strongly suggesting that this region is flexible. This flexibility can be observed in the ensemble of structures for EVIA determined by NMR (Figure 2); while this peptide was used as a template in our RosettaCM modeling,
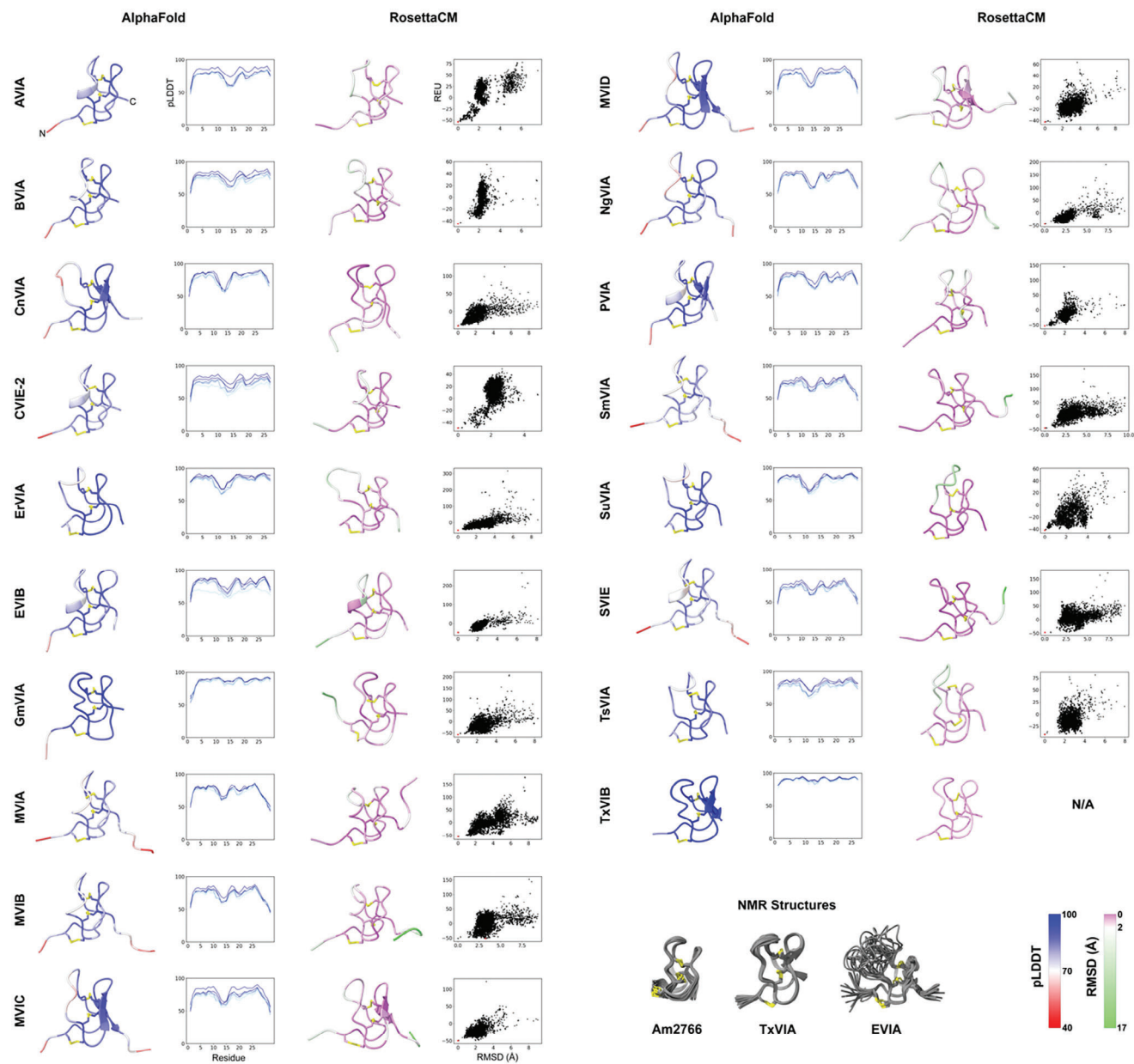
Figure 2. Results of the modeling pipeline for all peptides. Disulfide bonds are highlighted in yellow. From left: best-scoring model produced by AlphaFold, colored by per-residue pLDDT score; a plot of pLDDT per residue for the models generated by AlphaFold. Models with higher overall pLDDT scores are in darker blue; the best-scoring model produced by RosettaCM, colored by per-residue root-mean-square deviation (RMSD) to representatives of the top 5 best-scoring clusters; score-RMSD funnel plot for all 5000 decoys produced by RosettaCM, with score measured in Rosetta energy units (REU) and RMSD calculated by distance to the lowest-energy model, and with the displayed model indicated in red; Below: experimentally-determined NMR ensembles used as templates in this study for RosettaCM modeling. RMSD and pLDDT color bars are shown bottom right.

only a single conformation was used and, consequently all the structural diversity observed in this loop has been independently derived by the modeling algorithms.

While loop 2 is by far the longest loop in EVIA (9 amino acids), most of the other conotoxins in this study have a shorter loop 2 (6 amino acids) with the exceptions being ErVIA, SuVIA, and TsVIA from vermivorous snails (7 amino acids). It is therefore notable that the other peptides retain flexibility in this loop despite a shorter loop length; loop 1 has considerably lower pre-

dicted RMSDs than loop 2 (Figure S2, Supporting Information), despite also consisting of 6 amino acids in all cases.

This region is comparatively rigid in δ-conotoxins from molluscivorous snails. Examining the predicted structures for GmVIA reveals a role for the sidechain of Q16, which can form hydrogen bonds to the backbone carbonyl of P13 and the backbone amide of C24 (Figure S3A, Supporting Information). The AlphaFold models additionally show hydrogen bonding to the C24 backbone carbonyl. This glutamine is conserved across all

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SCIENCE**
Open Access

www.advancedscience.com

δ-conotoxins from molluscivorous peptides, and similar hydrogen bonding arrangements are found in some conformations of Am2766 (Figure S3B, Supporting Information) and TxVIA.[11,12] These peptides also lack the glycine immediately following C2 that is conserved across most δ-conotoxins from piscivorous cone snails which could contribute to the additional flexibility of loop 2 in these peptides.

2.5. Not All Peptides Show a Continuum of Positions in Loop 2

For some δ-conotoxins we do not observe a continuum of positions for loop 2 across the full range of motion, as might be expected for a disordered region where there are low energetic barriers to movement. In these cases, which are restricted to some δ-conotoxins from piscivorous cone snails, we observe that the models are clustered around a few discrete positions (Figure S4, Supporting Information). This suggests that the loop may be able to interconvert between different populations, with implications for bioactivity (see Discussion).

2.6. A Structural Role for Hydroxyproline Residues in RosettaCM Models

Post-translational modifications are common among conotoxins from all functional classes. Among the δ-conotoxins in this study, 4-hydroxyproline is modeled at position 6 (in the middle of loop 1) in five peptides, and at position 14 (middle of loop 2) in 12 peptides. Of these two Hyp residues, the first has a clear structural rationale: the additional hydroxyl group on the sidechain of Hyp6 forms a hydrogen bond with the sidechain of the conserved S19, where it mitigates any flexibility of loop 1 as well as further stabilizing S19, which has a role in forming the tight turn at the start of loop 3 (Figure S3C, Supporting Information).

While this Hyp residue has a clear structural justification, the same cannot be said for the Hyp residue on loop 2. This loop is flexible (see above) but in all conformations, the hydroxyl group is directed away from the rest of the peptide where it cannot form direct intramolecular interactions. This raises the possibility that it either contacts the target Na_v or influences the structure less directly.

2.7. δ-Conotoxins with an Extended Loop 2 Show a Mixture of Cis- and Trans-Proline

Cis-prolines are uncommon in protein structures due to the increased energy of this conformation relative to trans-proline, and their prevalence is estimated to be $\approx$6% of all prolines.[34] It is therefore significant that in the RosettaCM models of two δ-conotoxins in this study with an extended loop 2, a conserved Pro at position 11 shows a mixture of cis and trans conformations in the 5 best-scoring clusters (Figure S5, Supporting Information). This result extends published NMR studies of EVIA, which demonstrated a mixture of cis- and trans-prolines at the equivalent position in an $\approx$1:1 ratio with slow interconversion.[13,14] One NMR study also demonstrated that a P13A EVIA mutant produced a 100% trans conformation at this bond, and also

halved the potency of the peptide compared to the wild type in a competitive binding assay using ^{125}I-labeled TxVIA on rat brain synaptosomes.[13] This suggests that either the sidechain makes direct contact with the target, or that the cis-proline conformation is important for positioning other interacting residues.

Notably, the cis-proline bonds are only found in the RosettaCM models; the equivalent AlphaFold models show only the trans conformation. Comparing the top-scoring cluster for TsVIA (which has a cis-proline at position 11) with the corresponding top-ranked AlphaFold model shows that the cis-proline bond allows the L10 and V14 sidechains to project toward loop 3 and be partially buried in the peptide core (Figure S3D, Supporting Information). By contrast, the trans-proline conformation projects the sidechains away from the peptide core where they are fully exposed to solvent; it can therefore be hypothesized that the cis-trans isomerization at proline in this loop balances the higher-energy cis-conformation with the energetic penalty from fully exposed hydrophobic leucine and valine sidechains.

2.8. A Consensus Surface-Exposed Hydrophobic Patch

The presentation of hydrophobic sidechains on the solvent-accessible surface is disfavored in most protein structures but is a common feature of ICK peptides; it has been hypothesized that the reason for forming such a complex, covalently-linked peptide core is to force exposure of these residues, even against the energetic penalty.[35–37] These hydrophobic patches are often important for binding.[38,39] Examining the available experimental structures of δ-conotoxins,[11–13] together with our models, we find a common hydrophobic patch on the peptide surface (Figure S6, Supporting Information). Sequence alignments of the δ-conotoxins show that these regions are uniformly hydrophobic, across all types of cone snails. This consensus strongly suggests that these residues are important for binding the peptides to the target, including the "silent binding" by the δ-conotoxins from molluscivorous cone snails.

2.9. Some RosettaCM Models of Piscivorous Cone Snail δ-Conotoxins Show a Cis-nonPro Bond in Loop 2

Several of the top-ranked RosettaCM models (but none of the equivalent AlphaFold models) contained a non-proline cis-peptide bond at the conserved glycine-isoleucine motif in loop 2 immediately following the C2-C5 disulfide bond. Peptides that lacked the GI motif, such as GmVIA and TxVIB (derived from molluscivorous cone snails) and BVIA (from a piscivorous snail), did not contain a cis-peptide bond; neither did any of the peptides derived from vermivorous snails (TsVIA, ErVIA and SuVIA) even though these peptides contain a similar GL motif as part of a longer loop 2. Given the rarity of this structural feature in experimentally derived protein structures, we repeated the RosettaCM modeling protocol with a further two different score functions to investigate methodological explanations for its appearance in our model set (see Methods), only to obtain similar results (Table S1, Supporting Information). Although none of the AlphaFold models contained non-proline cis-peptide bonds, in some instances a twisted (non-planar) peptide bond was observed in the C-terminal regions which suggests some difficulties

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

ADVANCED
SCIENCE
Open Access

www.advancedscience.com

in modeling correct geometries in regions of higher uncertainty (see Discussion).[40]

3. Discussion

In this study we calculated models for δ-conotoxins without experimental structures, using both AlphaFold and RosettaCM. The models show many structural features that are well-known properties of all δ-conotoxins, such as the disulfide bond connectivity typical of ICK peptides and surface-exposed hydrophobic patch that arises from the rigidity of this network. However, several differences are observed between the δ-conotoxins derived from piscivorous cone snails compared with those from molluscivorous cone snails, which provide suggestions for the structural basis of the "silent binding" phenomenon.

The most notable difference between the two classes of peptides is the flexibility observed for loop 2, which is seen in all δ-conotoxins from piscivorous and vermivorous cone snails, but not among the δ-conotoxins from molluscivorous snails (Figure S2, Supporting Information). A previous structure-activity relationship study of PVIA noted that the mutants P9A and I12A – both highly conserved residues in δ-conotoxins from piscivorous snails but not molluscivorous snails – retained the ability to bind to Na$_v$s but lost their pharmacological properties.[9] Since these residues are located on loop 2, the flexibility of this loop observed in our models raises the question of how these residues are positioned to interact with the target. Since F9 is located on loop 1, the AlphaFold models show a largely consistent position for the sidechain. I12 shows greater variance; in some models, it is orientated away from loop 2 and the disulfide-bonded core and closer to the F9 sidechain, whereas in others it is directed inward and positioned closer to loop 3. We speculate that, given the relative exposure to the solvent of the I12 sidechain in the vertical position, as well as its closer positioning to the bioactive residues F9, the former orientation is more likely to be bioactive.

Examining the sequences of other classes of peptides known to elicit similar pharmacological properties when binding to Na$_v$s is also illustrative. A study of Australian funnel-web spiders uncovered 22 peptides of the δ-hexatoxin class.[41] Like the δ-conotoxins, these peptides similarly inhibit the inactivation of voltage-gated sodium channels (including in vertebrates) and are thought to share a binding site on Na$_v$ voltage-sensing domain (VSD) IV; although they share a common ICK disulfide bonding core, the δ-hexatoxins are typically larger (with extended loop 4 and N- and C-terminal domains) and possess an additional disulfide bond.[30,31] As with other ICK peptides, the in vivo folding of these peptides is likely mediated by the processing of the signal and propeptides,[42,43] and by enzymes including prolyl isomerases[44] and protein disulfide isomerases.[45] Sequence alignments of 14 representative δ-hexatoxins with the δ-conotoxins in this study show that an aromatic residue is conserved at the equivalent of F9 in all peptides (typically it is tryptophan), apart from the δ-conotoxins from molluscivorous cone snails, strongly implicating this residue in bioactivity (Figure S7, Supporting Information). Interestingly the equivalent position to I12, a conserved residue in δ-conotoxins that has also been shown

to affect bioactivity, is a charged residue (most commonly lysine) in the δ-hexatoxins.

The δ-conotoxins have a low net charge, ranging from −3 to +2 (Figure 1A); this contrasts with other peptide venom toxins, such as the μ-conotoxins or spider venom ICK peptides, which often have a very high positive net charge.[46] This likely contributes to the low aqueous solubility and refolding yields of these peptides, which has hindered their structural and functional characterization. Since a high net positive charge has been shown to be an important determinant of potency against voltage-gated sodium channels for related tarantula-derived ICK peptides due to a membrane-first binding mechanism,[47,48] the low charge of the δ-conotoxins would suggest that a different binding mechanism applies.

Although both piscivorous and molluscivorous snail-derived δ-conotoxins share a low overall charge, the charged residues are distributed differently across the loops, and recent docking studies have implicated these charged residues as determinants of channel binding.[14,49] Piscivorous snail δ-conotoxins usually have acidic residues at the N-terminus, on loop 3, and sometimes at the C-terminus, while the basic residues are limited to a single position on loop 2 (Figure 1A). Molluscivorous snail-derived δ-conotoxins, however, universally have a basic residue immediately after C1, while acidic residues are located near the end of loop 1, and immediately after C2. In cases where these residues are not negatively charged a polar residue is often found in its place. The conserved positioning of the basic residue in peptides from piscivorous snails, in a loop that is flexible and known to contain residues important for the inhibitory abilities of the peptide, suggests a role in the mechanism of action of these peptides. The supposed binding site for δ-conotoxins[6] (above VSD IV of voltage-gated sodium channels) permits access to the ladder of charged amino acids within VSD IV that are responsible for the movement of the domain on activation. Disruption of this ladder by basic residues is a known mechanism of channel inhibition by other venom peptides, and could play a role in inhibiting inactivation.[50–52]

Related to the positioning of these residues on loop 2 is the cis/trans-isomerization at the conserved proline on this loop among δ-conotoxins from piscivorous snails. The appearance of the cis-isomer is restricted to RosettaCM models of peptides with an extended loop 2 (i.e., longer than 6 residues) (Figure S5, Supporting Information), suggesting that the cis-conformation cannot form in the tighter and more constrained 6-residue loop that is most prevalent for loop 2. One possible functional explanation for the formation of the cis-isomer was suggested by Tietze et al.,[14] who hypothesized that the cis conformation was likely to increase binding affinity by presenting both a hydrophobic side chain and a backbone carbonyl (capable of accepting a hydrogen bond) to the Na$_v$.

All AlphaFold models showed exclusively trans-proline. To our knowledge, a systematic study of the propensity of AlphaFold to predict cis-prolines has not yet been performed, but in CASP14[53] it correctly predicted 42 trans-prolines and 3 out of 4 cis-prolines,[54] suggesting that AlphaFold performs well at identifying cis-prolines in its targets. However, a study of 91 conformationally heterogeneous proteins showed that AlphaFold predictions were consistently biased toward only a single ground

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SCIENCE**
Open Access

www.advancedscience.com

state model and failed to model alternative conformations in 94% of cases, even as the confidence metrics for the predicted ground state remained high (in contrast to intrinsically disordered proteins).[55] The authors attribute this to AlphaFold's preference for modeling the "most probable" conformer, noting that AlphaFold is trained predominantly on X-ray crystal structures of stably folded proteins and that structure prediction algorithms frequently assume that a single amino acid sequence generates a single stable protein fold.[56] In the case of the δ-conotoxins this would produce a strong preference for the trans-proline isomer since this conformation is predominant in experimental protein structures, even as experimental NMR studies of the δ-conotoxin EVIA show both cis- and trans-proline conformations, with slow interconversion.[13,14]

Cis-peptide bonds at residues other than proline are extremely unusual in protein structures and their appearance in our set of best-ranked RosettaCM models prompted further scrutiny. To verify the initial models we repeated the modeling using two further score functions and an alternative refinement strategy, but nonetheless obtained similar results across both new datasets. The non-proline cis-peptides seen in the top-ranked models all occur at the same position – the glycine-isoleucine bond immediately following the C2-C5 disulfide bond in conotoxins derived from piscivorous cone snails. Peptides that lack this motif or that have a longer loop 2 do not have cis-peptide bonds anywhere in the model. These results are consistent with the observation that ≈80% of characterized cis-Xaa-nonPro bonds have glycine prior to the cis-peptide bond,[57] since the steric penalty is minimized when the clashing sidechain is a proton. Despite their rarity, non-proline cis-peptide bonds have been characterized in NMR structures of α-scorpion toxins;[58–60] these peptides similarly inhibit Na$_v$ inactivation and also bind to the extracellular loops of VSD IV.

As previously noted, the AlphaFold models do not contain any cis-peptide bonds. However, in a handful of cases the AlphaFold models have "twisted" peptide bonds ($\omega > \pm 30°$ deviation from planar) which are practically nonexistent in experimental protein structures (Table S1, Supporting Information). It has been shown that AlphaFold is more likely to predict these model geometries in low-confidence regions (pLDDT < 60)[40] and this is also observed in our dataset, with the twisted peptides exclusively occurring in the disordered C-termini; similar features are seen in models of these peptides in the AlphaFold database.[28] Nonetheless, the conclusions in this paper are solely derived from models without cis or non-planar peptide bonds.

In this work, we use both a deep learning algorithm (AlphaFold) and a conventional homology and energy minimization strategy (RosettaCM) to generate our model sets. Each method has its strengths and weaknesses when modeling these complex peptides. Threading algorithms are at their most accurate when modeling sequences with high homology to the available templates, as is often the case with ICK peptides given the universality of this framework across different species; RosettaCM is particularly well suited for this given its ability to use multiple template structures simultaneously. Conotoxins often have extensive post-translational modifications; C-terminal amides and 4-hydroxyproline residues were present in our target set but other conotoxin classes are known to have different modifications, including D-amino acids, γ-carboxylation of glutamic acid residues, and L-6-bromination of tryptophan residues.[61] Rosetta has an extensive library of patches that enable these modifications to be handled natively during modeling which offers a significant advantage.

AlphaFold is a newer modeling algorithm that uses deep learning methods to calculate the structure using an MSA, residue-pair interaction data, and, optionally, experimental structures as templates. AlphaFold's chief advantage is its accuracy: in the CASP assessments of protein structure modeling methods in which AlphaFold has participated it considerably outperformed other algorithms.[53] It also has the advantage that it can predict structures based on the sequence data alone, given a sufficiently large MSA, and does not entirely rely on experimental structures as homology methods do. Given the challenges in experimental characterization presented by certain conotoxins, this could significantly improve model quality. While the original AlphaFold algorithm has been extended to improve modeling of protein-protein and protein-ligand complexes,[62] a full treatment of post-translational modifications is not available; this presents a drawback in the modeling of conotoxins given the frequency and importance of post-translational modifications in conotoxin structure and function.

A limiting factor in the development of drugs based on conopeptides is the prevalence of off-target activity caused by the high structural similarity of receptor subtypes, and a thorough understanding of the structure-activity relationship is essential for mitigating these side effects.[63] Structural information is also essential in developing agents capable of countering the effects of cone snail envenomation, for which there is currently no dedicated treatment.[64] We anticipate that the structural models presented in this work will provide a strong foundation for docking studies, molecular dynamics simulations, and experimental characterization of peptide-channel complexes that will further shed light on the mechanism of action of these peptides and contribute to drug discovery.

4. Conclusion

Our computational study has expanded the range of δ-conotoxins with modeled structures, including a class of δ-conotoxins from piscivorous cone snails for which experimental structures are sparse and challenging to obtain. We note intriguing similarities and differences between peptides originating from piscivorous, molluscivorous, and vermivorous cone snails and identify features likely to be of relevance to their binding and activity against their target Na$_v$s, including the "silent binding" phenomenon. Our protocol additionally suggests a roadmap for modeling similar disulfide-rich peptides. Further structural characterization of these δ-conotoxins, both in isolation and in complex, will shed additional light on their mechanism of action and provide new avenues for rational drug design.

5. Experimental Section

Database Searching, Template Selection, and Preparation: δ-conotoxins were identified using ConoServer.[65,66] Peptides annotated with the cysteine framework VI/VII and pharmacological family δ were selected; synthetic constructs were excluded. Three peptides, Am2766, TxVIA,

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SCIENCE**
Open Access

www.advancedscience.com

and EVIA, have experimental structures and were designated as templates for modeling the remaining δ-conotoxins in RosettaCM.[11–14] 4-Hydroxyproline residues were designated as prolines for sequence alignments. Due to the very high sequence similarity to TxVIA, conotoxin TxVIB was modeled using a different protocol (see below).

All template peptide structures were determined by nuclear magnetic resonance (NMR) and therefore comprise an ensemble of structures consistent with the calculated restraints. To prepare a single input model per template for RosettaCM, the ensemble structures were separated into individual PDB files and cleaned using the clean_PDB.py script supplied with Rosetta. For 1YZ2, which has a C-terminal amide, the OXT atom was replaced in one round of Rosetta FastRelax, and then edited to restore a C-terminal amide. All structures were then energy-minimized (5 models per input structure) using Rosetta Relax,[67,68] and the lowest-energy model was selected as the template.

Modeling with AlphaFold: AlphaFold v2.0 was obtained (https://github.com/google-deepmind/alphafold; 2021-07-14 parameters) and modeling was performed with the –max_template_date = 2021-09-01 and –preset = casp14 flags, and the use of experimental structures as templates was permitted. Hydroxyproline residues were modeled as unmodified prolines. The five ranked models produced were assessed for model quality by their per-residue pLDDT scores and model geometry was assessed with MolProbity.[69]

Model Generation in RosettaCM: Sequences were obtained from ConoServer. Since sequence data for some of these peptides is obtained through genomic analyses, which do not indicate the presence or absence of post-translational modifications, hydroxyproline residues, and C-terminal amides were modeled where annotations in ConoServer indicated their likely presence; note that some of these may be derived by similarity and do not guarantee a modification at that position.

Structure prediction in RosettaCM requires a sequence alignment between the target and template protein(s) and proceeds via an initial threading step that overlays the aligned regions of the target protein onto the template.[27] These threaded models are then hybridized to sample combinations of regions from the different threaded models and close missing loops by fragment insertion, followed by energy minimization and scoring of the final structures.

Sequence alignments were generated using ClustalΩ[70] and adjusted where necessary to ensure alignment of the cysteine residues. The phylogenetic tree (Figure 1B) was calculated using the phylogeny tool ClustalW2.[71,72]

All scripts and flags used are provided as Supplementary Material.

Rosetta version 3.12 was obtained from Rosetta Commons (www.rosettacommons.org) and installed from source. Initial threaded models were calculated for each peptide using the setup_RosettaCM.py script provided with Rosetta,[27] then 1000 initial models were calculated using the rosetta_cm.xml script. This script was modified where necessary to incorporate 4-hydroxyproline residues and C-terminal amides using the ModifyVariantType mover. The energy of the models was minimized using the Relax application with the current ref2015 score function,[73] generating 5 relaxed models per input model for a total of 5000 models per peptide. The -fix_disulfs flag was used, together with a parameters file defining the disulfide bond connectivity, to ensure the formation of the correct cysteine framework.

This protocol was similarly followed when modeling using the cartesian score functions beta_nov16_cart and ref2015_cart,[74] except the number of output hybridized models was set to 5000, and the -relax:dualspace flag was included to instigate dualspace refinement.

Model Clustering and Scoring: Models were clustered using Rosetta's energy-based clustering algorithm,[75] with a cluster radius of 1 Å. The 5 lowest-energy clusters were used for analysis, and decoys were aligned to the lowest-scoring model in ChimeraX.[76,77]

Modeling of TxVIB: The sequence of TxVIB is identical to that of TxVIA except for two mutations (L11V/L24F).[78,79] Modeling of TxVIB was therefore carried out using the mutagenesis tool in PyMOL[80] to mutate the two residues, followed by energy minimization using Rosetta Relax (500 models). The lowest-scoring model was selected for analysis. The standard AlphaFold modeling protocol was used for TxVIB.

Figures were produced in ChimeraX, PyMOL, Microsoft PowerPoint (microsoft.com), and Adobe Photoshop (adobe.com). Plots were produced using Matplotlib.[81]

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords

[1] J. Davis, A. Jones, R. J. Lewis, *Peptides* **2009**, *30*, 1222.

[2] R. J. Lewis, M. L. Garcia, *Nat. Rev. Drug Discov.* **2003**, *2*, 790.

[3] S. Dutertre, R. Lewis, *Snails: Biology, Ecology and Conservation*, (Eds: E. M. Hämäläinen, S. Järvinen) Nova Science Publishers, New York, NY **2013**, pp. 85–104.

[4] J. Barbier, H. Lamthanh, F. Le Gall, P. Favreau, E. Benoit, H. Chen, N. Gilles, N. Ilan, S. H. Heinemann, D. Gordon, A. Ménez, J. Molgó, *J. Biol. Chem.* **2004**, *279*, 4680.

[5] M. Fainzilber, O. Kofman, E. Zlotkin, D. Gordon, *J. Biol. Chem.* **1994**, *269*, 2574.

[6] E. Leipold, A. Hansel, B. M. Olivera, H. Terlau, S. H. Heinemann, *FEBS Lett.* **2005**, *579*, 3881.

[7] J. Wang, V. Yarov-Yarovoy, R. Kahn, D. Gordon, M. Gurevitz, T. Scheuer, W. A. Catterall, *Proc. Natl. Acad. Sci.* **2014**, *111*, 3644.

[8] S. Sudarslal, S. Majumdar, P. Ramasamy, R. Dhawan, P. P. Pal, M. Ramaswami, A. K. Lala, S. K. Sikdar, S. P. Sarma, K. S. Krishnan, P. Balaram, *FEBS Lett.* **2003**, *553*, 209.

[9] G. Bulaj, R. DeLaCruz, A. Azimi-Zonooz, P. West, M. Watkins, D. Yoshikami, B. M. Olivera, *Biochemistry* **2001**, *40*, 13201.

[10] M. Fainzilber, J. C. Lodder, K. S. Kits, O. Kofman, I. Vinnitsky, J. Van Rietschoten, E. Zlotkin, D. Gordon, *J. Biol. Chem.* **1995**, *270*, 1123.

[11] T. Kohno, T. Sasaki, K. Kobayashi, M. Fainzilber, K. Sato, *J. Biol. Chem.* **2002**, *277*, 36387.

[12] S. P. Sarma, G. S. Kumar, S. Sudarslal, P. Iengar, P. Ramasamy, S. K. Sikdar, K. S. Krishnan, P. Balaram, *Chem. Biodivers.* **2005**, *2*, 535.

[13] L. Volpon, H. Lamthanh, J. Barbier, N. Gilles, J. Molgó, A. Ménez, J. M. Lancelin, *J. Biol. Chem.* **2004**, *279*, 21356.

[14] D. Tietze, E. Leipold, P. Heimer, M. Böhm, W. Winschel, D. Imhof, S. H. Heinemann, A. A. Tietze, *Biochim. Biophys. Acta – Gen. Subj.* **2016**, *1860*, 2053.

[15] B. Webb, A. Sali, *Curr. Protoc. Bioinforma.* **2016**, *54*, 5.6.1.

[16] B. Kuhlman, P. Bradley, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681.

[17] J. Lee, P. L. Freddolino, Y. Zhang, in *From Protein Struct. to Funct. with Bioinforma* (Ed.: D. J. Rigden), Springer, Netherlands, Dordrecht **2017**, pp. 1–35.

[18] L. Aithani, E. Alcaide, S. Bartunov, C. D. O. Cooper, A. S. Doré, T. J. Lane, F. Maclean, P. Rucktooa, R. A. Shaw, S. E. Skerratt, *Curr. Opin. Struct. Biol.* **2023**, *80*, 102601.

[19] N. Sapoval, A. Aghazadeh, M. G. Nute, D. A. Antunes, A. Balaji, R. Baraniuk, C. J. Barberan, R. Dannenfelser, C. Dun, M. Edrisi, R. A. L. Elworth, B. Kille, A. Kyrillidis, L. Nakhleh, C. R. Wolfe, Z. Yan, V. Yao, T. J. Treangen, *Nat. Commun.* **2022**, *13*, 1728.

[20] N. Kumar, R. Srivastava, *Brief Bioinform* **2024**, https://doi.org/10.1093/bib/bbae042.

[21] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, et al., *Nature* **2021**, *596*, 583.

[22] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. Dustin Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. Van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. Christopher Garcia, N. V. Grishin, P. D. Adams, et al., *Science* **2021**, *373*, 871.

[23] R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, A. Kang, E. Brackenbrough, A. K. Bera, M. Baek, F. DiMaio, D. Baker, *Science* **2024**, https://doi.org/10.1126/science.adl2528.

[24] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, *Science* **2023**, *379*, 1123.

[25] R. Wu, F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, J. Ma, J. Peng, *bioRxiv* **2022**, https://doi.org/10.1101/2022.07.21.500999.

[26] Y. Fang, Y. Jiang, L. Wei, Q. Ma, Z. Ren, Q. Yuan, D. Q. Wei, *Bioinformatics* **2023**, https://doi.org/10.1093/bioinformatics/btad718.

[27] Y. Song, F. Di Maio, R. Y. R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, D. Baker, *Structure* **2013**, *21*, 1735.

[28] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Res.* **2022**, *50*, D439.

[29] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, et al., *Nature* **2021**, *596*, 590.

[30] V. Laverne, P. F. Alewood, M. Mobli, G. F. King, in *Venoms to Drugs Venom as a Source Dev Hum Ther* (Ed.: G. F. King), Royal Society Of Chemistry, Cambridge, **2015**, pp. 37–70.

[31] S. S. Pineda, Y. K. Y. Chin, E. A. B. Undheim, S. Senff, M. Mobli, C. Dauly, P. Escoubas, G. M. Nicholson, Q. Kaas, S. Guo, V. Herzig, J. S. Mattick, G. F. King, *Proc. Natl. Acad. Sci.* **2020**, *117*, 11399.

[32] A. O. Stevens, Y. He, *Biomolecules* **2022**, *12*, 985.

[33] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, S. K. Teixeira, *Front Bioinform.* **2023**, *3*, 1120370.

[34] A. A. Morgan, E. Rubenstein, *PLoS One* **2013**, *8*, 1.

[35] G. F. King, *Venom to Drugs: Venom as a Source for the Development of Human Therapeutics*, Royal Society of Chemistry, Cambridge **2015**.

[36] E. A. B. Undheim, M. Mobli, G. F. King, *BioEssays* **2016**, *38*, 539.

[37] M. H. Moreira, F. C. L. Almeida, T. Domitrovic, F. L. Palhano, *Comput. Struct. Biotechnol. J* **2021**, *19*, 6255.

[38] C. H. Y. Lau, G. F. King, M. Mobli, *Sci. Rep.* **2016**, *6*, 34333.

[39] J. J. Smith, T. R. Cummins, S. Alphy, K. M. Blumenthal, *J. Biol. Chem.* **2007**, *282*, 12687.

[40] C. J. Williams, D. C. Richardson, J. S. Richardson, *Comput Crystallogr Newsl* **2022**, *13*, 7.

[41] V. Herzig, K. Sunagar, D. T. R. Wilson, S. S. Pineda, M. R. Israel, S. Dutertre, B. S. McFarland, E. A. B. Undheim, W. C. Hodgson, P. F. Alewood, R. J. Lewis, F. Bosmans, I. Vetter, G. F. King, B. G. Fry, *Proc. Natl. Acad. Sci.* **2020**, *117*, 24920.

[42] G. Bulaj, B. M. Olivera, *Antioxidants Redox Signal* **2008**, *10*, 141.

[43] O. Buczek, B. M. Olivera, G. Bulaj, *Biochemistry* **2004**, *43*, 1093.

[44] H. Safavi-Hemami, G. Bulaj, B. M. Olivera, N. A. Williamson, A. W. Purcell, *J. Biol. Chem.* **2010**, *285*, 12735.

[45] H. Safavi-Hemami, D. G. Gorasia, A. M. Steiner, N. A. Williamson, J. A. Karas, J. Gajewiak, B. M. Olivera, G. Bulaj, A. W. Purcell, *J. Biol. Chem.* **2012**, *287*, 34288.

[46] R. J. Lewis, S. Dutertre, I. Vetter, M. J. Christie, *Pharmacol. Rev.* **2012**, *64*, 259.

[47] S. T. Henriques, E. Deplazes, N. Lawrence, O. Cheneval, S. Chaousis, M. Inserra, P. Thongyoo, G. F. King, A. E. Mark, I. Vetter, D. J. Craik, C. I. Schroeder, *J. Biol. Chem.* **2016**, *291*, 17049.

[48] E. Deplazes, S. T. Henriques, J. J. Smith, G. F. King, D. J. Craik, A. E. Mark, C. I. Schroeder, *Biochim. Biophys. Acta – Biomembr.* **2016**, *1858*, 872.

[49] D. Wang, S. W. A. Himaya, J. Giacomotto, M. M. Hasan, F. C. Cardoso, L. Ragnarsson, R. J. Lewis, *Mar. Drugs* **2020**, *18*, 343.

[50] H. Xu, T. Li, A. Rohou, C. P. Arthur, F. Tzakoniati, E. Wong, A. Estevez, C. Kugel, Y. Franke, J. Chen, C. Ciferri, D. H. Hackos, C. M. Koth, J. Payandeh, *Cell* **2019**, *176*, 702.

[51] T. Clairfeuille, A. Cloake, D. T. Infield, J. P. Llongueras, C. P. Arthur, Z. R. Li, Y. Jian, M. F. Martin-Eauclaire, P. E. Bougis, C. Ciferri, C. A. Ahern, F. Bosmans, D. H. Hackos, A. Rohou, J. Payandeh, *Science* **2019**, https://doi.org/10.1126/science.aav8573.

[52] K. George, D. Lopez-Mateos, T. M. Abd El-Aziz, Y. Xiao, J. Kline, H. Bao, S. Raza, J. D. Stockand, T. R. Cummins, L. Fornelli, M. P. Rowe, V. Yarov-Yarovoy, A. H. Rowe, *Front. Pharmacol.* **2022**, *13*, 846992.

[53] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, *Proteins Struct Funct Bioinforma* **2021**, *89*, 1607.

[54] O. Herzberg, J. Moult, *Proc. Natl. Acad. Sci.* **2023**, *120*, 2017.

[55] D. Chakravarty, L. L. Porter, *Protein Sci.* **2022**, *31*, e4353.

[56] K. Manalastas-Cantos, K. R. Adoni, M. Pfeifer, B. Märtens, K. Grünewald, K. Thalassinos, M. Topf, *Mol. Cell. Proteomics* **2024**, *23*, 100724.

[57] S. Das, S. Ramakumar, D. Pal, *FEBS J.* **2014**, *281*, 5602.

[58] K. S. Mineev, M. A. Chernykh, V. V. Motov, D. A. Prudnikova, D. M. Pavlenko, A. I. Kuzmenkov, S. Peigneur, J. Tytgat, A. A. Vassilevski, *FEBS Lett.* **2023**, *597*, 2358.

[59] R. J. Guan, Y. Xiang, X. L. He, C. G. Wang, M. Wang, Y. Zhang, E. J. Sundberg, D. C. Wang, *J. Mol. Biol.* **2004**, *341*, 1189.

[60] N. A. Kuldyushev, K. S. Mineev, A. A. Berkut, S. Peigneur, A. S. Arseniev, J. Tytgat, E. V. Grishin, A. A. Vassilevski, *Proteins Struct. Funct. Bioinforma.* **2018**, *86*, 1117.

[61] O. Buczek, G. Bulaj, B. M. Olivera, *Cell. Mol. Life Sci.* **2005**, *62*, 3067.

[62] AlphaFold Team, Google DeepMind Team, Isomorphic Labs, *Performance and Structural Coverage of the Latest, in-Development AlphaFold Model* **2023**, 1–20.

[63] R. J. Lewis, in *Venom to Drugs Venom as a Source Dev Hum Ther* (Ed.: G. F. King), Royal Society of Chemistry, Cambridge, **2015**, pp. 245.

[64] Z. A. Halford, P. Y. C. Yu, R. K. Likeman, J. S. Hawley-Molloy, C. Thomas, J. P. Bingham, *Diving Hyperb. Med.* **2015**, *45*, 200.

[65] Q. Kaas, R. Yu, A. H. Jin, S. Dutertre, D. J. Craik, *Nucleic Acids Res.* **2012**, *40*, D325.

[66] Q. Kaas, J. C. Westermann, R. Halai, C. K. L. Wang, D. J. Craik, *Bioinformatics* **2008**, *24*, 445.

[67] L. G. Nivón, R. Moretti, D. Baker, *PLoS One* **2012**, *8*, e59004.

[68] P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, *Protein Sci.* **2014**, *23*, 47.

[69] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res.* **2007**, *35*, W375.

[70] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.

[71] M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez, *Nucleic Acids Res.* **2010**, *38*, W695.

[72] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, *Bioinformatics* **2007**, *23*, 2947.

[73] R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, *J. Chem. Theory Comput.* **2017**, *13*, 3031.

[74] J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, et al., *Nat. Methods* **2020**, *17*, 665.

[75] P. Hosseinzadeh, G. Bhardwaj, V. K. Mulligan, M. D. Shortridge, T. W. Craven, F. Pardo-Avila, S. A. Rettie, D. E. Kim, D. A. Silva, Y. M. Ibrahim, I. K. Webb, J. R. Cort, J. N. Adkins, G. Varani, D. Baker, *Science* **2017**, *358*, 1461.

[76] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, *Protein Sci.* **2021**, *30*, 70.

[77] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, *Protein Sci.* **2018**, *27*, 14.

[78] M. Fainzilber, D. Gordon, A. Hasson, M. E. Spira, E. Zlotkin, *Eur. J. Biochem.* **1991**, *202*, 589.

[79] A. Hasson, M. Fainzilber, D. Gordon, E. Zlotkin, M. E. Spira, *Eur. J. Neurosci.* **1993**, *5*, 56.

[80] The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC, https://pymol.org/support.html.

[81] J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*, 90.

2404786 (10 of 10)