Received: 2 November 2017 | Revised: 16 January 2018 | Accepted: 29 January 2018

DOI: 10.1002/prot.25473

RESEARCH ARTICLE

WILEY PROTEINS

Large scale ab initio modeling of structurally uncharacterized antimicrobial peptides reveals known and novel folds

Mara Kozic[1,2] | Stephen J. Fox[2] | Jens M. Thomas[1] | Chandra S. Verma[2,3,4] | Daniel J. Rigden[1]

[1]Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, U.K.

[2]Agency for Science, Technology and Research (A*STAR), Bioinformatics Institute, Singapore

[3]Department of Biological Sciences, National University of Singapore, Singapore

[4]School of Biological Sciences, Nanyang Technological University, Singapore

Correspondence
Daniel J. Rigden, Institute of Integrative Biology, Room 101, Biosciences Building, University of Liverpool, Crown St., Liverpool L69 7ZB, U.K.
E-mail: drigden@liverpool.ac.uk

Abstract

Antimicrobial resistance within a wide range of infectious agents is a severe and growing public health threat. Antimicrobial peptides (AMPs) are among the leading alternatives to current antibiotics, exhibiting broad spectrum activity. Their activity is determined by numerous properties such as cationic charge, amphipathicity, size, and amino acid composition. Currently, only around 10% of known AMP sequences have experimentally solved structures. To improve our understanding of the AMP structural universe we have carried out large scale ab initio 3D modeling of structurally uncharacterized AMPs that revealed similarities between predicted folds of the modeled sequences and structures of characterized AMPs. Two of the peptides whose models matched known folds are Lebocin Peptide 1A (LP1A) and Odorranain M, predicted to form β-hairpins but, interestingly, to lack the intramolecular disulfide bonds, cation-π or aromatic interactions that generally stabilize such AMP structures. Other examples include Ponericin Q42, Latarcin 4a, Kassinatuerin 1, Ceratotoxin D, and CPF-B1 peptide, which have α-helical folds, as well as mixed αβ folds of human Histatin 2 peptide and Garvicin A which are, to the best of our knowledge, the first linear αββ fold AMPs lacking intramolecular disulfide bonds. In addition to fold matches to experimentally derived structures, unique folds were also obtained, namely for Microcin M and Ipomicin. These results help in understanding the range of protein scaffolds that naturally bear antimicrobial activity and may facilitate protein design efforts towards better AMPs.

KEYWORDS

ab initio modeling, antimicrobial peptide, antimicrobial resistance, protein structure-function, structure prediction

1 | INTRODUCTION

Antimicrobial resistance within a wide range of infectious agents is a severe and growing public threat.[1] A 2013 report from the American Centers for Disease Control estimated that over 23 000 deaths and more than 2 million cases of infections were caused by drug-resistant bacteria in the USA alone in 2013.[2] Antimicrobial proteins and peptides are among the leading alternatives to current antibiotics, exhibiting activity against a wide variety of bacteria and other microbes[3] and are of particular interest since they have maintained their effectiveness over hundreds of millions of years demonstrating that definitive resistance to them is not readily acquired by bacteria. Proteins with antimicrobial activity typically contain fewer than 200 residues, with most much shorter—12–100 residues—and hence are commonly known as antimicrobial peptides (AMPs).[4,5] They are produced by the immune systems of species from all domains of life. Most of them are cationic at physiological pH, with net positive charge ranging from +2 to +9, and hydrophobic with an amphipathic structure.[6] Another property

Institutions where work was performed: Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, U.K.; Agency for Science, Technology and Research (A*STAR), Bioinformatics Institute, Singapore

important in targeting bacterial membranes is amino acid composition. Trp residues are frequently found in AMPs and multiple studies have highlighted their importance in interactions with biological membranes. Peptides containing only Arg and Trp residues can be highly antimicrobial.[7] Trp residues are critical for anchoring and insertion of peptides into the membrane[8,9] and their removal can have drastic effects on the antimicrobial activity of peptides.[10] Simulations have been used extensively to probe these interactions.[11] Trp is stabilized by hydrogen bond interactions with water molecules and headgroups at the interface.[12,13] However, the Trp residues can equally easily lie inside the membranes where their bulky sidechains can disrupt the packed lipid chains.[8] Similar behavior is also seen for Tyr and to some extent for Phe side chains.[14,15] It is common to see the insertion of Trp residues in efforts to design AMPs.[16,17]

AMPs can contain secondary structures of all kinds—helices, β-sheets, extended, and loop regions. Generally, AMPs can be divided in 4 structural groups: α, β, αβ, and non-αβ.[18] The most abundant structural group of AMPs are amphipathic α-helices, followed by αβ and all-β structures.[19,20] Aside from short, linear α-helical peptides, more complex all-α folds have also been found. These include helix hairpins and helical bundles, commonly found in class II bacteriocins such as the well-known food preservative nisin. AMPs with αβ structure often have disulfide bonds, such as those seen in plant defensins' cysteine-stabilized αβ (CSαβ) motif. All-β AMPs have structures comprised of multiple β-strands, for instance a simple β-hairpin stabilized by a circular backbone and disulfide bonds (as seen in θ-defensins of non-human primates) or the cysteine-stabilized triple-stranded β-sheet seen in human defensins. Unlike conventional antibiotics, which generally target metabolic enzymes, AMPs act mainly by membrane-targeting mechanisms and are selective due to the difference in charge of prokaryotic and eukaryotic cell membranes. Furthermore, AMPs have faster antimicrobial activity than conventional antibiotics.[21,22] Generally speaking, AMPs can be divided into two mechanistic classes: membrane disruptive and non-membrane disruptive (acting on intracellular targets). Disruption of the negatively charged prokaryotic membrane is the predominant mode of action of AMPs, with three main mechanisms proposed: the barrel stave, toroidal, and carpet model.[23,24]

AMPs have therapeutic potential as bioactive coatings for needles, catheters, implants, surgical tools, bandages, and even contact lenses. However, only a few have been approved for clinical use, and only for topical application, mainly due to their toxic properties.[25,26] The main difficulty in AMP drug development is our lack of understanding of modes of action.[27] The availability of structural information is crucial in facilitating AMP design efforts to predict, understand and implement knowledge-based enhancement of activities yet the pace of structural determination lags far behind AMP discovery: currently, there are over 2000 AMP sequences known, but only about 10% of them are structurally characterized. Researchers have used different methods in order to optimize antimicrobial activity on known protein scaffolds. Quantitative structure-activity relationship (QSAR), Regression models and Machine Learning approaches such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forests (RF) and Hidden Markov Models (HMM) are some of the approaches employed.[28,29] However, most of these studies are sequence-based, and design efforts based on the structural properties of more complex folds, such as studies on β-hairpins by Edwards et al.[30] or Yang et al.,[31] are less common. Notably, sequence information on its own is not sufficient to determine relevant properties of folds, such as the amphipathicity or dipole moment of the molecule.

Due to coevolution with pathogens, AMP sequences are exceptionally diverse.[32] AMP genes have been found to evolve rapidly in both vertebrates and invertebrates as a result of rapid gene duplication, diversification, and positive selection. This has been documented for mammal, bird, amphibian, and insect AMPs. Positive selection in AMPs seems to be highest immediately after gene duplication, although there could be a limit on observing high number of nonsynonymous substitutions in distantly related sequences.[33] It is known that AMP and immune genes evolve much faster than non-immune genes[34,35] with other work showing that AMPs can evolve 3 times faster than other proteins.[36] Due to this rapid evolution, reconstruction of the evolutionary history of AMPs can be a challenging task.[37] Moreover, this limits the possibilities and scope of homology modeling that can be performed for known sequences: not only is the number of available templates limited (see above) but evolutionary relationships between targets and templates are often hard to discern. In such cases, where no experimentally solved homologous structures exist, or exist but cannot be identified, models have to be constructed from scratch by performing ab initio modeling. Successful ab initio modeling of proteins without structurally characterized homologs with RMSD values around 2–5 Å has been reported for sequences shorter than 100–120 residues,[38–40] with the most recent CASP free modeling experiments showing this limitation to be at 150 residues.[41,42] AMPs are generally small in size which makes them particularly suitable for ab initio modeling. Furthermore, they often contain disulfide bonds which, if their connectivity can be predicted, provide valuable additional data to guide modeling. Most often, cysteines in extracellular proteins come in even numbers.[43] In an overview of disulfide-containing AMPs, Lehrer[44] discusses peptides with intermolecular as well as intramolecular disulfide bonds. AMPs with one cysteine are quite rare and have been found to form hetero- or homodimers. While redox status is known to have an effect on antimicrobial activity,[45,46] Lehrer's overview[44] does not give examples of AMPs containing two reduced cysteines. Interestingly, when its cysteines are reduced by the host, human antimicrobial peptide β-defensin 1 shows increased activity.[46]

Here, we carry out a large scale ab initio modeling of the structures of structurally uncharacterized AMPs with Rosetta[47] aiming to improve our understanding of the AMP structural universe. Although it is well-known that membrane-active AMPs can undergo structural changes when adopting a functional conformation at the membrane,[48] prediction of structures in aqueous solution can be expected to illuminate non-obvious evolutionary relationships and shed light on structural determinants of initial membrane interaction. We assembled a protocol to create a representative set of AMP sequences which have no predicted homologs of known 3D structure, and predicted disulfide bonds in order to facilitate their modeling. Following ab initio modeling, we tested their stability, compared their 3D structures against characterized AMPs and found fold matches as well as several unique folds.

2 | MATERIALS AND METHODS

2.1 | Sequence assembling and processing

Sequences longer than 20 and shorter than 120 amino acids were collected from UniProt[49] and APD2[19] on 17 March 2015. APD2 was chosen from the several AMP databases available since it is manually curated, and comparatively large and up-to-date. The UniProt release at the time was 2015_03. Sequence redundancy was reduced to a threshold of 45% using CD-HIT[50] and its global alignment option. HHpred[51] was used to detect sequences with structurally characterized homologs in the PDB70 database as at 6 September 2014. PDB70 is a version of PDB that is redundancy-reduced to 70% sequence identity.[52] Upon inspection of the results, three conditions were required to be satisfied before a given AMP would be considered to have a homolog: (1) HHpred fold match probability higher than 90%, (2) alignment coverage of query sequence higher than 40%, and (3) absence of any mismatch greater than 2-fold in length of query and hit. Sequences with >35% residues with a IUpred[53] score of 0.5 or above were considered to be intrinsically disordered AMPs. Since ab initio modeling is not suitable for these proteins, they were not considered further.

2.2 | Disulfide prediction, ab initio modeling, ab initio benchmarking and clustering

Disulfide bond predictions were made using Disulfind, Dianna, and Dinosolve.[54–56] Disulfind and Dianna were run at their respective servers while Dinosolve was run locally (and used the nr database[57] from 17 March 2015 and PSI-BLAST v.2.2.26[58]). Consensus predictions deriving from 2 independent methods were obtained: there were no cases where all three programs agreed on all disulfides. For AMPs with 2 cysteines, it was assumed that the disulfide bond exists.[44] For those AMPs with 3 or 4 cysteines, we ran Rosetta[47] modeling with all possible combinations. AMPs with 5 and more cysteines were run with any consensus and without disulfide constraints. It should be noted, however, that although intermolecular disulfide bonds in AMPs are considered to be rare,[59] peptides with odd number of cysteines, could form dimers or oligomerize, such as recently seen in rodent α-defensin-related AMPs.[60]

Ab initio modeling of 184 AMPs was performed with Rosetta software using the `fix_disulf` and `relax` flags. Use of the `nohoms` flag to exclude homologous fragments was unnecessary since modeling was only done for targets which HHpred bore no obvious homology with PDB entries. Where an AMP target contained a modified amino-acid, it was modeled using the natural, unmodified version. Where the predicted structure for such an AMP proved to be of interest, the likely consequences of the unmodeled modification were considered later. For each AMP, 1000 models were made. Models defined as successful by Rosetta (meaning they passed the filters that eliminate models with non-protein like features) were clustered into 10 groups using SPICKER v. 2.0[61] in order to identify the likely near-native structure for each AMP. Larger and more homogeneous clusters indicate more reliable fold candidates. We performed Rosetta benchmarking on AMPs of known structure (fold matches of our modeled AMPs, see Tables S1-S3 and Figures S1-S3) to detect a threshold value for cluster size to refer to when inspecting our models. For the benchmarking, the `nohoms` flag was used when running Rosetta in order to exclude fragments from the target and homologous structures. We chose to include modeled structures where the largest cluster size was at least 25% of the total number of successfully modeled structures. However, structures with lower percentages were scrutinized and considered if the centroids of three largest clusters had similar folds (Figures S4-S7).

2.3 | Fold matching and visual representation of matches

We assembled a database of experimentally solved AMP structures to compare models against. To ensure that we collected as many structures as possible, PDB codes were collected from both APD2 and UniProt in the following manner: first, an APD2 (March 2015) search was run with filter *Original Location: PDB*, which resulted in 229 PDB entries. Then, a UniProt (April 2015) search was run with keywords *keyword: "Antimicrobial [KW-0929]" annotation:(type:peptide) database:(type:pdb)*. The results of this run were additionally filtered: only one PDB structure was taken for each UniProt entry, prioritized such that: (1) an X-ray structure was chosen if possible, (2) the X-ray structure with the highest resolution and matching start and end positions, compared to the AMP, was used, (3) if an X-ray structure was not available, the first NMR structure to match the start and end residue was chosen, and (4) the first chain was chosen. Structures with non-matching start and end residues were omitted if the mismatch was greater than 5 amino acids.

For each modeled AMP, a structure similarity search was carried out with GESAMT[62] to compare the three largest clusters' centroid models against the local database of AMP experimental structures. The results were then filtered so that only those modeled structures meeting all of the following conditions were left: (1) Q-score (a measure of structural similarity ranging from 0 to 1) ≥ 0.3, (2) query (AMP model) sequence length < 1.5 × match (experimental structure) sequence length, that is, query can't be more than 50% larger than match, (3) match sequence length <1.5 × model sequence length, that is, match cannot be more than 50% larger than query, and (4) number of aligned residues ≥0.7 × query sequence length, that is, the alignment covers at least 70% of the query. Additionally, after filtering out results, fold matches were inspected manually: a model was considered to have a fold match if the number and type of secondary structure elements was similar. Ab initio models with tertiary structure matching at least one of the top 3 filtered fold matches were considered further.

In order to visualize the structural similarity, GESAMT was run in all-vs-all fashion on a set of AMPs from the PDB90 and matching modeled structures. The resulting similarity matrix was used as input for CLANS software[63] to cluster the structures. The CLANS software was used to visualize clusters of modeled structures of AMPs and the matching folds in the PDB,[52] 56 structures in total. 100 152 rounds of convex clustering using a value of 0.4 for standard deviation cutoff and requiring a minimum of 2 sequences per cluster initially clustered 49 out of 56 structures into eight clusters.

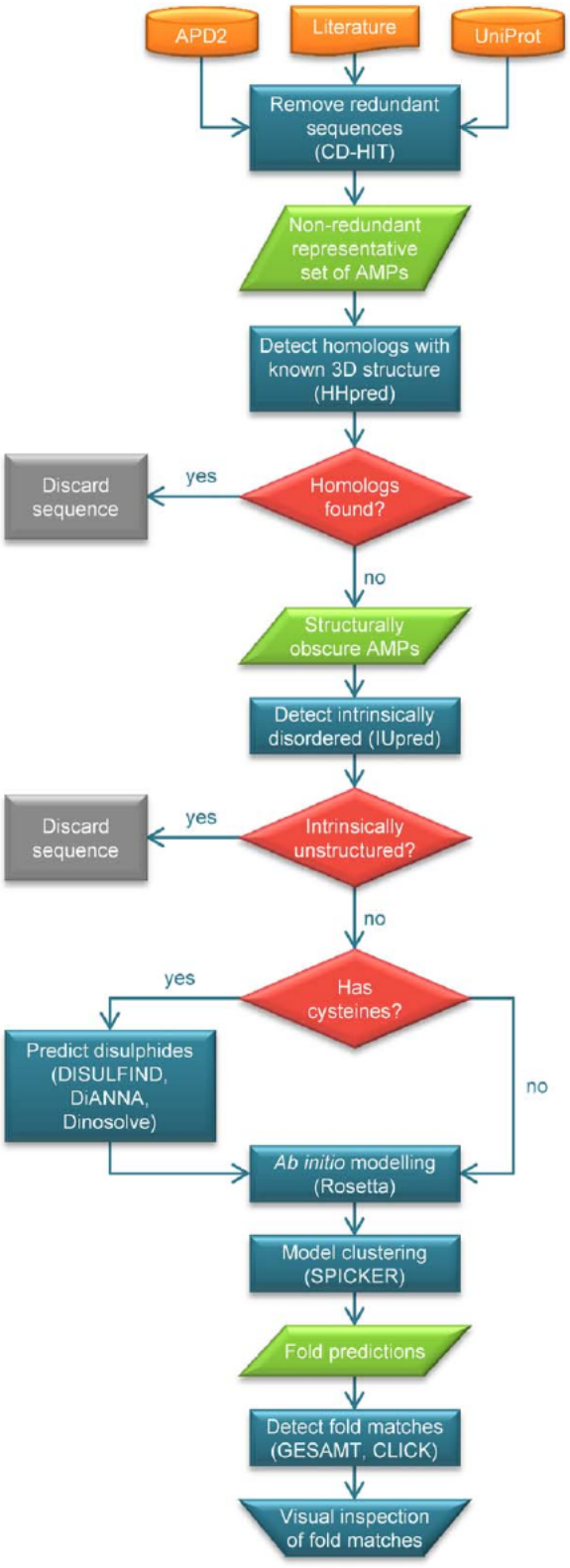

FIGURE 1 Bioinformatics pipeline used in this work. It leads from searches of sequence databases through representative set of structurally obscure AMPs to fold predictions and matches [Color figure can be viewed at wileyonlinelibrary.com]

Since GESAMT employs a topology-dependent algorithm we additionally used the topology-independent superposition method CLICK to search AMP models not matching by GESAMT against the same local database of experimentally determined AMP structures.[64] Again the results were filtered so that the query couldn't be more than 50% larger than the match, and the match could not be more than 50% larger than query. For models where no matches with AMPs were found, we ran additional CLICK database search on all protein chains from the PDB90 (not just AMP structures) and filtered results again in a similar manner, after which matches with Z-score values higher than 3 were taken forward. In cases where all of the Z-score values were lower than this threshold, we lowered this value to 2. Finally, all of the matches that were left after the filtering were visually inspected.

2.4 | Stability of peptides

Molecular Dynamics (MD) simulations were performed using the AMBER package and AMBER FF14SB force field.[65,66] Simulations with explicit solvent were performed using TIP3P water molecules with a 12 Å buffer between peptide atoms and the edge of a rectangular box. For each simulation, 10 000 steps of minimization were performed, with the first 5 000 using the steepest descent algorithm followed by 5 000 steps of conjugate gradient. The system was heated to 300 K in two steps; first heating from 0 to 100 K for 5 ps followed by heating from 100 to 300 K for another 100 ps, both using the Langevin thermostat. In the production step, we simulated the system at 300 K and 1 atm using the Berendsen barostat for 100 ns. Simulations with implicit solvent were run for 1 μs. All simulations were run in triplicate.

For the last 50 ns of each simulation, structural alignment was performed on Cα atoms of residues that formed regular secondary structure in the Rosetta model. RMSD clustering was carried out using MMTSB Tool Set based on those Cα atoms.[67] The structure closest to the centroid model was taken as a representative for each highly populated cluster.

3 | RESULTS AND DISCUSSION

In order to select and process AMPs, a workflow was implemented (Figure 1) to collect a non-redundant set of AMP sequences, eliminating those whose fold could be reliably inferred by homology detection and those predicted to be largely intrinsically disordered. Ab initio modeling of the resulting set, with or without predicted disulfide bonding as an additional constraint, was carried out using Rosetta. Clustering of the results determined candidate fold predictions, which were then compared to known AMP structures. Due to evolutionary constraints, protein folds can remain conserved even when there is an apparent lack of homology.[68] Similarity between our models and known AMP structures could therefore result from distant, unsuspected homology. Similar folds can also arise as a product of convergent evolution; the best example among AMPs are defensins, which are taxonomically widespread over insects, mammals and plants,[69,70] and are found to adopt a variety of folds such as β-sheets (triple-stranded β-sheet of Human Neutrophil α-defensin HNP-3 is an example), cyclic backbone

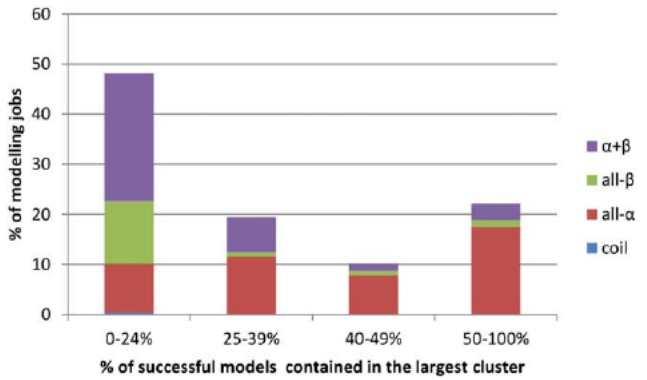

FIGURE 2 Distribution of modeling jobs per number of successful models contained in the largest cluster. Percentage of modelling jobs with largest cluster containing 0–24%, 25–39%, 40–49%, and 50–100% of the total number of successful models is shown in columns, while colors represent different fold classes [Color figure can be viewed at wileyonlinelibrary.com]

β-hairpins (θ-defensins), and cysteine-stabilized αβ (CSαβ) fold (commonly found in plants, fungi, and invertebrates). Indeed we found likely examples of both divergent and convergent evolution, as well as novel folds, strongly predicted, not hitherto associated with AMP activity.

After collecting AMPs from APD2[19] and UniProt,[49] elimination of identical sequences resulted in an initial count of 2131. CD-HIT[50] was used to further reduce this number and create a representative set with shared pairwise sequence identity of no more than 45%. 585 AMPs were then analyzed with HHpred[51] in order to discard AMPs with possible structural homologs, which left us with 235 peptides (see Methods for criteria). Furthermore, intrinsically disordered sequences were discarded and disulfide bridges were predicted in those peptides containing three or more cysteines. We proceeded with Rosetta ab initio modeling for 184 peptides, after which clustering of successful models was performed. Finally, a GESAMT[62] database structural similarity search was run on centroid models of three largest clusters for each AMP in order to obtain fold matches with experimental structures of AMPs from the PDB database.[52]

Psipred secondary structure prediction for 184 AMPs predicted 50.54% of peptides to have all-α structure, 33.7% to have mixed α and β, 15.22% to have all-β structure, and 0.54% to adopt a coil conformation.[71] Our set of 184 sequences contains some peptides with more than one pair of cysteines. For this subset, modeling was undertaken with different combinations of disulfides (see Methods), so that the overall number of modeling attempts grew from 184 to 216.

The reliability of our models was assessed based on the size of the largest cluster, and using the results of a benchmarking exercise in which structurally characterized AMPs were modeled ab initio (see Tables S1-S3 and Figures S1-S3). This revealed that, even without the use of disulfide contact information, the largest cluster usually predicted the fold correctly. Furthermore, the isolated failure (1CZ6) was distinguished by smaller largest cluster size. Where the largest cluster size exceeded 25% of the total models, that cluster always had the correct fold and so this was the major criterion used to judge the modeling results. Among 216 modeling runs, 48% of peptides had largest cluster containing less than 25% of the total number of successful models, 20% of peptides between 25 and 39% of the total, 10% of peptides between 40 and 49% and 22% had largest cluster containing between 50 and 100% of the total number of successful models (Figure 2). All-α structures were predominant in the most reliably modeled categories (Figure 2). This meant that, initially, 52% of structures were taken as reliable. Where the largest cluster was not larger than 25%, a comparison of the top, second and third largest cluster centroids was carried out. In two cases where these matched visually, the prediction was also considered of interest.

3.1 | Visual representation of fold matches

Modeled structures that matched AMP folds in the PDB were clustered using CLANS[63] in a semi-automated manner along with the corresponding fold matches (31 models and 25 fold matches making a total of 56 structures). Several modes of clustering were trialled but none proved capable of results that were fully in accord with expert assignment based on visual examination. For example, proteins with mixed αβ topologies sometimes allied more closely with β-hairpins, through a good fit of that portion, rather than with proteins with the same αβ overall topology but more poorly matching β-structure. Therefore, some manual (re)assignments were made to fine-tune an initial clustering of 8 groups for presentation purposes (Figure S8) and for discussion below. Three clusters contained β-hairpins and were combined and joined with a single modeled β-hairpin structure (Odorranain M1, AP01300)[72] left unassigned by the original clustering. One experimental structure (Mytilin B, PDB code 2EEM)[52] from this group was reassigned to a group of two combined clusters containing αββ folds joined by two modeled structures left unassigned, one αββ (Garvicin A, AP02402)[73] the other βαβ (Rattusin, AP02178).[74] Four helix hairpins that were left unassigned, namely experimental structures of Sublancin 168 (PDB code 2MIJ), Thurincin H (2LBZ), Thuricin CD (2L9X), and EcAMP1 (2L2R),[52] were joined with two clusters containing a continuum of v-shaped, helix-kink-helix and helix hairpin structures. As a result, the structures were finally clustered into the four groups shown in Figure S8: (1) v-shaped, helix-kink-helix and helix hairpins which form a continuum of structures shown in red, (2) αββ folds and βαβ folds shown in blue, (3) β-hairpins shown in magenta, and (4) helix bundles shown in green. These groups contain 32, 10, 11, and 3 structures, respectively. In Figure S8, structures with greater similarity (higher Q-scores) are positioned at shorter separations. We next discuss the results in each fold family.

3.2 | Fold matches

All the fold matches shown here had a Q-score ≥0.3 and were additionally manually screened so that matches were considered only when the AMP model and the experimentally determined matching structure were not too dissimilar in length and aligned over a majority of the model structure (see Methods).

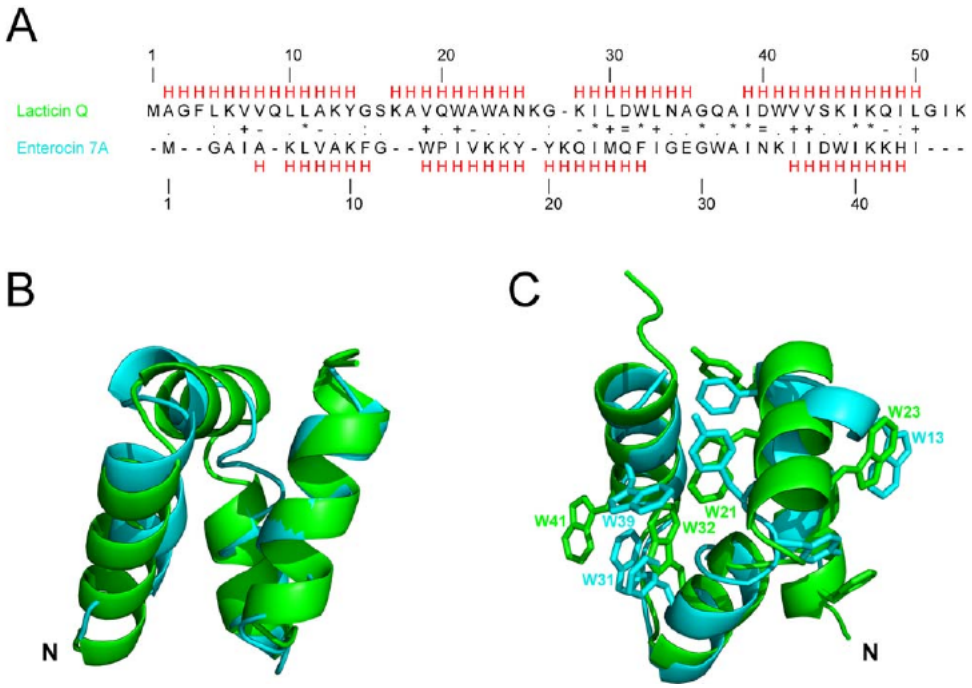

FIGURE 3 Alignments of modeled Lacticin Q and experimental structure of Enterocin 7A. A, GESAMT structure-based sequence alignment. H above and below indicates helical residues. Symbols indicate identity (*) through decreasing levels of similarity (+, =, –, :) to dissimilarity (.). B, The modeled structure of Lacticin Q is shown in green, aligned with the experimental structure of Enterocin 7A (PDB code 2M5Z) shown in cyan. N termini indicated with letter N. C, Rotated view with aromatic residues shown in stick representation. Trp residues are labeled. Figure created in PyMol

3.2.1 | Helix bundles

A single confidently modeled AMP, the bacteriocin Lacticin Q,[75,76] was predicted to fold as a helix bundle. Its ab initio modeling resulted in a large cluster comprising more than 30% of the successfully modeled structures, and suggested the presence of four helices, spanning residues 2–14, 16–25, 28–35, and 38–49, respectively. PSIPRED structure prediction for Lacticin Q identified two helices (residues 4–13 and 38–49) with high scores; and moderately high helical propensity for the other two helices predicted by Rosetta. GESAMT resulted in only two fold matches for our modeled structure of Lacticin Q with our local AMP structure database: Enterocins 7A and 7B,[77] both also bacteriocins. The fold matches had Q-scores of 0.589 and 0.558, RMSD values of 1.8 and 1.9 Å on Cα atoms, and sequence identities of 16% and 17% for Enterocins 7A and 7B, respectively. Similarly, the Dali server[78] gave *Z* scores of 4.4 and 4.2, a slightly higher RMSD of 1.9 and 2.3 Å on Cα atoms and sequence identities of 16% and 19% for Enterocins 7A and 7B, respectively.

Bacteriocins are antimicrobial peptides synthesized by the ribosomes of a variety of bacteria (both Gram-positive and Gram-negative).[79] Cotter et al.[80] categorized bacteriocins into three classes: class I, also known as lantibiotics, are post-translationally modified peptides containing amino acids called lanthionines; class II are a heterogeneous group of small heat-stable non-lanthionine containing peptides which may have disulfide/thioether bonds; and class III are large, heat-labile, lytic proteins called bacteriolysins.[18] Lacticin Q belongs to class II and shows selectivity for Gram-positive bacteria at the strain level suggesting that membrane lipid composition might not be the only determinant of its antimicrobial activity. It is also known that the peptide causes accumulation of hydroxyl radicals.[81]

It has been suggested that circular bacteriocins share a common overall structural motif of a saposin fold, that is, four helices surrounding a hydrophobic core, regardless of low shared sequence identity[82] and our results are consistent with this (Figure 3). Tryptophan residues are known to be involved in protein folding as well as to have a tendency for burial at the bilayer interface.[83,84] Another common feature of circular and leaderless bacteriocins is the presence of solvent-exposed tyrosine or tryptophan residues that are likely to facilitate membrane penetration.[85] A comparison of the modeled structure of Lacticin Q with the NMR structure of Enterocin 7A showed that the aromatic side chains located at different positions in the sequences were seen at structurally equivalent locations in the two structures. Out of three surface tryptophans in Enterocin 7A, two, namely W13 and W31, are in close proximity and have the same orientation as seen for W23 and W32, respectively, in Lacticin Q. Strikingly, these Trp pairs are found in corresponding positions in 3D space despite not aligning in the structure-based sequence alignment (Figure 3). Although Lohans et al.[77] had tentatively proposed a relationship between Lacticin Q and Enterocins 7A and 7B, based on weak sequence similarity in the N- and C-terminal helical regions, the striking structural similarity and structural correspondence between likely functional aromatic residues are strongly indicative of homology over the whole protein despite the low overall degree of sequence identity.

Lacticin Q shows selective antimicrobial activity against various Gram-positive bacteria. It is frequently compared to another class II

TABLE 1 Fold match results for β-hairpin modeled AMPs

Modeled AMP and the corresponding APD2 IDs	C1 size %	Fold matches	Fold matches' corresponding PDB and APD2/UniProt IDs	Identity %
Lebocin Peptide 1A (AP00030)	50	RTD-1 BTD-2 Retrocyclin-2	2LYF_A, AP00445 2LYE_A, AP00156 2LZI_A, AP01208	11 6 0
Odorranain M1 (AP01300)	29	Androctonin	1CZ6_A, AP00153	5
Silkworm 001 (AP01974)[a]	38	BTD-2pLR Tachyplesin I	2LYE_A, AP00156 2M3N_A, Q90WP7 1WO1_A, AP00214	6 0 13

Identity percentages were obtained through GESAMT. C1 size %—size of the largest cluster compared to the overall number of models.
[a]Data unpublished.

bacteriocin, nisin, for its nanomolar range antimicrobial activity, pore size and ATP efflux.[75,76,81,86] However, compared to nisin, lacticin Q is a leaderless bacteriocin—the peptide is synthesized without the N-terminal leader sequence that is otherwise removed when exporting from cells.[76,77] Lacticin Q, Enterocin 7A and 7B are unmodified leaderless N-formylated bacteriocins that adopt helical conformations in solution. They all have an overall net charge of +6 which induces binding to negatively charged lipids although a comparison of the electrostatic properties of the Lacticin Q model and Enterocin crystal structures shows no strong similarity (not shown). The authors suggest a huge toroidal pore (HTP) model as the antimicrobial mechanism of Lacticin Q, followed by lipid flip-flop and translocation of the peptide to the inner membrane leaflet.[76,77]

While this manuscript was in preparation, an NMR study of the Lacticin Q structure was published by Acedo et al.[87] The RMSD of the Cα atoms between our model and the NMR structure is 2.69 Å, while RMSD values of the Cα atoms between structures obtained with MD and the NMR structure range from 2.97 to 4.57 Å (Figure S9). The NMR model experimentally validates our model.

3.2.2 | β-hairpins

Three confidently modeled AMPs, Lebocin Peptide 1A,[88] Odorranain M1[72] and Silkworm 001 (unpublished; APD identifier 01974), were predicted to fold as β-hairpins (Table 1). The three models resulted in large clusters of 50%, 29%, and 38% of the successfully modeled structures, respectively, and their reliability was further tested by running 100 ns simulations in explicit solvent and performing clustering as described in Methods section. Structure representatives of highly populated clusters superimposed in Figure S10 show the stability of modeled folds in solvent.

With the reliability of our models confirmed, we compared their properties with those of the fold matches (Table 1), and other AMPs with experimentally determined hairpin structures. The simplest β-sheet structure is the β-hairpin fold, a structure comprised of two anti-parallel β-strands and a β-turn. Fold matches in this group have specific structural features stabilizing the structures such as disulfide bonds or aromatic residues, the latter being necessary for aromatic stacking or cation-π interactions. Examples of β-hairpins include both θ-defensins and structures with linear backbones, such as protegrins, thanatin, gomesin, tachyplesins, polyphemusins, and arenicins.[89]

Among the PDB fold matches to these 3 peptides, we find three θ-defensins: BTD-2, RTD-1, and HTD2 (Retrocyclin-2).[90] These are backbone cyclic β-hairpin AMPs containing three parallel disulfide bonds also known as the cystine ladder motif. A study on 18 residue long BTD-2 θ-defensin analogues by Conibear et al.[90] showed that a cyclic backbone appears to be essential for membrane activity resulting in antibacterial effects, as was also reported earlier by Tang et al.[91] for RTD-1 θ-defensin. However, the disulfide bonds have been shown to be essential for stability of these AMPs, as well as for resistance to the action of proteolytic enzymes.[90,92] Disulfides can be either essential and dispensable for the activity of β-hairpins: Protegrin-1 was as active in linear form as in cyclic,[93,94] while reducing disulfides in Arenicin-1 led to decreased activity.[95] Interestingly, it has been shown by Ma et al.[96] that disulfides were not only dispensable for Thanatin activity and toxicity, but that the secondary structure was maintained in their absence as well.

The other fold matches are β-hairpins with linear backbones: Androctonin (PDB code 1CZ6), Tachyplesin I (1WO1) and pLR (2M3N).[52] Androctonin and Tachyplesin I show sequence similarities and both contain two disulfide bonds with the same connectivity: 1–4, 2–3. Even though they are both active against Gram-positive and Gram-negative bacteria, Hetru et al.[97] suggest that a proper comparison of their modes of action cannot be made due to the different number of residues separating the cysteines. While Matsuzaki et al.[98] hypothesized that Tachyplesin I forms anion-selective pores and translocates, Androctonin seems to act via a detergent-like mechanism.[97] pLR β-hairpin fold match is a peptide with a linear backbone which has one disulfide and flexible terminal regions. It is likely that, as with other peptides, pLR forms oligomers in solution whose large size prevents effective traversing of the membrane in Gram-negatives.[99]

A recent review by Panteleev et al.[89] classified β-hairpins into subgroups by the number of disulfides but did not report any AMPs without this feature. Interestingly, none of our modeled structures contain intramolecular disulfides (Figure 4)—they are the first reported β-hairpins lacking disulfide bonds. Although a Conibear et al. study[90] on BTD-2 cyclic hairpin analogues speculates that the absence of disulfide bonds places a limit on the lengths of β-hairpin structures, our models suggest otherwise. Tachyplesin analogs including replacements

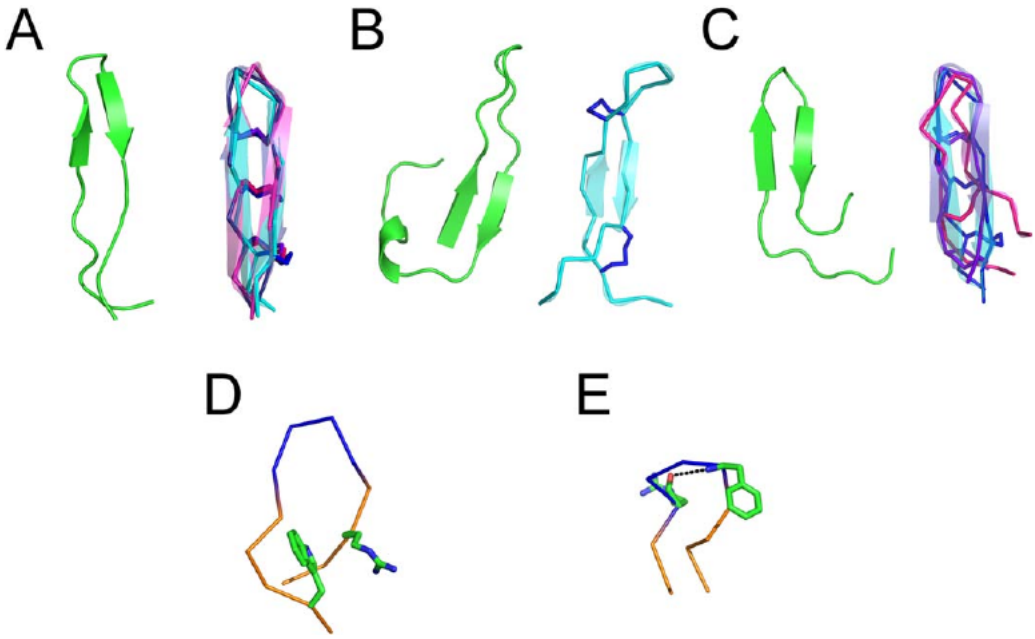

FIGURE 4 Rosetta models of AMPs aligned with their fold matches. A, Lebocin Peptide 1A (green) vs. RTD-1 (PDB code 2LYF, shown in cyan), BTD-2 (2LYE, magenta), and Retrocyclin 2 (2LZI, purple), B, Odorranain M1 (green) vs. Androctonin (1CZ6, cyan), and C, Silkworm 001 vs. BTD-2 (2LYE, cyan), pLR (2M3N, magenta) and Tachyplesin I (1WO1, purple). Disulfide bonds are shown as sticks in blue, pink, and purple, respectively. Alignments were made in GESAMT. β-hairpin-like structures shown in ribbon representation: D, CDT-LPS peptide (2LM8) featuring cation-π interaction between Trp2 and Arg11 (shown as sticks) and E, R11 peptide bound to LPS (2MQ4) with putative hydrogen bond within the β-turn. β-turns are shown in blue [Color figure can be viewed at wileyonlinelibrary.com]

of Cys residues with Tyr, Phe, or Ala maintain the fold as a result of aromatic stacking,[100] but our peptides are not spatially constrained by aromatic interactions either.

In order to assess how unusual the absence of disulfides in an AMP β-hairpin is, the PDB was mined using the mmCIF Keyword Search (Classification) "antimicrobial" and the hits were scanned visually for structural similarity. Shown in Figures S10D and S10E are 2 "β hairpin-like" AMPs that were found: entries 2LM8 (Cysteine Deleted analog of β-hairpin AMP Tachyplesin I in LPS, CDT-LPS)[101] and 2MQ4 (R11 peptide bound to LPS).[102] Both AMPs are of synthetic origin and very short, with 13 and 11 residues, respectively. CDT-LPS retains a β-hairpin-like structure due to cation-π interactions between residues W2 and R11, but such interactions are not present in our modeled AMPs. RR11 is an N-terminal truncated variant of Cys deleted protegrin-1, CDP-1. Both CDT-LPS and R11 are largely unstructured when free in aqueous solutions. NMR experiments of CDP-1 bound to LPS showed aromatic interactions. While the structure of RR11 bound to LPS somewhat resembles our β-hairpins, it lacks the characteristic hydrogen bonds to form the overall β-hairpin structure and is almost half the size of our modeled β-hairpins.

Since there are no intramolecular disulfides, cation-π or aromatic interactions to keep these structures stable, we looked for hydrogen bonds in the set of trajectory snapshots on which RMSD clustering was performed. Our Rosetta model of Lebocin Peptide 1A[88] showed two β-strands extending from residues 7 to 10 and 14 to 17. This feature was commonly present during the trajectory. For example, two hydrogen bonds between residues Phe8 and Thr16 were present ~81% and ~63% of the time. Similarly, the Odorranain M1 hairpin[72] has residues 2–4 and 14–16 in the β-strand conformation. Here, two hydrogen bonds between Ala3 and Arg15 were present ~75% and ~53% of the time. Finally, the Silkworm 001 AMP (unpublished; APD identifier 01974) Rosetta model showed two β-strands extending from residues 8 to 11 and from 14 to 17. Two hydrogen bonds between Ile9 and Tyr16 were present ~95% and ~86% of the time, and two hydrogen bonds between Ala11 and Tyr14 were present ~76% and ~74% of the time. Together, these findings suggest that LP1A and Odorranain-M1 are the first β-hairpin folds lacking intramolecular disulfide bonds, cation-π or aromatic interactions reported so far, stable only due to hydrogen bonding.

Although the β-fold and charge both appear to be key determinants of antimicrobial activity, their relative importance is not clear. For example, Mohanram and Bhattacharjya[102] associate the loss of charge associated with truncating the RGGR N-terminus in RR11 with the attenuation of activity rather than the loss of β-strands. However, Mani et al.[103] find for Protegrin-1 mutants that the β-hairpin fold is more important for activity than the cationicity.

Experimental structures of CDT-LPS and R11 peptides both have significant positive charge spatially arranged on one face of the molecule (Figures S11A and S11B). However, this feature was not seen in our modeled peptides. Silkworm 001 has a single negative charge at its C-terminus tail, and the overall charge of the molecule is zero. Odorranain M1 and Lebocin 1A are positively charged but, in contrast to the crystal structures, cluster positively charged residues (4 of 5 or 3 of 5, respectively) at one end, toward the bottom in the orientation shown in Figures S11C and S11D. However, it must be noted that these regions are very flexible and are therefore less reliably predicted.

TABLE 2 Fold match results for αββ and βαβ AMPs

Modeled AMP and the corresponding APD2 or UniProt IDs	C1 size %	Fold matches	Fold matches' corresponding PDB and APD2/UniProt IDs	Identity %
Human Histatin 2 (AP00799)	8[a]	Mytilin B Termicin AgaDef	2EEM_A, AP00333 1MM0_A, AP00403 2NY8_A, AP01363	8 4 0
Rattusin (AP02178)[b]	25	Palicourein Charybdotoxin Cycloviolacin	1R1F_A, AP01034 2CRD_A, AP00437 1NBJ_A, AP01035	0 11 14
Garvicin A (AP02402)	17	Mytilin B Termicin Eurocin	2EEM_A, AP00333 1MM0_A, AP00403 2LT8_A, AP02119	6 0 9

Identity percentages were obtained through GESAMT. C1 size %—size of the largest cluster compared to the overall number of models.
[a]C3 size %.
[b]Modeled structure matching an experimental structure with different topology.

Finally, Panteleev et al.[89] report that among several insect AMPs, only 1 was found to adopt a β-hairpin conformation. This was Thanatin, derived from the spined soldier bug *Podisus maculiventris*. Here, we predict that both insect peptides, Lebocin Peptide 1A (LP1A) and Silkworm 001, are also β-hairpins.

3.2.3 | αββ and βαβ folds

This group contains three confidently modeled AMPs. Human Histatin 2[104] and Garvicin A[73] were predicted to have αββ folds, while for Rattusin,[74] a βαβ fold was predicted. The first largest cluster for Human Histatin 2 peptide contained 35.4% of the total number of models but fold matches were found for the centroid structures of the second and the third largest clusters only. Nevertheless, the centroid structure of the largest cluster is highly similar to that of the second, and since the third cluster centroid has particularly pronounced secondary structure elements, this model was taken forward as a representative structure (Figure S4). Garvicin A modeling resulted in a small top cluster relative to the number of successfully modeled structures, and Rattusin fold was on the lower limit of 25% (Table 2, see Methods). Upon inspection of centroids of the remaining two largest clusters for Garvicin A and Rattusin, we found that these models displayed similar structures, confirming their reliability. In addition, we tested the stability of our modeled Rattusin fold by running 100 ns Molecular Dynamics simulations in explicit solvent and performing clustering as described in Methods section. Structure representatives of highly populated clusters superimposed are shown in Figure S12.

Human Histatin 2, similar to its fold matches Termicin (PDB code 1MM0) and AgaDef (2NY8) (Figure 5A),[52] shows antifungal activity. However, unlike the modeled peptide, its fold matches have three disulfide bonds: starting from the N-terminus, the first one is formed between the helix and first β-strand Cys residues, and the second and the third disulfide bonds are formed between the helix and second β-strand Cys residues.

The second modeled structure in this group is Rattusin, an α-defensin-related peptide with 5 cysteines. The peptide was modeled without disulfide constraints as there was no consensus on connectivity (see Methods). Here, we predict a βαβ fold which matches the αββ folds such as found in Palicourein (1R1F) and Cycloviolacin (1NBJ), AMPs with circular backbones, as well as Charybdotoxin (2CRD), a gamma-motif peptide targeting K+ channels (Figure 5B).

Disulfide bonds formed by Rattusin have been a subject of discussion. Whilst the newest NMR study by Min et al.[60] shows refolded Rattusin forming homodimers, each protomer being a β-hairpin, stabilized by five intermolecular disulfide bonds, Patil et al.[74] suggested that an odd number of cysteines may result in a fraction of Rattusin peptide forming dimers or multimers, which can

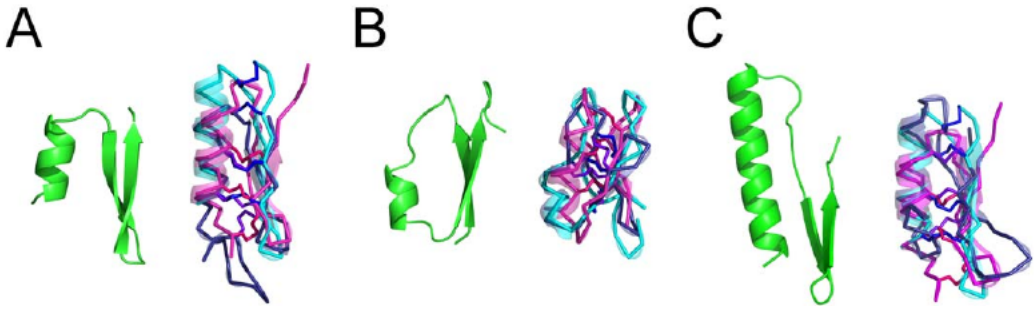

FIGURE 5 Rosetta models aligned with their fold matches. A, Human Histatin 2 peptide (green) aligned with Mytilin B (PDB code 2EEM, shown in cyan), Termicin (1MM0, magenta) and AgaDef peptide (2NY8, purple), B, Rattusin (green) aligned with Palicourein (PDB code 1R1F, shown in cyan), Charybdotoxin (2CRD, magenta), and Cycloviolacin (1NBJ, purple) peptides, and C, Garvicin A (green) aligned with Mytilin B (2EEM, cyan), Termicin (1MM0, magenta), and Eurocin (2LT8, purple) peptide. Disulfide bonds are shown as sticks in blue, pink, and purple, respectively. Alignments were made in GESAMT [Color figure can be viewed at wileyonlinelibrary.com]

in turn enhance its antibacterial effects. However, the peptide remains active even with the disulfides reduced, justifying a prediction of monomeric structure as performed here. It is well known that disulfide reduction can have a wide spectrum of effects from enhancement to total loss of activity (as seen in α and β-defensins, and θ-defensins, respectively), alteration of selectivity (as seen in cryptidin-related sequences) and so forth.[46,74,91,105]

Class II bacteriocins are a heterogeneous group of small heat-stable non-lanthionine containing peptides. One of the four modeled αββ folds with positive fold matches in the PDB was Garvicin A[73] isolated from *Lactococcus garvieae*, a class II bacteriocin. Although modeling produced a comparatively small largest cluster, the largest four clusters contained the same overall fold and exhibited pairwise RMSD on Cα atoms of 2.5 Å at most (Figure S5). These considerations suggest a reliably predicted fold. Garvicin A shares a similar fold with several cysteine-rich proteins and defensins, namely Mytilin B (PDB code 2EEM), Termicin (1MM0), Eurocin (2LT8), DEF-AAA (2NY8), Micasin (2LR5), and MGD-1 (1FJN) (Figure 5C).

Compared to Garvicin A, which is active only against *L. garvieae* strains, its fold matches are active against a wide spectrum of Gram-positive, Gram-negative bacteria and fungi. Interestingly, while Garvicin A has no cysteine residues, its matches are cysteine-rich peptides with 3 or 4 disulfide bonds stabilizing the structures. We mined the PDB with mmCIF Keyword Search (Classification) "antimicrobial" to see whether there was an αββ fold entry that didn't contain disulfides. Out of 339 structures, there was only one αββ AMP without disulfide bonds, namely Subtilisin A (1PXQ). This structure has a cyclized peptide backbone (with an amide between the N- and C-termini) and D-amino acids which have unusual sulfur to Cα cross-links. This suggests that Garvicin A and Human Histatin 2 are the first linear αββ fold AMPs lacking disulfide bonds.

3.2.4 | Helix-break-helix and helix-kink-helix continuum

By far the largest group of folds obtained was a continuum of v-shaped, helix-turn-helix and helix hairpins (Figure 6 and Table 3). Attempts to consistently subdivide these into separate helix-break-helix (containing helix-hairpins and v-shaped structures) and helix-kink-helix groups (Figure 6), were unsuccessful.

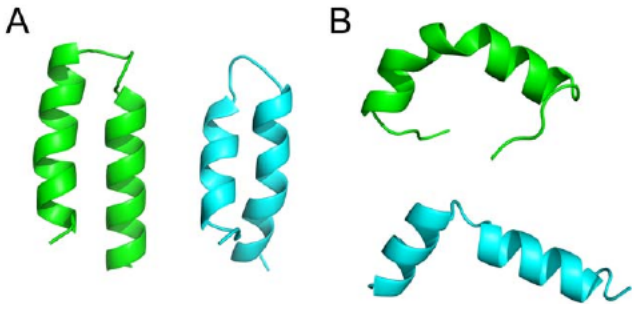

FIGURE 6 Simple helical folds. A, Helix-hairpins Ceratotoxin D and its fold match Thurincin H (PDB code 2LBZ), B, Helix-kink-helix Hymenochirin 2B and its v-shaped helix-break-helix fold match Latarcin 2a (PDB code 2G9P). Modeled structures are shown in green and the corresponding fold matches are shown in cyan [Color figure can be viewed at wileyonlinelibrary.com]

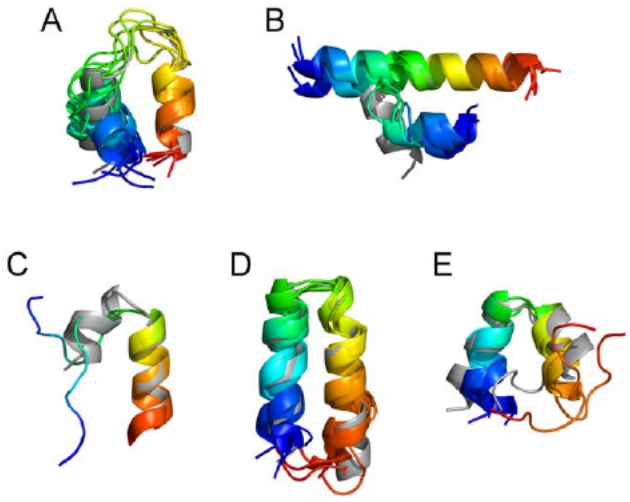

FIGURE 7 Stability of helical folds. A, Ponericin Q42, B, Latarcin 4a, C, Kassinatuerin 1, D, Ceratotoxin D, and E, CPF-B1 peptide. Structure representatives of the last 50 ns of triplicate experiments are shown superimposed and color coded from N (blue) to C terminus (red). Rosetta model is shown in grey. Structures were aligned with Theseus [Color figure can be viewed at wileyonlinelibrary.com]

For all of the structures in this group, the largest cluster contained high percentages of the total number of models, except for Andropin and Hymenochirin 2B. For these two peptides, we inspected the centroid structures of the remaining two largest clusters and found them to be similar to the largest cluster's centroid, which gave us confidence that all folds are modeled correctly. The stability of selected folds was tested by running 100 ns simulations in explicit solvent and performing clustering as described in Methods. Structure representatives of highly populated clusters superimposed are shown in Figure 7.

All of the peptides were modeled as monomers, including Cynthaurin peptide, which contains a single cysteine residue.[106] Cynthaurin is believed to predominantly form homodimers: however, both monomer and dimer are active against bacteria, whilst the monomer is non-hemolytic. Cynthaurin, along with Ponericin Q42[107], Ceratotoxin D,[108] and CPF-B1 peptides,[109] shown in Figures 7A, 7D, and 7E, respectively, has a helix-hairpin fold. Although the orientation of helices fluctuated somewhat, they were stable throughout the simulation (Figure 7). Latarcin 4a[110] (Figure 7B) showed 2 stable conformations, a single α-helix and a helix-break-helix fold. Interestingly, the short N-terminus helix of Kassinatuerin 1 peptide[111] (Figure 7C) rapidly unfolded in solution, while the C-terminal helix remained stable.

Helix-hairpins are formed by two antiparallel α-helices connected by a loop (Figure 6A). The helices interact through hydrophobic side-chain interactions at the interface. V-shaped structures can be defined as either: (1) two non-parallel α-helices whose angles intersect at angles from around 45° to 120° connected with a loop region (Figure 6B, bottom panel) or (2) a helical structure extending throughout the peptide but containing a kink. Usually, cationic peptides longer than ~20 amino acids contain a flexible hinge in the middle part promoted by helix-breaking residues such as glycine and/or proline.[112]

TABLE 3 Fold match results for helix-break-helix and helix-kink-helix continuum of structures

Modeled AMP and the corresponding APD2 or UniProt IDs	C1 size %	Fold matches	Fold matches' corresponding PDB and APD2/UniProt IDs	Identity %
Citropin 2.1.3. (AP00639)	46	EcAMP1 Thuricin CD Sublancin 168 Maximin 4	2L2R_A, AP01760 2L9X_A, AP01570 2MIJ_A, AP01606 2MHW_A, AP00061	12 8 4 5
Grammistin Gs A (P69845)	52	Maximin 4 Ltc2a	2MHW_A, AP00061 2G9P_A, AP01011	12 4
Andropin (Q8WSV4)	17	EcAMP1 Sublancin 168 Thuricin CDhIAPP	2L2R_A, AP01760 2MIJ_A, AP01606 2L9X_A, AP01570 2L86_A, AP02196	0 4 4 3
Casocidin I (P02663)	27	Thurincin H EcAMP1	2LBZ_A, AP02394 2L2R_A, AP01760	7 0
Dermaseptin 8 (P84928)	65	Thuricin CD EcAMP1	2L9X_A, AP01570 2L2R_A, AP01760	15 0
Ctri10033 (P0DME4)	69	Ltc2a Maximin 4 Fowlicidin 1	2G9P_A, AP01011 2MHW_A, AP00061 2AMN_A, AP00557	0 10 13
Ponericin Q42 (AP02435)	29	Thuricin CD EcAMP1 Maximin 4	2L9X_A, AP01570 2L2R_A, AP01760 2MHW_A, AP00061	4 4 15
Latarcin 4a (AP01014)	56	Ltc2a Fowlicidin 1 Maximin 4	2G9P_A, AP01011 2AMN_A, AP00557 2MHW_A, AP00061	22 8 6
RV 23 (AP01264)	78	Ltc2a Maximin 4	2G9P_A, AP01011 2MHW_A, AP00061	20 11
Lycotoxin I (AP00516)	91	EcAMP1 Ltc2a Maximin 4	2L2R_A, AP01760 2G9P_A, AP01011 2MHW_A, AP00061	4 20 14
Kassinatuerin 1 (AP00556)	64	Maximin 4 Thurincin H	2MHW_A, AP00061 2LBZ_A, AP02394	6 0
Ceratotoxin D (AP00417)	65	Thurincin H Sublancin 168 Thuricin CD EcAMP1	2LBZ_A, AP02394 2MIJ_A, AP01606 2L9X_A, AP01570 2L2R_A, AP01760	0 0 8 3
Pilosulin 5 monomer (AP00893)	88	Thurincin H Sublancin 168 EcAMP1hIAPP	2LBZ_A, AP02394 2MIJ_A, AP01606 2L2R_A, AP01760 2L86_A, AP02196	4 4 3 3
Hymenochirin 2B (AP01965)	24	Ltc2a Fowlicidin 1hIAPP Pardaxin P4	2G9P_A, AP01011 2AMN_A, AP00557 2L86_A, AP02196 1XC0, AP00644	33 19 8 27
Clavaspirin (AP00502)	65	Fowlicidin 1 Maximin 4 Ltc2a	2AMN_A, AP00557 2MHW_A, AP00061 2G9P_A, AP01011	20 15 18
Cryptonin (AP00722)	87	Maximin 4 Ltc2a Fowlicidin 1	2MHW_A, AP00061 2G9P_A, AP01011 2AMN_A, AP00557	17 25 10
Ocellatin 4 (AP00894)	98	Maximin 4 Thuricin CD Ltc2a Fowlicidin 1	2MHW_A, AP00061 2L9X_A, AP01570 2G9P_A, AP01011 2AMN_A, AP00557	15 11 6 20
Grammistin Pp 3 (P69847)	55	Ltc2a Maximin 4 Fowlicidin 1 Pardaxin P4	2G9P_A, AP01011 2MHW_A, AP00061 2AMN_A, AP00557 1XC0, AP00644	14 13 21 9

(Continues)

TABLE 3 (Continued)

Modeled AMP and the corresponding APD2 or UniProt IDs	C1 size %	Fold matches	Fold matches' corresponding PDB and APD2/UniProt IDs	Identity %
CPF B1 (AP01622)	65	Maximin 4 Thuricin CD Fowlicidin 1hIAPP	2MHW_A, AP00061 2L9X_A, AP01570 2AMN_A, AP00557 2L86_A, AP02196	23 4 10 4
Cynthaurin monomer (AP00510)	70	EcAMP1	2L2R_A, AP01760	4
Dinoponeratoxin Dq 3104 (C0HJH7)	57	Ltc2a Thuricin CDhIAPP	2G9P_A, AP01011 2L9X_A, AP01570 2L86_A, AP02196	10 11 9
Oreoch 3 (AP2140)	42	Maximin 4 EcAMP1	2MHW_A, AP00061 2L2R_A, AP01760	4 12
Ranatuerin 4 (AP00117)	61	Maximin 4 Ltc2a Fowlicidin 1	2MHW_A, AP00061 2G9P_A, AP01011 2AMN_A, AP00557	13 11 15
U4N938[a]	60	Thuricin CD Sublancin 168 EcAMP1	2L9X_A, AP01570 2MIJ_A, AP01606 2L2R_A, AP01760	0 21 19

Identity percentages were obtained through GESAMT. C1 size %—size of the largest cluster compared to the overall number of models.
[a]Disulfide connectivity known.

It has been speculated that v-shaped helix-break-helix structures with strongly amphiphilic α-helix at the N-terminus only are likely to be functional through the carpet mechanism, while structures with N- and C- termini that are both strongly amphiphilic are more likely to act via the pore-forming mechanism. For the first set, helix-break-helix AMPs with a hydrophobic gradient spanning from N- to C-terminus, it has been suggested that the amphipathic N-terminal helix is responsible for interaction with the membrane, while the C terminus, because of its lack of amphipathicity, lies on the membrane and only has a minimal interaction with it.[112] For this particular type of structures, Dubovskii et al.[112] proposed molecular hydrophobic potential (MHP) plots to be effective in sorting structures by mechanism. However, due to a lack of experimental data on structure and mechanisms of action of helix-kink-helix peptides, this method has not been particularly useful in gaining a clear picture for our modeled peptides and, for now, the structure-function relationship for these AMPs remains unclear.

3.3 | Potential novel AMP folds

The results discussed hitherto relate to fold matches between the AMPs we modeled and others of already determined structures. This revealed cases of likely cryptic homology, such as between Lacticin Q[87] and Enterocins 7A and 7B,[77] as well examples of recurrent simple folds that are more likely to be examples of structural analogy: a β-hairpin Lebocin Peptide 1A from a moth *Manduca sexta*[88] and θ-defensins BTD-2, RTD-2, and Retrocyclin-2 from baboon[90] would be likely examples. However, it is unlikely that currently determined AMP structures sample all naturally existing AMP fold space. We therefore examined results that appeared reliable, that is, with a large cluster of similar models and which represented well-packed protein structures, including considering whether they resembled any known folds (not just AMPs) in a topology-independent fashion that could be seen using the CLICK structure database search algorithm.[64] Here they are dealt with according to broad fold class and their specificities compared to those of AMPs with similar folds.

3.3.1 | All-β folds

A total of three all-β folds modeled (Figure 8) were scanned for fold matches using CLICK database search. The first, Cypemycin[113] (Figure 8A) is a bacteriocin active against Gram-positive bacteria. Although it has no lanthionine bridges present, it has some of the structural features of lantibiotics such as dehydrated amino acids, two L-allo-isoleucines, and a modified C-terminal D-cysteine that forms a ring structure with an L-cysteine. Only the disulfide bond constraint, set for the two cysteine residues, was included in our model. Since the two modeled isoleucines (which were modeled as the closest available representation of L-allo-isoleucine) are solvent exposed, and the D-cysteine is the

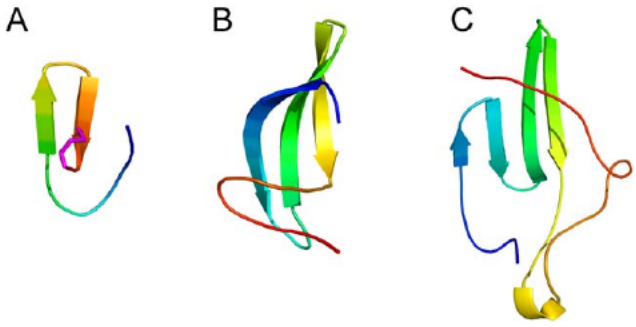

FIGURE 8 Plausible all-β folds. A, Cypemycin (AP01632, C1% 49.0, disulfide bond formed by Cys19 and Cys22), B, Scolopendin 1 (AP02453, C1% 24.2), C, Microcin M (AP01230, C1% 15.5). Structures were color-coded from N-terminus (blue) to C-terminus (red). Disulfide bonds are shown in magenta. C1%—size of the largest cluster compared to the overall number of models [Color figure can be viewed at wileyonlinelibrary.com]

terminal residue, we expect these modifications to have limited effect on the overall tertiary structure. The largest cluster of Cypemycin peptide contained 49% of the overall number of modeled structures. The centroid structure (Figure 8A) matched other β-hairpin AMPs such as Protegrin-2 (PDB code 2MUH), Tachyplesin I (1WO1, and wild-type 1MA2), θ-defensin HTD-2 (2LZI), Thanatin (8TFV) and Polyphemusin I peptide (1RKK), which are all, with the exception of antiviral HTD-2, active against both Gram-positives and Gram-negatives.[52] Due to strict thresholds upon filtering GESAMT results, β-hairpin top results such as Tachyplesin I (1WO1), Cyclic L27-11 (2M7I) and Retrocyclin-2 (2LZI) were filtered out and so were not considered with the other β-hairpin matches above.

Scolopendin 1[114] (Figure 8B) was found to have a fold similar to bovine neutrophil β-defensin 12 (PDB code 1BNB), chicken AvBD2 defensin (2LG5), human α-defensin HNP1 (3GNY), as well as Spheniscin-2 (1UT3) and Tricyclon A (1YP8). The largest cluster contained 24.2% of the overall number of models. Scolopendin 1 is active against a wide spectrum of Gram-positive and Gram-negative bacteria, as well as fungi, and in *Candida albicans* it was shown to induce reactive oxygen species (ROS) accumulation. With the exception of antiviral, negatively charged Tricyclon A, all of the fold matches exhibit activity against Gram-positives and Gram-negatives.

Microcin M[115,116] (Figure 8C) is a class IIb microcin, which contains a siderophore moiety C-terminal post-translational modification, that enables the uptake of peptide but was not included in our modeling. The largest cluster contained 15.5% of the overall number of modeled structures. It predicts Microcin M to contain a four-stranded β-sheet but no fold matches with known AMP structures or structures from the PDB90 were found.

3.3.2 | Mixed αβ folds

Upon performing CLICK database search, five mixed αβ folds were found to have fold matches that met the size and Z-score criteria (see Methods). The first, Propionicin-F[117] (Figure 9A) is an unmodified, heat-stable, negatively charged bacteriocin with a ββα fold. Although only 17.9% successfully modeled structures were clustered into the largest cluster, this peptide matched many αββ folds, such as mussel Mytilin (PDB code 2EEM), Plectasin (1ZFU), Termicin (1MM0), Cg-Def defensin (2B68), Micasin (2LR5), and Eurocin (2LT8). Interestingly, only a single ββα fold was found among the filtered results, namely Leucocin A (1CW6).[52] However, the short β-sheet does not pack against the helix in the same way as seen in our Rosetta model, leading to a much more elongated structure. In contrast, the topologically-distinct αββ folds share a similar mode of interaction between helix and β-sheet as Propionicin-F suggesting that they may be structural analogues. Like many bacteriocins, Propionicin-F has a nanomolar activity against strains of its producer organism, *Propionibacterium freudenreichii*.

ABP-118α[118] shown in Figure 9B is a type IIb, unmodified, heat-stable bacteriocin, active against both Gram-positives and Gram-negatives. These bacteriocins are comprised of two peptide chains, and the overall activity is obtained by the complementary activity of the two peptides. Interestingly, the α component inhibits bacterial growth on its own, and while the β component has no activity, it complements

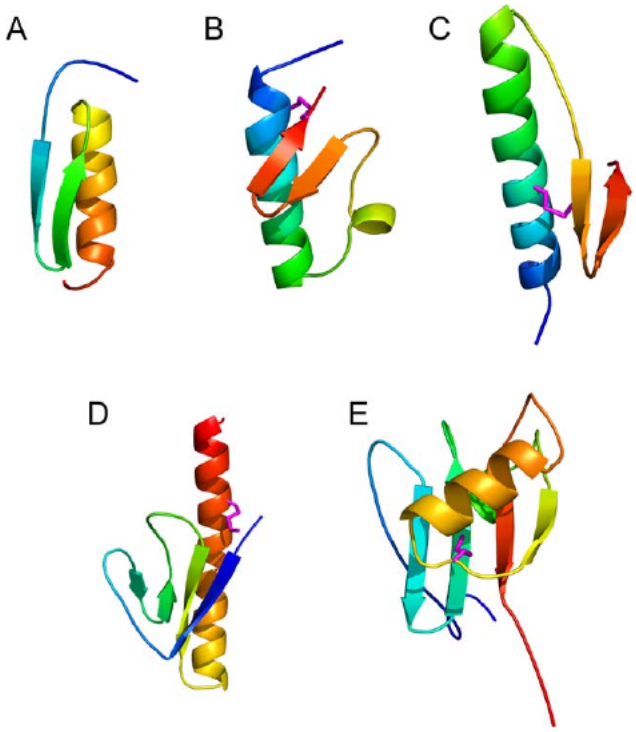

FIGURE 9 Plausible mixed αβ folds. A, Propionicin-F (Q6E3K9, C1% 17.9), B, ABP-118α (AP01172, C1% 26.0, disulfide bond connectivity was predicted to be Cys6-Cys44), C, Mutacin IV chain A (AP01174, C1% 20.2, disulfide bond connectivity was predicted to be Cys14-Cys34), D, Blp1 chain A (AP01995, C1% 28.1, disulfide bond connectivity was predicted to be Cys2-Cys63), and E, Ipomicin (AP01604, C1% 8.48, disulfide bond connectivity was predicted to be Cys32-Cys63). Structures were color-coded from N-terminus (blue) to C-terminus (red). Disulfide bonds are shown in magenta. C1%—size of the largest cluster compared to the overall number of models [Color figure can be viewed at wileyonlinelibrary.com]

the activity of the α chain.[118] Here, we modeled the α chain, with disulfide bond connectivity predicted to be Cys6-Cys44. 26% of successfully modeled structures were clustered in the top cluster. Its predicted αββ fold matches several AMPs, all containing multiple disulfide bonds: Termicin (1MM0), Beta-purothionin (1BHP) and and Viscotoxin A3 (1ED0). The latter two peptides show activity against Gram-positives, Gram-negatives and fungi, while Termicin is active against Gram-positive bacteria and fungi.

Mutacin IV[119] (Figure 9C) shows activity against the Mitis group of oral streptococci. The peptide was thought to be a type IIb bacteriocin,[119] but this has not been demonstrated unambiguously.[120] We have found the A chain of Mutacin IV to match Mytilin (PDB code 2EEM) and Termicin (1MM0), as well as the Mediterranean mussel defensin MGD-1 (1FJN) AMP. The disulfide connectivity for Mutacin IV A was predicted to be Cys14-Cys34 and largest cluster contained 20.2% of the successfully modeled structures.

This group also contains more complex folds, such as chain A Blp1 peptide—another type IIb bacteriocin (Figure 9D). The highly populated largest cluster (28.1% of models) obtained with disulfide connectivity Cys2-Cys63 suggests a successful fold prediction, but it gave no fold matches with AMPs. A CLICK database search against the PDB90

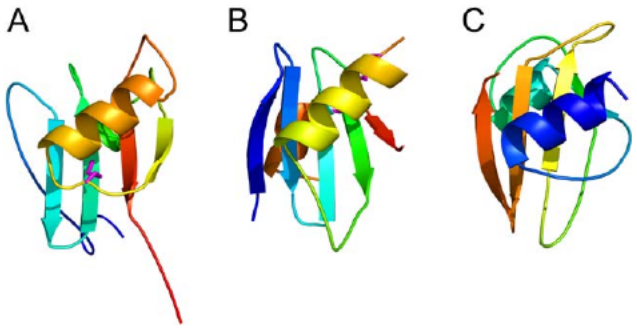

FIGURE 10 Ipomicin and its topologically distinct fold matches aligned. A, Ipomicin peptide (AP01604), B, BRICHOS domain of Lung Surfactant Protein C (PDB code 2YAD), and C, Uracil-DNA glycosylase protein (1UGI). Structures were color-coded from N-terminus (blue) to C-terminus (red) [Color figure can be viewed at wileyonlinelibrary.com]

database revealed similarity with Pseudopilin GspI, a chain from enterotoxigenic *Escherichia coli* secretion system (PDB code 3CI0, chain I),[121] supporting the biophysical plausibility of our modeled fold. However, the two are likely to have evolved independently: Pseudopilins are part of the type 2 secretion system found in Gram-negative bacteria, and GspI is known to be located at the pseudopilus base, interacting with the inner membrane components,[122] while Blp1 AMP is produced by the human oral strain *Lactobacillus salivarius* and is active against Gram-positive bacteria.

Our most complex fold, Ipomicin peptide (Figure 9E), is active against strains of the producer organism, *Streptomyces ipomoeae*. With its six secondary structure elements it did not match any known AMP folds. Although only 8.48% of successfully modeled structures were clustered in the top cluster, results of CLICK database search ran on structures from the PDB90 database revealed striking similarity to 2 known folds: BRICHOS domain of Lung Surfactant Protein C (2YAD) and an Uracil-DNA glycosylase protein (1UGI). Both of these are topologically different from Ipomicin, yet the secondary structure elements are positioned in a similar manner (Figure 10). We suggest that this packing is favorable, irrespective of the topological connection of the secondary structure elements, supporting the reliability of the Ipomicin fold prediction. This in turn suggests that the Ipomicin fold prediction adds a new and relatively complex architecture to the array of folds sampled by AMPs.

3.3.3 | All-α folds

Overall ten all-α folds modeled (Figure 11) were scanned for fold matches using the CLICK database search, and only one was found to have fold matches, namely Vejovine peptide[123] (Figure 11I). The largest cluster contained 18.9% of the total number of models, and the top three clusters all showed similar structure, supporting the reliability of the model.

Vejovine peptide, found in scorpion venom of *Vaejovis mexicanus*, is active against Gram-negative bacteria and hemolytic to human erythrocytes. The peptide is helical in 60% TFE, and the first 8 residues at N-terminus are crucial for its activity.[123] Our Rosetta model suggests a

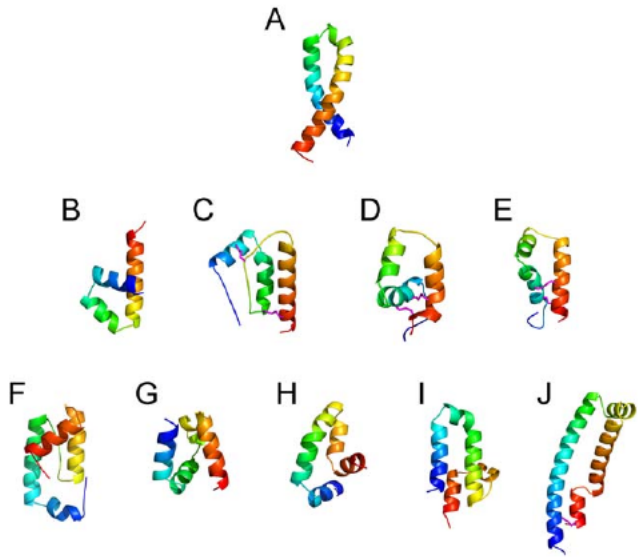

FIGURE 11 Plausible all-α folds. A, Pilosulin 1 (AP00535, C1% 26.2), B, Mutacin BHT-B (AP02251, C1% 40.8), C, PAMP (AP01600, C1% 28.0, disulfide bond connectivity was predicted to be Cys12-Cys37 and Cys29-Cys60), D, Perinerin (AP01235, C1% 31.1, disulfide bond connectivity was predicted to be Cys13-Cys44 and Cys16-Cys49), E, Perinerin with disulfide bond connectivity Cys13-Cys49 and Cys16-Cys44 and C1% 25.4, F, Halocin C8 (AP01193, C1% 21.3), G, BP100_gtag (C1% 43.1), H, Oxyopinin 1 (AP00772, C1% 44.8), I, Vejovine (AP01753, C1% 18.9), and J, Closticin 574 (AP00768, C1% 37.6, disulfide bond connectivity was predicted to be Cys9-Cys81). Structures were color-coded from N-terminus (blue) to C-terminus (red). Disulfide bonds are shown in magenta. C1%—size of the largest cluster compared to the overall number of models [Color figure can be viewed at wileyonlinelibrary.com]

structure containing four helices spanning residues 2–14, 17–38, 42–47, and 50–58, respectively. Neither GESAMT nor CLICK database searches against characterized AMPs returned results. However, an acyl carrier protein fold from *Thermus thermophilus*, with similar topology, was found to match Vejovine peptide through CLICK database search against PDB90 (Figure 12B).

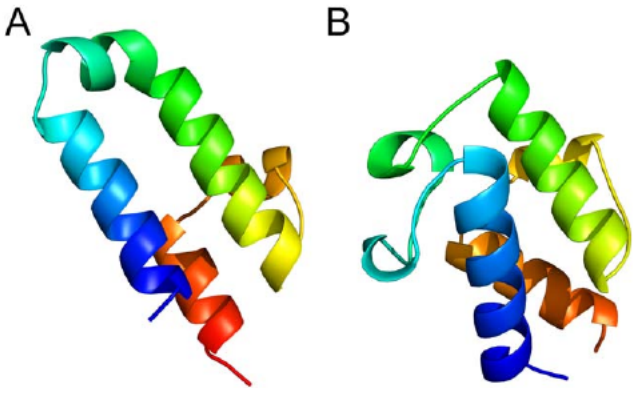

FIGURE 12 Vejovine and its topologically similar fold match from CLICK. A, Vejovine peptide and B, *T. thermophilus* acyl carrier protein (1X3O), color coded from N (blue) to C terminus (red) [Color figure can be viewed at wileyonlinelibrary.com]

4 | CONCLUSIONS

We have modeled the structures of a large set of previously structurally uncharacterized AMPs using a variety of computational methods to ensure robustness. The set was obtained from the APD2 database and a literature search, and contained 184 sequences of between 20 and 120 residues. The validity of the ab initio modeling was supported by both benchmarking against known AMP structures and fold stability testing using Molecular Dynamics experiments.

The newly mapped structural landscape reveals AMPs with similarities to existing folds across different classes, and also predicts new folds for several AMPs. So, for example, familiar α-helical folds were predicted for Ponericin Q42, Latarcin 4a, Kassinatuerin 1, Ceratotoxin D, and CPF-B1 peptide and β-hairpins for Lebocin Peptide 1A (LP1A), Odorranain M1, and Silkworm 001. Interestingly, LP1A and amphibian Odorranain-M1 are the first β-hairpin folds lacking intramolecular disulfide bonds, cation-π or aromatic interactions reported so far, stable only due to hydrogen bonding. Until now, tigerinins, which adopt a beta-turn fold due to a disulfide bridge between two cysteine residues forming a nonapeptide ring, were the only examples of non-helical amphibian antimicrobial peptides. Moreover, a single insect-derived β-hairpin AMP has been reported so far, and we predicted that both insect peptides, Lebocin Peptide 1A (LP1A), and Silkworm 001, are β-hairpins. Examples of mixed αβ folds are Garvicin A and Human Histatin 2, which contain no disulfide bonds, usually found in such AMPs as part of the cysteine-stabilized αβ (CSαβ) motif, making the modeled peptides the first linear αββ fold AMPs lacking intramolecular disulfide bonds. Novel folds were predicted for Microcin M and Ipomicin, the latter resembling the BRICHOS domain of Lung Surfactant Protein C, but with different topology between the secondary structure elements. These findings expand our knowledge of the AMP structural universe and may contribute to the structure-based development of more potent AMPs.

ACKNOWLEDGMENT

We acknowledge BMSI (A*STAR) and National Supercomputing Center Singapore for computational support.

ORCID

Daniel J. Rigden http://orcid.org/0000-0002-7565-8937

REFERENCES

[1] World Health Organization. Antimicrobial Resistance: Global Report on Surveillance [Internet]. World Health Organization; 2014. http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf?ua=1[TQ1]

[2] Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States, 2013 [Internet]. Centers for Disease Control and Prevention; 2013. https://www.cdc.gov/drugresistance/threat-report-2013/pdf/ar-threats-2013-508.pdf

[3] Narayana JL, Chen J-Y. Antimicrobial peptides: possible anti-infective agents. *Peptides*. 2015;72:88–94.

[4] Moreno-Habel DA, Biglang-Awa IM, Dulce A, Luu DD, Garcia P, Weers PMM. Inactivation of the budded virus of *Autographa californica* M nucleopolyhedrovirus by gloverin. *J Invertebr Pathol*. 2012;110(1):92–101.

[5] Sang Y, Blecha F. Antimicrobial peptides and bacteriocins: alternatives to traditional antibiotics. *Anim Health Res Rev*. 2008;9(02):227–235.

[6] Teixeira V, Feio MJ, Bastos M. Role of lipids in the interaction of antimicrobial peptides with membranes. *Prog Lipid Res*. 2012;51(2):149–177.

[7] Li J, Koh J-J, Liu S, Lakshminarayanan R, Verma CS, Beuerman RW. Membrane active antimicrobial peptides: translating mechanistic insights to design. *Front Neurosci*. 2017;11:73

[8] Yau WM, Wimley WC, Gawrisch K, White SH. The preference of tryptophan for membrane interfaces. *Biochemistry*. 1998;37(42):14713–14718.

[9] Vogel HJ, Schibli DJ, Jing W, Lohmeier-Vogel EM, Epand RF, Epand RM. Towards a structure-function analysis of bovine lactoferricin and related tryptophan- and arginine-containing peptides. *Biochem Cell Biol*. 2002;80(1):49–63.

[10] Fleury Y, Dayem MA, Montagne JJ, et al. Covalent structure, synthesis, and structure-function studies of mesentericin Y 105(37), a defensive peptide from gram-positive bacteria *Leuconostoc mesenteroides*. *J Biol Chem*. 1996;271(24):14421–14429.

[11] Shepherd CM, Vogel HJ, Tieleman DP. Interactions of the designed antimicrobial peptide MB21 and truncated dermaseptin S3 with lipid bilayers: molecular-dynamics simulations. *Biochem J*. 2003;370(1):233–243.

[12] Aliste MP, MacCallum JL, Peter Tieleman D. Molecular dynamics simulations of pentapeptides at interfaces: salt bridge and cation−π interactions. *Biochemistry*. 2003;42(30):8976–8987.

[13] Petersen FNR, Jensen MØ, Nielsen CH. Interfacial tryptophan residues: a role for the cation-pi effect?. *Biophys J*. 2005;89(6):3985–3996.

[14] Gleason NJ, Vostrikov VV, Greathouse DV, Grant CV, Opella SJ, Koeppe RE. 2nd,. Tyrosine replacing tryptophan as an anchor in GWALP peptides. *Biochemistry*. 2012;51(10):2044–2053.

[15] Sparks KA, Gleason NJ, Gist R, Langston R, Greathouse DV, Koeppe RE. 2nd,. Comparisons of interfacial Phe, Tyr, and Trp residues as determinants of orientation and dynamics for GWALP transmembrane peptides. *Biochemistry*. 2014;53(22):3637–3645.

[16] Chan DI, Prenner EJ, Vogel HJ. Tryptophan- and arginine-rich antimicrobial peptides: structures and mechanisms of action. *Biochim Biophys Acta*. 2006;1758(9):1184–1202.

[17] Bi X, Wang C, Ma L, Sun Y, Shang D. Investigation of the role of tryptophan residues in cationic antimicrobial peptides to determine the mechanism of antimicrobial action. *J Appl Microbiol*. 2013;115(3):663–672.

[18] Wang G. Antimicrobial Peptides: Discovery, Design and Novel Therapeutic Strategies. CABI, Wallingford, UK; 2010:230.

[19] Wang G, Li X, Wang Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res*. 2009;37(Database issue):D933–D937.

[20] Dathe M, Wieprecht T. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim Biophys Acta*. 1999;1462(1–2):71–87.

[21] Marr AK, Gooderham WJ, Hancock RE. Antibacterial peptides for therapeutic use: obstacles and realistic outlook. *Curr Opin Pharmacol*. 2006;6(5):468–472.

[22] Bai Y, Liu S, Li J, et al. Progressive structuring of a branched antimicrobial peptide on the path to the inner membrane target. *J Biol Chem*. 2012;287(32):26606–26617.

[23] Sani M-A, Separovic F. How membrane-active peptides get into lipid membranes. *Acc Chem Res*. 2016;49(6):1130–1138.

[24] Gazit E, Miller IR, Biggin PC, Sansom MS, Shai Y. Structure and orientation of the mammalian antibacterial peptide cecropin P1 within phospholipid membranes. *J Mol Biol*. 1996;258(5):860–870.

[25] Brandenburg KS, Rubinstein I, Sadikot RT, Önyüksel H. Polymyxin B self-associated with phospholipid nanomicelles. *Pharm Dev Technol*. 2012;17(6):654–660.

[26] Brooks BD, Brooks AE. Therapeutic strategies to combat antibiotic resistance. *Adv Drug Deliv Rev*. 2014;78:14–27.

[27] Wimley WC, Hristova K. Antimicrobial peptides: successes, challenges and unanswered questions. *J Membr Biol*. 2011;239(1–2):27–34.

[28] Kozić M, Vukičević D, Simunić J, et al. Predicting the minimal inhibitory concentration for antimicrobial peptides with Rana-box domain. *J Chem Inf Model*. 2015;55(10):2275–2287.

[29] Fernandes FC, Rigden DJ, Franco OL. Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. *Biopolymers*. 2012;98(4):280–287.

[30] Edwards IA, Elliott AG, Kavanagh AM, Zuegg J, Blaskovich MAT, Cooper MA. Contribution of amphipathicity and hydrophobicity to the antimicrobial activity and cytotoxicity of β-hairpin peptides. *ACS Infect Dis*. 2016;2(6):442–450.

[31] Yang N, Liu X, Teng D, et al. Antibacterial and detoxifying activity of NZ17074 analogues with multi-layers of selective antimicrobial actions against *Escherichia coli* and *Salmonella enteritidis*. *Sci Rep*. 2017;7(1):3392.

[32] Tassanakajon A, Somboonwiwat K, Amparyup P. Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Dev Comp Immunol*. 2015;48(2):324–341.

[33] Tennessen JA. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *J Evol Biol*. 2005;18(6):1387–1394.

[34] Erler S, Lhomme P, Rasmont P, Lattorff HMG. Rapid evolution of antimicrobial peptide genes in an insect host-social parasite system. *Infect Genet Evol*. 2014;23:129–137.

[35] Lazzaro BP, Clark AG. Rapid evolution of innate immune response genes. In: Singh RS, Xu J, Kulathinal RJ, eds. *Rapidly Evolving Genes and Genetic Systems*. United Kingdom: Oxford University Press, Oxford; 2012:203–210.

[36] Murphy PM. Molecular mimicry and the generation of host defense protein diversity. *Cell*. 1993;72(6):823–826.

[37] Hughes AL. Evolutionary diversification of the mammalian defensins. *Cell Mol Life Sci*. 1999;56(1–2):94–103.

[38] Klepeis JL, Wei Y, Hecht MH, Floudas CA. Ab initio prediction of the three-dimensional structure of a de novo designed protein: a double-blind case study. *Proteins*. 2004;58(3):560–570.

[39] Wu S, Skolnick J, Zhang Y. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol*. 2007;5:17.

[40] Zhang Y. Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10. *Proteins*. 2014;82:175–187.

[41] Zhang W, Yang J, He B, et al. Integration of QUARK and I-TASSER for Ab initio protein structure prediction in CASP11. *Proteins*. 2016;84(S1):76–86.

[42] Kinch LN, Li W, Monastyrskyy B, Kryshtafovych A, Grishin NV. Evaluation of free modeling targets in CASP11 and ROLL. *Proteins*. 2016;84(S1):51–66.

[43] Petersen MT, Jonson PH, Petersen SB. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng*. 1999;12(7):535–548.

[44] Lehrer RI. Evolution of antimicrobial peptides: a view from the cystine chapel. In: Hiemstra PS, Zaat SAJ, eds. *Antimicrobial Peptides and Innate Immunity*. Switzerland: Springer Basel; 2012: 1–27.

[45] Andrä J, Hammer MU, Grötzinger J, et al. Significance of the cyclic structure and of arginine residues for the antibacterial activity of arenicin-1 and its interaction with phospholipid and lipopolysaccharide model membranes. *Biol Chem*. 2009;390(4):337–349.

[46] Schroeder BO, Wu Z, Nuding S, et al. Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. *Nature*. 2011;469(7330):419–423.

[47] Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004;383:66–93.

[48] Yeaman MR, Yount NY. Mechanisms of antimicrobial peptide action and resistance. *Pharmacol Rev*. 2003;55(1):27–55.

[49] Apweiler R, Bairoch A, Wu CH, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004;32(Database issue):D115–119.

[50] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–1659.

[51] Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. *Proteins*. 2009;77(S9):128–132.

[52] Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28(1):235–242.

[53] Dosztányi Z, Csizmók V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*. 2005;347(4):827–839.

[54] Ceroni A, Passerini A, Vullo A, Frasconi P. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res*. 2006;34(Web Server):W177–W181.

[55] Ferrè F, Clote P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res*. 2005;33(Web Server issue):W230–W232.

[56] Yaseen A, Li Y. Dinosolve: a protein disulfide bonding prediction server using context-based features to enhance prediction accuracy. *BMC Bioinformatics*. 2013;14(S13):S9

[57] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2015;43(D1):D6–17.

[58] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402.

[59] Liu SP, Zhou L, Lakshminarayanan R, Beuerman RW. Multivalent antimicrobial peptides as therapeutics: design principles and structural diversities. *Int J Pept Res Ther*. 2010;16(3):199–213.

[60] Min HJ, Yun H, Ji S, et al. Rattusin structure reveals a novel defensin scaffold formed by intermolecular disulfide exchanges. *Sci Rep*. 2017;7:45282.

[61] Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem*. 2004;25(6):865–871.

[62] Krissinel E. Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem*. 2012;1(2):76–85.

[63] Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004;20(18):3702–3704.

[64] Nguyen MN, Tan KP, Madhusudhan MS. CLICK-topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res*. 2011;39(Web Server issue):W24–W28.

[65] Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. *Amber 14*. San Francisco, CA: University of California; 2014.

[66] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput*. 2015;11(8):3696–3713.

[67] Feig M, Karanicolas J, Brooks CL. 3rd,. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model*. 2004;22(5):377–395.

[68] Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. *Q Rev Biophys*. 2003;36(3):307–340.

[69] Broekaert WF, Terras FR, Cammue BP, Osborn RW. Plant defensins: novel antimicrobial peptides as components of the host defense system. *Plant Physiol*. 1995;108(4):1353–1358.

[70] Shafee TMA, Lay FT, Hulett MD, Anderson MA. The defensins consist of two independent, convergent protein superfamilies. *Mol Biol Evol*. 2016;33(9):2345–2356.

[71] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000;16(4):404–405.

[72] Li J, Xu X, Xu C, et al. Anti-infection peptidomics of amphibian skin. *Mol Cell Proteomics*. 2007;6(5):882–894.

[73] Maldonado-Barragan A, Cardenas N, Martinez B, et al. Garvicin A, a novel class IId bacteriocin from *Lactococcus garvieae* that inhibits septum formation in *L. garvieae* strains. *Appl Environ Microbiol*. 2013;79(14):4336–4346.

[74] Patil AA, Ouellette AJ, Lu W, Zhang G. Rattusin, an intestinal α-defensin-related peptide in rats with a unique cysteine spacing pattern and salt-insensitive antibacterial activities. *Antimicrob Agents Chemother*. 2013;57(4):1823–1831.

[75] Fujita K, Ichimasa S, Zendo T, et al. Structural analysis and characterization of lacticin Q, a novel bacteriocin belonging to a new family of unmodified bacteriocins of Gram-positive bacteria. *Appl Environ Microbiol*. 2007;73(9):2871–2877.

[76] Yoneyama F, Imura Y, Ichimasa S, et al. Lacticin Q, a lactococcal bacteriocin, causes high-level membrane permeability in the absence of specific receptors. *Appl Environ Microbiol*. 2009;75(2):538–541.

[77] Lohans CT, Towle KM, Miskolzie M, et al. Solution structures of the linear leaderless bacteriocins enterocin 7A and 7B resemble carnocyclin A, a circular antimicrobial peptide. *Biochemistry*. 2013;52(23):3987–3994.

[78] Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010;38(Web Server issue):W545–W549.

[79] Jack RW, Tagg JR, Ray B. Bacteriocins of Gram-positive bacteria. *Microbiol Rev*. 1995;59(2):171–200.

[80] Cotter PD, Hill C, Ross RP. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol*. 2005;3(10):777–788.

[81] Li M, Yoneyama F, Toshimitsu N, Zendo T, Nakayama J, Sonomoto K. Lethal hydroxyl radical accumulation by a lactococcal bacteriocin, lacticin Q. *Antimicrob Agents Chemother*. 2013;57(8):3897–3902.

[82] Martin-Visscher LA, Gong X, Duszyk M, Vederas JC. The three-dimensional structure of carnocyclin a reveals that many circular bacteriocins share a common structural motif. *J Biol Chem*. 2009;284(42):28674–28681.

[83] Killian JA, Salemink I, de Planque MR, Lindblom G, Koeppe RE, 2nd, Greathouse DV. Induction of nonbilayer structures in diacylphosphatidylcholine model membranes by transmembrane alpha-helical peptides: importance of hydrophobic mismatch and proposed role of tryptophans. *Biochemistry*. 1996;35(3):1037–1045.

[84] de Planque MRR, Bonev BB, Demmers JAA, et al. Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide−lipid interactions†. *Biochemistry*. 2003;42(18):5341–5348.

[85] Towle KM, Vederas JC. Structural features of many circular and leaderless bacteriocins are similar to those in saposins and saposin-like peptides. *Med Chem Commun*. 2017;8(2):276–285.

[86] Yoneyama F, Imura Y, Ohno K, et al. Peptide-lipid huge toroidal pore, a new antimicrobial mechanism mediated by a lactococcal bacteriocin, lacticin Q. *Antimicrob Agents Chemother*. 2009;53(8):3211–3217.

[87] Acedo JZ, van Belkum MJ, Lohans CT, Towle KM, Miskolzie M, Vederas JC. Nuclear magnetic resonance solution structures of lacticin Q and aureocin A53 reveal a structural motif conserved among leaderless bacteriocins with broad-spectrum activity. *Biochemistry*. 2016;55(4):733–742.

[88] Rayaprolu S, Wang Y, Kanost MR, Hartson S, Jiang H. Functional analysis of four processing products from multiple precursors encoded by a lebocin-related gene from *Manduca sexta*. *Dev Comp Immunol*. 2010;34(6):638–647.

[89] Panteleev PV, Bolosov IA, Balandin SV, Ovchinnikova TV. Structure and biological functions of β-hairpin antimicrobial peptides. *Acta Naturae*. 2015;7(1):37–47.

[90] Conibear AC, Johan Rosengren K, Daly NL, Henriques ST, Craik DJ. The cyclic cystine ladder in θ-defensins is important for structure and stability, but not antibacterial activity. *J Biol Chem*. 2013;288(15):10830–10840.

[91] Tang YQ, Yuan J, Osapay G, et al. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science*. 1999;286(5439):498–502.

[92] Ladenstein R, Ren B. Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. *FEBS J*. 2006;273(18):4170–4185.

[93] Mangoni ME, Aumelas A, Charnet P, et al. Change in membrane permeability induced by protegrin 1: implication of disulphide bridges for pore formation. *FEBS Lett*. 1996;383(1–2):93–98.

[94] Dawson RM, Liu C-Q. Disulphide bonds of the peptide protegrin-1 are not essential for antimicrobial activity and haemolytic activity. *Int J Antimicrob Agents*. 2010;36(6):579–580.

[95] Lee J-U, Kang D-I, Zhu WL, Shin SY, Hahm K-S, Kim Y. Solution structures and biological functions of the antimicrobial peptide, arenicin-1, and its linear derivative. *Biopolymers*. 2007;88(2):208–216.

[96] Ma B, Niu C, Zhou Y, et al. The disulfide bond of the peptide thanatin is dispensible for its antimicrobial activity in vivo and in vitro. *Antimicrob Agents Chemother*. 2016;60(7):4283–4289.

[97] Hetru C, Letellier L, Oren Z, Hoffmann JA, Shai Y. Androctonin, a hydrophilic disulphide-bridged non-haemolytic anti-microbial peptide: a plausible mode of action. *Biochem J*. 2000;345(3):653–664.

[98] Matsuzaki K, Yoneyama S, Fujii N, et al. Membrane permeabilization mechanisms of a cyclic antimicrobial peptide, tachyplesin I, and its linear analog†. *Biochemistry*. 1997;36(32):9799–9806.

[99] Mangoni ML, Papo N, Mignogna G, et al. Ranacyclins, a new family of short cyclic antimicrobial peptides: biological function, mode of action, and parameters involved in target specificity. *Biochemistry*. 2003;42(47):14023–14035.

[100] Laederach A, Andreotti AH, Fulton DB. Solution and micelle-bound structures of tachyplesin I and its active aromatic linear derivatives. *Biochemistry*. 2002;41(41):12359–12368.

[101] Saravanan R, Mohanram H, Joshi M, et al. Structure, activity and interactions of the cysteine deleted analog of tachyplesin-1 with

lipopolysaccharide micelle: Mechanistic insights into outer-membrane permeabilization and endotoxin neutralization. *Biochim Biophys Acta*. 2012;1818(7):1613–1624.

[102] Mohanram H, Bhattacharjya S. Cysteine deleted protegrin-1 (CDP-1): anti-bacterial activity, outer-membrane disruption and selectivity. *Biochim Biophys Acta*. 2014;1840(10):3006–3016.

[103] Mani R, Waring AJ, Lehrer RI, Hong M. Membrane-disruptive abilities of beta-hairpin antimicrobial peptides correlate with conformation and activity: a 31P and 1H NMR study. *Biochim Biophys Acta*. 2005;1716(1):11–18.

[104] Xu L, Lal K, Pollock JJ. Histatins 2 and 4 are autoproteolytic degradation products of human parotid saliva. *Oral Microbiol Immunol*. 1992;7(2):127–128.

[105] Wu Z, Hoover DM, Yang D, et al. Engineering disulfide bridges to dissect antimicrobial and chemotactic activities of human beta-defensin 3. *Proc Natl Acad Sci U S A*. 2003;100(15):8880–8885.

[106] Lee IH, Lee YS, Kim CH, et al. Dicynthaurin: an antimicrobial peptide from hemocytes of the solitary tunicate, *Halocynthia aurantium*. *Biochim Biophys Acta*. 2001;1527(3):141–148.

[107] Pluzhnikov KA, Kozlov SA, Vassilevski AA, Vorontsova OV, Feofanov AV, Grishin EV. Linear antimicrobial peptides from *Ectatomma quadridens* ant venom. *Biochimie*. 2014;107(Pt B):211–215.

[108] Rosetto M, De Filippis T, Manetti AG, Marchini D, Baldari CT, Dallai R. The genes encoding the antibacterial sex-specific peptides ceratotoxins are clustered in the genome of the medfly *Ceratitis capitata*. *Insect Biochem Mol Biol*. 1997;27(12):1039–1046.

[109] Mechkarska M, Ahmed E, Coquet L, et al. Antimicrobial peptides with therapeutic potential from skin secretions of the Marsabit clawed frog *Xenopus borealis* (Pipidae). *Comp Biochem Physiol C Toxicol Pharmacol*. 2010;152(4):467–472.

[110] Kozlov SA, Vassilevski AA, Feofanov AV, Surovoy AY, Karpunin DV, Grishin EV. Latarcins, antimicrobial and cytolytic peptides from the venom of the spider *Lachesana tarabaevi* (Zodariidae) that exemplify biomolecular diversity. *J Biol Chem*. 2006;281(30):20983–20992.

[111] Mattute B, Knoop FC, Conlon JM. Kassinatuerin-1: a peptide with broad-spectrum antimicrobial activity isolated from the skin of the hyperoliid frog, *Kassina senegalensis*. *Biochem Biophys Res Commun*. 2000;268(2):433–436.

[112] Dubovskii PV, Volynsky PE, Polyansky AA, Chupin VV, Efremov RG, Arseniev AS. Spatial structure and activity mechanism of a novel spider antimicrobial peptide. *Biochemistry*. 2006;45(35):10759–10767.

[113] Minami Y, Yoshida K-I, Azuma R, et al. Structure of cypemycin, a new peptide antibiotic. *Tetrahedron Lett*. 1994;35(43):8001–8004.

[114] Choi H, Hwang J-S, Lee DG. Identification of a novel antimicrobial peptide, scolopendin 1, derived from centipede *Scolopendra subspinipes mutilans* and its antifungal mechanism. *Insect Mol Biol*. 2014;23(6):788–799.

[115] Duquesne S, Destoumieux-Garzón D, Peduzzi J, Rebuffat S. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat Prod Rep*. 2007;24(4):708–734.

[116] Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. *Antimicrob Agents Chemother*. 2010;54(1):288–297.

[117] Brede DA, Faye T, Johnsborg O, Odegård I, Nes IF, Holo H. Molecular and genetic characterization of propionicin F, a bacteriocin from *Propionibacterium freudenreichii*. *Appl Environ Microbiol*. 2004;70(12):7303–7310.

[118] Flynn S, van Sinderen D, Thornton GM, Holo H, Nes IF, Collins JK. Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. salivarius UCC118. *Microbiology*. 2002;148(4):973–984.

[119] Qi F, Chen P, Caufield PW. The group I strain of *Streptococcus mutans*, UA140, produces both the lantibiotic mutacin I and a non-lantibiotic bacteriocin, mutacin IV. *Appl Environ Microbiol*. 2001;67(1):15–21.

[120] Hale JDF, Ting Y-T, Jack RW, Tagg JR, Heng NCK. Bacteriocin (mutacin) production by *Streptococcus mutans* genome sequence reference strain UA159: elucidation of the antimicrobial repertoire by genetic dissection. *Appl Environ Microbiol*. 2005;71(11):7613–7617.

[121] Korotkov KV, Hol WGJ. Structure of the GspK-GspI-GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system. *Nat Struct Mol Biol*. 2008;15(5):462–468.

[122] Michel GPF, Voulhoux R. The type II secretory system (T2SS) in Gram-negative bacteria: a molecular nanomachine for secretion of sec and tat-dependent extracellular proteins. In: Wooldridge K, editor. *Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis*. Norfolk, UK: Caister Academic Press; 2009.

[123] Hernández-Aponte CA, Silva-Sanchez J, Quintero-Hernández V, et al. Vejovine, a new antibiotic from the scorpion venom of *Vaejovis mexicanus*. *Toxicon*. 2011;57(1):84–92.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Kozic M, Fox SJ, Thomas JM, Verma CS, Rigden DJ. Large scale ab initio modeling of structurally uncharacterized antimicrobial peptides reveals known and novel folds. *Proteins*. 2018;86:548–565. https://doi.org/10.1002/prot.25473