Title

AMPing up the search: An *in silico* approach to identifying Antimicrobial Peptides (AMPs) with potential anti-biofilm activity

Authors

Shreeya Mhade*[1], Stutee Panse*[2], Gandhar Tendulkar*[3], Rohit Awate[4], Snehal Kadam[5], Karishma S Kaushik[6]
*These authors contributed equally

Affiliations

1 Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous), Mumbai, India

2 The Pennsylvania State University, University Park, Pennsylvania, USA

3 Sir Sitaram and Lady Shantabai Patkar College of Arts and Science and V P Varde College of Commerce and Economics (Autonomous), Mumbai, India

4 Northeastern University, Boston, USA

5 Hull York Medical School, University of Hull, United Kingdom

6 Department of Biotechnology, Savitribai Phule Pune University, Pune, India

Abstract

Antibiotic resistance is a public health threat, and the rise of multidrug-resistant bacteria, including those that form protective biofilms, further compounds this challenge. Antimicrobial peptides (AMPs) have been recognized for their anti-infective properties, including their ability to target processes important for biofilm formation. However, given the vast array of natural and synthetic AMPs, determining potential candidates for anti-biofilm testing is a significant challenge. In this study, we present an *in silico* approach, based on open-source tools, to identify AMPs with potential anti-biofilm activity. This approach is developed using the sortase-pilin machinery, important for adhesion and biofilm formation, of the multidrug-resistant, biofilm-forming pathogen *C. striatum* as the target. Using homology modeling, we modeled the structure of the *C. striatum* sortase C protein, resembling the semi-open lid conformation adopted during pilus biogenesis. Next, we developed a structural library of 5544 natural and synthetic AMPs from sequences in the DRAMP database. From this library, AMPs with known anti-Gram positive activity were filtered, and 100 select AMPs were evaluated for their ability to interact with the sortase C protein using *in-silico* molecular docking. Based on interacting residues and docking scores, we built a preference scale to categorize candidate AMPs in order of priority for future *in vitro* and *in vivo* biofilm studies. The considerations and challenges of our approach, and the resources developed, which includes a search-enabled repository of predicted AMP structures and protein-peptide interaction models relevant to biofilm studies (B-AMP), can be leveraged for similar investigations across other biofilm targets and biofilm-forming pathogens.

Keywords

Biofilms, antimicrobial peptides, sortase C, *C. striatum*, protein-peptide molecular docking

Introduction

The spread of multidrug-resistant bacteria, and the emergence of pathogens inherently resistant to a range of antibiotics, is a serious challenge. This problem is further exacerbated by the ability of bacteria to form protective communities or biofilms. In biofilms, bacteria are notoriously tolerant to antibiotics and often require long-term treatments, which further fuels the development of antibiotic resistance. Given this, there is an urgent need to search for novel anti-biofilm approaches to complement and potentially serve as alternatives to standard antibiotics. Antimicrobial peptides (AMPs) are a diverse class of peptides with a range of structural and functional properties [1]. In addition to their well-known effects on disrupting bacterial membrane integrity, several AMPs target specific bacterial components, and can thereby inhibit processes such as biofilm formation [2]. Bacterial cell-surface pili play an important role in the various stages of biofilm formation, including initial attachment and adhesion [3]. In Gram-positive bacteria [4], pilus biogenesis is mediated by a group of sortase enzymes, and inhibition of sortase activity results in defective biofilm formation [5–7]. AMPs have been shown to demonstrate activity as sortase inhibitors, with the potential to act in adjunct with standard antibiotics [8–10]. While this is promising, given the vast array of AMPs, identifying potential candidates for anti-biofilm testing remains a significant challenge.

In this study, we present an *in silico* approach to identify AMPs with a potential anti-biofilm activity using the sortase-pilin machinery of the multidrug-resistant, biofilm-forming pathogen *Corynebacterium striatum* as a target. Using homology modeling, we modeled the semi-open lid conformation of the *C. striatum* sortase C protein adopted during pilus biogenesis. With a combination of predictive peptide modeling tools, we developed structural models of 5544 natural and synthetic AMPs from the DRAMP database. From this library, AMPs with known anti-Gram positive activity were filtered, and 100 select AMPs were evaluated for their ability to interact with the sortase C protein using *in-silico* molecular docking. Docking results were used to build a preference scale to categorize AMPs in order of priority for future testing against *C. striatum* biofilms. In doing so, this *in silico* approach enables the identification and categorization of candidate AMPs with potential anti-biofilm activity, and provides a starting point for *in vitro* and *in vivo* biofilm testing. Finally, while our study focuses on *C. striatum*, we argue that the approach we present, considerations described and resources developed, which includes a search-enabled repository of predicted AMP structures and protein-peptide interaction models for biofilm studies (Biofilm-AMP or B-AMP), can be leveraged for similar investigations across other biofilm targets and biofilm-forming pathogens.

Methods

Homology modeling of the *C. striatum* sortase C protein

The three-dimensional structure of the sortase C protein of *Corynebacterium striatum* is not available in the RCSB Protein Data Bank (PDB) [11]. To construct the protein structure of the *C. striatum* sortase C enzyme, the amino acid sequence of the sortase C protein was retrieved from the GenPept Database (ID: WP_034656562) [12]. The physicochemical properties of the protein such as molecular weight, isoelectric point (pI), amino acid composition, estimated half-life, and instability index were determined using the ProtParam tool [13]. The secondary structure of the sortase C protein was predicted by the PSIPRED server [14]. Both physicochemical and secondary structure properties (helix, strand, or coil) of the protein were determined using the amino acid sequence as an input. The three-dimensional (3D) protein structure of the class C sortase enzyme was constructed using I-TASSER [15], by submitting the FASTA sequence retrieved from GenPept as an input (without any user-specific restraints). To mimic the semi-open lid conformation of the *C. striatum* sortase C protein, we generated a model that included a double mutation D93A/W95A (replacing Aspartate (93) and Tryptophan (95) with Alanine) in the protein. The input FASTA file was modified manually by replacing A in place of D and W.

Validation of the predicted 3D model of the *C. striatum* sortase C protein

The reliability of the sortase C protein model, including the modified protein model with semi-open lid conformation (generated with I-TASSER), was verified using Structure Analysis and Verification Server version 6.0 (SAVES) [16–18]. The SAVES 6.0 server integrates analysis from multiple, widely-used validation algorithms (including VERIFY 3D, ERRAT, and PROCHECK). The PDB file generated from I-TASSER was submitted as an input to SAVES 6.0. The quality and reliability of the generated model were checked by using the default quality parameters of the ERRAT and VERIFY3D software. In addition, the PROCHECK software was used to validate the generated model of the proteins by utilizing the Ramachandran plot.

Predictive modeling of AMP structures from the DRAMP database

From DRAMP V3.0 [19], we downloaded a dataset of 5562 general AMP sequences, with known antibacterial and antifungal activities. These AMPs were listed as Pep2 to Pep5563, where Pep1 is the pilin subunit (GenPept ID: WP_170219081.1) [20] **(Suppl File 1).** FASTA files were generated using an *in-house* developed Python script. The

Python script reads the .csv file from the DRAMP database and generates FASTA files which were used for predictive modeling.

Peptide modeling with PEP-FOLD3

For modeling AMPs with less than 50 amino acid residues, we used PEP-FOLD3 [21], which predicts 3D peptide structures from sequence information using a *de novo* approach. FASTA file for the structure to be generated was uploaded in the query form, following which a peptide structure for the given sequence was generated and saved in .pdb form.

Peptide modeling with Robetta

For modeling AMPs with more than 50 amino acids, we used the deep-learning-based method TrRosetta [22] available under Robetta [23]. FASTA files of the peptide were uploaded in the query submission interface and the resulting model was downloaded with a .pdb extension.

Peptide modeling using ab-initio I-TASSER modeling

For peptides with unusual or unknown (X) amino acids, we used the *ab-initio* modeling feature of I-TASSER [15,24] for generating 3D conformations. FASTA files of the peptide sequence were uploaded into the query submission form using the default parameters, and the resulting models were downloaded with a .pdb extension.

Protein-peptide *in silico* molecular docking of predicted AMP structures to the mutated *C. striatum* sortase C protein

For protein-peptide *in silico* molecular docking, we filtered out AMPs with annotated anti-Gram positive activity (based on information available in the DRAMP database) using an *in-house* developed Python script (https://github.com/KarishmaKaushikLab/B-AMP). The Python script employs a conditional statement to automatically scan the .csv file for the term “Anti-Gram+” from the activity column in the sheet **(Suppl File 2)**. From this filtered subset, 88 AMPs ranging in length from 2 to 8 amino acid residues, and 12 representative peptides with lengths ranging from 9 to 20 residues were selected for *in-silico* molecular docking. Before molecular docking to the mutated sortase C protein, the 3D AMP structures were subject to Energy Minimization in Swiss-PDBViewer SPDBV v4.1.0 [25] using the partial implementation of the GROMOS96 43B1 force field [26,27]. As a standard, the LPMTG motif of the pilin subunit was docked to the mutated sortase C protein. AutoDockTool1.5.6 [28] was used to generate the necessary input PDBQT files for docking, and AutoDock Vina [29] was used to performing protein-peptide docking on an Intel® Core™ i7-9750H CPU @ 2.60GHz system using the command:

vina --receptor M_Ala_Sortase.pdbqt --ligand Pep#.pdbqt --center_x 57.854 --center_y 47.915 --center_z 46.936 --size_x 28 --size_y 28 --size_z 28 --out vPep#.pdbqt --log vPep#.log. The run time for each docking varied depending on the length of the peptide and ranged from 2 minutes (for peptides 2-4 residues long) to 40 minutes (for peptides 20 residues long) with the system used.

Building a structural repository of AMPs and protein-peptide interactions for biofilm studies (B-AMP)

To enable the availability of resources and data generated in this study, we developed a user-friendly, and search-enabled repository of AMP structures and AMP interactions with biofilm targets (B-AMP) (https://b-amp.karishmakaushiklab.com/). The repository was built using HTML/CSS/JavaScript delivered over GitHub pages, and the scripts and other resources employed have been made available (https://github.com/KarishmaKaushikLab/B-AMP).

Results and Discussion

Sortase C as an anti-biofilm target

Sortase enzymes function to assemble pili or anchor them to the bacterial cell wall, and have been identified as promising antimicrobial targets, particularly in Gram-positive bacteria [30–33]. Across various species, inactivation of the genes encoding sortase enzymes, including sortase C, has been shown to result in a significant reduction in biofilm formation [6,34–36]. Given this, sortases have been recognized as a promising target for anti-biofilm approaches [37]. In *C. striatum*, cell-surface pili facilitate adhesion to host cells and abiotic surfaces, and mediate the formation of biofilms [3]. The sortase-pilin machinery includes the *spaDEF* operon that encodes a set of pilus proteins and their respective sortases [38,39]. In *Corynebacteria* spp, the sortase C enzyme functions as a pilus-specific sortase, recognizing and cleaving the sorting signal (LPMTG) in the pilin subunits, which is followed by their surface display [30,33]. Based on its relevance to *C. striatum* pilus biogenesis, the sortase C protein was selected as the anti-biofilm target.

Sequence-based prediction of physicochemical properties and secondary structure of the *C. striatum* sortase C protein

The amino acid sequence of the *C. striatum* sortase C protein was retrieved from the GenPept Database (ID: WP_034656562) [12]. Using the sequence, the physicochemical properties of the protein were predicted using ProtParam analysis. As shown in **Table 1,** the analysis revealed a molecular weight of 35134.72 Da, and a theoretical isoelectric

point of 5.38 and 5.46 respectively, indicating a negatively-charged protein. In particular, the sortase C protein contained 46 negatively charged residues (Asp+Glu) and 33 positively-charged residues (Arg+Lys). The computed instability index of 33.54 indicates that the protein is stable, as the obtained value is less than the cut-off of 40. The N-terminal sequence containing M (Met) indicates that the half-life of the protein is 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*), >10 hours (*Escherichia coli, in vivo*). Using the PSIPRED server [14], the secondary structure of the sortase C protein was observed to contain a combination of alpha-helices (41%), coils (44%), as well as beta sheets (15%) **(Suppl Figure 1).**

Homology model of the *C. striatum* sortase C protein

The three-dimensional structure of the sortase C protein of *C. striatum* is not available in the RCSB Protein Data Bank (PDB) [11]. To model the structure of the enzyme, I-TASSER selected the following crystal structures as templates: class C sortases of *Streptococcus pneumoniae* (PDB ID: 2W1J; 2W1K; 3G66; 3O0P), *Streptococcus agalactiae* (PDB ID: 3RBI; 4D7W; 4G1H), *Actinomyces oris* (PDB ID: 2XWG), *Clostridium perfringens* (PDB ID: 6IXZ) and Class A sortase from *Corynebacterium diphtheriae* (PDB ID: 5K9A). The sequence identity between *C. striatum* class C sortase and the selected templates of *Streptococcus pneumoniae, Streptococcus agalactiae, Actinomyces oris, Clostridium perfringens and Corynebacterium diphtheriae* was 50% respectively, which is more than the required 30% sequence homology for generating useful models [40]. As seen in **Figure 1A**, results from I-TASSER modeling of the sortase C protein predicted the presence of all the three secondary structure states (helices, coils, beta-sheets), correlating with the PSIPRED server results **(Suppl Figure 1)**. In Gram-positive bacteria, including *Corynebacterium* spp., the sortase C enzyme is known to possess a highly conserved active site cysteine residue, responsible for cleavage of pilin motifs. Present along with the cysteine residue, are histidine and arginine residues (His and Arg), that play essential roles in sortase activity, thereby forming a catalytic triad [30]. In the I-TASSER model of the *C. striatum* sortase C protein, the triad of Cys230-His168-Arg239 was found between the antiparallel beta-sheet strands **(Figure 1A and B)**. Further, based on molecular surface representation with amino acid hydrophobicity colored based on the Kyte-Doolittle scale [41], the active site is likely in a hydrophobic groove **(Figure 1B)**. The overall quality factor of the predicted sortase C model was 73.0263 by the ERRAT **(Suppl Figure 1)**. Moreover, the model was concurred by VERIFY3D showing 80.45% of the residues with an averaged 3D-1D score ≥ 0.2. Ramachandran plot analysis done by PROCHECK servers revealed that 95.6% of residues of the protein were in the most favored and allowed regions and only 4.4% residues in outlier regions.

Modeling the *C. striatum* sortase C protein in semi-open lid conformation

In addition to the characteristic beta-barrel structure and active site Cys-His-Arg triad, the homology model of the wild-type sortase C protein predicted the presence of a 'lid' consisting of a structural loop lodged in the putative active site. The 'lid' of the sortase C protein is suggested to play a role during polymerization of the pilin subunits [33,42], likely via the recognition of the sorting signal. However, there is no evidence that indicates that the pilin substrates are able to access the putative active site in this 'closed' confirmation, indicating that the lid likely undergoes displacement during pilus biogenesis [33,42]. As seen in **Figure 2A and B,** the 'lid' of the sortase C protein carries two important residues, which typically include an aspartate residue and a hydrophobic amino acid, two residues after aspartate, that interact with the putative catalytic site. In the 'closed' confirmation, the sortase C lid blocks access to the active site, observed as proximity between the Asp93 and Arg239 residues. The presence of the aromatic ring of the Trp95 residue in the lid, which is close to the Cys230 residue, renders the active site inaccessible [43]. Point mutations in the 'lid', such as a double mutation D93A/W95A, replacing Aspartate (93) and Tryptophan (95) with Alanine, are known to result in a 'semi-open' lid conformation and increased substrate accessibility [43]. To mimic the semi-open lid conformation, we generated a model that included a double mutation D93A/W95A in the sortase C protein. As seen in **Figure 2C and D**, I-TASSER modeling of the mutated sortase C protein (D93A/W95A) predicted the presence of a 'semi-open' lid conformation, with the lid displaced away from the triad residues and a cavity observed between the alanine residues and the putative active site triad. The mutated sortase C protein displayed similar physicochemical properties and secondary prediction as compared with the wild-type protein **(Table 1)**, with a distribution of alpha-helices (41%), coils (44%), as well as beta sheets (15%) **(Suppl Figure 2)**. Further, the sortase C protein model with 'semi-open' lid conformation generated a 3D score of 82.37% by the Verify3D Program under SAVES v6.0. The 'semi-open' conformation of the sortase C protein was used for protein-peptide *in silico* molecular docking studies.

Predicted structural models of AMPs using sequences from the DRAMP database

A dataset of 5562 general AMP sequences, including natural and synthetic AMPs, were downloaded from the DRAMP V3.0 database [19]. AMPs were listed from Pep2 to Pep5563 (Pep ID), where Pep1 is designated as the pilin subunit (GenPept ID: WP_170219081.1) **(Suppl Fig 3)** [20]. Using a combination of PEP-FOLD3, TrRosetta, and I-TASSER, we developed structural models of 5544 AMPs. Specifically, PEP-FOLD3 was used for peptide sequences with less than 50 amino acids, TrRosetta for sequences with more than 50 amino acids, and I-TASSER for sequences with undetermined 'X' amino acids. For 18 AMPs possessing unusual or exotic amino acids, models could not

be built as these amino acids were not recognized by the structure prediction tools. Predicted and chosen models were selected based on cluster ranking and scoring. As shown in **Figure 3**, AMPs were predicted to possess a range of structures from alpha helices, beta sheets, coils, or combinations of these. All 5544 predicted and selected peptide structures, searchable by their unique Pep ID, DRAMP ID, name, and course, and a separate subset of AMPs filtered by known anti-Gram positive activity are available in a structural AMP repository for biofilm studies (B-AMP) [44]. For each peptide, the FASTA files, PDB files, and images of the predicted and chosen 3D models are available.

Protein-peptide docking of candidate AMPs against the mutated sortase C protein of *C. striatum*

To identify candidate AMPs with potential to interact with the putative active site of the sortase C enzyme of *C. striatum*, 88 AMPs with a length ranging from 2-8 residues, and 12 representative AMPs each ranging from 9-20 residues in length were selected. Select peptides were subject to *in silico* molecular docking using AutoDock Vina [29], against the 'semi-open' conformation of the mutated sortase C protein. As a comparison, the LPMTG motif of the pilin subunit of *C. striatum* was considered as the standard (where M can be any amino acid). As seen in **Figure 4**, the LPMTG motif interacted with the mutated sortase C protein at the active site, forming three hydrogen bonds with His168, Asn236 and Gln143, with a binding affinity of -5.3 kcal/mol. Of the 100 candidate AMPs, 99 AMPs were predicted to form hydrogen bonds with the mutated sortase C protein (**Figure 4)**. We observed that, for AMPs more than 10 amino acids in length, as the length of the AMP increased the docking score decreased **(Figure 5)**. Further, AMPs more than 14 residues in length did not interact with the putative active site triad residues, but instead masked the active site by binding to protruding residues near the triad. Given this, we decided to analyze specific hydrogen interactions between the 100 candidate AMPs and residues of the mutated sortase C protein **(Figure 5)**. It is important to note that, based on our docking results, no interactions were predicted between the LPMTG motif or any of the AMPs with the active site Cys230 residue. This could be due to the semi-open lid conformation of the sortase C protein that limits access to the Cys230 or could relate to the chemical structure of cysteine that confers it selective interactions. All protein-peptide interaction models are available in B-AMP [44]. For each protein-peptide model, input and output PDBQT files, model images and bond information are provided. It is important to note that AutoDock Vina is typically suited for peptides with less than 32 bonds, and if the number of rotatable bonds exceeds 32 the tool flags select bonds as rigid or non-rotatable to fit parameters. Given this, we decided to work with small AMPs of lengths ranging from 2-8 residues, with a few select AMPs of 9-20 residues in length.

Preference scale of candidate AMPs based on docking interactions with mutated sortase C protein of *C. striatum*

Based on *in silico* molecular docking results, we categorized 100 candidate AMPs on a preference scale of 0-10, using interacting residues and docking scores **(Figure 5).** AMPs were categorized, in order of priority starting with interaction with at least two putative catalytic site residues, as seen with His168 and Arg239, which was given a preference score of 10. This was followed by preference score 9, which consisted of AMPs with multiple explicit interactions with His168 from the triad, as well as interactions with other amino acid residues. Similarly, AMPs with only one explicit interaction with His168, along with interactions with other amino acid residues, were given a preference score of 8. AMPs interacting with Arg239 (along with other amino acid residues) were designated a preference score of 7. In the preference score, exclusive interactions with His168 were given priority over Arg239, as the histidine residue is conserved across sortases, whereas arginine can be replaced with another residue. AMPs with interactions with residues surrounding the putative active site triad, such as ASN236 and GLN143, were categorized into preference scores from 6-3. Finally, interactions with ASP139, THR137, GLY170, ILE235, THR231 were categorized into preference scores from 2 and 1, based on their location further away from the triad. A single AMP showed no hydrogen bond interactions with the sortase C protein, and was given a preference score of 0. Finally, in each preference score, candidate AMPs were ordered based on docking scores, from highest to lowest scores **(Table 2)**.

Biofilm-AMP (B-AMP)

The resources and data generated in this approach are freely available in Biofilm-AMP (B-AMP), a user-friendly, search-enabled repository of AMP structures and AMP interactions with biofilm targets (https://b-amp.karishmakaushiklab.com/). The repository includes predicted structures for 5544 AMPs, searchable by their Pep ID, DRAMP ID, name and source, and a separate subset of 2534 AMPs filtered by known anti-Gram positive activity. For each peptide, the repository includes FASTA files, PDB files, and predicted and chosen 3D models. For the 100 AMPs selected for *in silico* molecular docking, the repository hosts input & output PDBQT files, images, and a text file detailing hydrogen interactions **(Figure 6)**. The Python scripts used for auto-generating FASTA files and filtering out peptides are available in a GitHub repository (https://github.com/KarishmaKaushikLab/B-AMP). B-AMP will be continually updated with new structural AMP models and AMP interaction models with potential biofilm targets.

Conclusions

AMPs are increasingly being recognized for their anti-biofilm potential, to act as alternatives to or in adjunct with conventional antibiotics [45–48]. Their ability to target specific and conserved cellular processes, hold particular promise in the context of broad-spectrum activity and decreased possibility of resistance development [47]. While several AMPs have been studied for their ability to inhibit biofilm formation [49], it is imperative to enable a steady pipeline of candidate AMPs for anti-biofilm evaluation. However, given the vast array of natural and synthetic AMPs, identifying potential candidates for *in vitro* and *in vivo* anti-biofilm testing remains a significant challenge.

Our approach uses open-source bioinformatics tools, to identify AMPs with potential anti-biofilm activity, using the sortase-pilin machinery of *C. striatum* as the target **(Figure 7).** With *in silico* protein-peptide molecular docking, we identified AMPs predicted to interact with putative catalytic site residues of the *C. striatum* sortase C protein, or residues in proximity with the triad. This could result in reduced accessibility of the pilin substrate to the catalytic site or inhibition of catalytic function of the protein [31,50,51]. Based on our approach, we propose a preference scale, based on which candidate AMPs can be taken up, in order of priority, for subsequent evaluation of anti-biofilm activity against *C. striatum*. This can include additional *in silico* approaches such as molecular dynamics simulations and RMSD analyses, in addition to standard *in vitro* and *in vivo* anti-biofilm testing approaches, such as assays for metabolic activity, biofilm eradication and inhibitory concentrations, advanced microscopy, and animal models of infection [51]. Further, evaluations can also focus on bioactivity, stability, and toxicity testing of candidate AMPs. To improve biocompatibility and minimize toxicity, candidate AMPs can also be subject to peptide design and engineering to produce suitable modifications [47,52–54]. While certainly not inclusive of all aspects critical to the evaluation of candidate AMPs as potential anti-biofilm agents, our approach uses open-source bioinformatics tools to narrow down candidate AMPs from a vast library of natural and synthetic AMPs. It thereby provides a starting point for subsequent *in silico*, as well as *in vitro* and *in vivo* biofilm testing approaches against an emerging, multidrug resistant, biofilm-forming pathogen.

While our study focused on the sortase-pilin machinery of *C. striatum*, the approach used and resources developed, which includes a unique structural repository of AMPs and protein-peptide models, can be leveraged for similar studies against other biofilm targets and biofilm-forming pathogens. Further, the vast structural library of AMPs in B-AMP, can be leveraged for a range of AMP-related investigations across a range of bacterial and fungal species.

References

[1] Mahlapuu M, Håkansson J, Ringstad L, Björn C. Antimicrobial Peptides: An Emerging Category of Therapeutic Agents. Front Cell Infect Microbiol 2016;0:194. https://doi.org/10.3389/FCIMB.2016.00194.

[2] Benfield AH, Henriques ST. Mode-of-Action of Antimicrobial Peptides: Membrane Disruption vs. Intracellular Mechanisms. Front Med Technol 2020;0:20. https://doi.org/10.3389/FMEDT.2020.610997.

[3] Souza C de, Faria YV, Sant'Anna L de O, Viana VG, Seabra SH, Souza MC de, et al. Biofilm production by multiresistant Corynebacterium striatumassociated with nosocomial outbreak. Mem Inst Oswaldo Cruz 2015;110:242. https://doi.org/10.1590/0074-02760140373.

[4] Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol 2008;16:33. https://doi.org/10.1016/J.TIM.2007.10.010.

[5] Thappeta KRV, Zhao LN, Nge CE, Crasta S, Leong CY, Ng V, et al. In-Silico Identified New Natural Sortase A Inhibitors Disrupt S. aureus Biofilm Formation. Int J Mol Sci 2020;21:1–18. https://doi.org/10.3390/IJMS21228601.

[6] Lévesque CM, Voronejskaia E, Huang Y-CC, Mair RW, Ellen RP, Cvitkovitch DG. Involvement of Sortase Anchoring of Cell Wall Proteins in Biofilm Formation by Streptococcus mutans. Infect Immun 2005;73:3773. https://doi.org/10.1128/IAI.73.6.3773-3777.2005.

[7] Hu P, Lv B, Yang K, Lu Z, Ma J. Discovery of myricetin as an inhibitor against Streptococcus mutans and an anti-adhesion approach to biofilm formation. Int J Med Microbiol 2021;311:151512. https://doi.org/10.1016/J.IJMM.2021.151512.

[8] Zrelovs N, Kurbatska V, Rudevica Z, Leonchiks A, Fridmanis D. Sorting out the Superbugs: Potential of Sortase A Inhibitors among Other Antimicrobial Strategies to Tackle the Problem of Antibiotic Resistance. Antibiotics 2021;10:1–14. https://doi.org/10.3390/ANTIBIOTICS10020164.

[9] Wang J, Li H, Pan J, Dong J, Zhou X, Niu X, et al. Oligopeptide Targeting Sortase A as Potential Anti-infective Therapy for Staphylococcus aureus. Front Microbiol 2018;9. https://doi.org/10.3389/FMICB.2018.00245.

[10] Guo Y, Cai S, Gu G, Guo Z, Long Z. Recent progress in the development of sortase A inhibitors as novel anti-bacterial virulence agents. RSC Adv 2015;5:49880–9. https://doi.org/10.1039/C5RA07568H.

[11] RCSB PDB: Homepage n.d. https://www.rcsb.org/ (accessed August 13, 2021).

[12] class C sortase [Corynebacterium striatum] - Protein - NCBI n.d. https://www.ncbi.nlm.nih.gov/protein/WP_034656562.1/ (accessed August 13, 2021).

[13] ExPASy - ProtParam tool n.d. https://web.expasy.org/protparam/ (accessed August

13, 2021).

[14] PSIPRED Workbench n.d. http://bioinf.cs.ucl.ac.uk/psipred/ (accessed August 13, 2021).

[15] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010 54 2010;5:725–38. https://doi.org/10.1038/nprot.2010.5.

[16] Bowie J, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science (80-) 1991;253:164–70. https://doi.org/10.1126/SCIENCE.1853201.

[17] C C, TO Y. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 1993;2:1511–9. https://doi.org/10.1002/PRO.5560020916.

[18] Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nat 1992 3566364 1992;356:83–5. https://doi.org/10.1038/356083a0.

[19] Kang X, Dong F, Shi C, Liu S, Sun J, Chen J, et al. DRAMP 2.0, an updated data repository of antimicrobial peptides. Sci Data 2019;6. https://doi.org/10.1038/S41597-019-0154-Y.

[20] SpaH/EbpB family LPXTG-anchored major pilin [Corynebacterium striatum] - Protein - NCBI n.d. https://www.ncbi.nlm.nih.gov/protein/WP_170219081.1 (accessed August 13, 2021).

[21] Lamiable A, Thévenet P, Thévenet T, Rey J, Vavrusa M, Derreumaux P, et al. PEP-FOLD3: faster denovo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 2016;44:449–54. https://doi.org/10.1093/nar/gkw329.

[22] Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci 2020;117:1496–503. https://doi.org/10.1073/PNAS.1914677117.

[23] Robetta - Baker Lab n.d. https://robetta.bakerlab.org/ (accessed August 13, 2021).

[24] I-TASSER server for protein structure and function prediction n.d. https://zhanglab.ccmb.med.umich.edu/I-TASSER/ (accessed August 13, 2021).

[25] N G, MC P. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–23. https://doi.org/10.1002/ELPS.1150181505.

[26] GROMOS – Group for Computer-Aided Chemistry | ETH Zurich n.d. https://www.igc.ethz.ch/gromos.html (accessed August 13, 2021).

[27] Gunsteren W. Biomolecular simulation : the GROMOS96 manual and user guide. Zürich ;;Groningen ;Zürich: Biomos ;;Vdf Hochschulverlag AG an der ETH Zürich; 1996.

[28] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al.

AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91. https://doi.org/10.1002/JCC.21256.

[29] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 2010;31:455. https://doi.org/10.1002/JCC.21334.

[30] Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. Mol Microbiol 2011;82:1044–59. https://doi.org/10.1111/J.1365-2958.2011.07887.X.

[31] Zhang J, Liu H, Zhu K, Gong S, Dramsi S, Wang Y-T, et al. Antiinfective therapy with a small molecule inhibitor of Staphylococcus aureus sortase. Proc Natl Acad Sci 2014;111:13517–22. https://doi.org/10.1073/PNAS.1408601111.

[32] Cossart P, Jonquières R. Sortase, a universal target for therapeutic agents against Gram-positive bacteria? Proc Natl Acad Sci 2000;97:5013–5. https://doi.org/10.1073/PNAS.97.10.5013.

[33] Jacobitz AW, Kattke MD, Wereszczynski J, Clubb RT. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. Adv Protein Chem Struct Biol 2017;109:223. https://doi.org/10.1016/BS.APCSB.2017.04.008.

[34] Guiton PS, Hung CS, Kline KA, Roth R, Kau AL, Hayes E, et al. Contribution of Autolysin and Sortase A during Enterococcus faecalis DNA-Dependent Biofilm Development. Infect Immun 2009;77:3626. https://doi.org/10.1128/IAI.00219-09.

[35] Kemp KD, Singh K V., Nallapareddy SR, Murray BE. Relative contributions of Enterococcus faecalis OG1RF sortase-encoding genes, srtA and bps (srtC), to biofilm formation and a murine model of urinary tract infection. Infect Immun 2007;75:5399–404. https://doi.org/10.1128/IAI.00663-07.

[36] Kumari P, Nath Y, Murty US, Ravichandiran V, Mohan U. Sortase A Mediated Bioconjugation of Common Epitopes Decreases Biofilm Formation in Staphylococcus aureus. Front Microbiol 2020;0:1702. https://doi.org/10.3389/FMICB.2020.01702.

[37] Cascioferro S, Totsika M, Schillaci D. Sortase A: an ideal target for anti-virulence drug development. Microb Pathog 2014;77:105–12. https://doi.org/10.1016/J.MICPATH.2014.10.007.

[38] Nudel K, Zhao X, Basu S, Dong X, Hoffmann M, Feldgarden M, et al. Genomics of Corynebacterium striatum, an emerging multidrug-resistant pathogen of immunocompromised patients. Clin Microbiol Infect 2018;24:1016.e7-1016.e13. https://doi.org/10.1016/J.CMI.2017.12.024.

[39] Ramos JN, Rodrigues I dos S, Baio PVP, Veras JFC, Ramos RTJ, Pacheco LG, et al. Genome sequence of a multidrug-resistant Corynebacterium striatum isolated from bloodstream infection from a nosocomial outbreak in Rio de Janeiro, Brazil. Mem Inst Oswaldo Cruz 2018;113:e180051. https://doi.org/10.1590/0074-02760180051.

[40] Xiang Z. Advances in homology protein structure modeling. Curr Protein Pept Sci

2006;7:217–27. https://doi.org/10.2174/138920306777452312.

[41] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12. https://doi.org/10.1002/JCC.20084.

[42] Khare B, Narayana SVL. Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. Protein Sci 2017;26:1458. https://doi.org/10.1002/PRO.3191.

[43] Jacobitz AW, Naziga EB, Yi SW, McConnell SA, Peterson R, Jung ME, et al. The "Lid" in the Streptococcus Pneumoniae SrtC1 Sortase Adopts a Rigid Structure that Regulates Substrate Access to the Active Site. J Phys Chem B 2016;120:8302. https://doi.org/10.1021/ACS.JPCB.6B01930.

[44] Antimicrobial Peptide Repository for Biofilms | B-AMP n.d. https://b-amp.karishmakaushiklab.com/ (accessed August 13, 2021).

[45] Zharkova MS, Orlov DS, Golubeva OY, Chakchir OB, Eliseev IE, Grinchuk TM, et al. Application of Antimicrobial Peptides of the Innate Immune System in Combination With Conventional Antibiotics—A Novel Way to Combat Antibiotic Resistance? Front Cell Infect Microbiol 2019;0:128. https://doi.org/10.3389/FCIMB.2019.00128.

[46] Chung PY, Khanum R. Antimicrobial peptides as potential anti-biofilm agents against multidrug-resistant bacteria. J Microbiol Immunol Infect 2017;50:405–10. https://doi.org/10.1016/J.JMII.2016.12.005.

[47] Dostert M, Belanger CR, Hancock REW. Design and Assessment of Anti-Biofilm Peptides: Steps Toward Clinical Application. J Innate Immun 2019;11:193–204. https://doi.org/10.1159/000491497.

[48] Lei J, Sun L, Huang S, Zhu C, Li P, He J, et al. The antimicrobial peptides and their potential clinical applications. Am J Transl Res 2019;11:3919.

[49] Yasir M, Mark Duncan Perry Willcox, Dutta D. Action of Antimicrobial Peptides against Bacterial Biofilms. Mater (Basel, Switzerland) 2018;11. https://doi.org/10.3390/MA11122468.

[50] Rebollo IR, McCallin S, Bertoldo D, Entenza JM, Moreillon P, Heinis C. Development of Potent and Selective S. aureus Sortase A Inhibitors Based on Peptide Macrocycles. ACS Med Chem Lett 2016;7:606–11. https://doi.org/10.1021/ACSMEDCHEMLETT.6B00045.

[51] Wang J, Li H, Pan J, Dong J, Zhou X, Niu X, et al. Oligopeptide Targeting Sortase A as Potential Anti-infective Therapy for Staphylococcus aureus. Front Microbiol 2018;0:245. https://doi.org/10.3389/FMICB.2018.00245.

[52] Sharma A, Gupta P, Kumar R, Bhardwaj A. dPABBs: A Novel in silico Approach for Predicting and Designing Anti-biofilm Peptides. Sci Reports 2016 61 2016;6:1–13. https://doi.org/10.1038/srep21839.

[53] dPABBs: A webserver for Designing of Peptides Against Bacterial Biofilms n.d. https://ab-openlab.csir.res.in/abp/antibiofilm/ (accessed August 13, 2021).

[54] Luca M Di, Maccari G, Maisetta G, Batoni G. BaAMPs: the database of biofilm-active antimicrobial peptides. Biofouling 2015;31:193–9. https://doi.org/10.1080/08927014.2015.1021340.

Figures and Figure Legends

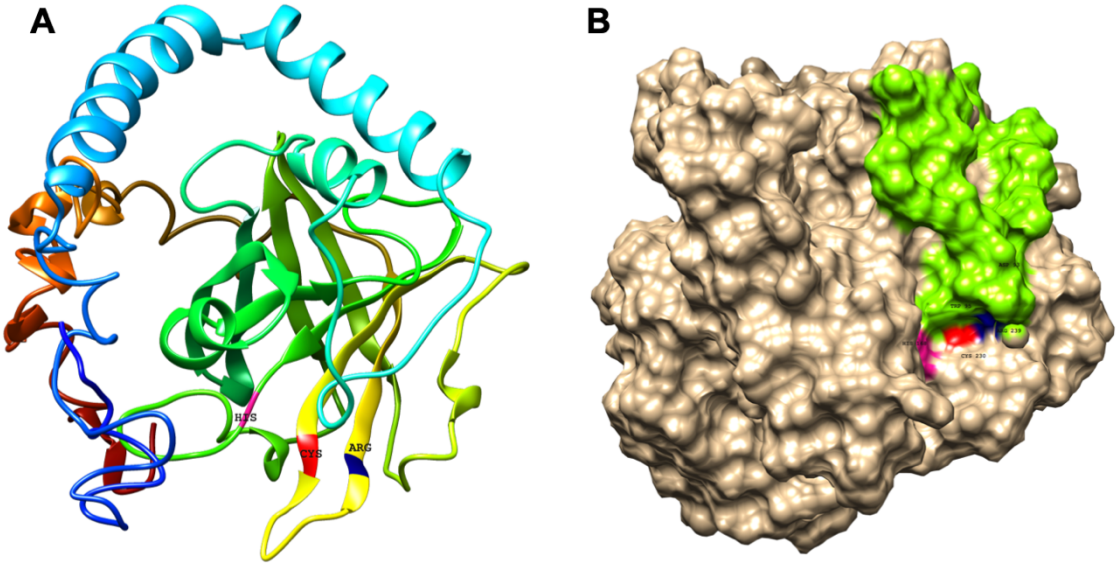


Figure 1: The predicted structure of the *C. striatum* sortase C protein reveals the presence of the putative active site triad, essential for catalytic activity, and a 'lid'. (A) The sortase C protein in ribbon form, showing the overall organization of the protein backbone in 3D space. The modeled structure reveals the presence of the putative Cys-His-Arg triad essential for catalytic activity. The Cys230 residue is shown in red, the His168 residue in green, and the Arg239 residue in blue. The default ribbon style is smooth (rounded). (B) Surface representation of the class C sortase protein showing the sortase 'lid' (green) occluding the active site residues. Homology modeling was done using the I-TASSER server using the amino acid sequence of the sortase C protein retrieved from GenPept (ID: WP_034656562).

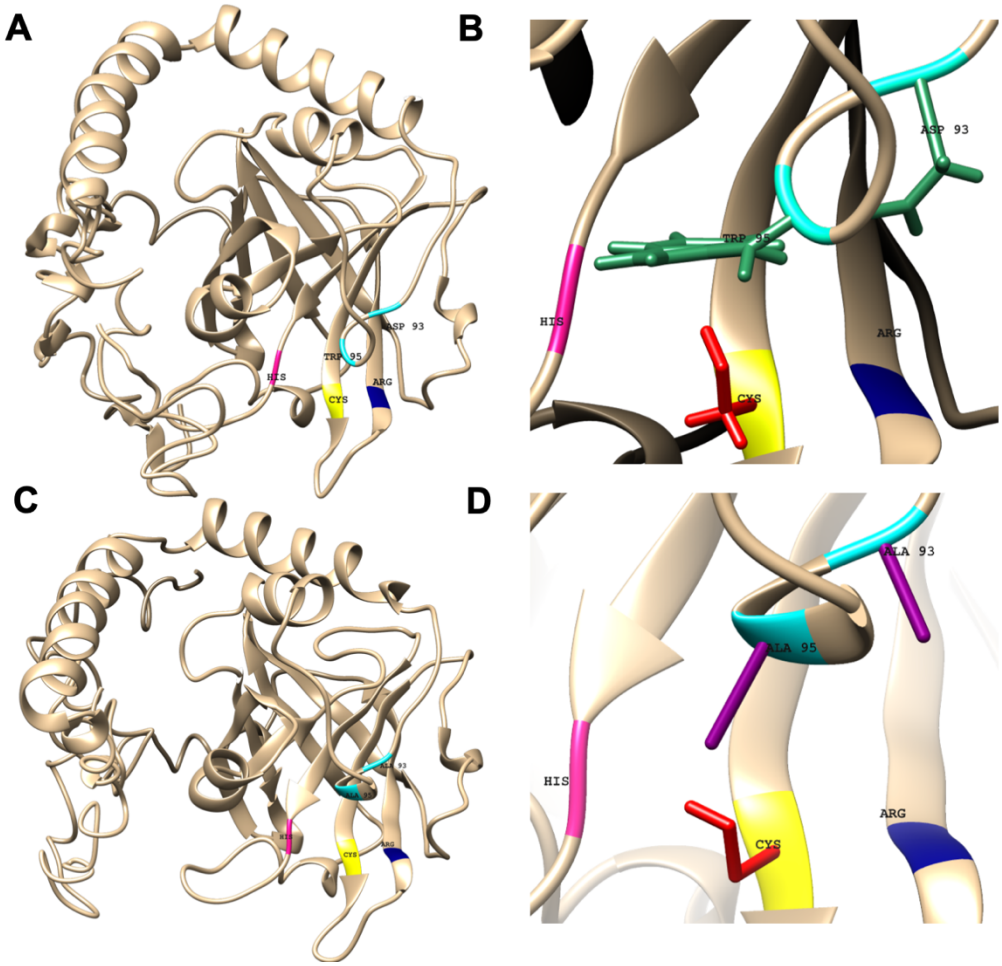


Figure 2: Predicted and chosen models of the wild-type *C. striatum* sortase C protein (closed conformation) and mutant protein (D93A/W95A; semi-open conformation) showing the relationship between the sortase 'lid' and putative active site triad residues. (A) In the wild-type sortase C protein, the sortase 'lid' is predicted as a loop-like extension, in close proximity with the putative active site Cys-His-Arg (yellow-pink-blue) residues (B) The closer version of the region, predicts interactions between the aromatic ring of the Trp residue (green) and the Asp residue (green) of the 'lid' with Cys230 and Arg239 residues respectively (C) In the mutant sortase C protein (D93A/W95A), the lid is predicted to be displaced from the triad residues (D) In the closer version, a cavity is predicted between the mutated alanine residues (purple) and the putative active site triad. This 'semi-open' lid conformation of the mutated sortase C protein was used for protein-peptide *in silico* molecular docking studies.

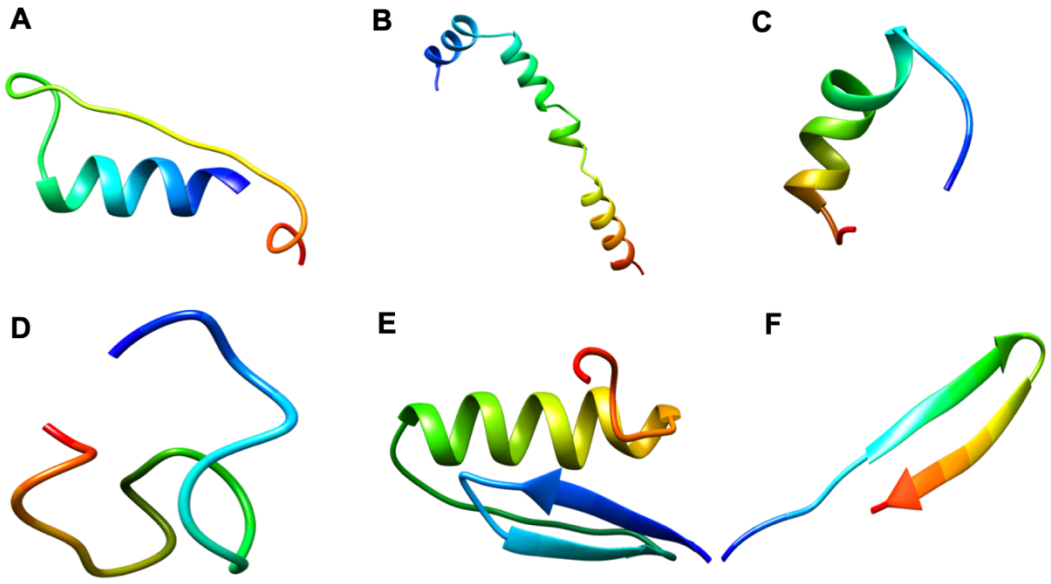


Figure 3: Representative predicted AMP structures with PEPFOLD3, TrRosetta and I-TASSER. FASTA sequences of AMPs from the DRAMP database were modeled using the PEPFOLD3 (for AMPs less than 50 amino acid residues in length), TrRosetta (for AMPs more than 50 amino acid residues in length) and I-TASSER (for AMPs with unusual or unknown amino acids). Predicted AMP structures were observed to possess a range of features such as alpha helices, beta sheets, coils or combinations of these. (A) Pep2 (DRAMP00005) showing a helix-coil structure (B) Pep6 (DRAMP00068) with amino acid sequence showing an alpha-helix structure C) Pep31 (DRAMP00244) showing an alpha-helix structure (D) Pep3 (DRAMP00017) showing a coil structure (E) Pep7 (DRAMP00069) showing an antiparallel beta sheet structure (F) Pep5 (DRAMP00063) showing a combination of alpha helices, beta sheets and coils. The entire structural library of 5544 AMPs is available in B-AMP. Models are colored from blue to red from N-to-C-terminal.

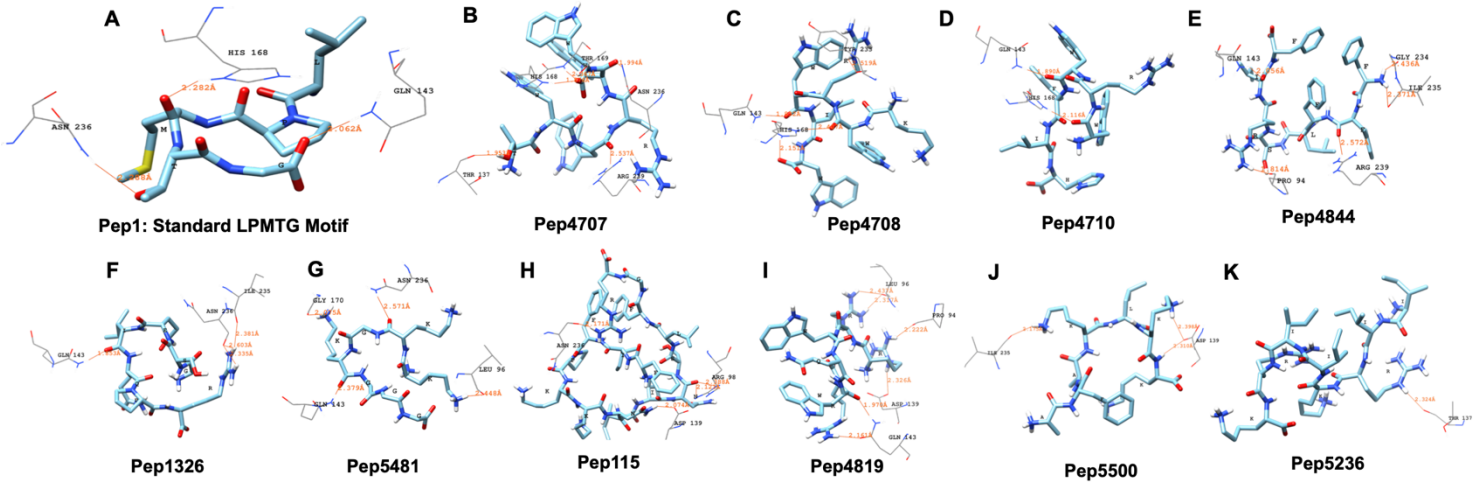


Figure 4: Representative results from *in silico* protein-peptide molecular docking for the standard pilin subunit LPMTG, and AMPs showing a range of hydrogen interactions, with the *C. striatum* mutated sortase C protein. (A) The standard LPMTG motif is predicted to form hydrogen bonds with His168, Asn23y, Gln143 (B) Pep4707 is predicted to interact with HIS168, ARG239, ASN236, THR137, THR169 (C) Pep4708 is predicted to interact with HIS168, HIS168, GLN143, TYR233 (D) Pep4710 is predicted to interact with HIS168, GLN143 (E) Pep4844 is predicted to interact with ARG239, GLN143, PRO94, ILE235, GLY234 (F) Pep1326 is predicted to interact with ASN236, ASN236, GLN143, ILE235 (G) Pep5481 is predicted to interact with ASN236, GLN143, LEU96, GLY170 (H) Pep115 is predicted to interact with ASN236, ASP139, ARG98, ARG98 (I) Pep4819 is predicted to interact with GLN143, GLN143, ASP139, LEU96, LEU96, PRO94 (J) Pep5500 is predicted to interact with ASP139, ASP139, ILE235 (K) Pep5236 is predicted to interact with THR137.

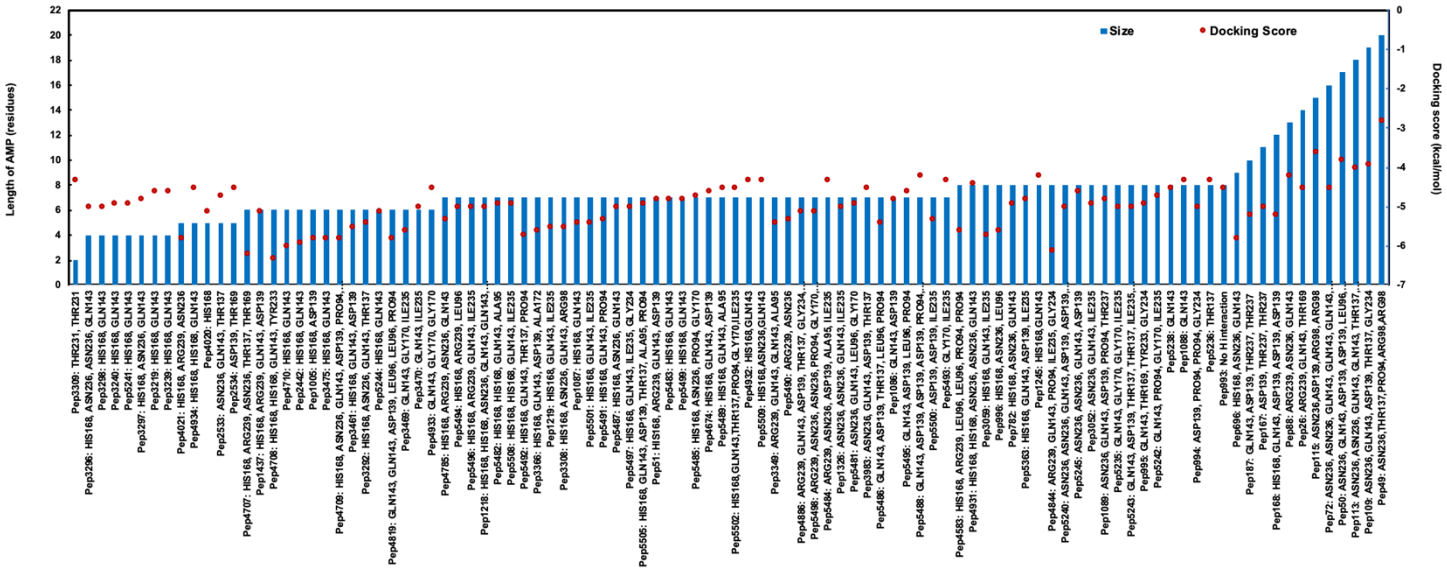


Figure 5: *In silico* molecular docking scores and interacting residues for 100 candidate AMPs and the semi-open (mutated) conformation of the *C. striatum* sortase C protein. AMPs were predicted to interact with the putative active site triad residues, protruding residues near the triad as well as residues further away from the triad. While docking scores were variable, for AMPs more than 10 amino acids in length, as the length of the AMP increased the docking score was observed to decrease.

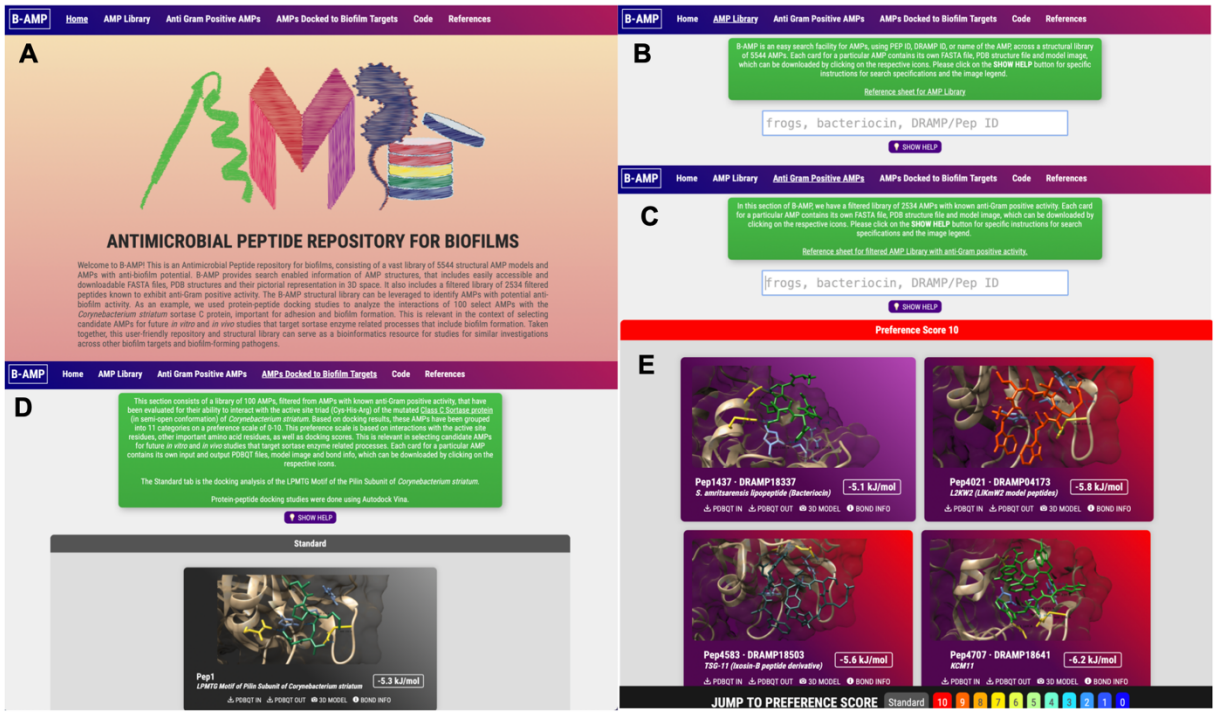


Figure 6: Features of Biofilm-AMP (B-AMP). (A) B-AMP is a structural repository of AMPs for biofilm studies (B) This user-friendly, search enabled repository includes predicted structures for 5544 AMPs, searchable by their Pep ID, DRAMP ID, name and source (C) It also includes a separate subset of 2534 AMPs filtered by known anti-Gram positive activity. For each peptide, the repository includes FASTA files, PDB files, and predicted and chosen 3D models (D and E) For the 100 AMPs selected for *in silico* molecular docking, the repository hosts protein-peptide docking models, input & output PDBQT files, images, and AMPs categorized in order of preference, based in docking scores and interacting residues. B-AMP will be continually updated with new structural AMP models and AMP interaction models with potential biofilm targets.

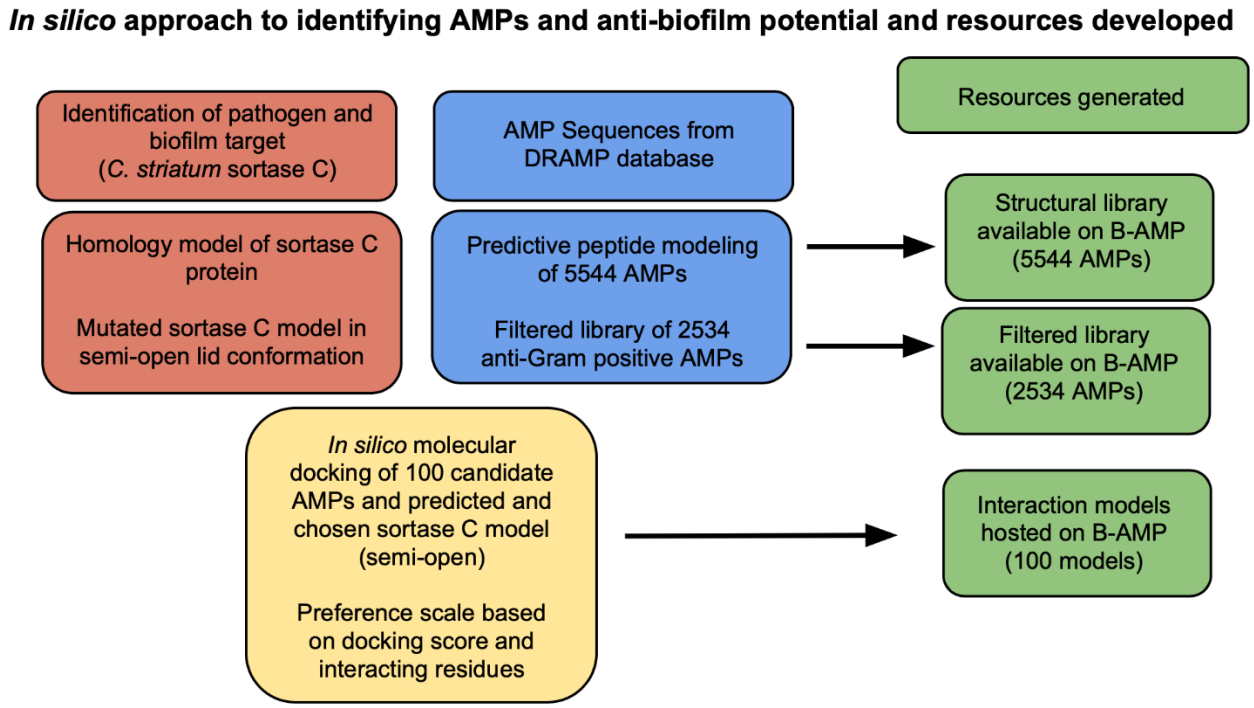
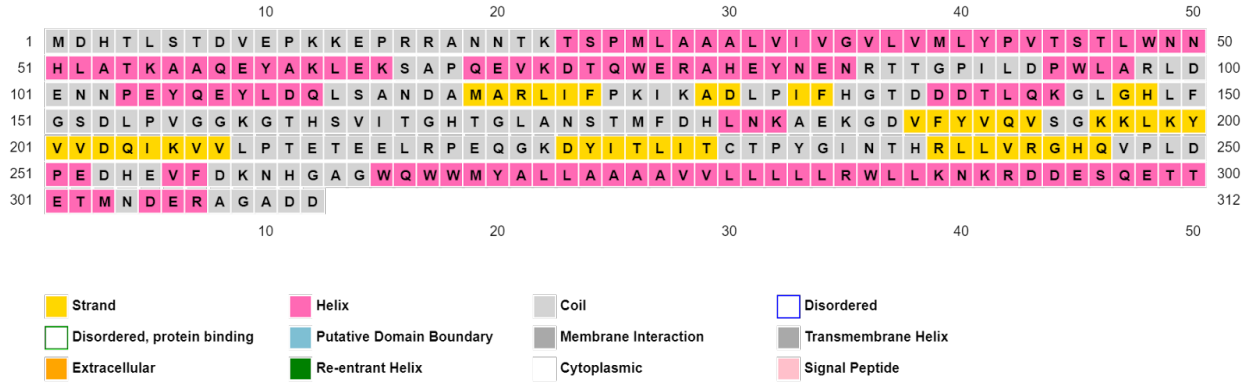
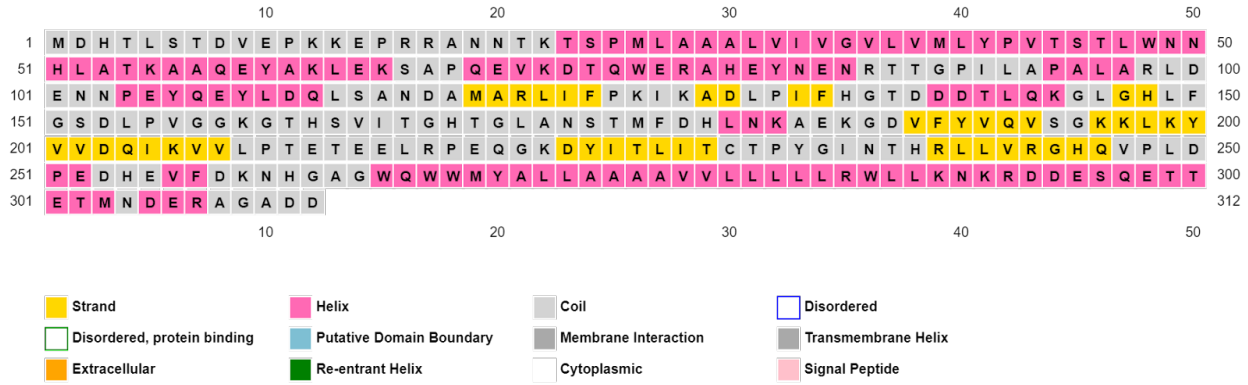


Figure 7: *In silico* approach to identifying AMPs with anti-biofilm potential and resources developed in this study. Our approach used a combination of homology modeling, predictive peptide modeling and protein-peptide molecular docking. The resources and data generated in this approach are freely available in Biofilm-AMP (B-AMP), a user-friendly, search-enabled repository of AMP structures and AMP interactions with biofilm targets. The approach used and resources developed can be leveraged for similar studies against other biofilm targets and biofilm-forming pathogens.

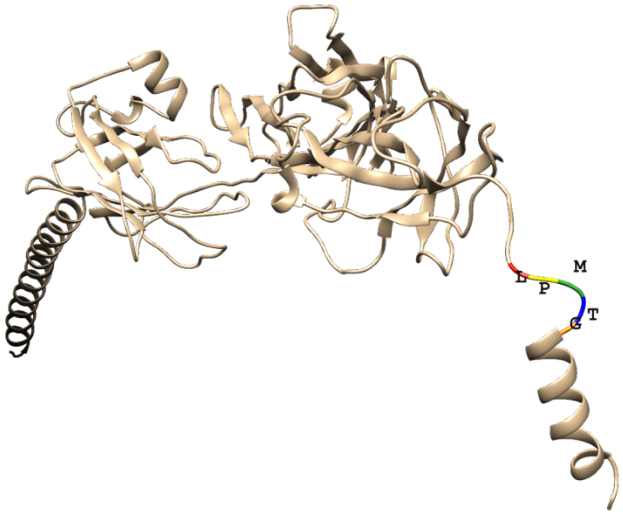
Supplementary figures



Suppl Fig 1: Secondary structure of the sortase C protein based on PSIPRED analysis. The secondary structure of the sortase C protein contains a combination of alpha-helices (41%), coils (44%), and beta strands (15%). The cysteine (Cys230) and histidine (His168) residues are found to be in between two beta-strands, constituting a turn connecting the strands and acting as a major groove for hydrophobic interaction. The arginine (Arg239) residue was predicted to initiate the second beta-strand of the hydrophobic groove. The image is a snapshot of the results obtained from the PSIPRED analysis program.



Suppl Fig 2: Secondary structure of the mutated sortase C protein based on PSIPRED analysis. The secondary structure of the mutated sortase C protein with Aspartate93 and Tryptophan95 replaced with Alanine (D93A/W95A). The image is a snapshot of the results obtained from the PSIPRED analysis program.



Suppl Fig 3: Predicted and chosen model of the *C. striatum* pilin subunit showing the presence of the cell sorting LPMTG motif. The *C. striatum* pilin subunit in ribbon form, showing the overall organization of the protein backbone in 3D space. The modeled structure rivals the presence of the cell sorting LPXTG motif (X is represented by M) at the C terminus of the protein. This motif is followed by a hydrophobic domain seen here as a helix. Homology modeling was done using the I-TASSER Server using the amino acid sequence of the pilin subunit retrieved from GenPept (ID: WP_170219081.1).

Funding:

Karishma S Kaushik's academic appointment is funded by the Ramalingaswami Re-entry Fellowship (BT/HRD/35/02/2006). Snehal Kadam was supported for a select duration of this project (June 2020-December 2020) on the Ramalingaswami Re-entry Fellowship (to KSK).

Acknowledgments: We thank Dr. Ragothaman Yenamalli, SASTRA University, Thanjavur, India, for inputs related to *in silico* molecular docking.

Author Contributions:

Shreeya Mhade: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Visualization, Writing the original draft, Editing draft. **Stutee Panse:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Visualization, Writing the original draft, Editing draft. **Gandhar Tendulkar:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Visualization, Writing the original draft, Editing draft. **Rohit Awate:** Methodology, Investigation. **Snehal Kadam:** Conceptualization, Investigation, Editing draft. **Karishma S Kaushik:** Conceptualization, Project administration, Supervision, Writing the original draft, Editing draft.

Conflict of Interest: None

Table 1
Physicochemical properties of the *C. striatum* sortase C protein (wild-type and double mutant D93A/W95A) using ProtParam analysis

Physicochemical Property		**Value**		
		Sortase C	**Mutant Sortase C (D93A/W95A)**	
Number of amino acids		312	312	
Molecular weight		35134.72	34975.57	
Theoretical pI (isoelectric point)		5.38	5.46	
Amino acid Composition				
Amino acid name	*Number of residues*	*Percentage*	*Number of residues*	*Percentage*
Ala (A)	25	8.0%	27	8.7%
Arg (R)	12	3.8%	12	3.8%
Asn (N)	15	4.8%	15	4.8%
Asp (D)	24	7.7%	23	7.4%
Cys (C)	1	0.3%	1	0.3%
Gln (Q)	12	3.8%	12	3.8%
Glu (E)	22	7.1%	22	7.1%
Gly (G)	19	6.1%	19	6.1%
His (H)	12	3.8%	12	3.8%
Ile (I)	10	3.2%	10	3.2%
Leu (L)	37	11.9%	37	11.9%
Lys (K)	21	6.7%	21	6.7%
Met (M)	7	2.2%	7	2.2%
Phe (F)	6	1.9%	6	1.9%
Pro (P)	16	5.1%	16	5.1%
Ser (S)	10	3.2%	10	3.2%
Thr (T)	25	8.0%	25	8.0%
Trp (W)	7		6	1.9%

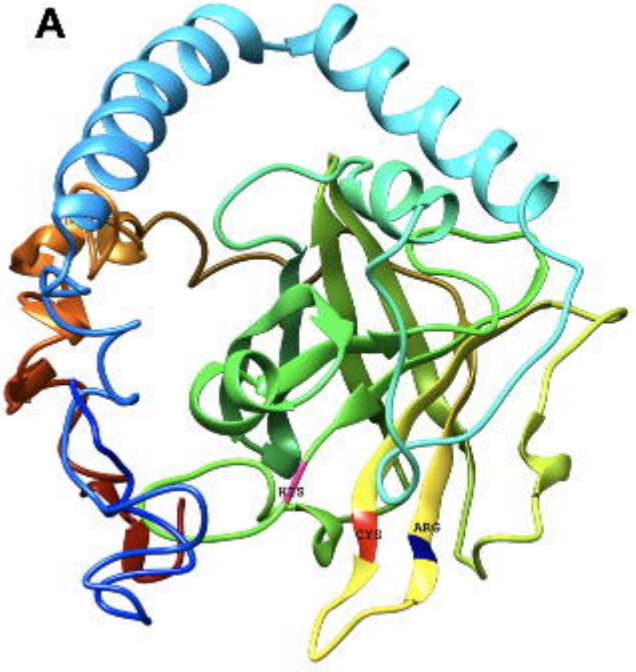
Tyr (Y) Val (V) Pyl (O) Sec (U)	10 21 0 0	2.2% 2.2% 6.7% 0.0% 0.0%	10 21 0 0	2.2% 6.7% 0.0% 0.0%
Total number of negatively charged residues (Asp + Glu):		46	45	
Total number of positively charged residues (Arg + Lys):		33	33	
Atomic composition				
Carbon	C	1565	1556	
Hydrogen	H	2449	2444	
Nitrogen	N	427	426	
Oxygen	O	477	475	
Sulfur	S	8	8	

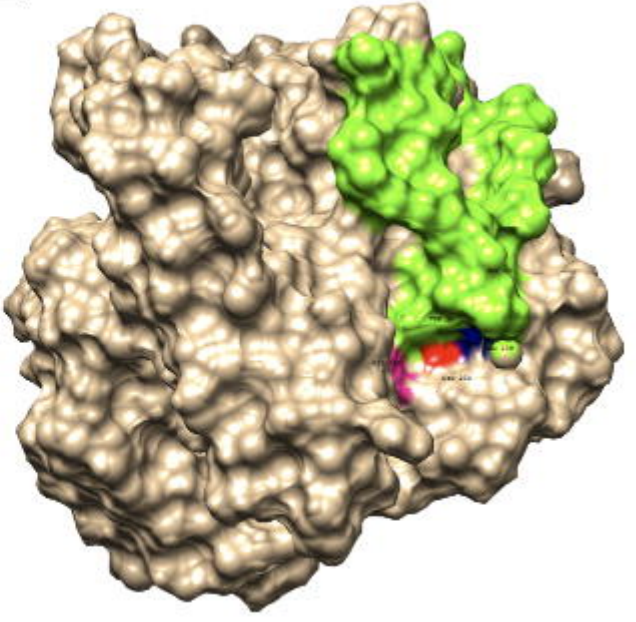
Table 2: Preference scale of candidate AMPs based on docking interactions with mutated sortase C protein of *C. striatum*
(where 10 represents the most optimum docking interactions and 0 represents no interactions with the protein residues)

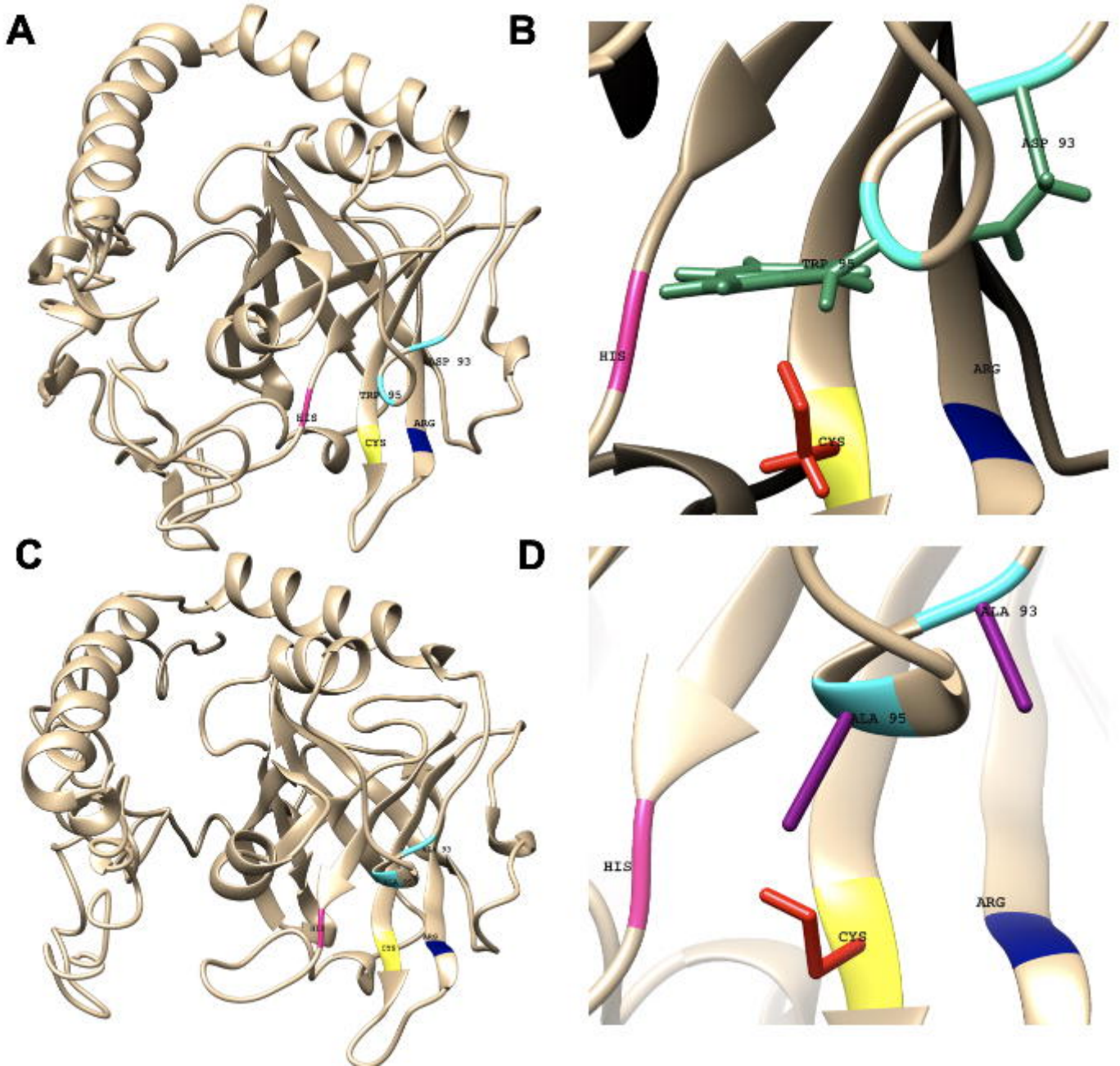
Serial Number	Pep ID	DRAMP ID	AMP length (residues)	Sequence	Name of AMP	TORSDOF	Docking Score (kcal/mol)	Interacting Residues	Essential criteria	Preference Score
1	1	--	5	LPMTG	LPMTG Motif of the Pilin Subunit	14	-5.3	HIS168, ASN236, GLN143		STANDARD
2	4707	18641	6	TWWRWW	KCM11	26	-6.2	HIS168, ARG239, ASN236, THR137, THR169	**Interactions with two putative catalytic site residues, seen as His168 and Arg239**	**10**
3	4021	04173	5	WLLKW	L2KW2	23	-5.8	HIS168, ARG239, ASN236		**10**
4	4583	18503	8	WKSYVRRW	TSG-11 (Ixosin-B peptide derivative)	39	-5.6	HIS168, ARG239, LEU96, LEU96, PRO94, PRO94		**10**
5	4785	18719	7	DDDDDDD	SAAP fraction 3 (Surfactant-associated anionic peptide)	28	-5.3	HIS168, ARG239, ASN236, GLN143		**10**
6	1437	18337	6	ATQSHQ	S. amritsarensis lipopeptide (Bacteriocin)	24	-5.1	HIS168, ARG239, GLN143, ASP139		**10**
7	5494	21437	7	KKLKAFA	Peptide 14 (Derived from B1)	33	-5	HIS168, ARG239, LEU96		**10**
8	5496	21439	7	KKKIAFA	Peptide 16 (Derived from B1)	33	-5	HIS168, ARG239, GLN143, ILE235		**10**
9	4708	18642	6	KWRWIW	KCM12	29	-6.3	HIS168, HIS168, GLN143, TYR233	**Multiple explicit interactions with His168 from the putative triad**	**9**
10	1218	04174	7	KWLKKWL	L2K3W2	37	-5	HIS168, HIS168, ASN236, GLN143, GLN143, THR137, PRO94		**9**
11	5482	21425	7	KKGKGGG	Peptide 2 (Derived from B1)	29	-4.9	HIS168, HIS168, GLN143, ALA95		**9**
12	5508	21451	7	KKKLAYA	Peptide 28 (Derived from B1)	34	-4.9	HIS168, HIS168, GLN143, ILE235		**9**
13	4934	20873	5	GLLKR	Pal-ano-5 (Pal-anoplin peptide derivative)	23	-4.5	HIS168, HIS168, GLN143		**9**
14	4931	20870	8	GLLKRIKT	Pal-ano-8 (Pal-anoplin peptide derivative)	38	-4.4	HIS168, HIS168, ASN236, ASN236, GLN143		**9**
15	4710	18644	6	WRWFIH	KRS22	26	-6	HIS168, GLN143	**Only one explicit interaction with His168 from the putative triad**	**8**
16	2442	18309	6	WAIVLL	Baceridin (Bacteriocin)	21	-5.9	HIS168, GLN143		**8**
17	1005	03881	6	RRWWWR	LFB-6RW	30	-5.8	HIS168, ASP139		**8**
18	3475	18218	6	TVVTNA	Fusaricidin A (Bacteriocin)	20	-5.8	HIS168, GLN143		**8**
19	4709	18643	6	KWWWRW	KCM21	29	-5.8	HIS168, ASN236, GLN143, ASP139, PRO94, PRO94		**8**
20	696	18371	9	VGVGGGFGR	Crinicepsin-1	26	-5.8	HIS168, ASN236, GLN143		**8**
21	5492	21435	7	AFALKKK	Peptide 12 (Derived from B1)	33	-5.7	HIS168, GLN143, THR137, PRO94		**8**
22	3059	18247	8	DWTXWSXL	Bacthuricin F4(Bacteriocin)	28	-5.7	HIS168, GLN143, ILE235		**8**
23	3366	18226	7	LLDVLLE	Gageostatin A (Bacteriocin)	28	-5.6	HIS168, GLN143, ASP139, ALA172		**8**
24	996	03869	8	RRWVIWRR	Bac8d (Bac2A variant)	39	-5.6	HIS168, ASN236, LEU96		**8**
25	3461	18221	6	TVYTQA	Fusaricidin D (Bacteriocin)	23	-5.5	HIS168, GLN143, ASP139		**8**
26	1219	04175	7	KWLLKWL	L3K2W2 (LIKmW2 model peptides)	34	-5.5	HIS168, GLN143, ILE235		**8**
27	3308	18228	7	LLDVLLE	Gageostatin C (Bacteriocin)	28	-5.5	HIS168, ASN236, GLN143, ARG98		**8**
28	3292	02930	6	HSPGGA	Antimicrobial protein 2	15	-5.4	HIS168, ASN236, GLN143, THR137		**8**
29	1087	03989	7	LLKWLLK	L4K2W4	34	-5.4	HIS168, GLN143		**8**
30	5501	21444	7	AFALKKK	Peptide 21 (Derived from B1)	33	-5.4	HIS168, GLN143, ILE235		**8**
31	5491	21434	7	AFAKLKK	Peptide 11 (Derived from B1)	33	-5.3	HIS168, GLN143, PRO94		**8**
32	168	01351	12	RVCFAIPLPICH	Tigerinin-2	41	-5.2	HIS168, GLN143, ASP139, ASP139, ASP139		**8**
33	4020	04172	5	WLKKW	LK2W2 (LIKmW2 model peptides)	26	-5.1	HIS168		**8**
34	5244	21214	6	IRIKIR	IK6-all D (Derived from IK8-all L)	31	-5.1	HIS168, GLN143		**8**
35	3296	18230	4	LLLE	Gageotetrin B (Bacteriocin)	17	-5	HIS168, ASN236, ASN236, GLN143		**8**
36	3298	18232	4	LLEL	Gageopeptide A(Bacteriocin)	17	-5	HIS168, GLN143		**8**
37	5487	21430	7	KKKLAFA	Peptide 7 (Derived from B1)	33	-5	HIS168, ASN236, GLN143		**8**
38	5497	21440	7	KKLKAFA	Peptide 17 (Derived from B1)	33	-5	HIS168, GLN143, ILE235, GLY234		**8**
39	3240	18234	4	LLEL	Gageopeptide C(Bacteriocin)	17	-4.9	HIS168, GLN143		**8**
40	5241	21215	4	IRIK	IK4-all D (Derived from IK8-all L)	21	-4.9	HIS168, GLN143		**8**
41	5505	21448	7	KKKLAYA	Peptide 25 (Derived from B1)	34	-4.9	HIS168, GLN143, ASP139, THR137, ALA95, PRO94		**8**
42	782	02999	8	PFKISIHL	Jellein-1	32	-4.9	HIS168, ASN236, GLN143		**8**
43	3297	18231	4	LLLE	Gageotetrin C (Bacteriocin)	17	-4.8	HIS168, ASN236, GLN143		**8**

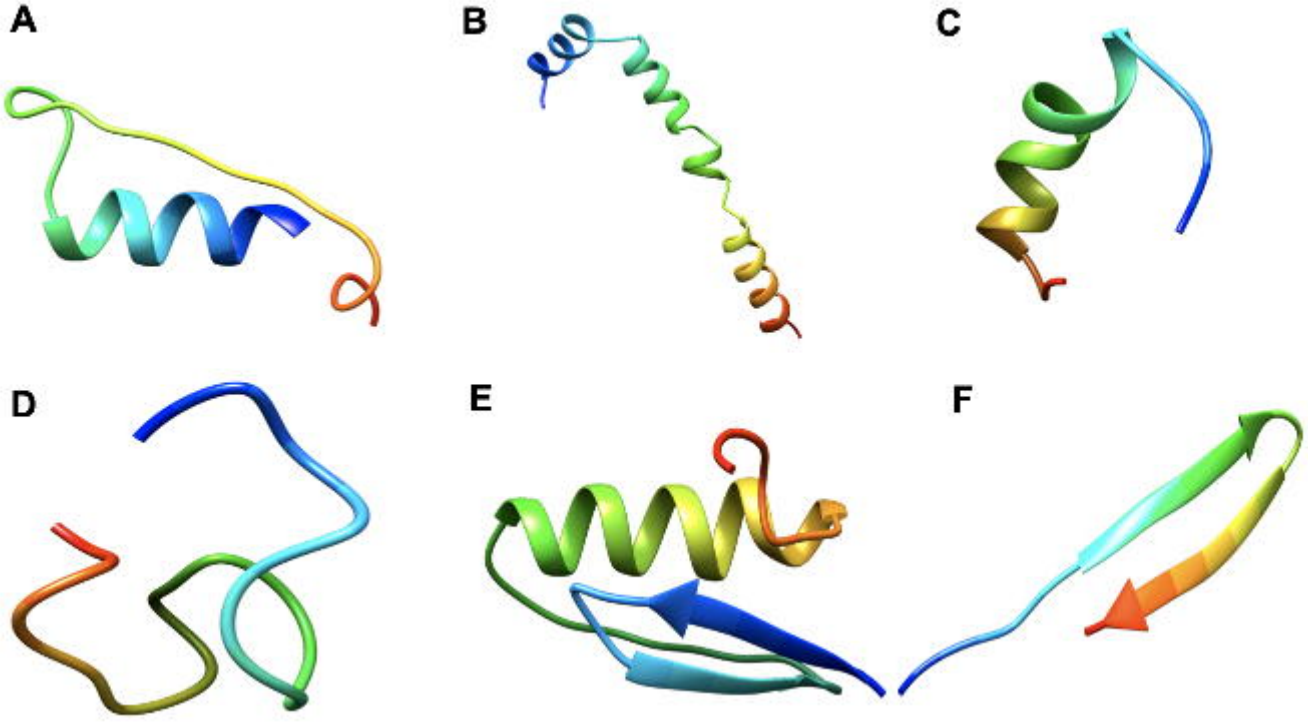
44	51	00766	7	KVFLGLK	JCpep7 (Plants)	31	-4.8	HIS168, ARG239, GLN143, ASP139		**8**
45	5483	21426	7	KKKGGGG	Peptide 3 (Derived from B1)	29	-4.8	HIS168, GLN143		**8**
46	5499	21442	7	AFALKKK	Peptide 19 (Derived from B1)	33	-4.8	HIS168, GLN143		**8**
47	5363	21305	8	RRRRRRRR	R8 (De novo synthesis)	48	-4.8	HIS168, GLN143, ASP139, ILE235		**8**
48	5485	21428	7	GGGKGKK	Peptide 5 (Derived from B1)	29	-4.7	HIS168, ASN236, PRO94, GLY170		**8**
49	3219	18235	4	LLEL	Gageopeptide D(Bacteriocin)	17	-4.6	HIS168, GLN143		**8**
50	3239	18233	4	LLLE	Gageopeptide B(Bacteriocin)	17	-4.6	HIS168, GLN143		**8**
51	4674	18612	7	KIIKVVK	LL-III/10	35	-4.6	HIS168, GLN143, ASP139		**8**
52	5489	21432	7	KKKLAFA	Peptide 9 (Derived from B1)	33	-4.5	HIS168, GLN143, ALA95		**8**
53	5502	21445	7	KKKLAYA	Peptide 22 (Derived from B1)	34	-4.5	HIS168, GLN143, THR137, PRO94, GLY170, ILE235		**8**
54	4932	20871	7	GLLKRIK	Pal-ano-7 (Pal-anoplin peptide derivative)	34	-4.3	HIS168, GLN143		**8**
55	5509	21452	7	KKKLAFA	Peptide 29 (Derived from B1)	33	-4.3	HIS168, ASN236, GLN143		**8**
56	1245	04240	8	KLKLLLLL	Synthetic 1	38	-4.2	HIS168, GLN143		**8**
57	4844	20778	8	FFFLSRIF	Temporin-SHf	34	-6.1	ARG239, GLN143, PRO94, ILE235, GLY234	**Interaction with Arg239 of the putative triad**	**7**
58	3349	18227	7	LLDVLLE	Gageostatin B (Bacteriocin)	28	-5.4	ARG239, GLN143, GLN143, ALA95		**7**
59	5490	21433	7	AFAKLKK	Peptide 10 (Derived from B1)	33	-5.3	ARG239, ASN236		**7**
60	4886	20820	7	FRIRVRV	FV7	32	-5.1	ARG239, GLN143, ASP139, THR137, GLY234, GLY234		**7**
61	5498	21441	7	AFAKLKK	Peptide 18 (Derived from B1)	33	-5.1	ARG239, ASN236, ASN236, PRO94, GLY170, TYR233		**7**
62	26	00201	14	SCNCVCGVCCSCSP	Amythiamicin A/B (Bacteriocin)	49	-4.5	ARG239, GLN143, THR169		**7**
63	5484	21427	7	GGGGKKK	Peptide 4 (Derived from B1)	29	-4.3	ARG239, ASN236, ASP139, ALA95, ILE235		**7**
64	88	01827	13	VLPLISMALGKLL	Temporin-1La	48	-4.2	ARG239, ASN236, GLN143		**7**
65	1326	00031	7	GSEIQPR	Lantibiotic carnocin-UI49 (Bacteriocin)	27	-5	ASN236, ASN236, GLN143, ILE235	**Multiple explicit interactions with Asn236 near the putative triad**	**6**
66	5240	21217	8	IRIKIRIK	IK8-2D (Derived from IK8-all L)	42	-5	ASN236, ASN236, GLN143, ASP139, ASP139, ARG98		**6**
67	5245	21212	8	IRIKIRIK	IK8-all L (De novo synthesis)	42	-4.6	ASN236, ASN236, GLN143, ASP139		**6**
68	72	01082	16	ILGKLLSTAAGLLSNL	Alyteserin-2a	59	-4.5	ASN236, ASN236, GLN143, GLN143, GLN143, ASP140		**6**
69	113	01192	18	FLPKLFAKITKKNMAHIR	Andersonin-Y1	80	-4	ASN236, ASN236, GLN143, GLN143, THR137, LEU96, ARG98, THR237		**6**
70	5481	21424	7	KKGKGGG	B1 (De Novo Synthesis)	29	-4.9	ASN236, GLN143, LEU96, GLY170	**Interactions with Asn236 and Gln143 near the putative triad**	**5**
71	3052	18248	8	KYGDVPLY	Bifidin I (Bacteriocin)	31	-4.9	ASN236, GLN143, ILE235		**5**
72	1089	03991	8	KLLKWLLK	L4K3W5 (LlKmWn model peptide)	41	-4.8	ASN236, GLN143, ASP139, PRO94, THR237		**5**
73	2533	04395	5	AMVGT	EP3	16	-4.7	ASN236, GLN143, THR137		**5**
74	3983	04037	7	LKLLKKL	Immobilized peptide E07LKK	37	-4.5	ASN236, GLN143, ASP139, THR137		**5**
75	109	01188	19	LALKSGGWLRLFGLKDKKH	Chensinin-1ZHa	82	-3.9	ASN236, GLN143, ASP139, THR137, GLY234		**5**
76	50	00765	17	RPRCWIKIKFRCKSLKF	Piceain 2	83	-3.8	ASN236, GLN143, ASP139, ASP139, LEU96, PRO94, ALA172		**5**
77	115	01195	15	FIFPKKNIINSLFGR	Andersonin-D1	63	-3.6	ASN236, ASP139, ARG98, ARG98	**Only one interaction with Asn236 near the putative triad**	**4**
78	49	00764	20	KSLRPRCWIKIKFRCKSLKF	Piceain 1	98	-2.8	ASN236, THR137, PRO94, ARG98, ARG98		**4**
79	4819	02842	6	RRWQWR	LfcinB (20-25)	31	-5.8	GLN143, GLN143, ASP139, LEU96, LEU96, PRO94	**Only one interaction with Gln143 near the putative triad**	**3**
80	3469	18220	6	TVYTNA	Fusaricidin C (Bacteriocin)	22	-5.6	GLN143, GLY170, ILE235		**3**
81	5486	21429	7	KKLKAFA	Peptide 6 (Derived from B1)	33	-5.4	GLN143, ASP139, THR137, LEU96, PRO94		**3**
82	187	03542	10	HKTDSFVGLM	Neurokinin A	41	-5.2	GLN143, ASP139, THR237, THR237		**3**
83	3470	18219	6	TVVTQA	Fusaricidin B (Bacteriocin)	21	-5	GLN143, ILE235		**3**
84	5235	21222	8	IIRKIIRK	Control-4D (Derived from IK12-all L)	42	-5	GLN143, GLY170, GLY170, ILE235		**3**

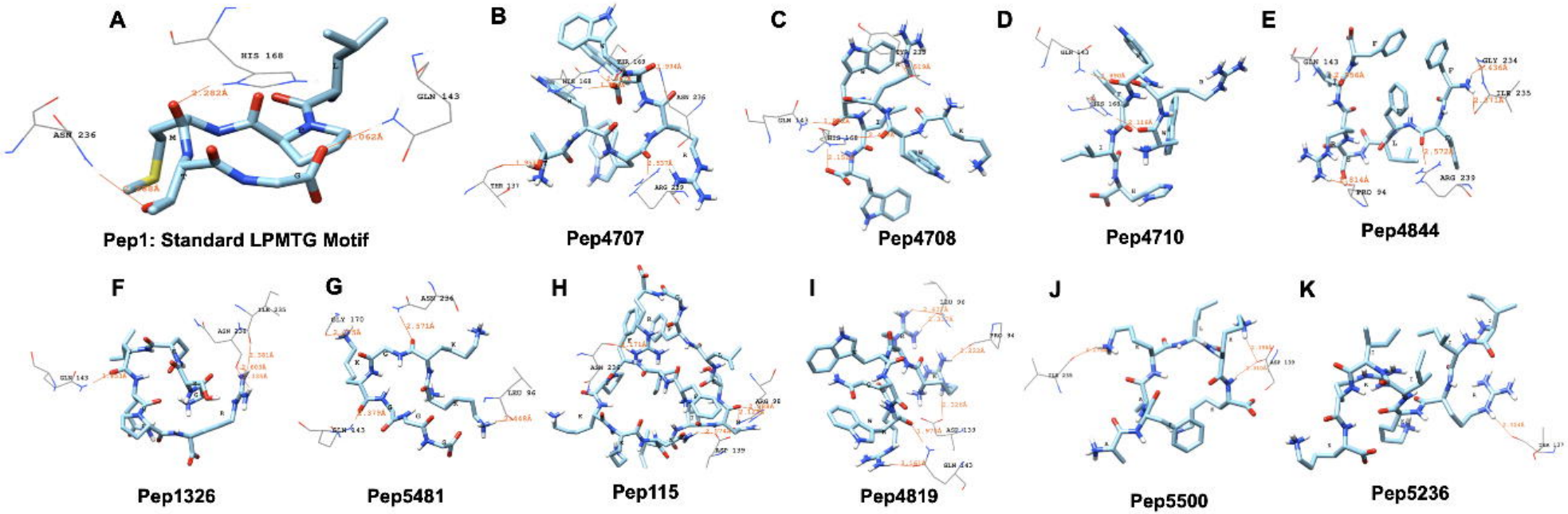
85	5243	21213	8	IRIKIRIK	IK8-all D (Derived from IK8-all L)	42	-5	GLN143, ASP139, THR137, THR137, ILE235, THR237		3
86	995	03868	8	RIWVIWRR	Bac8c (Bac2A variant)	37	-4.9	GLN143, THR169, TYR233, GLY234		3
87	1086	03988	7	LLKWLKK	L3K3W4 (LIKmWn model peptide)	37	-4.8	GLN143, ASP139		3
88	5242	21216	8	IRIKIRIK	IK8-4D (Derived from IK8-all L)	42	-4.7	GLN143, PRO94, GLY170, ILE235		3
89	5495	21438	7	KKKLAFA	Peptide 15 (Derived from B1)	33	-4.6	GLN143, ASP139, LEU96, PRO94		3
90	4933	20872	6	GLLKRI	Pal-ano-6 (Pal-anoplin peptide derivative)	27	-4.5	GLN143, GLY170, GLY170		3
91	5238	21220	8	IIRKIIRK	Control-all L (Derived from IK12-all L)	42	-4.5	GLN143		3
92	1088	03990	8	LLKWLKKL	L4K3W4 (LIKmWn model peptide)	41	-4.3	GLN143		3
93	5488	21431	7	KKLKAFA	Peptide 8 (Derived from B1)	33	-4.2	GLN143, ASP139, ASP139, ASP139, PRO94, ARG98		3
94	5500	21443	7	AFAKLKK	Peptide 20 (Derived from B1)	33	-5.3	ASP139, ASP139, ILE235	**Interactions with Asp139 away from the putative triad**	2
95	994	03867	8	RIWVIRWR	Bac8b (Bac2A variant)	37	-5	ASP139, PRO94, PRO94, GLY234		2
96	167	01350	11	FCTMIPIPRCY	Tigerinin-1	42	-5	ASP139, THR237, THR237		2
97	2534	04394	5	AMVSS	EP2	18	-4.5	ASP139, THR169		2
98	5236	21221	8	IIRKIIRK	Control-all D (From IK12-all L)	42	-4.3	THR137	**Interactions with other residues away from the putative triad**	1
99	5493	21436	7	AFALKKK	Peptide 13 (Derived from B1)	33	-4.3	GLY170, ILE235		1
100	3309	18229	2	LE	Gageotetrin A (Bacteriocin)	9	-4.3	THR231, THR231		1
101	993	03866	8	KIWVIRWR	Bac8a (Bac2A variant)	38	-4.5	No H interaction with any residues	**No interactions**	0


A
HIS
CYS
ARG

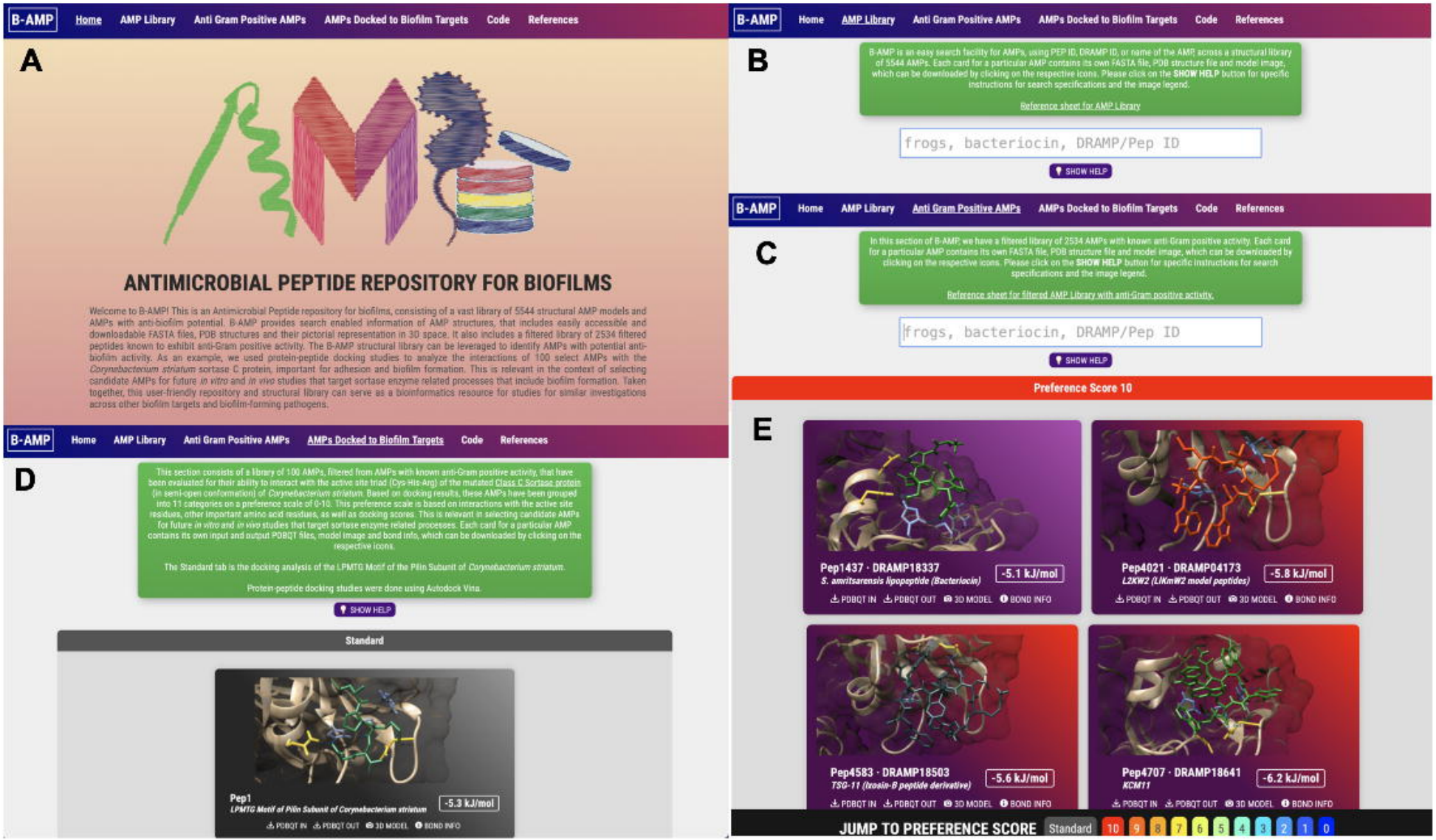
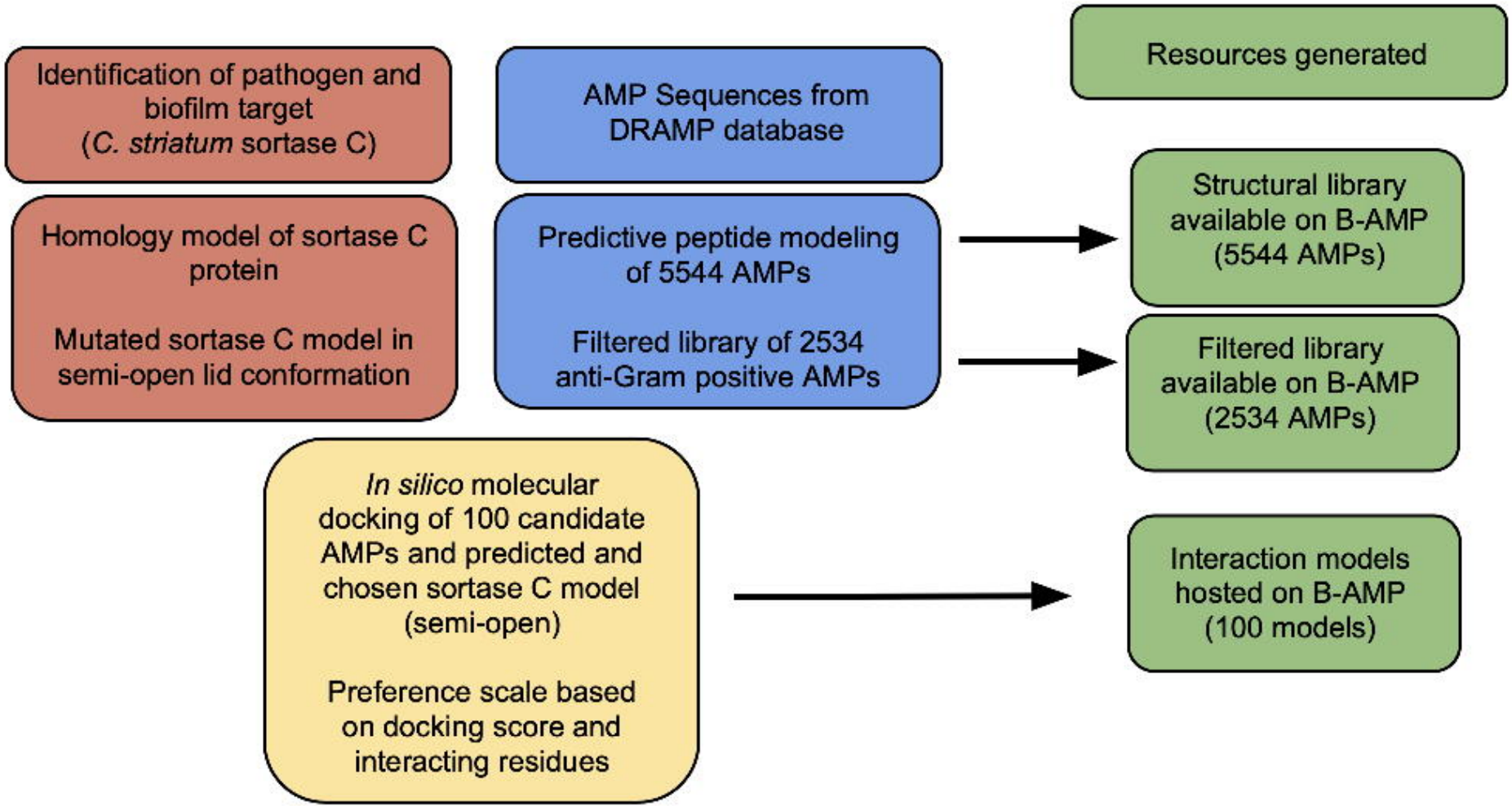

B


A
B
ASP 93
TRP 95
HIS
CYS
ARG
C
D
ALA 93
ALA 95
HIS
CYS
ARG


A
B
C
D
E
F


A
HIS 168
GLN 143
ASN 236
Pep1: Standard LPMTG Motif
B
Pep4707
C
Pep4708
D
Pep4710
E
Pep4844
F
Pep1326
G
Pep5481
H
Pep115
I
Pep4819
J
Pep5500
K
Pep5236

Size
Docking Score
Length of AMP (residues)
Docking score (kcal/mol)
Pep3309: THR231, THR231
Pep3296: HIS168, ASN236, ASN236, GLN143
Pep3298: HIS168, GLN143
Pep3240: HIS168, GLN143
Pep5241: HIS168, GLN143
Pep3297: HIS168, ASN236, GLN143
Pep3219: HIS168, GLN143
Pep3239: HIS168, GLN143
Pep4021: HIS168, ARG239, ASN236
Pep4934: HIS168, HIS168, GLN143
Pep4020: HIS168
Pep2533: ASN236, GLN143, THR137
Pep2534: ASP139, THR169
Pep4707: HIS168, ARG239, ASN236, THR137, THR169
Pep1437: HIS168, ARG239, GLN143, ASP139
Pep4708: HIS168, HIS168, GLN143, TYR233
Pep4710: HIS168, GLN143
Pep2442: HIS168, GLN143
Pep1005: HIS168, ASP139
Pep3475: HIS168, GLN143
Pep4709: HIS168, ASN236, GLN143, ASP139, PRO94
Pep3461: HIS168, GLN143, ASP139
Pep3292: HIS168, ASN236, GLN143, THR137
Pep5244: HIS168, GLN143
Pep4819: GLN143, GLN143, ASP139, LEU96, LEU96, PRO94
Pep3469: GLN143, GLY170, ILE235
Pep3470: GLN143, ILE235
Pep4933: GLN143, GLY170, GLY170
Pep4795: HIS168, ARG239, ASN236, GLN143
Pep5494: HIS168, ARG239, GLN143, LEU96
Pep5486: HIS168, ARG239, GLN143, ILE235
Pep1218: HIS168, HIS168, ASN236, GLN143, GLN143
Pep5482: HIS168, HIS168, GLN143, ALA95
Pep5508: HIS168, HIS168, GLN143, ILE235
Pep5492: HIS168, GLN143, THR137, PRO94
Pep3356: HIS168, GLN143, ASP139, ALA172
Pep1219: HIS168, GLN143, ILE235
Pep3308: HIS168, ASN236, GLN143, ARG98
Pep1087: HIS168, GLN143
Pep5501: HIS168, GLN143, ILE235
Pep5491: HIS168, GLN143, PRO94
Pep5487: HIS168, ASN236, GLN143
Pep5497: HIS168, GLN143, ILE235, GLY234
Pep5505: HIS168, GLN143, ASP139, THR137, ALA95, PRO94
Pep51: HIS168, ARG239, GLN143, ASP139
Pep5483: HIS168, GLN143
Pep5489: HIS168, GLN143
Pep5485: HIS168, ASN236, PRO94, GLY170
Pep4674: HIS168, GLN143, ASP139
Pep5488: HIS168, GLN143, ALA95
Pep5502: HIS168,GLN143,THR137,PRO94,GLY170,ILE235
Pep4932: HIS168,GLN143
Pep5509: HIS168,ASN236,GLN143
Pep3349: ARG239, GLN143, GLN143,ALA95
Pep5490: ARG239, ASN236
Pep4886: ARG239, GLN143, ASP139, THR137, GLY234
Pep5498: ARG239, ASN236, ASN236, PRO94, GLY170
Pep5484: ARG239, ASN236, ASP139, ALA95, ILE235
Pep1326: ASN236, ASN236, GLN143, ILE235
Pep5481: ASN236, GLN143, LEU96, GLY170
Pep3983: ASN236, GLN143, ASP139, THR137
Pep5486: GLN143, ASP139, THR137, LEU96, PRO94
Pep1086: GLN143, ASP139
Pep5495: GLN143, ASP139, LEU96, PRO94
Pep5488: GLN143, ASP139, ASP139, ASP139, PRO94
Pep5500: ASP139, ASP139, ILE235
Pep5493: GLY170, ILE235
Pep4583: HIS168, ARG239, LEU96, LEU96, PRO94, PRO94
Pep4931: HIS168, HIS168, ASN236, ASN236, GLN143
Pep3069: HIS168, GLN143, ILE235
Pep996: HIS168, ASN236, LEU96
Pep782: HIS168, ASN236, GLN143
Pep5363: HIS168, GLN143, ASP139, ILE235
Pep1245: HIS168,GLN143
Pep4844: ARG239, GLN143, PRO94, ILE235, GLY234
Pep5240: ASN236, ASN236, GLN143, ASP139, ASP139
Pep5245: ASN236, ASN236, GLN143, ASP139
Pep3052: ASN236, GLN143, ILE235
Pep1089: ASN236, GLN143, ASP139, PRO94, THR237
Pep5235: GLN143, GLY170, GLY170, ILE235
Pep5243: GLN143, ASP139, THR137, THR137, ILE235
Pep995: GLN143, THR169, TYR233, GLY234
Pep5242: GLN143, PRO94, GLY170, ILE235
Pep5238: GLN143
Pep1088: GLN143
Pep894: ASP139, PRO94, PRO94, GLY234
Pep5236: THR137
Pep993: No Interaction
Pep696: HIS168, ASN236, GLN143
Pep187: GLN143, ASP139, THR237, THR237
Pep167: ASP139, THR237, THR237
Pep168: HIS168, GLN143, ASP139, ASP139, ASP139
Pep88: ARG239, ASN236, GLN143
Pep26: ARG239, GLN143, THR169
Pep115: ASN236,ASP139,ARG98,ARG98
Pep72: ASN236, ASN236, GLN143, GLN143, GLN143
Pep50: ASN236, GLN143, ASP139, ASP139, LEU96
Pep113: ASN236, ASN236, GLN143, GLN143, THR137
Pep109: ASN236, GLN143, ASP139, THR137, GLY234
Pep49: ASN236,THR137,PRO94,ARG98,ARG98


B-AMP
Home
AMP Library
Anti Gram Positive AMPs
AMPs Docked to Biofilm Targets
Code
References
A
ANTIMICROBIAL PEPTIDE REPOSITORY FOR BIOFILMS
Welcome to B-AMP! This is an Antimicrobial Peptide repository for biofilms, consisting of a vast library of 5544 structural AMP models and AMPs with anti-biofilm potential. B-AMP provides search enabled information of AMP structures, that includes easily accessible and downloadable FASTA files, PDB structures and their pictorial representation in 3D space. It also includes a filtered library of 2534 filtered peptides known to exhibit anti-Gram positive activity. The B-AMP structural library can be leveraged to identify AMPs with potential anti-biofilm activity. As an example, we used protein-peptide docking studies to analyze the interactions of 100 select AMPs with the Corynebacterium striatum sortase C protein, important for adhesion and biofilm formation. This is relevant in the context of selecting candidate AMPs for future in vitro and in vivo studies that target sortase enzyme related processes that include biofilm formation. Taken together, this user-friendly repository and structural library can serve as a bioinformatics resource for studies for similar investigations across other biofilm targets and biofilm-forming pathogens.
B
B-AMP is an easy search facility for AMPs, using PEP ID, DRAMP ID, or name of the AMP, across a structural library of 5544 AMPs. Each card for a particular AMP contains its own FASTA file, PDB structure file and model image, which can be downloaded by clicking on the respective icons. Please click on the SHOW HELP button for specific instructions for search specifications and the image legend.
Reference sheet for AMP Library
frogs, bacteriocin, DRAMP/Pep ID
SHOW HELP
C
In this section of B-AMP, we have a filtered library of 2534 AMPs with known anti-Gram positive activity. Each card for a particular AMP contains its own FASTA file, PDB structure file and model image, which can be downloaded by clicking on the respective icons. Please click on the SHOW HELP button for specific instructions for search specifications and the image legend.
Reference sheet for filtered AMP Library with anti-Gram positive activity.
frogs, bacteriocin, DRAMP/Pep ID
SHOW HELP
D
This section consists of a library of 100 AMPs, filtered from AMPs with known anti-Gram positive activity, that have been evaluated for their ability to interact with the active site triad (Cys-His-Arg) of the mutated Class C Sortase protein (in semi-open conformation) of Corynebacterium striatum. Based on docking results, these AMPs have been grouped into 11 categories on a preference scale of 0-10. This preference scale is based on interactions with the active site residues, other important amino acid residues, as well as docking scores. This is relevant in selecting candidate AMPs for future in vitro and in vivo studies that target sortase enzyme related processes. Each card for a particular AMP contains its own input and output PDBQT files, model image and bond info, which can be downloaded by clicking on the respective icons.
The Standard tab is the docking analysis of the LPMTG Motif of the Pilin Subunit of Corynebacterium striatum.
Protein-peptide docking studies were done using Autodock Vina.
SHOW HELP
Standard
Pep1
LPMTG Motif of Pilin Subunit of Corynebacterium striatum
-5.3 kJ/mol
PDBQT IN
PDBQT OUT
3D MODEL
BOND INFO
Preference Score 10
E
Pep1437 · DRAMP18337
S. amritsarensis lipopeptide (Bacteriocin)
-5.1 kJ/mol
Pep4021 · DRAMP04173
L2KW2 (LlKmW2 model peptides)
-5.8 kJ/mol
Pep4583 · DRAMP18503
TSG-11 (trosin-B peptide derivative)
-5.6 kJ/mol
Pep4707 · DRAMP18641
KCM11
-6.2 kJ/mol
JUMP TO PREFERENCE SCORE
Standard
10
9
8
7
6
5
4
3
2
1
0

In silico approach to identifying AMPs and anti-biofilm potential and resources developed