Machine learning-enabled design features of antimicrobial peptides selectively targeting peri-implant disease progression

Kyle Boone[1,2], Natalia Tjokro[3], Kalea N. Chu[1,4], Casey Chen[3], Malcolm L. Snead[3,4] and Candan Tamerler[1,2,4]*

[1]Institute for Bioengineering Research, University of Kansas, Lawrence, KS, United States, [2]Department of Mechanical Engineering, University of Kansas, Lawrence, KS, United States, [3]Center for Craniofacial Molecular Biology, Herman Ostrow School of Dentistry of USC, University of Southern California, Los Angeles, CA, United States, [4]Bioengineering Program, University of Kansas, Lawrence, KS, United States

Peri-implantitis is a complex infectious disease that manifests as progressive loss of alveolar bone around the dental implants and hyper-inflammation associated with microbial dysbiosis. Using antibiotics in treating peri-implantitis is controversial because of antibiotic resistance threats, the non-selective suppression of pathogens and commensals within the microbial community, and potentially serious systemic sequelae. Therefore, conventional treatment for peri-implantitis comprises mechanical debridement by nonsurgical or surgical approaches with adjunct local microbicidal agents. Consequently, current treatment options may not prevent relapses, as the pathogens either remain unaffected or quickly re-emerge after treatment. Successful mitigation of disease progression in peri-implantitis requires a specific mode of treatment capable of targeting keystone pathogens and restoring bacterial community balance toward commensal species. Antimicrobial peptides (AMPs) hold promise as alternative therapeutics through their bacterial specificity and targeted inhibitory activity. However, peptide sequence space exhibits complex relationships such as sparse vector encoding of sequences, including combinatorial and discrete functions describing peptide antimicrobial activity. In this paper, we generated a transparent machine learning (ML) model that identifies sequence-function relationships based on rough set theory using simple summaries of the hydropathic features of AMPs. Comparing the hydropathic features of peptides according to their differential activity for different classes of bacteria empowered the predictability of antimicrobial targeting. Enriching the sequence diversity by a genetic algorithm, we generated numerous candidate AMPs designed for selectively targeting pathogens and predicted their activity using classifying rough sets. Empirical growth inhibition data are iteratively fed back into our ML training to generate new peptides, resulting in increasingly more rigorous rules for which peptides match targeted inhibition levels for specific bacterial strains. The subsequent top scoring candidates were empirically tested for their inhibition against keystone and accessory peri-implantitis pathogens as well as an oral commensal bacterium. A novel peptide, VL-13, was confirmed to be selectively active against a keystone pathogen. Considering the continually increasing number of oral implants placed each year and the complexity of

the disease progression, the prevalence of peri-implant diseases continues to rise. Our approach offers transparent ML-enabled paths towards developing antimicrobial peptide-based therapies targeting the changes in the microbial communities that can beneficially impact disease progression.

1 Introduction

Despite high success rates for dental implants, their bacterial plaque-associated inflammatory lesions, known as peri-implant diseases, still occur (1, 2). These lesions continue to degrade the stability of peri-implant soft and hard tissues, which can result in loss of the implant. While peri-implant mucositis is a reversible inflammatory condition, peri-implantitis is an irreversible pathological condition leading to loss of supporting alveolar bone (2). The reported prevalence of peri-implant mucositis and peri-implantitis shows a substantial increase over time following implant placement. Meta-analysis for patient-based peri-implant mucositis and peri-implantitis was reported as 46.83% and 19.83% by Lee *et al.* (3). In a separate study, meta-analyses estimated the peri-implant mucositis and peri-implantitis as 43% and 22%, respectively. Peri-implantitis is also reported to be in the range of 11%–47% among dental implants 10 years after their placement (4, 5). These numbers further increase in periodontally compromised patients (1, 6, 7).

Current treatments for peri-implantitis and periodontitis include mechanical debridement, disinfection of exposed implant surfaces, and antibiotic or antiseptic prescriptions to suppress the associated bacteria (5). The use of adjunctive antibiotics for treating peri-implantitis or periodontitis is debated mainly because of concerns about microbial antibiotic resistance, the non-selective suppression of both pathogenic and commensal species, and the adverse systemic reactions. Notably, these conventional treatment modalities may not prevent relapses, as the pathogens may either remain unaffected or quickly re-emerge after treatment.

The poor efficacy of antibiotic treatment in peri-implantitis may be explained by the non-specific suppression of dysbiotic biofilms. The adaptability and resiliency of pathogenic bacteria in biofilms is well documented (8). The unique structure and inter-species relationships within a biofilm enhance the individual strengths of the bacteria present, creating an unbalanced community organized to promote communal success at the expense of the host (9). Notably, keystone pathogens play an outsized role in shaping the community structure. Therefore, targeting keystone pathogens may be the most effective approach to reverse microbial dysbiosis and return to health-compatible eubiosis.

In peri-implantitis, *Porphyromonas gingivalis* (*P. gingivalis*) is widely acknowledged as a keystone pathogen (10). *P. gingivalis* is associated with increased levels of inflammation and subsequent alveolar bone loss (11). Once *P. gingivalis* has initiated biofilm

growth, other pathogens are free to flourish and further contribute to the dysbiotic community (11, 12). The interdependent-relationships among pathogens in a biofilm are a defining factor in their treatment difficulty. In peri-implantitis, this is evident by the coexistence of *Aggregatibacter actinomycetemcomitans*, another keystone pathogen associated with aggressive periodontitis, and *Streptococcus gordonii*, a commensal and accessory pathogen (13, 14). Microbial communities exhibiting both *P. gingivalis* and *S. gordonii* are linked to more severe cases of peri-implantitis, resulting in increased infection and bone loss as compared to others (15–17). Undoubtedly, in peri-implantitic biofilms, pathogens grow synergistically to promote each other's survival. Keystone pathogens, such as *P. gingivalis* and *A. actinomycetemcomitans,* play a pivotal role in shifting the oral microbiome to induce the host into a disease-oriented state. The presence of these pathogens is widely associated with intensified inflammation levels, prolonged infection, and enhanced alveolar bone loss in patients (18, 19). Successful mitigation of disease progression in peri-implantitis requires a specific mode of treatment capable of targeting keystone pathogens and restoring bacterial community balance toward commensal species. Broad-spectrum approaches have difficulty in preventing oral dysbiosis (11, 18, 20).

Antimicrobial peptides (AMPs) have been receiving increasing attention as promising therapeutic candidates since their use leads to no or low antibiotic resistance (21, 22). Moreover, AMPs with a short sequence domain offer straight-forward manufacturing, and relatively low-cost production (23). Within AMPs, antibacterial peptides account for the largest proportion of peptides with inhibitory activities ranging from broad- to specific-species (21). Complexity in structure-function relationships in AMPs is increasing with the increased number of peptide sequences that are isolated from a wide range of organisms, designed using computational search methods, or developed as peptide-mimics as potential candidates (24–27). Despite such progress, and the promise of AMPs as alternative treatments to antibiotics and antiseptics, still only a handful of AMPs have been applied to oral-craniofacial applications (28–33). We and other groups investigated AMPs that may serve to reduce biofilm load and/or to target emergent keystone pathogens on dental implants, and mitigation of bacterial-induced peri-implantitis has been demonstrated by rationally designed chimeric AMPs and peptides (30, 34–40).

As a result of growing interest in AMPs, large databases on their sequence and known functions are now readily available (41–46). With the increasing number of AMPs discovered at the

lab bench and via computational methods, determining the boundaries of similar AMPs and identifying their bacteria-specific function remains challenging. Machine learning models offer unique tools to find AMPs with targeted functionality (42). Through bioinformatics similarity tools, ML models are effective in identifying possible antimicrobial peptides among the large number of nucleic acid sequences (41). Recurrent neural networks (RNNs), long-short term models (LSTMs) and other deep learning methods have demonstrated success in peptide related-prediction models and these methods are now being used in constructive model approaches to design AMPs (47–51). However, when these methods are applied in generative approaches, they severely lack training adaptability. This is mainly due to their requirement for a relatively large number of training sets needed to learn specific paths in high-dimensional decision space. This makes them at risk for re-enforcing errors when the models incorrectly classify cases by correlated features. Customizing decisions for the sample distribution may optimize prediction performance, but this makes the ability of the model to adjust to the trends in unbiased sampling data extremely difficult to achieve. Gradient descent or backpropagation methods can help recognize the contribution from previously under-represented subgroups that are negatively impacting the models' applicability and overall efficacy. However, since retraining the whole model comes with a large computational cost, alternative methods are sought to address the bias. Still, no strategies are currently known that solve this issue.

In our previous work, we pioneered the use of rough set theory for the classification of peptide sequences and demonstrated how to achieve training adaptability by bypassing neural networks in the context of AMP identification (52). In this approach, different descriptors in accordance with antibacterial activity are analyzed by rough set theory (RST) which is used as a heuristic method to discover the rules distinguishing different outcomes. By combining the rough set theory approach with the algorithm of Modified Learning from Examples Module, Version 2 (MLEM2) and the Interesting Rule Induction Module (IRIM) algorithm, we achieved high-specificity performance. Our method provides a transparent selection approach to define explicit boundaries that distinguish between classes of AMPs by their activity. The model can adapt by using the explicit decision components and the related rules that are introduced by new hypotheses and labelled data. Non-linear categories distinguishing between active and inactive peptides reduce training time of the model, while preserving the structure of the explicit model choices. This improves training flexibility and avoids the cost barrier associated with re-training or the creation of new models. This method also guards against irrational decision relationships by maintaining transparency for each decision step throughout the decision process. In a separate study, we combined the RST based ML approach (CLN-MLEM2) with a codon-based genetic algorithm (CB-GA) and increased the variations of peptide sequences generated by RST ML search (53). Using the CB-GA combined ML approach, we identified an AMP sequence effective against *Staphylococcus epidermidis*. The training false discovery rate, i.e., probability of false positives, was approximately 5% (53).

In this study, we developed a transparent ML model and combined it with a genetic algorithm, that empowers AMP design targeted to a keystone pathogen. The predicted activity of the generated AMPs was classified using rough sets and the rules were improved using empirical growth inhibition data for specific pathogens. The validation tests were run with the peptides having the highest inhibition predictability score against the keystone pathogen, *A. actinomycetemcomitans*. A novel peptide VL-13 was confirmed to be active against the selected keystone pathogen without compromising the accessory-commensal species, *S. gordonii*.

2 Materials and methods

2.1 Materials

A. actinomycetemcomitans strain D7S-1, *S. gordonii* strain Challis, and *Streptococcus sanguinis* strain ATCC10556 were cultured using either modified Trypticase Soy Broth (mTSB) containing 3% trypticase soy broth and 0.6% yeast extract or on mTSB agar (mTSB with 1.5% agar Becton Dickinson and Company). In some experiments, the bacteria were cultured in SHI medium supplemented with hemin (5 µg/ml) (Sigma-Aldrich, St. Louis, MO, USA), menadione (1 µg/ml) (Sigma-Aldrich), human serum (10%) (Sigma-Aldrich), and sucrose (0.25%). Bacteria were cultured at 37 °C in a humidified atmosphere supplemented with 5% CO_2 (54). *P. gingivalis* strain ATCC33277 was cultured in the brain-heart infusion (BHI) broth at 37 °C under anaerobic conditions.

2.2 Bacterial viability tests

Bacterial viability tests were done by first adjusting the optical density of bacterial cultures to 0.2 (equivalent to approximately 10^7 CFU per ml) at 600 nm. The bacterial cultures were then diluted 1:20 to get to 5×10^5 CFU/ml and incubated with 100 µM AMPs. Bacterial viability was then determined at 0 and 6 h by CFU counts and at an additional 24-h time point for *P. gingivalis*.

2.3 Machine learning model

2.3.1 Initial datasets

The complete listing of the antimicrobial peptides used to generate the targeted rough set theory rules is given in Supplementary Table S1. In addition to literature-derived peptides, we included additional antimicrobial peptides previously studied for different applications shown in Table 1. The initial datasets for peptide generation were taken from the iAMP-2l database (55). This database was filtered to only include examples of anti-bacterial peptides, resulting in 1,274 unique peptides from the database. Of the 21 peptide sequences provided in Table 1 and Supplementary Table S1, 15 of them

TABLE 1 Rough set theory identities for *in vitro* inhibition against *A. actinomycetemcomitans* (Aa), *S. gordonii* (Sg), and *P. gingivalis* (Pg).

Name	Sequence	*Aa* inhibition	*Sg* inhibition	*Pg* inhibition
AMPa	KWKLWKKIEKWGQGIGAVLKWLTTW	High	High	Low
AMP1	LKLLKKLLKLLKKL	High	High	High
AMP2-NH2	KWKRWWWWR-NH2	None	None	None
TiBP-AH-AMP1	RPRENRGRERGLKGSVLSALKLLKKLLKLLKKL	None	None	None
TiBP-AH-GL13K	RPRENRGRERGLKGSVLSAGKIIKLKASLKLL	None	None	None
AMP7	ESYKKML	None	None	None
AMP10	GILGKLWEGVKSTF	None	None	None
TiBP-S5-AMPA	RPRENRGRERGLGSGGGKWKLWKKIEKWGQGIGAVLKWLTTW	None	None	None

were already included in the database. Therefore, a total of 1,280 AMP sequences are included in the initial set.

2.3.2 Peptide customization by rough set theory

We generated the rough set theory rules as described in our earlier publication on the CLN-MLEM2 method developed for classification of antimicrobial peptides with two enhancements (52). Previously, the rough set rules were designed to establish the boundaries simply between active and inactive antibacterial peptides, using non-correlated AAindex1 properties. The first enhancement we made is to combine the rough set rules from our previous paper with the targeting rough set theory rules and generate a multiple-dimensional view of the predicted activity. To generate these targeted activity rules, a set of AMPs with confirmed antibacterial activity against any of the three pathogens of interest, i.e., *P. gingivalis, S. gordonii, A. actinomycetemcomitans*, were used as positive data set (see Table 1, Supplementary Table S1 and Figures S1–S3). The second enhancement for targeting was introduced by focusing on the key physicochemical property features. We integrated eight indices proposed by a recent study as reduced AAindex (rAAindex) obtained from a subset of original 544 indices in the amino acid index database (56). Kibinge et al., applied a random forest (RF) algorithm for property reduction and maximizing metadata capturing. With the two enhancements introduced, our method creates collections of discriminating attributes separating targeted AMP activity trends focused on hydropathy variations for peptide sequences.

The resulting rules are each characterized by a hydropathy property of importance determined by the updated CLN-MLEM2 method, and a simple summary characteristic that portrays the features to be used to predict an AMP`s targeted ability. Sequence features that are most relevant for the observed peptide activity are collected as simple arithmetic summaries. The summary characteristics used included the sum of the property across the amino acids of the sequence, the mean of the property across the amino acids of the sequence and the maximum value of the property across three consecutive amino acids within a sequence (50). These summary characteristics correspond to non-linear boundaries between activity classes. Each time a rule set is generated, these boundaries and properties are chosen to separate the sequences into the desired classification groups. The heuristic goal of the method is to create definitions of activity classification with the minimum number of rules and conditions per rule possible. Rule sets contain descriptions of active and inactive peptides, but likely do not contain the set of all peptides in the union of the rule sets. Peptides that do not meet any of the rule sets are not identified directly by our classification method. Since specific peptide activity is not likely to be present for randomly selected sequences, we impute peptides as inactive if they do not belong to any rule set positive for activity.

2.3.3 Sequence expansion by codon-based genetic algorithm

We used the codon-based genetic algorithm for sequence expansion reported in one of our earlier publications (53). To begin this process, we used a total of 1,280 AMP sequences in the initial set, which contained a large variety of sequences to recombine and mutate through artificial genetic operations. These sequences were subsequently ranked by which generated sequences met the rough set theory rules for targeted antimicrobial activity. The rule sets have two separate descriptions of antimicrobial activity in this study. The first description is the rough sets previously used as a measure for broad-spectrum activity estimation. The second description is the newly generated rules targeting the chosen periodontal keystone pathogens. This second level of activity distinguishes between keystone-only activity and other antimicrobial activity with the goal of avoiding impacting commensal species. These two descriptions were weighed as components of the fitness function. There is a large disparity between the number of previously trained sequences, i.e., 2,347 sequences, and the empirically tested targeted sequences, i.e., 21 sequences (Table 1 and Supplementary Table S1). The rule counts were therefore independently calculated and normalized before being combined as separate terms in the fitness objective function. We defined fitness objective function for the codon-based genetic algorithm by the following equation in which *AB* is referred as antibacterial:

$$F(x) = \min\,(10*(\text{Max Target Rule Count} - \text{Sequence Target Rule Count})$$
$$+\;\; 0.01 * (\text{Max } AB \text{ Rule Count} - \text{Sequence } AB \text{ Rule Count}) \quad\quad (1)$$
$$+\; 5 * \min\,([10, 30] - \text{Sequence Amino Acid Length}))$$

The mutation rate for sequences to go to the next generation is 25%. Mutation changes a codon in the sequence, which may not result in an amino acid change or could result in a stop codon. The cross-over rate was 50%. Crossing over was completed with codon representation, often resulting in frameshifts for new candidate sequences compared to the parent codon sequences.

The generations were monitored for convergence, both with the best fitness between generations and the consistency of the targeting rules to generate the top-scoring sequences. Generations are deemed mature for identifying top candidates when the majority of sequences meeting the targeting rules are within one half of the maximum number of targeting rules.
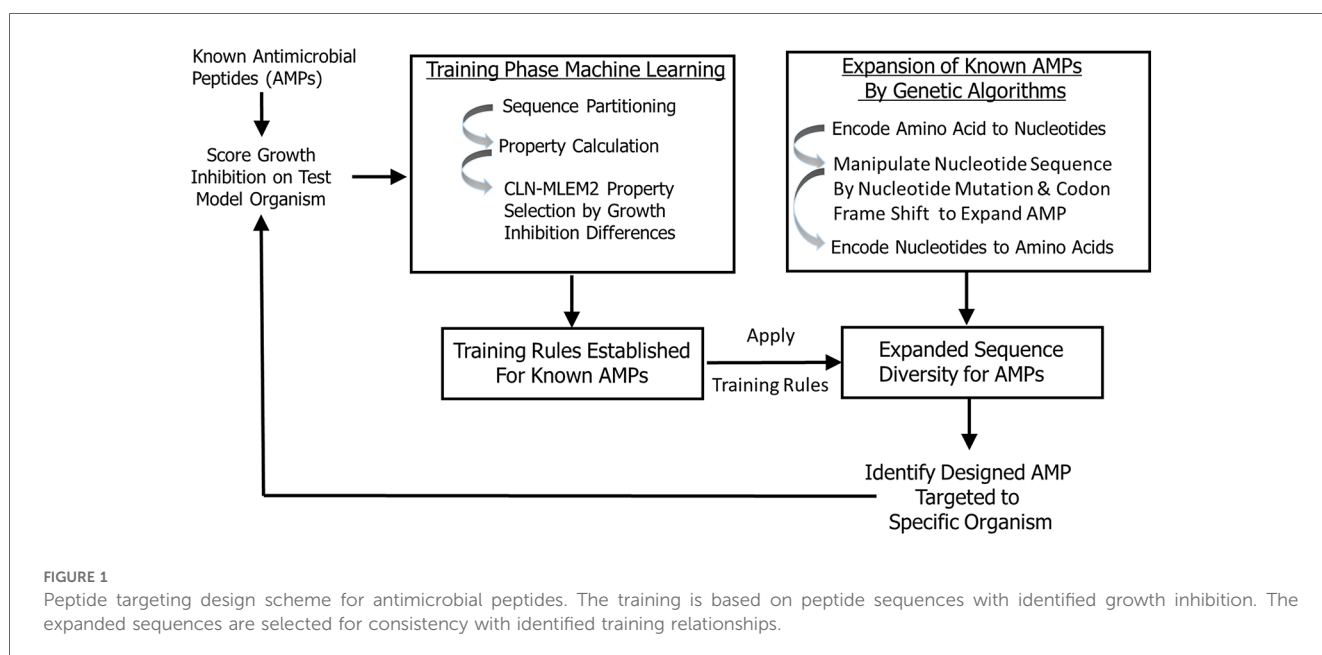
2.3.4 Peptide synthesis

Peptides were synthesized using Wang resin following a standard Fmoc chemistry method using an Aapptec Focus XC peptide synthesizer. Dimethylformamide (DMF) and 20%–40% piperidine in DMF were used for Fmoc deprotection with two repetitions. The peptide-resins were then washed with DMF. Activation of 0.2 M amino acids/DMF (2 equivalents) was performed by addition of 0.2M 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)/DMF. The coupling step was completed twice, and the procedure was repeated until the complete peptide was assembled on the solid resin support. Following synthesis, the peptide-resin was removed from the reaction vessel using DMF. Following the removal of DMF from the peptide-resin by washing with ethanol, a cleavage cocktail (15 ml/1 gram of resin) was added to the dried resin for 2 h with gentle stirring to remove the peptide from the solid support and remove the side chain protecting groups. The standard cleavage cocktail is composed of trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/water (95:2.5:2.5, % vol/vol/vol). To remove side chain protecting groups from peptides containing histidine or cysteine, 2.5% thioanisole and 2.5% 1,2 ethanedithiol were added to the cocktail and for peptides containing methionine, tyrosine, or arginine, 5% phenol was added. The cleavage products were filtered, and crude peptide product was isolated by precipitation in cold ether. The crude peptide was pelleted by centrifugation (2,000 rpm for 2 min), the supernatant

was removed, and the process was repeated for two to four times prior to lyophilization of the peptide products.

3 Results

In this study, we developed a transparent ML model that allowed us to incorporate iterative training sets and enrich our sequence space by a genetic algorithm to design antimicrobial peptides with inhibitory activity targeted to periodontal keystone and accessory pathogens. To identify antimicrobial peptides specific to oral pathogens, we first established rough set boundaries for training rules, building upon our initial iteration of known AMPs and their antimicrobial activity (Supplementary Figures S1–S3, Table 1, and Supplementary Table S1). Our rough set theory classifier was trained to identify possible antimicrobial peptides specific to oral keystone and supporting pathogens, with separate rule sets for each strain. We expanded sequence diversity using a genetic algorithm and selected novel candidates consistent with their predicted inhibitory activities using the identified training relationships. These candidates are generated by the second iteration. While the rough set rules generated apply to many other sequences than the sequences we trained on, the rules will not, in general, cover all sequences. Non-conforming sequences are imputed as non-targeted. Therefore, our second iteration is focused on finding the sequences which are like the sequences we identified as active against a targeted species in the first iteration considering the feature properties found to be discriminating between active and inactive against that single species. Training sequences do not need to have specific activity to generate sequences with targeted activity; multiple examples of non-specific peptides can still provide direction for what features are needed in generating rules. Figure 1 provides the selective targeting antimicrobial peptide design scheme which includes training the

FIGURE 1
Peptide targeting design scheme for antimicrobial peptides. The training is based on peptide sequences with identified growth inhibition. The expanded sequences are selected for consistency with identified training relationships.

next iteration of our model with known AMPs, establishing training rules, expansion of sequence diversity, candidate selection and verification of their predicted inhibition properties targeted to specific organisms. The targeting rules are heuristically made to be a minimal set that covers all our training sequences.

3.1 Establishing boundaries for growth inhibition rules

Our machine learning approach builds upon our previously developed CLN-MLEM2 method that utilizes rough set theory principles (52). The CLN-MLEM2 method separates sequences by their amino acid properties to establish functional classification. This method simultaneously filters which properties are key properties and provides boundaries for classification. The key properties identified by the CLN-MLEM2 method are provided in Supplementary Table S2. These properties were selected from the rAAindex (56), a reduced subset of the AAindex focused on hydropathy. We initiated the machine learning training by providing literature data to set up the initial inhibition descriptions for rough set participation to address selected pathogens (Supplementary Table S1). The rough sets are key property summary descriptions (e.g., property sum, property mean, property peak window) that identify peptides to have a certain activity level. The key property summary descriptions are used as features in our ML method (57). Both the full sequence length properties and short sequence segments are included as features. The short segments are summarized for a sequence by selecting the property peak window features. The full-sequence length features are included as the property sum and the property mean values. The CLN-MLEM2 method used 6 of the 8 properties in the rAAindex to build the rule sets. The selected properties and their short descriptions are included in Supplementary Table S2. The

inhibition activity data captured from the literature was also supplemented by incorporating the inhibitory activity from additional antimicrobial peptide sequences into the model. These sequences have been shown to be active in different contexts; therefore we evaluated their activity against selected oral pathogens, *P. gingivalis*, *A. actinomycetemcomitans*, and *S. gordonii* (see Supplementary Figures S1–S3). The level of minimum inhibitory activities is categorized from low to high and provided in Table 1.

The ratio of correctly identified cases to the total number of applicable cases for a set is identified as α ($0 \leq \alpha \leq 1$) (58–60). The CLN-MLEM2 method selects rules based on α ($0 \leq \alpha \leq 1$). Using higher values of α generates fewer rules with higher probability (Pr) values of training accuracy if generated rules do not meet the accuracy specification. Using lower values of α generates more rules with lower Pr values of training accuracy when rules do not meet the accuracy specification. For all our training peptide sets, we found every rule with $\alpha = 1.0$. Therefore, all generated rules meet the maximum accuracy specification of 1.0. Our current descriptions allow for enough discrimination to uniquely describe all peptide sequences when they differ in activity. We found hydropathy feature boundaries which explicitly classified our training examples. We next generated the CLN-MLEM2 rules for predicting the activity against keystone pathogen members *A. actinomycetemcomitans* and *P. gingivalis*, and accessory-commensal class member *S. gordonii*. These rules provide design criteria for either increasing or decreasing the probability of a peptide sequence having an antimicrobial activity against each strain. Our initial design strategy involved finding an antimicrobial peptide which heuristically has as many features as possible to be active through our genetic algorithm to satisfy the maximum count of non-linear boundary rules simultaneously through computational search.

Table 2 shows the selected sequence property rules that are associated with the inhibition of *A. actinomycetemcomitans* and *S. gordonii*. The sequence summary features in this table

TABLE 2 A set of CLN-MLEM2 rules for *A. actinomycetemcomitans* (Aa) and *S. gordonii* (Sg).

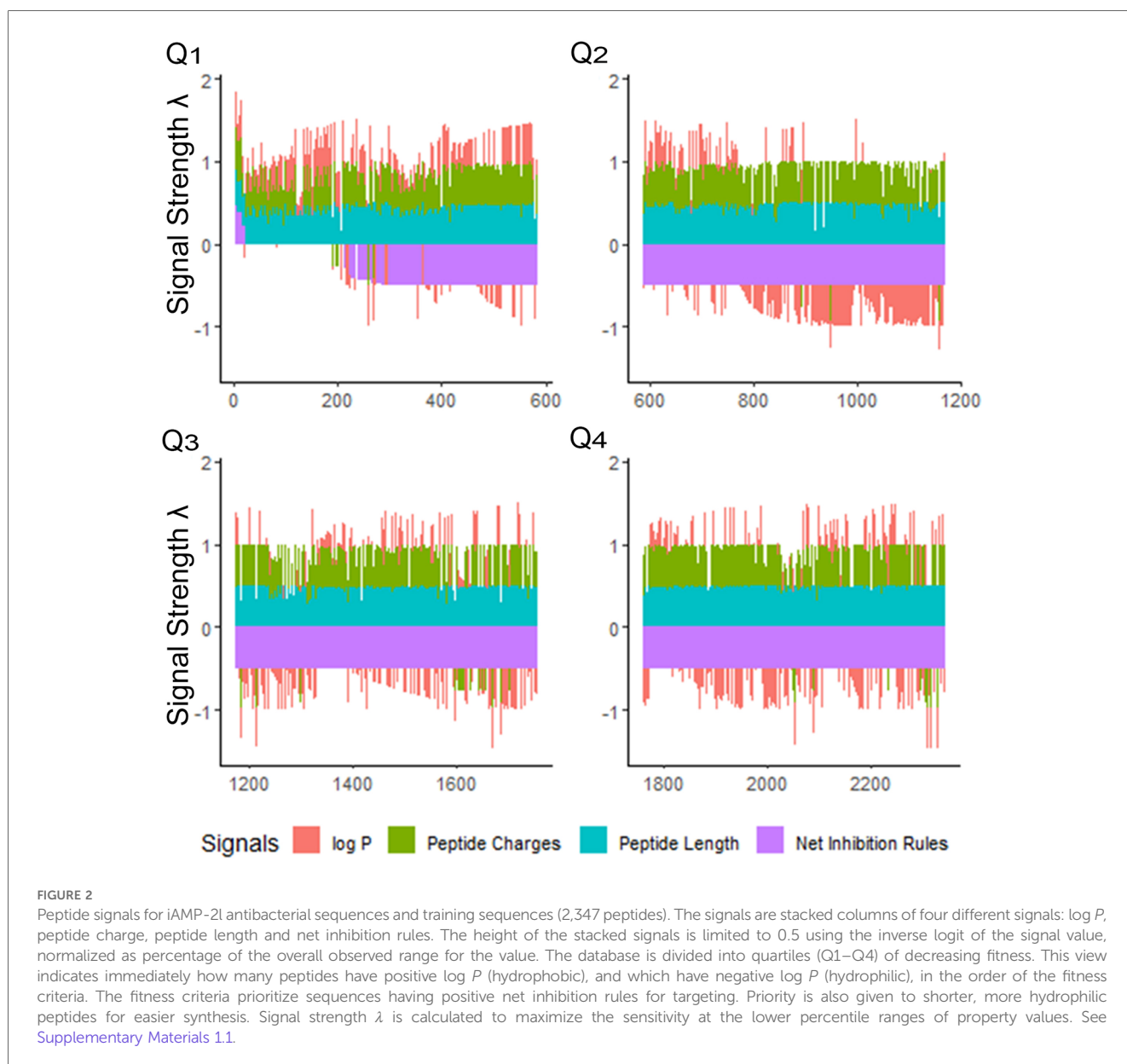| Activity identity	Rule	Training sequence support	Condition	Calculation	AAindex1 property	Lower-bound	Upper-bound
Growth inhibition of *A.a.*	1	3	1	Mean	ZIMJ680103	17.09	22.99
			2	Tripeptide Window	JACR890101	0.615	0.935
			3	Sum	COWR900101	−27.82	0.27
			4	Mean	JACR890101	−1.743	−1.218
			5	Mean	COWR900101	−0.359	−0.001
Growth inhibition of *A.a.*	2	2	1	Mean	WARP780101	6.401	7.141
			2	Mean	MEEJ810102	3.227	7.278
			3	Tripeptide Window	FAUJ880110	2.0	5.5
			4	Mean	FAUJ880110	0.40	0.69
No growth inhibition of *A.a.*	3	11	1	Tripeptide Window	FAUJ880110	6.5	12.0
			2	Tripeptide Window	COWR900101	3.78	5.41
			3	Tripeptide Window	WARP780101	26.825	30.12
			4	Tripeptide Window	JACR890101	0.89	1.32
Growth inhibition of *S.g.*	1	43	1	Tripeptide Window	FAUJ880110	3.5	12.0
			2	Mean	FAUJ880110	0.69	2.0
No growth inhibition of *S.g.*	2	35	1	Tripeptide Window	LIFS790102	3.505	7.890
			2	Tripeptide Window	WARP780101	24.64	30.12
			3	Tripeptide Window	COWR900101	2.9	5.41
			4	Tripeptide Window	JACR890101	0.89	1.32

provide conditions for classifying peptides with potential inhibitory properties. We provide a detailed analysis of the *A. actinomycetemcomitans* inhibition rules in Supplementary Figures S4–S9. We found that the rules positive for *A. actinomycetemcomitans* inhibition are consistent with the first iteration identities and apply to either slightly polar mean polarity or to slightly non-polar mean polarity sequence descriptions, depending on the scale used. Interestingly, instead of the descriptions being mutually exclusive, VL-13 combined membership of both rules into a single peptide sequence.

3.2 Ranking antimicrobial peptides by rough set theory relevance

Once "rough set" boundaries are established from our first iteration, antimicrobial peptides in the database can be compared in relation to the selected physical properties of their amino acids. We explored the selectivity of our training methods for the iAMP-2l database, which included 2,347 peptides (55). Using the CLN-MLEM2 rules, we ranked this database and generated a relatively small number of sequences of interest for further exploration (Figure 2). After combining the peptides in the iAMP-2l database and our initial testing set, only 20 peptide sequences (0.9%) met conditions for CLN-MLEM2 rules, the net of which were rules for inhibition instead of non-inhibition.

Overall, only 20 peptides in the iAMP-2l database (~2,200 peptides) had similar hydropathy features to the confirmed active AMPs and thus met the CLN-MLEM2 rules derived from our first iteration in antimicrobial activity prediction. We avoided selecting peptides with cysteine to enable rational cyclization studies using disulfide bonds in the future. Therefore, the third-ranked peptide KF-18 (KWKLFKKIPKFLHLAKKF) was selected

FIGURE 2
Peptide signals for iAMP-2l antibacterial sequences and training sequences (2,347 peptides). The signals are stacked columns of four different signals: log P, peptide charge, peptide length and net inhibition rules. The height of the stacked signals is limited to 0.5 using the inverse logit of the signal value, normalized as percentage of the overall observed range for the value. The database is divided into quartiles (Q1–Q4) of decreasing fitness. This view indicates immediately how many peptides have positive log P (hydrophobic), and which have negative log P (hydrophilic), in the order of the fitness criteria. The fitness criteria prioritize sequences having positive net inhibition rules for targeting. Priority is also given to shorter, more hydrophilic peptides for easier synthesis. Signal strength λ is calculated to maximize the sensitivity at the lower percentile ranges of property values. See Supplementary Materials 1.1.
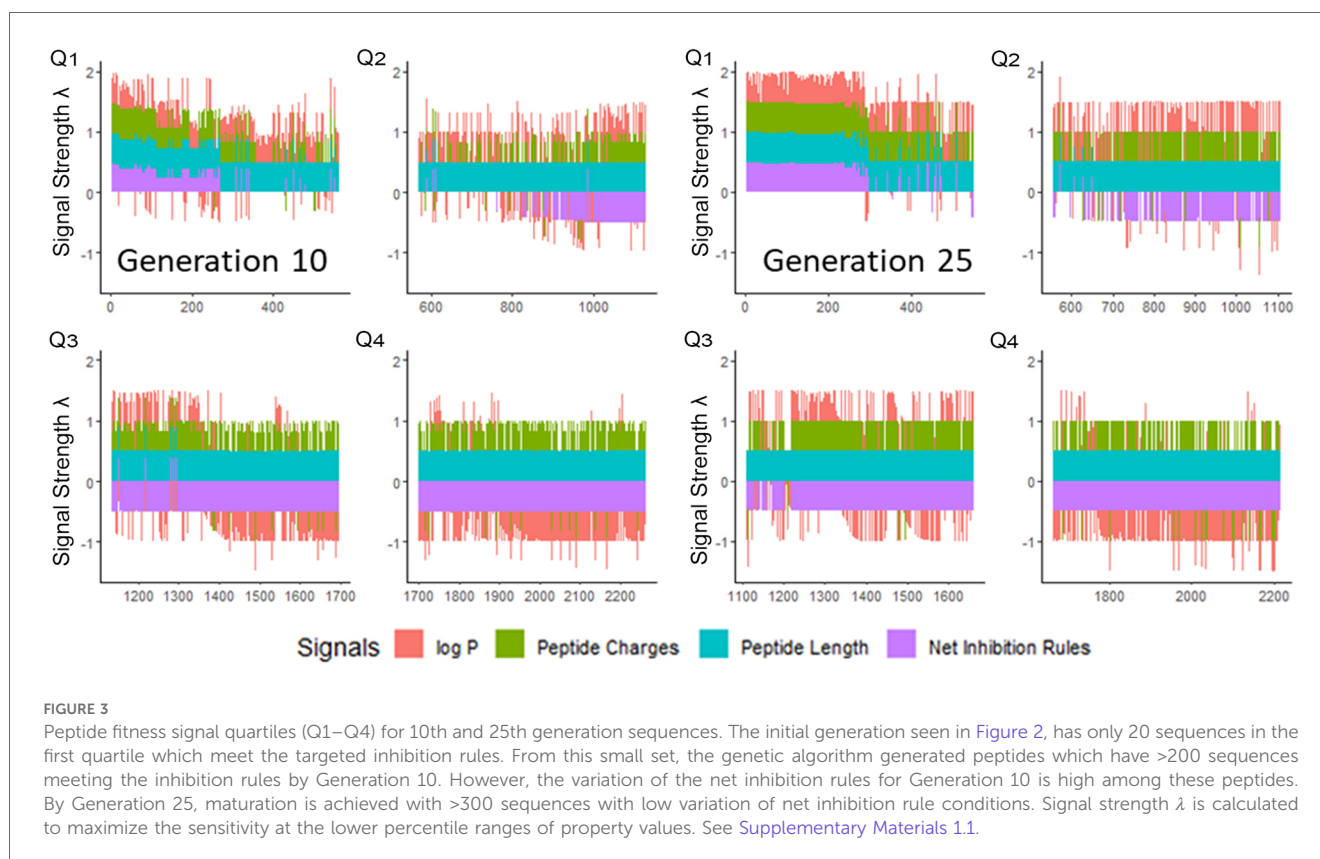
directly from the database to synthesize and to evaluate its *in vitro* activity. To our knowledge, inhibitory activity for this peptide is not reported for oral bacteria.

3.3 Identifying candidate antimicrobial peptides with enhanced relevance to existing rough sets
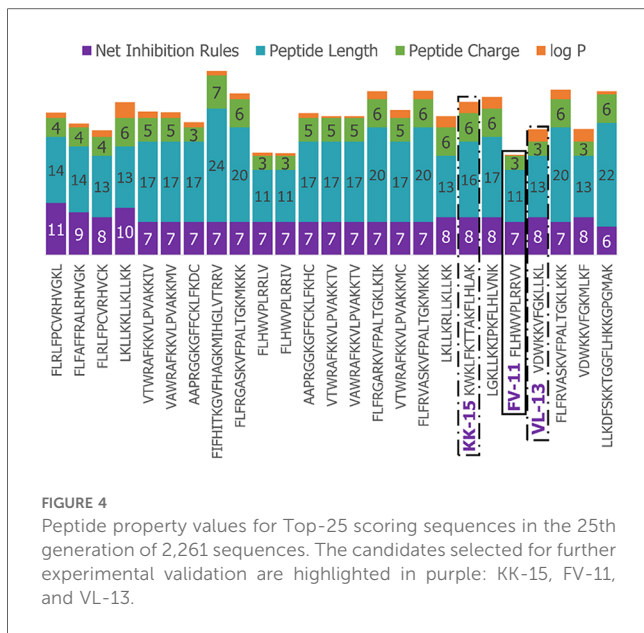
We recently developed a codon-based genetic algorithm (53) to identify antimicrobial peptides relevant to inhibition-related rough sets, as well as no growth inhibition rough sets. Our codon based-genetic algorithm uses reading frameshifts probabilistically to generate new amino acid sequences that have low sequence similarity to previously generated sequences. The codon-based operations supplement the recombination and mutation operators. We targeted peptides identified as possessing antimicrobial properties for at least one of three target bacterial strains. We further ranked the peptides by length and solubility estimates.

We next enriched the later-generation peptide sequences with sequences that met our rough set criteria. In Figure 3, this enrichment can be seen in the net-positive inhibition rule sequences where the initial generation has increased from 20 sequences (Figure 2) to 269 sequences by Generation 10. Generation 25 has a tighter distribution of net inhibition rules compared to Generation 10, which resulted in 329 peptide sequences. These sequences also contained the maximum number of net inhibition rules observed with the top scoring sequences transferred between generations. The end of 25th generation run resulted in 2,261 sequences, from these we selected three peptides to evaluate their inhibition activity: KK-15 (KWKLFKTTAKFLHLAK), FV-11 (FLHWVPLRRVV) and VL-13 (VDWKKVFGKLLKL) (Figure 4). Many of the top scoring sequences contained cysteine and they were avoided to enable future cyclization of candidates through rational placement of disulfide bonds. We chose to have a high number of rules (>7), which were satisfied by KK-15 and VL-13. We also did not select FLFAFFRALRHVGK (FK-14), LKLLKRLLKLLKK (LK-13-1) and LKLLKKLLKLLKK (LK-13-2) peptides. The LK-13 and LK-13-2 peptides were observed as close analogues of AMP1, only missing the last lysine residue or also having an arginine-lysine substitution. Since AMP1 is already included in the study, we selected other candidates. Peptide FK-14 was not selected because it has a GRAVY score of +0.72, indicating a solubility risk. However, we selected the shorter peptide, FV-11, because it has less of a solubility risk with a high score of 7. In summary, we focused our attention on potential solubility and possessing the most applicable rules. These candidates were among the top 25 sequences with 13 amino acids or less that fell within the 98th percentile of net inhibition rules. The maturation of generating new candidates with similar net inhibition rule counts was completed by the 25th generation. Noteworthy is the finding in Figure 3 that the 11th generation shows large variations of net inhibition rules among the candidates.
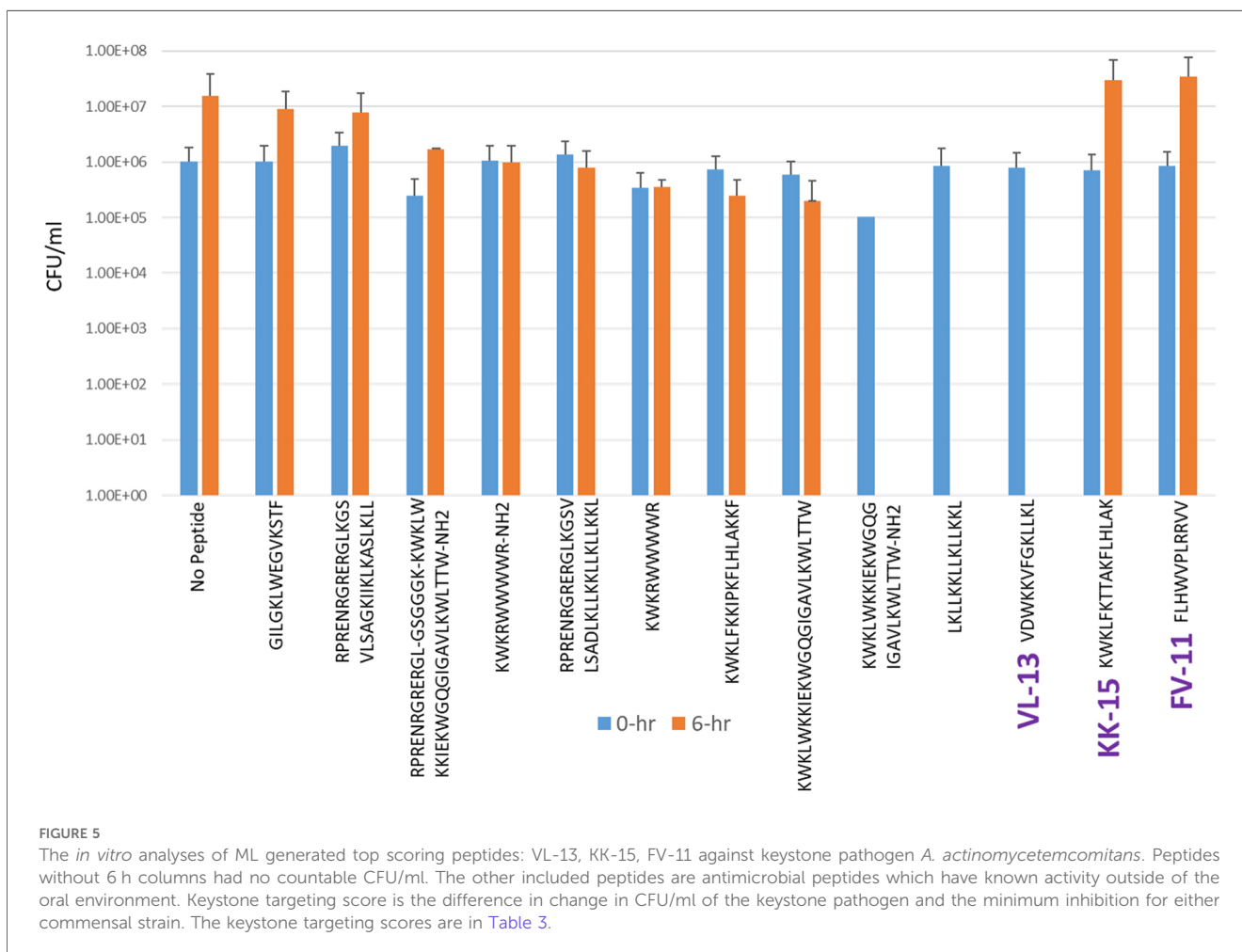

FIGURE 3
Peptide fitness signal quartiles (Q1–Q4) for 10th and 25th generation sequences. The initial generation seen in Figure 2, has only 20 sequences in the first quartile which meet the targeted inhibition rules. From this small set, the genetic algorithm generated peptides which have >200 sequences meeting the inhibition rules by Generation 10. However, the variation of the net inhibition rules for Generation 10 is high among these peptides. By Generation 25, maturation is achieved with >300 sequences with low variation of net inhibition rule conditions. Signal strength λ is calculated to maximize the sensitivity at the lower percentile ranges of property values. See Supplementary Materials 1.1.

FIGURE 4
Peptide property values for Top-25 scoring sequences in the 25th generation of 2,261 sequences. The candidates selected for further experimental validation are highlighted in purple: KK-15, FV-11, and VL-13.

Experimental evaluation of ML-generated sequences, KK-15, FV-11, and VL-13 was performed against keystone pathogen *A. actinomycetem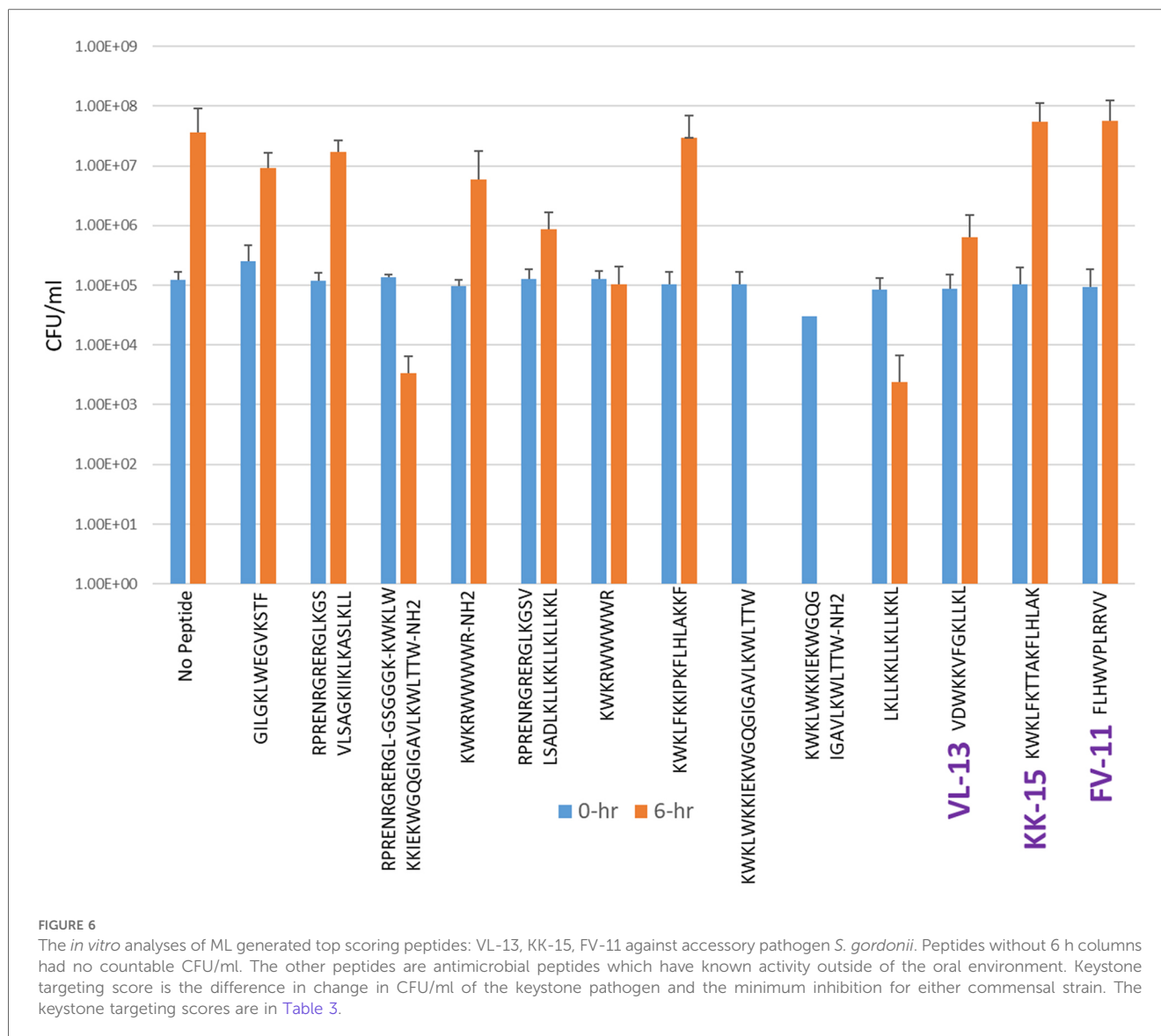comitans*, accessory/commensal *S. gordonii* and commensal *S. sanguinis* (Figures 5–7, respectively). We compared these second iteration peptides with two first iteration peptides, AMP1 and AMPa. Table 3 shows the scores for targeting keystone pathogens. VL-13 showed strong inhibition activity selectively against *A. actinomycetemcomitans* compared to the other peptides FV-11, KK-15, and KF-18. The score for targeting keystone pathogen is the difference between the number of CFU logs reduced after 6 h. VL-13 has the largest targeting score, having +7 more log reduction for the keystone pathogen *A. actinomycetemcomitans* than for the commensal *S. gordonii*. VL-13 also had a high score for targeting keystone pathogen of +4.5 when compared to the other commensal strain *S. sanguinis*. The other predicted peptides had positive targeting scores, but less than AMP1. Peptides used in the training sets AMPa resulted in the lowest targeting score of only a 0.5 log reduction difference between the two groups; AMP1 also resulted in activity against *S. sanguinis*. VL-13 was validated for predicted targeted activity against *A. actinomycetemcomitans*.

The superimposed structures generated for VL-13 are given in Figure 8. From the structures, a peptide structural feature with a high amount of hydrophobicity is indicated as a low-energy rotation barrier compared to the more hydrophilic structural features in the antimicrobial peptide. This hydrophobicity feature
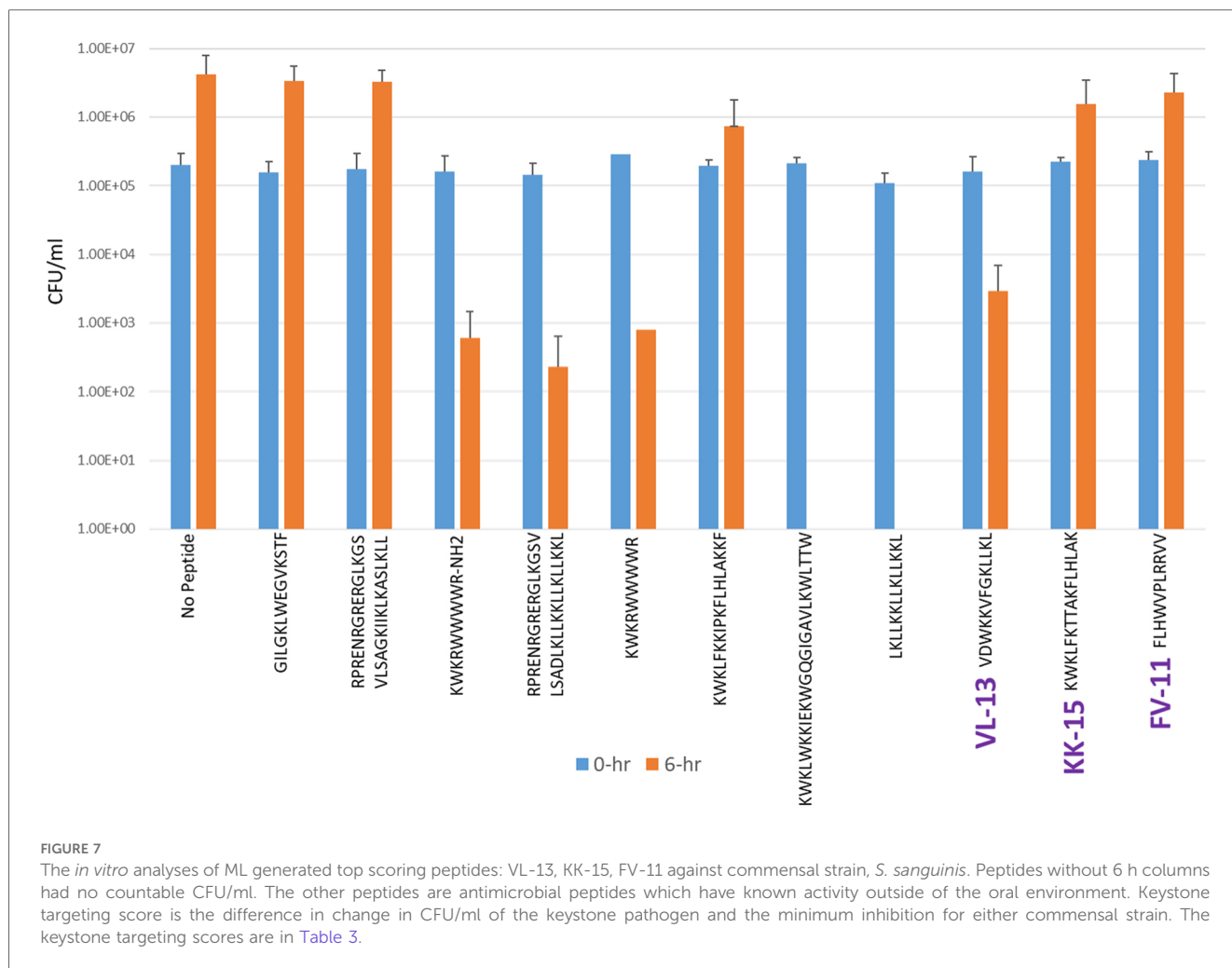

FIGURE 5
The *in vitro* analyses of ML generated top scoring peptides: VL-13, KK-15, FV-11 against keystone pathogen *A. actinomycetemcomitans*. Peptides without 6 h columns had no countable CFU/ml. The other included peptides are antimicrobial peptides which have known activity outside of the oral environment. Keystone targeting score is the difference in change in CFU/ml of the keystone pathogen and the minimum inhibition for either commensal strain. The keystone targeting scores are in Table 3.

FIGURE 6
The *in vitro* analyses of ML generated top scoring peptides: VL-13, KK-15, FV-11 against accessory pathogen *S. gordonii*. Peptides without 6 h columns had no countable CFU/ml. The other peptides are antimicrobial peptides which have known activity outside of the oral environment. Keystone targeting score is the difference in change in CFU/ml of the keystone pathogen and the minimum inhibition for either commensal strain. The keystone targeting scores are in Table 3.

is also discussed in the rough set theory sequence feature example with the JACR890101 3-amino acid window in Figure 9 and in Supplementary Figure S4.

As a parallel method to assess the progress of the CLN-MLEM2 model, we predicted the performance of additional peptides reported as potential therapeutics for oral bacteria and oral biofilms as the test set (61). In Supplementary Tables S3 and S4, we evaluate the performance of the rules specific to two keystone pathogens, *A. actinomycetemcomitans* and *P. gingivalis*, respectively. In Supplementary Table S5, we evaluate the performance of the rules specific to the commensal/accessory pathogen *S. gordonii*. The false discovery rate was found to be low for *A. actinomycetemcomitans* and the commensal pathogen, *S. gordonii* (Supplementary Tables S3 and S5) using prediction performance. Overall, the model performance shows good prediction for positive activity of the sequences for selected keystone pathogen and the commensal/accessory pathogen.

4 Discussion

4.1 Shifting focus toward pathogen *A. actinomycetemcomitans*

Treatment of periodontal and peri-implant diseases necessitates a targeted, polymicrobial approach that sufficiently inhibits progression of the disease state without detrimentally impacting commensal and otherwise opportunistic species. To address this, we attempted to develop a ML-enabled targeting AMP prediction approach. To train the CLN-MLEM2 model, an initial round of AMPs was tested against keystone pathogens *P. gingivalis* and *A. actinomycetemcomitans*, and opportunistic pathogen *S. gordonii*. Both *P. gingivalis* and *A. actinomycetemcomitans* display pathogenic characteristics that contribute significantly to biofilm prevalence and virulence (11). Although these species are referred as key stone pathogens, their involvement appears in different stages of the disease

FIGURE 7

The *in vitro* analyses of ML generated top scoring peptides: VL-13, KK-15, FV-11 against commensal strain, *S. sanguinis*. Peptides without 6 h columns had no countable CFU/ml. The other peptides are antimicrobial peptides which have known activity outside of the oral environment. Keystone targeting score is the difference in change in CFU/ml of the keystone pathogen and the minimum inhibition for either commensal strain. The keystone targeting scores are in Table 3.

progression. The ML-based tunability in this paper presents an opportunity to control the disease progression by targeting different keystones and other microbiome components.

A. actinomycetemcomitans is unique for its association with localized, aggressive cases of peri-implantitis and periodontitis —specifically those in younger individuals under the age of 35 (13, 62). This is extremely problematic not only for the livelihood of impacted individuals, but also for the whole of human health. Disease occurrence in younger individuals entails longer timelines of recurrent infection and treatment cycles. This makes these individuals, and thus bacterial species present in them, prime candidates for emergent bacterial resistance and innate microbiome depletion. It also makes them at heightened risk for implant installation failure and loss of oral functionality as the tissue and

TABLE 3 Net inhibition rule counts by bacterial species.

Name	Sequence	Aa NC	Sg NC	Pg NC	Aa LR	Sg LR	Ss LR	Keystone targeting
AMPa	KWKLWKKIEKWGQGIGAVLKWLTTW	2	4	−15	5	4.5	ND	+0.5
AMP1	LKLLKKLLKLLKKL	2	6	2	6	2	5	+4.0
KK-15	KWKLFKTTAKFLHLAK	2	4	2	−1.5	−3	−1	+1.5
FV-11	FLHWVPLRRVV	0	4	3	−1.5	−3	−1	+1.5
VL-13	VDWKKVFGKLLKL	2	6	0	6	−1	1.5	+7.0

Positive scores in the "NC" columns indicate predicted antibacterial activity for the strain. Zero or negative scores in these columns indicate no activity predicted. The "LR" columns indicate the decimal log change after 6 h of incubation with the peptide indicated in Figures 5–7. Keystone targeting is the difference between the *A. actinomycetemcomitans* (Aa) log reduction and the minimum of the *S. gordonii* (Sg) log reduction or the *S. sanguinis* (Ss) log reduction. Higher log reduction scores indicate more growth inhibition. Negative log reduction indicates growth during incubation instead of inhibition. Higher targeting scores indicate better targeting performance. The first two rows are first iteration sequences with known activity used to compare the targeting performance of our second iteration peptides generated by the codon-based genetic algorithm. The *Aa* inhibition rules describe transferred activity for VL-13 but not for KK-15, while the *Sg* inhibition rules did not describe transferred activity in any of our second iteration peptides. NC, net count of inhibition rules; LR, log reduction; and ND, not determined.
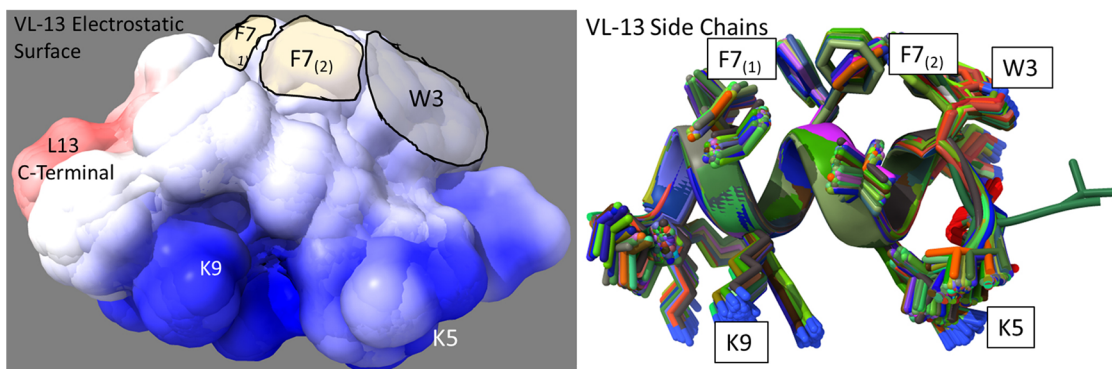
FIGURE 8
Ensemble structure description of VL-13 (VDWKKVFGKLLKL). The hydrophobic feature of VFG (see Figure 10) has very high solvent accessibility. This indicates high accessibility for potential binding partners. The locations of the side chains are stable through the ensemble. Only the Phenylalanine 7 seems to have multiple distinct clusters of orientations, which may indicate a local rotation between two smaller vibrations of the aromatic ring. The conserved backbone structure simplifies structure-function hypotheses for peptides. This lower folding entropy simplifies structure-function studies compared to using AMPa or AMP1. Superposition of 192 PEP-FOLD3 (66) structures aligned with MatchAlign (64) in UCSF Chimera (65).

bone surrounding their dental implants degrade. In our ML-predicted novel peptide generation, we therefore focused on assessment of the antimicrobial activity against *A. actinomycetemcomitans, S. gordonii,* and the commensal *S. sanguinis*.

FIGURE 9
JACR890101 hydrophobicity property trend as both a single amino acid property (blue) and as the sum of three consecutive amino acids (orange). In (**A**), VL-13 meets Rule 1 Condition 2 in Table 2 at residue F7 where the 3-aa window falls within the rule bounds (dashed lines). In (**B**), KF-18 does not have a hydropathy feature of the 3-aa sum that meets this condition. The dashed lines indicate the Rule 1 Condition 2 upper and lower boundaries in Table 2. Figure S4 compares Rule 2 and Rule 3 conditions in Table 2 with Rule 1.

4.2 Rule application and predicted vs. experimental antimicrobial activity in VL-13
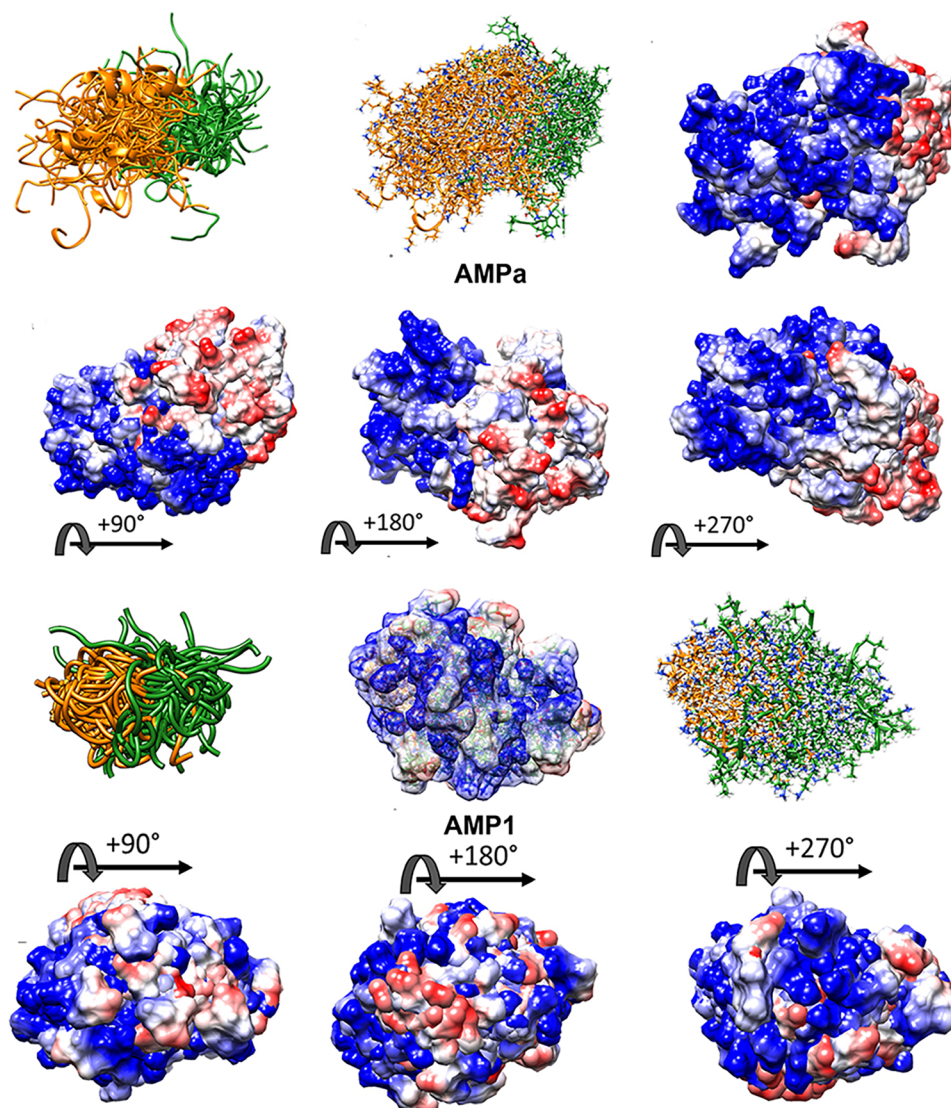
The rules in Table 2 show amphipathic descriptions of targeting growth inhibition of *A. actinomycetemcomitans*, both for descriptions of which peptides inhibit and which peptides do not inhibit growth. In Supplementary Figures S4–S9, we describe the feature characteristics for sequences that also have rule membership for rules which apply to VL-13 in Table 2. Our second-iteration decision system made incorrect predictions for KK-15 and FV-11. The next iteration of the decision system will avoid these incorrect predictions and generalize with these new cases to determine new decision boundaries. The next decision system draws more accurate boundaries for negative results but does not move boundaries when all cases within a sub-domain are accurate. This attribute of our transparent decision system shows the system development value of having test cases. This way the decision system challenges the rough set membership boundaries of the previous iteration rather than having test cases which are very likely to be accurate or peptides which have no rough set theory membership. Testing truly random peptides which do not belong to either rules for or against activity may not build on the knowledge gained from previously tested peptides. However, any peptide test results would start a new knowledge base to build on in future studies.

We used a nested-rough set rules approach by evaluating the rules generated when classifying peptides for having any antibacterial activity with rough set rules for peptides having targeted activity. This nested methodology is an example of transfer learning in our machine learning method. Further targeting of the activity of the peptides incorporating different design goals can also adapt this nesting approach in future studies.

In the first iteration of our codon-based genetic algorithm, generated a novel antimicrobial peptide against *S. epidermidis* (53). In our second iteration of this system in this work, we applied a nested version of our decision system to the *in vitro* oral environment for the mitigation of the progression of

FIGURE 10
High folding entropy of AMPa (A) and AMP1 (B) compared to VL-13. The orange ribbon is the first half of each sequence and green is the second half of the corresponding sequence. The divergent ribbons indicate a folding ensemble with many different orientations of backbone structure having similar energy values. In Figure 9, the conserved backbone shape among superposition structures indicates a relatively large folding energy barrier to other backbone shapes. While the high structural entropy complicates the structure-function analysis of these peptides, surface interactions are still facilitated through electrostatic means. The electrostatic surface indicates highly charged surface which is mostly electropositive (blue) for the first half of AMPa with most electronegative (red) surface for the second half of AMPa. The distribution of electronegativity for AMP1 is more heterogeneous compared to AMPa. Superposition of 100 PyRosetta structures (63) aligned with MatchAlign (64) in UCSF Chimera (65).

peri-implantitis. In this study, an antimicrobial peptide, VL-13, was demonstrated targeted growth inhibition against an oral keystone pathogen *A. actinomycetemcomitans* without inhibiting accessory/commensal species, *S. gordonii* (Figure 5).

VL-13 is both a positive result for this decision system iteration for activity against the keystone pathogen and a negative result for activity against the accessory/commensal pathogen *S. gordonii*. We designed peptides to meet multiple targeting classes, inferring that if the targeting criteria are relatively difficult to describe, finding one working targeting description would likely come before two working targeting descriptions. Our result does not limit building on the targeting criteria which could be introduced as new design

characteristics. In future studies, we can use both experimental results to draw boundaries which include VL-13 for *A. actinomycetemcomitans* inhibition and exclude VL-13 for *S. gordonii* inhibition. Our method learns boundaries from inhibition and non-inhibition results for each of the tested peptides. Superimposed structures of AMPa and AMP1 were evaluated to gain further insight, as both sequences were used in the training set based on their demonstrated activity (Figure 10). We show that the folding dynamics of these structures have relatively high folding entropy compared to VL-13, shown in Figure 8.

To discuss what the rough set theory rules imply about inhibition activity for the keystone pathogens, we begin with one

AAindex1 property. The three-amino acid window JACR890101 is the amino-acid wise component of hydrophobic interactions of amino acids at the bilayer (Figure 9, Supplementary Figure S4). While further study can investigate if this feature is necessary for the peptide's activity, we also note that this feature may be related to some motion allowing the peptide to efficiently attack/ bypass the membrane of *A. actinomycetemcomitans*, which is a gram-negative pathogen. In Rules 1 and 3 in Table 2, this feature was selected as a tripeptide window or a mean (see Figure 10). The rules both select tripeptide windows for this feature, in which Rule 1 has a left-shifted range (from 0.89–1.32 to 0.615–0.935). This window is very hydrophobic under the conditions of the bilayer (see Supplementary Figure S4–S9). The overall mean of the peptide for this property in Rule 1 was between −1.2 and −1.8, indicating that these tripeptide windows of greater than 0 are not likely to be common among peptides which are active against *A. actinomycetemcomitans*. The combination of these rules describes a preliminary description of amphipathy that is useful to inhibit the pathogen's growth. Having varying hydrophobicity has long been studied for antimicrobial peptides (57, 67, 68). Further non-linear boundary ranges are shown in Supplementary Figures S1–S6. The CLN-MLEM2 rough set theory method has added a process to identifying which hydrophobicity features relate to inhibiting the growth for a specific pathogen. The target that the antimicrobial peptide is affecting with the hydrophobicity feature is unknown.

4.3 Testing performance, limitations and future perspectives

This paper used data from a review article on peptides as therapeutics targeting oral bacteria by Sztukowska et al. (61) as a test set. We recognize there exist previous reviews of antimicrobial peptides in the oral environment with peptide sequences that were not included in our training set (69–72). The reasons to not include all experimentally tested peptides in the model training phase is to continue to develop our model through testing performance beyond training performance. Having peptides in the literature that are not included in our training set allows for testing the performance evaluation of our model. Indeed, the ML model should be validated with experimental results.

The literature peptide activity (61) was used as a test set for our rule sets describing targeted activity for the two keystone pathogens and the commensal/accessory pathogen. The first keystone pathogen rules for *A. actinomycetemcomitans* had higher accuracy than the commensal pathogen rules for *S. gordonii*, indicating a closer relationship between trained sequences and tested sequences by the rule set descriptions related to Supplementary Table S3 than to the rule set descriptions in Supplementary Table S5. The rule set for the second keystone pathogen had a high false discovery rate, indicating the rule set related to Supplementary Table S4 for *P. gingivalis* has a higher chance of leading to unexpected negative inhibitory results. These results also confirm the critical iterations of the enhanced rule sets for different pathogens including keystones such as

P. gingivalis. Our future studies focus on developing rule sets leading to low false discovery rates (<10%), then we will incorporate this bacterial strain in our ML-candidate evaluation process. Our ML-models are re-trained between iterations to avoid carrying over recognized performance errors.

Hydrophobicity trends among peptides are discovered during training and applied to the selection of new peptides. Our method is identifying database peptides that have similar hydropathy features to the peptides that are confirmed for their antimicrobial activity using *in vitro* evaluation against our targeted bacterial strains. These activities build better information for peptides with similar hydropathy features, rather than using brute-force testing on all related peptides to see if their inhibition activity changes in a useful way between our targeted bacterial strains. Incorporation of new information from *in vitro* results strengthens our approach for either positive or negative results. We recognize the fact that the peptides studied here need to be further evaluated for their toxicity. Our future work will include experimental evaluation of the potential peptide toxicity as well as other clinically relevant properties of these peptides. We also plan to incorporate multiple factors into the ML design of targeted peptides. Broader types of data, such as toxicity and stability, can be tightly integrated together with inhibitory activity because each new factor will have its own rule sets to simultaneously constrain the sequences the genetic algorithm will target. These distinct descriptions can be integrated with the inhibitory activity descriptions by building separate rule sets and using our genetic algorithm to find examples that combine multiple rule sets for distinct descriptions. The transparent approach of rough set theory allows us to re-classify sequences based on new results in a short time without using GPU-parallelized computation. Therefore, re-training of the entire dataset is feasible when integrating data between sources and extending data sources.

During the training, hydrophobicity trends among peptides were discovered and applied to the selection of new peptides. Our validation tests were run on three top candidates, one of which had the highest inhibition predictability score against the keystone pathogen, *A. actinomycetemcomitans*. Our *in vitro* test results demonstrated that peptide VL-13 is an antimicrobial peptide with the largest change of inhibition compared between the keystone pathogen and the accessory-commensal species, *S. gordonii*. Further, VL-13 has an added advantage in possessing the largest change of inhibition between the keystone pathogen and the commensal species, *S. sanguinis*.

5 Conclusion

We developed a transparent machine learning model with an iterative *in vitro* experimental validation approach to design antimicrobial peptides that selectively target keystone pathogens believed to play critical roles during biofilm dysbiogenesis leading to progression of peri-implant disease. The transparency of our machine learning method allows us to compare the discovered relationships with trends in literature. The non-linear nature of the boundaries also provides for rapid learning between iterations

of new sequences to explore. Through our transparent machine learning methods, we demonstrate the finding that antimicrobial peptide VL-13 inhibits the keystone pathogen *A. actinomycetemcomitans*, while the ML model also provided better learning descriptions for finding antimicrobial peptides inhibitory against the accessory pathogen *S. gordonii*. VL-13 was the most targeted antimicrobial peptide sequence we tested, with high levels of inhibition against *A. actinomycetemcomitans* and minimal impact on *S. sanguinis* and *S. gordonii*. The descriptions used to select the VL-13 sequence from our candidate sequence generation method were learned from tested peptides in this study combined with known sequences in the literature. From antimicrobial peptide sequences available in iAMP-2l antimicrobial peptide database, we chose KF-18 to test for activity. KF-18 has strong homology with AMPa, which we demonstrated is inhibitory toward *A. actinomycetemcomitans*. Since this peptide and a second generated candidate with homology to AMPa (KK-15) did not test as active, we have further insight into the features of AMPa which relate to inhibition activity for *A. actinomycetemcomitans* and *S. gordonii*. Future work aims to generate non-inhibitory sequences for commensal and accessory species, such as *S. gordonii*, while retaining inhibitory activity against keystone pathogens. Developing an engineering approach to iteratively discover targeted antimicrobials is a robust method for potential therapeutic treatment for peri-implant biofilm infections and thus to reduce the resulting host response of hyper-inflammation during disease progression. With increasing use of implants to replace missing teeth and support oral function, the number of patients suffering from peri-implantitis will continue to increase. It is critical to find targeted approaches to address this complex infectious disease. Antimicrobial peptides with selective bioactivity against keystone pathogens, while preserving commensals could be the next generation therapy that will respond to this urgent healthcare need.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

KB contributed to conceptualization, developed the computational methods and generated the associated data. NT

conducted the wet laboratory experiments. KC contributed computational analyses. KB, NT and KC contributed to manuscript preparation. CT, MLS and CC contributed to the conceptualization, computational and experimental design, analyses and writing as well as the supervision and the funding acquisition. All authors contributed to the review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdmed.2024.1372534/full#supplementary-material

References

1. Iacono VJ, Bassir SH, Wang HH, Myneni SR. Peri-implantitis: effects of periodontitis and its risk factors—a narrative review. *Front Oral Maxillofac Med*. (2022) 5:27. doi: 10.21037/fomm-21-63

2. Berglundh T, Wennström JL, Lindhe J. Long-term outcome of surgical treatment of peri-implantitis. A 2–11-year retrospective study. *Clin Oral Implants Res*. (2018) 29 (4):404–10. doi: 10.1111/clr.13138

3. Lee CT, Huang YW, Zhu L, Weltman R. Prevalences of peri-implantitis and peri-implant mucositis: systematic review and meta-analysis. *J Dent*. (2017) 62:1–12. doi: 10.1016/j.jdent.2017.04.011

4. Ardila CM, Vivares-Builes AM. Antibiotic resistance in patients with peri-implantitis: a systematic scoping review. *Int J Environ Res Public Health*. (2022) 19 (23):15609. doi: 10.3390/ijerph192315609

5. Roccuzzo A, Stähli A, Monje A, Sculean A, Salvi GE. Peri-implantitis: a clinical update on prevalence and surgical treatment outcomes. *J Clin Med*. (2021) 10 (5):1107. doi: 10.3390/jcm10051107

6. Schwarz F, Derks J, Monje A, Wang HL. Peri-implantitis. *J Clin Periodontol*. (2018) 45:S246–66. doi: 10.1111/jcpe.12954

7. Romandini M, Shin HS, Romandini P, Laforí A, Cordaro M. Hormone-related events and periodontitis in women. *J Clin Periodontol*. (2020) 47(4):429–41. doi: 10.1111/jcpe.13248

8. Wicaksono WA, Erschen S, Krause R, Müller H, Cernava T, Berg G. Enhanced survival of multi-species biofilms under stress is promoted by low-abundant but antimicrobial-resistant keystone species. *J Hazard Mater*. (2022) 422:126836. doi: 10.1016/j.jhazmat.2021.126836

9. Costerton JW. Introduction to biofilm. *Int J Antimicrob Agents*. (1999) 11(3–4):217–21; discussion 37–9. doi: 10.1016/s0924-8579(99)00018-7

10. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. *Nat Rev Microbiol*. (2012) 10(10):717–25. doi: 10.1038/nrmicro2873

11. Hajishengallis G, Lamont RJ. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (psd) model of periodontal disease etiology. *Mol Oral Microbiol*. (2012) 27(6):409–19. doi: 10.1111/j.2041-1014.2012.00663.x

12. Costalonga M, Herzberg MC. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunol Lett*. (2014) 162(2 Pt A):22–38. doi: 10.1016/j.imlet.2014.08.017

13. Fine DH, Markowitz K, Furgang D, Fairlie K, Ferrandiz J, Nasri C, et al. *Aggregatibacter actinomycetemcomitans* and its relationship to initiation of localized aggressive periodontitis: longitudinal cohort study of initially healthy adolescents. *J Clin Microbiol*. (2007) 45(12):3859–69. doi: 10.1128/jcm.00653-07

14. Zhu B, Macleod LC, Newsome E, Liu J, Xu P. *Aggregatibacter actinomycetemcomitans* mediates protection of *Porphyromonas Gingivalis* from *Streptococcus Sanguinis* hydrogen peroxide production in multi-Species biofilms. *Sci Rep*. (2019) 9(1):4944. doi: 10.1038/s41598-019-41467-9

15. Pollanen MT, Paino A, Ihalin R. Environmental stimuli shape biofilm formation and the virulence of periodontal pathogens. *Int J Mol Sci*. (2013) 14(8):17221–37. doi: 10.3390/ijms140817221

16. Lamont RJ, Koo H, Hajishengallis G. The oral Microbiota: dynamic communities and host interactions. *Nat Rev Microbiol*. (2018) 16(12):745–59. doi: 10.1038/s41579-018-0089-x

17. Brown JL, Yates EA, Bielecki M, Olczak T, Smalley JW. Potential role for *Streptococcus Gordonii*-derived hydrogen peroxide in heme acquisition by *Porphyromonas Gingivalis*. *Mol Oral Microbiol*. (2018) 33(4):322–35. doi: 10.1111/omi.12229

18. Yu XL, Chan Y, Zhuang L, Lai HC, Lang NP, Keung Leung W, et al. Intra-oral single-site comparisons of periodontal and peri-implant microbiota in health and disease. *Clin Oral Implants Res*. (2019) 30(8):760–76. doi: 10.1111/clr.13459

19. Cheng WC, van Asten SD, Burns LA, Evans HG, Walter GJ, Hashim A, et al. Periodontitis-associated pathogens *P. Gingivalis* and *A. Actinomycetemcomitans* activate human Cd14(+) monocytes leading to enhanced Th17/il-17 responses. *Eur J Immunol*. (2016) 46(9):2211–21. doi: 10.1002/eji.201545871

20. Mishra B, Reiling S, Zarena D, Wang G. Host defense antimicrobial peptides as antibiotics: design and application strategies. *Curr Opin Chem Biol*. (2017) 38:87–96. doi: 10.1016/j.cbpa.2017.03.014

21. Huan Y, Kong Q, Mou H, Yi H. Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front Microbiol*. (2020) 11:582779. doi: 10.3389/fmicb.2020.582779

22. Dini I, De Biasi MG, Mancusi A. An overview of the potentialities of antimicrobial peptides derived from natural sources. *Antibiotics (Basel)*. (2022) 11 (11):1483. doi: 10.3390/antibiotics11111483

23. Wibowo D, Zhao CX. Recent achievements and perspectives for large-scale recombinant production of antimicrobial peptides. *Appl Microbiol Biotechnol*. (2019) 103(2):659–71. doi: 10.1007/s00253-018-9524-1

24. Renaud S, Mansbach RA. Latent spaces for antimicrobial peptide design. *Digital Discovery*. (2023) 2(2):441–58. doi: 10.1039/D2DD00091A

25. Azmat M, Ghalandari B, Jessica J, Xu Y, Li X, Su W, et al. Pepdred: De Novo peptide design with strong binding affinity for target protein. *Anal Chem*. (2023) 95 (33):12264–72. doi: 10.1021/acs.analchem.3c01057

26. Martínez OF, Duque HM, Franco OL. Peptidomimetics as potential anti-virulence drugs against resistant bacterial pathogens. *Front Microbiol*. (2022) 13:831037. doi: 10.3389/fmicb.2022.831037

27. Mahlapuu M, Håkansson J, Ringstad L, Björn C. Antimicrobial peptides: an emerging category of therapeutic agents. *Front Cell Infect Microbiol*. (2016) 6:194. doi: 10.3389/fcimb.2016.00194

28. Griffith A, Mateen A, Markowitz K, Singer SR, Cugini C, Shimizu E, et al. Alternative antibiotics in dentistry: antimicrobial peptides. *Pharmaceutics*. (2022) 14 (8):1679. doi: 10.3390/pharmaceutics14081679

29. Fischer NG, Münchow EA, Tamerler C, Bottino MC, Aparicio C. Harnessing biomolecules for bioinspired dental biomaterials. *J Mater Chem B*. (2020) 8 (38):8713–47. doi: 10.1039/D0TB01456G

30. Holmberg KV, Abdolhosseini M, Li Y, Chen X, Gorr SU, Aparicio C. Bio-inspired stable antimicrobial peptide coatings for dental applications. *Acta Biomater*. (2013) 9(9):8224–31. doi: 10.1016/j.actbio.2013.06.017

31. Xie S-X, Song L, Yuca E, Boone K, Sarikaya R, VanOosten SK, et al. Antimicrobial peptide–polymer conjugates for dentistry. *ACS Appl Polym Mater*. (2020) 2(3):1134–44. doi: 10.1021/acsapm.9b00921

32. Spencer P, Ye Q, Misra A, Chandler JR, Cobb CM, Tamerler C. Engineering peptide-polymer hybrids for targeted repair and protection of cervical lesions. *Front Dent Med*. (2022) 3:1007753. doi: 10.3389/fdmed.2022.1007753

33. Yuca E, Xie SX, Song L, Boone K, Kamathewatta N, Woolfolk SK, et al. Reconfigurable dual peptide tethered polymer system offers a synergistic solution for next generation dental adhesives. *Int J Mol Sci*. (2021) 22(12):6552. doi: 10.3390/ijms22126552

34. Wisdom C, Chen C, Yuca E, Zhou Y, Tamerler C, Snead ML. Repeatedly applied peptide film kills Bacteria on dental implants. *Jom (1989)*. (2019) 71(4):1271–80. doi: 10.1007/s11837-019-03334-w

35. Wisdom EC, Zhou Y, Chen C, Tamerler C, Snead ML. Mitigation of peri-implantitis by rational design of bifunctional peptides with antimicrobial properties. *ACS Biomater Sci Eng*. (2020) 6(5):2682–95. doi: 10.1021/acsbiomaterials.9b01213

36. Zhang X, Geng H, Gong L, Zhang Q, Li H, Zhang X, et al. Modification of the surface of titanium with multifunctional chimeric peptides to prevent biofilm formation via inhibition of initial colonizers. *Int J Nanomedicine*. (2018) 13:5361–75. doi: 10.2147/IJN.S170819

37. Godoy-Gallardo M, Mas-Moruno C, Yu K, Manero JM, Gil FJ, Kizhakkedathu JN, et al. Antibacterial properties of Hlf1-11 peptide onto Titanium surfaces: a comparison study between silanization and surface initiated polymerization. *Biomacromolecules*. (2015) 16(2):483–96. doi: 10.1021/bm501528x

38. Godoy-Gallardo M, Mas-Moruno C, Fernández-Calderón MC, Pérez-Giraldo C, Manero JM, Albericio F, et al. Covalent immobilization of Hlf1-11 peptide on a Titanium surface reduces bacterial adhesion and biofilm formation. *Acta Biomater*. (2014) 10(8):3522–34. doi: 10.1016/j.actbio.2014.03.026

39. Yazici H, Fong H, Wilson B, Oren EE, Amos FA, Zhang H, et al. Biological response on a Titanium implant-grade surface functionalized with modular peptides. *Acta Biomater*. (2013) 9(2):5341–52. doi: 10.1016/j.actbio.2012.11.004

40. Zhou Y, Snead ML, Tamerler C. Bio-inspired hard-to-soft interface for implant integration to bone. *Nanomedicine*. (2015) 11(2):431–4. doi: 10.1016/j.nano.2014.10.003

41. Porto WF, Pires AS, Franco OL. Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnol Adv*. (2017) 35 (3):337–49. doi: 10.1016/j.biotechadv.2017.02.001

42. Wang G, Li X, Wang Z. Apd3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res*. (2016) 44(D1):D1087–93. doi: 10.1093/nar/gkv1278

43. Waghu FH, Idicula-Thomas S. Collection of antimicrobial peptides database and its derivatives: applications and beyond. *Protein Sci*. (2020) 29(1):36–42. doi: 10.1002/pro.3714

44. Ye G, Wu H, Huang J, Wang W, Ge K, Li G, et al. Lamp2: a major update of the database linking antimicrobial peptides. *Database (Oxford)*. (2020) 2020:baaa061. doi: 10.1093/database/baaa061

45. Fan L, Sun J, Zhou M, Zhou J, Lao X, Zheng H, et al. Dramp: a comprehensive data repository of antimicrobial peptides. *Sci Rep*. (2016) 6:24482. doi: 10.1038/srep24482

46. Azam MW, Kumar A, Khan AU. Acd: antimicrobial chemotherapeutics database. *PLoS One*. (2020) 15(6):e0235193. doi: 10.1371/journal.pone.0235193

47. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers*. (2021) 25(3):1315–60. doi: 10.1007/s11030-021-10217-3

48. Liu Z, Jin J, Cui Y, Xiong Z, Nasiri A, Zhao Y, et al. Deepseqpanii: an interpretable recurrent neural network model with attention mechanism for peptide-hla class ii binding prediction. *IEEE/ACM Trans Comput Biol Bioinform*. (2022) 19(4):2188–96. doi: 10.1109/TCBB.2021.3074927

49. Pertseva M, Gao B, Neumeier D, Yermanos A, Reddy ST. Applications of machine and deep learning in adaptive immunity. *Annu Rev Chem Biomol Eng*. (2021) 12:39–62. doi: 10.1146/annurev-chembioeng-101420-125021

50. Wang C, Garlick S, Zloh M. Deep learning for novel antimicrobial peptide design. *Biomolecules*. (2021) 11(3):471. doi: 10.3390/biom11030471

51. Khabbaz H, Karimi-Jafari MH, Saboury AA, BabaAli B. Prediction of antimicrobial peptides toxicity based on their physico-chemical properties using

machine learning techniques. *BMC Bioinform*. (2021) 22(1):549. doi: 10.1186/s12859-021-04468-y

52. Boone K, Camarda K, Spencer P, Tamerler C. Antimicrobial peptide similarity and classification through rough set theory using physicochemical boundaries. *BMC Bioinform*. (2018) 19(1):469. doi: 10.1186/s12859-018-2514-6

53. Boone K, Wisdom C, Camarda K, Spencer P, Tamerler C. Combining genetic algorithm with machine learning strategies for designing potent antimicrobial peptides. *BMC Bioinform*. (2021) 22(1):239. doi: 10.1186/s12859-021-04156-x

54. Edlund A, Yang Y, Hall AP, Guo L, Lux R, He X, et al. An in vitro biofilm model system maintaining a highly reproducible species and metabolic diversity approaching that of the human oral microbiome. *Microbiome*. (2013) 1(1):25. doi: 10.1186/2049-2618-1-25

55. Xiao X, Wang P, Lin WZ, Jia JH, Chou KC. Iamp-2l: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal Biochem*. (2013) 436(2):168–77. doi: 10.1016/j.ab.2013.01.019

56. Kibinge N, Ikeda S, Ono N, Altaf-Ul-Amin M, Kanaya S. Integration of residue attributes for sequence diversity characterization of terpenoid enzymes. *BioMed Res Int*. (2014) 2014:753428. doi: 10.1155/2014/753428

57. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. Aaindex: amino acid index database, progress report 2008. *Nucleic Acids Res*. (2008) 36(Database issue):D202–5. doi: 10.1093/nar/gkm998

58. Grzymala-Busse JW. Mining numerical data—a rough set approach. In: Kryszkiewicz M, Peters JF, Rybinski H, Skowron A, editors. *Proceedings of the International Conference on Rough Sets and Intelligent Systems Paradigms*. Berlin, Heidelberg: Springer-Verlag (2007). p. 12–21. doi: 10.1007/978-3-642-11479-3_1

59. Yao YY, Wong SKM, Lin TY. A review of rough set models. In: Lin TY, Cercone N, editors. *Rough Sets and Data Mining: Analysis of Imprecise Data*. Boston, MA: Springer US (1997). p. 47–75. doi: 10.1007/978-1-4613-1461-5_3

60. Pawlak Z. Rough sets: theoretical aspects of reasoning about data. *Theory and Decision Library Series D, System Theory, Knowledge Engineering, and Problem Solving*. Vol. 9. Netherlands: Springer Netherlands (1991). p. 1–19. doi: 10.1007/978-94-011-3534-4

61. Sztukowska MN, Roky M, Demuth DR. Peptide and non-peptide mimetics as potential therapeutics targeting oral Bacteria and oral biofilms. *Mol Oral Microbiol*. (2019) 34(5):169–82. doi: 10.1111/omi.12267

62. Claesson R, Höglund-Åberg C, Haubek D, Johansson A. Age-related prevalence and characteristics of *Aggregatibacter Actinomycetemcomitans* in periodontitis patients living in Sweden. *J Oral Microbiol*. (2017) 9(1):1334504. doi: 10.1080/20002297.2017.1334504

63. Chaudhury S, Lyskov S, Gray JJ. Pyrosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinform*. (2010) 26 (5):689–91. doi: 10.1093/bioinformatics/btq007

64. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with ucsf chimera. *BMC Bioinform*. (2006) 7(1):339. doi: 10.1186/1471-2105-7-339

65. 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(13):1605–12. doi: 10.1002/jcc.20084

66. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. Pep-Fold3: faster De Novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res*. (2016) 44(W1):W449–54. doi: 10.1093/nar/gkw329

67. Chen CH, Starr CG, Troendle E, Wiedman G, Wimley WC, Ulmschneider JP, et al. Simulation-guided rational De Novo design of a small pore-forming antimicrobial peptide. *J Am Chem Soc*. (2019) 141(12):4839–48. doi: 10.1021/jacs.8b11939

68. Kauffman WB, Fuselier T, He J, Wimley WC. Mechanism matters: a taxonomy of cell penetrating peptides. *Trends Biochem Sci*. (2015) 40(12):749–64. doi: 10.1016/j.tibs.2015.10.004

69. Lin B, Li R, Handley TNG, Wade JD, Li W, O'Brien-Simpson NM. Cationic antimicrobial peptides are leading the way to combat oropathogenic infections. *ACS Infect Dis*. (2021) 7(11):2959–70. doi: 10.1021/acsinfecdis.1c00424

70. da Silva BRD, Freitas VAAD, Nascimento-Neto LG, Carneiro VA, Arruda FVS, Aguiar ASWD, et al. Antimicrobial peptide control of pathogenic microorganisms of the oral cavity: a review of the literature. *Peptides*. (2012) 36(2):315–21. doi: 10.1016/j.peptides.2012.05.015

71. Niu JY, Yin IX, Mei ML, Mei ML, Wu WKK, Li Q-L, et al. The multifaceted roles of antimicrobial peptides in oral diseases. *Mol Oral Microbiol*. (2021) 36(3):159–71. doi: 10.1111/omi.12333

72. Gorr SU. Antimicrobial peptides of the oral cavity. *Periodontol 2000*. (2009) 51:152–80. doi: 10.1111/j.1600-0757.2009.00310.x