

BIONANO-INTERFACES THROUGH PEPTIDE DESIGN:

ANTIMICROBIAL INTERFACES

By

© 2018

Kyle Boone

Submitted to the graduate degree program in Bioengineering and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Candan Tamerler

Co-Chairperson Kyle Camarda

Dr. Paulette Spencer

Dr. Mark Richter

Dr. Suzanne Shontz

Date Defended: July 20, 2018

The dissertation committee for Kyle Boone certifies that this is the approved version of the following dissertation:

Bionano-Interfaces through Peptide Design

Chair: Candan Tamerler

Co-Chair: Kyle Camarda

Date Approved: July 20, 2018

ABSTRACT

The clinical success of restoring bone and tooth function through implants critically depends on the maintenance of an infection-free, integrated interface between the host tissue and the biomaterial surface. The surgical site infections, which are the infections within one year of surgery, occur in approximately 160,000-300,000 cases in the US annually. Antibiotics are the conventional treatment for the prevention of infections. They are becoming ineffective due to bacterial antibiotic-resistance from their wide-spread use. There is an urgent need both to combat bacterial drug resistance through new antimicrobial agents and to limit the spread of drug resistance by limiting their delivery to the implant site. This work aims to reduce surgical site infections from implants by designing of chimeric antimicrobial peptides to integrate a novel and effective delivery method.

In recent years, antimicrobial peptides (AMPs) have attracted interest as natural sources for new antimicrobial agents. By being part of the immune system in all life forms, they are examples of antibacterial agents with successfully maintained efficacy across evolutionary time. Both natural and synthetic AMPs show significant promise for solving the antibiotic resistance problems. In this work, AMP1 and AMP2 was shown to be active against three different strains of pathogens in Chapter 4. In the literature, these peptides have been shown to be effective against multi-drug resistant bacteria. However, their effective delivery to the implantation site limits their clinical use. In recent years, different groups adapted covalent chemistry-based or non-specific physical adsorption methods for antimicrobial peptide coatings on implant surfaces. Many of these procedures use harsh chemical conditions requiring multiple reaction steps. Furthermore, none of these methods allow the orientation control of these molecules on the surfaces, which is an essential consideration for biomolecules. In the last few decades, solid binding peptides attracted high interest due to their material specificity and self-assembly properties. These peptides offer robust surface adsorption and assembly in diverse applications. In this work, a design method for chimeric antimicrobial peptides that can self-assemble and self-orient onto biomaterial

surfaces was demonstrated. Three specific aims used to address this two-fold strategy of self-assembly and self-orientation are: 1) Develop classification and design methods using rough set theory and genetic algorithm search to customize antibacterial peptides; 2) Develop chimeric peptides by designing spacer sequences to improve the activity of antimicrobial peptides on titanium surfaces; 3) Verify the approach as an enabling technology by expanding the chimeric design approach to other biomaterials.

In Aim 1, a peptide classification tool was developed because the selection of an antimicrobial peptide for an application was difficult among the thousands of peptide sequences available. A rule-based rough-set theory classification algorithm was developed to group antimicrobial peptides by chemical properties. This work is the first time that rough set theory has been applied to peptide activity analysis. The classification method on benchmark data sets resulted in low false discovery rates. The novel rough set theory method was combined with a novel genetic algorithm search, resulting in a method for customizing active antibacterial peptides using sequence-based relationships.

Inspired by the fact that spacer sequences play critical roles between functional protein domains, in Aim 2, chimeric peptides were designed to combine solid binding functionality with antimicrobial functionality. To improve how these functions worked together in the same peptide sequence, new spacer sequences were engineered. The rough set theory method from Aim 1 was used to find structure-based relationships to discover new spacer sequences which improved the antimicrobial activity of the chimeric peptides.

In Aim 3, the proposed approach is demonstrated as an enabling technology. In this work, calcium phosphate was tested and verified the modularity of the chimeric antimicrobial self-assembling peptide approach. Other chimeric peptides were designed for common biomaterials zirconia and urethane polymer. Finally, an antimicrobial peptide was engineered for a dental adhesive system toward applying spacer design concepts to optimize the antimicrobial activity.

ACKNOWLEDGEMENTS

A thesis is not an individual achievement. I am very grateful for those who have guided me in my journey to complete it. First, I would like to thank my advisor, Professor Candan Tamerler for her understanding, support and guidance. She has inspired me to be inspired by the intricate systems found in nature. I also thank my co-advisor Professor Kyle Camarda for his support, guidance and friendship. He inspired me to push my coding skills to provide searching capabilities for large search spaces. What I have learned by being mentored by them I hope to invest in others.

I would also like to thank Professor Paulette Spencer, Professor Malcom Snead, and Professor Mark Richter for scientific discussions and for inspiring examples of research ideals. I would like to thank Dr. Charles Ye, Dr. Xueping Ge and Dr. Linyong Song for collaborations in dental materials and scientific group discussions in collaboration for developing advanced dental adhesives.

I am also grateful to the many who have accompanied me on my journey. Banu Karaca, Deniz Yucesoy, Esra Yuca and Phil Elrod showed me how to thrive in graduate school, as I continually learn more. Deniz, Cate Wisdom, Dr. Sheng-Xue Xie and Sarah VanOosten have contributed to my work immensely through their collaboration and generous sharing of their laboratory skills. Brandon Tomas contributed friendship and good feedback on my code. Next, I would like to thank Professor Camarda's group for their discussions: Dr. Rex Gaumer, Brock Roughton, Farhana Abedin and Thora Whitmore.

A special thank you to Professor Jerzy Grzymala-Busse for his work in rough set theory on the MLEM2 algorithm. Professor Luke Huan and his group have helped me with discussions on machine learning approaches, including a fantastic summer working on the KDD Cup competition. I would also like to thank the perspectives of visiting scholars Professor Ehren Piskin, Professor Turkan Haliloglu, Dr. Erkan Mozioglu, Arnaud Petit and Hassan Caglar.

I am honored to have been supported to do important work to reduce the risk of bacterial infections. This work has been mainly funded by research grants R01DE022054, 3R01DE022054-04S1 and R01DE025476 from the National Institute of Dental and Craniofacial Research, and from National Institute of Arthritis and Musculoskeletal and Skin Diseases 7R21AR062249-03, National Institutes of Health, Bethesda, Maryland.

I dedicate this work to my wife, Angie, and my daughters Sophia and Audrey. My wife is a critical source of strength, connection and patience. Her contributions on my journey have been not merely numerous but also substantial. She is my best partner and my best friend.

My daughters, Sophia and Audrey, are my energetic teachers in how to live with curiosity and passion. Sophia has shown a passion for creativity in her art that inspires me. Audrey has the curiosity and drive that shows me that resilience is worthwhile. I hope each one is as fortunate as I have been to be on a journey that is both collaborative and significant.

Contents

Abstract	iii
Acknowledgements	v
Figures.....	xii
Tables.....	xv
1.0 Introduction	1
1.1 Antimicrobial Peptide Classification by Rough Set Theory	8
1.1.1 Rough Set Theory	8
1.1.2 Antimicrobial Peptide Classification	14
1.2 Computer Aided Molecular Design of Peptides.....	15
1.3 Spacer Engineering through Peptide Structure	20
1.3.1 Polypeptide Structure Levels	21
1.3.2 PyRosetta Folding.....	22
1.3.3 Fragment Insertion.....	23
1.4 Specific Aims	24
2.0 Antimicrobial Peptide Classification through Rough Set Theory	27
2.1 Introduction	28
2.2 Background	29
2.3 Methods.....	32

2.3.1	Rule Induction by the MLEM2 Algorithm.....	32
2.3.2	Correlated AAindex1 Property Removal.....	35
2.3.3	Feature Generation.....	35
2.3.4	Performance Measures.....	38
2.4	Results.....	39
2.4.1	Rough Set Theory Analysis.....	41
2.4.2	Training Performance by Rule Coverage.....	43
2.4.3	Training Performance by Features Selected.....	44
2.4.4	Benchmarking CLN-MLEM2.....	46
2.5	Discussion.....	48
2.6	Conclusion.....	51
3.0	Peptide Design with Rough Set Theory and a Codon-Based Genetic Algorithm.....	53
3.1	Introduction.....	54
3.2	Background.....	54
3.3	Methods.....	58
3.3.1	Rule Induction by Rough Set Theory.....	58
3.3.2	Genetic Algorithm for Finding New Peptide Sequences.....	59
3.4	Results.....	61
3.4.1	Novel Antimicrobial Peptide Design.....	61
3.4.2	Novel Antimicrobial Peptide Experimental Test.....	64

3.4.3	Combining MLEM2 Rule Categories.....	66
3.5	Discussion.....	68
3.5.1	Increasing Variation through Codon Representation	68
3.5.2	Combining Antibacterial Classes.....	68
3.5.3	Targeting Specific Bacterial Strains.....	69
3.6	Conclusions	69
4.0	Chimeric Spacer Design for Titanium Implant Surfaces.....	71
4.1	Introduction	72
4.2	Background	72
4.3	Methods.....	74
4.3.1	PyRosetta Ensemble.....	74
4.3.2	Secondary Structure Feature Counts.....	74
4.4	Results.....	74
4.4.1	Chimeric Titanium Antimicrobial Peptides Reduced Efficacy	74
4.4.2	Designing New Chimeric Peptide Spacer for Improved Antimicrobial Activity	79
4.5	Discussion.....	81
4.6	Conclusions	82
5.0	Chimeric Antimicrobial Peptides across Multiple Materials.....	83
5.1	Introduction	84
5.2	Background	84

5.3	Methods.....	89
5.3.1	Chimeric Spacer Design Method.....	89
5.3.2	Secondary Structure and Spacer Frequency Error.....	90
5.3.3	Circular Dichroism.....	91
5.4	Results.....	92
5.4.1	Calcium Phosphate Surfaces.....	92
5.4.2	Zirconia Surfaces.....	105
5.4.3	Urethane Polymer Surfaces.....	108
5.5	Discussion.....	112
5.5.1	Calcium Phosphate Surfaces.....	112
5.5.2	Zirconia Surfaces.....	114
5.5.3	Urethane Polymer Surfaces.....	114
5.6	Conclusions.....	115
6.0	Engineered Antimicrobial Peptide Enhances Dental Adhesive.....	117
6.1	Introduction.....	118
6.2	Background.....	118
6.3	Methods.....	121
6.3.1	Inhibitory Concentration Assays.....	121
6.3.2	Circular Dichroism.....	122
6.3.3	Computational Structure Flexibility.....	122

6.3.4	Adhesive Disc Preparation	123
6.3.5	Antimicrobial Activity Assays of Adhesive Discs with Peptide Coupling.....	124
6.4	Results.....	125
6.4.1	Inhibitory Concentrations	125
6.4.2	CD Spectra and CD Deconvolution.....	126
6.4.3	Computational Structure Flexibility	128
6.4.4	Antibacterial Activity of Treated Discs.....	130
6.5	Discussion.....	131
6.6	Conclusion.....	133
7.0	Conclusions and Future Directions	137
8.0	References	143
9.0	Appendix A. Nomenclature.....	163
10.0	Appendix B. PyRosetta Folding Code	167

Figures

Figure 1.1 Self-assembled monolayers (SAM) attach to a surface to change the material's interfacial properties.....	4
Figure 1.2 Implant Surface at the Nanoscale.....	6
Figure 1.3 Computer-Aided Molecular Design Flow Chart.....	19
Figure 1.4 Classic relax protocol for three initial starting conformations of a titanium binding peptide (AVSPHGVHRSAHGGG).	23
Figure 1.5 Fragment insertion demonstrates large moves in conformational space compared to the moves from the Classic Relax protocol for a titanium binding peptide (TMTAPWPPVLDALGVVLSAVFLPVILPVIGKLLNGIL).....	24
Figure 2.1 CLN-MLEM2 Method.....	34
Figure 2.2 Rough Set Theory Rule Generation.....	37
Figure 2.3 Auto-Correlation and Selection of AAindex1 Properties.	40
Figure 2.4 Training accuracy by rule size.	44
Figure 2.5 The training set accuracy of CLN-MLEM2 rules by features selected.....	45
Figure 2.6 The frequency of selection of features by CLN-MLEM2 rules.	46
Figure 2.7 False discovery rates of antimicrobial peptide classification.	50
Figure 3.1 Computer Aided Molecular Design.....	56
Figure 3.2 CLN-MLEM2 Algorithm including codon-based genetic algorithm selection and evaluation. ...	59
Figure 3.3 Codon-Based Genetic Algorithm (CB-GA).....	60
Figure 3.4 Schematic of Codon Basis for CB-GA Algorithm	61
Figure 3.5 Fitness Distributions of Selected Generations.....	62
Figure 3.6 Improved sequence variation through codon-representation.....	64
Figure 3.7 Fitness Distributions of Selected Generations.....	66

Figure 4.1. Schematic of rule induction process by MLEM2 for discovering structure-function relationships for chimeric antimicrobial peptides.....	75
Figure 4.2. Aligned secondary structure decoys for 5-a.a.-helix rules generated by MLEM2 algorithm. ..	78
Figure 4.3. Aligned secondary structure decoys of 8-a.a.-helix rule generated by MLEM2.....	79
Figure 4.4. Improved spacer for antibacterial activity against <i>S. mutans</i>	80
Figure 5.1 Process schematic for chimeric spacer engineering design.....	90
Figure 5.2. Bacterial turbidity (OD_{600}) as a function of hydroxyapatite binding chimeric peptide concentration and its component domains.....	93
Figure 5.3. Helical structure frequencies for hydroxyapatite binding chimeric peptide and its components	94
Figure 5.4 Characteristic structures fitting each of the inducted rules in Table 5.2 for HABP-AMP and its component domains.....	96
Figure 5.5 Structure characterization of linear HABP1, HABP2, constrained HABP1 and constrained HABP2 through circular dichroism analysis.	97
Figure 5.6. Solvent engineering study with 2,2,2-trifluoroethanol (TFE) for HABP1.....	98
Figure 5.7 Estimation of Flexibility Study by computer-generated structures to estimate folding frequencies across the residues of the HABP1 sequences; A) HABP1, B) Constrained HABP1, C) HABP2, D) Constrained HABP2.....	99
Figure 5.8 Computational Intrinsic pKa Study to estimate effect of disulfide bond on peptide structure between linear peptides (A) HABP1 and (B) HABP2 and constrained peptides (C) cHABP1 and (D) cHABP2.....	100
Figure 5.9. Surface area changes of ensembles of structural decoys due to substitution of alanine at a single position.....	101

Figure 5.10 Surface area changes of ensembles of structural decoys due to substitution of histidine at a single position.....	102
Figure 5.11 Structure decoy of chimeric peptide combining fast mineralization kinetics with octacalcium phosphate mineral-forming morphology generated in PyRosetta.....	104
Figure 5.12: Summed error of chimeric peptides compared to ADP1 peptide alone of feature frequencies at each residue position.....	104
Figure 5.13: Summed error of chimeric peptides compared to HABP1 peptide alone of feature frequencies at each residue position.....	105
Figure 5.14. Structural decoy showing two 5-aa α helices in a chimeric peptide (ZrBP-IGVVISAV-AMP1) with a second-generation spacer domain (IGVVISAV).....	108
Figure 5.15. Secondary structure feature frequency by residue position of PUABP1.....	109
Figure 5.16 Secondary structure feature frequency by residue position of AMP1.	109
Figure 5.17 Secondary structure feature frequency by residue position of PUABP1-GSGGG-AMP1.....	110
Figure 5.18 Spacer frequency error (SFE) for selected chimeric peptides with AMPa (KWKLWKKIEKWGQGIGAVLKWLTTWL).....	111
Figure 5.19 Spacer frequency error (SFE) for selected chimeric peptides with AMP1 (LKKLLKLLKLLKLL).....	112
Figure 6.1 Dental primer adhesive system for antimicrobial barrier to prevent bacterial degradation ..	119
Figure 6.2 Antibacterial activity of CHX, antimicrobial peptides and ϵ -polylysine.....	126
Figure 6.3 Secondary structure characterization by circular dichroism of GH12 and GH12-M2.	127
Figure 6.4 Analysis of secondary structure features predicted through PyRosetta method and quantified through DSSP.	129
Figure 6.5 Inhibitory activity of GH12-M2 with the ϵ -polylysine added to the disc formulation.....	131

Tables

Table 1.1 Amino acid properties to distinguish amino acids based on charge at pH 7.4 of rough set theory rules.	10
Table 1.2 Rough set theory rules separating amino acids by charge at pH 7.4 based on their isoelectric point (pH at which they have no charge).....	11
Table 1.3 Rough set theory rules separating amino acids by charge based on their size.	12
Table 2.1 Description of summary functions to generate chemical property features as input for the MLEM2 method.	35
Table 2.2 Maximum Support Rough Set Theory Rules	41
Table 2.3 Description of AAindex features selected by CLN-MLEM2 method.	43
Table 2.4 Performance of rough set theory rule induction compared to motif-search in 10-fold cross validation.	47
Table 2.5 Performance comparison among prediction servers for antimicrobial peptides, a motif-based classification method and rough set theory approach.....	48
Table 3.1 Design targets for novel antimicrobial peptides for ease of Fmoc synthesis.	62
Table 3.2 Inhibition zone of <i>S. epidermidis</i> for candidate novel antimicrobial peptides.	65
Table 3.3 Design targets for novel antimicrobial peptides for combining MLEM2 rule categories.	66
Table 3.4 Inhibition zone of <i>S. epidermidis</i> as a screen test of antibacterial activity for combining MLEM2 rule categories.	67
Table 4.1 Minimum inhibitory concentrations for chimeric antimicrobial peptides with solid binding peptides for titanium and a GGG spacer.	75
Table 4.2. MLEM2 generated rules describing secondary structure and chimeric antimicrobial peptide activity.....	77

Table 4.3. PyRosetta method of peptide structure prediction frequencies of four and five amino acid alpha helices in computational generated structures.	81
Table 4.4 MIC of TiBP-Spacer5-AMP, TiBP-Spacer3-AMP and AMP	81
Table 5.1. Antimicrobial activity of peptides for the calcium phosphate surface.	93
Table 5.2 Inducted rules by MLEM2 for secondary structure features of chimeric hydroxyapatite binding peptide and its component domains.	95
Table 5.3: Initial Spacer Sequences as a Modeling Basis for Spacer Design Method	103
Table 5.4. Selected secondary structure feature rules from the MLEM2 method relating to high antibacterial activity from data mining structure ensembles.	106
Table 5.5. 5-a.a. α helix frequencies for the initial generation of chimeric spacers.....	107
Table 5.6 Second generation of spacers derived from initial generation of spacers in Table 5.5.....	107
Table 6.1 Commercial analogue components and ϵ -Polylysine Primer components.	124
Table 6.2 Properties of GH12 peptides, modifications and ϵ -polylysine	125

1.0 Introduction

This page is intentionally left blank.

This work develops new antibacterial peptides to bring infection-free interfaces to implanted surfaces. Bacterial infection of surgical sites has high costs for patient health and for healthcare. About 160,000-300,000 surgical site infections occur annually in the United States.¹ This represents 2%-5% of surgical procedures. The need for surgical procedures is increasing. For example, nearly 3 million annual total knee replacements are projected in the US by 2030.² Therefore, the expected surgical site infections for knee replacement surgeries alone will be between 60,000 - 150,000 if the average surgical site surgery rate is not improved. This number can be reduced if a new approach addresses this problem threatening the formation and maintenance of the integration of the implanted material and the host tissue.

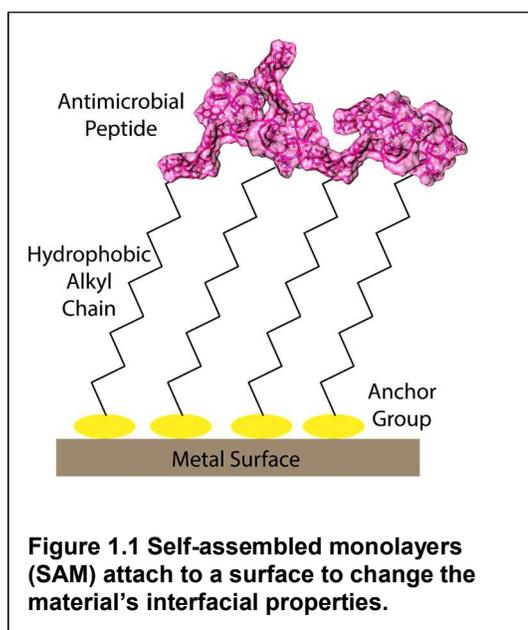
The recommended strategies for preventing surgical site infections are to administer antibiotic drugs pro-actively, to use topical antiseptics to kill bacteria on the skin and to oversee the adherence to approved protocols for post-operative care.¹ The role of bactericidal antibiotics can be put into a wider context than the killing of bacteria. Certain types of antibiotics have been observed for more than 15 years to have direct effects on immune cells.³ As these antibiotics require larger doses to kill bacteria due to drug resistance, their effects on the immune system and surrounding tissues become of greater importance. Therefore, synergistic solutions to treating bacterial infections while maintaining host-cell viability are important.⁴ New antibacterial agents are needed to replace current antibiotics to improve surgical outcomes.⁵⁻⁹

Antimicrobial peptides are promising antibacterial agents due to their maintained innate immune response against their bacterial targets.¹⁰⁻¹³ The specificity of the antimicrobial peptides is an advantage for the host to maintain the viability of the host cells. While antimicrobial peptides are promising treatments against bacteria for their potency and specificity among bacterial strains, the lack of delivery systems which can target specific biomaterial locations limit their use.⁵ Their delivery may result in immediate attack by the proteases even before they reach the implantation site. Localized delivery can solve the problem of sub-inhibitory concentrations or immediate protease attack when delivered by

systemic methods. Therefore, different groups have used covalent chemistries¹⁴⁻¹⁶ or physical adsorption¹⁷⁻¹⁹ to deliver antimicrobial peptides at the implant surface.^{20, 21}

The delivery of antimicrobial peptides to the implant surface through covalent chemistry approaches for surface chemical modification require harsh conditions and the availability of specific chemical groups on the implant surface. These approaches, therefore, are limited by the implant surface chemistry to which types of surfaces can be functionalized. Among these approaches, the use of self-assembled monolayers (SAM) has been the most commonly studied approach by different groups to deliver a single-layer of antimicrobial peptides to a material surface.²²⁻²⁴ The approach covalently bonds alkyl

chains with two chemical entities to a material. The first chemical entity is a functional group, called the anchor group, that binds directly with the material surface. The compatibility of the implant surface with this approach is restricted by the availability of these complementary chemical groups at the surface. The second chemical entity is the antimicrobial peptide. The SAM approach lacks self-organization and any orientation control of antimicrobial peptides which limits their efficacy. There are many



studies reporting that antimicrobial peptides immobilized through the SAM approach resulted in loss of function.^{25, 26} To solve this important issue, the chemical linkers between the surface and the biological function may need to move beyond being limited to the hydrophobic interactions of alkyl chains.

As an alternative approach for the delivery of antimicrobial peptides, physical adsorption approaches offer a wide variety of surfaces to be functionalized, removing chemical functional surface group restrictions. A common approach is to use high-surface area modifications, such as coating with nanometer-scale features such as tubes.²⁷⁻²⁹ Depending on the non-covalent interactions between the

implant surface and the antimicrobial peptide, the strength of the binding for the adsorption can be weak. If the antimicrobial peptide changes of conformation of the residues responsible for its activity, the peptide denatures as it adsorbs to the implant surface. Providing orientation control for active agents while non-specifically adsorbed is an unmet challenge.³⁰

In the recent years, solid binding peptides have attracted high interest due to the non-covalent nature of the biological self-assembly process.³¹⁻³⁵ Solid binding peptides are discovered through a combinatorial biology based biopanning process, namely phage and cell surface display. These systems rely on the genotype-to-phenotype relations where large molecular libraries are genetically engineered into the genomes of display vectors to express short peptides on the coat (phage display) or surface proteins (cell surface). When the host interacts with the material of technological interest, the displayed molecular library members of short peptides will have an opportunity to develop a stable, non-covalent interaction with the material. To differentiate the strength of the interactions among peptides, the stringency of the environment is increased throughout serial washing steps. When robust binding hosts are recovered, their DNA is extracted to identify the sequence. By translating the DNA sequence to amino acids, the sequences of the short peptides that interact with the material surface are identified. Before this work, Tamerler's group has also developed bioinformatics-based analysis through scoring matrices to improve the prediction of binding for peptides to material surfaces.³⁶⁻³⁸

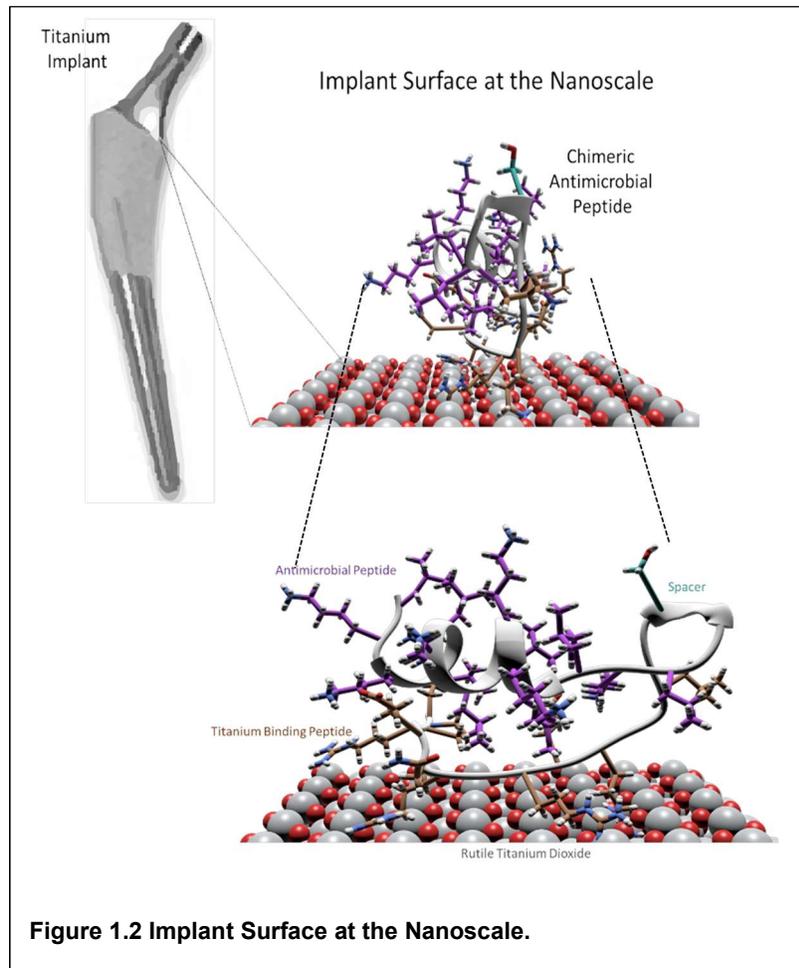
Using solid binding peptides to provide new functions to surfaces is an active area of research.^{33, 39, 40}

Solid binding peptides have multiple amino acid residues and structural mechanisms to adsorb specifically on materials through molecular recognition.^{5, 6, 30, 41, 42} In this work, the approach is to use chimeric antimicrobial peptides with solid binding peptides to coat a variety of biomaterials, while also providing self-organization capability to the antimicrobial peptides adsorbed onto the surface.

In this work, the hypothesis is that designing chimeric antimicrobial peptides that can self-assemble on implant surfaces will reduce bacterial viability on different implant surfaces. The approach is to develop chimeric antimicrobial peptides as single oligopeptide chains of amino acids with two distinct functional units that are combined with a spacer sequence between them. One functional unit is the solid binding peptide for the implantable materials surface. The other functional unit is an antimicrobial peptide. The spacer sequence provides self-organized orientation of the biological signal (Figure 1.2).

This design method includes tailoring the antimicrobial peptides specifically towards the problematic strains of drug-resistant bacteria as well as their self-assembly onto different biomaterial surfaces, i.e. titanium and calcium phosphate and dental adhesive systems as case studies.

Titanium is the most commonly used implantable material. It is used in a wide range of applications varying from joint



replacements, to dental implants, and to surgical tools.⁴³ The second material chosen to functionalize with antibacterial activity is calcium phosphate. This mineral, in different phases, has been used in the field to promote osseous integration to improve patient outcomes.⁴⁴⁻⁴⁷ The significance of this work is immediately to enable antibacterial functionality through the delivery of antimicrobial peptides to a

second surface type. The self-assembly of molecular recognition to titanium is expected to transfer to calcium phosphate surfaces when the solid binding peptide is exchanged. Beyond the examples of adding antimicrobial functionality in this work, biological signals that require orientation control can be effectively displayed on these surfaces if the antimicrobial peptide is exchanged for a biological signal. Prior to this work, Tamerler's group has established through integration of titanium binding chimeric peptides integrin-binding domain for mammalian cell recruitment⁴⁸ or *Wnt*-signaling peptide for controlling stem cell differentiation onto the implant surface.⁴¹ The current work explores the classification and design of chimeric antimicrobial peptides with spacer engineering to display biological functionality effectively onto the selected biomaterials. The efficacy of modular design principles with this approach is dependent on the interaction between the functional domains of the chimeric peptide. To address negative interactions between the functional domains seen in Tamerler's group's previous works,^{6, 49-51} this work provides a spacer design method to lessen the negative interactions.

Many antimicrobial peptide sequences have been reported in the literature include naturally occurring peptides, peptide mimics and *in silico* designed peptides. In the LAMP database, there are over 5,500 entries with sparse data about the activity of the antimicrobial peptides against specific bacterial strains. Therefore, there is an unsolved challenge to differentiate active peptides from non-active ones for targeted bacteria. First, this work attacks this problem by developing and implementing a novel antimicrobial peptide classification methodology. The classification method utilizes a rough set theory approach combined with different machine learning modules. Then this approach is integrated with computer aided molecular design (CAMD) for the discovery of functional antimicrobial peptides. Next, this work further explores structure-function relationships of peptide folding using a spacer as a tunable parameter to enhance the activity of the chimeric molecules. Finally, this work demonstrates the design of chimeric and active peptides on a variety of material surfaces. This work follows the summary of

these methodologies by the aims, i.e. explained in separate chapters. A dental adhesive system chapter follows showing another area of application of these methodologies.

1.1 Antimicrobial Peptide Classification by Rough Set Theory

Rough Set Theory

Rough set theory uses the intersections of sets to describe the relationship between data classification labels, called decisions, and columns of features in a data table. Sets are collections of cases, which are the rows of the data table. The cases are amino acids in this example and the features are amino acid properties. The classification label is the amino acid charge at pH 7.4.

A rule is a series of value ranges for columns in a data table that assigns a label. The support of a rule is the number of cases that the rule can correctly label in the data table. The coverage of the rule is the number of cases that the rule will label in the data table. The cases which are part of any rule's support are said to be distinguishable. Those cases that are part of a rule's coverage and not part of any rule's support are said to be indistinguishable. These cases are vague. Cases must be vague if all columns in a data table agree but the data classification labels do not agree.

Distinguishability in rough set theory is like the concept of correlation in linear regression analysis.

Highly correlated variables imply predictability of the value of one variable given the other variable of the linear relationship. Distinguishability is knowing the value of the classification label with certainty among the values in the data table given ranges of values for features. While distinguishability implies predictability, it does not imply or require a linear relationship to discover.

Before discussing the mechanics of rule induction by rough set theory, the relationship between the features and classification labels in Table 1.1 is discussed. The isoelectric point of an amino acid, Column 1, is the pH at which the amino acid is uncharged. The size of an amino acid is given in Column 2. This pH can be determined from the response to an electric field. The partition coefficient, in Column 3, is the

concentration ratio of the amino acid in octanol compared to water after mixing the system. The charge of an amino acid, Column 4, is the force applied by an electric field equal to the force on a single electron in the opposite direction that would be applied by the field. The isoelectric point directly relates to the charge of an amino acid at pH 7.4. Amino acids with isoelectric points much below 7.4 are negatively charged at pH 7.4, amino acids with isoelectric points close to 7.4 are not charged, and amino acids with isoelectric points higher than 7.4 are positively charged at pH 7.4.

There is no theoretical relationship between the size of an amino acid and its charge. There are however, empirical patterns between the size and the charge of the amino acid that can be discovered through rough set theory if combined with at least one other feature. The partition coefficient and charge do have a theoretical relationship in that a high partition coefficient implies low polarity to be soluble with octanol. Low polarity implies no charge. The partition coefficient does not create a distinguishable rule set for charge on its own. This is because the sign of the charge is not related to the partition coefficient, even if the presence of charge has an impact on the value of the partition coefficient. The direct relationship between isoelectric point and charge at pH 7.4 will be the first one examined by rough set theory to cover the rule induction mechanics. The second relationship will be the combination of two indirect relationships of size and partition coefficient to charge at pH 7.4 to explore the concept of overfitting.

Table 1.1 Amino acid properties to distinguish amino acids based on charge at pH 7.4 of rough set theory rules.

Amino Acid	Amino Acid Isoelectric Point⁵²	Amino Acid Size⁵³	Amino Acid Partition Coefficients⁵⁴	Amino Acid Charge at pH 7.4
Alanine	6.00	2.5	0.28	0
Arginine	10.76	7.5	0.10	+1
Asparagine	5.41	5.0	0.25	0
Aspartic Acid	2.77	2.5	0.21	-1
Cysteine	5.05	3.0	0.28	0
Glutamine	5.65	6.0	0.35	0
Glutamic Acid	3.22	5.0	0.33	-1
Glycine	5.97	0.5	0.17	0
Histidine	7.59	6.0	0.21	+1
Isoleucine	6.02	5.5	0.82	0
Leucine	5.98	5.5	1.00	0
Lysine	9.74	7.0	0.09	1
Methionine	5.74	6.0	0.74	0
Phenylalanine	5.48	6.5	2.18	0
Proline	6.30	5.5	0.39	0
Serine	5.68	3.0	0.12	0
Threonine	5.66	5.0	0.21	0
Tryptophan	5.89	7.0	5.70	0
Tyrosine	5.66	7.0	1.26	0
Valine	5.96	5.0	0.60	0

The simplest relationship between classification labels and features to find in a data table is a single feature column that can distinguish among all classification labels. To classify the charge of the amino acids, the isoelectric point can be used to distinguish among all amino acid charges at pH 7.4. Table 1.2 shows the rules that distinguish which isoelectric points relate to charge at pH 7.4. Rough set theory

accomplishes these inducted rules by selecting each classification label individually. Beginning with the classification label of -1, only two cases exist: glutamate and aspartate. Listing the amino acids sorted by isoelectric points in increasing order would put these two amino acids at the first and second positions. Cut points are introduced between each unique value in the sorted list, so there are twenty intervals ordered by increasing value sharing the minimum isoelectric point as the left bound and twenty intervals ordered by decreasing value sharing the maximum isoelectric point as the right bound. Of the forty intervals generated with the cut points, only twenty-two contain an amino acid with a -1 label. The other eighteen intervals are discarded. The remaining intervals are evaluated by the relevance, that is how many -1 labeled cases that they contain. Two of the intervals have one case, and twenty have two cases. The intervals with the maximum relevance of two are kept. Finally, to maximize the conditional probability of the inducted rules, the minimum number of cases from among those with maximum relevance is selected. Only one interval has two -1 cases and two total cases. This is the interval that uses the cut point between the second and third amino acid in the sorted order by isoelectric point. Therefore, the boundaries for this inducted rule is the first cut point, which is the minimum value, and the third cut point, which is the average of 3.22 and 5.05. The other two rules in Table 1.2 can be calculated in the same way.

Table 1.2 Rough set theory rules separating amino acids by charge at pH 7.4 based on their isoelectric point (pH at which they have no charge).

Rule Number	Condition Number	Property	Lower Value	Upper Value	Support	Coverage	Charge at pH 7.4
1	1	Isoelectric Point	2.77	4.135	2	2	-1
2	1	Isoelectric Point	4.135	6.945	15	15	0
3	1	Isoelectric Point	6.945	10.76	3	3	+1
				Total	20	20	

The second example of rule induction will include two features, the size of the amino acid and its partition coefficient in labeling the charge of the amino acid at pH 7.4. Rough set theory can distinguish among all amino acids in the data table because there are unique value pairs for each of the twenty amino acids using both features. If only the size feature is used, only eight amino acids can be correctly labeled with charge, as seen in Table 1.3.

Table 1.3 Rough set theory rules separating amino acids by charge based on their size.

Rule Number	Condition Number	Property	Lower Value	Upper Value	Support	Coverage	Charge
1	1	Size	5.25	5.75	3	3	0
2	1	Size	2.75	4.0	2	2	0
3	1	Size	0.5	1.5	1	1	0
4	1	Size	6.25	6.75	1	1	0
5	1	Size	7.25	7.5	1	1	+1
				Total	8	8	

Rules which distinguish amino acids based on partition coefficient and size are:

Rule 1: Partition Coefficient $\geq 0.34 \rightarrow$ Amino Acid Charge = 0 at pH 7.4

Rule 2: Partition Coefficient ≤ 0.23 & Size $\geq 5.5 \rightarrow$ Amino Acid Charge = +1 at pH 7.4

These rules allow for a new perspective on empirical relationships for the canonical amino acid between size, partition coefficient and amino acid charge. Rule 1 gives information which will likely hold for non-canonical amino acids. A review of commercially-available non-canonical amino acids for computational modeling has been published.⁵⁵ If the partition coefficient is greater than 0.34, then any amino acid is likely to be uncharged at pH 7.4. The partition coefficient is higher for molecules more soluble in octanol than in water. Because polar molecules are, in general, more soluble in water than in octanol, charged

amino acids will have smaller partition coefficients. Rule 2, because it is based on size, is an empirical relationship that may not generalize to non-canonical amino acids. If the partition coefficient is less than 0.23, then any amino acid will likely be charged at pH 7.4, but there are negatively charged non-canonical amino acids with sizes larger than 5.5, such as α -amino adipic acid.⁵⁵ Rough set theory can find empirical relationships within data tables. For the inducted rules to avoid overfitting, the underlying relationships between the features and classification labels must be applicable in the domain in which the data table is describing.

Often, there are multiple reductions for a single data table, demonstrated by the multiple variations of rough set theory algorithms in the literature.⁵⁶ The rule set from a data table is called a reduction. Finding all reducts for a data table is related to the Satisfiability Problem,^{56,57} in that if there is a polynomially-bound solution to finding all reducts, then there is a polynomially-bound solution to the Satisfiability Problem. Therefore, finding all reducts for a data table is known to be NP-Hard.^{58,59}

Using the principles of rough set theory in the literature, such as modified learning from experience module 2 (MLEM2),⁶⁰ open-loop feedback feature selection,⁶¹ and interesting rule induction module (IRIM),⁶² the CLN-MLEM2 method was built in this work for discovering which amino acid chemical properties relate clearly or which relate vaguely to antimicrobial activity. The rough set theory approach used in this work has heuristic methods to reduce overfitting. During rule induction, redundant features are removed. Value ranges are selected for maximizing the largest support. This work is the first time that rough set theory has been applied to describe peptide activity. In the future, many other different sequence-based and structure-based relationships can be explored for a wide variety of peptide functions.

1.1.1 Antimicrobial Peptide Classification

In the literature, the sequence representation used within classification approaches predicts which types of sequences are likely to be classified correctly.¹⁰ The first literature example, AntiBP,⁶³ is an antibacterial peptide prediction approach that uses a sliding window to represent sequences, leading to strong sensitivity to reordering amino acids of a given sequence. However, this representation leads to weak sensitivity for adding new amino acids at the ends of antibacterial regions that change key physicochemical properties of the overall sequence. Therefore, longer sequences are more difficult for sliding window representations to classify correctly.

Sliding window representations enable order sensitivity but are limited by a dependence on the length of the peptide.⁶⁴ CAMP^{65, 66} addresses length dependence by calculating overall physicochemical properties of sequences. While using overall sequence properties lead to strong sensitivity to adding new amino acids, this method has weak sensitivity for reordering amino acids in the same sequence. The overall sequence property representation is order insensitive but length independent. iAMP-2L⁶⁷ partially addresses order sensitivity by calculating autocorrelation of properties. This method is not sensitive to reordering among positions which are correlated and uses order-insensitive representations for many of its properties. EFC-FCBF^{12, 68} combines a motif-searching method that provides some order sensitivity with a support-vector machine (SVM) method that calculates overall properties for length independence. This method does not fully address order sensitivity because the twenty canonical amino acids are compressed to four categories in motifs. Therefore, this method is insensitive to reordering among the amino acids in the same category.

In work, a high-selectivity classification method named condition-limit number Modified Learning from Experience Module 2 (CLN-MLEM2) was developed. The condition-limit number refers to a user-defined parameter limiting the number of features in the rules inducted by the method. Therefore, the

performance of rule sets with less features can be directly compared to rule sets with more features. CLN-MLEM2 was developed from a rough set theory approach to discover key physicochemical properties for classifying peptides according to antibacterial activity. Order sensitivity is addressed by using the physicochemical properties of tripeptides within the full peptide sequences to distinguish among groups of similar active and inactive peptides by physicochemical properties. Length dependence is addressed by using the properties of full sequences. The method in this work combines order sensitivity with length independence through rules that determine classifications with both types of representations in a single rule. The observed low false positive rate is aided by inducing rules of inactive peptides.

1.2 Computer Aided Molecular Design of Peptides

More than three decades of research has been performed on antimicrobial peptides derived from natural sources. In the 1980s Merrifield's group, who developed solid phase peptide synthesis (SPPS) in the 1960s,⁶⁹ investigated antimicrobial peptides from the *Cecropia* moth.⁷⁰⁻⁷² Cantisani *et al* used alanine scanning of myxinidin to find new antimicrobial peptides by point substitutions showing improved activity across different bacteria.⁷³ While the rational design of antimicrobial peptide derivatives is a valuable tool for developing new antimicrobial peptides that are closely related by sequence to existing antimicrobial peptides, the approach limits the development of empirical rules that can more broadly explore peptide sequence space.

Concurrent with Merrifield's development of SPPS, QSAR models were pioneered by Hansch and Fujita.⁷⁴ An outstanding review of QSAR and its history has been recently published.⁷⁵ Traditional QSAR methods build molecular structure-function relationships. For peptides, the molecular structure is encoded by the amino acid sequence, which is often called the primary structure. Secondary structure, the hydrogen bonding of the backbone, and tertiary structure, the three-dimensional position of atoms, are complementary descriptions of the molecular structure of peptides. More recently, machine

learning methods have been used to generalize molecular structure relationships to binary or to categorical functions.⁷⁵ The rough set theory method in this work for describing antibacterial activity is an example.

Computer-aided molecular design (CAMD) is an approach to improve molecular design beyond the trial-and-error method.⁷⁶⁻⁷⁹ CAMD solves two problems simultaneously. The first problem is the forward problem, which is to understand which structures of the studied molecule leads to a targeted property. The second problem in CAMD is the reverse problem of using property values to search for candidate structures. In this work, the solutions for the forward problem are the rough set theory rules from the CLN-MLEM2 method.

The mathematical formulation for this work is of the form:

$$\min S = \sum_1^N \frac{1}{P_n^S} |P_n - P_n^*|$$

$$\text{Such that } P_n = f_n(x)$$

$$h_c(x) > 0$$

S is the “distance” of the peptide sequences from meeting all the objective targets. This optimization is a single-object approach that weighs multiple target property values to rank the overall value of the peptides. N is the number of the targeted properties. P^* is the target property vector, and P^S is the scaling vector weighing for normalizing the scale of each property. $f_n(x)$ is a function of the peptide sequence. Two examples of $f_n(x)$ explored in this work are the sum of the amino acid values for a property in the AAindex⁸⁰ and secondary structure frequencies resulting from a customized PyRosetta⁸¹ script. $h_c(x)$ is the difference between the sum of the rough set theory rules matched by the peptide sequence for activity and the sum of the rough set theory rules for inactivity. Since $h_c(x)$ is required to

be strictly positive, peptide sequences that do not match any rules or which match an equal number of rules for activity and inactivity are not included in the feasible region.

The second problem in CAMD is the reverse problem. In this work, the solutions for this problem are customized peptides which are antibacterial according to the boundaries of the rough set theory rules. The algorithm for the reverse problem must be a derivative-free approach to incorporate discontinuous activity descriptions from the rough set theory approach in this work.

Pattern Search Methods

Pattern search methods, which only rely on function evaluations for optimization, can be deterministic or stochastic.⁸² Derivative-free optimization methods have been used informally in experimental design and data analysis throughout the history of science.⁸³ In the 1960s and 1970s, many popular approaches for derivative-free optimization were developed including the direct search method,⁸⁴ the simplex method,⁸⁵ and genetic algorithms.⁸⁶ The direct search method and the simplex method may require too many function evaluations by partitioning the search space. The methods that partition the search space cover the entire search space at a coarse-grained level. However, due to the “curse of dimensionality”, the coarse-grained search space grows exponentially with an increasing number of dimensions. For peptide design, each amino acid is another dimension for each vector of amino acid sequences.

Genetic algorithms do not partition the search space, and they avoid evaluating some of the regions of the search space which are suboptimal. Instead, genetic algorithms rely on recombination operators to explore novel solutions for improved optimization. This property reduces the increase of the function evaluations when using recombination operators compared to partitioning approaches when the search space increases exponentially. For example, the increase in length of the peptide or protein being designed exponentially increases the size of the search space exponentially.

Genetic Algorithm

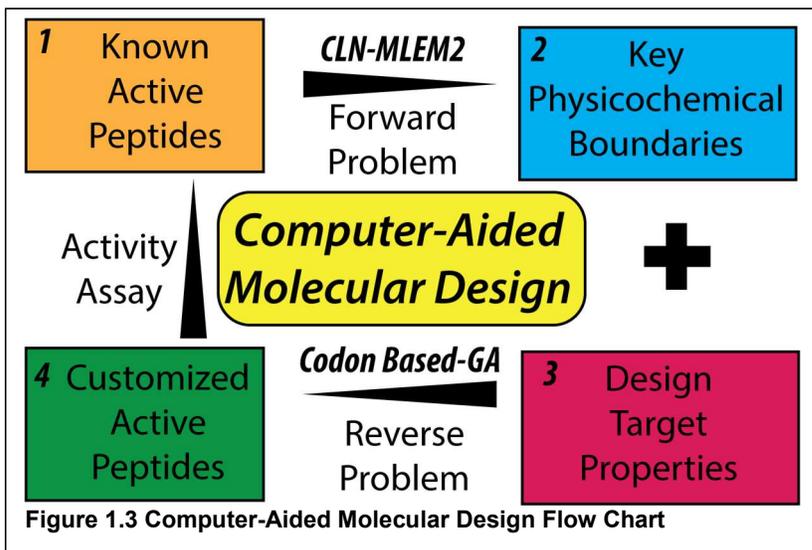
A genetic algorithm optimizes multiple generations of solutions toward defined goals through an evolutionary process. The genetic algorithm is a stochastic, or also called metaheuristic, optimization method. While deterministic optimization methods typically guarantee globally optimal solutions, many real-world combinatorial optimization problems do not fit currently available deterministic approaches. For this work, stochastic optimization methods are selected because of the description of antibacterial activity developed in this work is the form of inducted rules. The rough set theory rules form non-linear boundaries that are computationally fast to evaluate as the intersections of sets, but the rules are discontinuous discrete functions.

Among stochastic methods, simulated annealing and Tabu search are appropriate for the optimization of discontinuous functions. Simulated annealing uses a temperature parameter to shift from accepting non-improving moves in a strategy to escape local minima to only accepting improving moves as the annealing occurs once per simulation.⁸⁷ Genetic algorithms combine small and large moves with mutation to optimize for a local minimum and crossover to shift between local minima continually for each generation. Tabu search is a stochastic search method that combines neighborhood searches with intermediate and long-term moves to find new local minima until solutions meet a stopping condition, usually the repetition of previous solutions.⁸⁸ Genetic algorithms also incorporate multiple levels of moves to optimize local minima and to find new minima to optimize.

In the literature, genetic algorithms have been successfully integrated into the CAMD framework^{77, 78} and antimicrobial peptides design.⁸⁹ Genetic algorithms rely on evolutionary concepts to drive new solutions which are improved according to set fitness criteria.⁹⁰

Feasible solutions for antimicrobial peptides in this work are explored through a genetic algorithm. In

peptide design, genetic algorithms are a common approach,⁹¹⁻⁹⁷ partially due to the motivation of genetic algorithms relating to biological sequences. While genetic algorithms are inspired by nature's genetic code, not all aspects of how the genetic code functions in nature



has been incorporated into genetic algorithms in peptide design. Integrating the codon table concept generates novel sequences to take advantage of the transition probabilities encoded in reading frame shifts. Reading frame shifts are changes to the nucleic acid base position which results in different codons being read downstream in the nucleic acid code. The DNA code of ATGATG would result in two codons for methionine if read from the first letter or in a stop codon, ending the transcription, if read from the second letter.

As part of Aim 1, the codon-based genetic algorithm addressed the reverse problem of CAMD (Figure 1.3, Step 3 and 4). Earlier in Aim1, CLN-MLEM2 addressed the forward problem of CAMD finding key physicochemical properties to distinguish active peptides from inactive peptides. In Aim 2, Step 2 of the CAMD process is exchanged from boundaries calculated from primary peptide structure (sequences) to boundaries calculated from secondary peptide structure (hydrogen bonds in the peptide backbone). Secondary structure features are of interest in chimeric peptide design because the spacer sequence impacts the folding of the solid binding peptide and the antimicrobial peptide. To understand the effect of the spacer domain on the folding of the chimeric peptide, this work addresses how to identify low free-energy structures of peptides.

1.3 Spacer Engineering through Peptide Structure

Peptide folding is the rearranging of peptides in their environments into active or inactive states. The modeling of peptide folding is the optimization of arrangements of peptide atoms in 3D space fitting a combination of empirical data or thermodynamic principles. Currently, there is a critical opportunity to understand sequence-structure relationships for peptides which are being engineered, and often, have unsolved structures. Peptides are often intrinsically disordered,⁹⁸⁻¹⁰¹ having non-crystalized active structures. Examples of these peptides are subsequences of the microtubule-associated tau protein¹⁰² and alpha-synuclein.¹⁰³ Therefore, a single conformation or small collection of conformations may be limiting in their ability to describe the structure of peptides and how the structure relates to the function of the peptide. A large group of conformations, which would be large enough for stable statistical estimates, may be necessary to describe the dynamic, non-crystalized active states of intrinsically disordered peptides. This large group of conformations will be referred to as an ensemble.

A recent review of *de novo* peptide folding methods groups approaches into two main categories.¹⁰⁴ The first group uses molecular dynamics to generate peptide structural ensembles.^{105, 106} While this approach is robust and can take into account buffering salts and even surface adsorption,^{22, 107, 108} it is computationally expensive and not ideal for screening peptide designs.

For screening peptide designs, the second category of available approaches of specialized peptide-specific approaches currently produces few peptide conformations per model. While the native structure may be contained in these structure models, the intrinsically disordered nature of peptides may need more diversity to describe occurring structure features.^{98, 99} PEP-FOLD3 is a tool for generating structural conformations for peptides from 5 to 50 amino acids long.^{109, 110} While the method generates 100 structures to cluster for the best clusters, the tool's user cannot investigate trends for the original 100 structures generated. The five best structures, one from each of the best clusters, is the user output. For the user, this approach is not ensemble structure analysis. PEP-FOLD3 uses a coarse-

grain modeling approach which generates time-elapsd motions like molecular dynamics, but it lacks effects due to thermodynamic effects of solvation, electrostatics, or salt bridges. Structural optimization with full-atom descriptions such as the Rosetta approach addresses these issues with a computationally inexpensive solution that identifies likely structures.

Spacer domains on the same polypeptide chains as peptide domains influence the folding of the peptide domains. Figure 1.2 shows an example of a chimeric peptide with the spacer domain labeled.

Understanding the effect of the spacer on chimeric peptide folding is a method for rationally designing chimeric peptides to integrate two functions simultaneously. In this work, solid binding peptides are integrated with antimicrobial peptides. Interdomain interference, or loss of function, is mitigated through spacer design.

1.3.1 Polypeptide Structure Levels

The distinction between peptides and proteins, which are both polypeptide chains, is based on length. Peptides are polypeptides with less than fifty residues, and proteins are polypeptides have more than fifty residues. Domains are subsequences within polypeptide chains that have a functional activity that can be demonstrated without the rest of the polypeptide chain. While the topology of polypeptide chains is simple in that they are linear chains of amino acids, the interactions between the amino acids within the same chain gives rise to a wide variety of structural properties. Some proteins result in soft, globular enzymes which catalyze specific chemical reactions through folding dynamics, while other proteins form the motors and the structural support of muscle cells to create motion through their combination of rigid and dynamic structural properties. As a fundamental concept of molecular biology, polypeptide structure has been long studied, but it still holds many unsolved mysteries. One unsolved problem is how linker regions in multidomain proteins can add to or subtract from the activity of attached domains.

To investigate this effect, first, the four levels of polypeptide structure are described. The first level is primary structure, which is the amino acid sequence. All three of the other levels relate back to the information in the amino acid sequence. The second level is the secondary structure, the hydrogen bonding of the polypeptide backbone. These are the general shapes of the polypeptide chains such as α -helices or β -strands, which leave out descriptions of where the side-chains of residues exist. The tertiary structure specifies where both the backbone and side-chain atoms are, either at a given instant in time in computational models or as a time-average through experimental measurement. The tertiary structure, however, is not the complete description of the polypeptide structure. The quaternary structure is how the residues of polypeptide domains interact with the rest of the polypeptide sequence. The quaternary structure can also change depending on external interactions. A long-studied example of quaternary structure is domain activity changes of hemoglobin based on pH changes and the concentration of carbon dioxide.¹¹¹⁻¹¹⁴ As polypeptide chains, chimeric peptides also have multiple levels of structure which are based on the order and composition of the amino acid sequence. To estimate what the secondary and tertiary levels of structure are for candidate amino acid sequences, computational structure generation models were developed as customized scripts from PyRosetta, a publicly available set of tools.

1.3.2 PyRosetta Folding

The Rosetta project makes building computational protein folding estimates of secondary or tertiary structure a modular process.¹¹⁵⁻¹¹⁷ In fact, Foldit is a game for humans to learn how to fold proteins to develop better folding algorithms using the modular protein folding moves built into Rosetta.^{118, 119} In this work, a different branch of the Rosetta project was used. PyRosetta is a Python interface for the Rosetta project built on C source code. A custom PyRosetta script was built for this work, given in Appendix B. In developing the custom folding algorithm, changes in conformational space were evaluated with the Classic Relax method in PyRosetta. Representative structures are shown in Figure

1.4. While the peptide side chains seem to change orientation, almost no backbone movement occurs. Therefore, the Classic Relax method appears to be limited to local moves in peptide folding space.

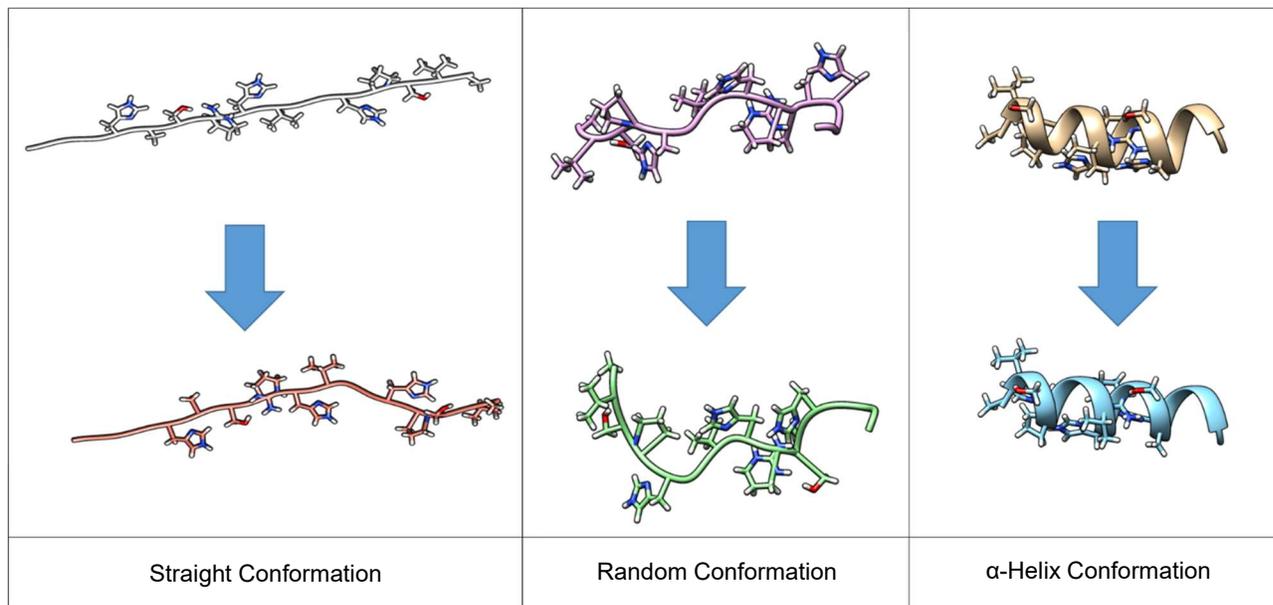


Figure 1.4 Classic relax protocol for three initial starting conformations of a titanium binding peptide (AVSPHGVHRS AHGGG).

1.3.3 Fragment Insertion

A solution to the limitation of small moves of the backbone conformations in Figure 1.4 is to use the fragment insertion methods built into Rosetta. Fragment insertion results in larger backbone conformation changes, resulting in a much wider search of the peptide conformational space. Using fragment insertion biases the search of the conformational space to conformations of solved protein structures. This bias is an advantage when searching for thermodynamically likely peptide folding in a short computational time.

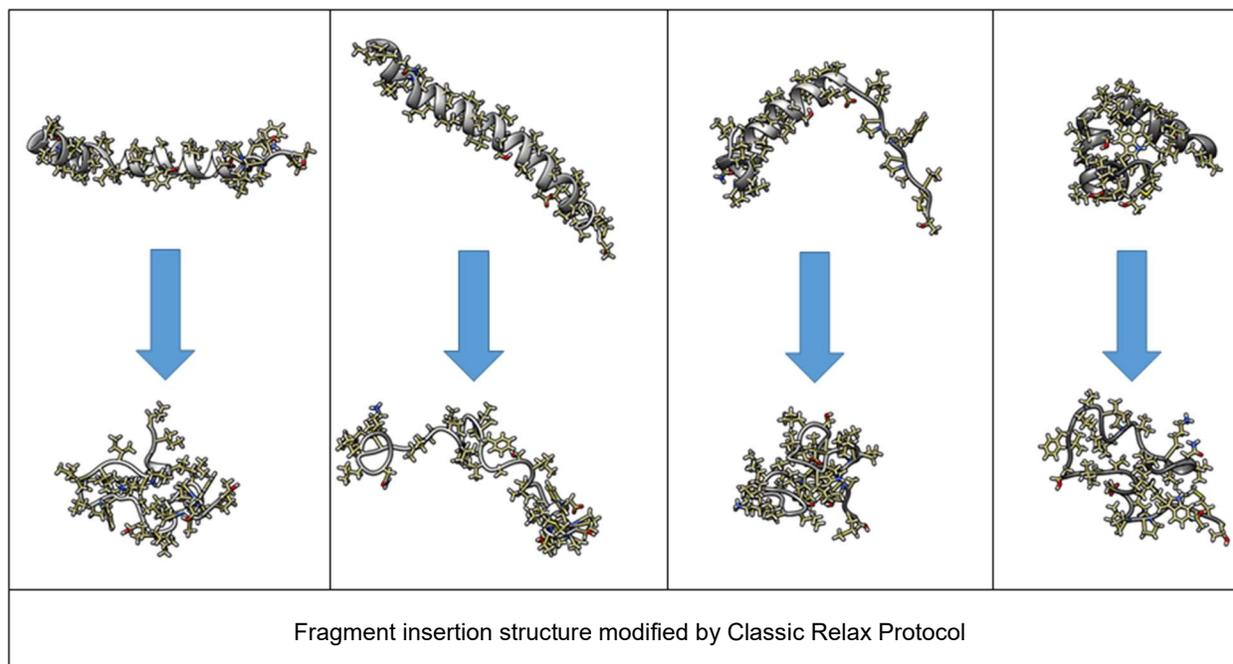


Figure 1.5 Fragment insertion demonstrates large moves in conformational space compared to the moves from the Classic Relax protocol for a titanium binding peptide (TMTAPWPPVLDALGVLSAVFLPVILPVIGKLLNGIL).

Structure-based relationships are integrated into the design of new chimeric peptide sequences as part of the CAMD process in Chapter 4. (Figure 1.3). In Chapter 2, a rough-set theory method not yet explored for peptide or protein classification is compared for antimicrobial peptide classification to other methods in the field. The sequence-function relationships discovered in Chapter 2 through the developed rough set theory method is part of how to design antimicrobial peptides that are effective against any bacterial species of interest through iterative cycles of computational analysis and efficacy determination in Chapter 3. How these methods fit together is described in Specific Aims.

1.4 Specific Aims

The success of implants requires the formation and maintenance of an integrated interface.¹²⁰⁻¹²²

Revision surgeries, the result of some failed implants, are up to 15% of hip replacements and 5% of dental implants.¹²³ The prevention of infections, which disrupt integrated interfaces, relies on the

application of antibiotics, which brings a major concern due to the rise of bacterial antibiotic-resistance.¹²⁴ There is an urgent need for a method to develop new antimicrobial agents as well as an improved method for their effective delivery at the implant site.^{125, 126}

Orientation control of antimicrobial peptides is an essential step for biomolecule activity. This work developed a design method for chimeric antimicrobial peptides that can self-assemble and self-orient onto implant surfaces in a single step while preserving their antimicrobial activity. Three specific aims to address this design challenge are: 1) Develop classification and design methods using rough set theory and intelligent search to find novel antibacterial peptides; 2) Develop chimeric peptides by designing spacer sequences to effectively display antimicrobial and self-assembling functions onto the Ti and Ti-Alloys; 3) Verify the approach as an enabling technology by expanding the chimeric AMP design to another common biomaterial, i.e. calcium phosphate mineral, and a dental adhesive interface.

In Aim 1, antibacterial peptides are classified as is shown in Chapter 2. Customized antibacterial peptides are designed, in Chapter 3, through computational tools using rough set theory and a genetic algorithm to find novel antibacterial peptides. The rough set theory approach^{60, 127-129} identifies boundaries of key chemical properties specific to antibacterial activity through inducted rules.

Inspired by the fact that spacer sequences play critical roles between functional protein domains, in Aim 2, chimeric antimicrobial peptides are designed in Chapter 4 with spacer sequences to reduce the interference between different peptide domain sequences. The chimeric antimicrobial peptides are first demonstrated with titanium implants, the most common biomaterial for implants. The spacer domain linking the solid binding peptide and the antimicrobial peptide is engineered through secondary structural features of the antimicrobial peptides.

In Aim 3, the platform developed in Aims 1 and 2 is demonstrated as an enabling technology. Chimeric peptides were designed and tested for another common biomaterial, calcium phosphate in Chapter 5.

Chimeric peptides of two different kinds were explored for calcium phosphate. Chimeric antimicrobial peptides, analogous to titanium, and chimeric mineral forming peptides were designed. The platform is also extended in Chapter 5 by the design of chimeric antimicrobial peptides for zirconia surfaces and urethane surfaces. A novel application of antimicrobial peptide integration in a dental adhesive system is demonstrated in Chapter 6.

2.0 Antimicrobial Peptide Classification through Rough Set Theory

2.1 Introduction

Antimicrobial peptides attract considerable interest as novel agents to combat infections. Their long-time potency across bacteria, viruses and fungi as part of diverse innate immune systems offers a solution to overcome the rising concerns from antibiotic resistance. With the rapid increase of antimicrobial peptides, peptide selection becomes a challenge. To address this problem, similarity analyses of the wide variation of the physicochemical properties of antimicrobial peptides was used to describe key properties that distinguish between active and non-active peptide sequences. An iterative supervised learning approach was used to identify separate groups of active peptides and inactive peptides.

By generating explicit boundaries, this method defines new categories of active and inactive peptides based on their physicochemical properties. Consequently, it describes physicochemical characteristics of similarity among active peptides and the physicochemical boundaries between active and inactive peptides in a single process. A rough set theory approach is used to build the similarity boundaries; to the author's knowledge, this is the first time that this approach has been used to classify peptides. The modified rough set theory method limits the number of values describing a boundary to a user-defined limit. This method is optimized for specificity over selectivity. Noting that false positives increase activity assays while false negatives only increase computational search time, this method provided a low false discovery rate. Published datasets are used to compare this rough set theory method to other published classification methods and based on this comparison, high selectivity and comparable sensitivity to currently available methods are achieved.

Rule sets are developed that define physicochemical boundaries which allow us to directly classify the active sequences from inactive peptides. Existing classification methods are either order insensitive or length-dependent, whereas this method generates the rule sets that combine order-sensitive descriptors with length-independent descriptors. The method provides comparable or improved

performance to currently available methods. Discovering the boundaries of physicochemical properties may lead to a new understanding of peptide similarity.

2.2 Background

In the US, over 23,000 deaths each year are associated with drug-resistant bacterial infections.¹³⁰ These types of infections are central to the projected increase in deaths globally by 2050, which are expected to reach 10 million annually.^{131, 132} The rise of antibiotic-resistant bacteria has prompted increasing interest in antimicrobial peptides as a solution to this critical issue.¹³³

Over 2,800 antimicrobial peptides have been discovered from natural sources in the last decade.^{65, 134-139} Antibacterial peptides derived from these natural sequences have shown both broad-spectrum and improved activity against targeted bacteria.¹⁴⁰⁻¹⁴² Antibacterial peptide-mimics are introduced as another source to the existing peptide libraries by incorporating additional backbone chain atoms for more structural flexibility and resistance to protease degradation.¹⁴³⁻¹⁴⁶ Post-translationally modified antimicrobial peptides are also being explored by going beyond the chemical properties of the naturally occurring amino acids.^{147, 148}

While many antimicrobial peptides have been discovered at the laboratory bench, computational methods have been integrated into this search to find many more candidates. Encrypted antimicrobial peptides are an example in which known active peptides are queried against DNA repositories to find new antimicrobial peptides.¹⁴⁹ Among many methods, grammar-based methods and regular-expression-based match sequence patterns are used to identify functional similarity.^{10, 64} Computer-aided molecular design^{78, 150-152} approaches using quantitative sequence activity relationships^{75, 153-155} (QSAR) predict the antibacterial level of peptides given key chemical properties. Artificial neural networks (ANN) have been used both to generate new sequences and to distinguish between active and inactive sequences.^{10, 89, 156-158} They are often used in the classification of antimicrobial peptide sequences.^{136, 159} While ANNs are

flexible enough to model many kinds of complex relationships, they lack transparency about how classification choices are made. Determining the boundaries of the similar antimicrobial peptide clusters remains difficult despite many existing machine learning methods.

Due to the ongoing need for improved antimicrobial peptide selection and design, many classification approaches have been developed with supervised machine learning methods. A recent review by Porto *et al* contrasts two different kinds of sequence representations for antibacterial classification.¹⁰ The first kind of representation preserves the order of the sequence which tends to lead to length-dependent predictions.¹⁶⁰ False positives may be produced if the overall chemical property of an antibacterial peptide is changed by adding amino acids with contradictory chemical properties. The second kind of sequence representation preserves overall sequence properties which tends to lead to order-insensitivity. False positives may be produced if the order of an active peptide is scrambled.⁶⁴

AntiBP¹⁶¹ was one of the first online available services for antibacterial peptide prediction. AntiBP uses a sliding window of 15 residues to predict the classification using support vector machines (SVM)¹⁶², quantitative matrices (QM)¹⁶³ and artificial neural networks (ANN).¹⁶⁴ The strength of this approach is that the order of amino acids impacts the prediction. However, the weakness to having a constant window of amino acids is that the predictions are peptide-length dependent.¹⁶⁰ To overcome the peptide length dependence, another sequence representation method was employed as CAMP,⁶⁵ with descriptors that summarized composition, physicochemical properties and structural features of the peptides. CAMP uses multiple machine learning approaches for these features such as SVM¹⁶⁵, ANN⁶⁶,¹⁶⁶, discriminate analysis (DA)¹⁶⁷ and random forest (RF).¹⁶⁸ However, the descriptor approach is insensitive to the sequence order arrangement. For example, full-length sequence descriptors can be sensitive to the overall charge of a peptide but not its charge distribution. iAMP-2L⁶⁷ uses some descriptors of correlations between residue positions of peptides, but some of the descriptors are order

insensitive.⁶⁴ This approach is order insensitive to sequence rearrangements that preserve the correlation structure from the original peptide.

The Evolutionary Feature Construction^{12, 68} (EFC) method, which achieves order-sensitive classification combines order sensitivity and length independence by selecting common chemical property sequence patterns for antimicrobial peptides. Length-independent classification is achieved with a support-vector machine method through physicochemical descriptors selected by FCBF (Fast-Correlation Based Filter selection). While this method does combine order-sensitivity and length-independence, it does not completely address either of these issues. Order-insensitivity is possible based on the rearrangements of amino acids that are indistinguishable by the pattern recognition scheme of compressing twenty amino acids into four categories.

In this work, a novel method addresses order sensitivity by calculating the physicochemical properties of subsequences in addition to using descriptors of physicochemical properties for length independence. This work, therefore, combines order-sensitivity and length independence as a new approach. These descriptors are analyzed using rough set theory (RST). Descriptors that are calculated for both physicochemical properties of the full peptide sequences, which are order insensitive, and descriptors of constant-length subsequences, which are order sensitive. RST selects combinations of both kinds of descriptors into a single rule. Each rule defines its own cluster including the classification of the peptide's activity or inactivity. To the author's knowledge, RST has not yet been studied to classify peptide sequences based on their activity.

Using a rough set theory approach that combines the algorithm of MLEM2 (modified learning from examples module, Version 2) with the algorithm IRIM (Interesting Rule Induction Module), a method is developed that investigates the sequence-function relationships. A high specificity performance by the condition-limit number MLEM2 is achieved with the fewest chemical property features among

benchmarked methods. This method is tested against publicly available prediction servers CAMP (Collection of Antimicrobial Peptides) AMP prediction⁶⁵, iAMP-2L (antimicrobial peptide prediction two-level)⁶⁷, and a motif-searching algorithm EFC (Evolutionary Feature Construction) method^{12, 68} with and without FCBF (Fast-Correlation Based Filter selection). This approach produces physicochemical boundaries that create definitions of similarity among antimicrobial peptides and non-antimicrobial peptides.

2.3 Methods

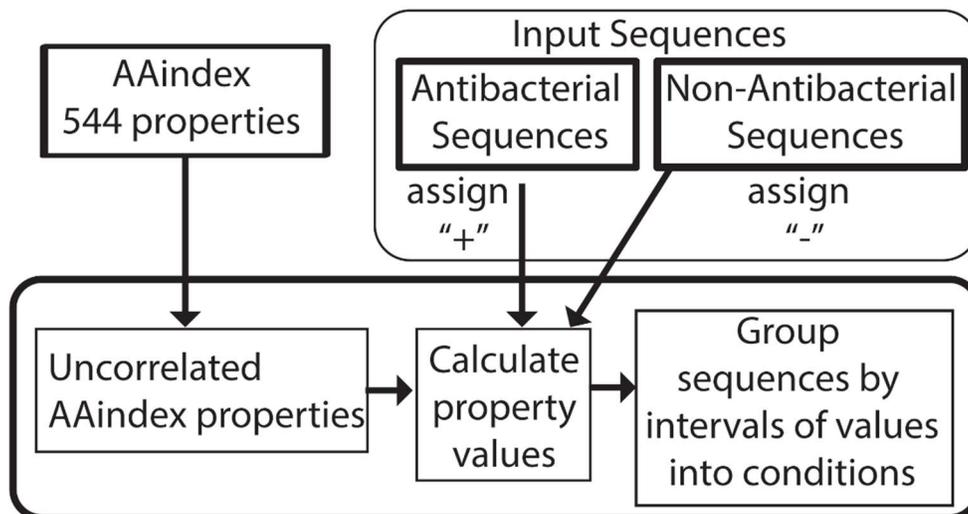
In this work, a rough set theory classification method is developed to differentiate antibacterial peptides from APD2¹³⁸ (Antimicrobial Peptide Database 2) and randomly selected peptides from the UniProt database.^{169, 170} These benchmark datasets are available online.^{67, 171}

2.3.1 Rule Induction by the MLEM2 Algorithm

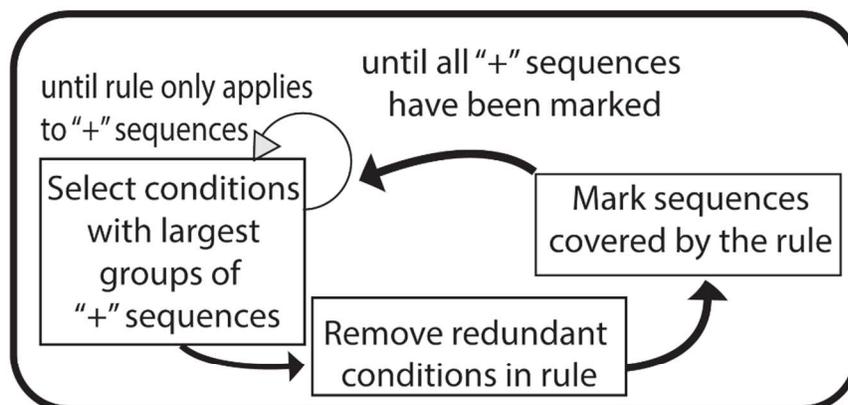
The MLEM2 rule induction method⁶⁰ is a classification method based on a rough set theory approach that uses local approximations when the available columns in the data table cannot perfectly separate the data. A local approximation is finding the value intervals for data table columns that generate rules with a user-set minimum accuracy. This accuracy parameter is called α . MLEM2 was modified to combine the polynomial run time growth rate of MLEM2 with the condition-limit number of the IRIM (Interesting Rule Induction Method) to find rules with small numbers of conditions in large datasets with many attributes. IRIM has an exponential run time growth rate with respect to attribute number. The maximum number of conditions was set to eight (8). Conditions are intervals of feature values. Each peptide sequence has one value for each feature. Rules are conjunctive expressions of conditions. Rules are built from conditions that contain the maximum number of peptide sequence of the desired antibacterial label. Ties are broken by the conditions that have the highest percentage of peptide sequences with the desired antibacterial label. Rules are refined by narrowing the interval of an included condition or by adding a new condition to the conjunctive expression. Rules are simplified by

omitting redundant conditions whose loss still results in a rule with no loss of accuracy. The minimum accuracy that a valid rule must have is a user-defined value, α . In this study, α is set to the accuracy of the majority class rule, which is to label all peptides with the non-antibacterial class. Figure 2.1 provides an overall schematic of the CLN-MLEM2 method.

1. Build Conditions



2. Build Rules



3. Refine Rule Sets

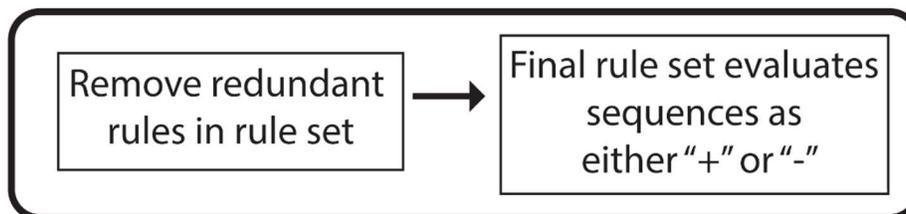


Figure 2.1 CLN-MLEM2 Method.

CLN-MLEM2 Rule induction process based on rough set theory approach to classify peptides with antibacterial activity.

2.3.2 Correlated AAindex1 Property Removal

The AAindex1 has 544 properties with one value for each of the twenty naturally occurring amino acids.⁸⁰ A database of all properties is available in the R package 'seqinr'.¹⁷² An autocorrelation matrix was constructed from these properties to provide pairwise correlation comparisons for all 544 properties. Properties were filtered using an absolute correlation value cutoff. Properties were compared in randomized order.

2.3.3 Feature Generation

Datasets for this classification system are lists of peptide sequences. Each peptide sequence is a list of letters representing amino acids in a polypeptide chain. The lengths of peptide sequences vary in the dataset. Features for this classification system are single number summaries of the entire peptide sequence. Calculations of the summaries are given Table 2.1.

Table 2.1 Description of summary functions to generate chemical property features as input for the MLEM2 method.

Summary Function	Description
Sum	$\sum a_i$, where a_i is the amino acid chemical property at position i from 1 to n , the length of the peptide sequence
Mean	$\sum \frac{a_i}{n}$, where a_i is the amino acid chemical property at position i from 1 to n
Window 3	$\max(\sum_i^{i+3} a_i)$, where a_i is the amino acid chemical property at position i from 1 to $n-3$

The explosion of available antimicrobial peptides brings the new challenge of selecting which antimicrobial peptides to use.^{159, 173-175} With the large increase in the number of available peptides, selecting specific peptides to use is aided by finding similarities among peptides. In this work, similarities are defined by boundaries between active peptides that also differentiate active peptides from inactive

peptides. The rule sets generated have at least one rule for every peptide in the training set, meaning that all peptides belong to at least one group.

The data tables for the peptide training sets include different columns. Each unique row value is linked to a list of peptides which share that value. For example, in case of examining the overall charge of the peptide, the column the sum of the overall charge. Each unique charge value, such as +1 would be linked to all peptides whose overall charge was +1. Value intervals form sets of peptides that are labeled with values in the interval. For overall charge, an interval would be all peptides with charges from +1 to +5. This peptide set whose values are included in the value interval is an example of a condition. An inducted rule is the set intersection of conditions. (Figure 2.2) Evaluating the performance of the rules being generated is done by calculating the Pr , the probability that a peptide covered by the rule has the activity assigned by the rule. The Pr is the ratio of the size of the sets of peptides described by the intersection of all the conditions in the rule that meet the targeted label to all the peptides described by the intersection of the conditions (Equation 2.1). The value of Pr must be at or above α , a user-defined cut-off limit for any rule that is part of the output of the method.

Equation 2.1

$$Pr = \frac{|\cap_1^{CLN} C_i|_{targeted\ label}}{|\cap_1^{CLN} C_i|_{any\ label}}$$

Figure 2.2 is a Venn diagram separating active and inactive peptides. A rule is the intersection of conditions (C_1 and C_2). Each rule must be selective for either active or inactive peptides. The minimum Pr value allowed for a rule is a user-defined parameter α . The selection of conditions that lead to rules is a feature selection process that chooses the most relevant conditions to describe the physicochemical boundaries. The CLN value is the user-defined condition-limit number which limits the number of

conditions in each rule. A rule set is the collection of all rules describing the boundaries for either activity or inactivity.

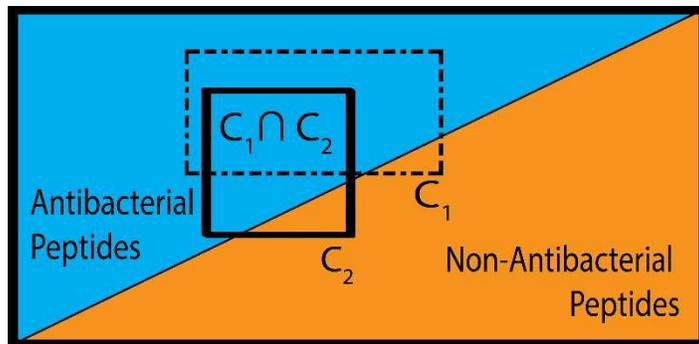


Figure 2.2 Rough Set Theory Rule Generation.

To create the CLN-MLEM2 method, features of MLEM2 (modified learning from examples module, Version 2) method^{176,177} are combined with a feature of the module IRIM (Interesting Rule Induction Module) to potentially improve the selectivity and specificity.¹⁷⁸ MLEM2 method is modified by adding the ability to limit the condition number for each of the rules, a feature of IRIM. Because the IRIM method exhaustively searches all possible rules given the number of conditions, it cannot be used for large numbers of conditions or large numbers of peptides because the runtime grows exponentially with the number of conditions.

This modified MLEM2 method uses the heuristics of the MLEM2 method to select condition combinations with a run time that grows polynomially in the number of peptides and in the number of conditions. This modified method includes a condition-limit number (CLN) which combines the polynomially-bound worst-case runtime of MLEM2 with the set number of conditions of IRIM. Because a

small number of conditions are selected from the available number of conditions, CLN-MLEM2 is an embedded feature selection method.¹⁷⁹ It attempts to use the most relevant conditions to describe the boundaries. The relevance of a condition is the number of peptides that are described by it in the training set. The CLN-MLEM2 method selects rules based on a user-defined minimum accuracy referred to as α ($0 \leq \alpha \leq 1$). Using higher values of α generates fewer rules with higher *Pr* values of training accuracy. Using lower values of alpha generates more rules with lower *Pr* values of training accuracy. CLN-MLEM2 generates rules until all peptides in the training set have at least one rule that applies to it. The collection of all rules for either active or inactive peptides is called a rule set.

2.3.4 Performance Measures

When measuring the performance of a prediction, there are four possible prediction outcomes for each sequence predicted: true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

True refers to correct classification while false refers to a misclassification. Positive refers to a classification of antibacterial and negative refers to a classification of non-antibacterial. Multiple performance measures were used to understand how well a classifier can predict positive or negative cases. The first is sensitivity, which is the frequency of predicting antibacterial sequences as antibacterial.

Equation 2.2

$$Sensitivity = \frac{TP}{TP + FN}$$

The second type is specificity, which is the frequency of predicting non-antibacterial sequences as non-antibacterial. False discovery rate is its complement, the frequency of predicting non-antibacterial sequences as antibacterial.

Equation 2.3

$$Specificity = \frac{TN}{TN + FP} = 1 - False\ Discovery\ Rate$$

Lastly, a performance measure is used that combines elements of sensitivity and specificity into a single measure called the Matthew's Correlation Coefficient (MCC). This measure is derived from the Pearson correlation coefficient for predictions when classifications are binary.¹⁸⁰ Because it is a special case of Pearson's correlation, the maximum value is 1 when the prediction is perfect, 0 when random and -1 when all predictions are false.

Equation 2.4

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

2.4 Results

To begin the condition-limit number MLEM2 (Modified Learning from Experience Module 2) method, multiple summaries are generated of the amino acid sequences of the given active and inactive peptides by selecting non-correlated amino acid properties in the AAindex1⁸⁰ (Amino Acid index 1). Of the 544 properties of the AAindex1, many of the properties are highly correlated. The autocorrelation among AAindex1 properties with each other is shown in Figure 2.3A. Positive correlation is magenta and negative correlation is teal. Non-correlated amino acid property pairs are white. The autocorrelation matrix shows that most amino acid properties are highly correlated. How many amino acid properties are below a correlation threshold was studied for all other amino acid properties (Figure 2.3B). 60 repetitions were performed with random initial properties of eliminating properties more correlated than a threshold. A very tight trend of how many uncorrelated properties there are for a given cut-off value was seen. For further study, a correlation cut-off of 0.65 was selected, which resulted in 74 properties remaining from the original 544 properties.

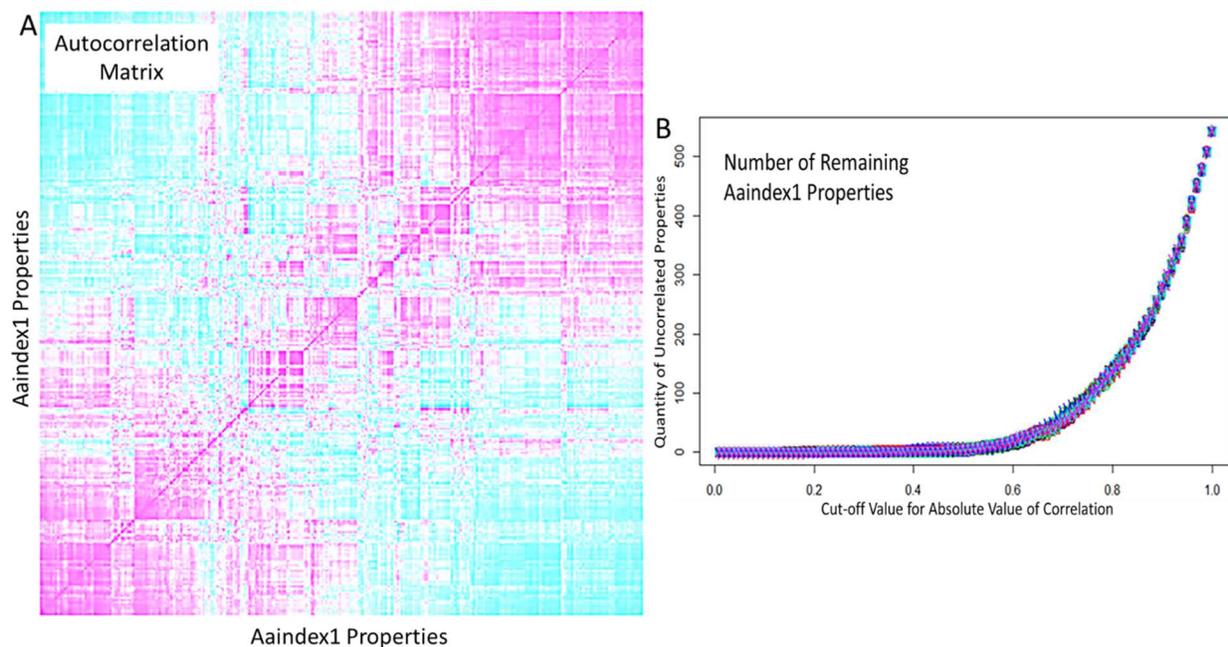


Figure 2.3 Auto-Correlation and Selection of AAindex1 Properties.

Combining overall sequence chemical properties with motif properties helped to build sequence representation that was sensitive to changes in order or changes in length that effect the peptide activity. If only chemical properties are evaluated by the sum or mean of the whole sequence, then the rules generated are order insensitive. By considering subsequences of the peptides, then the ordering of the chemical properties within the sequence can be used as a feature. Two types of sequence property summaries were calculated from the selected amino acid properties in the AAindex1 (Amino Acid index 1) after removing the correlated amino acid chemical properties. First, the overall property summaries were calculated as the mean and average of the properties of the amino acids present in the sequence. Secondly, motif properties were calculated as the maximal subsequence sum of a given length of the amino acid sequence. The CLN-MLEM2 method can combine overall sequence properties and motif properties within a single rule. Each rule forms a class of either active or inactive peptides.

Previously studied, publicly available datasets of antimicrobial peptides^{67, 171} are used to test the developed method of finding physicochemical boundaries for antibacterial activity. Table 2.2 shows the

inducted rule category with the largest membership of the studied dataset. The rule category is the conjunctive expression of each of the conditions up to the user-defined condition-limit number with the rule applying to antimicrobial peptides whose property values are within the range of the values given in Table 2.2.

Equation 2.5

$$\bigcap_1^{n=CLN} (Lower Value_{condition} \leq Value_{peptide} \leq Upper Value_{condition}) \xrightarrow{predicts} Antibacterial Activity$$

Table 2.2 has the rough set theory rules found with maximum support from large training dataset. The first rule describes antibacterial sequences. The accuracy of this rule is 97.8% (446/456) for the peptides that met the conditions from either the dataset from Xiao, *et al*¹⁸¹ or the dataset from Fernandes, *et al*.¹⁷¹ All sequences that do not match any rule for the applied rule set are classified as non-antibacterial.

Table 2.2 Maximum Support Rough Set Theory Rules

Calculation	AAindex1 Property	Lower Value	Upper Value
Window 3	NAKH900111	31.21	48.66
Window 3	FINA910104	3.45	5.10
Window 3	KUMS000101	23.6	28.20
Sum	GEIM800102	12.68	39.90
Window 3	VASM830102	1.67	2.12
Window 3	QIAN880139	0.38	0.98
Sum	FAUJ880112	0	3
Sum	CHAM820102	-0.61	19.51

2.4.1 Rough Set Theory Analysis

In this work, a data mining approach is taken with rough set theory to understand how to describe antibacterial activity in terms of amino acid properties. For defining active and inactive sequences,

training sets from the published iAMP-2L dataset were used.⁶⁷ The positive training set consisted of 1,274 sequences which originated from APD2¹³⁸ and the negative training set consisted of 1,440 sequences which originated from intracellular protein sequences from UniProt. To select from amino acid physicochemical properties, the AAindex¹⁸² was used from the 'seqinr' package in R.¹⁷² A correlation cut-off of 0.6 was selected, leaving 74 AAindex features from which to select. The overall properties of the sequences are described by one of three mathematical operations: sum, mean or window. These operations are described in Table 2.1. The CLN-MLEM selected from among 74 AAindex features x 3 operations = 222 features. CLN-MLEM2 is an embedded feature selection method with the limit of 8 conditions per rule was set. These AAindex features are listed in Table 2.3.

Table 2.3 Description of AAindex features selected by CLN-MLEM2 method.

Feature Selected	AAindex Description
AURR980118	Normalized positional residue frequency at helix termini C ¹⁸³
CHAM820102	Free energy of solution in water, kcal/mole ¹⁸⁴
FAUJ880112	Negative charge ¹⁸⁵
FINA910104	Helix termination parameter at position j+1 ¹⁸⁶
GEIM800102	Alpha-helix indices for alpha-proteins ¹⁸⁷
GEIM800103	Alpha-helix indices for beta-proteins ¹⁸⁷
GEOR030101	Linker propensity from all dataset ¹⁸⁸
KUMS000101	Distribution of amino acid residues in the 18 non-redundant families of thermophilic proteins ¹⁸⁹
NAKH900111	Transmembrane regions of non-mt-proteins ¹⁹⁰
QIAN880102	Weights for alpha-helix at the window position of -5 ¹⁹¹
QIAN880126	Weights for beta-sheet at the window position of 6 ¹⁹¹
QIAN880138	Weights for coil at the window position of 5 ¹⁹¹
QIAN880139	Weights for coil at the window position of 6 ¹⁹¹
RICJ880115	Relative preference value at C-cap ¹⁹²
ROBB760107	Information measure for extended without H-bond ¹⁹³
VASM830101	Relative population of conformational state A ¹⁹⁴
VASM830102	Relative population of conformational state C ¹⁹⁴
WERD780103	Free energy change of alpha(Ri) to alpha(Rh) ¹⁹⁵
YUTK870103	Activation Gibbs energy of unfolding, pH 7.0 ¹⁹⁶

2.4.2 Training Performance by Rule Coverage

The accuracy of the rules is the ratio of the support, the number of cases correctly classified by the rules, to the coverage, the number of cases to which the rule applies, as shown in Figure 2.4. Below a coverage of 30, all rules had 100% training accuracy. For the rules that applied to 100 or more cases, the training accuracy was below 100% but above the 95% threshold set for α , the user-set minimum requirement for accuracy.

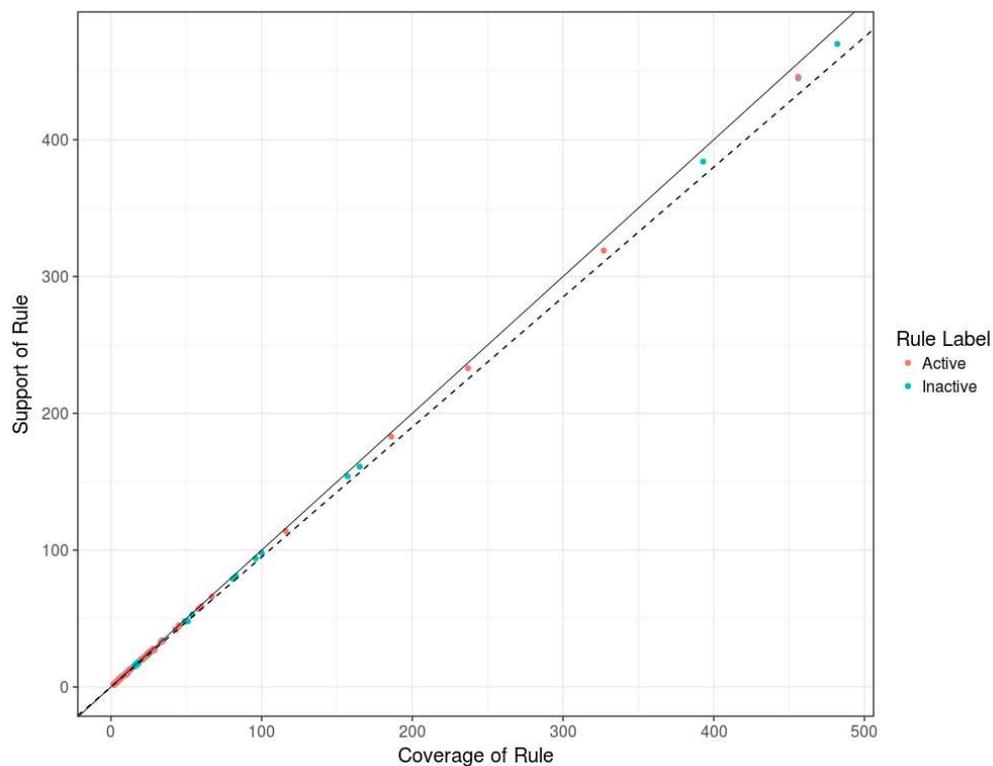


Figure 2.4 Training accuracy by rule size.

2.4.3 Training Performance by Features Selected

The training accuracy of the rules by the features selected is in Figure 2.5. The accuracy for each of the features was divided into rule sets for active peptides (+) or inactive peptides (-). Most of the features in Figure 2.5 have 100% accuracy, corresponding to rules with coverage of less than 30 cases. All features selected had an average training accuracy greater than 96%.



Figure 2.5 The training set accuracy of CLN-MLEM2 rules by features selected.

The distribution of each of the features among the rules selected is shown in Figure 2.6. The number of rules for each of the amino acid properties selected by each calculation made. The rules are an embedded feature selection approach. The grid is colored according to the relative times each feature was selected. The number of rules each feature occurs in is given in the colored grid for each of the rule sets (active or inactive).

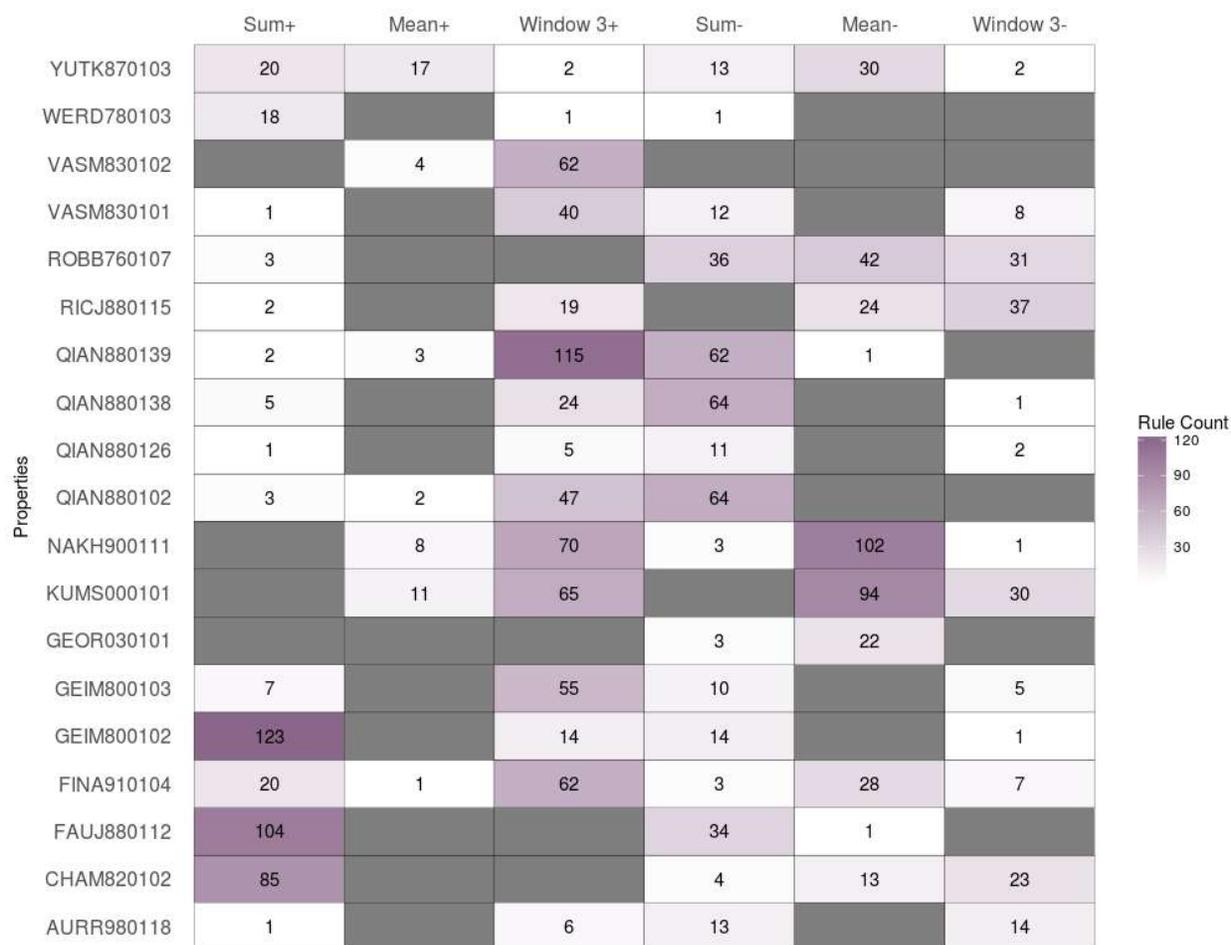


Figure 2.6 The frequency of selection of features by CLN-MLEM2 rules.

2.4.4 Benchmarking CLN-MLEM2

In this work, the method of using rough set theory to build rules that distinguish between active antibacterial peptides from inactive antibacterial peptides is benchmarked against a recently published method EFC⁶⁸, based on motif-recognition, and then against a larger set of methods from publicly available prediction servers. The first benchmark test is a ten-fold cross validation on a dataset used in previous studies^{68, 171} with the positive sequences clustered from the APD2 (Antimicrobial Peptide Database 2)¹³⁸ to 115 clusters and the negative sequences from the PDB¹⁹⁷ clustered to 116 clusters. Each cluster is represented by one sequence. The results are compared with EFC-based methods and support vector machines given subsequences of lengths 5 to 8 amino acids. Table 2.4 demonstrates that

the developed method has high selectivity and accuracy in comparison to the performance of the SVM methods, and comparable selectivity and accuracy in comparison to the EFC method. A trend of decreasing Mathew's Correlation Coefficient (0 for random guessing and 1 for perfect performance) as the length of the subsequence increases is seen in Table 2.4.

Table 2.4 Performance of rough set theory rule induction compared to motif-search in 10-fold cross validation.

Method	Sensitivity (%)	Specificity (%)	MCC
5-kmer SVM	75.7	75.0	0.54
6-kmer SVM	74.8	74.1	0.46
7-kmer SVM	73.0	72.4	0.40
8-kmer SVM	73.0	72.4	0.36
EFC-FCBF	87.1	87.2	0.76
CLN-MLEM2	86.9	86.3	0.75

The CLN-MLEM2 method is further tested against a larger variety of classification methods. The second benchmarking test uses the iAMP-2L dataset.⁶⁷ Like the dataset used for the first benchmark, this dataset is derived from the APD2 database. However, instead of choosing a single sequence from each cluster, the sequences are narrowed by removing sequences with greater than 40% similarity as measured by CD-HIT¹⁹⁸ only with cluster of more than 250 sequences. These sequences are reduced to a dataset of 848 unique sequences. The negative sequences are from a UniProt search of cytoplasmic proteins, also with less than 40% similarity. 2,405 unique sequences are included in the negative dataset.

Table 2.5 Performance comparison among prediction servers for antimicrobial peptides, a motif-based classification method and rough set theory approach.

Method	Sensitivity (%)	Specificity (%)	MCC
CAMP SVM	95.8	39.8	0.43
CAMP RF	97.1	33.5	0.40
CAMP ANN	89.1	70.9	0.61
CAMP DA	94.1	49.5	0.49
iAMP-2L	97.7	92.0	0.90
EFC-FCBF	92.0	90.0	0.73
EFC+307-FCBF (307 AAindex1 features)	92.4	96.1	0.86
CLN-MLEM2 (74 AAindex1 features)	88.0	95.4	0.85

2.5 Discussion

The feature selection process is embedded in the CLN-MLEM2 algorithm. Only 19 of the 74 AAindex features (25.7%) are represented in the rules generated. 46 of the possible 222 features (20.7%) are represented in the rules generated. While many of the 19 AAindex features in Table 2.3 have previously been identified as of interest in relating antimicrobial peptide activity,^{199, 200} several novel AAindex features relating to antibacterial activity have been identified with the CLN-MLEM2 method such as QIAN880139 and ROBB760107.

The training error of the rules, $1 - Pr$, increases as the coverage of the rules goes over 100 cases. There is a limit in the accuracy of the benchmarking data. While the positive training set cases have experimental evidence demonstrating antibacterial activity, the negative data set does not have direct experimental evidence for not having antibacterial activity. So, while the false positive rate is most likely very low from the experimental evidence supporting the labels, the false negative rate for the

benchmark set is unknown. The rough set theory rules identify candidates for experimental testing which are more likely to be false negatives in the negative training set used.

This work looks at which features have the most impact for finding either antimicrobial peptides or non-antimicrobial peptides. The sum of GEIM800102, the amino acid frequencies among alpha-helix indices for proteins dominated with alpha helix structure, was the most common feature selected for generating rules for describing antibacterial peptides appearing in 123 rules. The window of GEIM800102 only appeared in 14 rules. The AAindex feature QIAN880139, weights for coil at the window position of 6, had the opposite trend for rules describing active peptides. The window calculation was selected for 115 rules, while its sum was only selected for two rules.

No AAindex feature was used in more than 10 rules across all three calculation types for active peptides. ROBB760107, information measure for extended without H-bond, relates to the likelihood of extended conformations of the peptides. This AAindex feature was the only one to be selected for more and ten rules across all three calculation types for rules describing inactive peptides. This AAindex feature was only selected for three rules for active peptides. FINA910104 and YUTK870103 were the only AAindex features selected for all three calculation types for active and inactive peptide rules. YUTK870103 relates to the additive contributions of amino acids for free energy of folding at pH 7, near physiological conditions. FINA910104 relates to amino acid likelihood for being the terminal position of an α -helix.

The mean of NAKH900111 is the most common feature selected for the inactive peptide rule set, and the window of NAKH900111 was often selected to describe active peptides. NAKH900111 relates to the frequency of amino acids in membrane proteins that are not found in mitochondria. Therefore, one of the key differences between experimentally verified antimicrobial peptides from ADP2 and intracellular proteins listed in UniProt is the likelihood that amino acid sequences are stable within cellular membranes. For describing active peptides, the mean amino acid frequency is useful, while for

describing intracellular proteins, the 3-amino acid window sum is useful. The sum of the charge of the amino acid was often important, occurring in 104 rules describing active peptides and in 34 rules describing intracellular protein sequences.

The CLN-MLEM2 method has high specificity and similar accuracy for antibacterial classification as other current methods. When using a classification method for the discovery of antimicrobial peptides, the specificity of the method is more important than its selectivity.²⁰¹ The current method prioritizes specificity with a low false discovery rate (FDR) by classifying sequences that do not meet any rule in the applied rule set as inactive (Figure 2.7). In fact, there is only one method which provides lower FDR compared to the CLN-MLEM2 method, EFC+307-FCBF. However, the CLN-MLEM2 method results in similar specificity starting with fewer physicochemical properties. The method also acts as an embedded feature selection tool by limiting the physicochemical properties in the rules to a user-defined limit.¹⁷⁹

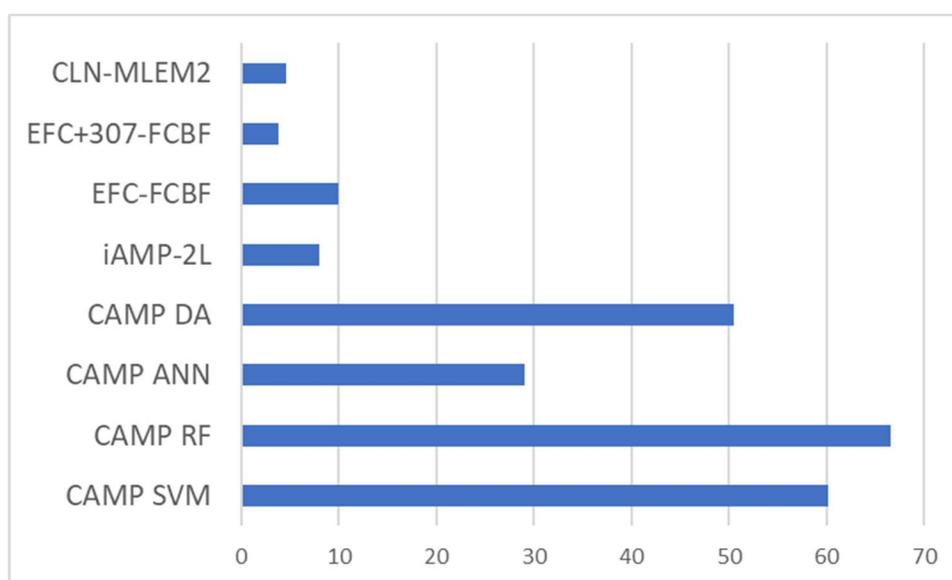


Figure 2.7 False discovery rates of antimicrobial peptide classification.

While the CLN-MLEM2 method has comparable selectivity in classification to current state-of-the-art method, the method is among the best in specificity (Table 2.5). The combination evolutionary algorithm with chemical properties (EFC+307-FCBF: EFC combined with FCBF (Fast Correlation Based

Features) using 307 features) is the only other state-of-the-art method with specificity that is comparable to CLN-MLEM2. Similar specificity is achieved using 74 AAindex1 features instead of 307 AAindex1 features. When removing the length-independent representation from the EFC method (EFC-FCBF: EFC without FCBF) results in almost no loss of sensitivity, but a loss of 6% in selectivity. Removing the order-sensitive representation for EFC in Table 2.2 results in lower sensitivity and selectivity performance (MCC = 0.54). While the datasets are different, between Table 2.4 and Table 2.5 results, the difference in the individual components of the EFC algorithm compared to the combined algorithm shows a dramatic improvement when integrating order-sensitive and length independent sequence representations. The CLN-MLEM2 method integrates these two types of representations at its most basic level of output, the rule.

A decrease in selectivity of the classification will cause longer computer search times, while a decrease in specificity will increase the number of necessary experimental activity assays. Since the cost of experimentally testing peptides is much greater than the computational time of searching for antimicrobial peptides, methods that have high specificity are preferred. In addition to the high specificity of the CLN-MLEM2 method, the method creates categories of antimicrobial peptides. Categorization of peptides aids in the selection and in the design of antimicrobial peptides by providing similarity groupings according to physicochemical property boundaries. Peptides that match multiple active categories can combine more physicochemical property values associated with activity.

2.6 Conclusion

The increase in multidrug resistant bacteria usage has prompted an intense search for agents that can be used to treat infectious diseases. There is a growing interest in antimicrobial peptides as novel agents to treat infections, and this interest has led to an exponential growth of known antimicrobial peptides. However, peptide selection is a challenge with the drastic increase in the number of these peptides discovered from natural resources, their modified versions, and computationally designed peptides.

Therefore, the CLN-MLEM2 method is developed for generating rule sets to describe the similarity among antimicrobial peptides by physicochemical boundaries. This method allows the user to limit the number of physicochemical properties used to set the boundaries. Discovering where the boundaries of physicochemical properties are among active peptides generates new categories of antimicrobial peptides.

This approach simultaneously groups peptides and classifies them. The rule set performance of CLN-MLEM2 was benchmarked against other classification methods. Some available classification methods are either order insensitive or length-dependent. The rule sets that the developed method generates combine order-sensitive descriptors with length-independent descriptors. The CLN-MLEM2 method achieves comparable or improved specificity and selectivity to currently available methods with lower false discovery rates. The high specificity of this method aids novel antibacterial peptide discovery because a low false discovery rate reduces the number of bacterial assays.

The CLN-MLEM2 algorithm (Figure 2.1) developed in this chapter provides non-linear boundaries of the differences between antibacterial and non-antibacterial peptides. In the next chapter, these boundaries are used to identify antibacterial peptides customized for ease of peptide synthesis through a novel genetic algorithm (Figure 3.3).

3.0 Peptide Design with Rough Set Theory and a Codon-Based Genetic Algorithm

3.1 Introduction

Over the last two decades, computer-aided molecular design of antimicrobial peptides has mainly combined a neural network to learn activity relationships with a genetic algorithm to discover new antimicrobial peptides. In this work, a rule-based method was developed, for the first time based on rough set theory, for transparent activity relationships. This rough set theory method was combined with a codon-based genetic algorithm to design novel antimicrobial peptides which are less than 16 amino acids, lessen aggregation and have no cysteine residues. The first step, the forward step, to discovering novel antimicrobial peptide sequences is to determine the key physicochemical properties a peptide needs to be active. The second step, the reverse step, is to search novel peptide sequences to customize sequences to meet the design targets and the key physicochemical property values. The codon-based genetic algorithm method developed takes advantage of reading frame shifts to increase the diversity of novel antimicrobial peptides produced. The codon-based genetic algorithm optimized multiple design targets simultaneously. Selected novel antimicrobial peptides were experimentally tested for zones of inhibition against *S. epidermidis*, a key pathogen in implant infections, and verify that the method in this work is a successful procedure for developing novel antimicrobial peptides.

3.2 Background

As discussed in Section 1.2, antimicrobial peptide research was revolutionized by Merrifield with the development of solid phase peptide synthesis (SPPS)⁶⁹ and some of the first investigations of antimicrobial peptides from the *Cecropia* moth.⁷⁰⁻⁷² That work pioneered rational design concepts for antimicrobial peptides, beginning with testing N-terminal and C-terminal analogues²⁰² and enantiomer variations.²⁰³⁻²⁰⁶ The modes of action of the peptides were studied with lysosomes as model bacterial membranes.²⁰⁵ Shortened antimicrobial peptides were shown to possess activity and adding the shortened sequences together into hybrid sequences was also shown to have antimicrobial activity.²⁰⁷⁻
²⁰⁹ A recent study by Cantisani *et al* has used rational design concepts with alanine scanning of myxinidin

to find new antimicrobial peptides with arginine and tryptophan substitutions for improved activity for different kinds of bacteria.⁷³ As discussed before, the rational design of antimicrobial peptides is a valuable tool for developing new antimicrobial peptides that are closely related by sequence to existing antimicrobial peptides, but the approach limits how broadly peptide sequence space can be explored.

Concurrent with the development of a generalized approach to synthesize peptides (SPPS), quantitative structure-activity relationship (QSAR) models were pioneered by Hansch and Fujita for the toxicity of some drugs.⁷⁴ For an excellent review of QSAR, the reader is referred to a recent review.⁷⁵ Computer-aided molecular design (CAMD), as shown in Figure 3.1, relies on a quantitative activity-structure relationship (QSAR) model to solve the forward problem of identifying key descriptors that relate molecular structure to activity.²¹⁰⁻²¹² CAMD also introduces the reverse problem of using these relationships to design novel molecules to meet specific design targets through search.^{78, 150, 151} Rapid peptide development approaches can combine QSAR and search approaches in a CAMD methodology with SPPS to iterate between design and testing quickly.^{78, 150, 151} The development of solutions to the forward problem CAMD in the past two decades for antimicrobial peptides has been neural network models. In 1998, Patel *et al* published a single-objective genetic algorithm search with a neural network model to describe antibacterial activity.²¹³ In 2009, Cherkasov *et al* added peptide array methods for an increase of the data for the neural network to train.¹⁵⁷ Fjell *et al* provided a follow up on this study in 2011 with improved cheminformatics descriptors and reported an outstanding 94% true positive rate when synthesizing the top-fifty predicted antimicrobial peptides. In 2016, Czyzewski *et al* extended this design approach for antimicrobial peptoid molecules,²¹⁴ and Wang *et al* demonstrated this type of CAMD model for antimicrobial peptides with unnatural amino acids.²¹⁵ A limitation of using neural networks to determine antibacterial activity is the lack of transparency in terms of how the decisions are made. This lack of transparency limits the scientific knowledge gained from these models and prevents using additive descriptions of activity to design new peptides. Fernandes *et al* developed a hybrid neural

network-fuzzy inference model.¹⁷¹ The hybrid model includes capabilities of heuristic data dimensionality reduction and outlier removal. However, this hybridization does not address the lack of transparency of using neural network models.

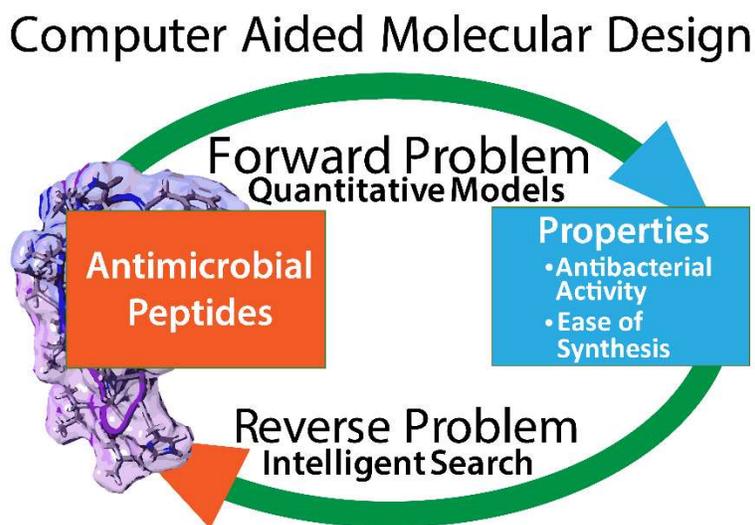


Figure 3.1 Computer Aided Molecular Design.

An alternative approach to neural networks for describing antimicrobial peptide activity is a rule-induction based approach. In 2013, Lira *et al* used a decision tree model approach for the forward problem of CAMD antimicrobial peptides. The decision tree approach addresses the lack of transparency of neural networks. However, a decision tree does not allow for a method of finding additive descriptions because the branch points of the tree are mutually exclusive. No peptide can be result of multiple traversals of the decisions tree. In this work, rough set theory is used for the first time for the forward problem in CAMD of antimicrobial peptides. The rough set theory method inducts rules based on the relevance of classification labels to feature descriptions. Membership of the peptides relating to the rules is not mutually exclusive. A peptide may meet multiple rules for activity or multiple rules for inactivity.

Feature selection is a critical part of building activity relationships that perform well beyond the training datasets. Building activity relationships has three different levels of feature selection integration: filter approaches, wrapper approaches and embedded approaches. Filter approaches score the relevance of the descriptors to the property or activity of interest without including the performance of the activity relationship. Using an analysis of variance between feature values and activity values to rank the relevance of features would be an example. However, the interaction between features for a specific model is ignored. Wrapped approaches offer ranking features with a specific activity relationship approach to use. Sequential forward selection,²¹⁶ backward elimination^{217, 218} and genetic algorithms^{219, 220} have been used as accessory algorithms to select features for other algorithms. With embedded approaches, the feature selection is completely integrated into the activity-relationship discovery algorithm. Decision trees^{221, 222} and random forests^{223, 224} are popular examples. Artificial neural networks (ANN) also have embedded feature selection, but they do not provide clear boundaries for making decisions like decision trees.

Rough set theory (RST) has received little interest in CAMD as part of solving the forward problem, even though it offers embedded feature selection and clear decision boundaries. A recent study has used RST for QSAR feature selection with small-molecule drug models and noted the improved performance to wrapper feature selection methods.²²⁵ For the first time, RST has been incorporated into a computer-aided molecular design approach as the solution to the forward problem for designing peptides with targeted properties.

For the first time in a CAMD approach, a rough set theory method is combined with a genetic algorithm search to tailor antimicrobial peptides for targeted properties. A genetic algorithm is a stochastic optimization technique using a fitness function to determine the likelihood of solution components in

newly generated solutions. Genetic algorithms are popular for search problems due to their highly modular nature and lack of dependence on gradient information.^{89, 226, 227} A variety of recombination operators have been developed in building genetic algorithms. Holland identified selected in this work are mutation and crossover.

For the first time in a genetic algorithm approach to designing peptides, a codon-basis will be used to increase the variation of peptide sequences generated for this intelligent search. The codon-based genetic algorithm (CB-GA) search completes the reverse problem of CAMD (Figure 3.1). The novel CAMD approach developed in this work was demonstrated as functional by designing antimicrobial peptides which showed activity against targeted against *S. epidermidis* and improved the ease of solid-state peptide synthesis for shorter sequences.

3.3 Methods

3.3.1 Rule Induction by Rough Set Theory

Rough set theory¹²⁷ is a heuristic method for describing the relationships between properties and classification labels. Given sequence properties calculated from amino acid chemical properties in the AAindex1,⁸⁰ rough set theory can describe which values of chemical properties separate antibacterial peptide sequences from non-antibacterial. The theory describes intervals of known values as conditions. Rules are the intersection of the conditions such that the remaining group of sequences is not vague: entirely antibacterial or non-antibacterial. Multiple rules are generated until all the sequences are described by at least one rule. The CLN-MLEM2 (Condition-Limited Number Modified Learning from Experience Module 2) is a method based on the rough set theory approach. Figure 3.2 shows the work flow of the rule induction process. CLN-MLEM2 builds categories with rules for antimicrobial peptides (Steps 2-3) by antibacterial level through chemical property features built from conditions (Step 1). These categories can be updated through incorporation of experimental results for more robust categories (Step 4).

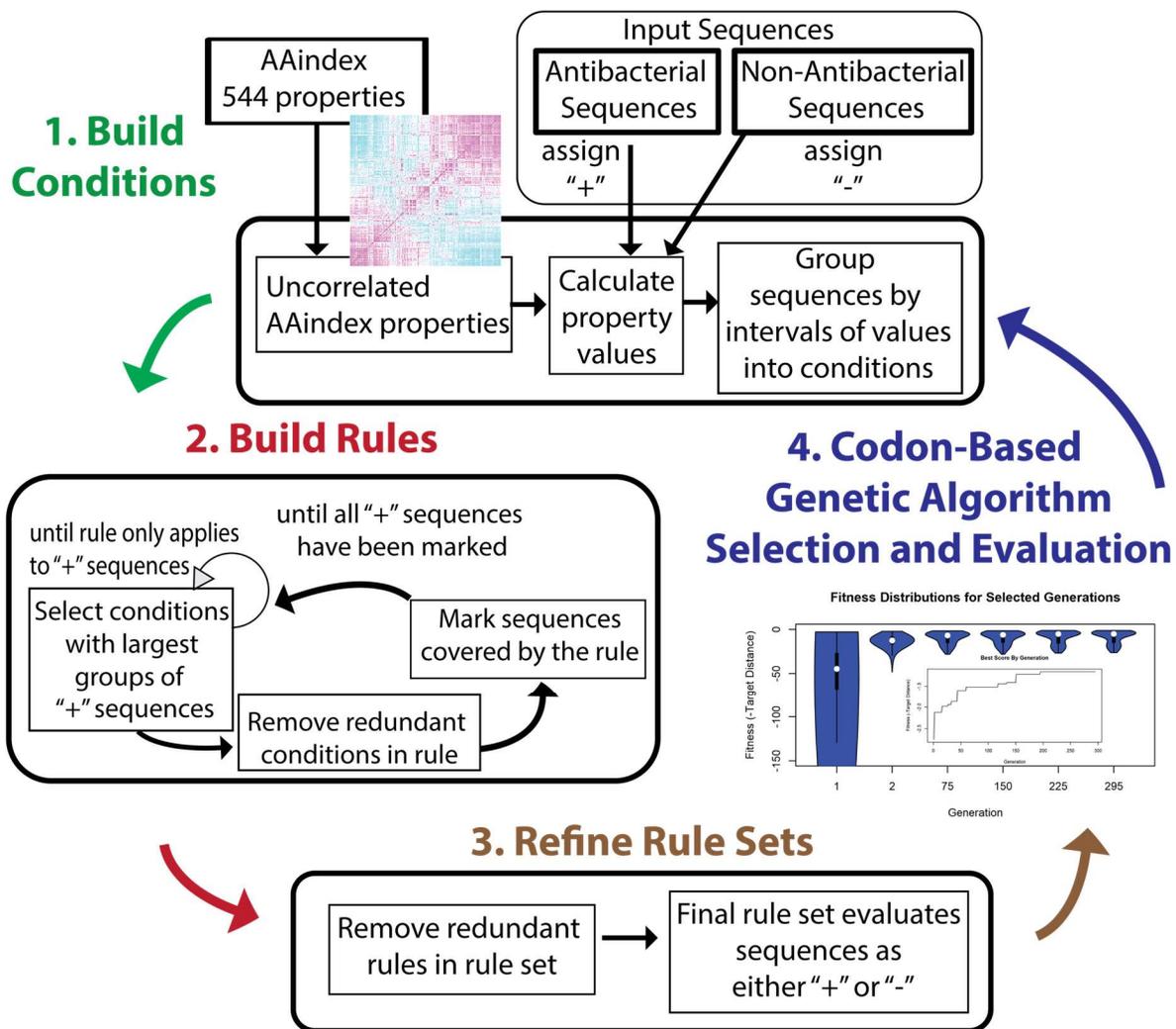


Figure 3.2 CLN-MLEM2 Algorithm including codon-based genetic algorithm selection and evaluation.

The rules discovered through rough set theory are used to classify the newly generated sequences for each new generation in the genetic algorithm. Once the peptide sequences from the genetic algorithm are experimentally characterized, new rule sets are made by adding the experimental results to the input peptide sequences.

3.3.2 Genetic Algorithm for Finding New Peptide Sequences

The genetic algorithm in this work begins by ranking known antimicrobial peptide according to a given set of design targets as seen in Figure 3.3. The initial step begins with a set of antimicrobial peptides.

The next step is to rank the peptides according to the design targets. The top 25% of scoring candidates

are selected to mutate and crossover by a DNA codon representation to generate novel peptide sequences. While removing the bottom 75% reduces the genetic diversity of future generations, it improves the convergence of the solutions to find new sequences with less computation time. The diversity lost with the filtering of the top candidates is replaced by recombination operators.

To minimize the computational time to find new antimicrobial peptide solutions, once the recombination operations are finished, the novel sequences are filtered by the antimicrobial peptide classifier. Since the best sequences are copied to the next generation, the highest scoring sequence across generations is in the final generation.

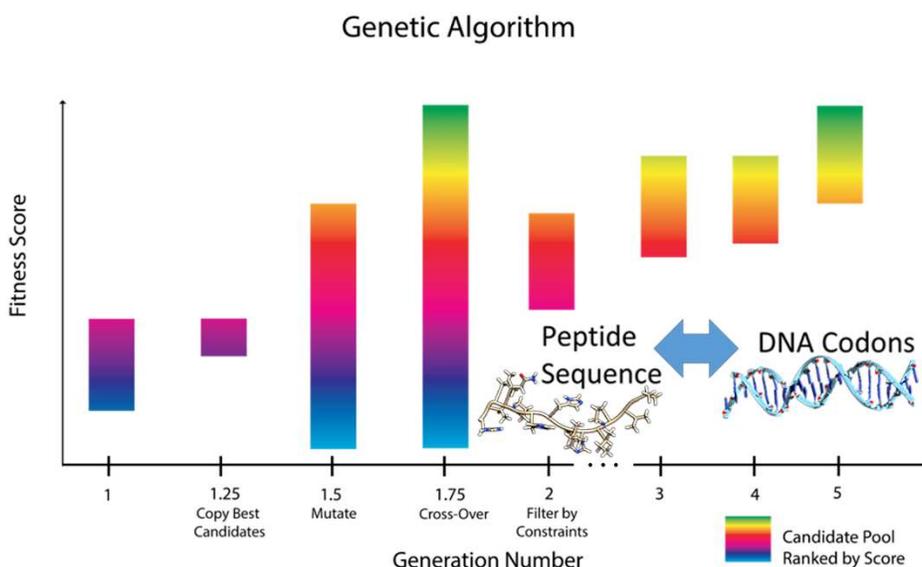


Figure 3.3 Codon-Based Genetic Algorithm (CB-GA).

The genetic algorithm implementation in this work gains flexibility in the moves it considers by using a codon-representation of peptides to direct the selection of sequences related in DNA-space (Figure 3.4). The process of peptide sequence conversion to DNA codons is the reverse of the information flow which occur in transcription and translation processes in biology.²²⁸ The information flow of the processes of transcription and translation of mapping nucleic acid sequences to amino acids are applied to the DNA

codon representation to recover the peptide sequence following the mutation and crossover events. Small moves in the DNA-space might be large moves in the protein sequence space due to reading frame shifts, as discussed in Section 1.2. To direct the genetic algorithm toward feasible answers, the most successful sequences are copied between generations. Making new candidates instead of modifying current candidate sequences builds in a historical property such that the best old sequences are propagated to future generations if they are competitive with the newly generated sequences.

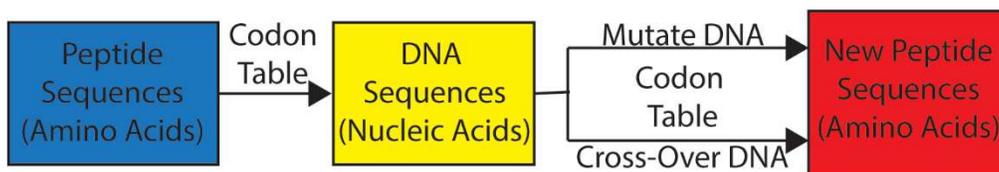


Figure 3.4 Schematic of Codon Basis for CB-GA Algorithm

3.4 Results

3.4.1 Novel Antimicrobial Peptide Design

The reverse problem of CAMD antimicrobial peptide design is solved with the CB-GA method developed in this work (Figure 3.3). For the first example, the first generation is selected from known antimicrobial peptides in the APD3¹³⁶ database and sorted by score using the design targets in Table 3.1. These targets are selected to improve the ease of synthesis with the Fmoc-based SPPS method. The scoring function for fitness is the negative weighted average of the distance from the targets. For ease of synthesis, short amino acid sequences are targeted. Cysteine residues were avoided in the targets to prevent inter-peptide bonding through disulfide bonds. Another target was no aggregation hotspots predicted through the Aggrescan method.²²⁹ The final target was to produce cationic peptide sequences for targeting bacterial membranes.^{64, 156, 230-232} The two classes of bacteria differentiated by Gram's stain both have negative charges displayed on their surfaces.²³³ Gram negative bacteria, like *E. coli*, have negatively charged lipopolysaccharides (LPS). Gram-positive bacteria like *S. epidermidis* and *S. mutans* have a peptidoglycan wall with acidic polymers, call teichoic acids radiating outside of the wall.

Table 3.1 Design targets for novel antimicrobial peptides for ease of Fmoc synthesis.

Property	Target
Amino Acid Length	7 to 15
Cysteine Count	0
Aggrescan Score	0
Net Positive Charge	+1 to +5

The newly generated sequences are filtered by the high specificity rough set theory classification method CLN-MLEM2 so that each completed generation only consists of either known antimicrobial peptides from the initial generation or predicted antimicrobial peptides. A violin plot²³⁴ of the fitness scores, where the center is a boxplot and the shape of the sides is formed from the probability density kernel, across selected generations is shown in Figure 3.5. Advancing generations become more skewed in favor of higher fitness scores.

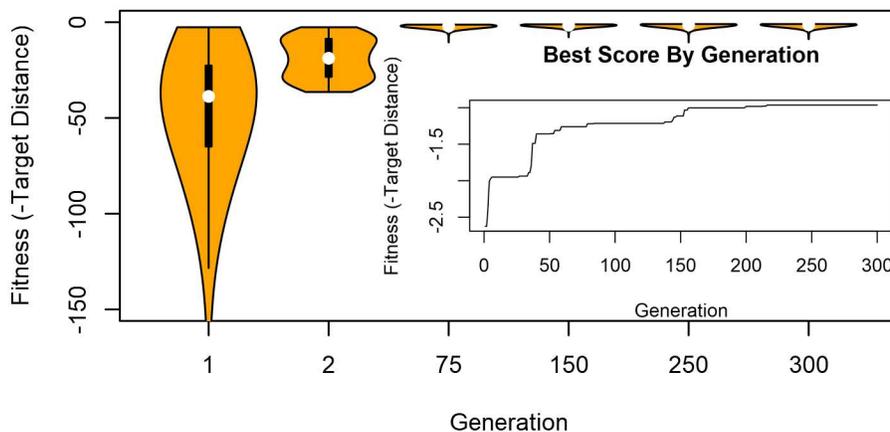


Figure 3.5 Fitness Distributions of Selected Generations.

Translation is the cellular process of nucleic acids becoming polypeptide sequences. To generate novel antimicrobial peptide sequences, peptide sequences were reverse-translated to a codon-representation to take advantage of reading frames for generating novel sequences. Mutation, as a local search

method, was the changing of a single codon through substitution, insertion or deletion. The large search changes in sequence space were achieved through crossover by slicing and recombining nucleic acid sequences reverse-translated from the peptide sequence candidates. Figure 3.6 shows that using the codon-representation increased the variance of the fitness scores generated while reaching similar maximum and mean fitness levels. The blue lines represent data for the codon-representation and the orange lines represent data without the codon-representation. The solid lines represent the average of 6 repeated genetic runs over 100 generations. The dotted lines represent the 95% CI using the student t test statistic of the repeated runs. In Figure 3.6a, the standard deviation of fitness scores is increased with codon representation. The maximum fitness converges independent of codon representation, in Figure 3.6b, and in Figure 3.6c the number of predicted antibacterial sequences decreases when using the codon representation. The codon-representation mean fitness converges in Figure 3.6d with the non-codon-representation mean fitness. Increasing the variance with the codon-representation resulted in a reduced number of predicted antibacterial peptides. To the author's knowledge, this is the first time a genetic algorithm to design peptides^{89, 156, 235-242} has used reading frame shifts for generating novel sequences.

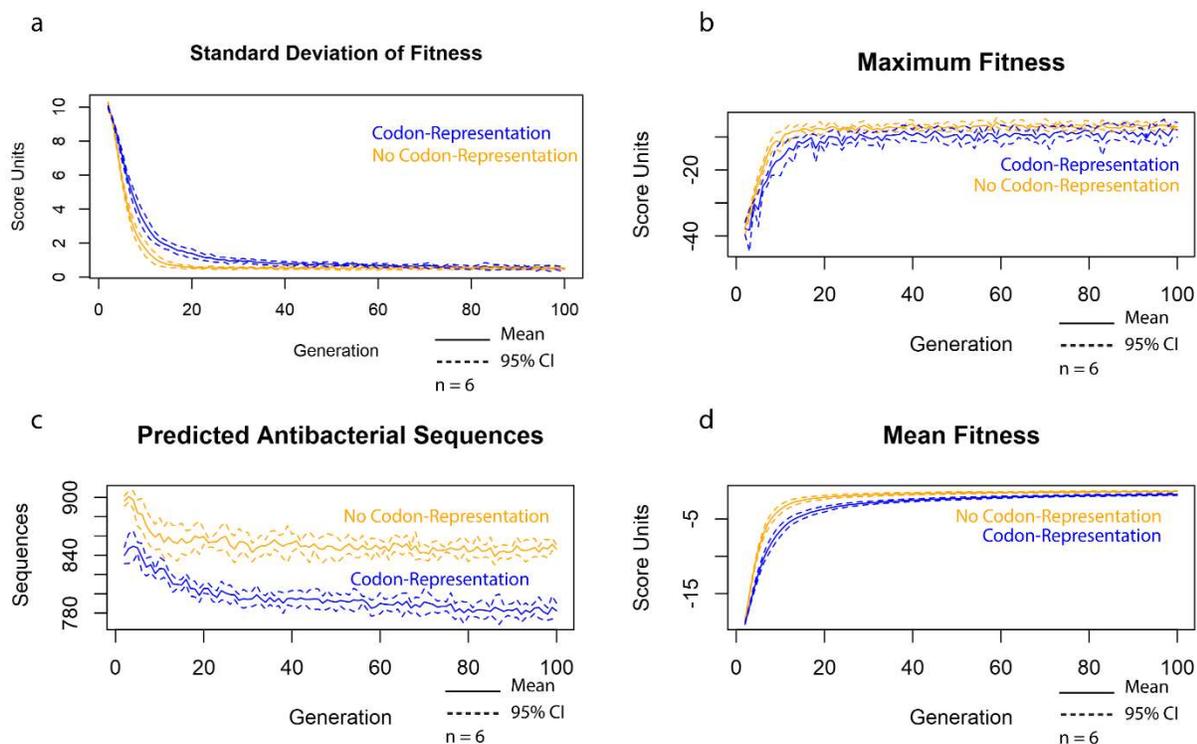


Figure 3.6 Improved sequence variation through codon-representation.

3.4.2 Novel Antimicrobial Peptide Experimental Test

S. epidermidis, a common pathogen for surgical site infections, was used by my colleague Cate Wisdom to evaluate antibacterial activity of the designed peptides on agar plates. The crude peptide synthesized by Cate Wisdom was utilized as screening.²⁴³⁻²⁴⁶ The screening test uses the diffusion of the peptides on agar plates to evaluate if the peptides can cause of zone of inhibition where the bacteria will not grow. Therefore, a large initial concentration of the peptides can be used to evaluate the range of peptide concentrations generated through diffusion out into the agar. Ampicillin was used as a positive control antibacterial agent.

Two of the three tested peptides for the top scoring peptides of the final generation of the genetic algorithm developed in this work show zones of inhibition against *S. epidermidis*, as shown in Table 3.2.

All three of the candidate peptides were within the boundaries described by a CLN-MLEM2 rule

category. AMP-1, AMP-2 and AMP-3 agree on residues for five of the seven residue positions. AMP-1 demonstrates that small sequence changes from AMP-2 can result in a loss of *S. epidermidis* activity, while AMP-3 demonstrates that a small sequence change from AMP-2 does not result in a loss of *S. epidermidis* activity. Applying the CLN-MLEM2 algorithm to the updated data set labeling AMP-1 as inactive and AMP-2 and AMP-3 will update the previous rule set which classified AMP-1 as active to be less likely to identify false positive peptides.

Table 3.2 Inhibition zone of *S. epidermidis* for candidate novel antimicrobial peptides.

Agent	Sequence	Concentration (mM)	Inhibition Zone (cm)
Ampicillin	n/a	0.028	1.6
ADP3 peptide 1 (crude)	DYHHGVRVL	0.377	0.0
ADP3 peptide 2 (crude)	GIHDILKYGKPS	0.407	1.0
AMP-1 (crude)	ESYKKML	0.468	0
AMP-2 (crude)	ESYKRMF	0.431	0.9
AMP-3 (crude)	ESYKHMF	0.439	1.1

Basic Local Alignment Search Tool (BLAST) is a fundamental tool in determining the homology of a sequence with other known sequences in databases.²⁴⁷ BLAST uses word search within sequences to find likely matches while focusing on words with high sequence complexity, that is a low frequency of character repeats. A tblastn search,²⁴⁸ a homology search of the designed peptide sequences against known peptides or peptide sequences translated from known nucleic acid sequences in the NCBI database, was performed. No significant matches were returned. These peptides do not have homology to previously studied peptides or peptide sequences translated from sequenced nucleic acids in the NCBI database.

3.4.3 Combining MLEM2 Rule Categories

To investigate if meeting multiple CLN-MLEM2 rule categories increases the likelihood of activity for peptides, the MLEM2 rule category count was added as a design target (Table 3.3).

Table 3.3 Design targets for novel antimicrobial peptides for combining MLEM2 rule categories.

Property	Target
Amino Acid Length	7 to 15
Cysteine Count	0
Aggrescan Score	0
Net Positive Charge	+1 to +5
Matching MLEM2 Rule Count	8-12

Again, the genetic algorithm developed in this work found peptides that were closer to the design targets than any of the antimicrobial peptides in the APD3 database. (Figure 3.7).

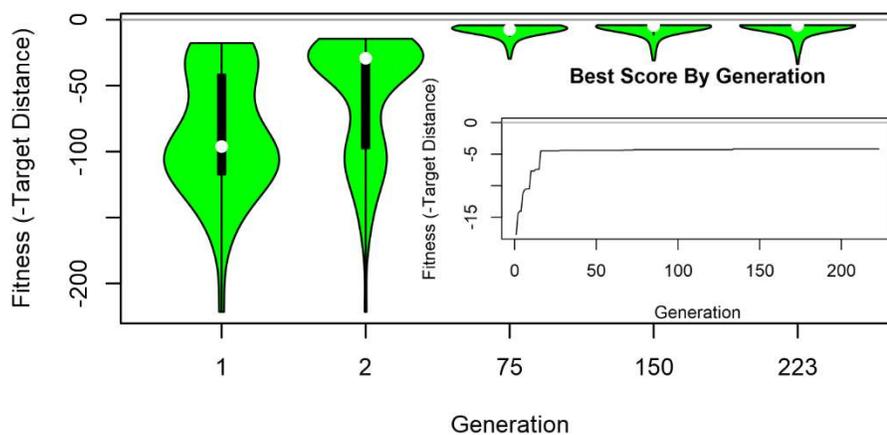


Figure 3.7 Fitness Distributions of Selected Generations.

The antibacterial screening with *S. epidermidis* showed that one of the three antimicrobial peptides showed antibacterial activity with a slightly larger inhibition zone than the peptides chosen without the multiple MLEM2 rule category design target, as seen in Table 3.4. Hp1404,²⁴⁹ was the highest-ranking peptide from the APD3 database according to the design targets in Table 3.3. Hp1404 has known activity against gram positive bacteria. The middle of the HP1404 sequence was conserved among novel peptides for this genetic algorithm search. Conserved residues are underlined in Table 3.4. The target of meeting multiple MLEM2 rules produced search results with more conserved residues than was shown in Table 3.2.

A tblastn search of the designed peptide sequences in Table 3.4 against the NCBI translated nucleotide database returned no matches. Therefore, the designed peptides were not found to have high-homology from translated DNA sequences. Sequences that do have high homology from the translated nucleotide database with no previous annotation of antimicrobial activity are referred to as encrypted antimicrobial peptides.²⁵⁰ As seen with examples from Table 3.2, close sequence similarity relationships between AMP-4 and AMP-5 resulted in different antibacterial activity against *S. epidermidis*. New MLEM2 rule categories can be generated to separate sequences with these differences.

Table 3.4 Inhibition zone of *S. epidermidis* as a screen test of antibacterial activity for combining MLEM2 rule categories.

Agent	Sequence	Concentration (mM)	Inhibition Zone (cm)
Ampicillin	n/a	0.028	1.6
Hp1404 (crude)	<u>G</u> <u>I</u> <u>L</u> <u>G</u> <u>K</u> <u>L</u> <u>W</u> <u>E</u> <u>G</u> <u>V</u> <u>K</u> <u>S</u> <u>T</u> <u>F</u>	2.61	2.3
AMP-4 (crude)	AT <u>L</u> <u>G</u> <u>V</u> <u>L</u> <u>W</u> <u>E</u> <u>S</u> <u>I</u> <u>R</u> <u>G</u> <u>H</u> <u>R</u>	2.51	0
AMP-5 (crude)	AT <u>L</u> <u>G</u> <u>V</u> <u>L</u> <u>W</u> <u>E</u> <u>G</u> <u>A</u> <u>R</u> <u>G</u> <u>H</u> <u>T</u>	2.73	1.2
AMP-6 (crude)	<u>G</u> <u>T</u> <u>L</u> <u>A</u> <u>N</u> <u>G</u> <u>W</u> <u>E</u> <u>G</u> <u>V</u> <u>R</u> <u>T</u> <u>N</u> <u>H</u>	2.65	0

3.5 Discussion

3.5.1 Increasing Variation through Codon Representation

Genetic algorithms for solving the reverse problem of CAMD have been used for a variety of problems.^{251, 252} Protein or peptide design is a natural application of genetic algorithms since the process of genetic evolution designs new biological proteins. Computational protein and peptide design have been accomplished through genetic algorithms⁹¹⁻⁹⁷ before but not all aspects of the genetic system that inspired genetic algorithms have been explored in protein and peptide design. A DNA codon was applied to the representation of peptides within the genetic algorithm in this work to take advantage of reading frame shifts. Generating novel solutions in a genetic algorithm is a balance between viability, finding solutions that meet some criteria, and adaptability, finding solutions that meet all criteria. Increasing the viability of each generation often involves using small moves in sequence space to avoid the loss of viability of large, random moves. Increasing the adaptability relies on the ability to make bigger moves in sequence space to preserve genetic diversity among generations. Reading frame shifts in biology are large moves which balance viability and adaptability as shown by how biological proteins develop in nature. Single-codon mutations in DNA, either deletions or insertions, cause reading frame shifts. Reading frame shifts encode transition probabilities for which new amino acids replace the previous amino acids. While most-reading frame shifts are nonviable, the viable frame shifts in nature lead to the development of new proteins. While using a codon-representation reduces the viability of this method by generating less antibacterial sequences (Figure 3.6), the genetic diversity is increased because of the increase of the standard deviation of scores. Since the method developed in this work filters out non-antibacterial sequences, this increased genetic diversity is among predicted antimicrobial peptides.

3.5.2 Combining Antibacterial Classes

Each CLN-MLEM2 rule for antibacterial activity describes a set of physicochemical properties that separates a set of antibacterial peptides from all given non-antibacterial peptides in the training set.

Some peptides may meet more than one rule for antibacterial activity. These peptides may act in multiple ways to achieve antibacterial activity. Measuring the number of CLN-MLEM2 rules a peptide meets is a measure of its robustness for having broad spectrum antibacterial activity.

3.5.3 Targeting Specific Bacterial Strains

Multiple peptides were developed with antibacterial activity against *S. epidermidis* without having labels for which peptides were effective against this strain. Several peptides were also found which have no activity against *S. epidermidis*. These negative activity examples are a critical resource to developing antibacterial peptides which can preserve or restore balance to microbiomes. A microbiome is the complement of bacteria that is vital for the health and wellness of an ecosystem.²⁵³ The method developed in this work can identify which physicochemical property rules relate to activity against some strains over other strains. Just as combining CLN-MLEM2 rules for activity may lead to a broader spectrum of activity, combining CLN-MLEM2 rules for inactivity can lead to more targeted antibacterial peptides.

3.6 Conclusions

To avoid a post-antibiotic era in which bacterial infections become untreatable, a computer-aided molecular design (CAMD) method was developed in this work to design antibacterial peptides which can be targeted for desired properties and strain specificity. This method found novel antibacterial peptides of seven amino acids that are easier to synthesize than antimicrobial peptides in the APD3 database. An improvement in antibacterial activity was found when adding together multiple rules for activity from the rough set theory method from the previous chapter. For the forward problem of quantifying sequence-activity relationships, a rough set theory method (CLN-MLEM2) as a quantitative structure-activity relationship (QSAR) model for peptides. For the first time, RST has been applied to designing peptides or proteins. For the reverse problem of finding novel peptide sequences, the codon-based

genetic algorithm developed in this work was applied to discover novel antibacterial sequences against *S. epidermidis*, a key pathogen for implant infections.

Genetic algorithms favor more fit solutions over less fit solutions in reproduction. The reduced diversity of the reproduced solutions may lead to only searching suboptimal parts of the search space. To replace diversity that is lost when killing off solutions between generations, recombination operators used in this work are mutation and crossover. Other recombination operators from genetics have been studied for genetic algorithms such as gene dominance and permutation crossover.²⁵⁴⁻²⁵⁷ These operators attempt to use pattern recognition within generated solutions to recognize blocks of solutions which can be recombined for improved solutions. The current work prioritizes the computational efficiency of finding new antimicrobial peptides over finding the most optimized solutions, but future work may use recombination operators which exploit pattern recognition schemes.

In Aim 1, new antimicrobial peptides were customized according to ease of synthesis design criteria. Aim 2 of this work was to create chimeric antimicrobial peptides which can create antibacterial activity interfaces. In Chapter 4, titanium-binding antimicrobial peptides are demonstrated and engineered for improved antibacterial activity.

4.0 Chimeric Spacer Design for Titanium Implant Surfaces

4.1 Introduction

The integration of titanium implants in the host tissue depends on the creation and maintenance of an infection-free interface.^{19, 258} The conventional approach to providing an infection-free environment is the systemic delivery of antibiotics.^{124, 130} This approach results in sub-inhibitory antibiotic concentrations throughout the body as the antibiotic is metabolized, leading to further antibiotic drug resistance for the bacteria associated with the body. This resistance will lead to reduced efficacy of the antibiotic if the bacteria becomes pathogenic for the patient or also for other patients if the resistance is genetically transferred.

Active research in the literature is ongoing about how to deliver antibacterial agents to the titanium implant surface to overcome the challenge of spreading bacterial drug resistance.²⁵⁸⁻²⁶¹ The two main approaches are covalent-chemistry attachment^{26, 262} and non-specific physical adsorption²⁶³⁻²⁶⁶ of antibacterial agents. Both approaches require modification to the titanium surface, preventing the use of currently developed implants. Also, none of the current studies for either of these approaches have provided orientation control for the applied antibacterial agents.

Chimeric peptides are an innovative solution that can provide orientation control while using currently developed titanium implants.^{5, 6} In this work, using current implants is achieved through solid binding peptide technology. Combinatorial biology protocols run by my colleague Deniz Yucesoy discovered solid binding peptide sequences that conforms to titanium surfaces through non-covalent interactions. Orientation control was achieved through engineering spacer sequences to improve the functionality of the antimicrobial peptide domain of the chimeric peptide through peptide folding analysis.

4.2 Background

The reliability of the interface between proteins and material surfaces is often poor due to non-specific adsorption and orientation.^{267, 268} The proteins that adhere to a material surface often have randomized orientation. One approach is to develop surfaces which encourage proper protein orientation. Wang *et*

al use co-block polymers to improve the interactions between surfaces and proteins at the nanoscale.²⁶⁹⁻
²⁷² The surfaces are intended to guide understanding of how protein resistance, that is the lack of protein adsorption, can be designed into material surfaces for different kinds of proteins. These design rules, once understood, could also lead to the design of nanoscale surfaces which could enhance protein binding. However, these surfaces would need to be transferred to currently available devices to have the broadest impact. Another approach to loss of protein functionality from randomized orientation is to increase the number of immobilized proteins by adsorbing multiple layers. In certain applications, such as keratins, multiple layering is advantageous.²⁷³ However, multiple layering is not advantageous for many protein classes such as enzymes. Such multiple layering methods lead to reduced specific activity of the enzymes on the surface due to a variety of factors. Another example of reduced specific activity is antimicrobial peptides. A recent study has used calcium phosphate coatings and titanium nanotubes to deliver a dosage of cationic antimicrobial peptides at the implant surface.²⁷⁴ While the release profile may help prevent bacteria near the implant site, the delivery system does not facilitate the proper display of the antimicrobial peptides so that they reach maximum specific activity on the surface.

Methods are in development to improve the specific activities of peptides and proteins on the surface of materials.²⁷⁵⁻²⁷⁹ The activity of proteins and peptides on the surface of materials depends both on their secure attachment to the surface and on preserving their orientation for their function.²⁸⁰⁻²⁸³ Using non-specific adsorption onto the surface can lead to problems with insecure attachment and degraded orientation. Physical release systems can address the insecure attachment problem but cannot improve the orientation degradation without any additional mechanisms.²⁸⁴⁻²⁸⁹ Self-assembled monolayers and other irreversible bonding techniques²⁹⁰⁻²⁹² provide secure attachment, but the orientation degradation challenge remains. Solid binding peptides are a promising solution to overcome both secure attachment and orientation degradation problems as a next-generation physical adsorption technique.^{5, 6, 33, 48, 275, 293-}

²⁹⁵ For this work, two studies have been published demonstrating chimeric antimicrobial peptides functioning on titanium surfaces.^{6, 293}

4.3 Methods

4.3.1 PyRosetta Ensemble

To build an ensemble, a fragment-insertion based method was used to fold the backbone of the structural decoy using Robetta server fragments.²⁹⁶ The fragments are inserted into the backbone if they improve the energy score of the decoy. If not, the fragments may be inserted if they are close to improving the energy score. This barrier is gradually reduced as an annealing method. Once the backbone structure is generated, the Fast Relax protocol is used to determine the lowest energy rotamers.¹¹⁹ The process is repeated up to 1,000 times to generate an ensemble for further statistical analysis. An example implementation of this procedure is included in Appendix B.

4.3.2 Secondary Structure Feature Counts

Each ensemble of computational structures was generated according to customized PyRosetta scripts such as the script included in Appendix B. Each structural decoy within a structural ensemble was analyzed by the Dictionary of Secondary Structure in Proteins (DSSP) program²⁹⁷ to identify secondary structure features within the sequence. The feature counts of every DSSP analysis for a single decoy was tabulated into a single text file in the LERS format. The feature counts were divided by the number of structural decoys generated to calculate the structural feature frequencies.

4.4 Results

4.4.1 Chimeric Titanium Antimicrobial Peptides Reduced Efficacy

The chimeric peptide antibacterial activity is dependent on the interaction of the solid-binding domain and the antimicrobial peptide domain, data is provided in Table 4.1.⁶ These activity experiments were performed by my colleague Deniz Yucesoy. For example, the *S. mutans* minimum inhibitory

concentration increased over four times for AMP1 when the titanium binding peptide TIBPS1 was added, and over nine times when titanium binding peptide TIBSP2 was added.

Table 4.1 Minimum inhibitory concentrations for chimeric antimicrobial peptides with solid binding peptides for titanium and a GGG spacer.

Peptide	<i>E. coli</i> MIC (μM)	<i>S. epidermidis</i> MIC (μM)	<i>S. mutans</i> MIC (μM)
AMP1	9.45	4.72	37.81
AMP2	21.08	0.66	10.54
TIBPS1-AMP1	9.58	4.78	153.25
TIBPS2-AMP1	21.0	5.23	336.5
TIBPS1-AMP2	80.8	2.52	80.8
TIBPS3-AMP2	167.4	5.23	83.7
TIBPS4-AMP2	167.6	5.22	83.8

The folding dynamics of the antimicrobial peptide domain may change in the chimeric peptide compared to the single domain. To learn which secondary structure features in the antimicrobial peptides are the most relevant to antibacterial activity, a rough set theory analysis of computationally predicted structures for single-domain antimicrobial peptides and chimeric antimicrobial peptides was performed with MLEM2, as diagramed in Figure 4.1.

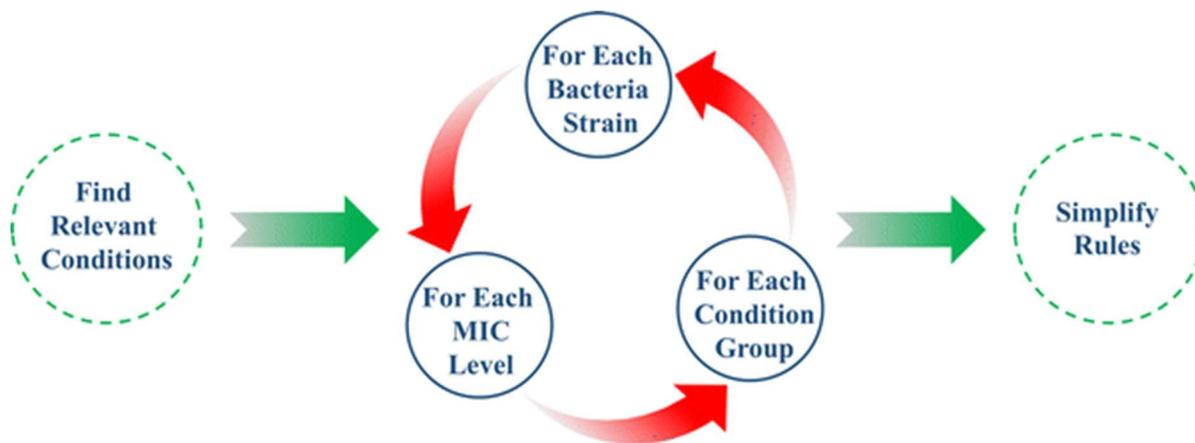


Figure 4.1. Schematic of rule induction process by MLEM2 for discovering structure-function relationships for chimeric antimicrobial peptides.

The MLEM2 analysis showed that the most relevant differences observed among computational structure predictions were helices. Table 4.2 describes the secondary structure features associated with low and high levels of antibacterial activity shown in Table 4.1. Three different pathogens were evaluated by my colleague Deniz Yucesoy for the activity against the seven peptide sequences in Table 4.1. *E. coli* was selected for its role in urinary device infections, *S. epidermidis* for its role in orthopedic implant infections and *S. mutans* for its for in dental implant infections.

Table 4.2. MLEM2 generated rules describing secondary structure and chimeric antimicrobial peptide activity.

Alpha Helix Property	Pathogen	MIC Interval (μM)	Correction Cases / Applicable Cases
4-a.a.-right-handed helix & 5-a.a.-helix	<i>E. coli</i>	9.45-21	164/179
4-a.a.-right-handed helix & 5-a.a.-helix	<i>S. epidermidis</i>	4.72-5.23	164/179
4-a.a.-right-handed helix & 5-a.a.-helix	<i>S. mutans</i>	10.54-37.81	151/179
8-a.a.-helix & 6 or 8-a.a.-right-handed helix	<i>S. mutans</i>	336.5	8/11

Figure 4.2 shows the helical formations which are most relevant to the peptide folding of the most active sequences against the pathogens listed in Table 4.2. 200 different structural decoys were generated for each of the seven peptide sequences to discover the most relevant secondary structures to antibacterial inhibition in Table 4.1.

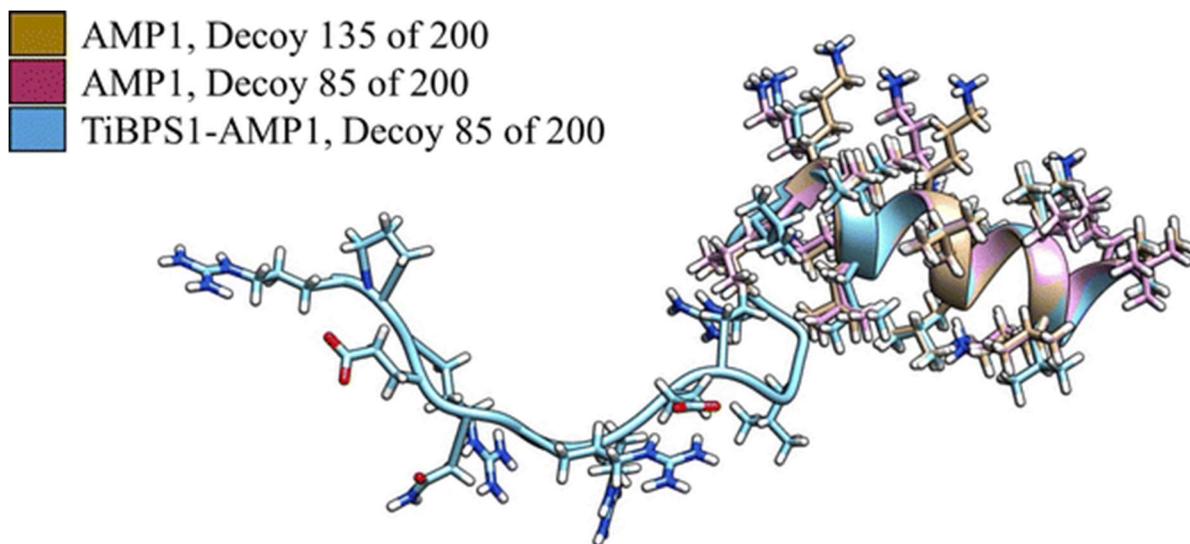


Figure 4.2. Aligned secondary structure decoys for 5-a.a.-helix rules generated by MLEM2 algorithm.

Figure 4.3 shows the structures that were observed for the peptides with higher inhibitory concentrations in Table 4.1. These structural trends shown in Figure 4.2 and 4.3 are structure-function relationships of activity for chimeric peptide sequences. The structures shown in Figure 4.3 were used to build improved chimeric antimicrobial peptide function through the design of a new spacer.

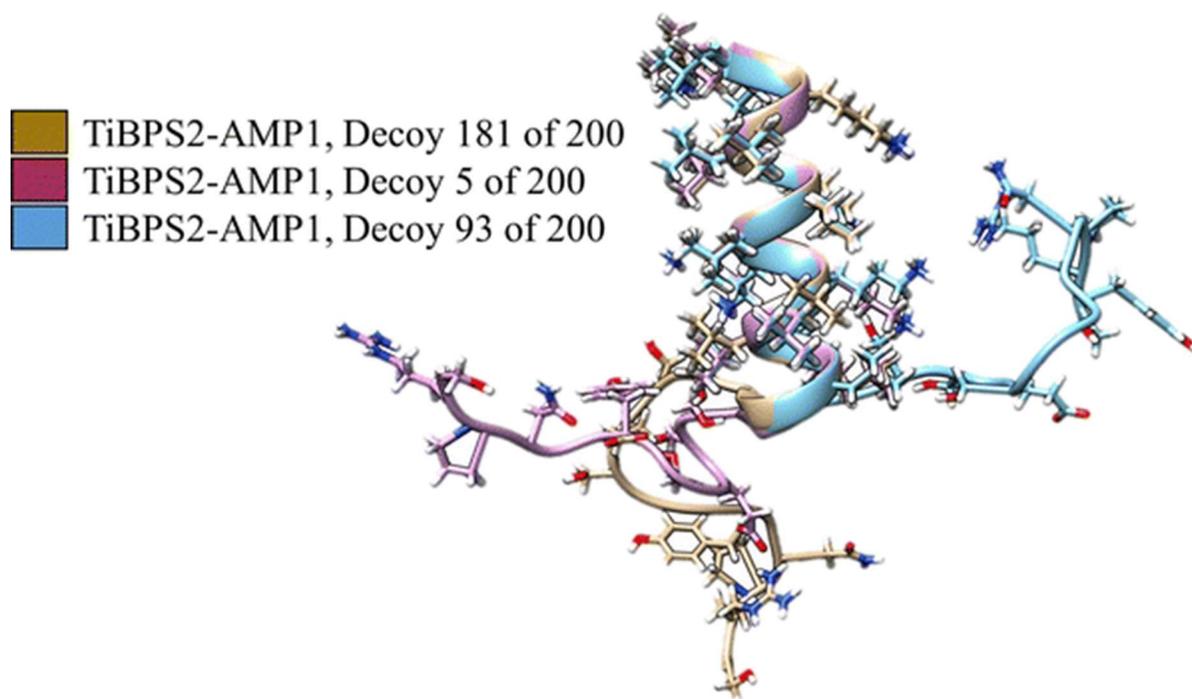


Figure 4.3. Aligned secondary structure decoys of 8-a.a.-helix rule generated by MLEM2.

4.4.2 Designing New Chimeric Peptide Spacer for Improved Antimicrobial Activity

Different chimeric peptide sequence combinations of the same antimicrobial peptide and different solid-binding peptides resulted in different computationally predicted secondary structure frequencies.

Different spacer domains were investigated to see if they would also change the computationally predicted secondary structure frequencies. The GGG linking domain used for the sequences in Table 4.1 was changed to GSGGG due to serine/glycine combinations in peptide display systems. Computer-generated structures of these sequences are in Figure 4.4.

In Figure 4.4, the lowest energy structures modeled in solution for (a) TiBP-Spacer3-AMP chimeric peptide; (b) TiBP-Spacer5-AMP chimeric peptide; (c) Spacer3 (GGG); (d) Spacer5 (GSGGG); (e) AMP; (f) TiBP. The peptide backbone is represented as a ribbon to show secondary structure for peptides with

side chains represented by full atoms. TiBP domains, spacer domain and AMP domains are designated with blue-, black- and red-shading, respectively.

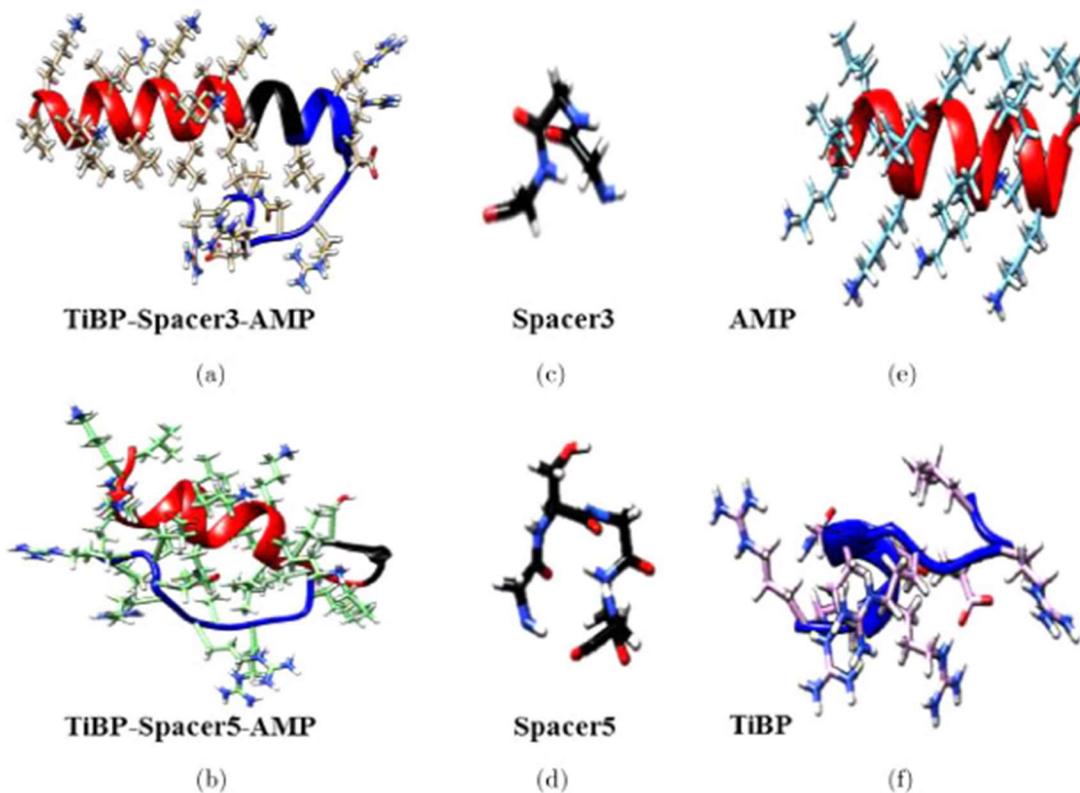


Figure 4.4. Improved spacer for antibacterial activity against *S. mutans*

The secondary structure frequencies were calculated for TiBP-Spacer3-AMP and TiBP-Spacer5-AMP, and a shift was observed for features that were the most relevant for high antimicrobial activity against *S. mutans* and *S. epidermidis* for peptide sequences in Table 4.1. TiBP-Spacer3-AMP is the same sequence as TiBPS1-AMP1 in Table 4.1. Table 4.3 gives the secondary structure frequency comparison between these two sequences. The secondary structure function relationship predicts that TiBP-Spacer5-AMP possesses a secondary structure more associated with antimicrobial activity than the secondary structure of the TiBP-Spacer3-AMP.

Table 4.3. PyRosetta method of peptide structure prediction frequencies of four and five amino acid alpha helices in computational generated structures.

Peptide	4 aa α -helix Frequency (%)	5 aa α -helix Frequency (%)
TiBP-Spacer3-AMP (RPRENRGRERGL-GGG-KLLKLLKLLKLL)	10.4	5.6
TiBP-Spacer5-AMP (RPRENRGRERGL-GSGGG-KLLKLLKLLKLL)	17.6	8.0

Along with a shift in secondary structural features seen in Table 4.3, a shift in antibacterial activity was observed in Table 4.4. The TiBP-Spacer5-AMP had a minimum inhibitory concentration much closer to the antimicrobial peptide domain alone than the MIC of the TiBP-Spacer3-AMP. This example shows that spacer design in chimeric peptides can help close the antibacterial activity gap between the activity of the single domain peptide and the chimeric peptide which adds the additional functionality of coating the titanium surface through solid binding peptide technology.

Table 4.4 MIC of TiBP-Spacer5-AMP, TiBP-Spacer3-AMP and AMP

Peptide	<i>S. mutans</i> MIC (μ M)	<i>S. epidermidis</i> MIC (μ M)
AMP	38	4
TiBP-Spacer3-AMP	153	5
TiBP-Spacer5-AMP	50	8

4.5 Discussion

Because chimeric antimicrobial peptides are polypeptide chains, they have multiple levels of structure.

This work has shown that small changes in the primary level of structure has led to detectable differences in modeled secondary structure through the customized PyRosetta script built for this work.

The secondary structure features of antimicrobial peptides such as helices has been shown to be related to their antibacterial function.²⁹⁸⁻³⁰⁰ This work relates the modeled secondary structure features to the observed antibacterial activity of antimicrobial peptides and chimeric antimicrobial peptides. For

chimeric peptides, some loss of activity was observed. A novel spacer was shown to improve the activity of the chimeric peptide.

4.6 Conclusions

A structure description tool (PyRosetta Peptide Folding Method, Appendix B) was developed for evaluating structure-function relationships and the effect of the spacer sequences of chimeric peptides on the titanium surface. Specific secondary structure features were discovered through a data mining technique (CLN-MLEM2, Figure 2.1) to find spacers which increase the frequencies of secondary structures related to antibacterial activity. The second aim of designing chimeric antimicrobial peptides for the titanium implant surface was achieved. The third aim of this work to explore multiple types of implant surfaces to apply antibacterial activity. In the next chapter, chimeric antimicrobial peptides are designed for new materials: calcium phosphate, zirconia and urethane polymers.

5.0 Chimeric Antimicrobial Peptides across Multiple Materials

5.1 Introduction

The second aim of this work is to engineer the interface between implant surfaces and the host tissue for titanium surfaces with chimeric peptides in the previous chapter. The third aim of this work is to design chimeric antimicrobial peptides for multiple materials. Calcium phosphate, zirconia and urethane polymers are provided as case studies.

Within both disciplines of material science and molecular biology, the interface between biological molecules and engineered material surfaces is among the most active research areas. Investigations in this area are applicable to a wide range of problems such as biological sensors, industrial processing of biological feedstocks, and medical device interfaces.

5.2 Background

In Chapter 4, the most common metal in biomaterial implants was selected for chimeric antimicrobial peptide development (titanium). In this chapter, common biomaterials of different types were selected for chimeric antimicrobial peptide design. A mineral (calcium phosphate), a ceramic (zirconia) and a polymer surface (acrylic urethane) have been selected for developing chimeric antimicrobial peptides.

Material Surfaces

Calcium phosphate is the main mineral component of hard tissues. The main mineral phases of calcium phosphate are amorphous calcium phosphate (ACP), octacalcium phosphate (OCP) and hydroxyapatite (HAP).³⁰¹ Like titanium, calcium phosphate is widely used as a biomaterial in orthopedics and dental applications.³⁰² Calcium phosphate is used as an implant coating because of its osteoinductivity across a variety of mineral phases.^{44, 47, 302-306} Therefore, due to the risk of surgical site infections, the protection of the calcium phosphate surface against drug-resistant bacteria must go beyond what prophylactic systemic antibiotics provide.^{5, 19, 305}

In Tamerler's group's previous work, many calcium phosphate binding peptides have been selected through combinatorial biology protocols. In 2008, a library of peptides was selected against hydroxyapatite.³⁰⁷ From this study, HABP1 was identified as a strong binding peptide can direct the formation of octacalcium phosphate from the ion stage of mineralization. In 2012, the library of over 100 hydroxyapatite binding peptides from the 2008 study was used to calculate which subsequences of amelogenin, a protein used in forming hydroxyapatite mineral phase in tooth enamel, likely bind to hydroxyapatite³⁸ through sequence similarity bioinformatics methods.³⁸ Three separate subsequences of amelogenin were identified as likely to bind to hydroxyapatite (ADP1, ADP2, and ADP4). The range of dissociation constants for amelogenin subsequences identified as likely to bind to hydroxyapatite was in the low micromolar range. The dissociation constant of HABP1 is also in the low micromolar range. The calcium phosphate solid binding peptides have comparable binding kinetics to the titanium solid binding peptides used in Chapter 4.^{6, 48}

In dentistry, titanium and zirconia are most commonly used for connection between the implant and the mandible or the maxilla. Therefore, the bony integration of the implant is critical to avoiding implant failure in challenging cases with comorbidities that reduce bone healing such as diabetes and osteoporosis. In this work, the primary approach for an antibacterial barrier is to develop chimeric antimicrobial peptides at interface. The secondary approach is to provide a mechanical barrier to bacteria and pH buffering capacity by mineralizing the implant interface with a calcium phosphate mineral phase. The solubility of the formed mineral depends on the ratios of these phases,^{308, 309} which are tunable based on the capping properties of the peptide-mediated mineral formation process.^{307, 310-}

313

Amelogenin is a protein secreted by ameloblasts in the formation of tooth enamel.^{314, 315} The distinct roles of amelogenin-subsequence peptides in the formation of hydroxyapatite in enamel was elegantly described in a 2012 study.³¹¹ With this recent study, two functions were differentiated: the nucleation of

calcium phosphate particles and the formed morphologies. This study was used as inspiration for combining fast formation kinetics with octacalcium phosphate-forming morphology. Characterization of calcium consumption rate of HBP1 in the bio-enabled mineralization assay was published by Gungormus *et al* in 2008.³⁰⁷ The calcium consumption rate of amelogenin derived peptide 1 (ADP1) was published by Gungormus *et al* in 2012.³¹¹

Zirconia, or zirconium dioxide, has three phases: monoclinic, tetragonal and cubic. Ceramic material properties are dependent on the mixture of these phases. The mixture is controlled by alloying different oxides.³¹⁶ Zirconia has been used in many new uses in medicine and dentistry in the past decade.³¹⁷ Example include orthopedic hip implants, endosseous dental implants and all-ceramic crowns.

Polymer surfaces in biomedical devices are common among many device types and categories. Coatings of medical devices are an important opportunity to create and maintain infection-free interfaces. Polyurethanes are versatile family of polymers in many medical devices such as pacemaker lead insulation, vascular prostheses and breast implants. Grafting polyurethane with poly(acrylic acid) can change the hydrophilic and lubricating properties of the polymer surface.³¹⁸ Acrylic urethanes were selected to demonstrate the feasibility of using chimeric antimicrobial peptides to create antibacterial surfaces on urethane surfaces, specifically, and polymer surfaces, generally.

Computational Structure Characterization Methods

To achieve chimeric peptide sequences that retain the functions of the component domains, interdomain interference must be minimized. Structural modeling of chimeric peptides provides an approach to estimate when domains are folding similarly in the chimeric peptide or much differently. Computational structure modelling of proteins can be achieved through a wide variety of tools such as Rosetta, Amber, CHARMM, GROMACS and others.^{117, 319-321} These tools allow for the modeling of structural changes in a domain due to sequence changes external to the domain. In this work, structures were modelled through

a PyRosetta script in Appendix B. This method uses a Rosetta approach to generate computational structure decoys with less computational work than molecular dynamics (MD) approaches. The reduced computational time compared to MD allows for screening more peptides in the design process.

The study of allostery,^{114, 322-325} how different binding events regulate the functionality of proteins, gives insight into a possible solution to the reduced performance of combining antimicrobial peptides with solid binding peptide seen in Table 4.1. Inspiration from the linking domains of multi-domain proteins was used to design spacers for chimeric peptides. Within a single polypeptide chain, the combination of a solid binding peptide and a displayed functional domain enables functional nanoscale interfaces between inorganic materials and biological systems.^{5, 6, 41, 48, 293} While the solid-binding peptide allows for the self-assembly of the chimeric peptide onto the inorganic surface through molecular recognition,^{30, 326} it is the linking sequence, or spacer, between the solid binding domain and the displayed domain that allows for self-organization of the antimicrobial peptide to retain its antibacterial activity.

The potential of spacer design is shown through the growing literature of the complex allosteric effects observed in biologically evolved proteins discovered both through structure determination techniques such as NMR and CD and through computational structure modeling techniques such as molecular dynamics and coarse grain models.^{323, 327, 328}

Spacer Sequence Design

The design of linking domains in proteins has been studied for more than 25 years.³²⁹⁻³³¹ A recent review compared linker design for naturally occurring, multi-domain proteins and those which are empirically designed.³³² Empirical linkers are further divided into flexible, rigid and *in vivo* cleavable. Two databases of naturally occurring linkers with designing tools were discussed. LINKER takes in queries with multiple inputs, including length and proteases to avoid.³³³ In this review, the development of databases of empirical linkers are discussed, concluding that a need exists for new search algorithms for linkers.³³²

This need is addressed in this work for new search algorithms by developing a codon-based genetic algorithm to find short peptide sequences of targeted length according to user specifications (Figure 5). In addition to the limitations noted by the reviewers, the design of the linking domains is intended for fusion proteins, making the use of the designs for combining peptide domains high risk. A method for mitigating this risk is to estimate the structure folding consequences of the choice of spacer domain in chimeric peptides according to the procedures in this chapter.

While many designs of linking domains in proteins exist in the literature and in databases, methods for designing spacer sequences for chimeric solid-binding peptide have not been reported. To design peptide sequences, a computer aided molecular design (CAMD) approach was developed for this work. A genetic algorithm was used (Figure 3.3) to consistently improve the spacer-induced structural features, as seen in Figure 5.1. The spacer design method begins with design targets and initial sequences to investigate. The candidate spacer sequences are scored by the distance from the design targets, then filtered by MLEM2 rules from the method in Section 2.3.1. Further customization is accomplished through adding target properties to the single-objective of the genetic algorithm. Structure preservation is evaluated for the active domains desired for the application through spacer frequency error (SFE). The lower the SFE, the closer the frequencies of the secondary structures are between the single domain decoys and the chimeric peptide domain decoys. The resulting spacer sequences are used to build chimeric peptide sequences with the newly designed spacers linking the given pair of domains.

The design of chimeric peptides for mineralization is approached by characterize the folding of solid binding peptides through three different methods. The first method is to characterize the peptide folding, both experimentally and computationally, for secondary structural features. The second approach is to characterize the peptide folding by protonation state distribution. The third approach is

to characterize the peptide folding through the surface area distributions. This work explores these characterization methods to serve as design targets for chimeric peptides.

5.3 Methods

5.3.1 Chimeric Spacer Design Method

The initial sequences studied here are from short peptide sequences based on flexibility and secondary structural ideals. The new candidate spacer sequences are generated through the genetic algorithm developed in this work (Figure 3.3). As part of the genetic algorithm, the spacer sequences are filtered by MLEM2 rules generated through the rough set theory method in Figure 2.1. Spacers are further customized with new design constraints. The overall methodology is detailed in Figure 5.1.

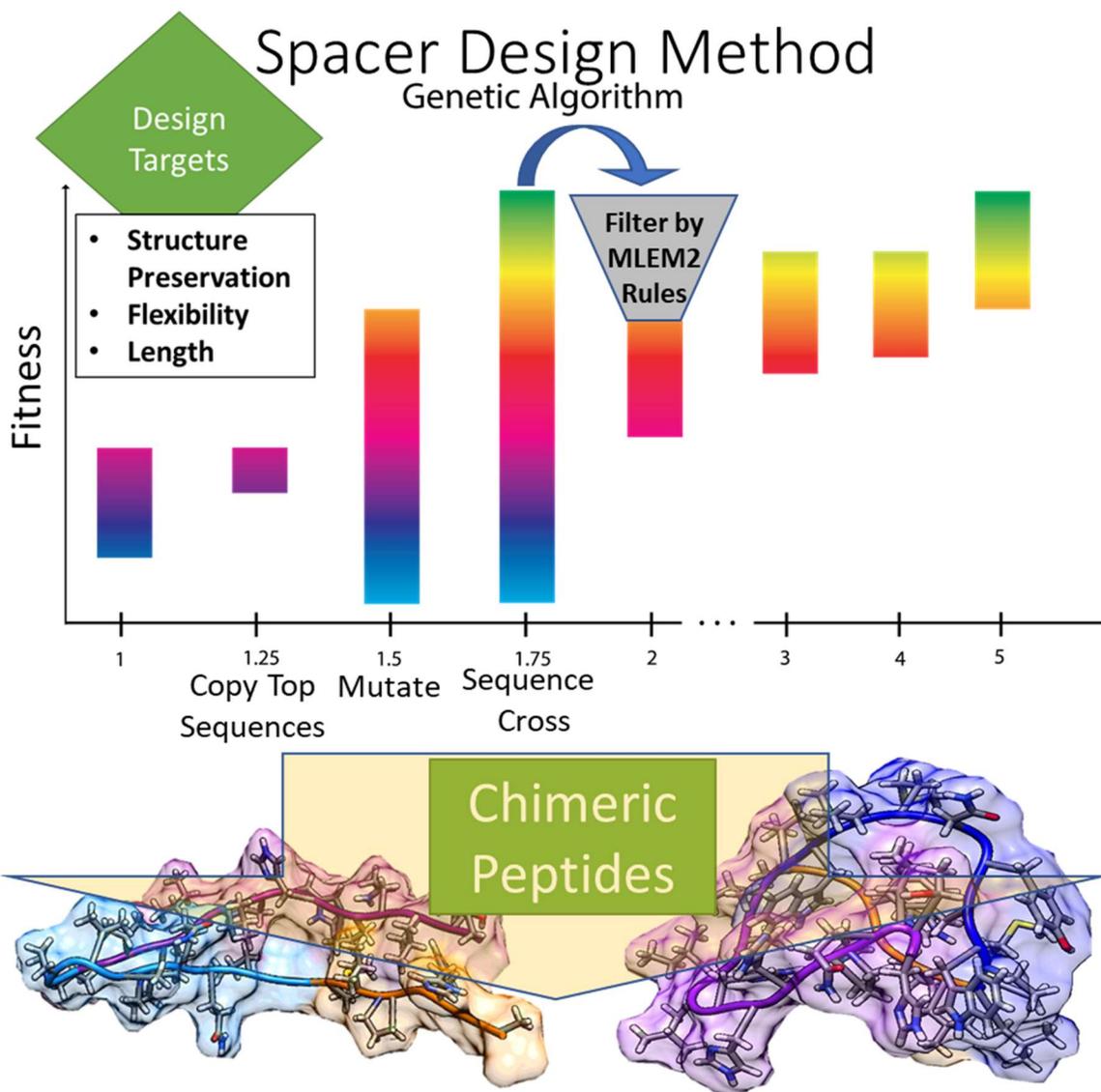


Figure 5.1 Process schematic for chimeric spacer engineering design.

5.3.2 Secondary Structure and Spacer Frequency Error

Ensembles of up to 1,000 structural decoys were generated for candidate chimeric peptide sequences and the single domain sequences. Each structural decoy was analyzed by the DSSP program²⁹⁷ to identify secondary structure within the sequence. The results of every DSSP analysis of single decoys are tabulated into a single text file, as in Chapter 4 methods (Section 4.3.2). This method differs from the

Chapter 4 method by including the residue position in the description of the secondary structural feature. For the residue position calculation, at each residue position each unique symbol (B, E, G, H, I, S and T) was summed for the ensemble, then divided by the number of decoys generated in the ensemble. To estimate the difference in the folding between the chimeric domain and the single domain, the frequencies at each residue position are summed in the spacer frequency error (SFE). This error was calculated as the sum of the absolute value of the differences between the feature frequencies of the chimeric residues from the single domain residues at each residue position.

Equation 5.1

$$SFE = \sum_{i=0}^n |f_{chimeric\ residue} - f_{domain\ residue}|$$

5.3.3 Circular Dichroism

Circular dichroism (CD) is a method for experimentally observing the differential absorption of polarized light across a spectrum in molecules.³³⁴ This technique is widely used for the characterization to secondary structure in peptides and proteins.³³⁵⁻³³⁹ Depending on the electron density of the carbonyl bond in the polypeptide backbone, circularly polarized light is absorbed differently. The secondary structure folding patterns of peptides and proteins can be inferred through CD decomposition analysis. CD Pro is a set of tools built for this purpose.³⁴⁰

The CD spectra were recorded with a CD spectrometer (JASCO, J-815) at room temperature, using a 1.0 mm cuvette. Each peptide sample was dissolved at 0.2 mg/mL in 100 mM Tris (pH 7.4) with and without TFE at 4°C for overnight. CD spectra were acquired from 190 to 260 nm at a scanning speed of 60 nm/min and were averaged from three runs per each sample. The secondary structure Far-UV CD spectra were processed with the tools of CD Pro.³⁴⁰ The mean residue absorbance was processed with

CRDATA.exe to create the input file for SELCON3.exe, CDSSTR.exe and CONTILL.exe. For each set of data, the reference set selected was SMP50. The fractions of secondary structure (Regular Helix, Distorted Helix, Regular Sheet, Distorted Sheet, Turns and Unordered) were averaged for all three CD Pro tools (SELCON3, CDSSTR, and CONTILL). This decomposition of secondary structure is used to compare with the PyRosetta computational structure estimation method (Section 5.3.2).

5.4 Results

5.4.1 Calcium Phosphate Surfaces

The chimeric antimicrobial peptide approach on titanium from Chapter 4 is being extended in Chapter 5 to calcium phosphate, zirconia and urethane polymer surfaces to demonstrate that the chimeric antimicrobial peptide technology is applicable to different types of biomaterials.

Structure-Function Relationship for Chimeric Antimicrobial Peptide

My colleague Hilal Yazici performed the synthesis of three peptides (cHABP1, HHC-36 and cHABP1-GGG-HHC-36). These peptides were chosen to build a chimeric antimicrobial peptide for calcium phosphate surfaces. cHABP1 is a solid binding peptide for calcium phosphate,^{307, 341} HHC-36 is an antimicrobial peptide with known activity against multi-drug resistant bacteria,¹⁵⁷ and the chimeric peptide cHABP1-GGG-HHC-36 is a chimeric peptide analogous to the titanium chimeric peptides in Table 4.1 in Chapter 4. Hilal Yazici also performed activity characterization with these three peptides against two strains of bacteria common in areas of surgical site infections, *E. coli* and *S. mutans*. In Figure 5.2, two different pathogens were evaluated by the optical density of well volumes at 600 nanometers (OD₆₀₀). The optical density measurement quantifies the scattering of the solution and is a measure of the quantity of bacteria present. As shown in Figure 5.2, (A) shows the relative growth of *E. coli* with various concentrations the studied peptides, and (B) shows the relative growth of *S. mutans* with various concentrations of the studied peptides.

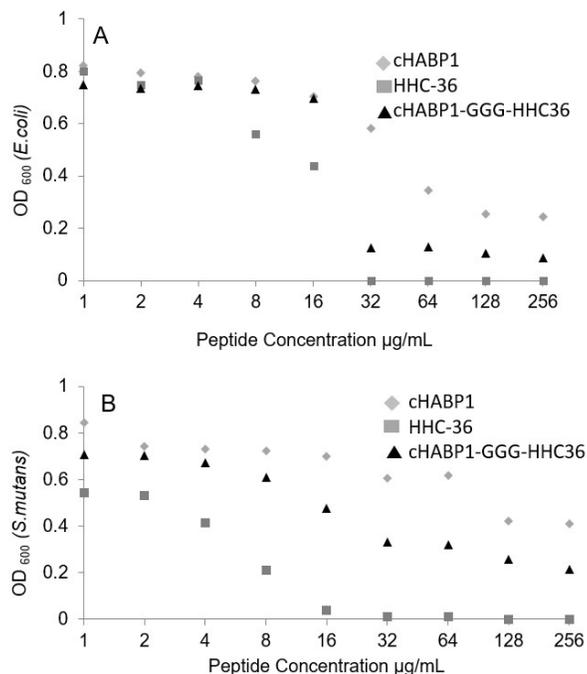


Figure 5.2. Bacterial turbidity (OD₆₀₀) as a function of hydroxyapatite binding chimeric peptide concentration and its component domains.

Table 5.1 displays the sequences of the peptides and their minimum inhibitory concentration. MIC values against some *E. coli* strains for HHC-36 have been previously reported, ranging from 2.5 µM – 5.4 µM.^{157, 342} As in Chapter 4, the MIC values for the chimeric antimicrobial peptides are higher than the MIC values for the antimicrobial peptide domain alone, indicating some loss of activity.

Table 5.1. Antimicrobial activity of peptides for the calcium phosphate surface.

Peptide Name	Sequence	<i>S. mutans</i> IC ₅₀	<i>E. coli</i> IC ₅₀
cHABP1	CMLPHHGAC	128 µg/mL	48 µg/mL
HHC-36	KRWWKWWRR	3 µg/mL	12 µg/mL
cHABP1-GGG-HHC36	CMLPHHGACGGGKRWWKWWRR	16 µg/mL	18 µg/mL

A secondary structure study on these peptides was performed using the PyRosetta script discussed in Appendix B.^{31, 343} The secondary structure frequencies were graphed in R. The α-helix formation frequency

trends positively with antibacterial activity level: the highest for HHC-36, lower for cHABP1-GGG-HHC36 and the lowest for cHABP1, as shown in Figure 5.3. The structure generation data analysis for these three peptides reinforces the trend of increasing α -helical structure frequency with increased antibacterial activity.

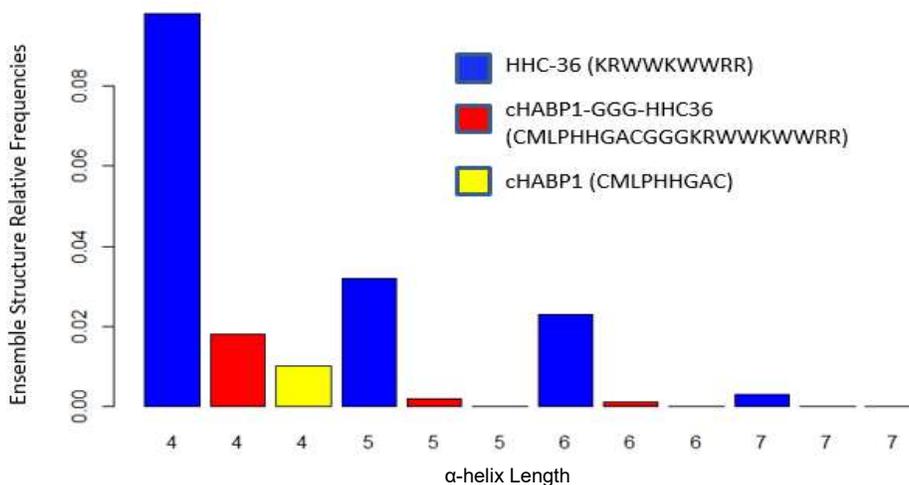


Figure 5.3. Helical structure frequencies for hydroxyapatite binding chimeric peptide and its components

The trend of α -helix length to antibacterial activity among these three peptides combines the results shown in Table 5.1 with Figure 5.3. For an analysis that relates many kinds of secondary structure features to antibacterial activity, the MLEM2 algorithm, as discussed in Chapter 2, is used. 500 structure decoys for each sequence generated from the PyRosetta script method (Appendix B) were reviewed with the MLEM2 algorithm. Six secondary structure rules were discovered to be specific between the peptides cHABP1, HHC-36 and cHABP1-GGG-HHC36, and these are shown in Table 5.2.

Table 5.2 Inducted rules by MLEM2 for secondary structure features of chimeric hydroxyapatite binding peptide and its component domains.

Rule	Secondary Structure Feature	Sequence	Feature Frequency for Sequence	Feature Frequency for Other Sequences
1	4-a.a.- α helix & 2-a.a.-turn	HHC-36 (KRWWKWWRR)	40/500	4/1000
2	6-a.a.- α helix	HHC-36 (KRWWKWWRR)	35/500	0/1000
3	5-a.a.- α helix & 2-a.a.-turn	HHC-36 (KRWWKWWRR)	37/500	0/1000
4	4-a.a.-bend & 2-a.a.-bend	cHABP1-CCC-HHC36 (CMLPHHGAC GGGKRWWKWWRR)	43/500	0/1000
5	4-a.a.-bend & 2-a.a.-bend	cHABP1-CCC-HHC36 (CMLPHHGAC GGGKRWWKWWRR)	49/500	0/1000
6	2-a.a.-turn	cHABP1 (CMLPHHGAC)	85/500	14/1000

Figure 5.4 shows the distinguishing secondary structure features for each of the three peptides.

Secondary Structure Features 1-3 are unique to HHC-36, while Secondary Structure Features 4 and 5 are unique to cHABP1-CCC-HHC36. Secondary Structure Feature 6 is unique to cHABP1.

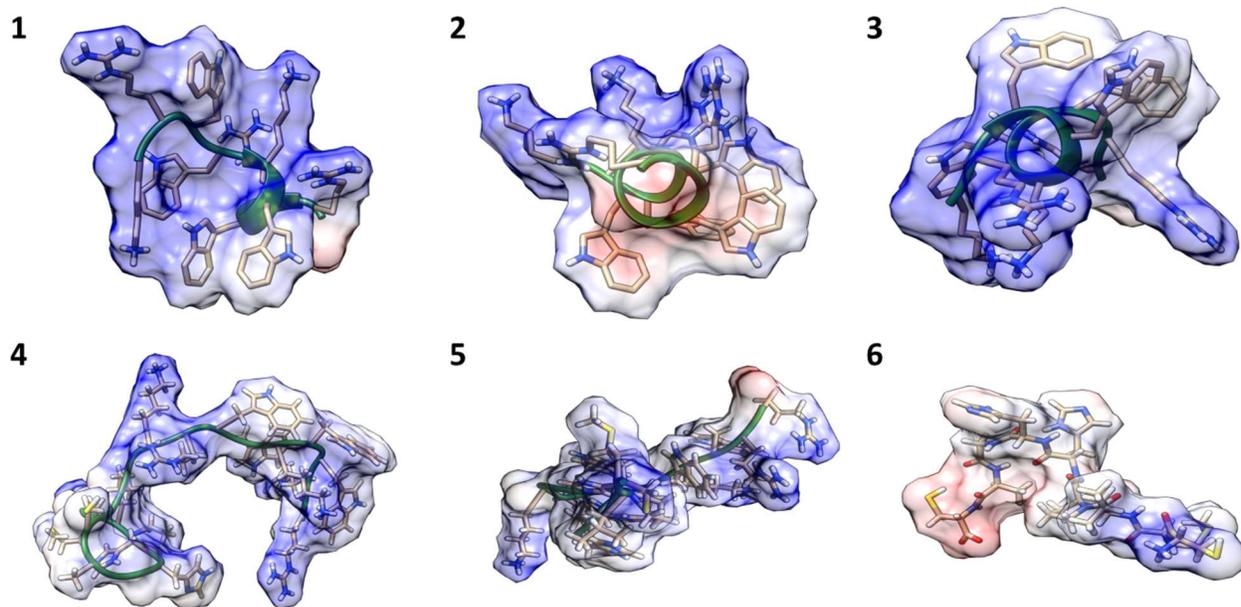


Figure 5.4 Characteristic structures fitting each of the inducted rules in Table 5.2 for HABP-AMP and its component domains.

Peptide Characterization for Mineralization Design

In this work, a secondary approach for infection-free interfaces is to mineralize the interface from calcium and phosphate ion sources. HABP1 was identified as a mineral forming peptide by Tamerler's group.³⁰⁷ Three separate approaches for characterizing the interfacial properties of this peptide in two forms are presented compared to a non-mineral forming peptide. The first approach is to characterize the secondary structure of the peptide through CD analysis and computational structure generation. The second approach is to characterize the distribution of protonation state likelihoods with intrinsic pK_a analysis. The third approach is to characterize the surface area distribution change due to a hydrophobic or a hydrophilic side chain substitution.

Secondary Structure Characterization of Mineralization Peptide

The secondary structure folding patterns of HABP1 are characterized, in both linear and constrained forms, through experimental measurement and computational structure generation methods. The folding patterns serve as design guides to know when peptide structure is changing with spacer

engineering for chimeric peptide applications. Figure 5.5 shows the folding characterization with (A) referring to the CD spectrum of each of the peptides in Tris buffer at pH 7.4 and (B) referring to the CD Pro decomposition of secondary structures as an average of three methods (SELCON, CONTILL, CDSSTR).

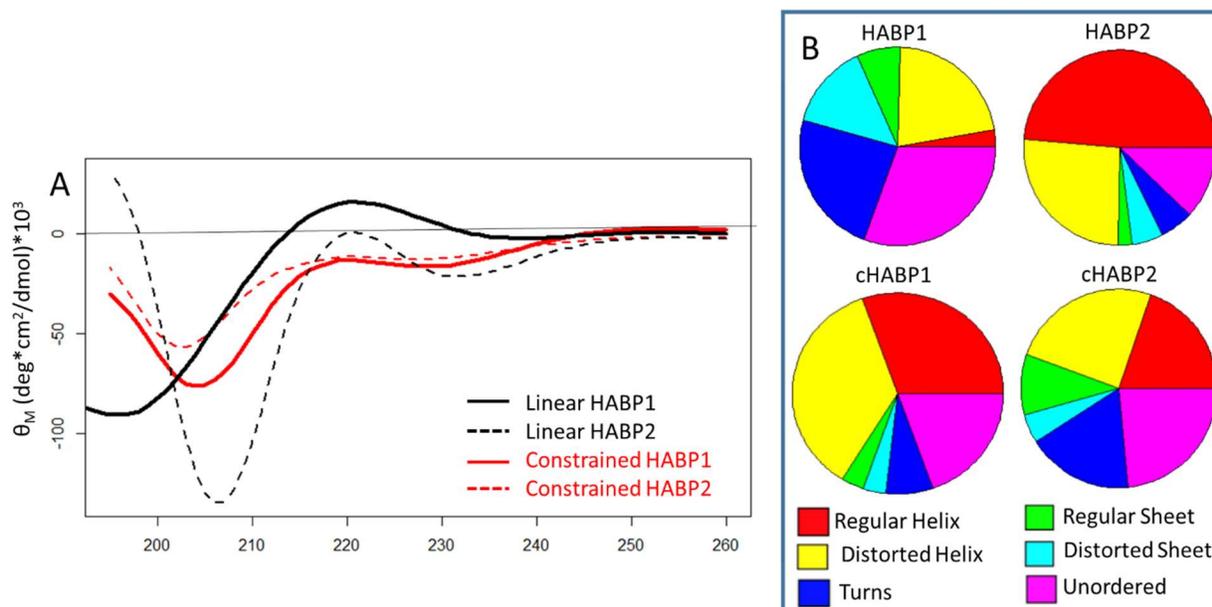


Figure 5.5 Structure characterization of linear HABP1, HABP2, constrained HABP1 and constrained HABP2 through circular dichroism analysis.

The folding of the peptides near mineral surfaces can be inferred using a kosmotropic (water-ordering) agent. In this work, 2,2,2-trifluoroethanol (TFE) was selected.³⁰⁷ The TFE environment results in a more ordered secondary structure of linear HABP1, as seen in Figure 5.6. However, changes in TFE concentrations for cHABP1 TFE had little effect on its CD spectrum.³⁰⁷ In Figure 5.6, linear HABP1 changes more in response to changes in TFE than cHABP1. The trend of increasing TFE concentration for linear HABP1 shows increasing similarity of secondary structure decomposition to constrained HABP (Figure 5.6B). This observation supports that there may be similarities between the folding of HABP1 on the mineral surface and the folding of cHABP1 on the mineral surface.

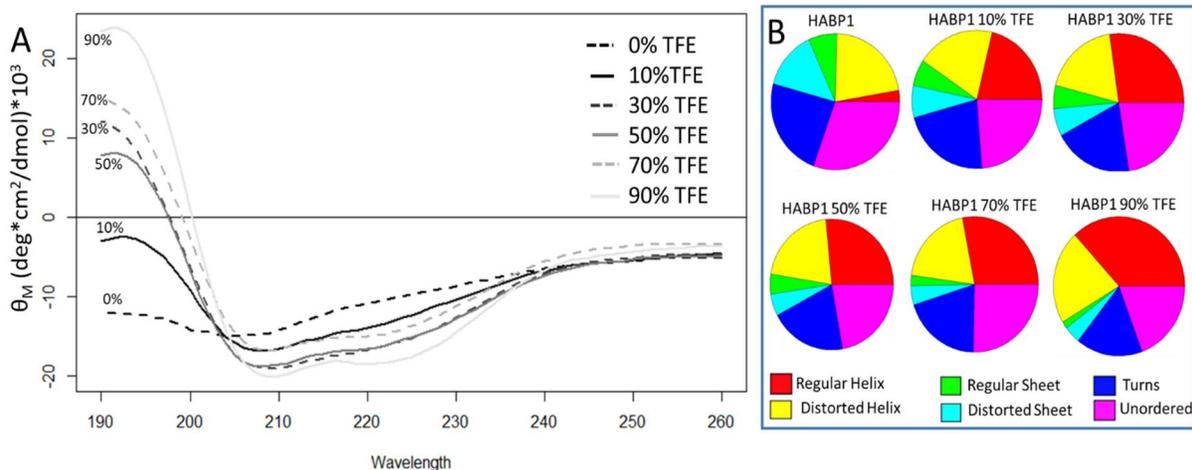


Figure 5.6. Solvent engineering study with 2,2,2-trifluoroethanol (TFE) for HABP1.

Computational Secondary Structure Patterns

Flexibility is the structural entropy of the peptide ensemble. Secondary structures represent a small percentage of possible structural states, and thus are low entropy states. The flexibility of the peptide sequences is shown by secondary structure frequencies by residue position.

In this work, the folding properties of HABP1, HABP2, cHABP1 and cHABP2 are explored through computational structure decoy ensembles using the PyRosetta method (Appendix B). These ensembles are used to estimate the flexibility of the peptide by estimating the frequencies of structures at a given residue position, as seen Figure 5.7 for HABP1 and HABP2 in linear and constrained forms.

The changes in peptide folding due to the disulfide bond difference between linear and constrained forms of HABP1 and HABP2 can be seen in Figure 5.7. The constraint of the disulfide bond and increase in peptide length from 7 to 9 amino acids reduces the frequency of right handed residue decoys for leucine at residue position 2 and proline at residue position 3 for HABP1 (Figure 5.7A) compared to cHABP1 (Figure 5.7B). The constrained HABP2 peptide shows some changes to the dominate orientation for P, G and F (Residues 3-5 in cHABP2) compared to linear HABP2. These differentiated secondary

structure features provide design targets for study in spacer design within the cHABP1 sequence to tune the interactions between the varied geometries of the faces of forming calcium phosphate mineral.

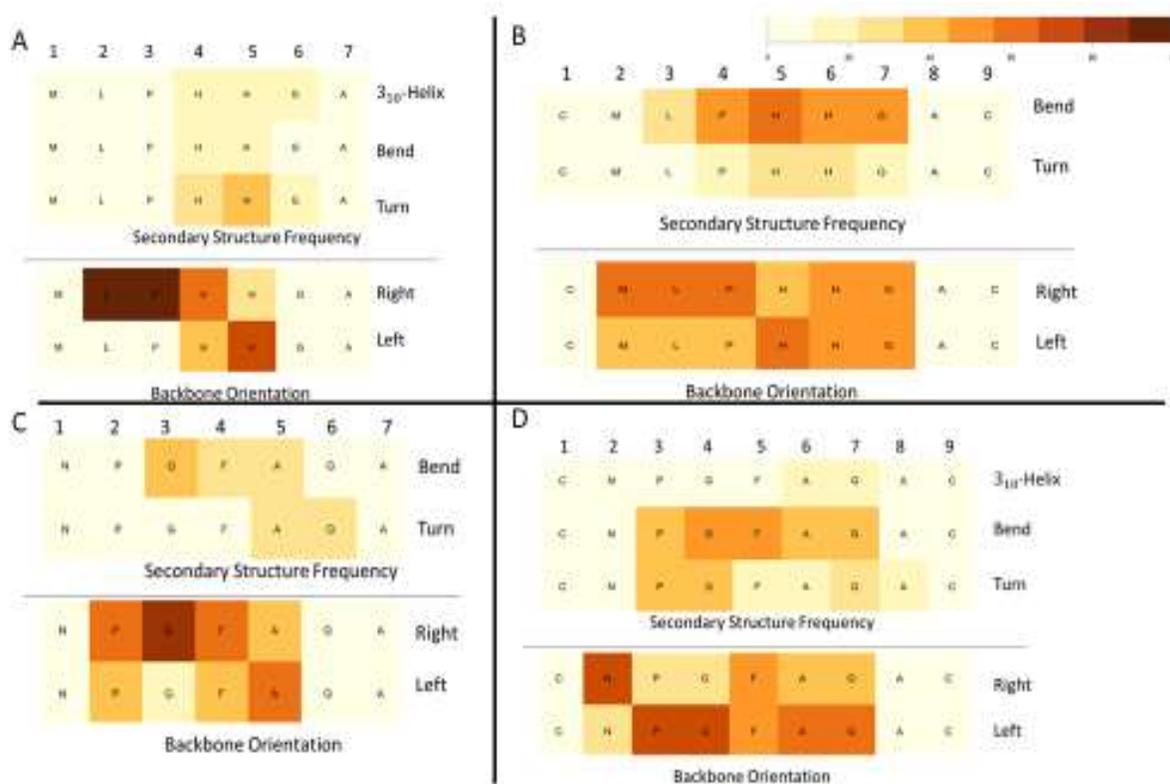


Figure 5.7 Estimation of Flexibility Study by computer-generated structures to estimate folding frequencies across the residues of the HABP1 sequences; A) HABP1, B) Constrained HABP1, C) HABP2, D) Constrained HABP2.

Protonation State Distribution of Mineralization Peptide

The second approach to characterize the mineralization peptide interface is the estimation of the pK_a , the pH at which the frequency of two adjacent protonation states of a peptide are equal. The estimation is the intrinsic pK_a which estimates the peptide pK_a from a folded structure. This value is calculated using Propka 3.1³⁴⁴ as a function of the generated structural ensembles, as seen in Figure 5.8.

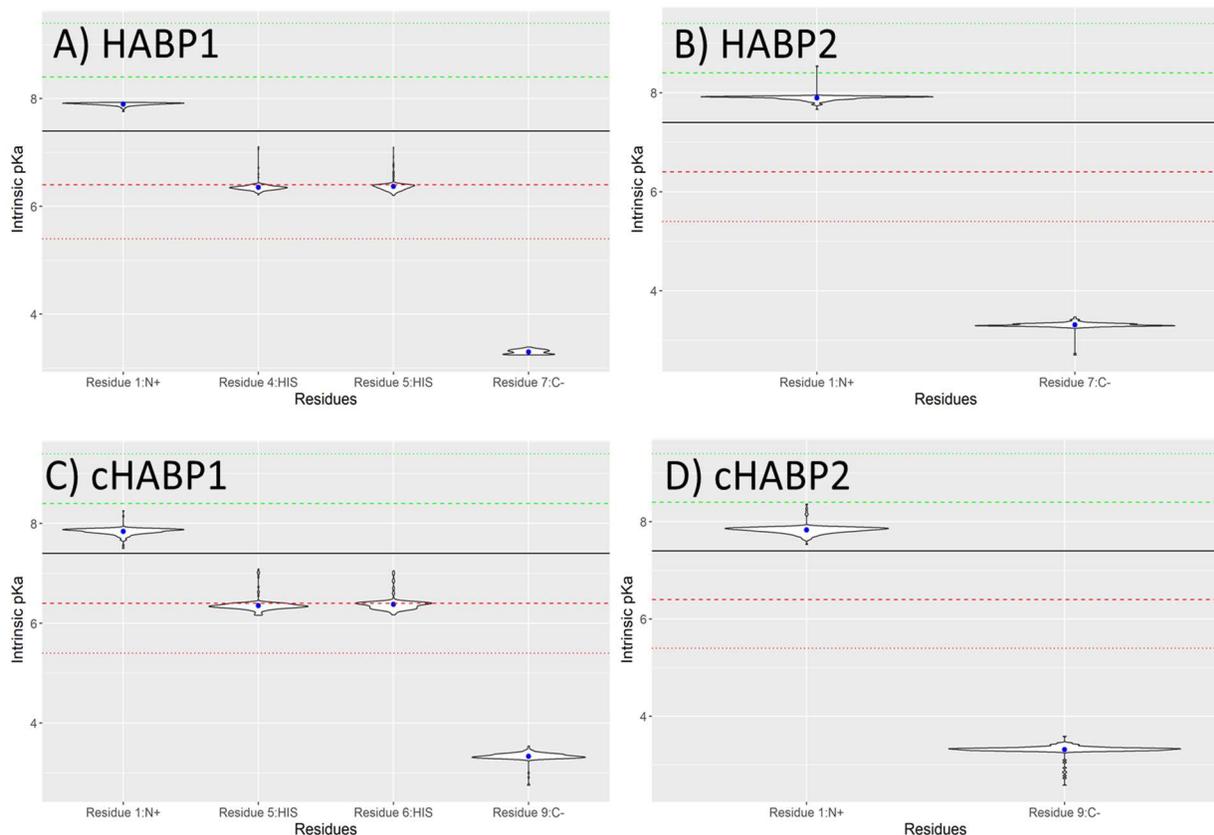


Figure 5.8 Computational Intrinsic pKa Study to estimate effect of disulfide bond on peptide structure between linear peptides (A) HABP1 and (B) HABP2 and constrained peptides (C) cHABP1 and (D) cHABP2.

The solid black line in Figure 5.8 indicates that the bioenabled mineralization assay occurs at pH 7.4 which is the normal physiological pH. The dotted lines indicate one or two logarithmic units away from the pH of the assay. Species with pK_a values within the first dotted line have 10% or more of their protonation states changed. The green dotted line are species with deprotonated states at pH 7.4, and the red dotted line are species with protonated states. Species between the dotted lines are 1% to 10% changed with higher pH values indicating deprotonation. Species outside of the dotted lines have less than 1% of the species with a changed protonation state. The terminals of cHABP1 and cHABP2 have slightly wider ranges of pK_a values compared to HABP1 and HABP2.

Surface Area Distribution as a Function of Side Chain Substitution

The third approach to characterize the mineralization peptide interface is the surface area distribution. To determine how the surface area affects the mineralization process, this work calculates the surface area changes of linear HABP1 for side chain substitutions. First, an alanine scanning study shows the surface area for a substitution of alanine along the peptide sequence in Figure 5.9. The position substitution with the maximum hydrophilic surface area was HABP1-M1A.

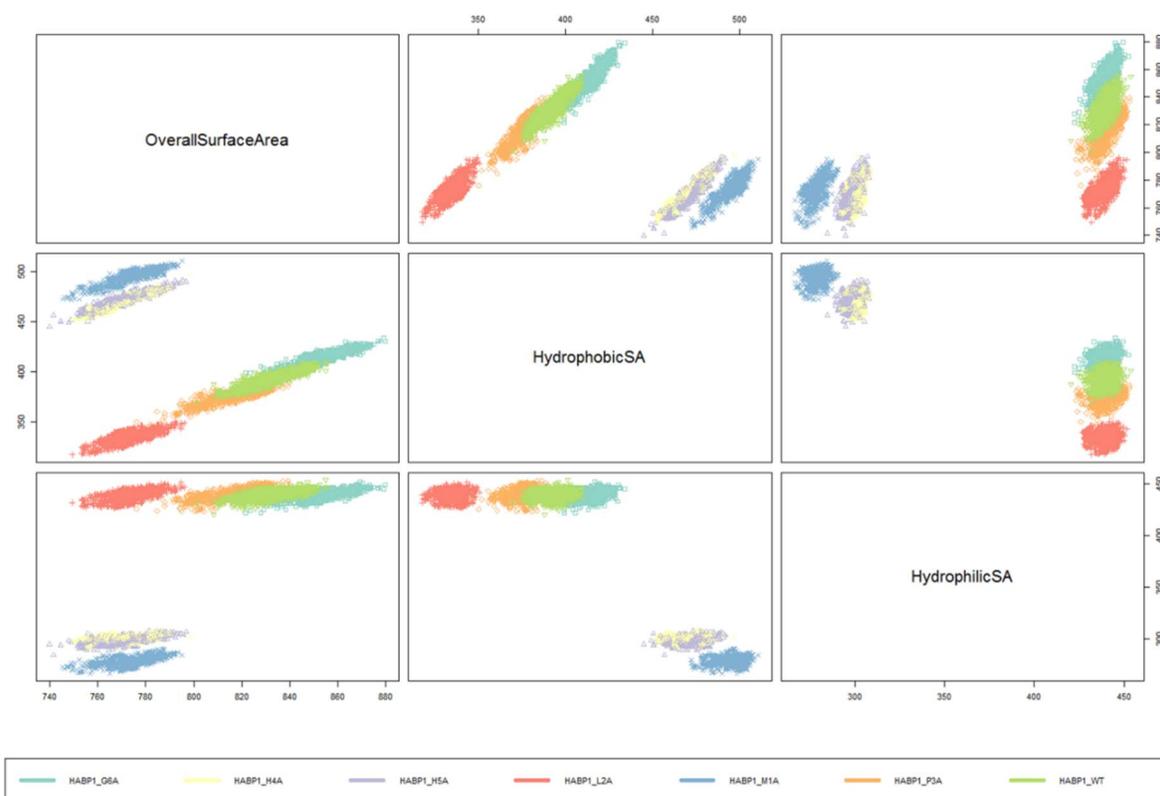


Figure 5.9. Surface area changes of ensembles of structural decoys due to substitution of alanine at a single position.

Alanine has a hydrophobic side chain. Substituting histidine at each residue position is used to explore hydrophilic side chain substitutions as a contrast. Figure 5.10 shows the clustering of surface area by sequence mutation. The position substitution with the maximum hydrophilic surface area was HABP1-A7H.

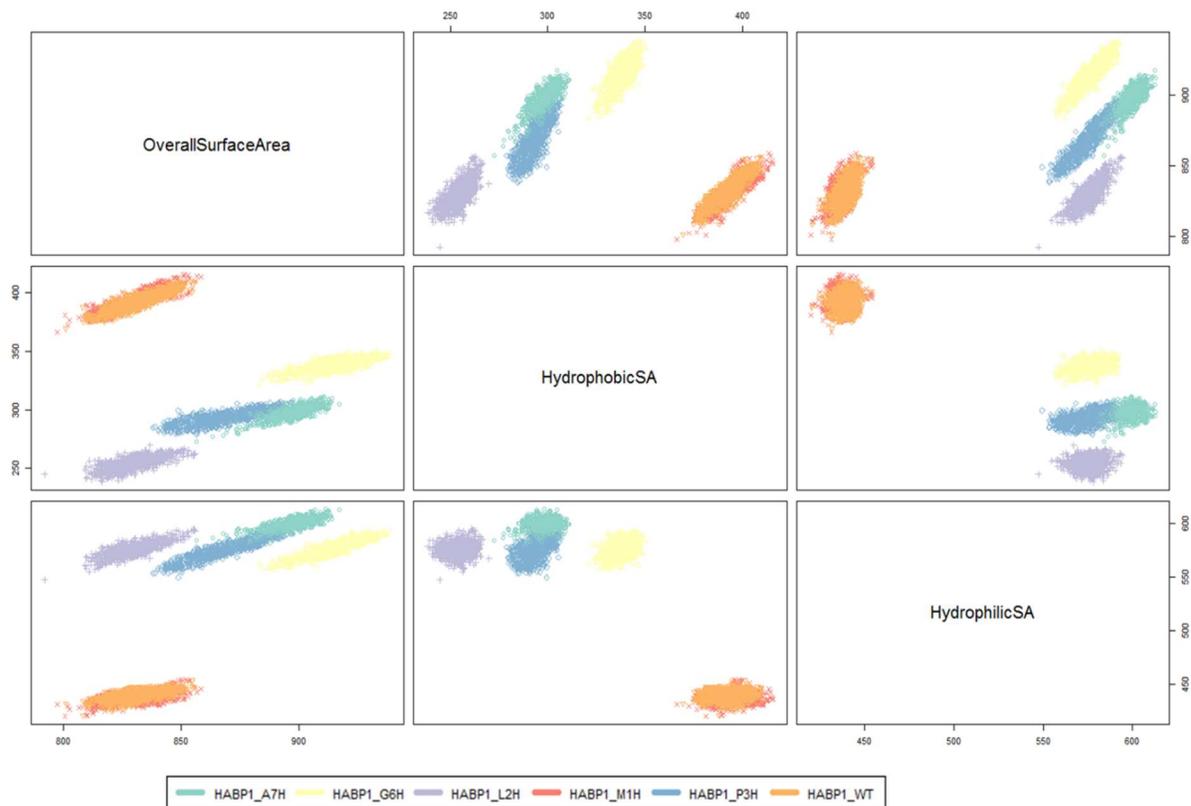


Figure 5.10 Surface area changes of ensembles of structural decoys due to substitution of histidine at a single position.

Fast and Slow Mineral-Formation Kinetic Domains in a Single Chimeric Peptide

A chimeric peptide designed for mineralization control can combine the functions of distinct peptide domains. One example is to combine peptides that have slow and fast kinetics to investigate if the mineral morphology seen with the slow kinetic peptide can be sped up by the fast-kinetic peptide without changing the mineral morphology. Table 5.3 shows the sequences for spacer evaluation and the two peptides selected, one for its fast kinetics (ADP1)³¹¹ and one for its formed mineral morphology (HABP1).³⁰⁷

Table 5.3: Initial Spacer Sequences as a Modeling Basis for Spacer Design Method

Peptide Name	Peptide Sequence
ADP1	HTLQPHHHIPVV
HABP1	MLPHHGA
ADP1-Rigid Spacer-HABP	HTLQPHHHIPVV PAPAP MLPHHGA
ADP1-ADP8 Spacer-HABP	HTLQPHHHIPVV GPVAPQQPG MLPHHGA
ADP1-Soluble AH1 Spacer-HABP	HTLQPHHHIPVV KGSVLSA MLPHHGA
ADP1-Soluble AH2 Spacer-HABP	HTLQPHHHIPVV PKSALQEL MLPHHGA
ADP1-Hydrophobic AH1 Spacer-HABP	HTLQPHHHIPVV GLALLGWG MLPHHGA
ADP1-Hydrophobic AH2 Spacer-HABP	HTLQPHHHIPVV LGWLSAV MLPHHGA
ADP1-Pi Helix 1 Spacer-HABP	HTLQPHHHIPVV WLMNYFWPL MLPHHGA
ADP1-Pi Helix 2 Spacer-HABP	HTLQPHHHIPVV YLMNYLLPY MLPHHGA
ADP1-Original Linker-HABP	HTLQPHHHIPVV GGG MLPHHGA
ADP1-Flexible Spacer-HABP	HTLQPHHHIPVV GGGSGGG MLPHHGA

Figure 5.11 shows a structural decoy of one of the chimeric peptides. Structure ensembles were built for each peptide in Table 5.3 to compare the structural frequencies. The spacer frequency error (SFE) was calculated as the sum of the absolute value of the frequency differences between the single domain and the same sequence in the chimeric peptide. The SFE results are grouped by functional domain and ranked by median SFE. The boxplots represent repetitions of building structural ensembles. Three repeats were made for all sequences in Table 5.3.

Green: ADP1 (amelogenin-derived peptide 1)
 Blue: GGG spacer
 Cyan: HABP1

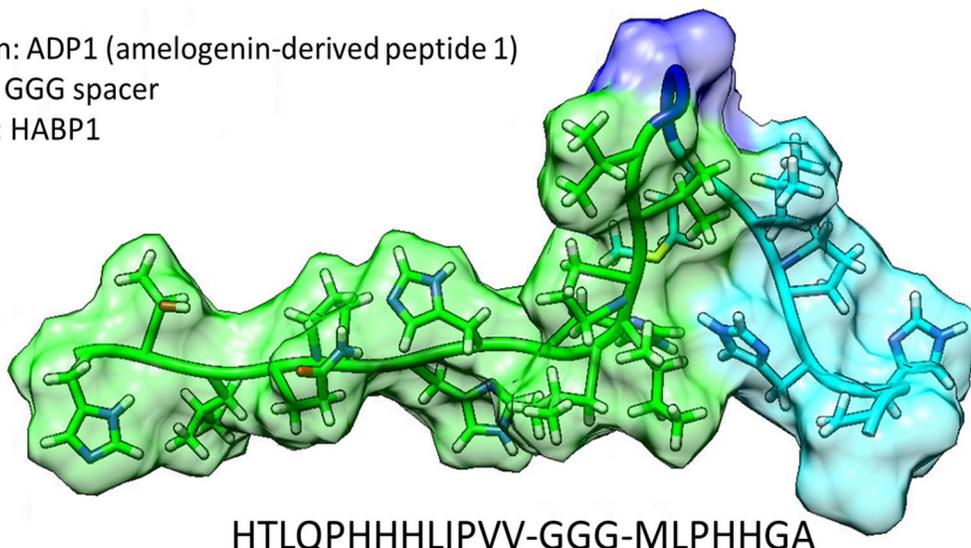


Figure 5.11 Structure decoy of chimeric peptide combining fast mineralization kinetics with octacalcium phosphate mineral-forming morphology generated in PyRosetta.

In Figure 5.12, The feature frequencies are calculated by DSSP and averaged by category (helix, beta sheet, turn/bend and right/left orientation). Each category is scaled to level the contribution from each category to the error score. The box plots represent repetitions of generating sets of 1,000 computational structures with the same Robetta fragment files (N=3).

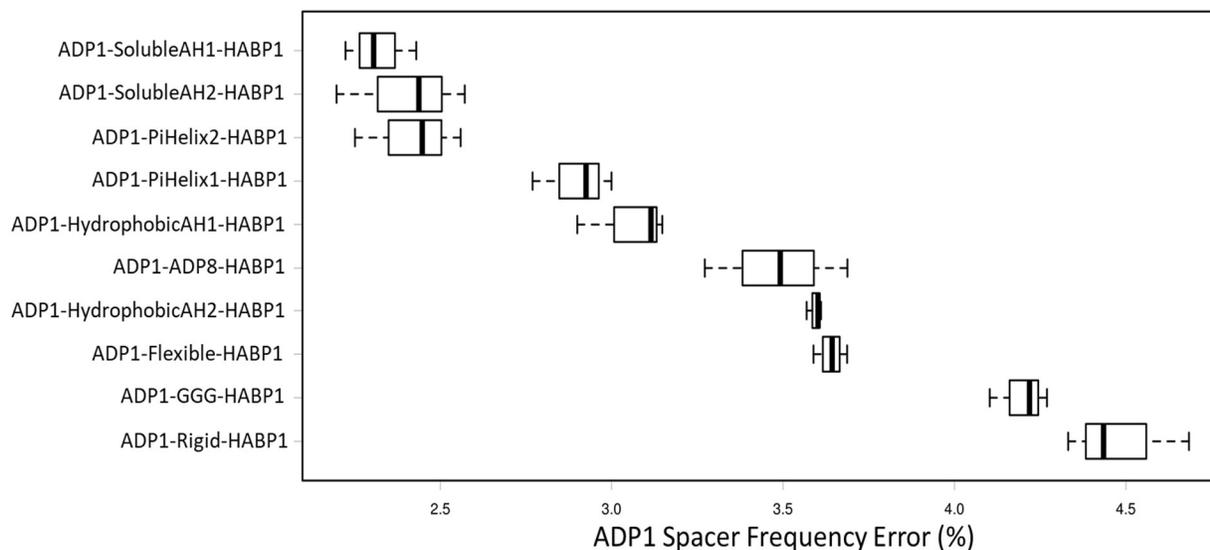


Figure 5.12: Summed error of chimeric peptides compared to ADP1 peptide alone of feature frequencies at each residue position.

In Figure 5.13, the feature frequencies were calculated by DSSP and averaged by category (helix, beta sheet, turn/bend and right/left orientation). Each category was scaled to level the contribution from each category to the error score. The box plots represent repetitions of generating sets of 1,000 computational structures with the same Robetta fragment files (N=3).

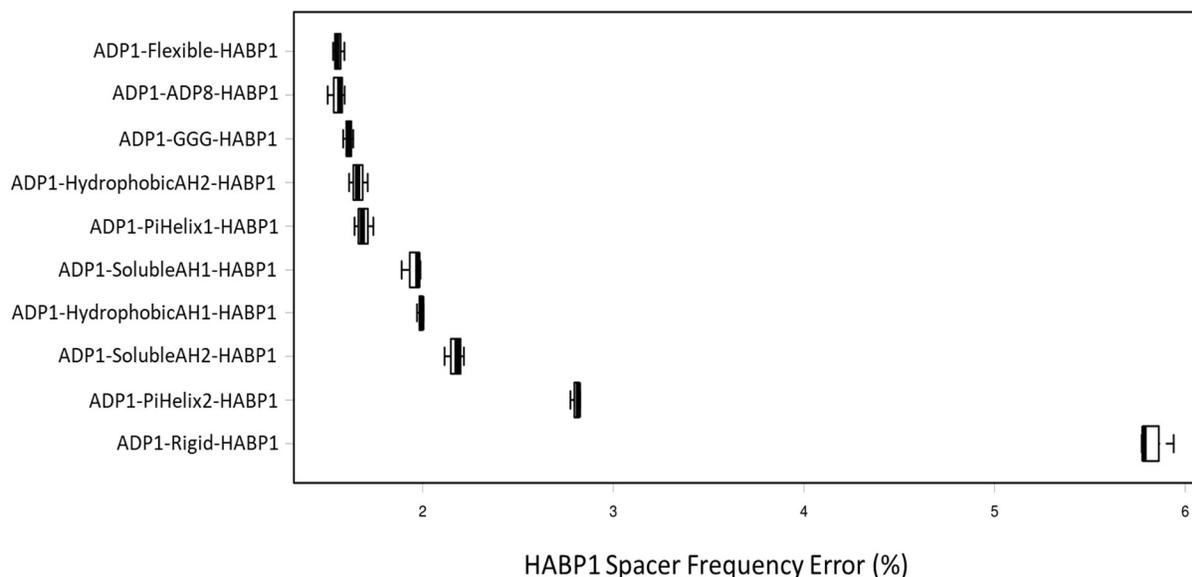


Figure 5.13: Summed error of chimeric peptides compared to HABP1 peptide alone of feature frequencies at each residue position.

5.4.2 Zirconia Surfaces

For the chimeric antimicrobial peptide concept, changing the surface for antibacterial activity requires changing the solid-binding peptide and customizing the spacer for each pairing of a solid binding peptide and an antimicrobial peptide. The computer-aided molecular design (CAMD) approach is used for designing chimeric peptide spacers for antimicrobial peptides. The forward problem of quantitative relationships are structure-function relationships discovered through the MLEM2 method (Figure 2.1).^{60, 345} Structure ensembles of 1,000 energy minimized structures are formed for low, medium and high antibacterial activity antimicrobial peptides using PyRosetta tools.^{81, 343} Secondary structure feature rules for high antibacterial activity are shown in Table 5.4. The MLEM2 induced rules in this table were

calculated by combining the data tables from which rules in Table 4.2 and Table 5.2 are generated. The chimeric peptides are for targeting two surfaces, titanium and calcium phosphate. The antimicrobial peptides are AMP1 and AMP2 from Yucesoy *et al.*⁶

Table 5.4. Selected secondary structure feature rules from the MLEM2 method relating to high antibacterial activity from data mining structure ensembles.

Secondary Structure Feature	Pathogen	Antibacterial Activity	Ensemble Frequency
12-a.a.- α helix	<i>E. coli</i>	High	4.2%
11-a.a.- α helix	<i>E. coli</i>	High	3.3%
5-a.a.-α helix	<i>E. coli / S. mutans</i>	High	2.2%

As shown in Figure 5.1, the chimeric spacer design process begins with spacer sequences of different putative secondary structures. The initial basis is to start with evaluating the structural frequencies of different kinds of spacers: flexible, putative π helices, and putative α helices, both hydrophobic and hydrophilic. To find chimeric peptides which have high antibacterial activity, the 5-aa α -Helix frequency was calculated as a function of spacer sequence. The average among the chimeric peptides with each of the two antimicrobial peptide domains, AMP1 (LKLLKLLKLLKLL) and AMP2 (KWKRWWWWWR), and each of two zirconia-binding domains, ZrBP (RPRENRGRERF) and ZrBPM1 (RPREQRGRER). These average frequencies are in Table 5.5.

Table 5.5. 5-a.a. α helix frequencies for the initial generation of chimeric spacers.

Spacer	Sequence	5-aa α -Helix Frequency
π -Helix 1 Spacer	WLMNYFWPL	5.4%
π -Helix 2 Spacer	YLMNYLLPY	5.0%
Soluble α -Helix Spacer 2	PKSALQEL	4.3%
Original Spacer	GGG	4.0%
Flexible Spacer	GGGSGGG	3.9%
Hydrophobic α -Helix Spacer 1	GLALLGWG	3.8%
Hydrophobic α -Helix Spacer 2	LGVVLSAV	2.5%
Polyalanine Spacer	AAAAAAA	0.1%
Soluble α -Helix Spacer 1	KGSVLSAD	0.0%

The crossover and mutation operators are applied to the top half of the initial generation to populate the second generation. The secondary structure feature frequency was averaged across four separate chimeric peptide sequences: ZrBPS1-Spacer-AMP1, ZrBPS1-Spacer-AMP2, ZrBPS2-Spacer-AMP1 and ZrBPS2-Spacer-AMP2. The crossover and mutation operators are applied to the top half of the initial generation to populate the second generation. The resulting average 5-aa α helix frequencies of three second generation spacers is in Table 5.6. A structural decoy with the 5-aa α helix occurrences labeled for the IGVVISAV spacer is in Figure 5.14.

Table 5.6 Second generation of spacers derived from initial generation of spacers in Table 5.5.

Spacer	Sequence	5-aa α Helix Frequency
L1I & L5I Soluble AH2	IGVVISAV	7.3%
Soluble AH2 x π Helix 2	KGSVYLLPY	7.1%
π Helix 1 x Soluble AH2	WLMNLSAD	5.9%

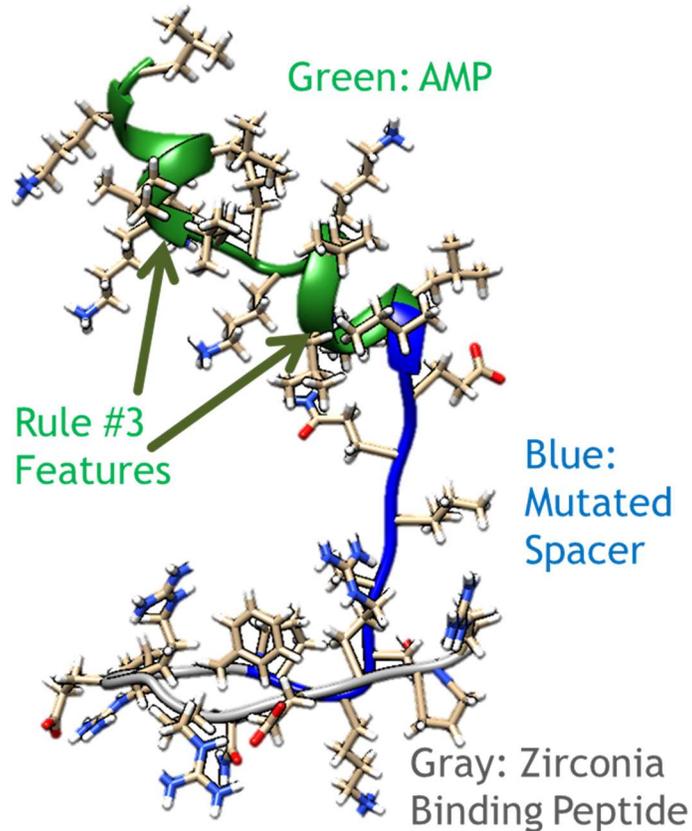


Figure 5.14. Structural decoy showing two 5-aa α helices in a chimeric peptide (ZrBP-IGVVISAV-AMP1) with a second-generation spacer domain (IGVVISAV).

5.4.3 Urethane Polymer Surfaces

This case study seeks to design a chimeric antimicrobial peptide for a urethane polymer surface.³⁴⁶ The secondary structure features as a function of residue position are shown for a solid binding peptide, PUABP1 (GRAVRRSIRRRV),³⁴⁷ and antimicrobial peptide AMP1 (LKLLKLLKLLKLL), and a chimeric peptide PUABP1-GSGGG-AMP1 in Figures 5.15-5.17. PUABP1-GSGGG-AMP1 (GRAVRRSIRRRV GSGGG LKLLKLLKLLKLL) is designed as a chimeric antimicrobial peptide to coat urethane polymeric materials with antimicrobial activity. The structures are generated according to the PyRosetta method (4.3.1) and

analyzed according to the position-specific secondary structure method (5.3.2).

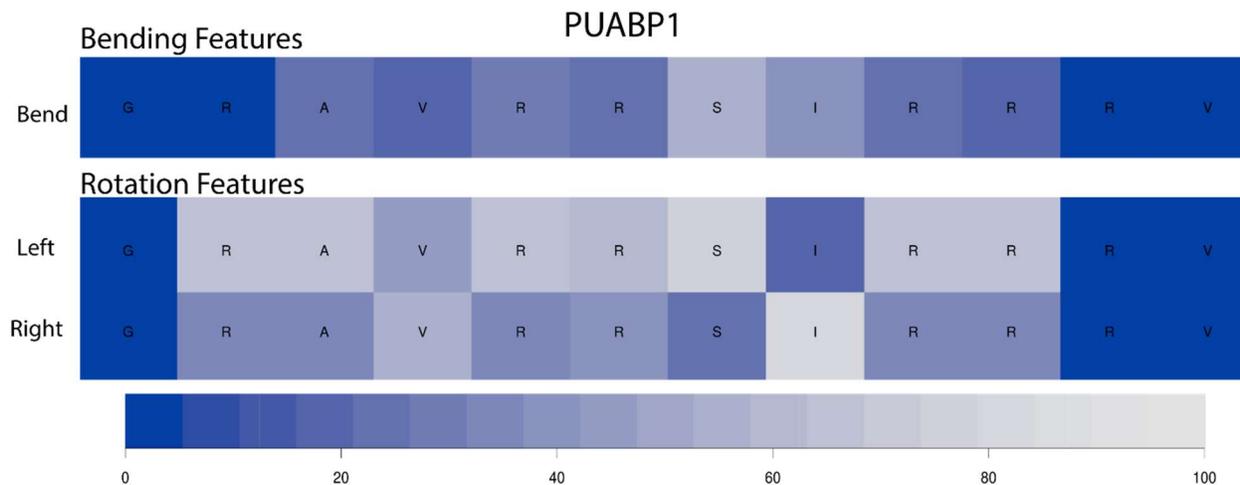


Figure 5.15. Secondary structure feature frequency by residue position of PUABP1.

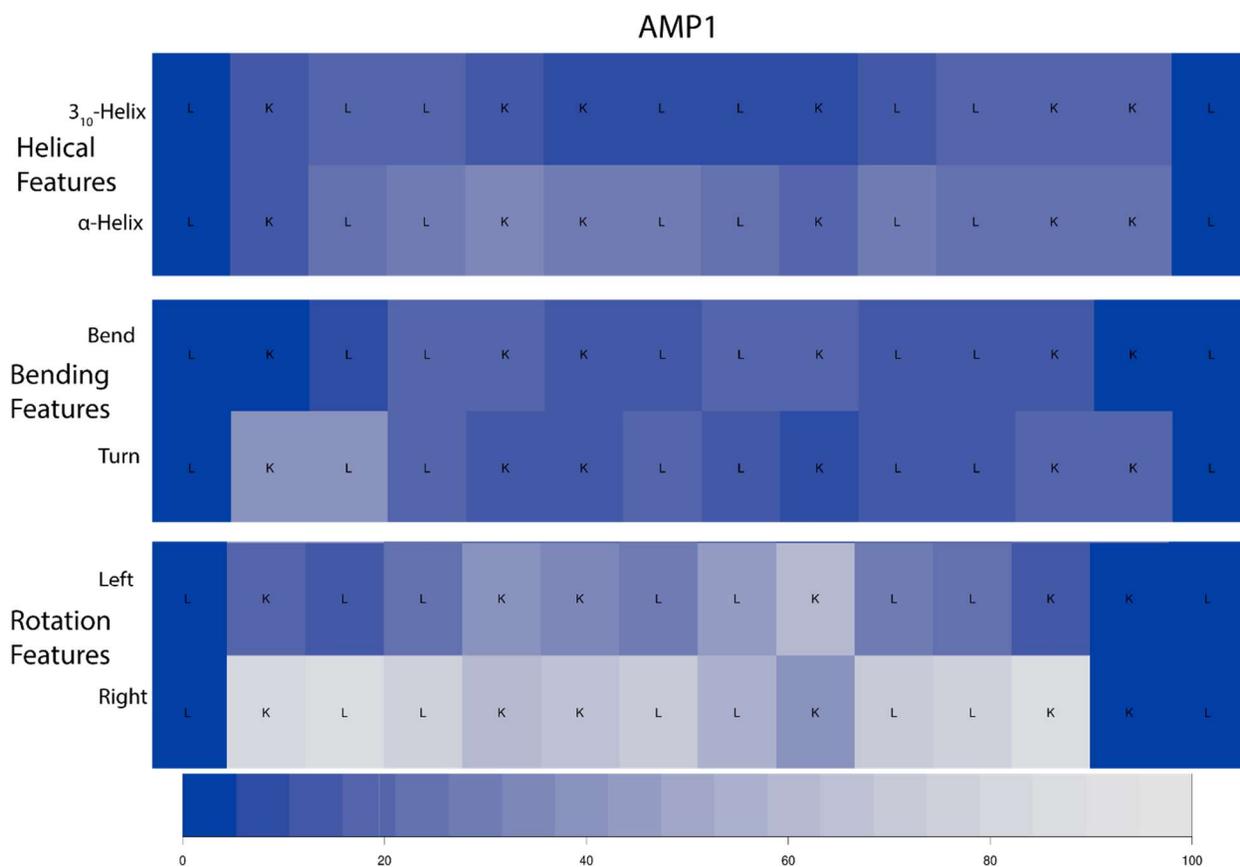


Figure 5.16 Secondary structure feature frequency by residue position of AMP1.

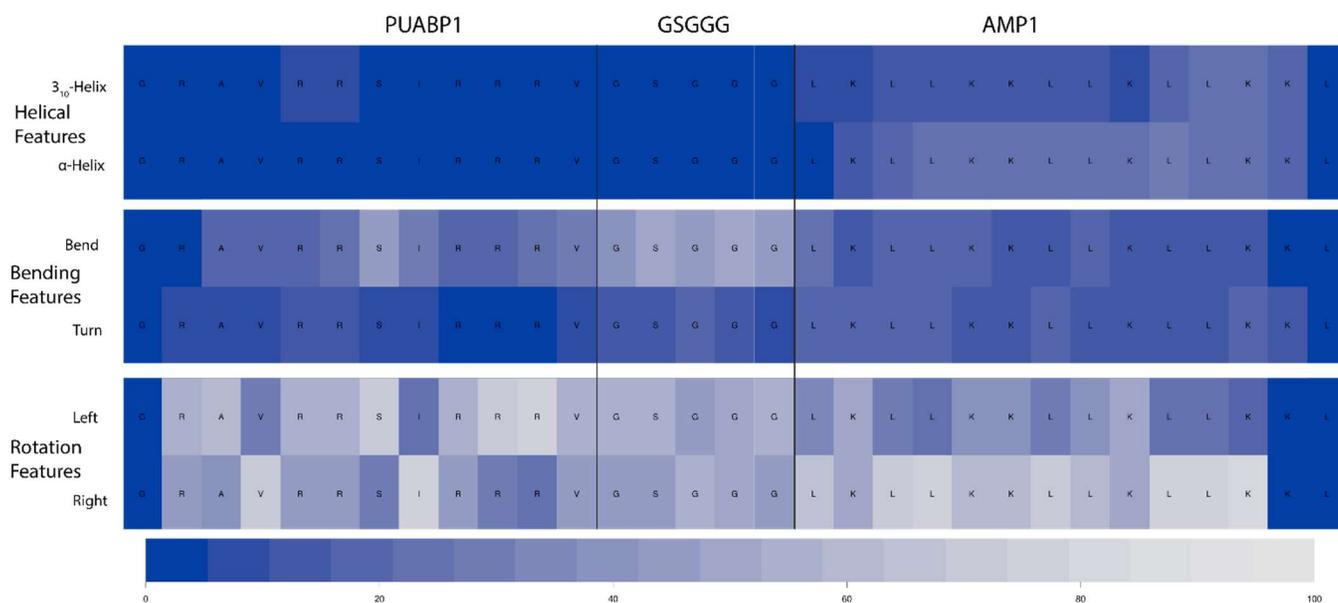


Figure 5.17 Secondary structure feature frequency by residue position of PUABP1-GSGGG-AMP1.

The spacer design scheme for preserving as much of the predicted secondary structure as possible minimizes the differences between the predictions of the chimeric peptide domain and the single domain. The Manhattan distance was used to define the differences between the individual domain frequencies and the chimeric domain as the spacer frequency error (SFE), which is defined in Equation 5.1 in Section 5.3.2.

SFE is calculated by feature category (helical features, beta sheet features, bends and turns, and left/right orientation). These features are normalized and the summed to calculate the overall spacer frequency error as described in Section 5.3.2. These calculations are plotted for repetitions of structure generations of 1,000 decoys (Figure 5.19-5.20). The solid binding peptide candidates PUABP1 and PUABP2 both have GSGGG as a highly ranked spacer for either antimicrobial peptide tested. GSGGG is the top spacer for both antimicrobial peptides for PUABP1 and for the top spacer for PUABP2 and AMPa. The top scoring spacer for PUABP2 and AMP1 was Soluble Alpha Helix 1 (KGSVLSAD). These rankings may be further improved by reducing the variation of SFE estimates by incorporating more structure decoy sets or by incorporation of larger structure decoy sets whose statistical frequencies of secondary

structures converge. The wide boxplot bars in Figures 5.18-5.19 have estimates of SFE that vary enough to lead to uncertainty into the proper ranking of the spacers because of the overlapping boxplot ranges.

In Figure 5.18, the chimeric peptides are ranked by the median SFE. The box plots are repetitions of the SFE, created by generating 1,000 structure decoys and generating the secondary structure frequencies (N=6). PUABP1 (GRAVRRSIRRRV) and PUABP2 (AIRGIRGIRGIR) are the candidate solid-binding peptide domains. AMP1 (LKLLKLLKLLKLL) and AMPa (KWKLWKKIEKWGQGIGAVLKWLTTWL) are the candidate antimicrobial peptide domains.

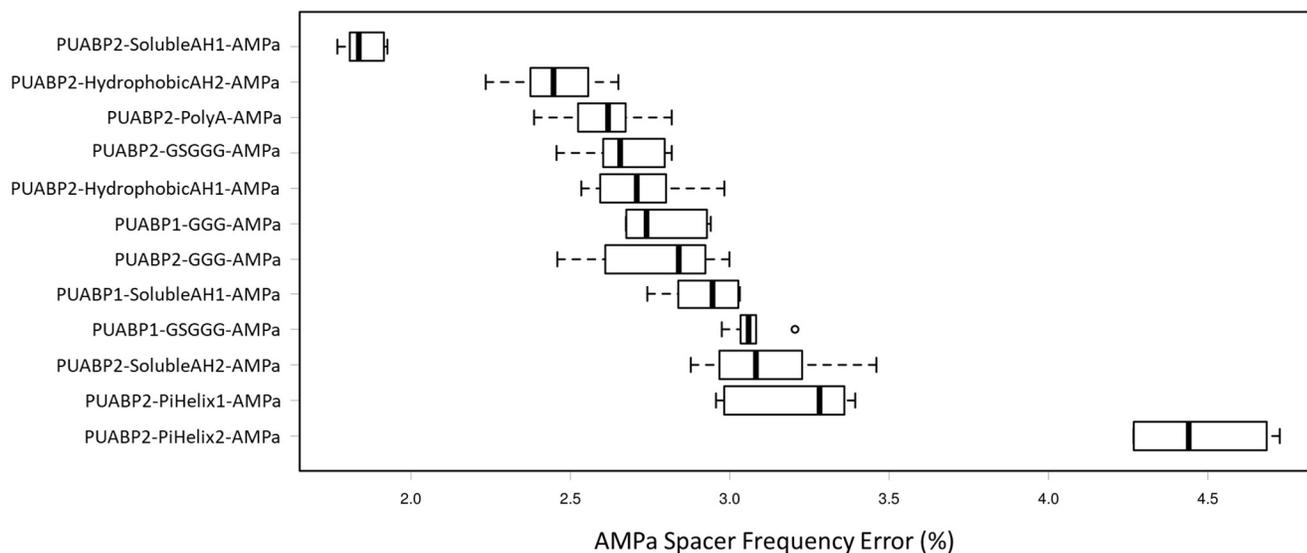


Figure 5.18 Spacer frequency error (SFE) for selected chimeric peptides with AMPa (KWKLWKKIEKWGQGIGAVLKWLTTWL).

In Figure 5.19, the chimeric peptides are ranked by the median SFE. The box plots are repetitions of the SFE by generating 1,000 structure decoys and generating the secondary structure frequencies (N=6).

PUABP1 (GRAVRRSIRRRV) and PUABP2 (AIRGIRGIRGIR) are the candidate solid-binding peptide domains.

AMP1 (LKLLKLLKLLKLL) and AMPa (KWKLWKKIEKWGQGIGAVLKWLTTWL) are the candidate

antimicrobial peptide domains.

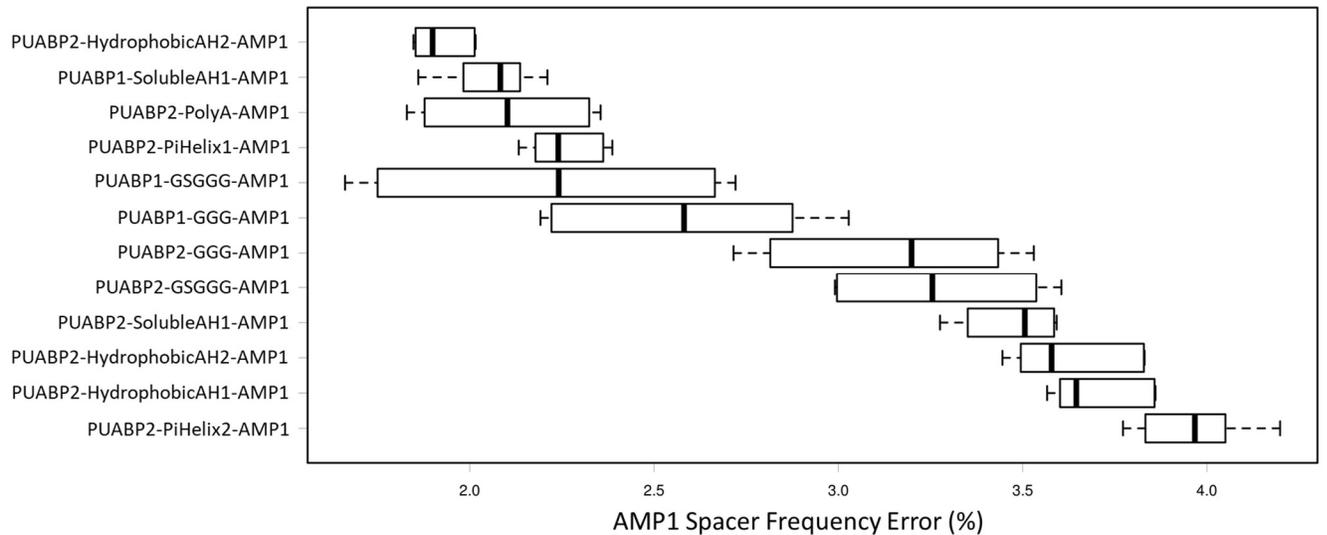


Figure 5.19 Spacer frequency error (SFE) for selected chimeric peptides with AMP1 (LKLLKLLKLLKLL).

The candidate spacers are ranked according to the preservation of secondary structures in single domain sequences compared to the chimeric domain sequences. The spacer frequency error is spacer dependent. Future work is needed to determine the number of structure decoy sets or the number of structure decoys in a set to get convergence of these frequency calculations.

5.5 Discussion

5.5.1 Calcium Phosphate Surfaces

In this work, chimeric peptide technology was designed for three different kinds of surfaces: mineral, ceramic and polymer. The mineral selected is calcium phosphate. Chimeric antimicrobial peptide designs were supplemented with a secondary antibacterial strategy of a formed mineral barrier. Three different properties of the mineralization peptides were explored as design targets for chimeric peptide designs.

Chimeric Antimicrobial Peptide

The secondary structural features in Table 5.2 can be used to design new chimeric antimicrobial peptide spacers that close the gap in activity between HHC-36 and cHABP1-GGG-HHC36 in Table 5.1 using the methods described in Chapter 4. As part of closing the gap, secondary structure design structures are

found to be helpful in the spacer design in Chapter 4. This work in Chapter 5 discovers the secondary structural features to target for chimeric antimicrobial peptides for calcium phosphate surfaces.

Mineralization Peptide Protonation State Distribution

Mineralization is a secondary antibacterial barrier approach. The protonation state effects the attraction or repulsion of the mineralization peptide to ions and mineral components in the mineral formation process. If the pK_a of a side chain or terminus changes, the distribution of the protonation state of the peptide will change at a stable pH. This change will be larger the closer the pK_a value is to the given pH. This work calculates the intrinsic pK_a distributions of a mineralization peptide due to the variation of folding in the structural decoys generated. This work has provided protonation state distributions for which to target chimeric peptide design.

The shielding effect of larger peptide sequences may provide new trends in the intrinsic pK_a values. For these short peptides, no interior is shielded from solvent; the entire peptide sequence is the exterior surface. For larger peptides with interior ionizable amino acids, larger variations of intrinsic pK_a values are expected. Large variations of pK_a values have been measured in many proteins, including Staphylococcal Nuclease.³⁴⁸

Mineralization Peptide Surface Area Distribution by Substitution

The surface area distribution of a mineralization peptide is compared to mutants which either have a hydrophobic residue substitution of alanine or a hydrophilic residue substitution with histidine. The mutant with the largest hydrophobic surface area was HABP1-M1A, and the mutant with the maximum hydrophilic surface area was HABP1-A7H.

Slow and Fast Mineralization Kinetics in a Chimeric Peptide

Different mineralization mechanisms are designed to be combined by adding together slow and fast mineralization kinetics in a single chimeric peptide. Amelogenin has demonstrated that this target is

achievable. The chimeric peptide combines an amelogenin subsequence ADP1 with fast mineralization kinetics with the slow-formation kinetics of HBP1 selected from combinatorial biology protocols. To maintain the original folding of both domains within the chimeric peptide, SFE was calculated. Low SFE values indicate secondary structure features that closely match the secondary structure features in the single mineralization peptide. This work estimates the SFE to provide guidance in ranking candidate spacers. The results of the spacer error rankings do not agree between domains in that the same spacer did not result in the minimum error for both domains.

5.5.2 Zirconia Surfaces

Chimeric antimicrobial peptides are designed with novel spacers through the genetic algorithm approach in Figure 5.1. The design targets for the secondary structure feature were calculated from combining the data bases of titanium and calcium phosphate chimeric antimicrobial peptides. The initial generation of spacers were selected for variety of secondary structures. Some of these initial spacers increased the targeted secondary structures compared to the GGG spacer. A second generation of spacers were formed through recombination operators of mutation and crossover to produce a spacer which improved on the initial generation for the targeted secondary structure.

5.5.3 Urethane Polymer Surfaces

Chimeric antimicrobial peptides are designed for polyurethane acrylic surfaces with a new approach to describing structural features. A limitation of the previous approach in Chapter 4.3.2 was to count structure features independent of where they occurred along the chimeric sequence. The secondary structures with a specific side chain type at a specific residue position would not be separated from the same structure at a different residue position. New types of structure-function relationships can be investigated by accounting for the positions of the secondary structure within the sequence. The

position-based secondary structure feature differences are summarized by the SFE. An initial generation of spacers were evaluated according to the SFE.

This work has shown two methods of using structure-function relationships to design spacer sequences between peptide domains. The first method uses the existence of secondary structure of a specific length and the second method uses the existence of a secondary structure feature at a specified position within the sequence to determine the structure description.

5.6 Conclusions

Despite the design of linking domains in protein and peptides being a long-studied topic in the literature, there does not exist a universally applicable method of design for peptide domain interactions. For this work, a structure description tool was developed as a customized script (PyRosetta Peptide Folding Method, Appendix B) for evaluating structure-function relationships and the effect of the spacer sequences of chimeric peptides. This structure generation tool may be combined with data mining techniques (CLN-MLEM2, Figure 2.1) and a genetic algorithm search (codon-based genetic algorithm, Figure 3.3) to engineer spacers which optimize the frequencies of secondary structure in the connected domains of the chimeric peptides. This work explores four different types of materials: metals, minerals, ceramics and polymers. The chimeric solid binding peptide concept is a platform technology.

In Aim 1, a novel method for customizing the design of antimicrobial peptides was developed through solving the forward problem of computer-aided molecular design (CAMD) in Chapter 2 and the reverse problem of CAMD in Chapter 3. The second aim of this work, in Chapter 4, was to engineer chimeric peptides to improve the formation and maintenance of an infection-free interface between titanium implant surfaces and the host tissue. In Chapter 5, chimeric antimicrobial peptides were demonstrated as a platform technology across calcium phosphate, zirconia and urethane polymer surfaces. In Chapter 6, this platform was further explored through the integration into a dental adhesive system.

This page is intentionally left blank.

6.0 Engineered Antimicrobial Peptide Enhances Dental Adhesive

6.1 Introduction

Due, in part, to aesthetic appeal, dental composites are overtaking amalgam for the repair of posterior teeth. However, this transition from amalgam to composite is not without consequences. The clinical lifetime of dental composites is a half to a third that of amalgam. The result for dentists is that more than half of dental practice time is spent on dental composite repairs and replacements, and the result for patients is increasing tooth loss whenever composites are repaired or replaced. The most common cause of dental failure are secondary carries due to oral bacteria invading the adhesive/dentin interface. Therefore, the interface between the remaining tooth and dental composite is a weak interface.

In this work, an engineered antibacterial primer adhesive is developed that is hydrophilic to completely wet the demineralized dentin layer as part of an integrated dental composite system. While commercial dental adhesives with antibacterial properties are available using quaternary amines, toxicity to dental tissues and oral bacterial resistance are concerns. These concerns are addressed by coupling an antimicrobial peptide, inspired by the many native oral antimicrobial peptides in saliva, to the primer adhesive. The folding of the engineered antimicrobial peptide was characterized using a kosmotropic agent. The dental adhesive material system developed in this work shows antibacterial activity against *S. mutans* UA159 antimicrobial peptide. This work provides a path to conjugate the antimicrobial peptide to the primer adhesive polymer for longer term availability of antimicrobial activity.

6.2 Background

Traditionally, direct dental restorations for posterior teeth were made from a silver-tin alloy known as dental amalgam. Composite restorations are overtaking amalgam as the most popular material for posterior teeth. These restorations are often very stable materials, but they have reduced average clinical lifetimes. These reduced lifetimes are due to bacteria degrading the adhesive/dentin interface. The current reported longevity for composite restorations is 7.8 years as compared to 12.8 years for amalgam.³⁴⁹

Clinical data suggest that over half of all dental restorations are replacements for failed restorations.³⁵⁰ Recurrent caries are the largest contributor (36.5%) followed by restoration loss (19.6%) and unknown reason (17.7%).³⁵¹ While many reviews stress patient factors in the approach to recurrent caries³⁵², material engineering can also address the prevention of recurrent caries. Biodegradation by bacteria of the adhesive bond interface is considered a critical contributor to secondary loss of adhesion, microleakage and decay.³⁵³ This interface is the weakest link in the composite restoration.³⁵⁰ Therefore, adhesives with antibacterial interfaces should provide protection against bacteria.

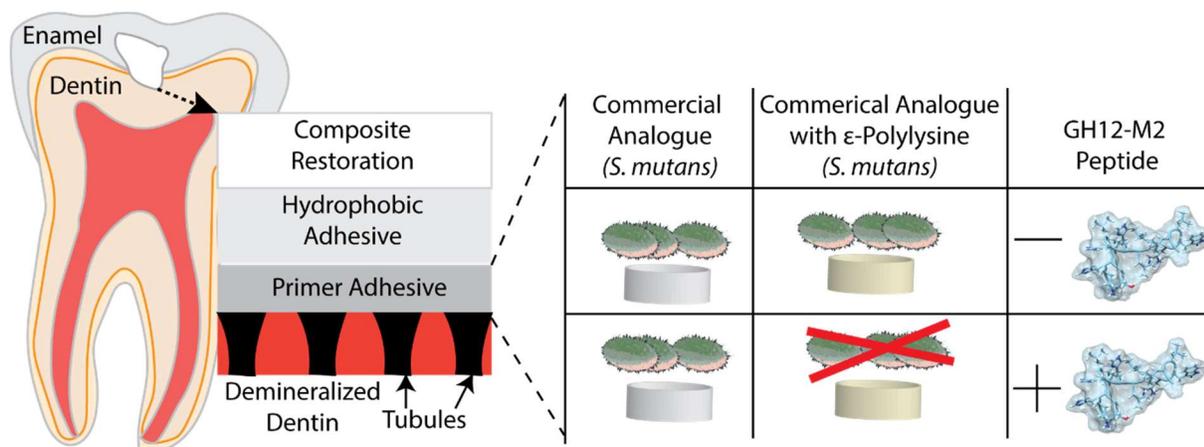


Figure 6.1 Dental primer adhesive system for antimicrobial barrier to prevent bacterial degradation

Antimicrobial approaches are being investigated to improve the lifetimes of composite restorations. One option is to add metallic ions to build an antibacterial adhesive/dentin interface.³⁵⁴ A limitation of this method, as with amalgam, is the reduced aesthetic appearance of the dental restoration. Discoloration due to metallic ions may be mistaken for tooth infection or decay. Small amounts of chlorhexidine gluconate (CHX) have also been used.³⁵⁵ However, long-term exposure to CHX has known adverse effects such as mouth discoloration, mucosal irritation and desquamation of the gums.³⁵⁶ A method of more directly targeting the enzymes that *S. mutans* uses to break down adhesive is to mix in

dextranomer microspheres to competitively remove the enzymes.³⁵⁷ While removing the enzyme would prolong the life of the adhesive, this approach would also store the enzyme for eventual release.

Antimicrobial peptides are a nature-inspired approach to antibacterial activity. In the oral environment, many different antimicrobial peptides are present.³⁵⁸⁻³⁶⁰ A recent review has catalogued which antimicrobial peptides have been evaluated for use in the oral environment against which pathogens.³⁶¹ None of these studies, however, attempted to integrate the antimicrobial peptide into a dental adhesive system. A recent study has incorporated nisin, a peptide food preservative, in a dental adhesive mixture at 1% without significantly lowering the bond strength of the adhesive.³⁶² At a mixture of 3% or 5% nisin, the bond strength was reduced. While inhibition of *S. mutans* was shown within the material, no zone of inhibition surrounding the mixture was observed.

To address the critical problem of bacterial degradation at the adhesive/dentin interface, a dental adhesive system is developed in this work with multiple mechanisms for resisting bacterial degradation and inhibiting bacterial growth near the interface. An amine-functionalized dental adhesive system builds on Spencer's group previous work to address the challenges of the hydration difference between the dental adhesive and the demineralized dentin.³⁶³⁻³⁶⁵ The major components in dental adhesives include monofunctional monomers, such as 2-hydroxyethylmethacrylate (HEMA), and difunctional monomers, such as bisphenol A-glycidyl methacrylate (BisGMA) and triethylene glycol dimethacrylate (TEGDMA).^{366, 367} HEMA is a common component of most dental adhesives, due to its unique solubility in both hydrophilic and hydrophobic environments.³⁶⁸ Difunctional monomers are used in adhesive to enhance the properties of the polymers to withstand the rigors of the oral environment.^{369, 370} TEGDMA functions as a crosslinker and shows a relatively high degree of conversion.³⁶⁹ Polymers bearing a phosphorylcholine group have been studied as biocompatible materials,³⁷¹⁻³⁷³ and 2-methacryloyloxyethyl phosphorylcholine (MPC) has been successfully used as a monomer for preparation of such biomaterials^{371, 372, 374-376} and in dental adhesives.³⁷⁷⁻³⁸¹ Adding MPC was shown to

improve the neutralization performance without the degradation of the adhesive. A secondary strategy for including buffering capacity in the dental adhesive is to add a form of lysine, an amino acid with a weak base side chain. A commercial food preservative and a natural antimicrobial agent,^{382, 383} ϵ -polylysine is added to the dental adhesive primer formulation. Finally, an engineered derivative of an antimicrobial peptide with known activity against *S. mutans*, GH12 was coupled to the dental adhesive system. This work demonstrates that the addition of the ϵ -polylysine and engineered antimicrobial peptide components results in bacterial inhibition near the dental adhesive material. CHX was used as a positive control for the bacterial inhibition experiments in this work. Because of the recognized limitations of CHX in the mouth, this work develops a dental adhesive system with an antimicrobial peptide to provide antibacterial activity.

6.3 Methods

6.3.1 Inhibitory Concentration Assays

The inhibitory assays were conducted by Sheng-Xue Xie. This work follows the standard broth microdilution method, according to the recommendations of the Clinical and Laboratory Standards Institute (CLSI).³⁸⁴⁻³⁸⁶ Briefly, two-fold serial dilutions of compounds are prepared with sterile deionized water to achieve concentrations ranging from 2500 to 4.9 $\mu\text{g/ml}$ in a volume of 20.0 μl . After these dilutions, 80.0 μl of BHI broth and 100.0 μl of bacterial culture containing 1.0×10^6 CFU/ml are added. Thus, each well contains final peptide concentrations ranging from 0.5 to 250 $\mu\text{g/ml}$ in a final volume of 200 μl and a final bacterial concentration of 5.0×10^5 CFU/ml.

For control samples for positive antibacterial activity, microwells containing chlorhexidine digluconate (CHX) with concentrations ranging from 0.04 to 20 $\mu\text{g/ml}$ were also combined with BHI broth and bacterial culture. The blank well contains 20 μl of H_2O and 180 μl of medium without cells. The 96-well plate is incubated in the presence of 5% CO_2 at 37°C overnight. The MIC value is defined as the lowest peptide, CHX and ϵ -polylysine concentration corresponding to total inhibition. The metabolic activity of

bacteria is indicated by alamar blue.³⁸⁷ The activity is tracked by conversion of resazurin to the fluorescent molecule, resorufin. Upon adding bacteria, the reduction of resazurin produces a bright red color.

6.3.2 Circular Dichroism

This method is similar to the CD method previously discussed in Section 5.3.3. The CD spectra are recorded with a CD spectrometer (JASCO, J-815) at room temperature, using a 1.0 mm cuvette. Each peptide sample was dissolved at 0.2 mg/mL in 10 mM potassium phosphate containing 100 mM $(\text{NH}_4)_2\text{SO}_4$ (pH 7.4) with and without TFE at -4°C for overnight. CD spectra were acquired from 185 to 300 nm at a scanning speed of 60 nm/min and were averaged from three runs per each sample. The secondary structure Far-UV CD spectra were processed with the tools of CD Pro.³⁴⁰ The mean residue ellipticity of each sample was converted to mean residue absorbance. The mean residue absorbance was processed with CRDATA.exe to create the input file for SELCON3.exe, CDSSTR.exe and CONTILL.exe. For each set of data, the reference set selected was SMP50. The fractions of secondary structure (Regular Helix, Distorted Helix, Regular Sheet, Distorted Sheet, Turns and Unordered) were averaged over all three CD Pro tools (SELCON3, CDSSTR, and CONTILL). These fractions can be compared to the statistical estimates of across residue positions from the computational structure flexibility method.

6.3.3 Computational Structure Flexibility

Ensembles of structure decoys are generated with PyRosetta tools using a script developed for folding peptides in Appendix B.^{6, 293} Briefly, fragments of lengths of 3 and 9 were inserted from a Robetta database of structures indexed from the Protein Databank. Side chains were positioned according to the Fast Relax Protocol. The percentages of each secondary structure defined in the DSSP program is calculated at each peptide residue for each structure decoy. The combined percentages of the ensemble

are compared with the circular dichroism decomposition values from CD Pro tools to determine the agreement of feature composition for the entire sequence.

6.3.4 Adhesive Disc Preparation

The dental adhesive discs are synthesized and prepared by Linyong Song. Disc specimens are prepared by injecting the resin into a standard aluminum hermetic lid (Tzero®, P/N:901600.901) and covering them with a glass cover (22 mm×30 mm, Fisherfinest®) to prevent oxygen exposure. The specimens are light-cured for 40 s at 23±2 °C with a commercial visible-light-polymerization unit (Spectrum®, Dentsply, Milford, DE) at an intensity of 550 mW/cm². The polymerized disc specimens are stored in the dark at 23±2 °C for at least 48 h. Then, the disc specimens are prewashed by soaking in water for 5 days to remove the unreacted components before antimicrobial testing. The components for the adhesive disc are shown in Table 6.1.

Table 6.1 Commercial analogue components and ϵ -Polylysine Primer components.

Component	Commercial Analogue (wt%)	ϵ -Polylysine (wt%)
HEMA	64	59
TEGDMA	15	15
MPC	5	5
ϵ -Polylysine	/	5
Water	14	14
Photo-Initiators	2	2
Total	100	100

Photo-Initiators: CQ 0.5 wt%, EDMAB 0.5 wt%, and DPIHP 1.0 wt%.

After post-curing, each group primer discs were leached out in a large volume of water (250 ml) for 4 days and monitored by HPLC. The water was changed 3-4 times during the leaching out phase.

6.3.5 Antimicrobial Activity Assays of Adhesive Discs with Peptide Coupling

In-solution antimicrobial activity of peptide-treated resin discs is assessed in this work by Sheng-Xue Xie.

For these studies, each disc is soaked overnight in 100 μ l sterile deionized water, peptide, or CHX solution at 4°C. The soaked discs are rinsed 4 times with 100 μ l of H₂O, with excess H₂O on the disc surface gently removed by blotting with Kimwipes. Each soaked and rinsed disc are carefully transferred to a microwell of a 96-well plate. 20 μ l H₂O, 80 μ l of BHI broth and 100 μ l of BHI broth containing 1×10^6 CFU/ml *S. mutans* UA159 cells are added to each well. The final bacterial concentration in each well is 5.0×10^5 CFU/ml.³⁸⁵ *S. mutans* UA159 bacteria without any disc present serves as the positive control and the negative control was 20 μ l of 63 μ g/ml CHX solution instead of H₂O. The blank well contains 20 μ l of H₂O and 180 μ l of media without cells. The 96-well plate is incubated in the presence of 5% CO₂ at 37°C overnight.

The metabolic activity of bacteria is indicated by alamar blue.³⁸⁷ Samples are incubated for 1-4 hours before fluorescence is taken.

6.4 Results

6.4.1 Inhibitory Concentrations

To incorporate antimicrobial functionality to the dental adhesive system in this work, an antimicrobial peptide with known activity against *S. mutans* was synthesized and characterized by my colleague Sheng-Xue Xie. Table 5.2 shows the GH12 peptide sequence³⁸⁶ and its modifications in this work. In the literature, the MIC value of GH12 with an amide at the C-terminus was around 6.7 µg/ml.³⁸⁶

Table 6.2 Properties of GH12 peptides, modifications and ε-polylysine

Peptide	Sequence	<i>S. mutans</i> IC ₅₀ (µg/ml)	<i>S. mutans</i> MIC (µg/ml)
GH12	GLLWHLHLLH	11.55	31.3
GH12-M1	K_GGGSG_GLLWHLHLLH	17.05	31.3
GH12-M2	GLLWHLHLLH_GSGGG_K	12.75	31.3
ε-Polylysine	(ε-Lysine) _n	71.27	125

GH12-M1 is designed with the spacer domain on the N-terminus side of the antimicrobial peptide.

GH12-M2 is designed with the spacer and a single lysine residue on the C-terminus side. IC₅₀ concentrations of GH12 and GH12-M2 were 11.55 and 12.75 µg/ml, as seen in Table 6.2. The IC₅₀ concentration of GH12-M1 was 17.05 µg/ml, an increase compared to GH12. Due to the lower IC₅₀ concentration, GH12-M2 was selected for further antibacterial studies with the dental adhesive discs.

IC₅₀ values of designed peptides, ε-polylysine, and CHX are shown in Figure 6.2.

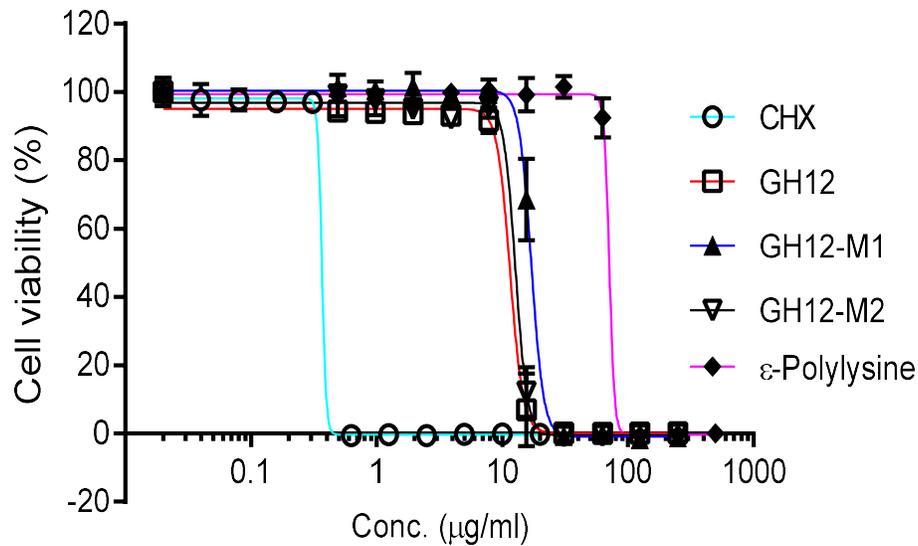


Figure 6.2 Antibacterial activity of CHX, antimicrobial peptides and ϵ -polylysine.

6.4.2 CD Spectra and CD Deconvolution

TFE was used as a kosmotropic agent to simulate ordered water near the bacterial membrane. The circular dichroism spectrum was relatively flat for GH12 without TFE in ammonium sulfate and potassium phosphate buffer at pH 7.4, with a negative peak near 230 nm, as shown in Figure 6.3. The CD spectrum for GH12-M2 was also relatively flat with a negative peak near 235 nm. Adding 10% TFE resulted in flattening the spectra for both peptides while increasing the MRE near 185 nm. Increasing the TFE to 20% showed a large difference in the CD spectrum between the peptides. The GH12-M2 peptide at 20% TFE shows the characteristic large positive peak near 190 nm and negative peaks near 205 and 225 nm, representative of helical secondary structure formation. The GH12 peptide does not show this behavior at 20% TFE, but both peptides show this behavior at 40% TFE and above concentrations.

In Figure 6.3, A) represents the mean residue ellipticity (MRE) of far-UV CD Spectra for GH12 and GH12-M2 with varying levels of 2,2,2-trifluoroethanol (TFE) and B) represents circular dichroism spectra

deconvolution through CD Pro analysis. The percentages given in the figure indicate the amount of TFE in the solution.

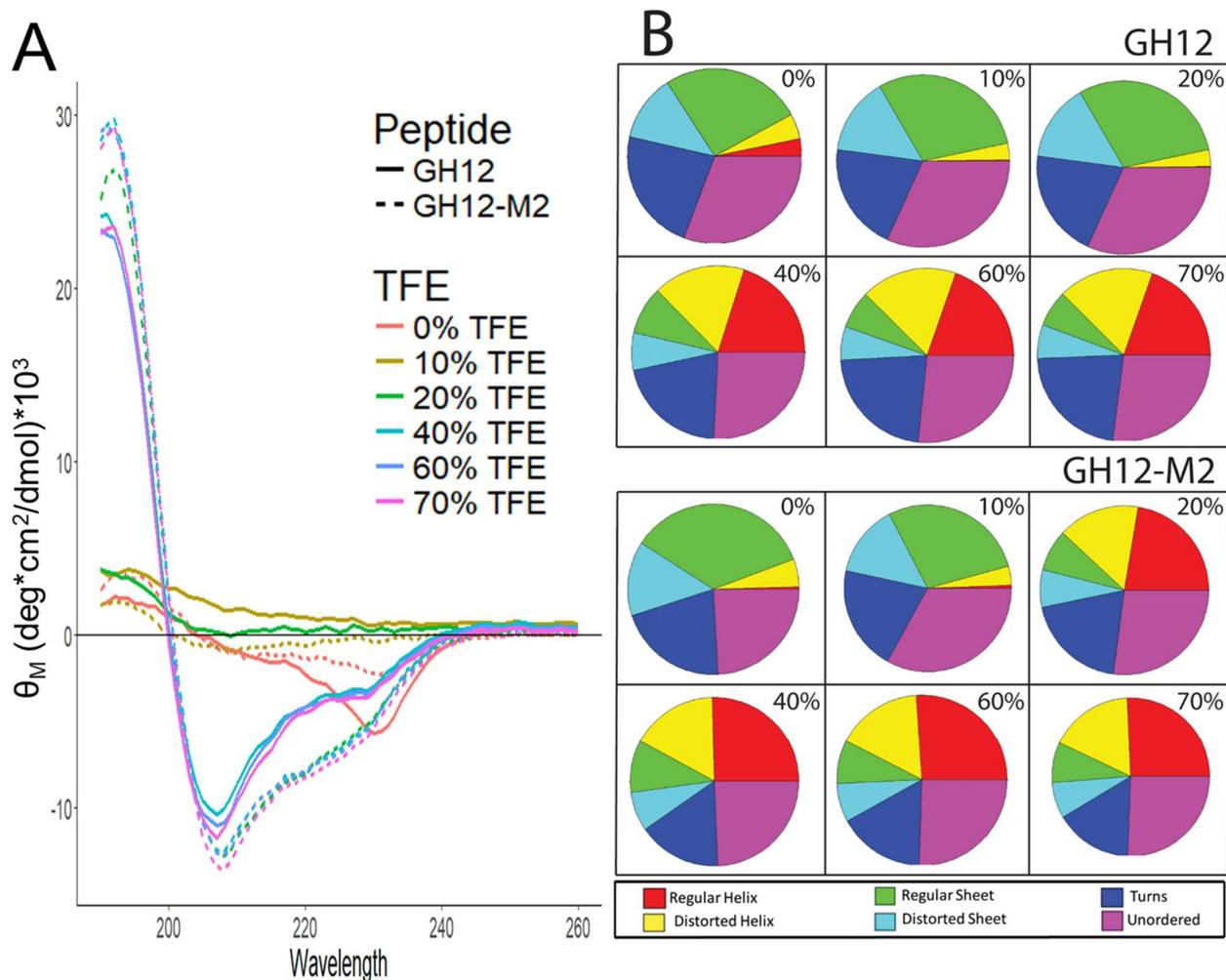


Figure 6.3 Secondary structure characterization by circular dichroism of GH12 and GH12-M2.

With no TFE present, GH12 and GH12-M2 have similar secondary structure fractions. The regular helix structure is slightly larger (3% vs 0%) for the GH12 peptide. With 10% TFE in the buffer solution, the result is a similar distribution of secondary structure fractions with the GH12 reducing the regular helix so that all the categories are within 1%-2%. At 20% TFE, a large difference between the secondary structure fractions for GH12 and GH12-M2 was observed in the original CD spectra results. GH12-M2 has much more regular helix and distorted helix (38%) than GH12 (4%). The regular sheet and distorted

sheet are higher for GH12 (44%) versus GH12-M2 (15%). At high TFE concentrations (>20%), the two peptides stabilize in the fractions of their CD deconvolution. GH12 has a larger turn fraction (21%-23%) and smaller helix fraction (37%-38%) than GH12-M2 for turns (16%-17%) and helix fraction (42%-43%).

6.4.3 Computational Structure Flexibility

In this work, secondary structures of peptide sequences are predicted when generating structure decoys through PyRosetta. In Figure 6.4, the GH12 lowest energy score structure decoy generated via the PyRosetta method is shown in A. GH12 flexibility analysis of percentages of secondary structures occurring in the PyRosetta-generated decoys by residue calculated by DSSP is shown in B. GH12-M2 lowest energy structure is shown in C. GH12-M2 flexibility analysis of secondary structure of decoys by residue using PyRosetta and DSSP are shown in D. The green ribbon indicates the GH12 residues, and the purple ribbon indicates the modification for GH12-M2. Figure 6.4C shows a helix in the polypeptide backbone ribbon for the lowest energy score decoy. The distribution of secondary structure features as a function of sequence residue for GH12 is shown in Figure 6.4A. Darker colors denote higher frequencies of the feature labeled at the right-hand side. The residue positions with the highest occurrence of alpha helix are residues 3-7 (LWHLL). The residues with the highest predicted three-helix are residues 3-5 (LWH). The first residue and the last two residues cannot be labeled with secondary structure features because the definition of the features used by DSSP uses multiple adjacent residues to label structures. Bends (which are not helices) peaked at residues 5, 9 and 10. Turns (which are not bends or helices) peaked at residues 2 and 3. The bottom two rows are defined by the direction in which the polypeptide backbone rotates. For about 80% of the decoys, the leucine at residue 2 is turned right. From residue 3 to 6 about 70% of decoys are turned right, while for residues 7-10 about 60% of the decoys are turned right.

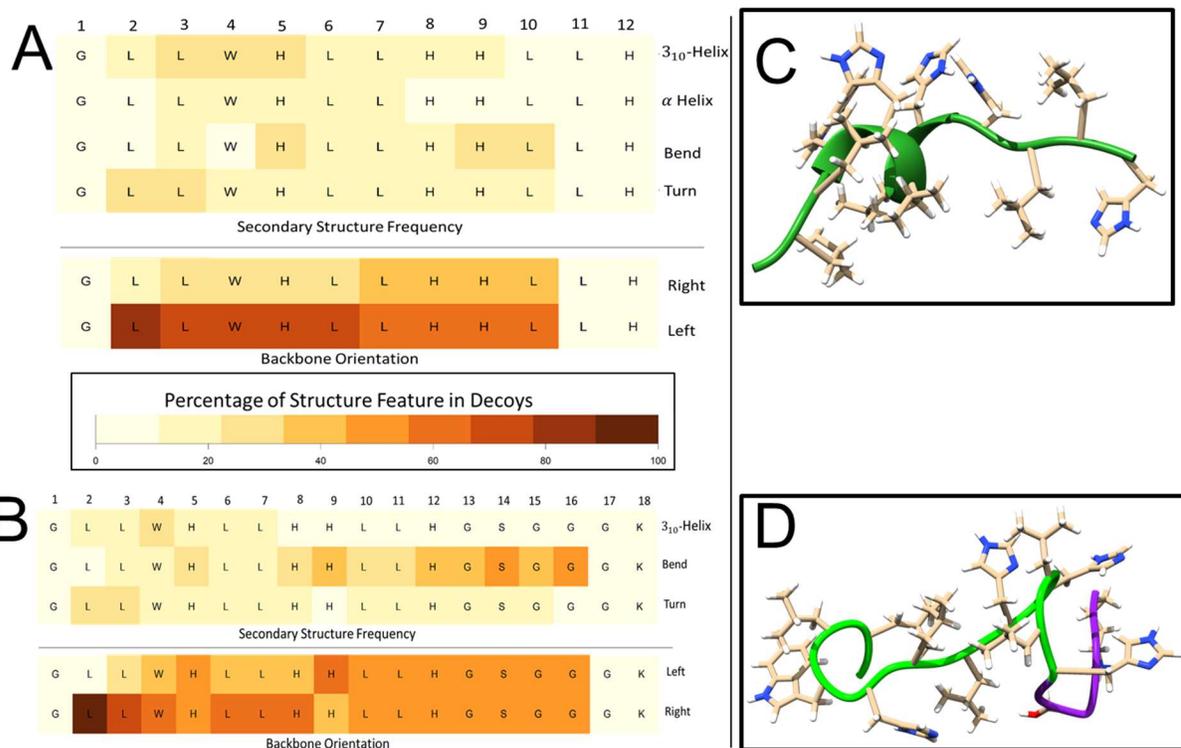


Figure 6.4 Analysis of secondary structure features predicted through PyRosetta method and quantified through DSSP.

Figure 6.4B shows the lowest energy decoy for GH12-M2. No α -helical secondary structure was predicted for this decoy. The M2 spacer (GSGGGK) does induce a change in the flexibility, as seen by comparing residues 2-10 from GH12 with residues 2-10 from GH12-M2. Even though no helix is predicted in the lowest energy score decoy, a 3_{10} -helix is predicted for about 10-20% of decoys for residues 2-7. A structure estimation approach that only finds a few low energy structures may have missed this secondary structural feature.

This work investigates the differences in secondary structural features that occur by adding the spacer sequence on GH12 resulting in GH12-M2 to understand what feature differences lead to similar antibacterial activity for future chimeric peptide design. The percentages of alpha helix drop below 10% for any residue of GH12-M2, while the percentages of alpha helix are above 10% for residues 3-7 for GH12. Flexibility features of bends are predicted in residues 14 and 16 in about 50% of the decoys. The

right and left percentages seem to be like the GH12 percentages, with a 50%/50% split for the spacer residues. These noted structural differences lead to no change in the measured MIC values between GH12 and GH12-M2 in Table 6.2, and a small change in IC₅₀ values.

6.4.4 Antibacterial Activity of Treated Discs

Coupling the GH12-M2 engineered antimicrobial peptide to the dental adhesive discs built in this work showed antibacterial activity against *S. mutans* UA159 if the ϵ -polylysine additive was also included, as shown in Figure 6.5. The minimum effective initial concentration for the peptide to use with the coupling process is between 6.25 and 12.5 times the MIC. Sheng-Xue Xie's coupling process is to soak the disc overnight in a solution of peptide with the indicated concentration. The discs are rinsed with water to remove excess solution on the outside of the disc. The discs are incubated with *S. mutans* bacteria to determine viability.

With ϵ -polylysine in the dental adhesive discs and neither GH12-M2 or CHX, no inhibition was detected, as shown in Figure 6B for 0x concentration. Without ϵ -polylysine, the coupling of the GH12-M2 did not result in detected antibacterial activity against *S. mutans* UA159 using an initial concentration of peptide that was up to 50 times the measured MIC (data not shown). Adding CHX to the dental primer adhesive system resulted in antibacterial activity with or without ϵ -polylysine. The minimum effective initial concentrations of CHX was the same multiplication factor of the MIC as GH12-M2.

In Figure 6.5A, the schematic of the primer adhesive disc samples starting with a commercial analogue. The two test conditions are to add ϵ -polylysine to the disc formulation or to couple GH12-M2 antimicrobial peptide by soaking. When both conditions are met, bacterial inhibition is seen. As a positive control for antimicrobial activity, chlorhexidine gluconate was added at a multiple of its measured MIC value. In Figure 6.5B, *S. mutans* UA159 viability is shown based on alamar blue conversion for dental adhesive discs using the conditions indicated in Figure 6.5A.

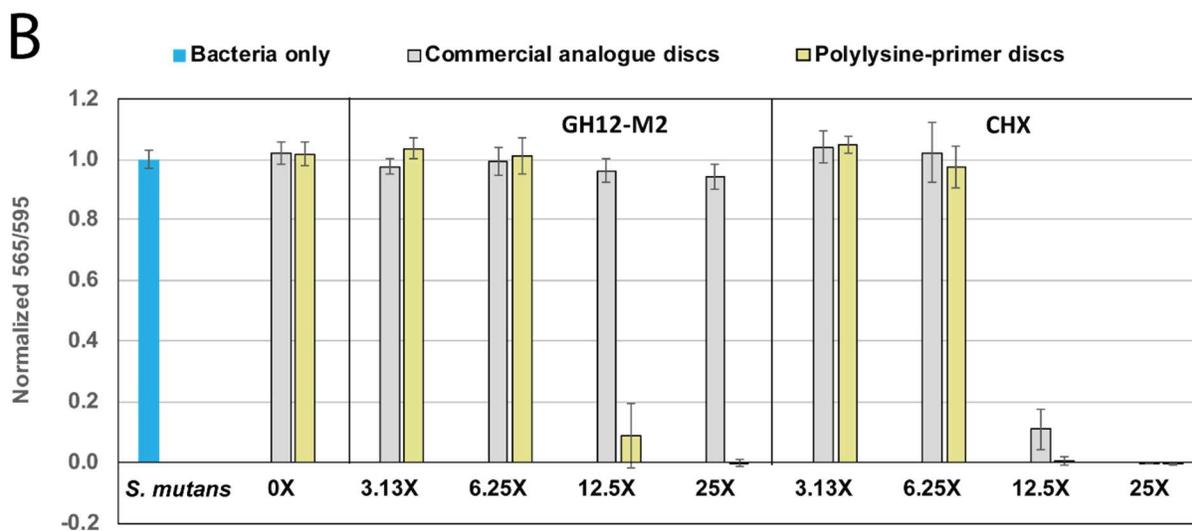
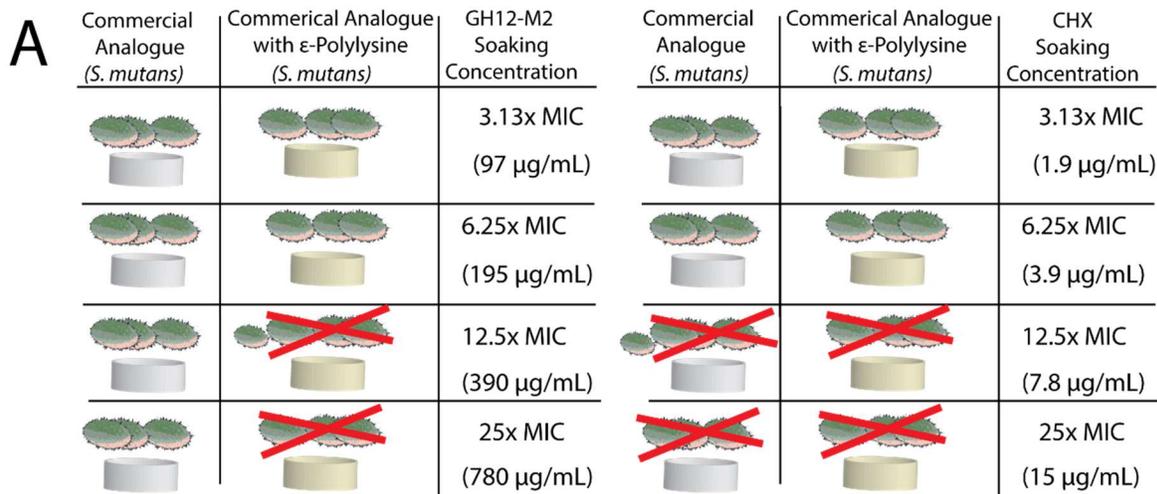


Figure 6.5 Inhibitory activity of GH12-M2 with the ϵ -polylysine added to the disc formulation.

6.5 Discussion

An antimicrobial peptide for the oral environment was engineered for the interface of the adhesive and the remaining tooth structure. GH12 was selected because of its known activity against *S. mutans*, a key bacterium for dental caries. GH12 was also selected for its lack of lysine to aid the conjugation of the antimicrobial peptide to the carboxyl functional group (-COOH) of the primer adhesive. This work shows that the antimicrobial peptide can remain active in this environment with an ϵ -polylysine additive to the dental adhesive material. While further study is needed to confirm why the ϵ -polylysine additive is necessary for the antimicrobial peptide to provide inhibition, the most likely reasons the additive plays

an essential role in the antimicrobial peptide function is change in hydrophathy and the change of porosity of the adhesive discs. The adhesive is more hydrophilic and may have larger pore size when ϵ -polylysine is present. The antimicrobial peptide must have a concentration near the MIC surrounding the adhesive disc to inhibit the growth of bacteria. Since this did not occur without the presence of the ϵ -polylysine, this work concludes that the diffusion of the antimicrobial peptide through the material was too limited to see activity. CHX was effective with the same trend with and without ϵ -polylysine, and thus was able to diffuse through the material to reach its MIC surrounding the adhesive disc. This is evidence for CHX diffusing more easily than GH12-M2 out of the material.

The lack of lysine in GH12 allows for the placement of an added lysine at a terminal of the peptide when adding a spacer sequence. Thus, the conjugation reaction avoids attaching the monomer to the middle of the peptide, likely changing its folding patterns and reducing its activity. The lysine at the terminal is attached to the GH12 peptide through the spacer which was shown in Chapter 4 to improve the folding of a chimeric titanium antimicrobial peptide.²⁹³

The next design choice was which terminal of the peptide to use, the N-terminal or C-terminal. The activity of each of the two peptides was evaluated against *S. mutans* UA159. The N-terminal modification (GH12-M1) has an increased IC_{50} which shows reduced activity compared to the unmodified peptide, as seen in Figure 6.2. The C-terminal modification (GH12-M2) showed a small increase in the IC_{50} .

The difference in flexibility between GH12 and GH12-M2 was evaluated according to the fragment insertion method used by PyRosetta to understand structural difference that lead to similar antibacterial activity. The decoy generation method is stochastic, and therefore cannot predict if the lowest energy structures possible are found. However, the lowest energy structure may differ from the native structure.^{388, 389} The parallel optimization of many decoys at once may also provide insight into the

distribution of the ensemble of structures sampled by peptides in solution. Peptides are not static when floating in solution, but they are quite dynamic angstrom-to-nanometer scale fuzzy masses constantly in motion.³³⁴

The computational folding model and the circular dichroism measurements have conflicting results for the buffer-only measurements but agree as the TFE amount in the solution increases. The computational folding model predicts helical structures, while the deconvolution of the CD spectra by CD Pro predicts mainly beta-sheet structures for buffer-only conditions. However, the deconvolution method predicts large increases in the helical structures with increasing amounts of 2,2,2-trifluoroethanol (TFE) in the solution. TFE is a known kosmotropic agent, increasing the order of the water near the surface of biomolecules.³⁹⁰ TFE has long been used in the crystallization of proteins.^{391, 392} Therefore, increasing the TFE concentration likely puts the peptide into a more ordered, structured state. The computational folding model is biased toward native, ordered structures because it uses fragments from experimentally determined structures, often crystallized structures. Therefore, fragment-insertion structures generated using these kinds of fragments are more likely to be more representative of peptide or protein structures associated with kosmotropic agents such as glucose and TFE than with chaotropic agents such as urea and guanidinium chloride. The water near bacterial surfaces is likely also highly ordered according to the biomolecules on the bacterial membrane.³⁹³ Since the structure generation method developed in this work searches portions of the structural space with ordered fragments, it provides structures which more likely to be relevant for structure-function relationships for how the peptide interacts with the bacterial surface.

6.6 Conclusion

An antimicrobial peptide GH12-M2 was engineered from GH12, an antimicrobial peptide with known activity against *S. mutans*, for future conjugation to a dental adhesive to enhance its antibacterial properties. This work coupled GH12-M2 to a dental adhesive by a soaking method to evaluate the

remaining antibacterial activity. The engineered peptide was structurally characterized to obtain a path for structure-function relationships by measuring folding the peptide with a kosmotropic agent through circular dichroism and by using a computational structure generation folding method biased toward ordered-water states.

Commercial analogue coupled with the engineered peptide not inhibit against *S. mutans* UA159 up to an initial concentration of 1560 $\mu\text{g/ml}$ (50 times MIC). However, when ϵ -polylysine was added to the dental adhesive as a component, the engineered peptide completely inhibited the growth of *S. mutans* for initial concentrations at or above 390 $\mu\text{g/ml}$ (12.5 times MIC). CHX showed a similar trend inhibiting the growth of *S. mutans* with or without ϵ -polylysine in the dental adhesive discs. CHX inhibited the growth of *S. mutans* at 7.8 $\mu\text{g/ml}$ (12.5 times MIC). Because CHX has a variety of limitations for oral use, an antimicrobial peptide (GH12-M2) is engineered for integrating into a dental adhesive.

Antimicrobial peptides, such as GH12, can be integrated into dental adhesives at the interface between the remaining tooth structure and the composite restoration if the peptides can be engineered to be attached to the dental adhesive and remain active. This interface is currently vulnerable to bacterial attack, resulting in shorter clinical lifetimes for composite restorations compared to dental amalgam for posterior restorations. Chapter 4 provides examples of chimeric peptides which can non-covalently attach to titanium surfaces but have reduced activity. Engineering GH12 with principles in Chapters 4 and 5 is expected to result in an antimicrobial peptide that can attach within the dental adhesive and remain active.

In Aim 1, computer-aided molecular design (CAMD) was used to customize antimicrobial peptides for the ease of synthesis. In Chapter 4, chimeric peptides were engineered for infection-free interfaces at titanium implant surfaces. Chimeric antimicrobial peptides were shown as a platform across several materials in Chapter 5. In this chapter, the coupling of an antimicrobial peptide was achieved in a dental

adhesive leading toward conjugating the antimicrobial peptide within the adhesive. Chimeric antimicrobial peptides show promise for creating antibacterial interfaces through solid binding peptides and through conjugation approaches.

This page is intentionally left blank.

7.0 Conclusions and Future Directions

Biomimetics, the study of biological systems to advance engineering principles, recognizes that biological systems utilize nanoscale interfaces between biomolecules and inorganic materials. The structure of biologically-made materials results from orientation control at multiple scales to form nanometer-sized minerals with many levels of hierarchical structure to form bone. In current synthetic materials engineering, high energy costs are required to form materials through melting, such as glass, because of the lack of orientation control, such as in the protein-mediated formation of silicon dioxide spicules in siliceous sponges. In medicine, drugs are often delivered systemically when the active agent is meant for targeted areas, such as in preventing medical implant infections. This project advances the engineered interfaces between active peptides and inorganic materials through building computational analysis tools to understand sequence-function and structure-function relationships to aid in novel peptide design.

The first tool built for this work was a classification method to distinguish active and inactive peptides based on physicochemical boundaries, CLN-MLEM2. This tool is built using the principles of rough set theory. This novel extension combines the polynomial worst-case run time of MLEM2 with the feature number limit of the IRIM method. Comparing to other antimicrobial peptide classification methods, CLN-MLEM2 achieves good selectivity and state-of-the-art specificity. This supervised learning method may be used on other molecule types besides peptides, such as polymers and small molecule drugs.

Because a set theory approach was used in this work to describe antibacterial activity, selective targets of antibacterial activity can be expressed through the grouping of antibacterial activity labels against one bacterial species and inactivity against another bacterial species. For example, if different sets of antimicrobial peptides are effective against *E. coli* than *S. mutans*, such as the bacteriocins these bacteria produce, then the CLN-MLEM2 method can make boundaries to customize new antimicrobial peptides that are selectively active against either *E. coli* or *S. mutans*. As a future direction, the

selectivity approach allows engineering bacterial environments into balanced systems among bacterial strains.

This work has not exhausted the peptide functionalities that can be explored with this approach. In the future, many other different sequence-function relationships can be discovered. A speed improvement of CLN-MLEM2 is to compute the maximum relevance condition in parallel when multiple cores are available. Using the current implementation in Python 2.7, many different parallelization packages, like pp (parallepython.com), can be implemented starting with the code produced for this work.

The next tool built for this work was a codon-based genetic algorithm method to create new sequences that fit the classification boundaries identified with the rough set theory tool also developed for this work. The codon-based genetic algorithm is a single-objective optimization that addresses multiple target properties simultaneously. While this genetic algorithm did not find peptide sequences that met all design targets, the algorithm was effective at finding customized antimicrobial peptides with detectable activity against *S. epidermidis* that were closer to the design targets than any sequence in the initial generation. The genetic algorithm was necessary to find the solutions in this work because in the first trial, the output peptides with seven residues had little sequence similarity with the most fit sequence in the original generation. In the second trial, the output peptides had higher sequence similarity, but only half of the fourteen residues were conserved. For seven non-conserved residues with twenty possible amino acids each from either trial, there are $\sum_{x=1}^7 20^x = 1.35 \times 10^9$ possible sequences, if the residues may also be deleted from the sequence. If which amino acids are conserved is not known, there are 3,432 ways to choose 7 out of 14 amino acids, making the number of possibilities 4.62×10^{12} . If how many residues to conserve is not known before hand, there are $\sum_{x=1}^{14} 20^x = 1.72 \times 10^{18}$ possible sequences. Enumerating 10^{18} sequences, i.e. a million-billion sequences, is infeasible. This analysis also ignores non-canonical amino acids. The target lengths of seven to fifteen amino acids used in Chapter 3 are short sequence examples for peptide design. Longer sequences up to around 50 amino acids

exponentially increase the number of sequences to consider. Peptide design space is too large to enumerate. Metaheuristic approaches, like genetic algorithms are a good fit for this design space.

In future work, more parameter optimization is considered. The selection rate has currently been tested between 10% to 25% of the population of each generation. The number of the sequences in each generation is not further capped. To reduce the computation time of the genetic algorithm, only unique sequences are selected for new generations. The fitness of the sequence is accounted for in the probability of being selected by a recombination operator. To avoid losing fit sequences, sequences from the previous generation are considered for selection along with the newly generated sequences. The increased presence of highly fit sequence is incorporated by using recombination operators to select more fit sequences more often. Mutation rates of up to 50% and crossover rates up to 30% have been explored. The codon-based representation is another technique used to improve the genetic diversity of the solutions considered by the genetic algorithm, which was confirmed to improve the diversity of the solutions in Figure 3.6.

Other metaheuristic methods besides genetic algorithms may also be considered in future work. Tabu search could be implemented with an edit distance function to avoid generating repetitive solutions. Other metaheuristics based on natural metaphors have had some success such as ant colony, particle swarm optimization, and directed evolutionary algorithms. Physics-based metaphors include simulated annealing, stochastic fractal search and other non-linear physics approaches.

Further development of codon-based genetic algorithm would be to incorporate other recombination operators based on genetics concepts, from such as allele dominance and chromosome translocation. These concepts will give new insight of how to make efficient, large changes in sequences while converging to solutions quickly. Evolutionary biology has demonstrated many adaptations of protein and peptide sequences for new purposes. A recent study has shown how apply concept of B-cell evolution in

adaptive immunity to understanding how to direct the evolution of functional sequences.³⁹⁴

Understanding the mechanisms of nucleic acid sequence evolution will lead to more informed peptide and protein design approaches.

Finally, a peptide structure generation script was built for this work by incorporating algorithms from PyRosetta. This integrates with the rough set theory tool to form a method for generating relevant features to discover structure-function relationships. The peptide folding tool built using PyRosetta gives insight into the thermodynamic likelihood of peptide folded structures. This method is biased toward generating ordered structures, which occur more often when kosmotropic agents are present or at certain kinds of material interfaces. In Chapter 4, the chimeric peptide folding change due to the spacer sequence showed that this structure generation method is a useful approach for peptide spacer design. In future work, combining the peptide folding tool with coarse-folding models can enable the improved prediction of larger peptide sequences. Another future direction for this structure generation method would be to validate the folding predictions through empirical information from NMR, X-ray crystallography, CD and other structure determination tools.

This work has begun applying the above tools to the design of an antimicrobial peptide for integration into a dental adhesive. It has demonstrated the successful integration of an antimicrobial peptide, coupled within the dental adhesive. This integration is a novel approach for the development of dental adhesives because it provides inhibitory activity near the dental adhesive without discoloration effects. The appearance of infection is difficult to distinguish from discoloration. To further develop this approach, one could design spacers for antimicrobial peptides conjugated with polymers, using the approaches discussed in Chapters 4 and 5. A genetic algorithm approach to spacer design for antimicrobial peptides would be evaluated without a secondary domain. The candidate spacers with the minimum spacer frequency error could be evaluated for antibacterial activity and compared to the antimicrobial peptide alone.

The structure-function relationships present in biological systems provide clear evidence that peptides and proteins can have well-integrated interfaces with inorganic materials. Hard tissues, such as bones and teeth, are organ-level examples. As the field at the intersection of molecular biology and materials science becomes increasingly able to characterize and to understand how peptides can build novel interfaces for engineering materials, a path can be built toward artificial hierarchical structures with integrated interfaces at multiple levels.

8.0 References

1. Anderson DJ, Podgorny K, Berríos-Torres SI, Bratzler DW, Dellinger EP, Greene L, Nyquist A-C, Saiman L, Yokoe DS, Maragakis LL. Strategies to prevent surgical site infections in acute care hospitals: 2014 update. *Infection Control & Hospital Epidemiology*. 2014;35(06)
2. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am*. 2007;89(4)
3. Anuforom O, Wallace GR, Piddock LV. The immune response and antibacterial therapy. *Med Microbiol Immunol*. 2015;204(2)
4. Yang JH, Bhargava P, Mccloskey D, Mao N, Palsson BO, Collins JJ. Antibiotic-induced changes to the host metabolic environment inhibit drug efficacy and alter immune function. *Cell Host & Microbe*. 2017;22(6)
5. Yazici H, O'neill MB, Kacar T, Wilson BR, Oren EE, Sarikaya M, Tamerler C. Engineered chimeric peptides as antimicrobial surface coating agents toward infection-free implants. *ACS Appl Mater Interfaces*. 2016;8(8)
6. Yucesoy DT, Hnilova M, Boone K, Arnold PM, Snead ML, Tamerler C. Chimeric peptides as implant functionalization agents for titanium alloy implants with antimicrobial properties. *JOM*. 2015;67(4)
7. Roux D, Danilchanka O, Guillard T, Cattoir V, Aschard H, Fu Y, Angoulvant F, Messika J, Ricard JD, Mekalanos JJ, Lory S, Pier GB, Skurnik D. Fitness cost of antibiotic susceptibility during bacterial infection. *Sci Transl Med*. 2015;7(297)
8. Hawkey PM. Multidrug-resistant Gram-negative bacteria: a product of globalization. *J Hosp Infect*. 2015;89(4)
9. Ayhan DH, Tamer YT, Akbar M, Greenberg DE, Toprak E. A synthetic knob for modulating antibiotic resistance. *Biophys J*. 2016;110(3)
10. Porto WF, Pires AS, Franco OL. Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnol Adv*. 2017;35(3)
11. Chernysh S, Gordya N, Suborova T. Insect antimicrobial peptide complexes prevent resistance development in bacteria. *PLoS One*. 2015;10(7)
12. Veltri D, Kamath U, Shehu A. Improving recognition of antimicrobial peptides and target selectivity through machine learning and genetic programming. *IEEE/ACM Trans Comput Biol Bioinform*. 2017;14(2)
13. Peschel A, Sahl HG. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat Rev Microbiol*. 2006;4(7)
14. Li LL, Qi GB, Yu F, Liu SJ, Wang H. An adaptive biointerface from self-assembled functional peptides for tissue engineering. *Adv Mater*. 2015;27(20)
15. Mccloskey AP, Gilmore BF, Lavery G. Evolution of antimicrobial peptides to self-assembled peptides for biomaterial applications. *Pathogens*. 2014;3(4)
16. Chen R, Willcox MD, Ho KK, Smyth D, Kumar N. Antimicrobial peptide melimine coating for titanium and its in vivo antibacterial activity in rodent subcutaneous infection models. *Biomaterials*. 2016;85
17. Ozcelik B, Chen R, Glattauer V, Kumar N, Willcox MP, Thissen H. Crosslinked platform coatings incorporating bioactive signals for the control of biointerfacial interactions. *Macromol Biosci*. 2017;17(4)
18. Yu K, Lo JC, Yan M, Yang X, Brooks DE, Hancock RE, Lange D, Kizhakkedathu JN. Anti-adhesive antimicrobial peptide coating prevents catheter associated infection in a mouse urinary infection model. *Biomaterials*. 2017;116
19. Raphael J, Holodniy M, Goodman SB, Heilshorn SC. Multifunctional coatings to simultaneously promote osseointegration and prevent infection of orthopaedic implants. *Biomaterials*. 2016;84
20. Bosco R, Van Den Beucken J, Leeuwenburgh S, Jansen J. Surface engineering for bone implants: A trend from passive to active surfaces. *Coatings*. 2012;2(3)

21. Nejadnik MR, Yang X, Bongio M, Alghamdi HS, Van Den Beucken JJ, Huysmans MC, Jansen JA, Hilborn J, Ossipov D, Leeuwenburgh SC. Self-healing hybrid nanocomposites consisting of bisphosphonated hyaluronan and calcium phosphate nanoparticles. *Biomaterials*. 2014;35(25)
22. Heinz H, Pramanik C, Heinz O, Ding Y, Mishra RK, Marchon D, Flatt RJ, Estrela-Lopis I, Llop J, Moya S, Ziolo RF. Nanoparticle decoration with surfactants: Molecular interactions, assembly, and applications. *Surface Science Reports*. 2017
23. Gimenez-Marques M, Hidalgo T, Serre C, Horcajada P. Nanostructured metal-organic frameworks and their bio-related applications. *Coordination Chem Rev*. 2016;307
24. Yang X, Yang M, Pang B, Vara M, Xia Y. Gold nanomaterials at work in biomedicine. *Chem Rev*. 2015;115(19)
25. Wang Q, Wei S, Wu J, Zou X, Sieggreen O, Liu Y, Xi C, Brooks CL, Chen Z. Interfacial behaviors of antimicrobial peptide Cecropin P1 immobilized on different self-assembled monolayers. *The Journal of Physical Chemistry C*. 2015;119(39)
26. Costa F, Carvalho IF, Montelaro RC, Gomes P, Martins MC. Covalent immobilization of antimicrobial peptides (AMPs) onto biomaterial surfaces. *Acta Biomater*. 2011;7(4)
27. Simchi A, Tamjid E, Pishbin F, Boccaccini AR. Recent progress in inorganic and composite coatings with bactericidal capability for orthopaedic applications. *Nanomedicine*. 2011;7(1)
28. Popat KC, Eltgroth M, Latempa TJ, Grimes CA, Desai TA. Titania nanotubes: a novel platform for drug-eluting coatings for medical implants? *Small*. 2007;3(11)
29. Ma M, Kazemzadeh-Narbat M, Hui Y, Lu S, Ding C, Chen DD, Hancock RE, Wang R. Local delivery of antimicrobial peptides using self-organized TiO₂ nanotube arrays for peri-implant infections. *J Biomed Mater Res A*. 2012;100(2)
30. Sarikaya M, Tamerler C, Schwartz DT, Baneyx F. Materials assembly and formation using engineered polypeptides. *Annual Review of Materials Research*. 2004;34(1)
31. Tamerler C, Khatayevich D, Gungormus M, Kacar T, Oren EE, Hnilova M, Sarikaya M. Molecular biomimetics: GEPI-based biological routes to technology. *Biopolymers*. 2010;94(1)
32. Sarikaya M, Tamerler C, Jen AKY, Schulten K, Baneyx F. Molecular biomimetics: nanotechnology through biology. *Nat Mater*. 2003;2(9)
33. Care A, Bergquist PL, Sunna A. Solid-binding peptides: smart tools for nanobiotechnology. *Trends Biotechnol*. 2015;33(5)
34. Seker UO, Wilson B, Kulp JL, Evans JS, Tamerler C, Sarikaya M. Thermodynamics of engineered gold binding peptides: establishing the structure-activity relationships. *Biomacromolecules*. 2014;15(7)
35. Corni S, Hnilova M, Tamerler C, Sarikaya M. Conformational behavior of genetically-engineered dodecapeptides as a determinant of binding affinity for gold. *Journal of Physical Chemistry C*. 2013;117(33)
36. Oren EE, Notman R, Kim IW, Evans JS, Walsh TR, Samudrala R, Tamerler C, Sarikaya M. Probing the molecular mechanisms of quartz-binding peptides. *Langmuir*. 2010;26(13)
37. Notman R, Oren EE, Tamerler C, Sarikaya M, Samudrala R, Walsh TR. Solution study of engineered quartz binding peptides using replica exchange molecular dynamics. *Biomacromolecules*. 2010;11
38. Oren EE, Tamerler C, Sahin D, Hnilova M, Seker UO, Sarikaya M, Samudrala R. A novel knowledge-based approach to design inorganic-binding peptides. *Bioinformatics*. 2007;23(21)
39. Adams BL, Finch AS, Hurley MM, Sarkes DA, Stratis-Cullum DN. Genetically engineered peptides for inorganics: study of an unconstrained bacterial display technology and bulk aluminum alloy. *Adv Mater*. 2013;25(33)
40. Briggs BD, Knecht MR. Nanotechnology meets biology: Peptide-based methods for the fabrication of functional materials. *Journal of Physical Chemistry Letters*. 2012;3(3)

41. Zhou Y, Snead ML, Tamerler C. Bio-inspired hard-to-soft interface for implant integration to bone. *Nanomedicine*. 2015;11(2)
42. Hassert R, Beck-Sickinger AG. Tuning peptide affinity for biofunctionalized surfaces. *Eur J Pharm Biopharm*. 2013;85(1)
43. Salou L, Hoornaert A, Louarn G, Layrolle P. Enhanced osseointegration of titanium implants with nanostructured surfaces: An experimental study in rabbits. *Acta Biomater*. 2015;11
44. Philippart A, Boccaccini AR, Fleck C, Schubert DW, Roether JA. Toughening and functionalization of bioactive ceramic and glass bone scaffolds by biopolymer coatings and infiltration: A review of the last 5 years. *Expert Rev Med Devices*. 2015;12(1)
45. Jiang S, Pan H, Chen Y, Xu X, Tang R. Amorphous calcium phosphate phase-mediated crystal nucleation kinetics and pathway. *Faraday Discuss*. 2015;179(0)
46. Santander S, Alcaine C, Lyahyai J, Perez MA, Rodellar C, Doblare M, Ochoa I. In vitro osteoinduction of human mesenchymal stem cells in biomimetic surface modified titanium alloy implants. *Dent Mater J*. 2014;33(3)
47. Surmenev RA, Surmeneva MA, Ivanova AA. Significance of calcium phosphate coatings for the enhancement of new bone osteogenesis--A review. *Acta Biomater*. 2014;10(2)
48. Yazici H, Fong H, Wilson B, Oren EE, Amos FA, Zhang H, Evans JS, Snead ML, Sarikaya M, Tamerler C. Biological response on a titanium implant-grade surface functionalized with modular peptides. *Acta Biomater*. 2013;9(2)
49. Tamerler C, Sarikaya M. Genetically designed peptide-based molecular materials. *ACS Nano*. 2009;3(7)
50. Sarikaya M, Tamerler C. Biohesion: Coupling synthetic and biological entities using genetically engineered peptides. *Abstracts of Papers of the American Chemical Society*. 2009;237
51. Kacar T, Ray J, Gungormus M, Oren EE, Tamerler C, Sarikaya M. Quartz binding peptides as molecular linkers for co-assembling nanoentities on multifunctional micropatterned substrates. *Abstracts of Papers of the American Chemical Society*. 2009;237
52. Zimmerman JM, Eliezer N, Simha R. The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol*. 1968;21(2)
53. Dawson D, Brock D, Mayo O. *The biochemical genetics of man*. Brock, DJH, Mayo, O(eds). 1972
54. Garel JP, Filliol D, Mandel P. Partition coefficients of amino acids, nucleobases, nucleosides and nucleotides in a saline solvent system. *J Chromatogr*. 1973;78(2)
55. Renfrew PD, Choi EJ, Bonneau R, Kuhlman B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS One*. 2012;7(3)
56. Ullman JD. NP-complete scheduling problems. *Journal of Computer and System sciences*. 1975;10(3)
57. Cocco S, Monasson R. Trajectories in phase diagrams, growth processes, and computational complexity: How search algorithms solve the 3-satisfiability problem. *Physical review letters*. 2001;86(8)
58. Thangavel K, Pethalakshmi A. Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing*. 2009;9(1)
59. Yao Y, Wong SM, Lin TY. *A review of rough set models*. Rough sets and data mining: Springer; 1997.
60. Grzymala-Busse JW, Rzasza W. A local version of the MLEM2 algorithm for rule induction. *Fundamenta Informaticae*. 2010;100(1-4)
61. Swiniarski RW, Skowron A. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*. 2003;24(6)
62. Lindsey TS. *Interesting rule induction module: Adding support for unknown attribute values*: University of Kansas; 2016.

63. Lata S, Mishra NK, Raghava GP. AntiBP2: improved version of antibacterial peptide prediction. *BMC Bioinformatics*. 2010;11 Suppl 1(1)
64. Loose C, Jensen K, Rigoutsos I, Stephanopoulos G. A linguistic model for the rational design of antimicrobial peptides. *Nature*. 2006;443(7113)
65. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S. CAMP: A useful resource for research on antimicrobial peptides. *Nucleic Acids Res*. 2010;38(Database issue)
66. Waghu FH, Gopi L, Barai RS, Ramteke P, Nizami B, Idicula-Thomas S. CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res*. 2014;42(Database issue)
67. Xiao X, Wang P, Lin W-Z, Jia J-H, Chou K-C. iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal Biochem*. 2013;436(2)
68. Veltri D, Kamath U, Shehu A, editors. A novel method to improve recognition of antimicrobial peptides through distal sequence-based features. *IEEE International Conference on Bioinformatics and Biomedicine; 2014 2-5 Nov. 2014*.
69. Mitchell AR. Bruce Merrifield and solid-phase peptide synthesis: A historical assessment. *Biopolymers*. 2008;90(3)
70. Fink J, Merrifield RB, Boman A, Boman HG. The chemical synthesis of Cecropin D and an analog with enhanced antibacterial activity. *Journal of Biological Chemistry*. 1989;264(11)
71. Andreu D, Merrifield RB, Steiner H, Boman HG. Solid-phase synthesis of Cecropin A and related peptides. *Proc Natl Acad Sci U S A*. 1983;80(21)
72. Merrifield RB, Vizioli LD, Boman HG. Synthesis of the antibacterial peptide Cecropin A (1-33). *Biochemistry*. 1982;21(20)
73. Cantisani M, Finamore E, Mignogna E, Falanga A, Nicoletti GF, Pedone C, Morelli G, Leone M, Galdiero M, Galdiero S. Structural insights into and activity analysis of the antimicrobial peptide myxinidin. *Antimicrob Agents Chemother*. 2014;58(9)
74. Hansch C, Fujita T. Rho-Sigma-Pi analysis, method for correlation of biological activity and chemical structure. *J Am Chem Soc*. 1964;86(8)
75. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin, Ii, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A. QSAR modeling: where have you been? Where are you going to? *J Med Chem*. 2014;57(12)
76. Gani R, Nielsen B, Fredenslund A. A group contribution approach to computer-aided molecular design. *AIChE Journal*. 1991;37(9)
77. Sundaram A, Venkatasubramanian V. Parametric sensitivity and search-space characterization studies of genetic algorithms for computer-aided polymer design. *Journal of Chemical Information and Computer Sciences*. 1998;38(6)
78. Lin B, Chavali S, Camarda K, Miller DC. Computer-aided molecular design using Tabu search. *Computers & Chemical Engineering*. 2005;29(2)
79. Boone K, Abedin F, Anwar MR, Camarda KV. Molecular design in the pharmaceutical industries. In: Martín M, Eden MR, Chemmangattuvalappil NG, editors. *Tools for Chemical Product Design. Computer Aided Chemical Engineering*. 39: Elsevier; 2016.
80. 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)
81. Chaudhury S, Lyskov S, Gray JJ. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*. 2010;26(5)
82. Rios LM, Sahinidis NV. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*. 2013;56(3)
83. Dennis J, Torczon V, editors. *Derivative-free pattern search methods for multidisciplinary design problems. 5th Symposium on Multidisciplinary Analysis and Optimization; 1994*.

84. Hooke R, Jeeves TA. "Direct Search" solution of numerical and statistical problems. *J of ACM*. 1961;8(2)
85. Nelder JA, Mead R. A Simplex-Method for Function Minimization. *Computer Journal*. 1965;7(4)
86. Holland J. *Adaptation in natural and artificial systems: an introductory analysis with application to biology*. Control and artificial intelligence. 1975
87. Kirkpatrick S, Gelatt CD, Jr., Vecchi MP. Optimization by simulated annealing. *Science*. 1983;220(4598)
88. Glover F, Laguna M. *Tabu search*. Handbook of combinatorial optimization: Springer; 1998.
89. Fjell CD, Jenssen H, Cheung WA, Hancock RE, Cherkasov A. Optimization of antibacterial peptides by genetic algorithms and cheminformatics. *Chem Biol Drug Des*. 2011;77(1)
90. Eiben AE, Hinterding R, Michalewicz Z. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*. 1999;3(2)
91. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 1997;267(3)
92. Lazar GA, Desjarlais JR, Handel TM. De novo design of the hydrophobic core of ubiquitin. *Protein Science*. 1997;6(6)
93. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*. 2009;77(4)
94. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. *Protein Science*. 1995;4(10)
95. Yeh CT, Brunette TJ, Baker D, McIntosh-Smith S, Parmeggiani F. Elfin: An algorithm for the computational design of custom three-dimensional structures from modular repeat protein building blocks. *J Struct Biol*. 2018;201(2)
96. Movahedi M, Zare-Mirakabad F, Arab SS. Evaluating the accuracy of protein design using native secondary sub-structures. *BMC Bioinformatics*. 2016;17(1)
97. Brunk E, Perez MaS, Athri P, Rothlisberger U. Genetic-algorithm-based optimization of a peptidic scaffold for sequestration and hydration of CO₂. *Chemphyschem*. 2016;17(23)
98. Jeschke G. Ensemble models of proteins and protein domains based on distance distribution restraints. *Proteins*. 2016;84(4)
99. Tompa P, Varadi M. Predicting the predictive power of IDP ensembles. *Structure*. 2014;22(2)
100. Luo Y, Ma B, Nussinov R, Wei G. Structural insight into tau protein's paradox of intrinsically disordered behavior, self-acetylation activity, and aggregation. *Journal of Physical Chemistry Letters*. 2014;5(17)
101. Uversky VN. Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta*. 2013;1834(5)
102. Levine ZA, Larini L, Lapointe NE, Feinstein SC, Shea JE. Regulation and aggregation of intrinsically disordered peptides. *Proc Natl Acad Sci U S A*. 2015;112(9)
103. Uversky VN. A protein-chameleon: Conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J Biomol Struct Dyn*. 2003;21(2)
104. Thévenet P, Rey J, Moroy G, Tuffery P. *De novo peptide structure prediction: An overview*. Computational Peptidology: Springer; 2015.
105. Patwardhan SV, Emami FS, Berry RJ, Jones SE, Naik RR, Deschaume O, Heinz H, Perry CC. Chemistry of aqueous silica nanoparticle surfaces and the mechanism of selective peptide adsorption. *J Am Chem Soc*. 2012;134(14)
106. Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, Van Gunsteren WF. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*. 2005;26(16)

107. Lin TJ, Heinz H. Accurate force field parameters and pH resolved surface models for hydroxyapatite to understand structure, mechanics, hydration, and biological interfaces. *Journal of Physical Chemistry C*. 2016;120(9)
108. Briggs BD, Bedford NM, Seifert S, Koerner H, Ramezani-Dakhel H, Heinz H, Naik RR, Frenkel AI, Knecht MR. Atomic-scale identification of Pd leaching in nanoparticle catalyzed C-C coupling: Effects of particle surface disorder. *Chem Sci*. 2015;6(11)
109. Shen Y, Maupetit J, Derreumaux P, Tuffery P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J Chem Theory Comput*. 2014;10(10)
110. 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)
111. Tsubaki M, Srivastava RB, Yu NT. Resonance Raman investigation of carbon monoxide bonding in (carbon monoxy) hemoglobin and-myoglobin: detection of iron-carbon monoxide stretching and iron-carbon-oxygen bending vibrations and influence of the quaternary structure change. *Biochemistry*. 1982;21(6)
112. Lukin JA, Kontaxis G, Simplaceanu V, Yuan Y, Bax A, Ho C. Quaternary structure of hemoglobin in solution. *Proc Natl Acad Sci U S A*. 2003;100(2)
113. Silva MM, Rogers PH, Arnone A. A third quaternary structure of human hemoglobin A at 1.7-A resolution. *Journal of Biological Chemistry*. 1992;267(24)
114. Nussinov R, Tsai CJ. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol*. 2015;30
115. Lindert S, Meiler J, Mccammon JA. Iterative molecular dynamics-Rosetta protein structure refinement protocol to improve model quality. *J Chem Theory Comput*. 2013;9(8)
116. Simoncini D, Berenger F, Shrestha R, Zhang KY. A probabilistic fragment-based protein structure prediction algorithm. *PLoS One*. 2012;7(7)
117. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol*. 2011;487
118. Khoury GA, Liwo A, Khatib F, Zhou H, Chopra G, Bacardit J, Bortot LO, Faccioli RA, Deng X, He Y, Krupa P, Li J, Mozolewska MA, Sieradzan AK, Smadbeck J, Wirecki T, Cooper S, Flatten J, Xu K, Baker D, Cheng J, Delbem AC, Floudas CA, Keasar C, Levitt M, Popovic Z, Scheraga HA, Skolnick J, Crivelli SN, Foldit P. WeFold: a competition for protein structure prediction. *Proteins*. 2014;82(9)
119. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popovic Z, Baker D, Players F. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A*. 2011;108(47)
120. Rodriguez Sanchez F, Rodriguez Andres C, Arteagoitia I. Which antibiotic regimen prevents implant failure or infection after dental implant surgery? A systematic review and meta-analysis. *J Craniomaxillofac Surg*. 2018;46(4)
121. Wang JT, Wang AY, Psarros C, Da Cruz M. Rates of revision and device failure in cochlear implant surgery: a 30-year experience. *Laryngoscope*. 2014;124(10)
122. Wang DA, Varghese S, Sharma B, Strehin I, Fermanian S, Gorham J, Fairbrother DH, Cascio B, Elisseeff JH. Multifunctional chondroitin sulphate for cartilage tissue-biomaterial integration. *Nat Mater*. 2007;6(5)
123. Sadoghi P, Liebensteiner M, Agreiter M, Leithner A, Bohler N, Labek G. Revision surgery after total joint arthroplasty: A complication-based analysis using worldwide arthroplasty registers. *J Arthroplasty*. 2013;28(8)
124. Aminov R. History of antimicrobial drug discovery: Major classes and health impact. *Biochem Pharmacol*. 2017;133

125. Belanger CR, Mansour SC, Pletzer D, Hancock REW. Alternative strategies for the study and treatment of clinical bacterial biofilms. *Emerging Topics in Life Sciences*. 2017
126. Bechinger B, Gorr SU. Antimicrobial peptides: Mechanisms of action and resistance. *J Dent Res*. 2017;96(3)
127. Pawlak Z. Rough set theory and its applications to data analysis. *Cybernetics and Systems*. 1998;29(7)
128. Clark PG, Grzymala-Busse JW, Rzasa W. A comparison of two MLEM2 rule induction algorithms extended to probabilistic approximations. *Journal of Intelligent Information Systems*. 2015
129. Cohagan C, Grzymala-Busse JW, Hippe ZS, editors. Mining inconsistent data with the bagged MLEM2 rule induction algorithm. *IEEE Conference on Granular Computing*; 2010: IEEE.
130. Ventola CL. The antibiotic resistance crisis: Part 1: Causes and threats. *P T*. 2015;40(4)
131. Mishra B, Reiling S, Zarena D, Wang G. Host defense antimicrobial peptides as antibiotics: Design and application strategies. *Curr Opin Chem Biol*. 2017;38
132. O'neill J. Tackling drug-resistant infections globally: final report and recommendations. *Review on Antimicrobial Resistance 2016*. 2016
133. Al-Tawfiq JA, Laxminarayan R, Mendelson M. How should we respond to the emergence of plasmid-mediated colistin resistance in humans and animals? *Int J Infect Dis*. 2017;54
134. Fan L, Sun J, Zhou M, Zhou J, Lao X, Zheng H, Xu H. DRAMP: a comprehensive data repository of antimicrobial peptides. *Sci Rep*. 2016;6
135. Di Luca M, Maccari G, Maisetta G, Batoni G. BaAMPs: the database of biofilm-active antimicrobial peptides. *Biofouling*. 2015;31(2)
136. Wang G, Li X, Wang Z. APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res*. 2016;44(D1)
137. Zhao X, Wu H, Lu H, Li G, Huang Q. LAMP: A database linking antimicrobial peptides. *PLoS One*. 2013;8(6)
138. Wang G, Li X, Wang Z. APD2: The updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res*. 2009;37(Database issue)
139. Wang J, Dong X, Yu Q, Baker SN, Li H, Larm NE, Baker GA, Chen L, Tan J, Chen M. Incorporation of antibacterial agent derived deep eutectic solvent into an active dental composite. *Dent Mater*. 2017;33(12)
140. Chen Y, Mant CT, Farmer SW, Hancock RE, Vasil ML, Hodges RS. Rational design of alpha-helical antimicrobial peptides with enhanced activities and specificity/therapeutic index. *Journal of Biological Chemistry*. 2005;280(13)
141. Jodoin J, Hincke MT. Histone H5 is a potent antimicrobial agent and a template for novel antimicrobial peptides. *Sci Rep*. 2018;8(1)
142. Tajbakhsh M, Karimi A, Tohidpour A, Abbasi N, Fallah F, Akhavan MM. The antimicrobial potential of a new derivative of cathelicidin from *Bungarus fasciatus* against methicillin-resistant *Staphylococcus aureus*. *J Microbiol*. 2018;56(2)
143. Vasudev PG, Chatterjee S, Shamala N, Balaram P. Structural chemistry of peptides containing backbone expanded amino acid residues: Conformational features of beta, gamma, and hybrid peptides. *Chem Rev*. 2011;111(2)
144. Sang P, Shi Y, Teng P, Cao A, Xu H, Li Q, Cai J. Antimicrobial A-peptides. *Curr Top Med Chem*. 2017;17(11)
145. Seebach D, Beck AK, Bierbaum DJ. The world of beta- and gamma-peptides comprised of homologated proteinogenic amino acids and other components. *Chem Biodivers*. 2004;1(8)
146. Porter EA, Weisblum B, Gellman SH. Mimicry of host-defense peptides by unnatural oligomers: antimicrobial beta-peptides. *J Am Chem Soc*. 2002;124(25)

147. Knerr PJ, Van Der Donk WA. Discovery, biosynthesis, and engineering of lantipeptides. In: Kornberg RD, editor. Annual review of biochemistry. Vol 81,2012.
148. Brogden NK, Brogden KA. Will new generations of modified antimicrobial peptides improve their potential as pharmaceuticals? *Int J Antimicrob Agents*. 2011;38(3)
149. Candido-Ferreira IL, Kronenberger T, Sayegh RSR, Batista IDC, Da Silva PI. Evidence of an antimicrobial peptide signature encrypted in HECT E3 ubiquitin ligases. *Frontiers in Immunology*. 2017;7
150. Boone K, Abedin F, Anwar MR, Camarda KV. Molecular design in the pharmaceutical industries. *Computer Aided Chemical Engineering*. 2017;39
151. Ng LY, Chong FK, Chemmangattuvalappil NG. Challenges and opportunities in computer-aided molecular design. *Computers & Chemical Engineering*. 2015;81
152. Roughton BC, Christian B, White J, Camarda KV, Gani R. Simultaneous design of ionic liquid entrainers and energy efficient azeotropic separation processes. *Computers & Chemical Engineering*. 2012;42
153. Riera-Fernandez P, Martin-Romalde R, Prado-Prado FJ, Escobar M, Munteanu CR, Concu R, Duardo-Sanchez A, Gonzalez-Diaz H. From QSAR models of drugs to complex networks: State-of-art review and introduction of new Markov-spectral moments indices. *Curr Top Med Chem*. 2012;12(8)
154. Prado-Prado FJ, Uriarte E, Borges F, Gonzalez-Diaz H. Multi-target spectral moments for QSAR and Complex Networks study of antibacterial drugs. *Eur J Med Chem*. 2009;44(11)
155. Du Q-S, Huang R-B, Chou K-C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Current Protein and Peptide Science*. 2008;9(3)
156. Fjell CD, Hiss JA, Hancock RE, Schneider G. Designing antimicrobial peptides: Form follows function. *Nat Rev Drug Discov*. 2011;11(1)
157. Cherkasov A, Hilpert K, Jenssen H, Fjell CD, Waldbrook M, Mullaly SC, Volkmer R, Hancock RE. Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS Chem Biol*. 2009;4(1)
158. Claro B, Bastos M, Garcia-Fandino R. Design and applications of cyclic peptides. *Peptide Applications in Biomedicine, Biotechnology and Bioengineering*: Elsevier; 2018.
159. Muller AT, Kaymaz AC, Gabernet G, Posselt G, Wessler S, Hiss JA, Schneider G. Sparse neural network models of antimicrobial peptide-activity relationships. *Mol Inform*. 2016;35(11-12)
160. Gabere MN, Noble WS. Empirical comparison of web-based antimicrobial peptide prediction tools. *Bioinformatics*. 2017;33(13)
161. Lata S, Sharma BK, Raghava GP. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics*. 2007;8
162. Bhasin M, Raghava GP. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics*. 2004;20(3)
163. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*. 2004;22(23-24)
164. Saha S, Raghava G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*. 2006;65(1)
165. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab-An S4 package for kernel methods in R. *Journal of Statistical Software*. 2004;11(9)
166. Venerables W, Ripley B. *Modern applied statistics with S*. Springer; 2002.
167. Norušis MJ. *SPSS/PC+ advanced statistics V2. 0: for the IBM PC/XT/AT and PS/2: SPSS Incorporated*; 1988.
168. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3)
169. Magrane M, Uniprot C. UniProt Knowledgebase: A hub of integrated protein data. *Database (Oxford)*. 2011;2011

170. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'donovan C, Redaschi N, Yeh LS. The universal protein resource (UniProt). *Nucleic Acids Res.* 2005;33(Database issue)
171. Fernandes FC, Rigden DJ, Franco OL. Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. *Biopolymers.* 2012;98(4)
172. Charif D, Lobry JR. SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Structural approaches to sequence evolution: Springer; 2007.*
173. Yu D, Sheng Z, Xu X, Li J, Yang H, Liu Z, Rees HH, Lai R. A novel antimicrobial peptide from salivary glands of the hard tick, *Ixodes sinensis*. *Peptides.* 2006;27(1)
174. Wang G, Watson KM, Peterkofsky A, Buckheit RW, Jr. Identification of novel human immunodeficiency virus type 1-inhibitory peptides based on the antimicrobial peptide database. *Antimicrob Agents Chemother.* 2010;54(3)
175. Menousek J, Mishra B, Hanke ML, Heim CE, Kielian T, Wang G. Database screening and in vivo efficacy of antimicrobial peptides against methicillin-resistant *Staphylococcus aureus* USA300. *Int J Antimicrob Agents.* 2012;39(5)
176. Grzymala-Busse JW, Rzasa W. A Local Version of the MLEM2 Algorithm for Rule Induction. *Fundamenta Informaticae.* 2010;100(1)
177. Clark PG, Gao C, Grzymala-Busse JW, editors. Complexity of rule sets induced by two versions of the MLEM2 rule induction algorithm. *International Conference on Artificial Intelligence and Soft Computing; 2017: Springer.*
178. Grzymala-Busse JW, Hamilton J, Hippe ZS. Diagnosis of melanoma using IRIM, a data mining system. In: Rutkowski L, Siekmann JH, Tadeusiewicz R, Zadeh LA, editors. *International Conference on Artificial Intelligence and Soft Computing.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2004.
179. Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chemical Engineering Research and Design.* 2016;116
180. Baldi P, Brunak S, Chauvin Y, Andersen CA, Nielsen H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics.* 2000;16(5)
181. 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)
182. Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res.* 2000;28(1)
183. Aurora R, Rose GD. Helix capping. *Protein Science.* 1998;7(1)
184. Charton M, Charton BI. The structural dependence of amino acid hydrophobicity parameters. *J Theor Biol.* 1982;99(4)
185. Fauchere JL, Charton M, Kier LB, Verloop A, Pliska V. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int J Pept Protein Res.* 1988;32(4)
186. Finkelstein AV, Badretdinov AY, Ptitsyn OB. Physical reasons for secondary structure stability: alpha-helices in short peptides. *Proteins.* 1991;10(4)
187. Geisow MJ, Roberts RD. Amino acid preferences for secondary structure vary with protein class. *International Journal of Biological Macromolecules.* 1980;2(6)
188. George RA, Heringa J. SnapDRAGON: a method to delineate protein structural domains from sequence data. *J of Mol Biol.* 2002;316(3)
189. Kumar S, Tsai CJ, Nussinov R. Factors enhancing protein thermostability. *Protein Eng.* 2000;13(3)
190. Nakashima H, Nishikawa K, Ooi T. Distinct character in hydrophobicity of amino acid compositions of mitochondrial proteins. *Proteins.* 1990;8(2)
191. Qian N, Sejnowski TJ. Predicting the secondary structure of globular proteins using neural network models. *J Mol Biol.* 1988;202(4)

192. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of alpha helices. *Science*. 1988;240(4859)
193. Robson B, Suzuki E. Conformational properties of amino acid residues in globular proteins. *J Mol Biol*. 1976;107(3)
194. Vasquez M, Nemethy G, Scheraga HA. Computed conformational states of the 20 naturally occurring amino acid residues and of the prototype residue α -aminobutyric acid. *Macromolecules*. 1983;16(7)
195. Wertz DH, Scheraga HA. Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. *Macromolecules*. 1978;11(1)
196. Yutani K, Ogasahara K, Tsujita T, Sugino Y. Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins substituted at a unique position of tryptophan synthase alpha subunit. *PNAS*. 1987;84(13)
197. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Research*. 2000;28(1)
198. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13)
199. Georgiev AG. Interpretable numerical descriptors of amino acid space. *Journal of Computational Biology*. 2009;16(5)
200. Veltri DP. A computational and statistical framework for screening novel antimicrobial peptides: George Mason University; 2015.
201. Porto WF, Pires AS, Franco OL. Antimicrobial activity predictors benchmarking analysis using shuffled and designed synthetic peptides. *J Theor Biol*. 2017;426
202. Andreu D, Merrifield RB, Steiner H, Boman HG. N-terminal analogs of Cecropin-A - synthesis, antibacterial activity and conformational properties. *Biochemistry*. 1985;24(7)
203. Wade D, Boman A, Wahlin B, Drain CM, Andreu D, Boman HG, Merrifield RB. All-D amino acid-containing channel-forming antibiotic peptides. *Proc Natl Acad Sci U S A*. 1990;87(12)
204. Merrifield EL, Mitchell SA, Ubach J, Boman HG, Andreu D, Merrifield RB. D-enantiomers of 15-residue Cecropin A-melittin hybrids. *Int J Pept Protein Res*. 1995;46(3-4)
205. Steiner H, Andreu D, Merrifield RB. Binding and action of cecropin and cecropin analogs - antibacterial peptides from insects. *Biochim Biophys Acta*. 1988;939(2)
206. Juvvadi P, Vunnam S, Merrifield RB. Synthetic melittin, its enantio, retro, and retroenantio isomers, and selected chimeric analogs: Their antibacterial, hemolytic, and lipid bilayer action. *J Am Chem Soc*. 1996;118(38)
207. Juvvadi P, Vunnam S, Yoo B, Merrifield RB. Structure-activity studies of normal and retro pig cecropin-melittin hybrids. *J Pept Res*. 1999;53(3)
208. Andreu D, Ubach J, Boman A, Wahlin B, Wade D, Merrifield RB, Boman HG. Shortened Cecropin A-melittin hybrids. Significant size reduction retains potent antibiotic activity. *FEBS Lett*. 1992;296(2)
209. Juvvadi P, Vunnam S, Merrifield EL, Boman HG, Merrifield RB. Hydrophobic effects on antibacterial and channel-forming properties of Cecropin A-melittin hybrids. *Journal of Peptide Science*. 1996;2(4)
210. Yousefinejad S, Hemmateenejad B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemometrics and Intelligent Laboratory Systems*. 2015;149
211. Xu L, Fu HY, Yin QB, Fan Y, Goodarzi M, She YB. Interpretable linear and nonlinear quantitative structure-selectivity relationship (QSSR) modeling of a biomimetic catalytic system by particle swarm optimization based sparse regression. *Chemometrics and Intelligent Laboratory Systems*. 2016;159
212. Pirhadi S, Shiri F, Ghasemi JB. Multivariate statistical analysis methods in QSAR. *RSC Advances*. 2015;5(127)

213. Patel S, Stott IP, Bhakoo M, Elliott P. Patenting computer-designed peptides. *Journal of Computer-Aided Molecular Design*. 1998;12(6)
214. Czyzewski AM, Jenssen H, Fjell CD, Waldbrook M, Chongsiriwatana NP, Yuen E, Hancock RE, Barron AE. In Vivo, In Vitro, and In Silico Characterization of Peptoids as Antimicrobial Agents. *PLoS One*. 2016;11(2)
215. Wang Y, Yang YJ, Chen YN, Zhao HY, Zhang S. Computer-aided design, structural dynamics analysis, and in vitro susceptibility test of antibacterial peptides incorporating unnatural amino acids against microbial infections. *Comput Methods Programs Biomed*. 2016;134
216. Araujo SC, Maltarollo VG, Silva DC, Gertrudes JC, Honorio KM. ALK-5 Inhibition: A Molecular Interpretation of the Main Physicochemical Properties Related to Bioactive Ligands. *Journal of the Brazilian Chemical Society*. 2015;26(9)
217. Xu L, Zhang WJ. Comparison of different methods for variable selection. *Analytica Chimica Acta*. 2001;446(1-2)
218. So SS, Karplus M. Genetic neural networks for quantitative structure-activity relationships: improvements and application of benzodiazepine affinity for benzodiazepine/GABAA receptors. *J Med Chem*. 1996;39(26)
219. So SS, Karplus M. Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks. *J Med Chem*. 1996;39(7)
220. Kubinyi H. Variable selection in QSAR studies. 1. An evolutionary algorithm. *Quantitative Structure-Activity Relationships*. 1994;13(3)
221. Tong W, Xie Q, Hong H, Shi L, Fang H, Perkins R. Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity. *Environ Health Perspect*. 2004;112(12)
222. Bruce CL, Melville JL, Pickett SD, Hirst JD. Contemporary QSAR classifiers compared. *J Chem Inf Model*. 2007;47(1)
223. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci*. 2003;43(6)
224. Palmer DS, O'boyle NM, Glen RC, Mitchell JB. Random forest models to predict aqueous solubility. *J Chem Inf Model*. 2007;47(1)
225. Maji P, Paul S. Rough sets for selection of molecular descriptors to predict biological activity of molecules. *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews*. 2010;40(6)
226. Shehu A, Jong KD. Evolutionary algorithms for protein structure modeling. *Annual Conference on Genetic and Evolutionary Computation; Madrid, Spain*. 2756569: ACM; 2015.
227. Jain A, Castelli IE, Hautier G, Bailey DH, Jacobsen KW. Performance of genetic algorithms in search for water splitting perovskites. *Journal of Materials Science*. 2013;48(19)
228. Crick F. Central dogma of molecular biology. *Nature*. 1970;227(5258)
229. Conchillo-Sole O, De Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S. AGGRESKAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*. 2007;8(1)
230. Han HM, Gopal R, Park Y. Design and membrane-disruption mechanism of charge-enriched AMPs exhibiting cell selectivity, high-salt resistance, and anti-biofilm properties. *Amino Acids*. 2016;48(2)
231. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol*. 2005;3(3)
232. Scott MG, Hancock RE. Cationic antimicrobial peptides and their multifunctional role in the immune system. *Crit Rev Immunol*. 2002;20
233. Tille P. *Bailey & Scott's Diagnostic Microbiology-E-Book*: Elsevier Health Sciences; 2015.

234. Hintze JL, Nelson RD. Violin plots: A box plot-density trace synergism. *The American Statistician*. 1998;52(2)
235. Opuu V, Silvert M, Simonson T. Computational design of fully overlapping coding schemes for protein pairs and triplets. *Sci Rep*. 2017;7(1)
236. Lorenzo-Ginori JV, Rodriguez-Fuentes A, Abalo RG, Rodriguez RS. Digital signal processing in the analysis of genomic sequences. *Current Bioinformatics*. 2009;4(1)
237. Giannerini S, Gonzalez DL, Rosa R. DNA, dichotomic classes and frame synchronization: A quasi-crystal framework. *Philos Trans A Math Phys Eng Sci*. 2012;370(1969)
238. Lebre S, Gascuel O. The combinatorics of overlapping genes. *J Theor Biol*. 2017;415
239. Patel S, Stott IP, Bhakoo M, Elliott P. Patenting computer-designed peptides. *J Comput Aided Mol Des*. 1998;12(6)
240. Jones DT. De novo protein design using pairwise potentials and a genetic algorithm. *Protein Science*. 1994;3(4)
241. D'costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding GB, Poinar HN, Wright GD. Antibiotic resistance is ancient. *Nature*. 2011;477(7365)
242. Maximova T, Plaku E, Shehu A, editors. Computing transition paths in multiple-basin proteins with a probabilistic roadmap algorithm guided by structure data. *IEEE International Conference on Bioinformatics and Biomedicine*; 2015 9-12 Nov. 2015.
243. Lopez-Perez PM, Grimsey E, Bourne L, Mikut R, Hilpert K. Screening and optimizing antimicrobial peptides by using SPOT-synthesis. *Front Chem*. 2017;5
244. Borquaye LS, Darko G, Ocansey E, Ankomah E. Antimicrobial and antioxidant properties of the crude peptide extracts of *Galatea paradoxa* and *Patella rustica*. *SpringerPlus*. 2015;4
245. Bluhm ME, Knappe D, Hoffmann R. Structure-activity relationship study using peptide arrays to optimize Api137 for an increased antimicrobial activity against *Pseudomonas aeruginosa*. *Eur J Med Chem*. 2015;103
246. Minervini F, Algaron F, Rizzello CG, Fox PF, Monnet V, Gobetti M. Angiotensin I-converting-enzyme-inhibitory and antibacterial peptides from *Lactobacillus helveticus* PR4 proteinase-hydrolyzed caseins of milk from six species. *Appl Environ Microbiol*. 2003;69(9)
247. Mount DW. Using the basic local alignment search tool (BLAST). *Cold Spring Harbor Protocols*. 2007;2007(7)
248. Mcginnis S, Madden TL. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*. 2004;32(Suppl 2)
249. Li Z, Xu X, Meng L, Zhang Q, Cao L, Li W, Wu Y, Cao Z. Hp1404, a new antimicrobial peptide from the scorpion *Heterometrus petersii*. *PLoS One*. 2014;9(5)
250. Brand GD, Magalhães MT, Tinoco ML, Aragão FJ, Nicoli J, Kelly SM, Cooper A, Bloch Jr C. Probing protein sequences as sources for encrypted antimicrobial peptides. *PLoS One*. 2012;7(9)
251. Korichi M, Gerbaud V, Floquet P, Meniai AH, Nacef S, Joulia X. Computer aided aroma design I- Molecular knowledge framework. *Chemical Engineering and Processing*. 2008;47(11)
252. Karami Y, Khakzad H, Arab S, Fathy M, Shirazi H, editors. Protein structure prediction using bio-inspired algorithm: A review. *Artificial Intelligence and Signal Processing (AISP)*, 2012 16th CSI International Symposium on; 2012 2-3 May 2012.
253. Goldwater PN. Gut microbiota and immunity: Possible role in sudden infant death syndrome. *Front Immunol*. 2015;6
254. Holland JH. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT press; 1992.
255. Oliver I, Smith D, Holland JR, editors. Study of permutation crossover operators on the traveling salesman problem. *Genetic algorithms and their applications: proceedings of the second International*

- Conference on Genetic Algorithms: July 28-31, 1987 at the Massachusetts Institute of Technology, Cambridge, MA; 1987: Hillsdale, NJ: L. Erlbaum Associates, 1987.
256. Parsons RJ, Forrest S, Burks C. Genetic algorithms, operators, and DNA fragment assembly. *Machine Learning*. 1995;21(1-2)
 257. Islam SM, Das S, Ghosh S, Roy S, Suganthan PN. An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization. *IEEE Trans Syst Man Cybern B Cybern*. 2012;42(2)
 258. Cheng H, Yue K, Kazemzadeh-Narbat M, Liu Y, Khalilpour A, Li B, Zhang YS, Annabi N, Khademhosseini A. Mussel-inspired multifunctional hydrogel coating for prevention of infections and enhanced osteogenesis. *ACS Appl Mater Interfaces*. 2017;9(13)
 259. Dobson AJ, Purves J, Kamysz W, Rolff J. Comparing selection on *S. Aureus* between antimicrobial peptides and common antibiotics. *PLoS One*. 2013;8(10)
 260. Hickok NJ, Shapiro IM. Immobilized antibiotics to prevent orthopaedic implant infections. *Adv Drug Deliv Rev*. 2012;64(12)
 261. Onaizi SA, Leong SS. Tethering antimicrobial peptides: current status and potential challenges. *Biotechnol Adv*. 2011;29(1)
 262. Hilpert K, Elliott M, Jenssen H, Kindrachuk J, Fjell CD, Korner J, Winkler DF, Weaver LL, Henklein P, Ulrich AS, Chiang SH, Farmer SW, Pante N, Volkmer R, Hancock RE. Screening and characterization of surface-tethered cationic peptides for antimicrobial activity. *Chem Biol*. 2009;16(1)
 263. Polivkova M, Hubacek T, Staszek M, Svorcik V, Siegel J. Antimicrobial treatment of polymeric medical devices by silver nanomaterials and related technology. *Int J Mol Sci*. 2017;18(2)
 264. Cyphert EL, Von Recum HA. Emerging technologies for long-term antimicrobial device coatings: advantages and limitations. *Exp Biol Med*. 2017;242(8)
 265. Bauer S, Schmuki P, Von Der Mark K, Park J. Engineering biocompatible implant surfaces: Part I: Materials and surfaces. *Progress in Materials Science*. 2013;58(3)
 266. Tamerler C. Surfaces and their interfaces meet biology at the bio-interface. *JOM*. 2015;67(11)
 267. Luo QY, Lin Y, Peng J, Liu SL, Zhang ZL, Tian ZQ, Pang DW. Evaluation of nonspecific interactions between quantum dots and proteins. *Phys Chem Chem Phys*. 2014;16(17)
 268. Fu YY, Guan EL, Liang JG, Ren GL, Chen L. Probing the effect of Ag₂S quantum dots on human serum albumin using spectral techniques. *Journal of Nanomaterials*. 2017
 269. Wang K, Chen Y, Gong X, Xia J, Zhao J, Shen L. A mobile precursor determines protein resistance on nanostructured surfaces. *Phys Chem Chem Phys*. 2018;20(18)
 270. Shen L, Zhu XY. Evidence of a mobile precursor state in nonspecific protein adsorption. *Langmuir*. 2011;27(11)
 271. Shen L, Zhu J. Oriented protein nanoarrays on block copolymer template. *Macromol Rapid Commun*. 2016;37(6)
 272. Shen L, Zhu J. Heterogeneous surfaces to repel proteins. *Adv Colloid Interface Sci*. 2016;228
 273. Marzinek JK, Lian GP, Marzinek JK, Mantalaris A, Pistikopoulos EN, Zhao YY, Han LJ, Chen LJ, Bond PJ, Noro MG. Molecular and thermodynamic basis for EGCG-keratin interaction-part I: Molecular dynamics simulations. *AIChE Journal*. 2013;59(12)
 274. Kazemzadeh-Narbat M, Lai BF, Ding C, Kizhakkedathu JN, Hancock RE, Wang R. Multilayered coating on titanium for controlled release of antimicrobial peptides for the prevention of implant-associated infections. *Biomaterials*. 2013;34(24)
 275. Cetinel S, Caliskan HB, Yucesoy DT, Donatan AS, Yuca E, Urgan M, Karaguler NG, Tamerler C. Addressable self-immobilization of lactate dehydrogenase across multiple length scales. *Biotechnol J*. 2013;8(2)
 276. Gao W, Hu CM, Fang RH, Luk BT, Su J, Zhang L. Surface functionalization of gold nanoparticles with red blood cell membranes. *Adv Mater*. 2013;25(26)

277. Zhang K, Sugawara A, Tirrell DA. Generation of surface-bound multicomponent protein gradients. *Chembiochem*. 2009;10(16)
278. Sharma V, Blackwood K, Haddow D, Dye DJ. Protein binding functionalization of plasma-derivatized silicone surfaces. *Wound Repair and Regeneration*. 2013;21(6)
279. Chen RJ, Zhang Y, Wang D, Dai H. Noncovalent sidewall functionalization of single-walled carbon nanotubes for protein immobilization. *J Am Chem Soc*. 2001;123(16)
280. Gkaniatsou E, Sicard C, Ricoux R, Mahy JP, Steunou N, Serre C. Metal-organic frameworks: A novel host platform for enzymatic catalysis and detection. *Materials Horizons*. 2017;4(1)
281. Barbosa O, Ortiz C, Berenguer-Murcia A, Torres R, Rodrigues RC, Fernandez-Lafuente R. Strategies for the one-step immobilization-purification of enzymes as industrial biocatalysts. *Biotechnol Adv*. 2015;33(5)
282. Kuila T, Bose S, Khanra P, Mishra AK, Kim NH, Lee JH. Recent advances in graphene-based biosensors. *Biosens Bioelectron*. 2011;26(12)
283. Garcia-Galan C, Berenguer-Murcia A, Fernandez-Lafuente R, Rodrigues RC. Potential of different enzyme immobilization strategies to improve enzyme performance. *Advanced Synthesis & Catalysis*. 2011;353(16)
284. Deforest CA, Tirrell DA. A photoreversible protein-patterning approach for guiding stem cell fate in three-dimensional gels. *Nat Mater*. 2015;14(5)
285. Censi R, Di Martino P, Vermonden T, Hennink WE. Hydrogels for protein delivery in tissue engineering. *J Control Release*. 2012;161(2)
286. Tully J, Yendluri R, Lvov Y. Halloysite clay nanotubes for enzyme immobilization. *Biomacromolecules*. 2016;17(2)
287. Shi XW, Du YM, Wei XQ, Wu HP. Electrical signals guided protein immobilization and controlled release. *Abstracts of Papers of the American Chemical Society*. 2011;241
288. Choi JS, Yoo HS. Nano-inspired fibrous matrix with bi-phasic release of proteins. *J Nanosci Nanotechnol*. 2010;10(5)
289. Chen B, Qi W, Li X, Lei C, Liu J. Heated proteins are still active in a functionalized nanoporous support. *Small*. 2013;9(13)
290. Liu Y, Ai K, Lu L. Polydopamine and its derivative materials: synthesis and promising applications in energy, environmental, and biomedical fields. *Chem Rev*. 2014;114(9)
291. Hudson S, Cooney J, Magner E. Proteins in mesoporous silicates. *Angew Chem Int Ed Engl*. 2008;47(45)
292. Brady D, Jordaan J. Advances in enzyme immobilisation. *Biotechnol Lett*. 2009;31(11)
293. Wisdom C, Vanoosten SK, Boone KW, Khvostenko D, Arnold PM, Snead ML, Tamerler C. Controlling the biomimetic implant interface: Modulating antimicrobial activity by spacer design. *J Mol Eng Mater*. 2016;4(1)
294. Care A, Bergquist PL, Sunna A. Solid-binding peptides in biomedicine. *Peptides and Peptide-based Biomaterials and their Biomedical Applications*: Springer; 2017.
295. Donatan S, Sarikaya M, Tamerler C, Urgen M. Effect of solid surface charge on the binding behaviour of a metal-binding peptide. *J R Soc Interface*. 2012;9(75)
296. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004;32(Web Server issue)
297. Kabsch W, Sander C. DSSP: definition of secondary structure of proteins given a set of 3D coordinates. *Biopolymers*. 1983;22
298. Daniele-Silva A, Machado RJ, Monteiro NK, Estrela AB, Santos EC, Carvalho E, Araujo Junior RF, Melo-Silveira RF, Rocha HA, Silva-Junior AA, Fernandes-Pedrosa MF. Stigmurin and TsAP-2 from *Tityus stigmurus* scorpion venom: Assessment of structure and therapeutic potential in experimental sepsis. *Toxicol*. 2016;121

299. Ma QQ, Dong N, Shan AS, Lv YF, Li YZ, Chen ZH, Cheng BJ, Li ZY. Biochemical property and membrane-peptide interactions of de novo antimicrobial peptides designed by helix-forming units. *Amino Acids*. 2012;43(6)
300. Nguyen LT, Haney EF, Vogel HJ. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol*. 2011;29(9)
301. Mann S. *Biomineralization: principles and concepts in bioinorganic materials chemistry*: Oxford University Press on Demand; 2001.
302. Paital SR, Dahotre NB. Calcium phosphate coatings for bio-implant applications: Materials, performance factors, and methodologies. *Materials Science and Engineering: R: Reports*. 2009;66(1–3)
303. Mcentire BJ, Bal BS, Rahaman MN, Chevalier J, Pezzotti G. Ceramics and ceramic coatings in orthopaedics. *Journal of the European Ceramic Society*. 2015;35(16)
304. Utku FS, Seckin E, Goller G, Tamerler C, Urgen M. Electrochemically designed interfaces: Hydroxyapatite coated macro-mesoporous titania surfaces. *Applied Surface Science*. 2015;350
305. Goodman SB, Yao Z, Keeney M, Yang F. The future of biologic coatings for orthopaedic implants. *Biomaterials*. 2013;34(13)
306. Legeros RZ. Properties of osteoconductive biomaterials: calcium phosphates. *Clin Orthop Relat Res*. 2002;395(395)
307. Gungormus M, Fong H, Kim IW, Evans JS, Tamerler C, Sarikaya M. Regulation of in vitro calcium phosphate mineralization by combinatorially selected hydroxyapatite-binding peptides. *Biomacromolecules*. 2008;9(3)
308. Combes C, Bareille R, Rey C. Calcium carbonate-calcium phosphate mixed cement compositions for bone reconstruction. *J Biomed Mater Res A*. 2006;79(2)
309. Chow LC. *Solubility of calcium phosphates. Octacalcium phosphate*. 18: Karger Publishers; 2001.
310. Ye Q, Spencer P, Yuca E, Tamerler C. Engineered peptide repairs defective adhesive-dentin interface. *Macromol Mater Eng*. 2017;302(5)
311. Gungormus M, Oren EE, Horst JA, Fong H, Hnilova M, Somerman MJ, Snead ML, Samudrala R, Tamerler C, Sarikaya M. Cementomimetics-constructing a cementum-like biomineralized microlayer via amelogenin-derived peptides. *Int J Oral Sci*. 2012;4(2)
312. Chiu D, Zhou W, Kitayaporn S, Schwartz DT, Murali-Krishna K, Kavanagh TJ, Baneyx F. Biomineralization and size control of stable calcium phosphate core-protein shell nanoparticles: potential for vaccine applications. *Bioconjug Chem*. 2012;23(3)
313. Tsuji T, Onuma K, Yamamoto A, Iijima M, Shiba K. Direct transformation from amorphous to crystalline calcium phosphate facilitated by motif-programmed artificial proteins. *Proc Natl Acad Sci U S A*. 2008;105(44)
314. Shaw WJ, Campbell AA, Paine ML, Snead ML. The COOH terminus of the amelogenin, LRAP, is oriented next to the hydroxyapatite surface. *Journal of Biological Chemistry*. 2004;279(39)
315. White SN, Luo W, Paine ML, Fong H, Sarikaya M, Snead ML. Biological organization of hydroxyapatite crystallites into a fibrous continuum toughens and controls anisotropy in human enamel. *J Dent Res*. 2001;80(1)
316. Cristache CM, Burlibasa M, Cristache G, Drafta S, Popovici IA, Iliescu AA, Zisi S, Burlibasa L. Zirconia and its biomedical applications. *Metalurgia International*. 2011;16(7)
317. Chen YW, Moussi J, Drury JL, Wataha JC. Zirconia in biomedical applications. *Expert Rev Med Devices*. 2016;13(10)
318. Xuefeng Z, Tianzhu Z, Xiaoli J, Ning G. The surface modification of medical polyurethane to improve the hydrophilicity and lubricity: The effect of pretreatment. *Journal of Applied Polymer Science*. 2010;116(3)
319. Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput*. 2012;8(5)

320. Cui Q, Nussinov R. Making biomolecular simulations accessible in the post-Nobel Prize era. *PLoS Comput Biol.* 2014;10(8)
321. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry.* 2009;30(10)
322. Ribeiro AA, Ortiz V. A chemical perspective on allostery. *Chem Rev.* 2016;116(11)
323. Papaleo E, Saladino G, Lambrugh M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The role of protein loops and linkers in conformational dynamics and allostery. *Chem Rev.* 2016;116(11)
324. Liu J, Nussinov R. Allostery: An overview of its history, concepts, methods, and applications. *PLoS Comput Biol.* 2016;12(6)
325. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature.* 2014;508(7496)
326. Tamerler C, Sarikaya M. Molecular biomimetics: Genetic synthesis, assembly, and formation of materials using peptides. *Mrs Bulletin.* 2008;33(5)
327. Liu J, Nussinov R. The role of allostery in the ubiquitin-proteasome system. *Crit Rev Biochem Mol Biol.* 2013;48(2)
328. Ma B, Tsai CJ, Haliloglu T, Nussinov R. Dynamic allostery: linkers are not merely flexible. *Structure.* 2011;19(7)
329. Zdanov A, Wu S, Dimaio J, Konishi Y, Li Y, Wu X, Edwards BF, Martin PD, Cygler M. Crystal structure of the complex of human alpha-thrombin and nonhydrolyzable bifunctional inhibitors, hirutinin-2 and hirutinin-6. *Proteins.* 1993;17(3)
330. Szewczuk Z, Gibbs BF, Yue SY, Purisima E, Zdanov A, Cygler M, Konishi Y. Design of a linker for trivalent thrombin inhibitors - interaction fo the main chain of the linker with thrombin. *Biochemistry.* 1993;32(13)
331. Lin X, Koelsch G, Loy JA, Tang J. Rearranging the domains of pepsinogen. *Protein Science.* 1995;4(2)
332. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013;65(10)
333. Crasto CJ, Feng JA. LINKER: a program to generate linker sequences for fusion proteins. *Protein Eng.* 2000;13(5)
334. Berova N, Nakanishi K, Woody R. Circular dichroism: principles and applications: John Wiley & Sons; 2000.
335. Mavridis L, Janes RW. PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics.* 2017;33(1)
336. Molski MA, Goodman JL, Chou F-C, Baker D, Das R, Schepartz A. Remodeling a small beta-peptide bundle. *Chemical Science.* 2013;4(1)
337. Wiedemann C, Bellstedt P, Gorkach M. CAPITO--a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics.* 2013;29(14)
338. Hoang HN, Abbenante G, Hill TA, Ruiz-Gómez G, Fairlie DP. Folding pentapeptides into left and right handed alpha helices. *Tetrahedron.* 2012;68(23)
339. Burck J, Roth S, Wadhwani P, Afonin S, Kanithasen N, Strandberg E, Ulrich AS. Conformation and membrane orientation of amphiphilic helical peptides by oriented circular dichroism. *Biophys J.* 2008;95(8)
340. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem.* 2000;287(2)

341. Yuca E, Karatas AY, Seker UO, Gungormus M, Dinler-Doganay G, Sarikaya M, Tamerler C. In vitro labeling of hydroxyapatite minerals by an engineered protein. *Biotechnology and Bioengineering*. 2011;108(5)
342. Bagheri M, Arasteh S, Haney EF, Hancock RE. Tryptic stability of synthetic bactenecin derivatives is determined by the side chain length of cationic residues and the peptide conformation. *J Med Chem*. 2016;59(7)
343. Gray JJ, Chaudhury S, Lyskov S. The PyRosetta interactive platform for protein structure prediction and design: a set of educational modules: Lulu. com; 2010.
344. Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput*. 2011;7(2)
345. Grzymala-Busse JW. MLEM2—discretization during rule induction. *Intelligent Information Processing and Web Mining: Springer*; 2003.
346. Kwon S, Kim B, Noh H. Study of Physical Properties of UV Protective Film with Acrylate Polymers. *Polymer-Korea*. 2017;41(2)
347. Carson D, Hnilova M, Yang X, Nemeth CL, Tsui JH, Smith AS, Jiao A, Regnier M, Murry CE, Tamerler C, Kim DH. Nanotopography-Induced Structural Anisotropy and Sarcomere Development in Human Cardiomyocytes Derived from Induced Pluripotent Stem Cells. *ACS Appl Mater Interfaces*. 2016;8(34)
348. Baran KL, Chimenti MS, Schlessman JL, Fitch CA, Herbst KJ, Garcia-Moreno BE. Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J Mol Biol*. 2008;379(5)
349. Leprince JG, Palin WM, Hadis MA, Devaux J, Leloup G. Progress in dimethacrylate-based dental composite technology and curing efficiency. *Dent Mater*. 2013;29(2)
350. Spencer P, Ye Q, Park J, Topp EM, Misra A, Marangos O, Wang Y, Bohaty BS, Singh V, Sene F, Eslick J, Camarda K, Katz JL. Adhesive/Dentin interface: The weak link in the composite restoration. *Annals of Biomedical Engineering*. 2010;38(6)
351. Chisini LA, Collares K, Cademartori MG, De Oliveira LJC, Conde MCM, Demarco FF, Correa MB. Restorations in primary teeth: A systematic review on survival and reasons for failures. *Int J Paediatr Dent*. 2018;28(2)
352. Demarco FF, Correa MB, Cenci MS, Moraes RR, Opdam NJ. Longevity of posterior composite restorations: not only a matter of materials. *Dent Mater*. 2012;28(1)
353. Spencer P, Ye Q, Misra A, Goncalves SE, Laurence JS. Proteins, pathogens, and failure at the composite-tooth interface. *J Dent Res*. 2014;93(12)
354. Zhang K, Melo MA, Cheng L, Weir MD, Bai Y, Xu HH. Effect of quaternary ammonium and silver nanoparticle-containing adhesives on dentin bond strength and dental plaque microcosm biofilms. *Dent Mater*. 2012;28(8)
355. Jedrychowski JR, Caputo AA, Kerper S. Antibacterial and mechanical properties of restorative materials combined with chlorhexidines. *J Oral Rehabil*. 1983;10(5)
356. Backlund CJ, Worley BV, Schoenfisch MH. Anti-biofilm action of nitric oxide-releasing alkyl-modified poly(amidoamine) dendrimers against *Streptococcus mutans*. *Acta Biomater*. 2016;29
357. Mashburn-Warren L, Downey JS, Goodman SD. Novel method for the depletion of cariogenic bacteria using dextranomer microspheres. *Molecular Oral Microbiology*. 2017;32(6)
358. Gorr SU, Abdolhosseini M. Antimicrobial peptides and periodontal disease. *J Clin Periodontol*. 2011;38 Suppl 11
359. Gorr SU. Antimicrobial peptides of the oral cavity. *Periodontol 2000*. 2009;51
360. Tao R, Jurevic RJ, Coulton KK, Tsutsui MT, Roberts MC, Kimball JR, Wells N, Berndt J, Dale BA. Salivary antimicrobial peptide expression and dental caries experience in children. *Antimicrob Agents Chemother*. 2005;49(9)

361. Mai S, Mauger MT, Niu LN, Barnes JB, Kao S, Bergeron BE, Ling JQ, Tay FR. Potential applications of antimicrobial peptides and their mimics in combating caries and pulpal infections. *Acta Biomater.* 2017;49
362. Su M, Yao S, Gu L, Huang Z, Mai S. Antibacterial effect and bond strength of a modified dental adhesive containing the peptide nisin. *Peptides.* 2018;99
363. Breschi L, Mazzoni A, Ruggeri A, Cadenaro M, Di Lenarda R, De Stefano Dorigo E. Dental adhesion review: aging and stability of the bonded interface. *Dent Mater.* 2008;24(1)
364. Spencer P, Wang Y. Adhesive phase separation at the dentin interface under wet bonding conditions. *Journal of Biomedical Materials Research Part A.* 2002;62(3)
365. Abedin F, Ye Q, Good HJ, Parthasarathy R, Spencer P. Polymerization-and solvent-induced phase separation in hydrophilic-rich dentin adhesive mimic. *Acta Biomater.* 2014;10(7)
366. Moszner N, Salz U. Recent developments of new components for dental adhesives and composites. *Macromolecular Materials and Engineering.* 2007;292(3)
367. Moszner N. New monomers for dental application. *Macromolecular Symposia.* 2004;217
368. Asmussen E, Hansen EK, Peutzfeldt A. Influence of the solubility parameter of intermediary resin on the effectiveness of the gluma bonding system. *J Dent Res.* 1991;70(9)
369. Peutzfeldt A. Resin composites in dentistry: the monomer systems. *European Journal of Oral Sciences.* 1997;105(2)
370. Moszner N, Salz U. New developments of polymeric dental composites. *Progress in Polymer Science.* 2001;26(4)
371. Murphy EF, Lu JR, Lewis AL, Brewer J, Russell J, Stratford P. Characterization of protein adsorption at the phosphorylcholine incorporated polymer-water interface. *Macromolecules.* 2000;33(12)
372. Kiritoshi Y, Ishihara K. Preparation of cross-linked biocompatible poly(2-methacryloyloxyethyl phosphorylcholine) gel and its strange swelling behavior in water/ethanol mixture. *J Biomater Sci Polym Ed.* 2002;13(2)
373. Yuan B, Chen Q, Ding WQ, Liu PS, Wu SS, Lin SC, Shen J, Gai Y. Copolymer coatings consisting of 2-methacryloyloxyethyl phosphorylcholine and 3-methacryloxypropyl trimethoxysilane via ATRP to improve cellulose biocompatibility. *ACS Appl Mater Interfaces.* 2012;4(8)
374. Lewis AL, Hughes PD, Kirkwood LC, Leppard SW, Redman RP, Tolhurst LA, Stratford PW. Synthesis and characterisation of phosphorylcholine-based polymers useful for coating blood filtration devices. *Biomaterials.* 2000;21(18)
375. Lewis AL. Phosphorylcholine-based polymers and their use in the prevention of biofouling. *Colloids and Surfaces B-Biointerfaces.* 2000;18(3-4)
376. Clarke S, Davies MC, Roberts CJ, Tendler SJB, Williams PM, O'byrne V, Lewis AL, Russell J. Surface mobility of 2-methacryloyloxyethyl phosphorylcholine-co-lauryl methacrylate polymers. *Langmuir.* 2000;16(11)
377. Zhang N, Zhang K, Melo MA, Chen C, Fouad AF, Bai Y, Xu HH. Novel protein-repellent and biofilm-repellent orthodontic cement containing 2-methacryloyloxyethyl phosphorylcholine. *J Biomed Mater Res B Appl Biomater.* 2016;104(5)
378. Zhang N, Weir MD, Romberg E, Bai Y, Xu HH. Development of novel dental adhesive with double benefits of protein-repellent and antibacterial capabilities. *Dent Mater.* 2015;31(7)
379. Zhang N, Melo MA, Chen C, Liu J, Weir MD, Bai Y, Xu HH. Development of a multifunctional adhesive system for prevention of root caries and secondary caries. *Dent Mater.* 2015;31(9)
380. Zhang N, Melo MA, Bai Y, Xu HH. Novel protein-repellent dental adhesive containing 2-methacryloyloxyethyl phosphorylcholine. *Journal of Dentistry.* 2014;42(10)
381. Song L, Ye Q, Ge X, Misra A, Tamerler C, Spencer P. Probing the neutralization behavior of zwitterionic monomer-containing dental adhesive. *Dent Mater.* 2017;33(5)

382. Liang C, Yuan F, Liu F, Wang Y, Gao Y. Structure and antimicrobial mechanism of varepsilon-polylysine-chitosan conjugates through Maillard reaction. *International Journal of Biological Macromolecules*. 2014;70
383. Pranantyo D, Xu LQ, Hou Z, Kang ET, Chan-Park MB. Increasing bacterial affinity and cytocompatibility with four-arm star glycopolymers and antimicrobial alpha-polylysine. *Polymer Chemistry*. 2017;8(21)
384. Ferraro MJ. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically: NCCLS; 2000.
385. Min KR, Galvis A, Williams B, Rayala R, Cudic P, Ajdic D. Antibacterial and antibiofilm activities of a novel synthetic cyclic lipopeptide against cariogenic *Streptococcus mutans* UA159. *Antimicrob Agents Chemother*. 2017;61(8)
386. Tu H, Fan Y, Lv X, Han S, Zhou X, Zhang L. Activity of synthetic antimicrobial peptide GH12 against oral *Streptococci*. *Caries Res*. 2016;50(1)
387. Page B, Page M, Noel C. A new fluorometric assay for cytotoxicity measurements in-vitro. *Int J Oncol*. 1993;3(3)
388. Simoncini D, Nakata H, Ogata K, Nakamura S, Zhang KY. Quality assessment of predicted protein models using energies calculated by the fragment molecular orbital method. *Molecular Informatics*. 2015;34(2-3)
389. Sankar K, Jia K, Jernigan RL. Knowledge-based entropies improve the identification of native protein structures. *Proc Natl Acad Sci U S A*. 2017;114(11)
390. Casanova-Morales N, Alavi Z, Wilson CaM, Zocchi G. Identifying Chaotropic and Kosmotropic Agents by Nanorheology. *Journal of Physical Chemistry B*. 2018;122(14)
391. Henry GD, Sykes BD. [18] Methods to study membrane protein structure in solution. *Methods in enzymology*. 239: Elsevier; 1994.
392. Buck M. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quarterly Reviews of Biophysics*. 1998;31(3)
393. Pandey R, Usui K, Livingstone RA, Fischer SA, Pfaendtner J, Backus EH, Nagata Y, Frohlich-Nowoisky J, Schmuser L, Mauri S, Scheel JF, Knopf DA, Poschl U, Bonn M, Weidner T. Ice-nucleating bacteria control the order and dynamics of interfacial water. *Sci Adv*. 2016;2(4)
394. Balelli I, Milisic V, Wainrib G. Random walks on binary strings applied to the somatic hypermutation of B-cells. *Math Biosci*. 2018;300

9.0 Appendix A. Nomenclature

AAindex1: amino acid index 1

ADP1: amelogenin-derived peptide 1

APD2: antimicrobial peptide database 2

APD3: antimicrobial peptide database 3

AMP: antimicrobial peptide

ANN: artificial neural network

BLAST: basic local alignment search tool

BisGMA: bisphenol A-glycidyl methacrylate

Ca: calcium

CAMD: computer-aided molecular design

CAMP: collections of antimicrobial peptides

CD: circular dichroism

CHX: chlorhexidine gluconate

CLN: condition limit number

DA: discriminant analysis

DSSP: dictionary of secondary structure in proteins

CLN-MLEM2: condition-limit number modified learning from experience method 2

EFC: evolutionary feature construction

EFC-FBCF: evolutionary feature construction without fast correlation-based filter selection

EFC+307-FCBF: evolutionary feature construction and fast correlation-based filter selection with 307 features

FBCF: fast correlation-based filter selection

FDR: false discovery rate

FMOC: fluorenylmethoxycarbonyl protecting group

FN: false negative

FP: false positive

GA: genetic algorithm

GH12: 12 amino acid antimicrobial peptide

GH12-M2: GH12 modification 1

GH12-M2: GH12 modification 2

GRAVY: grand average of hydropathy

HABP1: hydroxyapatite binding peptide 1

HEMA: 2-hydroxyethylmethacrylate

HMM: hidden Markov model

iAMP-2L: antimicrobial peptide prediction two-level

IC₅₀: concentration of 50% inhibition

IRIM: interesting rule induction method

LR: logistic regression

MCC: Matthew's correlation coefficient

MIC: minimum inhibitory concentration

MLEM2: Modified Learning from Experience Module 2

MPC: 2-methacryloyloxyethyl phosphorylcholine

MRE: mean residue ellipticity

NCBI: National Center for Biotechnology Information

NIH: National Institutes of Health

NMR: nuclear magnetic resonance

PUABP: poly-urethane acrylic binding peptide

QM: quantitative matrix

QSAR: quantitative structure activity relationship

RST: rough set theory

SAM: self-assembled monolayer

SFE: spacer frequency error

SPPS: solid-phase peptide synthesis

SVM: support vector machine

TEGDMA: triethylene glycol dimethacrylate

Ti: titanium

TFE: trifluoroethanol

TN: true negative

TP: true positive

10.0 Appendix B. PyRosetta Folding Code

The following is an implementation of the PyRosetta folding method in Python 2.7 for generating peptide structure decoys.

```
#####  
# input for script is a FASTA file with peptides to be modelled  
#####  
###Import  
import os, sys, random, time, copy, math, subprocess  
import rosetta as r  
#####  
#####  
#provides access to Rosetta pose data structure and Rosetta score function  
r.init()  
scorefxn = r.ScoreFunction()  
scorefxn.set_weight(fa_atr, 0.8) #lennard-jones attractive forces  
scorefxn.set_weight(fa_rep, 0.4) #lennard-jones repulsive forces  
scorefxn.set_weight(fa_sol, 0.75) #lazaridis-jarplus solvation energy  
scorefxn.set_weight(fa_pair, 1.0) #favors salt bridges  
#scorefxn.set_weight(fa_plane, 0.25) #pi-pi interactions, by default is zero  
scorefxn.set_weight(rama, 1.0) #weights ramachandran plot information  
scorefxn.set_weight(hbond_lr_bb, 1.2) #long-range hydrogen bonds  
scorefxn.set_weight(hbond_sr_bb, 0.585) #short-range hydrogen bonds  
scorefxn.set_weight(hbond_bb_sc, 1.2) #backbone-side chain hydrogen bonds  
scorefxn.set_weight(hbond_sc, 1.1) #side chain-side chain hydrogen bonds  
scorefxn.set_weight(dslf_ss_dst, 1.5) #distance score in current disulfide  
scorefxn.set_weight(dslf_cs_ang, 1.5) #csangles score in current disulfide  
scorefxn.set_weight(dslf_ss_dih, 1.5) #dihedral score in current disulfide  
scorefxn.set_weight(dslf_ca_dih, 1.5) #ca dihedral score in current disulfide
```

```

coarseFastRelax = r.FastRelax()
coarseFastRelax.set_scorefxn(scorefxn)

scorefxn2 = r.get_fa_scorefxn()
fastRelax = r.FastRelax()
fastRelax.set_scorefxn(scorefxn2)
#classicRelax = rosetta.ClassicRelax()
#classicRelax.set_scorefxn(scorefxn2)

###Global Variables
date = 'dateTag'
scriptDir = 'dirPath'
if ('annealedPDB' + date) not in os.listdir(scriptDir):
    os.makedirs(scriptDir + '/annealedPDB'+date)

fileTagList = ['fastaFileTag']
fragmentFile9List = ['fragmentID']
fragmentFile3List = ['fragmentID']

disulfideTagList = [f + '_disulfide.txt' for f in fileTagList]

#####

###Classes
class Sequence:
    def __init__(self,name,seq):
        self.name = name
        self.seq = seq

```

```

class DisulfidePairs:

    def __init__(self,name, pairs):

        self.name = name

        self.pairs = pairs #list of paired tuples for known disulfide bonds

#imported classes from swarmIntFolding

#####

#Move defines how the backbone angles of a single residue are optimized

class Move:

    def __init__(self, pose, residue, initialAngleNum, disulfidePair, isStochastic=False):

        self.pose = pose

        self.residue = residue

        self.disulfidePair = disulfidePair

        self.isStochastic = isStochastic

        self.offset = random.randint(0,360)%(360/float(initialAngleNum))

        self.choice = 0

        self.degreesOfFreedom = initialAngleNum

        self.angleOpt() #returns empty

        self.disulfideDist = self.pickFromPool() #sets energy but also changes pose to chosen state

    def angleOpt(self): #returns energy of optimized state

        phiList = [360/float(self.degreesOfFreedom)*x + self.offset for x in range(self.degreesOfFreedom)]

        psiList = [360/float(self.degreesOfFreedom)*x + self.offset for x in range(self.degreesOfFreedom)]

        #C-terminus has no omega bond

        if self.residue == self.pose.n_residue():

            omegaList = [0]

        else:

```

```

    omegaList = [180,-180]
angleList = [phiList, psiList, omegaList]

#Initial case of recursive search
self.pool = self.crossPools(angleList)
self.genSearch(self.degreesOfFreedom/2)

def pickFromPool(self):
    #Pick the tuple based on kind
    #Default kind is greedy
    if self.isStochastic:
        self.choice = self.selectLogChoice()
    self.setAngles(self.pool[self.choice][1], self.pool[self.choice][2], self.pool[self.choice][3])
    return self.pool[self.choice][0]

def crossPools(self, angleList):
    #print angleList
    pool = [(self.poseScore(phi, psi, omega), phi, psi, omega) for phi in angleList[0] for psi in angleList[1]
for omega in angleList[2]]
    pool.sort(key=lambda t: t[0])
    return pool

def mutatePool(self, increment):
    phiList = list(set([self.pool[x][1] for x in range(len(self.pool))]))
    phiList += [x + increment for x in phiList] + [x - increment for x in phiList]
    psiList = list(set([self.pool[x][2] for x in range(len(self.pool))]))
    psiList += [x + increment for x in psiList] + [x - increment for x in psiList]
    omegaList = list(set([self.pool[x][3] for x in range(len(self.pool))]))

```

```

    return [phiList, psiList, omegaList]

def poseScore(self, phi, psi, omega):
    self.setAngles(phi, psi, omega)
    return disulfideDistance(self.pose, self.disulfidePair)

def setAngles(self, phi, psi, omega):
    self.pose.set_phi(self.residue, phi)
    self.pose.set_psi(self.residue, psi)
    if omega != 0:
        self.pose.set_omega(self.residue, omega)

def genSearch(self, increment):
    N = min(75, len(self.pool))
    #print "N"
    #print N
    tempPool = self.pool + self.crossPools(self.mutatePool(increment))
    #print "temp pool length"
    #print len(tempPool)
    tempPool.sort(key=lambda t: t[0])
    self.pool = tempPool[0:N]
    del tempPool
    #print "self.pool length"
    #print len(self.pool)
    if increment <= 1:
        return
    else:
        self.genSearch(increment/2)

```

```

def selectLogChoice(self):
    if len(self.pool) == 1:
        return self.pool[0]

    #calculates energy differences between tuples
    deltaEnergy = [self.pool[x][0]- self.pool[0][0] for x in range(len(self.pool))]

    #By definition of lognormvariate, r > 0 and r grows to infinity with a logarithmic probability
    extraEnergy = random.lognormvariate(0, 1)

    #Picks highest energy choice that is suboptimal by at most of the amount of extra energy
    return len(filter(lambda x: x <= extraEnergy, deltaEnergy))-1

#####
# The instances are lists of numbers from 1 to N | N is the length of the sequence
bulkiness = {'A': 11.500, 'R': 14.280, 'N': 12.820, 'D': 11.680, 'C': 13.460, 'Q': 14.450, 'E': 13.570, 'G':
3.400, 'H': 13.690, 'I': 21.400,
            'L': 21.400, 'K': 15.710, 'M': 16.250, 'F': 19.800, 'P': 17.430, 'S': 9.470, 'T': 15.770, 'W': 21.670,
'Y': 18.030, 'V': 21.570}

class MoveOrder:
    def __init__(self, kind, pose):
        self.kind = kind
        self.pose = pose
        self.moveList = []
        if kind == 'singleResidue':
            self.moveList = self.singleResidue()
        if kind == 'smallFragment':

```

```

    self.moveList = self.smallFragment()

    print self.moveList

if kind == 'foldInHalfOrder':

    self.moveList = self.foldInHalfOrder()

if kind == 'shuffleOrder':

    self.moveList = self.shuffleOrder()

if kind == 'linearOrder':

    self.moveList = self.linearOrder()

if kind == 'bulkinessOrder':

    self.moveList = self.bulkinessOrder()

def foldInHalfOrder(self): #sorts by distance from the middle residue

    mid = (self.pose.n_residue()-1)/2 + 1

    midSeqTuples = [(x, abs(x-mid)) for x in range(1, self.pose.n_residue()+1)]

    midSeqTuples.sort(key=lambda t:t[1])

    return [t[0] for t in midSeqTuples]

def shuffleOrder(self):

    residues = range(1,self.pose.n_residue()+1)

    random.shuffle(residues)

    return residues

def linearOrder(self):

    return range(1, self.pose.n_residue()+1)

def bulkinessOrder(self):

    bulkSeqTuples = [(y, bulkiness[x]) for x,y in zip(self.pose.sequence(), range(1,
self.pose.n_residue()+1))]

```

```
bulkSeqTuples.sort(key=lambda t:t[1])
return [t[0] for t in bulkSeqTuples]
```

```
def smallFragment(self):
    n = random.randint(2,5)
    r = random.randint(1,self.pose.n_residue())
    fragment1 = [(r+x) % self.pose.n_residue() + 1 for x in range(n)]
    fragment2 = sorted(fragment1,reverse=True)
    return random.choice([fragment1, fragment2])
```

```
def singleResidue(self):
    return [random.randint(1,self.pose.n_residue())]
```

```
#####
#####
```

```
#####
#####
```

```
#####
```

```
###Functions
```

```
def copyPoseList(poseList):
    copyList = list()
    for pose in poseList:
        poseCopy = r.Pose()
        poseCopy.assign(pose)
        copyList.append(poseCopy)
    return copyList
```

```

def pdbDump(poseList, prependString, scorefxn):
    global date
    global fileTag
    #outLog = open(os.getcwd() + '\\\\' + outfile, 'a')
    for pose in poseList:
        name = prependString + pose.pdb_info().name().rstrip('*')
        pose.dump_scored_pdb('annealedPDB'+date+'\\' + name + '_' + fileTag + '.pdb', scorefxn)
        print "PDB File Created: " + name + '_' + fileTag + '.pdb'
    return

```

```

def fromFASTAtoSeqList(fileName):
    fastaFileName = fileName
    seqList = list()
    #Open fasta file
    fastafile = open(fastaFileName)
    fastaLines = fastafile.readlines()
    #Make each fasta object a Sequence object
    for line in fastaLines:
        if line[0] == '%' or line[0] == '\n':
            continue
        if line[0] == '>':
            seqName = line.split('>')[1].rsplit('\n')[0]
        else:
            seqSequence = line.rstrip('\n')
            #Add it to the sequence list
            seqList.append(Sequence(seqName,seqSequence))
    return seqList

```

```
def fromSeqListToPoseList(seqList):
```

```
    poseList = list()
```

```
    for sequence in seqList:
```

```
        pose = r.pose_from_sequence(sequence.seq)
```

```
        pose.pdb_info().name(sequence.name)
```

```
        print pose.sequence()
```

```
        poseList.append(pose)
```

```
    return poseList
```

```
def coarseRelaxPose(poseList, iterNum):
```

```
    for pose in poseList:
```

```
        coarseFastRelax.apply(pose)
```

```
    prependNames(poseList, 'coarseRelax_' + str(iterNum) + '_')
```

```
    #pdbDump(poseList, ",coarseFastRelax.get_scorefxn())
```

```
    return
```

```
def relaxPose(poseList, iterNum):
```

```
    for pose in poseList:
```

```
        fastRelax.apply(pose)
```

```
    prependNames(poseList, 'relax_' + str(iterNum) + '_')
```

```
    pdbDump(poseList, ", fastRelax.get_scorefxn())
```

```
    return
```

```
def fragmentMover(poseList, fragset9, fragset3, iterNum, knownDisulfideBondList):
```

```
    movemap = r.MoveMap()
```

```
    movemap.set_bb(True)
```

```
    mover_9mer = r.ClassicFragmentMover(fragset9, movemap)
```

```

mover_3mer = r.ClassicFragmentMover(fragset3, movemap)
#Combines sequences into a single long protein to get fragments
totalSequence = ''
for pose in poseList:
    totalSequence = totalSequence + pose.sequence()
totalPose = r.pose_from_sequence(totalSequence)
#Inserts fragments at an average of one change per position
n = len(totalSequence)
repeat_9mer = r.RepeatMover(mover_9mer,n)
repeat_9mer.apply(totalPose)
insertAllBackBoneAngles(poseList, totalPose)

```

```

prefold(poseList, knownDisulfideBondList, iterNum)

```

```

trialFrag(poseList, totalPose, mover_9mer, movemap)
trialFrag(poseList, totalPose, mover_3mer, movemap)
seqMover = r.SequenceMover()
seqMover.add_mover(mover_9mer)
seqMover.add_mover(mover_3mer)
seqMover.add_mover(mover_3mer)
trialFrag(poseList, totalPose, seqMover, movemap)
#pdbDump(poseList, "",coarseFastRelax.get_scorefxn())
return

```

#Finds the next set of fragments based on lowering energy via Metropolis criterion

```

def trialFrag(poseList, totalPose, mover, movemap):

```

```

    numOfPasses = 10

```

```

    kT = 4.0

```

```

for i in range(numOfPasses):
    kT = max([0.5*kT, 0.0001])
    totalIndex = 1
    for pose in poseList:
        movemap.set_bb_true_range(totalIndex, totalIndex + pose.total_residue()-1)
        currentEnergy = scorefxn(pose)
        for j in range(pose.total_residue()):
            copyPose = r.Pose()
            copyPose.assign(pose)
            #fragment insertion
                for k in range(i):
                    mover.apply(totalPose)
            #copyAngles to pose
            insertBackBoneAngles(pose, totalPose, totalIndex, totalIndex + pose.total_residue()-1)
            newEnergy = scorefxn(pose)
                if newEnergy < currentEnergy:
                    currentEnergy = newEnergy
                    copyBackBoneAngles(pose, totalPose, totalIndex, totalIndex + pose.total_residue()-1)
            elif math.exp(-(newEnergy-currentEnergy)/kT) > random.random():
                currentEnergy = newEnergy
                copyBackBoneAngles(pose, totalPose, totalIndex, totalIndex + pose.total_residue()-1)
            else:
                pose.assign(copyPose)
                copyBackBoneAngles(pose, totalPose, totalIndex, totalIndex + pose.total_residue()-1)

        totalIndex += pose.total_residue()
return

```

```

def copyBackBoneAngles(pose, totalPose, start, stop):
    for i in range(1, stop-start):
        totalPose.set_phi(start+i-1, pose.phi(i))
        totalPose.set_psi(start+i-1, pose.psi(i))
        totalPose.set_omega(start+i-1, pose.omega(i))
    totalPose.set_phi(stop, pose.phi(stop-start))
    totalPose.set_psi(stop, pose.psi(stop-start))

def insertBackBoneAngles(pose, totalPose, start, stop):
    for i in range(1, stop-start):
        pose.set_phi(i, totalPose.phi(start+i-1))
        pose.set_psi(i, totalPose.psi(start+i-1))
        pose.set_omega(i, totalPose.omega(start+i-1))
    pose.set_phi(stop-start, totalPose.phi(stop))
    pose.set_psi(stop-start, totalPose.psi(stop))

def insertAllBackBoneAngles(poseList, totalPose):
    #Read the backbone angles from totalPose back into the original poses
    totalIndex = 0
    for pose in poseList:
        for i in range(1, pose.total_residue()+1):
            totalIndex += 1
            #print str(i) + '/' + str(totalIndex)
            if i == pose.total_residue():
                pose.set_phi(i, totalPose.phi(totalIndex))
                pose.set_psi(i, totalPose.psi(totalIndex))
            else:
                pose.set_phi(i, totalPose.phi(totalIndex))

```

```

        pose.set_psi(i, totalPose.psi(totalIndex))
        pose.set_omega(i, totalPose.omega(totalIndex))
        #chi angles are not included in fragment insertion file
    return

def prependNames(poseList, prependString):
    for pose in poseList:
        pose.pdb_info().name(prependString + pose.pdb_info().name())
    return

def getPMM():
    pmm = r.PyMOL_Mover()
    #pmm.keepHistory(True)
    return pmm

def updatePyMOLPoses(poseList, pmm):
    for pose in poseList:
        pmm.apply(pose)

def disulfideDistance(pose, pair):
    return (pose.residue(pair[0]).xyz(6) - pose.residue(pair[1]).xyz(6)).norm

#Initial code from: D060_Folding.py in PyRosetta/test/
# locates all cysteines, make bonds between them, output the bonds that lower
# the score

```

```

def guess_disulfides(pose, cutoff = 10.0):
    """
    A quick method for probing a protein for cysteine residues close to each
    other (within <cutoff> )
    """
    disulfide = False

    # find all cysteine residues and consider possible disulfides
    cys = [i for i in range(1, pose.total_residue() + 1) if pose.residue(i).name1() == 'C']
    partners = [0]*sum( range(len(cys)) ) # all disulfides possible
    i = 0

    # create all combinations
    for first in range(len(cys[:-1])):
        for second in cys[first + 1:]:
            partners[i] = (cys[first], second)
            i += 1

    # try each disulfide, if it lowers the score, print it to screen
    print '='*80
    print 'Potential Disulfides:'
    for pair in partners:
        # for a fullatom cysteine in PyRosetta, the 6th atom is sulfur
        separation = (pose.residue(pair[0]).xyz(6) - pose.residue(pair[1]).xyz(6)).norm
        if separation < cutoff:
            print 'between (pose numbered) residue %d and %d | %4g Angstrom separation'%(pair[0],pair[1],
            separation)
            r.form_disulfide(pose.conformation(), pair[0], pair[1])
            disulfide = True
    print '='*80
    if disulfide:

```

```

    pose.pdb_info().name(pose.pdb_info().name() + '_disulfideBond')
# to manipulate disulfide bonds, use:
# formation:  form_disulfide(pose.conformation(), 6, 16)
# cleavage:  change_cys_state(6, 'CYS', pose.conformation() )
#           change_cys_state(16, 'CYS', pose.conformation() )

def prefold(poseList, knownDisulfideBondList, iterNum):
    disulfideLength = 4.5
    bondDict = dict()
    for disulfideList in knownDisulfideBondList:
        bondDict[disulfideList.name] = disulfideList

    for pose in poseList:
        disulfidePairList = bondDict[pose.pdb_info().name()]
        #all residues between cysteines can be moved
        ntermC = min([min(pair) for pair in disulfidePairList.pairs])
        ctermC = max([max(pair) for pair in disulfidePairList.pairs])
        moveList = range(ntermC, ctermC+1)
        numPairs = 0
        print "Pose: " + pose.pdb_info().name()
        for pair in disulfidePairList.pairs:
            if pair[0] == 0 or pair[1] == 0:
                continue

            maxIter=10
            iteration = 0
            for iteration in range(maxIter):
                if disulfideDistance(pose, pair) <= disulfideLength:
                    break

```

```

print "\tDisulfide Pair: " + str(pair) + "\tOptimization Iteration: " + str(iteration+1)
moveOrdering = MoveOrder('shuffleOrder', pose)
for residue in moveOrdering.moveList:
    if residue in moveList:
        Move(pose,residue,6,pair)
        if disulfideDistance(pose, pair) <= disulfideLength:
            break
    if disulfideDistance(pose, pair) > disulfideLength:
        pose.pdb_info().name('prefoldFail' + pose.pdb_info().name())
        print "Pre-fold failure: ran out of iterations"
        continue
    if pair[0] != 0 and pair[1] != 0:
        r.form_disulfide(pose.conformation(), pair[0], pair[1])
        numPairs += 1
        print "Disulfides formed: " + str(numPairs)
        for residue in pair:
            if residue in moveList:
                moveList.remove(residue)
prependNames(poseList, 'prefold_')
#pdbDump(poseList, str(iterNum),coarseFastRelax.get_scorefxn())

def fragmentToAnnealed(poseList, i, fragset9, fragset3, knownDisulfideBondList):
    fragmentMover(poseList, fragset9, fragset3, i, knownDisulfideBondList)
    coarseRelaxPose(poseList, i)
    relaxPose(poseList, i)

return

```

```

def createDisulfideBonds(poseList, knownDisulfideBondList):

    #knownDisulfideBondList is a list with an entry for each pose, where the entry is a list of binary
    tuples of residue numbers

    #sorting by name is to match the poses with the correct disulfide bonds

    bondDict = dict()

    for disulfideList in knownDisulfideBondList:

        bondDict[disulfideList.name] = disulfideList

    for pose in poseList:

        if pose.pdb_info().name() in bondDict:

            disulfideList = bondDict[pose.pdb_info().name()]

            for pair in disulfideList.pairs:

                if not 0 in pair:

                    print pair

                    r.form_disulfide(pose.conformation(), pair[0], pair[1])

    return

def getDisulfideBonds(disulfideFile):

    #disulfideFile is a txt file in which each entry begins with a > and a name (which will be used to match
    with the name of the pose)

    #The data line is of the form: x,y; ... where x and y are residue numbers of cysteines that form a
    disulfide bond and ; separates pairs.

    #the function returns a disulfideList, a list of sequence name in the order in the file with the known
    disulfide pairs

    disulfideList = list()

    #Open file

    f = open(disulfideFile)

```

```

fLines = f.readlines()

#Make each dataline a DisulfidePairs object with a name

for line in fLines:

    if line[0] == '%' or line[0] == '\n':

        continue

    if line[0] == '>':

        seqName = line.split('>')[1].rsplit('\n')[0]

    else:

        if line.startswith('na'):

            disulfideList.append(DisulfidePairs(seqName,[]))

        else:

            disulfidePairArray = line.rsplit('\n')[0].split(';')

            pairs = [(int(pair.split(',')[0]), int(pair.split(',')[1])) for pair in disulfidePairArray]

            #Add it to the sequence list

            disulfideList.append(DisulfidePairs(seqName,pairs))

return disulfideList

#####
#####

#####

###Script

for fileTag,disulfideFile,fragmentFile9,fragmentFile3 in
zip(fileTagList,disulfideTagList,fragmentFile9List,fragmentFile3List):

    seqList = fromFASTAtoSeqList(fileTag + '.fasta')

    for seq in seqList:

        print seq.name

        print seq.seq

poses = fromSeqListToPoseList(seqList)

```

```

knownDisulfideBondList = getDisulfideBonds(disulfideFile)
start = time.time()

fragset9 = r.ConstantLengthFragSet(9)
fragset9.read_fragment_file('fragments/' + fragmentFile9)
fragset3 = r.ConstantLengthFragSet(3)
fragset3.read_fragment_file('fragments/' + fragmentFile3)

iterStart = 1
iterEnd = 1000
iterNum = iterEnd - iterStart + 1
print "Starting " + str(iterNum) + " modeling iterations"
finTimeFile = 'finishFile.txt'
finTime = open(finTimeFile,'w')
finTime.close()

for i in range(iterStart,iterEnd+1):
    finTime = open(finTimeFile,'a')
    copiedList = copyPoseList(poses)
    fragmentToAnnealed(copiedList, i, fragset9, fragset3, knownDisulfideBondList)
    #print "\tModeling Iteration: " + str(i) + " of " + str(iterNum)
    #print "\t\tElapsed Time: %.4f minutes"%((time.time()-start)/60)
    #print "Time Left: %.4f minutes\n\n"%(((time.time()-start)*(iterNum-i)/float(60))/float(i))
    finTime.write("Modeling Iteration: " + str(i) + " of " + str(iterNum) + "\n")
    finTime.write("Elapsed Time: %.4f minutes\n\n"%((time.time()-start)/60))
    finTime.write("Time Left: %.4f minutes\n\n"%(((time.time()-start)*(iterNum-i)/float(60))/float(i)))
    finTime.close()

```