Relationship Between Biophysical Properties of Antimicrobial Peptides (AMPs) and their Associated Drug Efficacies

Alisha Zhu

❋ Abstract

Antibiotic resistance is a growing global health threat. One consequence is that patients with cystic fibrosis (CF) are prone to developing antibiotic resistant lung infections caused by multiple strains of bacteria, including *Pseudomonas aeruginosa*. Due to the limited number of treatment options for patients with chronic antibiotic resistant infections, there is a need for finding new antibiotics that allow for effective eradication of bacterial infections, such as those in the CF lung. Many antimicrobial peptides (AMPs) have been annotated in databases and are considered as potential alternatives for current antibiotics. However, in many instances, the suitability of AMPs as drug molecules has not been extensively explored. Here, we propose that certain molecular properties of AMPs favor high antibiotic efficacy. Using information from AMP databases, we combined statistical analyses and machine learning techniques to identify relationships between various biophysical properties of AMPs and their drug efficacies. Analyses from classification and regression trees (CART) and random forests suggest that net charge and maximum average hydrophobic moment are the most important properties in determining if a peptide is useful against *P. aeruginosa* infections in CF patients. Maximum average hydrophobic residue, average alpha helix propensity score, hydrophobic proportion, and peptide length still contribute to this determination but to lesser degrees. Cation-pi interactions, on the other hand, do not appear to factor into this decision at all. Based on these properties, our current work is focused on designing and experimentally testing new peptides that may have activity against *P. aeruginosa* infections.

2 Introduction

Antibiotic resistance is a persistent issue in patient care that affects more than 2.8 million individuals, with an estimated occurrence of ~35,000 deaths in the U.S. each year (CDC, 2019). Deemed as a serious threat by the CDC, *Pseudomonas aeruginosa*, a type of rod-shaped, Gram-negative bacteria, is a major source of infection that requires greater attention. Gram-negative bacterial infections tend to be difficult to treat because they have an extra outer protective membrane layer. Due to buildup of mucus and formation of bacterial biofilms that prevent the delivery of antibiotics at the levels needed for bacteria eradication, patients with cystic fibrosis (CF) are susceptible to developing antibiotic resistant lung infections caused by bacteria, of which the most common in adults is *P. aeruginosa*. CF is a recessive genetic disorder that arises from mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Loss of function of the CFTR decreases hydration in the airways, leading to decreased mucociliary clearance and bacterial retention (Allen et al., 2020). Consequently, 80% of CF patients develop chronic infections (Moreau-Marquis et al., 2008). Thus, there is a need for the discovery of new antibiotic agents. Antimicrobial peptides (AMPs) are a class of small peptides that exist in nature. Because they exhibit a broad range of activity against various bacteria, AMPs hold potential as small molecule antibiotics (Ageitos et al., 2017).

AMPs have various modes of action. First, AMPs can directly kill bacteria by disrupting the cell membrane. AMPs can destabilize and permeabilize the bacterial cell membrane through both hydrophobic and electrostatic interactions. In many cases, the positive net charge of AMPs allows them to interact with the negatively charged regions of the

bacterial surface. These interactions often lead to the formation of membrane pores resulting in lysis, or the rupture of the cell membrane (Li et al., 2021). Second, AMPs can prevent the formation of bacterial cell walls by blocking peptidoglycan elongation. Peptidoglycan is a polymer consisting of sugar chains interlinked with peptides. Since the bacterial cell wall is an essential protective barrier and reinforces cell shape, inhibiting cell wall biosynthesis effectively kills the bacteria (Chen et al., 2020). Third, AMPs can penetrate the cell membrane and act on intracellular targets, including RNA, DNA, and ribosomes. Although this mode of action does not result in cell lysis, it interferes with bacterial bioprocesses. For instance, some AMPs block translation initiation by binding to the P-site on ribosomes, while others inhibit protein synthesis by preventing RNA splicing post-transcription (Ageitos et al., 2017).

Promising AMPs proposed in literature, particularly cationic antimicrobial peptides (CAMPs)—i.e., AMPs that have an overall net positive charge—suggest that certain characteristics of peptides favor activity against Gram-negative bacterial infections, such as those caused by *P. aeruginosa* (Lei et al., 2019). Although AMPs have been documented in databases, in many cases, their microbiological activity is not established. The rules governing which peptides are likely to be effective are not understood, especially for a specific species of bacteria such as *P. aeruginosa* (Ramazi et al., 2022). Thus, we propose to use database analysis to explore the relationship between biophysical properties of AMPs and their corresponding effectiveness. We define effectiveness based on a peptide's minimum inhibitory concentration (MIC), which is the lowest concentration of peptide that prevents visible bacterial growth. We selected the Database of Antimicrobial Activity and Structure of Peptides (DBAASP) for our studies. A set of candidate peptides designed to be efficacious will be identified and tested against clinical strains of *P. aeruginosa*.

3   Methodology

Database Selection and Filtering

Numerous databases have been constructed to provide classified information for productive AMP design and research. The DBAASP is one of the larger AMP databases, since it contains information on both synthetic and natural peptides (Pirtskhalava et al., 2021). The inclusion of both synthetic and natural peptides widens the scope of antimicrobial research and development by allowing for a more comprehensive exploration of the peptide design space. At the time of use, the DBAASP contained information on 18,000+ peptides. However, the complete DBAASP data set file contained 150,000+ entries; several peptides had multiple data entries containing results from different experiments. We implemented initial conditions to acquire a suitable subset. For the DBAASP, selected peptides must be monomers, cannot contain any D-amino acids (as we do not wish to work with synthetic amino acids), and cannot contain any uncommon or variable amino acids. These conditions ensure that the subset consists of peptides, not proteins, that are more likely to be compatible with biological systems. In addition, the peptides must have lengths less than or equal to 50 amino acids but also greater than or equal to 11 amino acids. Shorter lengths enable AMPs to have greater efficiency and flexibility when interacting with bacterial cell membranes, which is why an upper limit for peptide length was specified. However, the AMPs also had to have lengths of at least 11 amino acids long, since this represents three turns of an alpha helix spanning the minimum thickness of a Gram-negative bacterial cell membrane. Most importantly, the peptide must have known activity against *P. aeruginosa*, but its MIC value cannot exceed 500 µM. A high MIC value indicates that more peptide is required for inhibiting growth; we defined a MIC > 500 µM as ineffective for treating *P. aeruginosa* infections. If a peptide's reported MIC was a lower threshold value (i.e., greater than some value), then we considered its MIC value to be double its lower threshold. This was done to account for one more factor of two in the two-fold serial dilution performed in a MIC assay. If a peptide's reported MIC was an upper threshold value (i.e., lower than some value), then we considered its MIC value to be equal to its upper threshold. If a peptide had numerous reported MIC values, the lowest MIC was

selected. After these initial conditions were implemented, we arrived at a subset of 4,218 unique peptides.

MOLECULAR PROPERTIES OF BIOPHYSICAL SIGNIFICANCE

Based on what is known about AMPs, there are several molecular properties that are relatively simple to compute and are likely to influence AMP activity:

PEPTIDE LENGTH. The peptide length is the total number of amino acid residues in the peptide sequence. AMPs are typically quite short; most of them have lengths less than 50 amino acids, but some can have lengths as large as 100 amino acids (Ageitos et al., 2017).

NET CHARGE. A peptide's net charge is calculated by subtracting the total number of negatively charged amino acid residues from the total number of positively charged amino acid residues.

$$Net\ Charge = Lys + Arg - Asp - Glu \quad (1)$$

Net charge indicates how an AMP will electrostatically interact with the bacterial membrane. Studies have shown that decreasing the net charge to less than +4 rendered peptides inactive, and increasing net charge improved antimicrobial activity. However, an increase to net charges greater than +9 led to a dramatic rise in hemolytic activity (Jiang et al., 2008). This is undesirable because greater hemolytic activity leads to red blood cell damage and can result in anemia, inflammation, and organ dysfunction.

HYDROPHOBIC PROPORTION. The hydrophobic proportion refers to the percentage of hydrophobic amino acid residues within the peptide sequence. Hydrophobic residues facilitate interactions with the phospholipid fatty acyl chains of the bacterial membrane (Edwards et al., 2016). AMPs typically contain 40%-60% hydrophobic residues (Ye et al., 2019). We considered an amino acid to be hydrophobic if it exceeds 0.700 on Fauchere and Pliska's (1983) hydrophobicity scale based on octanol-water partition coefficients, which is defined as the ratio of the peptide's concentration in the octanol phase to its concentration in the aqueous phase. If the coefficient

value is greater than one, then the peptide is more lipophilic (dissolves in lipids); if the coefficient value is less than one, then the peptide is more hydrophilic (dissolves in water).

$$H_{prop} = \frac{Cys+Phe+Ile+Leu+Met+Pro+Val+Trp+Tyr}{Peptide\ Length} \quad (2)$$

MAXIMUM AVERAGE HYDROPHOBIC MOMENT. The hydrophobic moment is a measure of the amphiphilicity, or asymmetry of hydrophobicity, of a peptide's structure. A large hydrophobic moment corresponds to a structure that is primarily hydrophobic on one side and primarily hydrophilic on the other side. Eisenberg et al. defined the hydrophobic dipole moment as

$$\mu = \sqrt{[\sum_i H_i \cos(i\delta)]^2 + [\sum_i H_i \sin(i\delta)]^2} \quad (3)$$

where δ represents the angle between the amino acid side chains (δ = 100 for an alpha helix), *i* represents the residue number in the *i*th position, and H_i represents the *i*th amino acid's hydrophobicity based on an averaged consensus scale (Eisenberg et al., 1982, 1984). We modified this equation to determine the maximum average hydrophobic moment across an 11 amino acid window, which represents three turns of an alpha helix, with δ ranging from 90° to 110° instead of fixing it at 100°. By adding these modifications, we hope to account for peptides that have alpha helical segments but may not have a complete alpha helical structure. Additionally, instead of using the averaged consensus scale, values from Fauchere and Pliska's (1983) scale were used.

$$< \mu_{max} \\ > = \frac{\sqrt{[\sum_i H_i \cos(i\delta)]^2 + [\sum_i H_i \sin(i\delta)]^2}}{11} \quad (4)$$

MAXIMUM AVERAGE HYDROPHOBIC RESIDUE. The hydrophobic residue calculation provides a measure of the average hydrophobicity across an 11 amino acid window. As in the maximum average hydrophobic moment calculation, H_i represents the *i*th amino acid's hydrophobicity based on Fauchere and Pliska's scale (1983).

$$< H_{max} > \; = \frac{\sum_i H_i}{11} \quad (5)$$

AVERAGE ALPHA HELIX PROPENSITY SCORE. Alpha-helices are the most common secondary structure seen in AMPs. Studies have demonstrated that increased alpha-helical content correlates to greater antimicrobial activity (Li et al., 2021). The alpha helix propensity score describes how likely a peptide will form an alpha helix. Pace and Scholtz (1998) created a helix propensity scale based on experimental studies of proteins and peptides. The propensity score is given in terms of the amount of energy required for a given amino acid to fold into an alpha helix, with a higher propensity score corresponding to a smaller probability that the peptide will form an alpha helical shape. We modified Porto et al.'s (2018) alpha helix propensity score calculation by dividing it by peptide length to standardize the propensity score across the entire peptide.

$$< \alpha > \; = \frac{\sum_i e^{H_{x_i}}}{Peptide\ Length} \quad (6)$$

Hxi represents the ith amino acid's helix propensity, based on the Pace–Schols scale.

CATION-PI INTERACTIONS. Cation-pi interactions between the positively charged side chain of an arginine amino acid residue and the aromatic side chain of a tryptophan amino acid residue are of interest because studies suggest that tryptophan containing peptides have higher affinity for deeper insertion into bacterial cell membranes (Hicapie et al, 2018). Assuming that the peptide adopts an alpha helix secondary structure, a cation-pi interaction exists only for the configuration where an arginine is four residues after a tryptophan (Shi et al., 2022). We computed a simple count for the number of cation-pi interactions within a peptide sequence.

$$Trp \rightarrow Arg\ (i,\ i+4) \quad (7)$$
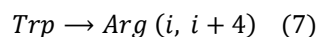
STATISTICAL ANALYSES AND MACHINE LEARNING TECHNIQUES

SCATTERPLOT MATRIX. Using the psych R package, a scatterplot matrix was generated to visualize the pairwise relationships between the selected biophysical features. A scatterplot matrix helps to graphically and quantitatively identify if multicollinearity exists among the various molecular properties of interest.

PRINCIPAL COMPONENT ANALYSIS (PCA). As a dimensionality reduction technique, PCA is appropriate for our purposes because the DBAASP subset is a large data set in which each observation contains a high number of features (each peptide is associated with seven molecular properties). PCA was conducted using the factoextra R package and is accomplished by linearly transforming the data onto a new coordinate system. Information is projected onto a two-dimensional space, increasing interpretability while preserving the maximum amount of information (Jolliffe & Cadima, 2016).

DECISION TREES. Classification and Regression Tree (CART) is a predictive modeling approach. It illustrates how peptide effectiveness (response variable) can be predicted based on the seven biophysical properties (explanatory variables). While classification trees are used for predicting a qualitative outcome, regression trees are used for predicting a quantitative outcome (Krzywinski & Altman, 2017). Using the rpart and rpart.plot R packages, we generated CART for the DBAASP subset. For the classification tree, the peptides were separated into two classes based on a MIC threshold that we specified: noneffective (MIC ≥ 4 μM) and effective (MIC < 4 μM). For the regression tree, the peptides were partitioned based on their biophysical properties, and an average $\log_2(MIC)$ was reported for each subgroup. $\log_2(MIC)$ was used in place of the MIC to reduce the skewness of the MIC data. $\log_2(MIC)$ was the most sensible transformation to apply because MIC assays are based on two-fold serial dilutions.

RANDOM FORESTS. The random forest algorithm is an ensemble learning method that aggregates multiple decision trees to create more accurate predictions (Brieman, 2001). Like decision trees, random forests can handle both regression and classification tasks. However, it is a more robust approach than CART because it controls overfitting. Functions from the randomForest R package were implemented to construct a random forest

regression model. The model included all seven biophysical properties as the explanatory variables with $\log_2$(MIC) as the response variable. Based on the method used to generate the prediction algorithm by Thomas et al. (2010), 70% of the 4,218 peptides in the DBAASP data set were used as the training set and 30% were used as the testing set. The training set was used to build a random forest regression model that would predict the $\log_2$(MIC) of peptide sequences.

DESIGNING AND TESTING PEPTIDES

DESIGNING PEPTIDES. Based on the features learned from the bioinformatic analyses of the DBAASP, we wanted to design new, synthetic peptide sequences. We generated a set of 50,000 random peptide sequences that had the same length and amino acid distributions as those in our DBAASP subset (FIGURE 1). To control the amino acid distribution, we assigned each amino acid with a probability of occurrence based on the frequency that each appeared in the DBAASP subset. We then computed the seven biophysical properties of interest for all 50,000 peptide sequences. From there, we were able to use the random forest regression model to generate predictions for their $\log_2$(MIC) values.

SELECTING PEPTIDES FOR EXPERIMENTAL TESTING. Our goal was to determine if any of the designed peptides were effective *in vitro* against *P. aeruginosa* infections. Assuming that the peptides with the lowest predicted MICs have the greatest likelihoods for being effective, we decided to examine 20 generated peptide sequences with the lowest predicted $\log_2$(MIC) values more closely. For the 20 peptide sequences, we used the Hemolytic Activity Prediction for Peptides Employing Neural Networks (HAPPENN) tool to calculate the predicted hemolytic activity of the peptides (Timmons & Hewage, 2020). The HAPPENN tool provides a PROB score between 0 and 1. A peptide with a score of 0 is predicted to be most likely non-hemolytic, while a peptide with a score of 1 is predicted to be most likely hemolytic. Because the HAPPENN tool can only make predictions on sequences with lengths between 7 and 35 amino acids long, we were only able to deduce the potential hemolytic activity for 11 out of the 20 peptides. Of those 11 peptides, 8 peptides were predicted to be most likely non-hemolytic. Therefore, we decided to select those 8 peptides for experimental testing.
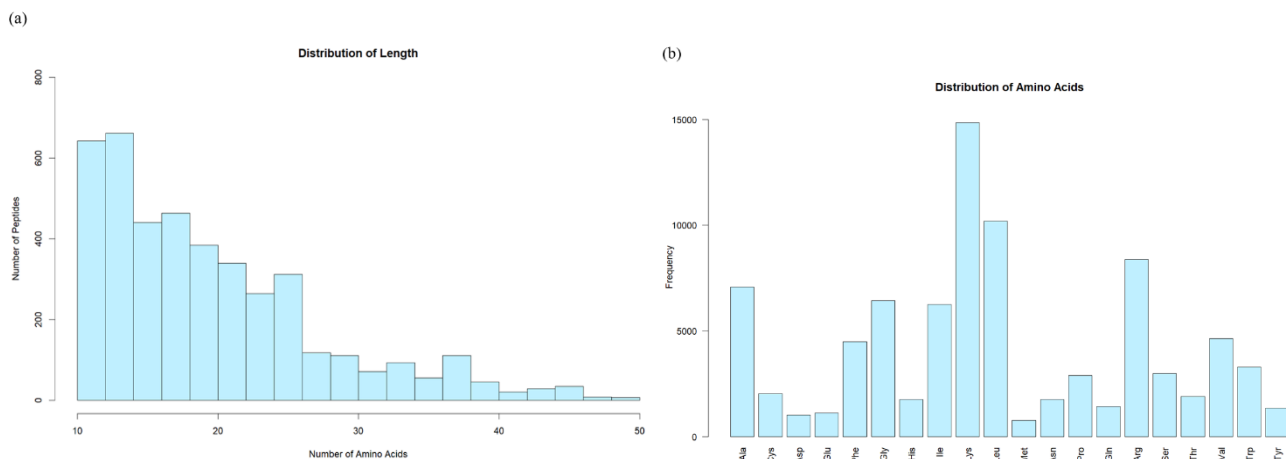

FIGURE 1: *(a) Distribution of lengths for n = 4,218 peptides. (b) Distribution of amino acid residues that make up the n = 4,218 peptides*

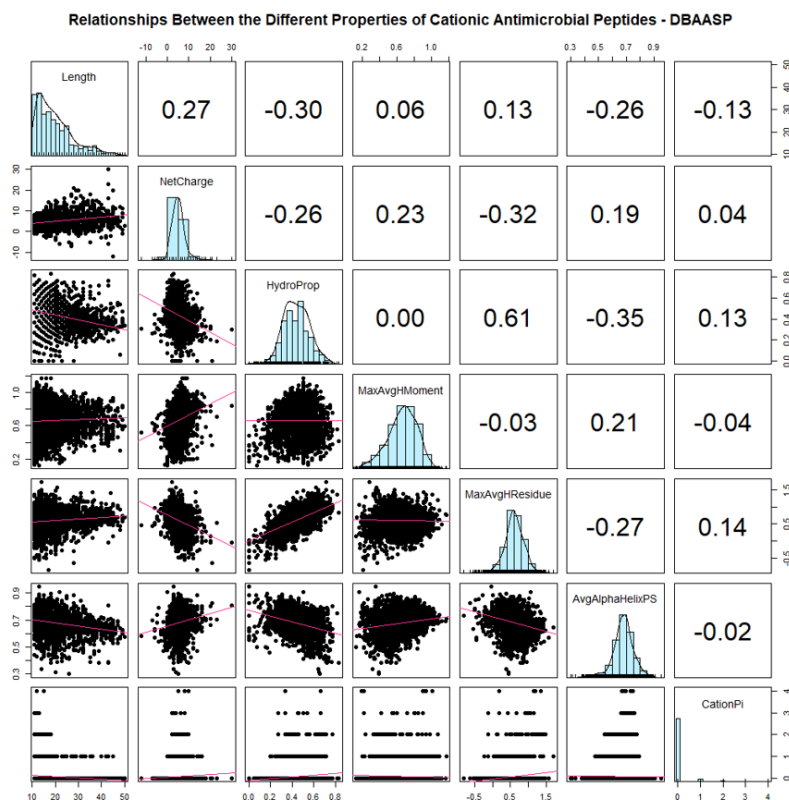| Sequence	Predicted log_2(MIC)
AKRTQRFPRWKKCLRLGFVGCKGNILKAA	-0.175
RVRKGAGTSRKLKIVKNLGRHIVWFKGIP	0.105
IRKYRPGLFAKFHKLLKNRKIGGKNL	0.141
MKSKPTVIMRYRFWRVGKL	0.307
RQACGTKAKARLRKPRCLTIGRRVVRKFSKWR	0.329
KMVAKKVKIKRCKRKVHKLPGFGSISIL	0.335
LIAKYGKHAKFKAKKKPQGISGVPKRFYKALWIG	0.432
PRRIKTGAAKRKPKLSKKWNQKLLKKLTPFGW	0.464

TABLE 1: *List of 8 peptide sequences selected from 50,000 computer generated sequences. These peptides are predicted to be both antimicrobial and nonhemolytic.*

4 RESULTS AND DISCUSSION

RELATIONSHIP BETWEEN MOLECULAR PROPERTIES

The scatterplot matrix illustrates the relationship between each pair of molecular properties (FIGURE 2). Since none of the correlation coefficients exceeds 0.70, it is generally agreed upon that none of the molecular properties is strongly linearly correlated with each other (Ratner, 2009). This demonstrates that the properties are non-redundant, and they may all influence a peptide's effectiveness. However, it is important to note that there exists a moderate, positive correlation between hydrophobic proportion and maximum average hydrophobic residue. This was expected because both hydrophobicity-based calculations were dependent on Fauchere and Pliska's scale.

FIGURE 2: *Scatterplot Matrix. The diagonal panels provide histograms for each molecular property. The panels below the diagonal depict bivariate scatter plots, which provide a visual representation of the relationship between two variables, allowing for pattern identification, outlier detection, and linear correlation assessment. The panels above the diagonal report the Pearson correlation coefficients associated with the regression line in each scatter plot—a value close to +1 or -1 signifies strong correlation while a value close to 0 signifies weak correlation.*

Relationships Between the Different Properties of Cationic Antimicrobial Peptides - DBAASP

The PCA projects information from the seven-dimensional space onto orthogonal axes that are linear combinations and capture the variation of the original variables. The first two dimensions, also called principal components, account for 30.1% and 19.3% (FIGURE 3), respectively, of the total variation in the DBAASP subset. This means that the two-dimensional scatter plot, often used in PCA due to its facility for visual interpretation, is not a faithful representation of the original data in the seven-dimensional space, since it only includes 50% of the variation in the DBAASP data set. To account for greater variation, higher dimensions are needed. Nevertheless, the PCA provides insight into how variation is distributed among the molecular properties. The PCA shows that the greatest contributors to the variation are length, hydrophobic proportion, and maximum average hydrophobic residue.

(a)

(b)

FIGURE 3: *Principal Component Analysis. (a) Plot of molecular properties and their respective contributions. (b) Biplot of individual peptide data points and their molecular properties.*

(a)

(b)

FIGURE 4: *CART. For both classification and regression trees, each decision node is associated with a condition. If the condition is true (or yes) for a peptide, the peptide will be sorted down the left branch and either reach a terminal node or proceed to the next condition. (a) Classification tree built for an n = 4,218 sample, where partitions are based on MIC = 4 μM threshold. Each box displays the overall classification at that node, and the left value represents the number of peptides that are effective while the right value represents the number of peptides that are not effective. (b) Regression tree built for an n = 4,218 sample, where the response variable is log2(MIC). Each node box displays the predicted log2(MIC) value and the n number of peptides that contributed to that predicted value.*

RELATIONSHIP BETWEEN MOLECULAR PROPERTIES AND PEPTIDE EFFECTIVENESS

The conditions used in the classification tree to split the DBAASP subset indicate that all the molecular properties, except for cation-pi interactions, are important in partitioning the data. For classification purposes, we defined a peptide as effective if

its MIC value for *P. aeruginosa* is less than 4 μM. Based on the classification tree that we generated, it appears that the lowest ratio of noneffective to effective peptides (69 to 139) is associated with a group of peptides that satisfies the following conditions: net charge ≥ 4.5, maximum average hydrophobic moment ≥ 0.6981, and length ≥ 29.5 amino acids (FIGURE 4A). Regarding the regression tree that we generated, the subgroup of peptides with the lowest average $\log_2$(MIC) is composed of 208 peptides and satisfies the following conditions: net charge ≥ 4.5, maximum average hydrophobic moment ≥ 0.6977, and length ≥ 29.5 amino acids (FIGURE 4B). These conditions are similar to those described for the classification tree, suggesting that the information provided by the two trees agree. Using these conditions, we can make predictions for new peptide sequences that were not included in the DBAASP subset.

 While the decision trees provide simple visualizations and predictive methods, the predictions are not always accurate. By using multiple trees, random forests produce more accurate predictions and provide more reliable information of variable importance. The feature importance attribute of the random forest regression model indicates that net charge and the maximum average hydrophobic moment are the most distinctive biophysical properties influencing peptide effectiveness (Figure 5). The feature importance attribute also confirms that the presence of cation-pi interactions contributes marginally to the fit of the regression model. The other four properties were of similar importance. Overall, the random forest regression model had a mean of squared residuals of 4.409, which suggests that the model is incorrect by 2.1 $\log_2$(MIC) units on average, and the % variance explained by the model is 37.83%. While the fit of the model is lacking, the regression model is still worth exploring, as our goal is not to be able to predict exact MICs, but to develop a selection method for identifying effective peptides based on their biophysical properties.
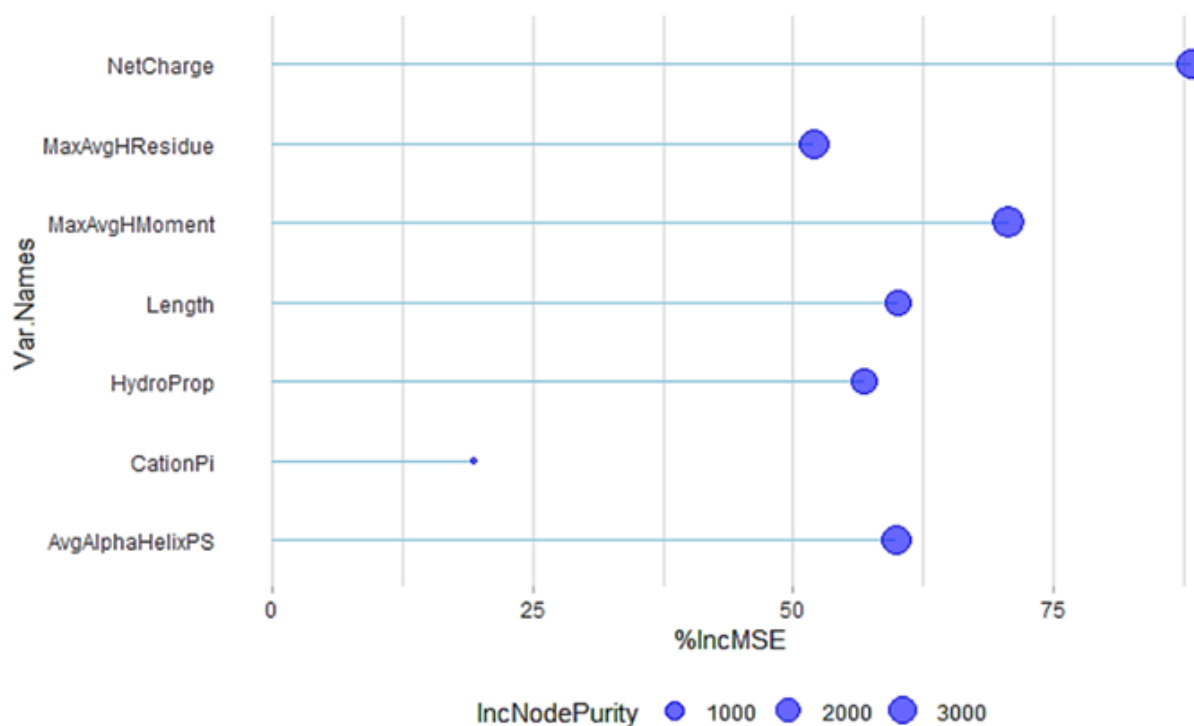

FIGURE 5: *The importance of a biophysical property is determined by measuring the increase in mean square error and the increase in node purity when that property is randomly permuted. Important properties result in substantial changes when randomly permuted, leading to greater increases in mean square error and node purity.*

5 CONCLUSION

AMPs serve as a rich source of alternatives to conventional antibiotics and provide a potential solution for the antibiotic resistance crisis. Existing AMP databases are continuously being updated to account for the expanding knowledge known about them. However, a comprehensive understanding of the relationship between molecular properties of AMPs and their effectiveness against bacterial infections, especially from a clinical perspective, remains elusive. To develop this comprehension and to accelerate the process for screening peptides, this study aimed to create a biophysical understanding and predictive model for classifying peptides as effective or noneffective. To do so, we used a relevant subset of 4,218 AMPs from the DBAASP.

To ensure that the selected molecular properties are not correlated with each other and thus all contribute separately to peptide effectiveness, we constructed a scatterplot matrix. It is apparent that none of the properties are strongly linearly correlated with each other. This was unexpected, considering that three of the seven selected properties were related to the hydrophobicity of the peptide. At the same time, this was promising because it indicates that multicollinearity was not present in our model, and thus each property individually contributes to peptide effectiveness. To further examine the contributions of each molecular property to the total variation within the DBAASP, we performed a PCA. Length, hydrophobic proportion, and maximum average hydrophobic residue had the greatest amount of information captured in the first two dimensions.

We used CART to gain insight into which biophysical properties were most important for deciding peptide effectiveness. Both the classification tree and regression tree reached a consensus that peptides with the highest potential to become efficacious drugs have a maximum average hydrophobic moment ≥ 0.75 and have a net charge ≥ 4. The finding about net charge is not surprising and is consistent with available literature (Jiang et al., 2008).

A predictive regression model using the random forest algorithm was generated to assess the potential antimicrobial activity of a given peptide sequence. This predictive tool was used to evaluate a set of 50,000 computer generated peptide sequences. Based on the predicted antimicrobial and hemolytic activities of those peptide sequences, we identified a set of eight peptides that may have activity against *P. aeruginosa* lung infections. In the future, we wish to assess the accuracy of the random forest predictive tool, and we hope to confirm the activity of those 8 peptides experimentally by conducting MIC assays. Furthermore, we wish to expand upon our understanding of AMPs by exploring other methods, including non-linear dimensionality reduction, PCA regression, and implementing a more formal search for designing peptides through following a genetic algorithm framework∎

6 ACKNOWLEDGEMENTS

7 REFERENCES

[1] Ageitos, J. M., Sánchez-Pérez, A., Calo-Mata, P., & Villa, T. G. (2017). Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria. Biochemical pharmacology, 133, 117–138.
HTTPS://DOI.ORG/10.1016/J.BCP.2016.09.018

[2] Allen, P., Borick, J., & Borick, J. (2020). Acute and Chronic Infection Management in CF. *Cystic Fibrosis in* Primary Care, 69–87.
HTTPS://DOI.ORG/10.1007/978-3-030-25909-9_8

[3] Breiman, L. (2001). Random Forests. Machine Learning 45, 5–32.
HTTPS://DOI.ORG/10.1023/A:1010933404324

[4] Centers for Disease Control and Prevention. (2019). Antibiotic resistant threats in the United States. U.S. Department of Health and Human Services.
HTTPS://WWW.CDC.GOV/DRUGRESISTANCE/PDF/THREATS-RE-PORT/2019-AR-THREATS-REPORT-508.PDF

[5] Chen, C.H.; Lu, T.K. (2020). Development and Challenges of Antimicrobial Peptides for Therapeutic Applications. Antibiotics, 9, 24.

[6] Edwards, I. A., Elliott, A. G., Kavanagh, A. M., Zuegg, J., Blaskovich, M. A., & Cooper, M. A. (2016). Contribution of Amphipathicity and Hydrophobicity to the

Antimicrobial Activity and Cytotoxicity of β- Hairpin Peptides. ACS infectious diseases, 2(6), 442–450. HTTPS://DOI.ORG/10.1021/ACSINFECDIS.6B00045

[7] Eisenberg, D., Weiss, R.M., Terwilliger, T.C., & Wilcox, W. (1982). Hydrophobic moments and protein structure. Faraday Symp. Chem. Soc., 17, 109-120.

[8] Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. Proceedings of the National Academy of Sciences of the United States of America, 81(1), 140–144. HTTPS://DOI.ORG/10.1073/PNAS.81.1.140

[9] Fauchere, J., Pliska, V (1983). Hydrophobic parameters p of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. Eur. J. Med. Chem, 8, 369–375.

[10] Hincapié, O., Giraldo, P., & Orduz, S. (2018). In silico design of polycationic antimicrobial peptides active against Pseudomonas aeruginosa and Staphylococcus aureus. Antonie van Leeuwenhoek, 111(10), 1871–1882. HTTPS://DOI.ORG/10.1007/S10482-018-1080-2

[11] Jiang, Z., Vasil, A. I., Hale, J. D., Hancock, R. E., Vasil, M. L., & Hodges, R. S. (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic alpha-helical cationic antimicrobial peptides. Biopolymers, 90(3), 369–383. HTTPS://DOI.ORG/10.1002/BIP.20911

[12] Jolliffe I.T. and Cadima J. (2016). Principal component analysis: a review and recent developments. Phil. Trans. R. Soc. A.3742015020220150202. HTTPS://DOI.ORG/10.1098/RSTA.2015.0202

[13] Krzywinski, M., Altman, N. (2017). Classification and regression trees. Nat Methods 14, 757–758. HTTPS://DOI.ORG/10.1038/NMETH.4370

[14] Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D. H., & He, Q. (2019). The antimicrobial peptides and their potential clinical applications. American journal of translational research, 11(7), 3919–3931.

[15] Li, S., Wang, Y., Xue, Z., Jia, Y., Li, R., He, C., & Chen, H. (2021). The structure-mechanism relationship and mode of actions of antimicrobial peptides: A review. Trends in Food Science & Technology, 109, 103-115. DOI:10.1016/J.TIFS.2021.01.005

[16] Moreau-Marquis, S., Stanton, B. A., & O'Toole, G. A. (2008). Pseudomonas aeruginosa biofilm formation in the cystic fibrosis airway. Pulmonary Pharmacology & Therapeutics, 21(4), 595-599. HTTPS://DOI.ORG/10.1016/J.PUPT.2007.12.001

[17] Pace, C.N., Scholtz, J.M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. Biophysical Journal 75, 422-427. HTTPS://DOI.ORG/10.1016/S0006-3495(98)77529-0

[18] Pirtskhalava, M., Amstrong, A. A., Grigolava, M., Chubinidze, M., Alimbarashvili, E., Vishnepolsky, B., Gabrielian, A., Rosenthal, A., Hurt, D. E., & Tartakovsky, M. (2021). DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. Nucleic acids research, 49(D1), D288–D297. HTTPS://DOI.ORG/10.1093/NAR/GKAA991

[19] Porto, W.F., Irazazabal, L., Alves, E.S.F. et al. (2018). In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. Nat Commun 9, 1490. doi: HTTPS://DOI.ORG/10.1038/S41467-018-03746-3

[20] Ramazi, S., Mohammadi, N., Allahverdi, A., Khalili, E., & Abdolmaleki, P. (2022). A review on antimicrobial peptides databases and the computational tools. Database: the journal of biological databases and curation, baac011. HTTPS://DOI.ORG/10.1093/DATABASE/BAAC011

[21] Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they?. J Target Meas Anal Mark 17, 139–142. HTTPS://DOI.ORG/10.1057/JT.2009.5

[22] Shi, Z., Olson, C. A., & Kallenbach, N. R. (2002). Cation-pi interaction in model alpha-helical peptides. Journal of the American Chemical Society, 124(13), 3284–3291. HTTPS://DOI.ORG/10.1021/JA0174938

[23] Thomas, S., Karnik, S., Barai, R. S., Jayaraman, V. K., & Idicula-Thomas, S. (2010). CAMP: a useful resource for research on antimicrobial peptides. Nucleic acids research, 38(Database issue), D774–D780. HTTPS://DOI.ORG/10.1093/NAR/GKP1021

[24] Timmons, P. B., & Hewage, C. M. (2020). HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks. Scientific reports, 10(1), 10869. HTTPS://DOI.ORG/10.1038/S41598-020-67701-3

[25] Ye, Z., & Aparicio, C. (2019). Modulation of supramolecular self-assembly of an antimicrobial designer peptide by single amino acid substitution: Implications on peptide activity. Nanoscale advances, 1(12), 4679–4682. HTTPS://DOI.ORG/10.1039/C9NA00498J
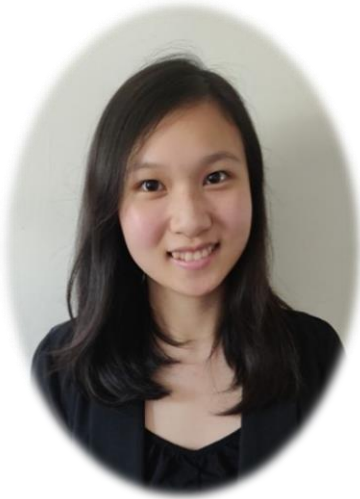
Alisha Zhu is a student in the Honors College at Rutgers University – New Brunswick. She is majoring in Biomedical Engineering and minoring in Statistics. Over the past three years, she has been working as an undergraduate researcher in Dr. Charles Roth's Lab, which she first joined through the Aresty Research Assistant Program. Her research focuses on using computational statistics and machine learning techniques to study the relationship between the biophysical properties of antimicrobial peptides and their drug efficacies. Outside of research, she is the President of the Society of Women Engineers (SWE) and is the External Vice President of the Chinese Student Organization (CSO).

Contact: ALISHA.ZHU@RUTGERS.EDU