THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Antimicrobial Peptides as Broad-Spectrum Therapeutics: Computational Analysis to Identify Universal Physical-Chemical Features Responsible for Multitarget Activity

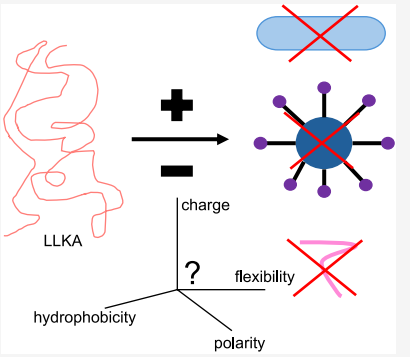
Angela Medvedeva, Catherine Vasnetsov, Victor Vasnetsov, and Anatoly B. Kolomeisky*

Cite This: *J. Phys. Chem. Lett.* 2024, 15, 12416−12424

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Antimicrobial peptides (AMPs) hold significant potential as broad-spectrum therapeutics due to their ability to target a variety of different pathogens, including bacteria, fungi, and viruses. However, the rational design of these peptides requires the molecular understanding of properties that enable such broad-spectrum activity. In this study, we present a computational analysis that utilizes machine-learning methods to distinguish peptides with single-target activity from those with activity against multiple pathogens. By optimizing a feature-selection procedure, the most relevant physical-chemical properties, such as dipeptide compositions, solvent accessibility, charge distributions, and optimal hydrophobicity, that differentiate between narrow-spectrum and broad-spectrum peptides are identified. Possible molecular scenarios responsible for the universality of these features are discussed. These findings provide valuable insights into the molecular mechanisms and rational design of multitarget AMPs.



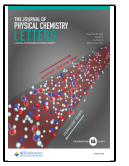
Antimicrobial peptides (AMPs) received significant attention recently as potential therapeutic agents against various pathogens, including bacteria, viruses, fungi, and parasites. These predominantly cationic peptides are an integral part of the innate immune system across many prokaryotic and eukaryotic organisms, with the ability to target multiple pathogens simultaneously.[1,2] Given the rising threat of antimicrobial resistance and the increasing number of infectious diseases, there is a pressing need for novel, efficient therapeutic strategies with minimal side effects. Broad-spectrum AMPs present a promising solution by potentially reducing the toxicity associated with traditional treatments and lowering the probability of resistance development.[3,4]

The design of AMPs with multitargeting capabilities relies heavily on understanding the microscopic properties that enable them to be effective across multiple pathogen classes. Quantitative structure−activity relationship (QSAR) studies have proven invaluable in these efforts, helping to uncover the key molecular features that correlate with antimicrobial activity against single or multiple targets.[5−7] Using statistical and machine learning techniques, QSAR models can predict the biological activity of some peptides based on their chemical structure, streamlining the discovery of AMPs with desirable properties.[8] A critical step in QSAR studies is a feature selection process, which involves identifying the most relevant physical-chemical properties that might be responsible for the antimicrobial effectiveness of peptides.[9−15]

Several characteristics have already been associated with the broad-spectrum activity of AMPs. These peptides are generally cationic and amphipathic, allowing them to interact with negatively charged microbial membranes, thereby compromising membrane integrity.[16,17] Hydrophobic regions enable AMPs to insert themselves into lipid bilayers, leading to membrane disruption and cell death.[18] Structural flexibility further enhances their ability to adapt to different pathogen types, contributing to their multitargeting properties.[19] Additionally, AMPs often target multiple intracellular components, enhancing their effectiveness against a broad spectrum of pathogens.[20,21]

Recent advancements in QSAR modeling have aided the design of AMPs with multitargeting properties. For example, 3D-QSAR models have been successfully employed to predict AMP activity based on AMPs' three-dimensional structures, resulting in the development of peptides with high efficacy against both Gram-positive and Gram-negative bacteria while minimizing toxicity to host cells.[9] Moreover, techniques such as evolutionary algorithms and genetic programming, which mimic natural selection to iteratively improve peptide sequences, have been used to optimize AMP sequences, improving their activity and stability.[8] Despite these advancements, however, significant challenges remain. One of the primary obstacles is the limited availability of high-quality training data, which can improve the performance of machine

Received: November 5, 2024
Revised: December 5, 2024
Accepted: December 9, 2024
Published: December 11, 2024



https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

learning models.[22] Furthermore, the specific features that differentiate broad-spectrum AMPs from narrow-spectrum or single-target AMPs remain not well understood. While significant progress has been made in designing dual-target AMPs,[11,23,24] designing peptides with either broad-spectrum or pathogen-specific activity remains a serious challenge. Studying peptides with experimentally validated multitargeting activity offers critical insights for the rational design of new AMPs. However, a major problem here is the scarcity of data regarding peptides with confirmed multitarget activity across different pathogen classes. While many peptides exhibit broad-spectrum activity, validating their efficacy across all relevant pathogens often requires extensive and labor-intensive experimentation.

The primary aim of this study is to identify universal physical-chemical features that correlate with broad-spectrum activity across different pathogen classes. We focus on classifying peptides as either single-target (e.g., antibacterial, antifungal, antiviral, or antiparasitic) or broad-spectrum (targeting multiple pathogen types). Then, by studying peptides with triple-action activity against bacteria, fungi, and viruses or parasites, we identified the specific physical-chemical features associated with broad-spectrum efficacy. The possible microscopic picture associated with these universal properties is also discussed. The insights gained from this study can potentially inform the rational design of AMPs for future therapeutic development. These findings should also improve understanding of the microscopic mechanisms underlying AMP functionality.

For our analysis, pathogens such as bacteria, fungi, viruses, and parasites, were chosen not only because they are common and well-documented threats, providing ample data for analysis, but also because they represent practical targets for developing antimicrobial peptides.[25−28] We identified the AMPs that are effective against these pathogens. The peptides were classified into two main groups: single-action peptides, which target a single type of pathogen, and triple-action peptides, which are active against multiple pathogens. Triple-action peptides were further categorized into Set 1 (active against bacteria, viruses, and fungi) and Set 2 (active against bacteria, fungi, and parasites). The single-action peptides were grouped by pathogen type: antibacterial, antiviral, antifungal, and antiparasitic. Now we employed a feature selection procedure to identify the properties that have the strongest correlations with the activities of AMPs. This was done using the Least Absolute Shrinkage and Selection Operator (LASSO) method, as detailed in ref 10. For each system, the most relevant physical-chemical properties were determined.

To perform LASSO-based feature selection, we employed a two-stage hyperparameter tuning process. First, we manually tuned hyperparameters by focusing on the regularization parameter α, repeating the manual tuning process (see ref 11 and the detailed section on hyperparameter tuning methods in the Supporting Information). From this process, we selected the value of α that produced the highest classification accuracy. This manually selected value was then used as the input for an automated hyperparameter tuning process, which iteratively refined the model to minimize classification error and further reduce the number of selected features. The automated tuning process adjusted α iteratively to optimize both accuracy and feature selection. The final value of α maximized classification performance while minimizing the feature set. Accuracy plots for all tested values of α are provided in the Supporting Information.

To evaluate the accuracy of our computational approach, we utilized several evaluation metrics, as illustrated in Table 1.

Table 1. Accuracy and Matthews Correlation Coefficient (MCC) for Different Classification Tasks

classification task	*N* features	accuracy	mcc
	Set 1		
antibacterial-only	1547	0.63	0.28
after feature selection	58	0.85	0.70
antiviral-only	1547	0.80	0.60
after feature selection	74	0.88	0.76
antifungal-only	1547	0.78	0.57
after feature selection	39	0.88	0.76
	Set 2		
antibacterial-only	1547	0.67	0.34
after feature selection	45	0.85	0.71
antifungal-only	1547	0.76	0.52
after feature selection	39	0.90	0.80
antiparasitic-only	1547	0.95	0.90
after feature selection	15	0.96	0.93

Matthews correlation coefficient (MCC), which lies between −1 and 1 and is similar to other correlation coefficients, is a robust metric under various circumstances and provides a balanced, minimally biased score; a score close to 1 indicates perfect prediction, while a score close to 0 indicates that the prediction is random, and a score close to −1 indicates that all predictions are the opposite of the true class (e.g., all single-action peptides are incorrectly predicted to be triple-action peptides).[29]

Table 1 shows the accuracy and MCC before and after feature selection procedure. Across all tasks, accuracy and MCC improved significantly after the feature selection. This indicates that the removal of irrelevant or redundant properties enhances the model's ability to properly classify the peptides. These results demonstrate that focusing on a reduced set of features not only simplifies the model but also enhances its predictive power, especially in distinguishing between single-action and triple-action peptides. While previous studies have also highlighted the importance of feature selection in improving classification performance for high-dimensional data sets, the improvement in accuracy observed here was greater overall than in earlier studies employing a similar procedure.[10,11]

Our computational method allowed us to identify the specific physical-chemical properties that are responsible for the broad-spectrum activities of AMPs. Feature importance plots derived from LASSO coefficients are used to understand the relevance of each feature in predicting the class of a peptide as single-action or triple-action. In these plots, the importance of each feature is directly proportional to the absolute value of its coefficient. The coefficients themselves have units corresponding to the scale of the dependent variable and the independent variables. The scale of the LASSO coefficients can vary widely and is not restricted to a fixed range like −1 to 1. Instead, the coefficients represent the strength and direction of the relationship between each feature and the target variable. Larger absolute values of coefficients indicate more significant features, while a coefficient of zero implies that the feature has been excluded from the model due to its lack of relevance. This

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

method allows for the identification of key features that contribute most significantly to the prediction, aiding in the interpretability and refinement of the model. The *x*-axis for the feature importance plots are presented in Figures 1, 2, and 3 is identical for each panel and represents feature importance.

Figure 1 shows the most important features that distinguish single-action from triple-action peptides in Set 1, where triple-action peptides are active against bacteria, viruses, and fungi. Features with positive coefficients are associated with triple-action peptides, while negative coefficients are linked to single-action peptides. Table S1 in the Supporting Information shows the complete list of features and definitions for all classification tasks.

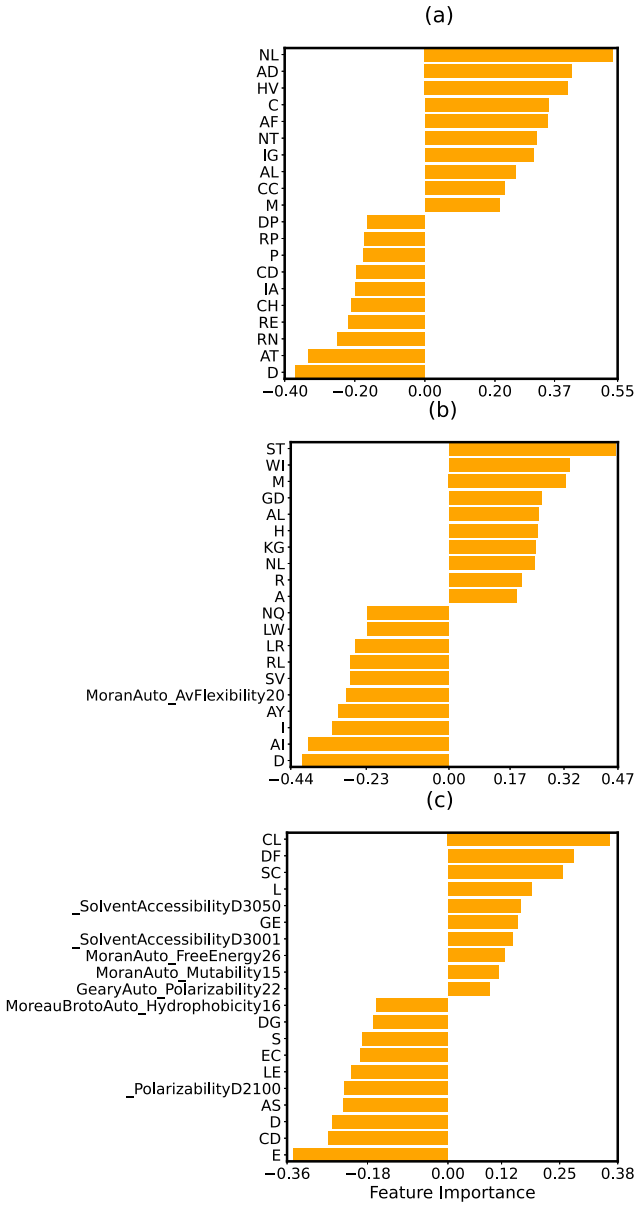


Figure 1. Relative importance of different physical-chemical features distinguishing single-action from triple-action peptides in Set 1 (where triple-action peptides are active against bacteria, viruses, and fungi): (a) antibacterial-only, (b) antiviral-only, and (c) antifungal-only. For each analysis, the top 10 positive and bottom 10 negative features are displayed after sorting the features by their contribution magnitude.

For the antibacterial-only classification (Figure 1a), positively weighted features such as NL (asparagine-leucine dipeptide) and AD (alanine-aspartate dipeptide) suggest that dipeptide composition plays a critical role in determining broad-spectrum activity. These dipeptides have been previously associated with membrane interactions, which are crucial for the initial steps of multipathogen targeting.[30] Additionally, the presence of hydrophobic and amphipathic residues such as valine (HV dipeptide) and cysteine (C) highlights the importance of structural properties that facilitate peptide interactions with different microbial membranes.[21,31] In contrast, negatively weighted features like DP (aspartate-proline) and RP (arginine-proline) are more closely linked to antibacterial-only peptides. Proline-containing dipeptides often disrupt secondary structures, which may limit their effectiveness against multiple pathogen types.[32] The presence of aspartate, which introduces a negative charge, likely reduces the peptide's ability to interact with various pathogen membranes, further confining it to antibacterial activity.[33] In other words, the dominating physical-chemical properties for the single-target peptides are too specific and too strong, and limiting AMPs in being active against other threats.

For the antiviral-only classification (Figure 1b), positively weighted features such as ST (serine-threonine), WI (tryptophan-isoleucine), and M (methionine) are associated with peptides that can target multiple pathogens. These features reflect the structural flexibility and membrane interaction properties required for broad-spectrum activity. Tryptophan, for example, is known for its role in enhancing peptide-membrane interactions due to its hydrophobicity and ability to form strong contacts with lipid bilayers.[34,35] On the other hand, negatively weighted descriptors like LW (leucine-tryptophan) and LR (leucine-arginine) are more characteristic of antiviral-only peptides. These dipeptides likely limit broad-spectrum activity due to their specialized interactions with viral components, reducing their effectiveness against other pathogen types.

In the antifungal-only classification (Figure 1c), descriptors like CL (cysteine-leucine), DF (aspartate-phenylalanine), and SC (serine-cysteine) are positively associated with triple-action peptides. These combinations of hydrophobic and polar amino acids likely contribute to the peptides' ability to function against diverse pathogens. Solvent accessibility features such as *SolventAccessibilityD*3050 and *SolventAccessibilityD*3001, which measure the intermediate solvent exposure of residues, further suggest that surface-exposed residues play a key role in broad-spectrum antimicrobial activity. Specifically, *SolventAccessibilityD*3050 is the fraction of the sequence that contains 50% of amino acids with buried solvent accessibility (M,P,S,T,H,Y), and *SolventAccessibilityD*3001 is the fraction of the sequence that contains the first amino acid with intermediate solvent accessibility (M).

Conversely, negatively weighted features like *Moreau−Broto−Auto−Hydrophobicity*16 (hydrophobicity correlation between amino acids 16 residues apart) and DG (aspartate-glycine) suggest that certain hydrophobic properties, while effective for fungal targeting, may limit the peptide's ability to target other pathogen types.[32] The negative impact of solvent accessibility descriptors indicates that certain residue exposure patterns are associated with more specialized, single-pathogen activity, probably supporting the specificity of such interactions.

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

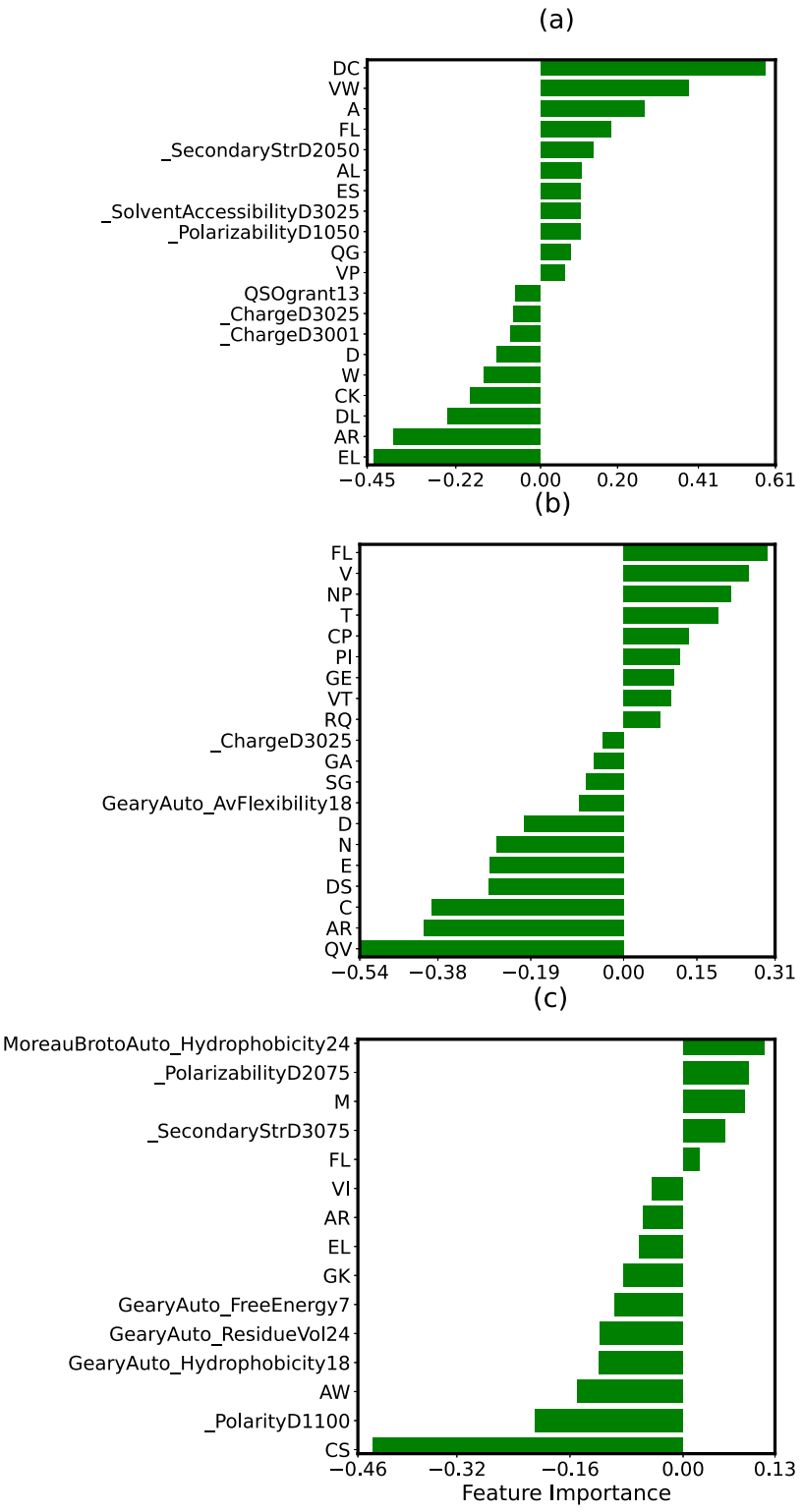


Figure 2. Relative importance of different physical-chemical features distinguishing single-action from triple-action peptides in Set 2 (where triple-action peptides are active against bacteria, fungi, and parasites): (a) antibacterial-only, (b) antifungal-only, and (c) antiparasitic-only. For parts a and b, the top 10 positive and bottom 10 negative features are displayed after sorting the features by their contribution magnitude. For part c, all features are shown.

In Set 2, where triple-action peptides are active against bacteria, fungi, and parasites, the feature importance for antibacterial-only, antifungal-only, and antiparasitic-only classifications (Figure 2) reveals similar patterns of positively and negatively weighted descriptors. For antibacterial-only peptides (Figure 2a), features such as DC (correlation in free energy between residues 12 amino acids apart) and AL (alanine-leucine dipeptide) positively contribute to triple-action activity. The importance of solvent accessibility and secondary structure, as seen in descriptors like *SecondaryStrD*2050 and

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

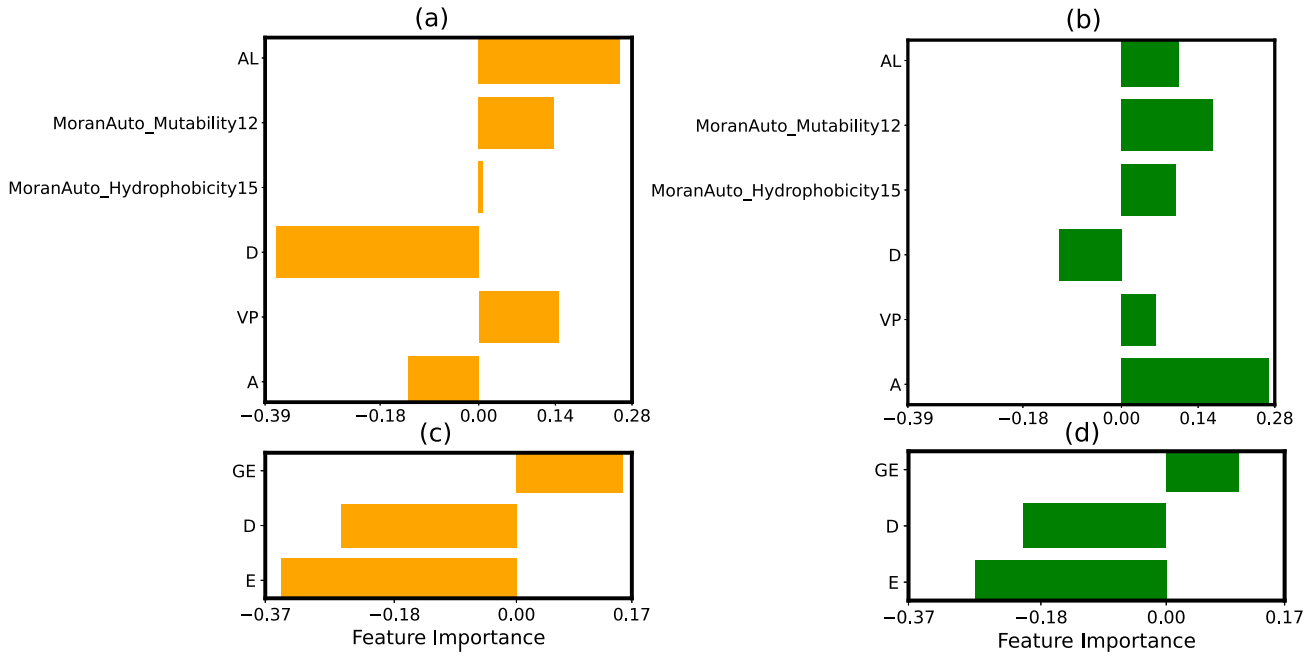


Figure 3. Comparison of overlapping features for antibacterial-only and anti0fungal-only classifications for Triple Action Set 1 and Set 2. Antibacterial-only features from Triple-Action Set 1 in part a (orange bars) and from Triple-Action Set 2 in part b (green bars). Antifungal-only features from Triple-Action Set 1 in part c (orange bars) and from Triple-Action Set 2 in part d (green bars).

*Solvent Accessibility D*3025, emphasizes the need for structural flexibility and surface exposure to target multiple pathogens.[36,37] *SecondaryStrD*2050 is the fraction of the sequence that contains 50% of the amino acids associated with the beta strand secondary structure (group 2: V,I,Y,C,W,F,T) and *SolventAccessibilityD*3025 is the fraction of the sequence that contains 25% of the amino acids associated with intermediate solvent accessibility (group 3: M,P,S,T,H,Y). Negative features like *QSOgrant*13 (quasi-sequence order descriptor based on chemical distance) and *ChargeD*3025 (distribution of negatively charged residues) suggest that charge distribution and specific sequence orders may limit the peptide's ability to target a wide range of pathogens.[38,39]

For antifungal-only peptides (Figure 2b), positively selected features include FL (phenylalanine-leucine dipeptide), VT (valine-threonine dipeptide), and NP (asparagine-proline dipeptide). These features suggest that hydrophobicity and specific dipeptide compositions enhance the peptide's ability to target both fungi and other pathogens. However, negatively weighted descriptors such as RQ (arginine-glutamine) and GA (glycine-alanine) indicate that excessive polar or charged dipeptides may reduce the effectiveness of the peptide against other types of pathogens. In the antiparasitic-only classification (Figure 2c), hydrophobicity and polarizability descriptors, such as *MoreauBrotoAuto_Hydrophobicity*24 (hydrophobicity correlation between amino acids 24 residues apart) and *PolarizabilityD*2075 (fraction of the sequence that contains 75% of the amino acids with polarizability values of 0.128−120.186: C,P,N,V,E,Q,I,L), contribute positively to triple-action classification. Methionine (M), known for its membrane-disrupting capabilities, is also a key feature for broad-spectrum activity in this classification.[40,41]

Figure 3 highlights the shared features between Set 1 and Set 2 when comparing antibacterial-only and antifungal-only peptides against the corresponding triple-action peptides. Set 1 includes peptides that are active against bacteria, viruses, and fungi, while Set 2 contains peptides that are active against bacteria, fungi, and parasites. These properties can be associated with the most universal features that distinguish single-target and multitarget peptides. Note that some of the shared features in Figure 3 are not displayed in Figures 1 and 2 because they are not among the top 10 positive or negative features for those comparisons. However, they still belong to the set of optimal selected features that are associated with the highest classification accuracy for the corresponding task.

Notable universal features include AL (alanine-leucine) and *MoranAuto_ Hydrophobicity*15 (hydrophobicity correlation over 15 residues), which consistently contribute to broad-spectrum activity. These features underscore the importance of optimal hydrophobicity and structural flexibility across different pathogen types.[16,36,37] While the same features appear in both sets, their magnitude and influence vary, reflecting subtle differences in the peptides' interaction with pathogen membranes. For example, the feature D (aspartate) shows a stronger negative influence in Set 1 compared to Set 2, suggesting that charge distribution plays a more critical role in limiting broad-spectrum activity in some cases. Similarly, positively weighted features like VP (valine-proline dipeptide) and *MoranAuto_Mutability*12 (correlation in mutability (susceptibility of an amino acid to mutate during evolution) between amino acids 12 amino acids apart) suggest that hydrophobic and mutable regions within the peptide sequence are fundamental for targeting multiple pathogens. These findings agree with previous studies that highlight the importance of moderate hydrophobicity, appropriate charge distribution, and structural adaptability in broad-spectrum AMPs.[42,43] Features related to solvent accessibility and secondary structure, such as *SecondaryStrD*2050 (the distribution of amino acids associated with the β strand secondary structure) and *SolventAccessibilityD*3025 (the distribution of amino acids associated with intermediate solvent accessibility), also indicate that the accessible surfaces and adaptable

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

secondary structures are key to interacting with a variety of microbial cell walls.[44,45]

The computational analysis identified key physical-chemical features that distinguish broad-spectrum (triple-action) antimicrobial peptides from single-action AMPs using LASSO for feature selection and SVM for classification. By automating the manual hyperparameter tuning process, we systematically optimized the model for feature selection, leading to important insights into the mechanisms of broad-spectrum peptide activity. This approach not only improved the model's ability to classify triple-action peptides but also increased accuracy compared to earlier studies using similar methods.[10,11] The identified features provide valuable insights into how broad-spectrum AMPs function across different pathogen types. The importance of solvent accessibility descriptors, such as *Solvent AccessibilityD*3025 (the distribution of amino acids associated with intermediate solvent accessibility), suggests that greater surface exposure is required for these peptides to interact with diverse microbial membranes. Similarly, secondary structure descriptors like *SecondaryStrD*2050 (the distribution of amino acids associated with the beta strand secondary structure) and *SecondaryStrD*3075 (the fraction of the sequence that contains 75% of the amino acids associated with the coil secondary structure: G,N,P,S,D) highlight the importance of conformational flexibility, which enables these peptides to penetrate and disrupt a range of different membrane types.[46] These results agree with previous findings showing that flexibility and surface exposure are crucial for antimicrobial activity.[17,47]

Our theoretical analysis also suggests that charge distribution plays a critical role in these processes. Negative features such as *ChargeD*3025 (distribution of negatively charged residues) indicate that an even distribution of negative charges across the peptide sequence reduces the broad-spectrum efficacy. This finding is consistent with prior observations showing that positively charged peptides, where charges are clustered together, interact more effectively with negatively charged microbial membranes, emphasizing the importance of balancing charge and hydrophobic properties for multi-pathogen targeting.[36,48,49] Additionally, broad-spectrum AMPs require specific distributions of some properties, including flexibility, polarizability, and hydrophobicity. Descriptors like *GearyAuto_ Polarizability*22 (correlation in polarizability between amino acids 22 amino acids apart) and *MoreauBrotoAuto_Hydrophobicity*24 (hydrophobicity correlation between amino acids 24 residues apart) suggest that peptides with adaptable polarizability and hydrophobicity are better equipped to target multiple pathogen types. This adaptability is essential for breaching structurally distinct pathogen membranes, supporting the peptides' action against bacteria, fungi, and viruses.[36,37,50,51]

A key finding of this study is the overlap in certain features between different sets, such as AL (alanine-leucine dipeptide) and *MoranAuto_Hydrophobicity*15 (hydrophobicity correlation across 15 residues). The consistent presence of these features in both sets suggests that hydrophobicity and structural adaptability are central to broad-spectrum activity. These overlaps indicate that while peptides may differ in the specifics of their pathogen targeting, certain core physical-chemical properties are universally important for targeting a range of pathogens. This finding suggests that the difference between broad-spectrum and narrow-spectrum peptides lies not just in the presence of certain features, such as hydrophobicity, charge distribution, or flexibility, but in how these features are distributed, balanced, and fine-tuned for interaction with diverse microbial membranes. Broad-spectrum peptides appear to have these properties optimized in a way that allows them to adapt and function effectively across different types of pathogens, such as bacteria, fungi, and viruses. In contrast, narrow-spectrum peptides may possess similar features, but these are more specifically tuned for interacting with a single pathogen type, limiting their versatility to target a broader range of microorganisms. Thus, peptides with more generalized properties may have the flexibility to function across multiple pathogen types, whereas peptides with more narrowly defined properties may be optimized for specific targets.

A notable limitation of the study is related to the difficulty of ruling out potential targets for peptides unless demonstrated inactivity is proven. Many peptides may possess activity against other pathogens but may not have been tested beyond a specific pathogen type, such as bacteria. Without evidence of inactivity, we cannot conclusively classify some peptides as single-target in our data sets. As multitargeting peptides become more thoroughly tested, more accurate classifications and feature analysis can be achieved. Additionally, multitargeting peptides, though not inherently rare, are experimentally challenging to verify due to the extensive testing required across multiple pathogens. The relatively small data set used in this study may limit the generalizability of the findings. As more data on triple-action peptides become available, future models could incorporate more sophisticated tuning methods or leverage advanced machine learning techniques, such as deep learning or ensemble methods, to enhance prediction accuracy.

While Least Absolute Shrinkage and Selection Operator (LASSO) remains a widely used method for feature selection, recent developments have introduced several promising alternatives, such as Recursive Feature Elimination (RFE), Boruta, and Shapley Additive exPlanations (SHAP) values. RFE operates by iteratively removing the least important features based on model weights until an optimal feature subset is identified.[52,53] Boruta, an extension of Random Forest, ranks feature importance by comparing each feature's relevance to that of random permutations.[54,55] SHAP values, derived from cooperative game theory, quantify the contribution of each feature to a model's prediction, offering insights into feature importance through marginal contribution analysis.[56,57] However, these traditional methods often face issues like computational inefficiency and potential biases, particularly in high-dimensional data environments.[58]

To address these challenges, more sophisticated techniques, including Deep Neural Network (DNN)-based feature selection, have emerged. DNNs leverage their capacity to learn complex patterns and identify salient features, which enhances both prediction accuracy and model interpretability.[59] Additionally, methods such as mixed-integer programming and local false discovery rate estimation have been developed to improve the robustness and interpretability of DNN-based feature selection.[59,60] These advanced approaches help mitigate overfitting, reduce biases, and provide clearer insights into model behavior. In summary, while traditional methods like those implemented in this study remain valuable, the integration of DNN architectures and optimization techniques represents a promising direction for future feature selection in complex data environments.

It is important to note here that although our theoretical analysis does not fully uncover the microscopic mechanisms of

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424

multitarget activities of some AMPs, it provides an important first step in this direction by identifying specific features that are most probably responsible for these phenomena. Addressing microscopic mechanisms directly will require more advanced theoretical and experimental studies to get more details of the molecular picture. The value of our theoretical method is that it suggests the directions in which these studies should be directed.

In conclusion, this study represents a significant step toward understanding the microscopic picture that defines multitargeting AMPs. By developing and automating the tuning process, we optimized feature selection and identified key physical-chemical properties associated with broad-spectrum activity, emphasizing the universality of certain features. While our approach primarily focused on feature selection, future studies should expand this method to include more specific prediction models that can guide the rational design of multitargeting AMP-based therapies. Ultimately, this work should provide a foundation for understanding the molecular mechanisms of AMPs functioning.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.4c03197.

Additional explanation of methods, results of hyperparameter tuning, and description of the selected features for each analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Anatoly B. Kolomeisky − *Department of Chemistry, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Department of Chemical and Biomolecular Engineering, and Department of Physics and Astronomy, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Angela Medvedeva − *Department of Chemistry, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Catherine Vasnetsov − *Department of Chemistry, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Victor Vasnetsov − *Department of Chemistry, Rice University, Houston, Texas 77005, United States; Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.4c03197

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745). We thank Zlata Krutova and Sofya Domakhina for their contributions to this project.

■ REFERENCES

(1) Chen, T.; Sun, T.; Bian, Y.; Pei, Y.; Feng, F.; Chi, H.; Li, Y.; Tang, X.; Sang, S.; Du, C.; et al. The Design and Optimization of Monomeric Multitarget Peptides for the Treatment of Multifactorial Diseases. *J. Med. Chem.* **2022**, *65*, 3685−3705.

(2) Wiradharma, N.; Sng, M. Y.; Khan, M.; Ong, Z.-Y.; Yang, Y.-Y. Rationally designed α-helical broad-spectrum antimicrobial peptides with idealized facial amphiphilicity. *Macromol. Rapid Commun.* **2013**, *34*, 74−80.

(3) Zhou, C.; Qi, X.; Li, P.; Chen, W. N.; Mouad, L.; Chang, M. W.; Leong, S. S. J.; Chan-Park, M. B. High potency and broad-spectrum antimicrobial peptides synthesized via ring-opening polymerization of α-aminoacid-N-carboxyanhydrides. *Biomacromolecules* **2010**, *11*, 60−67.

(4) Gou, S.; Li, B.; Ouyang, X.; Ba, Z.; Zhong, C.; Zhang, T.; Chang, L.; Zhu, Y.; Zhang, J.; Zhu, N.; et al. Novel broad-spectrum antimicrobial peptide derived from anoplin and its activity on bacterial pneumonia in mice. *J. Med. Chem.* **2021**, *64*, 11247−11266.

(5) Bouarab-Chibane, L.; Forquet, V.; Lantéri, P.; Clément, Y.; Léonard-Akkari, L.; Oulahal, N.; Degraeve, P.; Bordes, C. Antibacterial properties of polyphenols: characterization and QSAR (Quantitative structure−activity relationship) models. *Frontiers in microbiology* **2019**, *10*, 829.

(6) Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in Pharmacology* **2018**, *9*, 1275.

(7) Sadasivam, D.; Nambiar, P.; Dutta, A.; Mitra, D. Rational design of antimicrobial peptides: an optimization approach. *Molecular Systems Design & Engineering* **2024**, *9*, 311−322.

(8) Cardoso, M. H.; Orozco, R. Q.; Rezende, S. B.; Rodrigues, G.; Oshiro, K. G. N.; Cândido, E. S.; Franco, O. L. Computer-Aided Design of Antimicrobial Peptides: Are We Generating Effective Drug Candidates? *Frontiers in Microbiology* **2020**, *10*, 3097.

(9) Li, J.; Hu, S.; Jian, W.; Xie, C.; Yang, X. Plant antimicrobial peptides: structures, functions, and applications. *Botanical Studies* **2021**, *62*, 5.

(10) Teimouri, H.; Medvedeva, A.; Kolomeisky, A. B. Bacteria-specific feature selection for enhanced antimicrobial peptide activity predictions using machine-learning methods. *J. Chem. Inf. Model.* **2023**, *63*, 1723−1733.

(11) Medvedeva, A.; Domakhina, S.; Vasnetsov, C.; Vasnetsov, V.; Kolomeisky, A. Physical−Chemical Approach to Designing Drugs with Multiple Targets. *J. Phys. Chem. Lett.* **2024**, *15*, 1828−1835.

(12) Lee, E. Y.; Wong, G. C.; Ferguson, A. L. Machine learning-enabled discovery and design of membrane-active peptides. *Bioorganic & medicinal chemistry* **2018**, *26*, 2708−2718.

(13) Lee, E. Y.; Fulan, B. M.; Wong, G. C.; Ferguson, A. L. Mapping membrane activity in undiscovered peptide sequence space using machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13588−13593.

(14) Singh, O.; Hsu, W.-L.; Su, E. C.-Y. Co-AMPpred for in silico-aided predictions of antimicrobial peptides by integrating composition-based features. *BMC bioinformatics* **2021**, *22*, 389.

(15) Olcay, B.; Ozdemir, G. D.; Ozdemir, M. A.; Ercan, U. K.; Guren, O.; Karaman, O. Prediction of the synergistic effect of antimicrobial peptides and antimicrobial agents via supervised machine learning. *BMC Biomedical Engineering* **2024**, *6*, 1.

(16) Huan, Y.; Kong, Q.; Mou, H.; Yi, H. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Frontiers in Microbiology* **2020**, *11*, 582779.

(17) Mabrouk, D. M. Antimicrobial peptides: features, applications and the potential use against COVID-19. *Molecular Biology Reports* **2022**, *49*, 10039−10050.

(18) Bhattacharjya, S.; et al. Structure−activity relationship analysis of plant AMPs. *Botanical Studies* **2009**, *50*, 205.

(19) Zhu, Y.; Zhao, L.; Wen, N.; Wang, J.; Wang, C. DataDTA: a multi-feature and dual-interaction aggregation framework for drug−target binding affinity prediction. *Bioinformatics* **2023**, *39*, btad560.
(20) Le, C.-F.; Fang, C.-M.; Sekaran, S. D. Intracellular Targeting Mechanisms by Antimicrobial Peptides. *Antimicrob. Agents Chemother.* **2017**, *61*, e02340-16.
(21) Sharma, S.; Zhang, X.; Azhar, G.; Patyal, P.; Verma, A.; KC, G.; Wei, J. Y. Valine improves mitochondrial function and protects against oxidative stress. *Biosci., Biotechnol., Biochem.* **2024**, *88*, 168−176.
(22) Yan, J.; Cai, J.; Zhang, B.; Wang, Y.; Wong, D. F.; Siu, S. W. Recent progress in the discovery and design of antimicrobial peptides using traditional machine learning and deep learning. *Antibiotics* **2022**, *11*, 1451.
(23) Luo, Y.; Song, Y. Mechanism of Antimicrobial Peptides: Antimicrobial, Anti-Inflammatory and Antibiofilm Activities. *International Journal of Molecular Sciences* **2021**, *22*, 11401.
(24) Parchebafi, A.; Tamanaee, F.; Ehteram, H.; Ahmad, E.; Nikzad, H.; Kashani, H. H. The dual interaction of antimicrobial peptides on bacteria and cancer cells; mechanism of action and therapeutic strategies of nanostructures. *Microbial Cell Factories* **2022**, *21*, 118.
(25) Sabtu, N.; Enoch, D. A.; Brown, N. M. Antibiotic resistance: what, why, where, when and how? *British Medical Bulletin* **2015**, *116*, cr4c00681.
(26) Fishman, J. A. Infection in solid-organ transplant recipients. *New England Journal of Medicine* **2007**, *357*, 2601−2614.
(27) Pariente, N.; Editors, P. B. S. The antimicrobial resistance crisis needs action now. *PLOS Biology* **2022**, *20*, e3001918.
(28) Kramer, A.; Lexow, F.; Bludau, A.; Köster, A. M.; Misailovski, M.; Seifert, U.; Eggers, M.; Rutala, W.; Dancer, S. J.; Scheithauer, S. How long do bacteria, fungi, protozoa, and viruses retain their replication capacity on inanimate surfaces? A systematic review examining environmental resilience versus healthcare-associated infection risk by "fomite-borne risk assessment". *Clin. Microbiol. Rev.* **2024**, *0*, e00186-23.
(29) Boughorbel, S.; Jarray, F.; El-Anbari, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PloS one* **2017**, *12*, e0177678.
(30) Batool, T.; Makky, E. A.; Jalal, M.; Yusoff, M. M. A Comprehensive Review on l-Asparaginase and Its Applications. *Appl. Biochem. Biotechnol.* **2016**, *178*, 900−923.
(31) Väth, K.; Mattes, C.; Reinhard, J.; Covino, R.; Stumpf, H.; Hummer, G.; Ernst, R. Cysteine cross-linking in native membranes establishes the transmembrane architecture of Ire1. *J. Cell Biol.* **2021**, *220*, e202011078.
(32) Benítez-Chao, D. F.; León-Buitimea, A.; Lerma-Escalera, J. A.; Morones-Ramírez, J. R. Bacteriocins: An Overview of Antimicrobial, Toxicity, and Biosafety Assessment by in vivo Models. *Frontiers in Microbiology* **2021**, *12*, 630695.
(33) Doi, Y.; Bonomo, R. A.; Hooper, D. C.; Kaye, K. S.; Johnson, J. R.; Clancy, C. J.; Thaden, J. T.; Stryjewski, M. E.; van Duin, D. Gram-Negative Bacterial Infections: Research Priorities, Accomplishments, and Future Directions of the Antibacterial Resistance Leadership Group. *Clinical Infectious Diseases* **2017**, *64*, S30−S35.
(34) Barik, S. The Uniqueness of Tryptophan in Biology: Properties, Metabolism, Interactions and Localization in Proteins. *International Journal of Molecular Sciences* **2020**, *21*, 8776.
(35) Kanova, M.; Kohout, P. Tryptophan: AUnique Role in the Critically Ill. *International Journal of Molecular Sciences* **2021**, *22*, 11714.
(36) Gagat, P.; Ostrówka, M.; Duda-Madej, A.; Mackiewicz, P. Enhancing Antimicrobial Peptide Activity through Modifications of Charge, Hydrophobicity, and Structure. *International Journal of Molecular Sciences* **2024**, *25*, 10821.
(37) Zhang, Q.-Y.; Yan, Z.-B.; Meng, Y.-M.; Hong, X.-Y.; Shao, G.; Ma, J.-J.; Cheng, X.-R.; Liu, J.; Kang, J.; Fu, C.-Y. Antimicrobial peptides: mechanism of action, activity and clinical potential. *Military Medical Research* **2021**, *8*, 48.
(38) Sitbon, E.; Pietrokovski, S. Occurrence of protein structure elements in conserved sequence regions. *BMC structural biology* **2007**, *7*, 3.
(39) Khrustalev, V. V. Random Coils of Proteins Situated Between a Beta Strand and an Alpha Helix Demonstrate Decreased Solvent Accessibility. *Protein Journal* **2020**, *39*, 308−317.
(40) Chaturvedi, D.; Mahalakshmi, R. Methionine mutations of outer membrane protein X influence structural stability and beta-barrel unfolding. *PLoS One* **2013**, *8*, e79351.
(41) Aledo, J. C. The role of methionine residues in the regulation of liquid-liquid phase separation. *Biomolecules* **2021**, *11*, 1248.
(42) Mechesso, A. F.; Zhang, W.; Su, Y.; Xie, J.; Wang, G. Segment-Based Peptide Design Reveals the Importance of N-Terminal High Cationicity for Antimicrobial Activity Against Gram-Negative Pathogens. *Probiotics and Antimicrobial Proteins* **2024**, *14*, 10376.
(43) Nayab, S.; Aslam, M. A.; ur Rahman, S. u.; ud Din Sindhu, Z. u. D.; Sajid, S.; Zafar, N.; Razaq, M.; Kanwar, R.; Amanullah.; et al. Amanullah A Review of Antimicrobial Peptides: Its Function, Mode of Action and Therapeutic Potential. *International Journal of Peptide Research and Therapeutics* **2022**, *28*, 46.
(44) Walter, A.; Mayer, C. *Extracellular Sugar-Based Biopolymers Matrices*; Springer, 2019; pp 237−299.
(45) Garde, S.; Chodisetti, P. K.; Reddy, M. Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus* **2021**, *9*, ESP-0010-2020.
(46) Singh, O.; Hsu, W.-L.; Su, E. C.-Y. Co-AMPpred for in silico-aided predictions of antimicrobial peptides by integrating composition-based features. *BMC Bioinformatics* **2021**, *22*, 389.
(47) Cavallo, L.; Kleinjung, J.; Fraternali, F. POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res.* **2003**, *31*, 3364−3366.
(48) Moretta, A.; Van Eijk, M. Antimicrobial Peptides: Potential Alternative to Antibiotics and Overcoming Limitations for Future Therapeutic Applications. *Journal of Peptide Science* **2021**, *27*, e10623.
(49) Van Eijk, M.; et al. Antimicrobial Peptides: Mechanism of Action, Activity and Clinical Potential. *Journal of Peptide Science* **2020**, *26*, e10343.
(50) Nooranian, S.; Oskuee, R. K.; Jalili, A. Antimicrobial Peptides, a Pool for Novel Cell Penetrating Peptides Development and Vice Versa. *International Journal of Peptide Research and Therapeutics* **2021**, *27*, 1205−1220.
(51) Karmakar, S.; Das, S.; Banerjee, K. K. Interaction of antimicrobial peptides with model membranes: a perspective towards new antibiotics. *European Physical Journal Special Topics* **2024**, *2024*, 01105-6.
(52) Han, Y.; Huang, L.; Zhou, F. A dynamic recursive feature elimination framework (dRFE) to further refine a set of OMIC biomarkers. *Bioinformatics* **2021**, *37*, 2183−2189.
(53) Priyatno, A. M.; Widiyaningtyas, T.; et al. A Systematic Literature Review: Recursive Feature Elimination Algorithms. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)* **2024**, *9*, 196−207.
(54) Kursa, M. B.; Rudnicki, W. R. Feature selection with the Boruta package. *Journal of statistical software* **2010**, *36*, 1−13.
(55) Kursa, M. B.; Jankowski, A.; Rudnicki, W. R. Boruta−a system for feature selection. *Fundamenta Informaticae* **2010**, *101*, 271−285.
(56) Wang, H.; Liang, Q.; Hancock, J. T.; Khoshgoftaar, T. M. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data* **2024**, *11*, 44.
(57) Kraev, E.; Koseoglu, B.; Traverso, L.; Topiwalla, M. Shap-Select: Lightweight Feature Selection Using SHAP Values and Regression. *arXiv* 2024, arXiv:2410.06815. DOI: 10.48550/arXiv.2410.06815
(58) Strimmer, K. A unified approach to false discovery rate estimation. *BMC bioinformatics* **2008**, *9*, 303.
(59) Cao, Z.; Sun, X.; Fu, Y. Deep neural network-based feature selection with local false discovery rate estimation. *Applied Intelligence* **2025**, *55*, 32.
(60) Zhao, S.; Tsay, C.; Kronqvist, J. Model-based feature selection for neural networks: A mixed-integer programming approach.

International Conference on Learning and Intelligent Optimization **2023**, *14286*, 223−238.

https://doi.org/10.1021/acs.jpclett.4c03197
J. Phys. Chem. Lett. 2024, 15, 12416−12424