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

pubs.acs.org/JPCL Letter

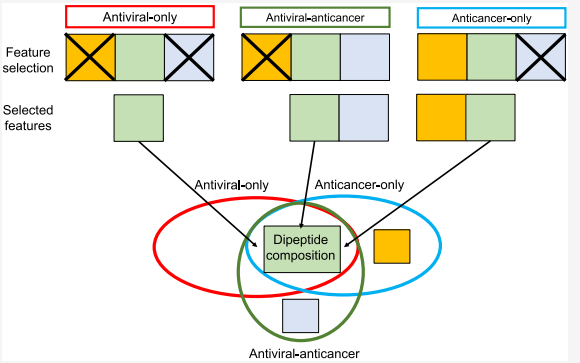
# Physical–Chemical Approach to Designing Drugs with Multiple Targets

Angela Medvedeva, Sofya Domakhina, Catherine Vasnetsov, Victor Vasnetsov, and Anatoly Kolomeisky*

Cite This: J. Phys. Chem. Lett. 2024, 15, 1828–1835 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Many people simultaneously exhibit multiple diseases, which complicates efficient medical treatments. For example, patients with cancer are frequently susceptible to infections. However, developing drugs that could simultaneously target several diseases is challenging. We present a novel theoretical method to assist in selecting compounds with multiple therapeutic targets. The idea is to find correlations between the physical and chemical properties of drug molecules and their abilities to work against multiple targets. As a first step, we investigated potential drugs against cancer and viral infections. Specifically, we investigated antimicrobial peptides (AMPs), which are short positively charged biomolecules produced by living systems as a part of their immune defense. AMPs show anticancer and antiviral activity. We use chemoinformatics and correlation analysis as a part of the machine-learning method to identify the specific properties that distinguish AMPs with dual anticancer and antiviral activities. Physical–chemical arguments to explain these observations are presented.






A constantly growing number of people simultaneously suffer from multiple diseases or chronic conditions such as diabetes, arthritis, hypertension, depression, asthma, and many others.[1] Any combination of diseases and conditions significantly complicates effective medical treatments. One of the most striking examples is that certain types of cancer reduce the number of immune cells available to fight infections, so patients with these diseases may need treatments for the infections as well as the cancer.[2−4] Moreover, medical therapies that work against a specific disease can produce harsh side effects that lead to other medical complications. For example, certain treatments against cancer can also lead to a weakened immune system and greater susceptibility to viral infections.[2] Developing efficient drugs with multiple therapeutic targets remains difficult.

We propose a new theoretical approach to assist in developing new drugs that could be efficient against several diseases. Our idea is that such compounds should exhibit a specific spectrum of physical–chemical properties responsible for such versatility, and this could help in the rational design of new medicines. The goal is to identify the correlations between the most relevant physicochemical features of potential drugs and their abilities to work against multiple targets. For this reason, machine-learning methods that "learn" about these correlations should be an important part of the specific application of the proposed method. To illustrate our approach, we present a first step in a procedure that investigates compounds that simultaneously work against cancer and viral infections. Analyzing the most relevant physical–chemical features that correlate with the dual action of drugs allows us to better understand the underlying microscopic mechanisms of complex biological processes and provides specific directions for selecting new medical approaches.

Cancer is defined as a set of genetic diseases characterized by the uncontrolled growth of mutated cells that hijack the resources from normal cells, invade healthy tissues, and might even lead to the death of the organism.[5] It is typically triggered by DNA mutations that result in the activation of growth-promoting genes and the deactivation of tumor-suppressing genes.[6] One of the negative effects of certain types of cancer is the direct influence on the immune system[3,4] that can make a person much more susceptible to various infections.[7] Certain cancer treatments also negatively affect the immune system and can increase the likelihood of infection.[2] The unchecked spread of cancer is terminal, but a simultaneous viral infection could accelerate this fatal outcome.[8,9] Recent studies suggest that >60% of deaths of cancer patients are due to infections.[10] Infections can also disrupt cancer treatment, necessitating

Received: December 28, 2023
Revised: February 6, 2024
Accepted: February 7, 2024





https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828–1835

careful management to avoid interruptions and potential fatalities.

Viruses are complex infectious pathogens that can enter originally healthy cells and exploit their metabolic processes for their replication.[11] They replace the fraction of the host cell's genome with their genetic material, and cell death occurs when the newly replicated viruses (virions) burst out and cause the host cell to undergo lysis.[12,13] Viruses can be highly lethal because they modify the host cell's functions[14] and immune responses[15] to facilitate rapid spread. Similar to cancer, uncontrolled viral proliferation within a host can be fatal, particularly when the immune system is already compromised due to side effects of cancer treatments such as chemotherapy.[16,17] Medical therapies that can simultaneously address the dangers of both cancer and viruses are desired.

Antimicrobial peptides (AMPs), produced by most living organisms as part of their immune defense, have been proposed recently as potentially powerful antibiotic drugs that might also be useful in the fight against other diseases. They exhibit broad-spectrum activity against various pathogens while maintaining low toxicity toward the host organism's cells, which can mean fewer side effects of potential treatment.[18] These properties are hypothesized to be due to the amphipathic nature of AMPs;[19] i.e., these biopolymers have spatially separated hydrophobic and hydrophilic segments. Specifically, the cationic regions of the AMPs can assist them in binding to the anionic lipids in the pathogen's cell membrane through electrostatic interactions and thus selectively disrupting pathogen but not host cells.[20] Concurrently, the hydrophobic regions of the AMPs enable them to break into the cellular membranes or to enter the pathogenic cell and disrupt their functions with a limited likelihood of inducing resistance,[21] ultimately inhibiting pathogenic cell replication and leading to cell death.[22] Cancer cell membranes contain anionic lipids that can be targeted by AMPs.[23,24] Certain AMPs can also target the distinct cell surface receptors and acidic microenvironment of cancer cells.[25−28] These observations suggest that AMPs might be effectively utilized in fighting both cancer and viral infections.

Experimental studies found that AMPs with antiviral or anticancer abilities are relatively short (20−30 amino acids long),[29] and they can easily adapt $\alpha$-helical or $\beta$-sheet conformations upon electrostatic interaction with the cell membranes.[30,31] These characteristics have been further investigated through computational studies as part of efforts to efficiently screen thousands of known AMPs as potential drug candidates against cancer and viruses. In addition, various machine-learning approaches have been developed to predict whether an AMP would demonstrate only anticancer or only antiviral activity.[32−36] Furthermore, machine-learning-based feature selection approaches have been implemented to identify the most important physical−chemical properties associated with the functioning of AMPs against viruses or cancer.[37] However, very little is known about the features of AMPs that are associated with simultaneous dual anticancer−antiviral activity, although such possibilities have been mentioned.[38,39] There is a huge advantage in finding such AMPs because of the reduced toxicity, lower probability of resistance development, and weaker side effects. This would also potentially lead to more efficient chemotherapy treatments without interruptions due to infections.

To understand the molecular mechanisms of functioning of AMPs as potential drugs, it is important to note that not all molecular properties might be equally relevant for understanding and predicting activity against viruses and cancer. It is reasonable then to utilize a feature selection, which is the process of reducing the number of redundant properties and retaining only relevant ones for a subsequent theoretical analysis. In this paper, we utilize the least absolute shrinkage and selection operator (LASSO) model that performs feature selection by identifying the most significant physicochemical properties (e.g., the correlation between amino acids in hydrophobicity) that contribute to a peptide's activity against viruses and cancer. The selected features can then be used as inputs to a model, such as the support vector machine (SVM), to predict whether a peptide is active against viruses and cancer. The SVM learns the association between the selected features and the peptide type (antiviral, anticancer, or dual activity) using a subset of data, known as the training data set. The remaining data, termed the test data set, are used to evaluate the SVM's ability to predict the peptide type using the selected features corresponding to each peptide in the test data set. This process facilitates the development of robust and accurate models for peptide classification. By combining feature selection (LASSO) and classification (SVM), we can construct robust predictive models, better guiding experimental work, validating hypotheses, or even leading to the discovery of new AMPs. This approach can simultaneously improve our understanding of the physicochemical properties of peptides and make predictions about peptide behavior.

The most important part of our theoretical method is identifying the physicochemical properties of AMPs related to their dual anticancer−antiviral actions. For this purpose, a list of anticancer and antiviral AMPs was obtained from the Antimicrobial Peptide Database (APD).[40] We then extracted 1547 physical−chemical features utilizing the *propy* package[41] in Python for every peptide. This package contains a very large number of physicochemical properties collected from different experimental sources. Note that these properties are not independent, and many of them correlate with each other. This information was utilized in the SVM classification models to predict the AMP type (anticancer, antiviral, and dual activity) to "learn" about the most important physical−chemical properties of peptides associated with their functioning. The details of these features, the *propy* Python package, the SVM classification model, and the details of the computational analysis are presented in the Supporting Information.

Because the traditional SVM compares only two classes of objects at any time, known as positive (1) and negative (0), in this work, the SVM was used to consider pairs of classes for the three most relevant cases: (1) antiviral but not anticancer activity (0) versus anticancer but not antiviral activity (1), (2) antiviral but not anticancer activity (0) versus both antiviral and anticancer activity (1), and (3) anticancer but not antiviral activity (0) versus both antiviral and anticancer activity (1). In the next step of the machine-learning analysis, we reduced the number of features using the LASSO method (as explained in detail in ref 42 and the Supporting Information) and repeated the prediction procedure with the SVM models. Using these machine-learning methods, we explicitly estimated the correlations in the system.

Various statistical metrics have been utilized to evaluate the prediction performance of our method. First, we used a parameter called accuracy, which measures the number of correct predictions, and its values lie between 0 (worst possible


https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828−1835

prediction performance) and 1 (perfect prediction performance)

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{n_t} \quad (1)$$

where $n_t$ is the total number of samples being predicted, TP is the number of accurate predictions for the positive class (class label of 1), and TN is the number of correct predictions for the negative class (class label of 0).

Next, we explored a parameter called recall, which measures the number of correct predictions but in a less biased way than accuracy (its values are also between 0 and 1)

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

where FN is the number of incorrect predictions for the positive class (class label of 1). The third utilized parameter is called the area under the curve (AUC), and it provides a measure of correct and incorrect predictions

$$\text{AUC} = \frac{\text{TP}}{\text{FP}} \quad (3)$$

where FP is the number of incorrect predictions for the negative class (class label of 0). Finally, we also explored a parameter known as Matthews' correlation coefficient (MCC), which is similar to other correlation coefficients (its values are between −1 and 1)

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (4)$$

Next, as explained in detail in the Supporting Information, we evaluated predictions of our theoretical procedure for comparison of different classes of peptides, and the results of our analysis are listed in Table 1. One can see that a reasonable accuracy can be achieved if all available physical−chemical properties are used in the analysis (accuracy ∼ 0.72−0.76). However, using a smaller subset of the total number of features, for example, 24 in the antiviral versus anticancer AMP analysis, could improve the accuracy (accuracy ∼ 0.79−0.83). For this reason, hyperparameter tuning has been applied to the feature selection models to determine the optimal subset of properties in which the lowest number of features corresponds to the highest accuracy. Table 1 presents the comparison of prediction performance between the baseline all-features-included SVM models and the selected-features SVM models (after hyperparameter tuning and feature selection are applied). As one can see, the accuracy of the selected-features SVM models was always higher than that of the all-features SVM models. Using only a few properties also decreased the likelihood of overfitting of data.

**Table 1. Performance Predictions of the Models for Different Cases**[a]

| analysis | no. of features | accuracy | recall | AUC | MCC |
|---|---|---|---|---|---|
| antiviral (0) vs anticancer (1) | 1547 | 0.72 | 0.61 | 0.61 | 0.38 |
| | 24 | 0.79 | 0.75 | 0.75 | 0.53 |
| anticancer (0) vs antiviral and anticancer (1) | 1547 | 0.76 | 0.55 | 0.55 | 0.21 |
| | 7 | 0.83 | 0.70 | 0.70 | 0.50 |
| antiviral (0) vs antiviral and anticancer (1) | 1547 | 0.75 | 0.67 | 0.67 | 0.44 |
| | 18 | 0.81 | 0.78 | 0.78 | 0.60 |

[a]The numbers next to the type of peptide (0 or 1) refer to the class label assigned to each type of peptide, which is relevant for interpreting the feature importance in the feature selection results.

The selected physical−chemical features are determined through hyperparameter tuning to minimize the error (or maximize the accuracy) in predicting one class or another in the SVM model. In this procedure, additional constraints have been added to minimize the number of selected features to increase the interpretability of the model. The idea here is that the anticancer and antiviral activities of peptides can be improved more efficiently by changing only a few physical−chemical features of the peptides that are known to be associated with antiviral and anticancer activity, and it is more difficult to change many features at once. The complete procedure and results of hyperparameter tuning are explained in detail in the Supporting Information.

The main results of our theoretical analysis are presented in Figure 1, which shows the relative importance of specific physical−chemical properties for each type of peptide (anticancer, antiviral, and dual activity) based on the sign and magnitude of the coefficient estimated for the given feature. For each analysis, any peptide in class 0 is associated more closely with the properties that exhibit negative coefficients, while peptides in class 1 are more closely associated with the properties that exhibit positive coefficients. To be more specific, let us consider Figure 1a in which the antiviral peptides are labeled as class 0 and the anticancer peptides as class 1. This means that the physical−chemical properties with negative coefficients better identify the antiviral AMPs while the physical−chemical properties with positive coefficients better identify the anticancer AMPs.

The higher magnitude of the coefficient suggests that this feature is assigned a greater importance by the model in distinguishing one type of peptide from another, but one should also note that while the linear weights provide a measure of the model's estimation of a particular feature, the real-world ability of a given metric to explain the differences between two classes may not be solely determined by these weights. Other factors, such as the data set characteristics and the overall model architecture, might also play a crucial role in assessing the reliability of a feature's contribution to class differentiation.

As one can see in Figure 1a, our theoretical analysis suggests that the most important feature for distinguishing the antiviral peptides is labeled as NK, which is a dipeptide composition referring to the fraction of specific amino acid pairs in a specific order (in this case N, asparagine, followed by K, lysine) in the sequence. Some other dipeptide compositions such as LW (L, leucine, followed by W, tryptophan), GY (G, glycine, followed by Y, tyrosine), and QY (Q, glutamine, followed by Y, tyrosine) also distinguish the antiviral peptides, although to a lesser degree. The anticancer peptides are distinguished the most by the dipeptide compositions KY (K, lysine, followed by Y, tyrosine) and FK (F, phenylalanine, followed by K, lysine).

Our computational procedure identifies the correlations with specific physical−chemical properties, allowing us to discuss the microscopic origin of these observations. More specifically, our method indicates the preference for both antiviral and anticancer peptides to contain lysine (K) amino acids in dipeptide composition features. This finding is consistent with previous studies that have found an abundance of lysine in single-amino acid and dipeptide composition features


https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828−1835

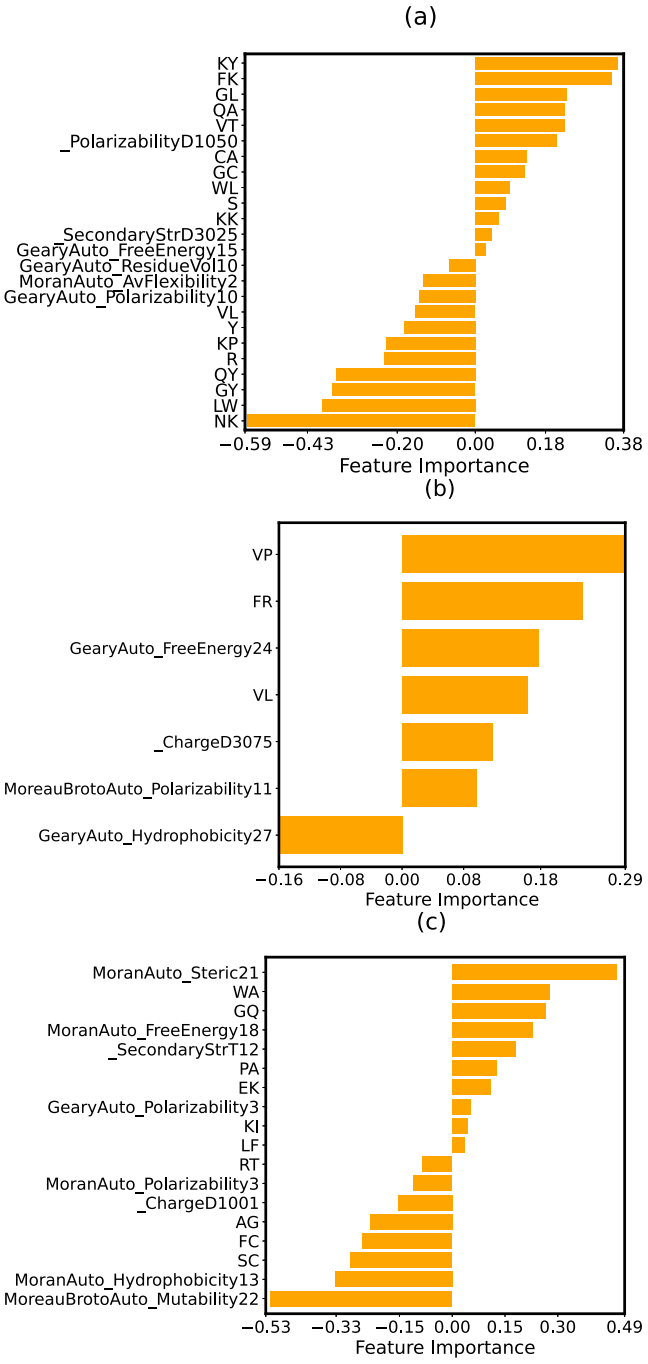



**Figure 1.** Relative importance of different physicochemical properties in distinguishing different classes of AMPs: (a) for AMPs with antiviral but not anticancer activity from AMPs with anticancer but not antiviral activity, (b) for AMPs with anticancer but not antiviral activity from AMPs with dual-action activity, and (c) for AMPs with antiviral but not anticancer activity from AMPs with dual-action activity. The properties labeled with two letters correspond to different dipeptide combinations, as explained in the text. Other properties are specified in the Supporting Information.

associated with antiviral and anticancer peptides.[43−46] Moreover, a greater fraction of K residues in hydrophilic regions has been already associated with more selective activity against cancer.[47] This result is not surprising given the aliphatic, polar, and positively charged nature of this amino acid. This amino acid can also be viewed as amphipathic, because it simultaneously contains the charged side chain and the hydrophobic tail. In addition, lysines easily establish hydrogen and covalent bonding. All of these properties make lysine very important for antiviral and anticancer peptides because they allow AMPs to associate better and enter the cell membranes, which is a crucial task in their functioning.

Similar analysis can be done to identify the physical−chemical properties that distinguish anticancer AMPs from dual-action AMPs (Figure 1b). Peptides with only anticancer activities are distinguished by a specific hydrophobic property called GearyAutoHydrophobicity27 (explained in the Supporting Information). At the same time, the dual-action peptides can be distinguished well by dipeptide composition features VP (V, valine, followed by P, proline) with the highest magnitude of the coefficient and FR (F, phenylalanine, followed by R, arginine) with a slightly smaller coefficient.

The results for identifying physical−chemical properties that distinguish antiviral AMPs from dual-action AMPs are presented in Figure 1c. One can see that the most important feature of antiviral peptides is the correlation between amino acids in mutability at distances of 22 amino acids apart. Mutability here refers to the degree to which an amino acid has been replaced by another amino acid at the same position in a protein sequence throughout evolution.[48,49] Other important features include several dipeptide compositions such as AG (A, alanine, followed by G, glycine), FC (F, phenylalanine, followed by C, cysteine), and SC (S, serine, followed by C, cysteine), as well as the property related to hydrophobicity. The most important for distinguishing dual-action peptides is the correlation between amino acids in steric hindrance at distances of 21 amino acids apart. Steric hindrance refers here to the phenomenon in which the presence of large or bulky functional groups in an amino acid, such as sizable R groups, restricts the flexibility of the protein chain.[50,51] This can limit the protein's ability to fold into specific secondary structures like $\alpha$ helices, which can play a role in the mechanisms by which amino acids interact with cancer and viral cells.[50,51] Other important features for dual-action peptides are dipeptide compositions such as WA (W, tryptophan, followed by A, alanine) and GQ (G, glycine, followed by Q, glutamine).

Our theoretical analysis indicates that there are other physical−chemical properties beyond the dipeptide compositions that are important for distinguishing different types of AMPs, and these are described in detail in the Supporting Information. At the same time, the analysis clearly shows the dominant role of several specific dipeptide compositions in distinguishing AMPs with dual action (Figure 1b,c). Identifying specific features that correlate most with the dual action is critically important for clarifying the molecular mechanisms of potential drugs.

The advantage of our theoretical approach is that we can not only identify dipeptide compositions as the most relevant physical−chemical properties for characterizing dual-action AMPs but also specify which specific combinations of residues are the most important. We found that the antiviral AMPs distinguish themselves from the anticancer AMPs by having a preference for VL double residues and R single residues. At the same time, the anticancer AMPs distinguish themselves from the antiviral AMPs by preferring WL, QA, CA, and FK double residue combinations. However, the dual-action AMPs are associated with a stronger preference for VP, FR, and VL residue combinations as compared to those of the anticancer peptides, and the dual-action AMPs are associated with a stronger preference for EK and KI as compared to those of the


https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828−1835

antiviral peptides. These results are consistent with previous studies,[29,44−46] which connected the anticancer peptides with dipeptide compositions consisting of hydrophobic leucine (L), phenylalanine (F), tryptophan (W), and alanine (A), while the antiviral peptides are connected with dipeptide compositions consisting of cationic arginine (R), anionic glutamic acid (E), and hydrophobic isoleucine (I) and valine (V).[52]

The interesting observation from our analysis is that the dual-action peptides prefer certain dipeptide compositions, such as FR or VL, that can be viewed as additive combinations of amino acids that are more relevant to antiviral activity (R and V) with amino acids that are more relevant to anticancer activity (L and F). This "additivity" is easy to understand from the point of view that dual-action AMPs must have two main functions, against viruses and tumors, and it might be explored for developing more efficient dual-action AMPs. Thus, our theoretical method suggests a concept of additivity that might be explored in designing new drugs with multiple targets, especially if these targets are "orthogonal" to each other, i.e., do not affect each other much.

Another important observation from our analysis is that the antiviral peptides in particular seem to prefer the cationic R and Y residues, and this agrees with previous studies that arginine is an important amino acid for antiviral activity.[29] The larger fraction of R residues could also be associated with the greater flexibility of antiviral peptides. It is known that the peptides based on magainin-2 were found to be more flexible when the amino acid arginine (R) was at position 10 in the sequence than when tryptophan (W) was at the same position.[53] It was argued that this is due to the arginine allowing for kinking, while tryptophan made the peptide stiffer and less flexible. The flexibility of AMPs might also be important for their antiviral functioning, because it could allow them to associate with a larger number of protein receptors, blocking viruses from associating with them.

Our results also show that dipeptide features with V, valine, are more frequently associated with the antiviral and antiviral−anticancer AMPs compared to just anticancer AMPs, and this agrees with existing studies. Specifically, we predicted that VL (valine, V, followed by leucine, L, in the amino acid sequence) is a selected feature with a stronger impact on the antiviral peptides and the antiviral−anticancer peptides than on the anticancer peptides (Figure 1a,b), while the dipeptide composition features with V are absent from the analysis of antiviral versus dual-action AMPs in Figure 1c. This is likely because dipeptide composition features with V are important for both the antiviral peptides and antiviral−anticancer peptides.

The importance of VL residue combinations can be understood from the following arguments. Valine (V) is an aliphatic, nonpolar amino acid with intermediate polarizability associated with $\beta$-strand secondary structures. Leucine (L) is also an aliphatic, nonpolar amino acid with intermediate polarizability, but it is mostly associated with $\alpha$-helical structures. The difference between leucine and valine in secondary structures is related to another selected feature associated with the dual-action peptides compared to the anticancer peptides, the higher frequency of secondary structure transitions between single amino acids from the $\alpha$ helix to the $\beta$ strand (see the Supporting Information). The difference in the secondary structures can contribute to the greater flexibility of the peptide molecules.[54,55] The presence of VL residue combinations in the peptide can increase the amphipathicity if this pair of amino acids is clustered between hydrophobic and hydrophilic clusters.

Considering biophysically similar dipeptides WL and LW, the differences observed in their significance for antiviral and anticancer activities likely originate from their specific placement within the peptide sequences. For instance, in the WL dipeptide, W (tryptophan) could be adjacent to a hydrophobic region, while L (leucine) could be near a hydrophilic region. Altering the order of this dipeptide to LW might impact the overall peptide structure and, consequently, its function. The weights observed in the model are influenced by a combination of factors, including the composition of amino acids, the context of the sequence, and the tendencies toward certain secondary structures. Furthermore, the orientation of the peptide and the precise positions of these dipeptides within the sequence can affect their interaction with membranes.[56−61]

In this Letter, we introduce a new theoretical procedure that could help rationally design medical drugs with multiple targets. More specifically, as a first step of the procedure, we investigated what physical−chemical properties are most relevant for identifying AMPs that can simultaneously act against viruses and cancer. The theoretical method combined correlations and chemoinformatics analysis as part of machine-learning investigations to predict the most relevant properties associated with dual-action peptides. Theoretical analysis also suggested that several strategies can be effectively utilized in the rational design process. For example, we introduced and illustrated the concept of additivity as a possible strategy for targets that are very different from each other.

Furthermore, the method allowed us to determine the dominant effect of dipeptide residue compositions that clearly distinguish the dual-action peptides. While we are not able to clarify the microscopic details of the mechanisms of AMPs, we can speculate about the possible origin of the dipeptide effects. Antiviral and anticancer peptides can exert their inhibitory effects by interacting with specific structures or proteins on the surfaces of viruses and cancer cells, respectively. Dipeptide motifs within these peptides might also play a crucial role in recognizing and interacting with components of these cells. For example, dipeptide motifs in antiviral peptides can assist in disrupting the function of envelope proteins on the surface of viral cells that are important for the entry of viruses into host cells, potentially preventing that entry.[62−64] In addition, dipeptide motifs can interfere with fusion proteins involved in enveloped viruses merging their envelope with the host cell membrane during entry into the host cell, preventing successful entry.[65,66] Certain dipeptide motifs might also enhance the cell-penetrating ability of anticancer peptides.[67,68] Once inside cancer cells, dipeptide motifs may facilitate interactions with intracellular proteins or organelles, contributing to the disruption of essential cellular processes. For example, dipeptide motifs have been involved in the binding to mitochondrial proteins, leading to the induction of apoptosis.[67,69]

While the proposed theoretical procedure might be useful in selecting new drugs with multiple targets, as illustrated in our analysis, it is important to discuss its limitations and shortcomings (also discussed with more technical details in the Supporting Information). First, it implicitly assumes that there is a 1:1 correspondence between the multitarget activities of the drug and certain physical−chemical properties that they exhibit. However, this might not be always the case because the



https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828−1835

complexity of underlying mechanisms might lead to the appearance of some random correlations under specific conditions. In other words, the correlations might not be universal. More microscopic information about how drugs function is needed. Second, our method relies on the existence of some molecules that already possess dual functions, so that the machine-learning methods can "learn" from them. However, in many situations, there are no such drugs or the information is very limited, and the predictions made with the help of machine learning might be not reliable. In addition, there is always a danger of overfitting the limited amount of data. Furthermore, in our analysis of anticancer−antivirus peptides, we did not distinguish different types of cancer and viruses that might exhibit very different mechanisms of action. However, despite these issues, the proposed method provides a clear physical−chemical approach to finding new drugs with controlled properties.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The source code and data used to produce the results and analyses presented in this work are available from Figshare (https://figshare.com/s/fc13c4f1f41516a62919).

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c03624.

Additional explanation of methods, considerations for the design of future peptides, limitations, results of hyperparameter tuning, and description of the selected features for each analysis (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Anatoly Kolomeisky** − *Department of Chemistry, 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 and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

**Sofya Domakhina** − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

**Catherine Vasnetsov** − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

**Victor Vasnetsov** − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c03624


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), the National Science Foundation (NSF) (CHE-1953453 and CHE-2246878), and the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

## ■ REFERENCES


(1) Boersma, P.; Black, L. I.; Ward, B. W. Prevalence of Multiple Chronic Conditions Among US Adults, 2018. 2020.

(2) Hota, D.; Tripathy, A. *Cancer Diagnostics and Therapeutics: Current Trends, Challenges, and Future Perspectives*; Springer, 2022; pp 287−302.

(3) Kwon, J.; Bakhoum, S. F. The cytosolic DNA-sensing cGAS−STING pathway in cancer. *Cancer discovery* **2020**, *10*, 26−39.

(4) Allen, B. M.; Hiam, K. J.; Burnett, C. E.; Venida, A.; DeBarge, R.; Tenvooren, I.; Marquez, D. M.; Cho, N. W.; Carmi, Y.; Spitzer, M. H. Systemic dysfunction and plasticity of the immune macroenvironment in cancer models. *Nat. Med.* **2020**, *26*, 1125−1134.

(5) Hanahan, D.; Weinberg, R. A. Hallmarks of cancer: the next generation. *cell* **2011**, *144*, 646−674.

(6) Cetraro, P.; Plaza-Diaz, J.; MacKenzie, A.; Abadía-Molina, F. A review of the current impact of inhibitors of apoptosis proteins and their repression in cancer. *Cancers* **2022**, *14*, 1671.

(7) Blaylock, R. L. Viruses and tumor cell microenvironment: a brief summary. *Surgical Neurology International* **2019**, *10*, 160.

(8) Shlomai, A.; de Jong, Y. P.; Rice, C. M. Virus associated malignancies: the role of viral hepatitis in hepatocellular carcinoma. *Semin. Cancer Biol.* **2014**, *26*, 78−88.

(9) Yeo, W.; Chan, P. K.; Zhong, S.; Ho, W. M.; Steinberg, J. L.; Tam, J. S.; Hui, P.; Leung, N. W.; Zee, B.; Johnson, P. J. Frequency of hepatitis B virus reactivation in cancer patients undergoing cytotoxic chemotherapy: a prospective study of 626 patients with identification of risk factors. *J. Med. Virol.* **2000**, *62*, 299−307.

(10) Elfaituri, M. K.; Morsy, S.; Tawfik, G. M.; Abdelaal, A.; El-Qushayri, A. E.; Faraj, H. A.; Hieu, T. H.; Huy, N. T. Incidence of Infection-related mortality in cancer patients: Trend and survival analysis. *J. Clin. Oncol.* **2019**, *37*, e23095.

(11) De Clercq, E. Strategies in the design of antiviral drugs. *Nat. Rev. Drug Discovery* **2002**, *1*, 13−25.

(12) Correa, A. M.; Howard-Varona, C.; Coy, S. R.; Buchan, A.; Sullivan, M. B.; Weitz, J. S. Revisiting the rules of life for viruses of microorganisms. *Nat. Rev. Microbiol.* **2021**, *19*, 501−513.

(13) Villanueva, R. A.; Rouillé, Y.; Dubuisson, J. Interactions between virus proteins and host cell membranes during the viral life cycle. *Int. Rev. Cytol.* **2005**, *245*, 171−244.

(14) Song, B.; Sheng, X.; Justice, J. L.; Lum, K. K.; Metzger, P. J.; Cook, K. C.; Kostas, J. C.; Cristea, I. M. Intercellular communication within the virus microenvironment affects the susceptibility of cells to secondary viral infections. *Sci. Adv.* **2023**, *9*, No. eadg3433.

(15) Mohamadzadeh, M.; Chen, L.; Schmaljohn, A. L. How Ebola and Marburg viruses battle the immune system. *Nat. Rev. Immunol.* **2007**, *7*, 556−567.

(16) Mueller, S. N.; Rouse, B. T. Immune responses to viruses. *Clin. Immunol.* **2008**, 421.

(17) Ison, M. G.; Hayden, F. G. Viral infections in immunocompromised patients: what's new with respiratory viruses? *Current opinion in infectious diseases* **2002**, *15*, 355−367.

(18) Zaiou, M. Multifunctional antimicrobial peptides: therapeutic targets in several human diseases. *Journal of Mol. Med.* **2007**, *85*, 317−329.

(19) Tossi, A.; Sandri, L.; Giangaspero, A. Amphipathic, α-helical antimicrobial peptides. *Pept. Sci.* **2000**, *55*, 4−30.

(20) Zhong, C.; Zhang, L.; Yu, L.; Huang, J.; Huang, S.; Yao, Y. A review for antimicrobial peptides with anticancer properties: repurposing of potential anticancer agents. *BIO Integr.* **2021**, *1*, 156−167.

(21) 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. *Mil. Med. Res.* **2021**, *8*, 1−25.

(22) Ahmed, T. A.; Hammami, R. Recent insights into structure−function relationships of antimicrobial peptides. *J. Food Biochem.* **2019**, *43*, No. e12546.

(23) Motsa, B. B.; Stahelin, R. V. Lipid−protein interactions in virus assembly and budding from the host cell plasma membrane. *Biochem. Soc. Trans.* **2021**, *49*, 1633−1641.
(24) Skotland, T.; Kavaliauskiene, S.; Sandvig, K. The role of lipid species in membranes and cancer-related changes. *Cancer Metastasis Rev.* **2020**, *39*, 343−360.
(25) He, H.; Sun, L.; Ye, J.; Liu, E.; Chen, S.; Liang, Q.; Shin, M. C.; Yang, V. C. Enzyme-triggered, cell penetrating peptide-mediated delivery of anti-tumor agents. *J. Controlled Release* **2016**, *240*, 67−76.
(26) Norouzi, P.; Mirmohammadi, M.; Houshdar Tehrani, M. H. Anticancer peptides mechanisms, simple and complex. *Chem.-Biol. Interact.* **2022**, *368*, 110194.
(27) Naeimi, R.; Bahmani, A.; Afshar, S. Investigating the role of peptides in effective therapies against cancer. *Cancer Cell Int.* **2022**, *22*, 1−10.
(28) Han, W.; Meng, F.; Gan, H.; Guo, F.; Ke, J.; Wang, L. Targeting self-assembled F127-peptide polymer with pH sensitivity for release of anticancer drugs. *RSC Adv.* **2021**, *11*, 1461−1471.
(29) Shoombuatong, W.; Schaduangrat, N.; Nantasenamat, C. Unraveling the bioactivity of anticancer peptides as deduced from machine learning. *EXCLI Journal* **2018**, *17*, 734.
(30) Jafari, A.; Babajani, A.; Sarrami Forooshani, R.; Yazdani, M.; Rezaei-Tavirani, M. Clinical applications and anticancer effects of antimicrobial peptides: from bench to bedside. *Front. Oncol.* **2022**, *12*, 819563.
(31) Vilas Boas, L. C. P.; Campos, M. L.; Berlanda, R. L. A.; de Carvalho Neves, N.; Franco, O. L. Antiviral peptides as promising therapeutic drugs. *Cell. Mol. Life Sci.* **2019**, *76*, 3525−3542.
(32) Yao, L.; Li, W.; Zhang, Y.; Deng, J.; Pang, Y.; Huang, Y.; Chung, C.-R.; Yu, J.; Chiang, Y.-C.; Lee, T.-Y. Accelerating the Discovery of Anticancer Peptides through Deep Forest Architecture with Deep Graphical Representation. *InterNatl. J. (Wash.) of Molecular Sciences* **2023**, *24*, 4328.
(33) Grisoni, F.; Neuhaus, C. S.; Gabernet, G.; Müller, A. T.; Hiss, J. A.; Schneider, G. Designing anticancer peptides by constructive machine learning. *ChemMedChem.* **2018**, *13*, 1300−1302.
(34) Ghaly, G.; Tallima, H.; Dabbish, E.; Badr ElDin, N.; Abd El-Rahman, M. K.; Ibrahim, M. A.; Shoeib, T. Anti-Cancer Peptides: Status and Future Prospects. *Molecules* **2023**, *28*, 1148.
(35) Ge, R.; Feng, G.; Jing, X.; Zhang, R.; Wang, P.; Wu, Q. Enacp: an ensemble learning model for identification of anticancer peptides. *Front. Genet.* **2020**, *11*, 760.
(36) Adari, G. K.; Raja, M.; Vijaya, P. *Data Science for Genomics*; Elsevier, 2023; pp 25−68.
(37) Boopathi, V.; Subramaniyam, S.; Malik, A.; Lee, G.; Manavalan, B.; Yang, D.-C. mACPpred: a support vector machine-based meta-predictor for identification of anticancer peptides. *InterNatl. J. (Wash.) of molecular sciences* **2019**, *20*, 1964.
(38) Piotrowska, U.; Sobczak, M.; Oledzka, E. Current state of a dual behaviour of antimicrobial peptides—Therapeutic agents and promising delivery vectors. *Chemical Biology & Drug Design* **2017**, *90*, 1079−1093.
(39) Parchebafi, A.; Tamanaee, F.; Ehteram, H.; Ahmad, E.; Nikzad, H.; Haddad Kashani, H. The dual interaction of antimicrobial peptides on bacteria and Cancer Cells; mechanism of action and therapeutic strategies of nanostructures. *Microb. Cell Fact.* **2022**, *21*, 118.
(40) Wang, G.; Li, X.; Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **2016**, *44*, D1087−D1093.
(41) Cao, D.-S.; Xu, Q.-S.; Liang, Y.-Z. propy: a tool to generate various modes of Chou's PseAAC. *Bioinformatics* **2013**, *29*, 960−962.
(42) 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.
(43) Manavalan, B.; Basith, S.; Shin, T. H.; Choi, S.; Kim, M. O.; Lee, G. MLACP: machine-learning-based prediction of anticancer peptides. *Oncotarget* **2017**, *8*, 77121.
(44) Chen, W.; Ding, H.; Feng, P.; Lin, H.; Chou, K.-C. iACP: a sequence-based tool for identifying anticancer peptides. *Oncotarget* **2016**, *7*, 16895.
(45) Agrawal, P.; Bhagat, D.; Mahalwal, M.; Sharma, N.; Raghava, G. P. AntiCP 2.0: an updated model for predicting anticancer peptides. *Briefings Bioinf.* **2021**, *22*, bbaa153.
(46) Chang, K. Y.; Yang, J.-R. Analysis and prediction of highly effective antiviral peptides based on random forests. *PloS one* **2013**, *8*, No. e70166.
(47) Hadianamrei, R.; Tomeh, M. A.; Brown, S.; Wang, J.; Zhao, X. Rationally designed short cationic $\alpha$-helical peptides with selective anticancer activity. *J. Colloid Interface Sci.* **2022**, *607*, 488−501.
(48) Brooks, D. J.; Fresco, J. R.; Lesk, A. M.; Singh, M. Evolution of Amino Acid Frequencies in Proteins Over Deep Time: Inferred Order of Introduction of Amino Acids into the Genetic Code. *Mol. Biol. Evol.* **2002**, *19*, 1645−1655.
(49) Tourasse, N. J.; Li, W.-H. Selective Constraints, Amino Acid Composition, and the Rate of Protein Evolution. *Mol. Biol. Evol.* **2000**, *17*, 656−664.
(50) Childers, M. C.; Towse, C.-L.; Daggett, V. The effect of chirality and steric hindrance on intrinsic backbone conformational propensities: tools for protein design. *Protein Eng. Des. Sel.* **2016**, *29*, 271−280.
(51) Cieplak, A. S. Protein folding, misfolding and aggregation: The importance of two-electron stabilizing interactions. *PLoS One* **2017**, *12*, No. e0180905.
(52) Schaduangrat, N.; Nantasenamat, C.; Prachayasittikul, V.; Shoombuatong, W. Meta-iAVP: a sequence-based meta-predictor for improving the prediction of antiviral peptides using effective feature representation. *InterNatl. J. (Wash.) of molecular sciences* **2019**, *20*, 5743.
(53) Balleza, D. Peptide Flexibility and the Hydrophobic Moment are Determinants to Evaluate the Clinical Potential of Magainins. *J. Membr. Biol.* **2023**, *256*, 317.
(54) Lee, J. H.; Yin, R.; Ofek, G.; Pierce, B. G. Structural Features of Antibody-Peptide Recognition. *Front. Immunol.* **2022**, *13*, 910367.
(55) Brandt, G. S. *Molecular Life Science*; SpringerLink, 2018.
(56) Juhl, D. W.; Glattard, E.; Aisenbrey, C.; Bechinger, B. Antimicrobial peptides: mechanism of action and lipid-mediated synergistic interactions within membranes. *Faraday Discuss.* **2021**, *232*, 419.
(57) Schrank, E.; Wagner, G. E.; Zangger, K. Solution NMR Studies on the Orientation of Membrane-Bound Peptides and Proteins by Paramagnetic Probes. *Molecules* **2013**, *18*, 7407−7435.
(58) Galdiero, S.; Falanga, A.; Cantisani, M.; Vitiello, M.; Morelli, G.; Galdiero, M. Peptide-Lipid Interactions: Experiments and Applications. *InterNatl. J. (Wash.) of Molecular Sciences* **2013**, *14*, 18758−18789.
(59) Pino-Angeles, A.; Lazaridis, T. Effects of Peptide Charge, Orientation, and Concentration on Melittin Transmembrane Pores. *Biophys. J.* **2018**, *114*, 2865.
(60) Holt, A.; Killian, J. A. Orientation and dynamics of transmembrane peptides: the power of simple models. *Eur. Biophys. J.* **2010**, *39*, 609−621.
(61) Al Musaimi, O.; Mercado-Valenzo, O. M.; Williams, D. R. Factors Influencing the Prediction Accuracy of Model Peptides in Reversed-Phase Liquid Chromatography. *Separations* **2023**, *10*, 81.
(62) Schoeman, D.; Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virol. J.* **2019**, *16*, 69.
(63) Thakur, N.; Qureshi, A.; Kumar, M. AVPpred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res.* **2012**, *40*, W199−W204.
(64) Vilas Boas, L. C. P.; Campos, M. L.; Berlanda, R. L. A.; de Carvalho Neves, N.; Franco, O. L. Antiviral peptides as promising therapeutic drugs. *Cell. Mol. Life Sci.* **2019**, *76*, 3525−3542.
(65) Joardar, A.; Pattnaik, G. P.; Chakraborty, H. Mechanism of Membrane Fusion: Interplay of Lipid and Peptide. *J. Membr. Biol.* **2022**, *255*, 211−224.


https://doi.org/10.1021/acs.jpclett.3c03624
J. Phys. Chem. Lett. 2024, 15, 1828−1835

(66) Sardar, A.; Dewangan, N.; Panda, B.; Bhowmick, D.; Tarafdar, P. K. Lipid and Lipidation in Membrane Fusion. *J. Membr. Biol.* **2022**, *255*, 691−703.
(67) Kolchina, N.; Khavinson, V.; Linkova, N.; Yakimov, A.; Baitin, D.; Afanasyeva, A.; Petukhov, M. Systematic search for structural motifs of peptide binding to double-stranded DNA. *Nucleic Acids Res.* **2019**, *47*, 10553−10563.
(68) Karami Fath, M.; Babakhaniyan, K.; Zokaei, M.; Yaghoubian, A.; Akbari, S.; Khorsandi, M.; Soofi, A.; Nabi-Afjadi, M.; Zalpoor, H.; Jalalifar, F.; Azargoonjahromi, A.; Payandeh, Z.; Alagheband Bahrami, A. Anti-cancer peptide-based therapeutic strategies in solid tumors. *Cell. Mol. Biol. Lett.* **2022**, *27*, 33.
(69) Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018**, *46*, 8038−8056.



https://doi.org/10.1021/acs.jpclett.3c03624
*J. Phys. Chem. Lett.* 2024, 15, 1828−1835