JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2018, 58, 1141−1151

pubs.acs.org/jcim

Predictive Model of Linear Antimicrobial Peptides Active against Gram-Negative Bacteria

Boris Vishnepolsky,*[,†] Andrei Gabrielian,[‡] Alex Rosenthal,[‡] Darrell E. Hurt,[‡] Michael Tartakovsky,[‡] Grigol Managadze,[†] Maya Grigolava,[†] George I. Makhatadze,[§] and Malak Pirtskhalava*[,†]

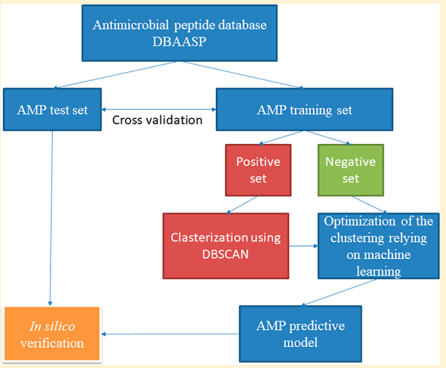
[†]Ivane Beritashvili Center of Experimental Biomedicine, Tbilisi 0160, Georgia

[‡]Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

[§]Rensselaer Polytechnic Institute, Troy, New York 12180, United States

Supporting Information

ABSTRACT: Antimicrobial peptides (AMPs) have been identified as a potential new class of anti-infectives for drug development. There are a lot of computational methods that try to predict AMPs. Most of them can only predict if a peptide will show any antimicrobial potency, but to the best of our knowledge, there are no tools which can predict antimicrobial potency against particular strains. Here we present a predictive model of linear AMPs being active against particular Gram-negative strains relying on a semi-supervised machine-learning approach with a density-based clustering algorithm. The algorithm can well distinguish peptides active against particular strains from others which may also be active but not against the considered strain. The available AMP prediction tools cannot carry out this task. The prediction tool based on the algorithm suggested herein is available on https://dbaasp.org



INTRODUCTION

Antimicrobial peptides (AMPs) are a diverse group of molecules produced by the innate immune system in response to infectious agents. They have recently been identified as a potential new class of anti-infectives for drug development. A common feature of the antimicrobial mechanism of action is its interaction with the membrane, whereby AMPs disturb the membrane bilayer integrity with the consequential collapse of the transmembrane electrochemical gradients.[1]

Widespread resistance to AMPs has not been reported, but there are several ways for resistance to appear.[2,3] For instance, in some species there is the possibility of changing membrane charge by modifying the lipid polar group. In other species, peptidases that can degrade the AMP are synthesized. Additionally, efflux of AMPs by the pump is another potential resistance mechanism. Because mechanisms of resistance are known, strategies to prevent resistance through these pathways can be engineered. For example, stable peptide mimetics can be used against peptidases, resistance caused by modification of lipids can be avoided by using combinations of differently charged AMPs, or pump inhibitors can be used to prevent efflux.

Besides avoiding resistance, another main challenge in developing AMPs is the comprehensive knowledge and understanding of the amino acid "code" of antimicrobial activity. Revealing motives that are responsible for antimicrobial activity in sequences is difficult because the same peptide can act differently on various targets. Also, different AMPs may use different mechanisms of action against a certain target. Therefore, it is challenging to uncover common features, like physical-chemical characteristics, that predict the effectiveness of peptides against certain strains. To address this challenge and gain a comprehensive knowledge of AMPs' mechanisms of action, a task-oriented design of anti-infective drug approach can be performed.

Task-oriented rational *ab initio* design, which involves sequence-based computational prediction of antimicrobial properties with the subsequent synthesis and experimental testing of peptides for antimicrobial activity, is the most cost-effective way of developing novel antibiotics against drug-resistant bacteria. While some AMPs are already in clinical and commercial use, future design of novel AMPs requires further quantitative structure−activity relationship (QSAR) studies to optimize sequence-based predictive models.

There are many computational methods to predict AMP potency. The methods combine different physical-chemical and structural descriptors together with various approaches for data analysis such as principal component analysis (PCA), partial least-squares (PLS),[4] artificial neural networks (ANN),[5−8] support vector machines (SVM),[9−12] decision trees (DT),[13] k-mean clustering (kMC),[14] fuzzy k-nearest neighbor (fKNN),[15] Random Forest (RF),[16,17] Long Short-Term Memory (LSTM) recurrent neural networks,[18] and many others.[19−28] Based on

Received: February 28, 2018
Published: May 2, 2018

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

these approaches, computer-assisted peptide design was developed.[16,29−32] However, most of the algorithms for AMP prediction do not take into account differences in mechanisms of action, structures, and modes of interaction with membrane.

There are at least four kinds of AMPs: linear cationic antimicrobial peptides (LCAP), cationic peptides stabilizing structure by intrachain covalent bonds (CICP), proline- and arginine-rich peptides (PAP), and anionic antimicrobial peptides (AP).[33] According to DBAASP,[34] which is a manually curated database of AMPs, linear peptides (LCAP) represent the largest class of AMPs (see Figure 1). As such, knowledge-based prediction tools for this AMP class might be statistically more reliable than for other classes.

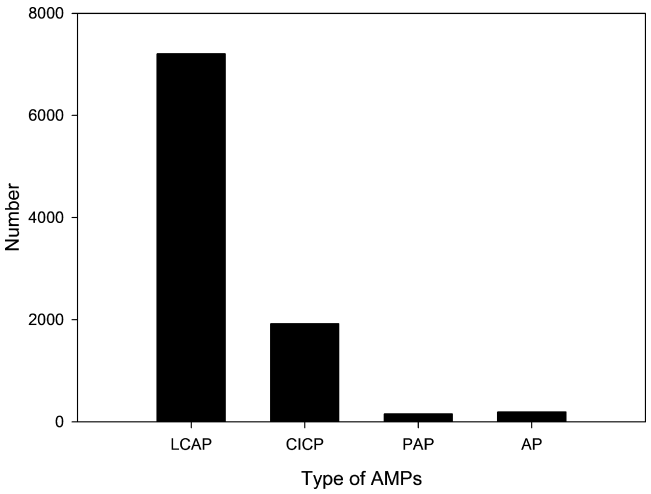


Figure 1. Distribution of four types of AMPs in DBAASP.

Recently the common features of AMPs have been revealed, and a predictive model using these features has been developed.[4−29] However, such models have several drawbacks. First, the methods do not distinguish targeted strains during the model development, although the antimicrobial potency of AMPs strongly depends on bacterial envelope types and, thus, on the strains it targets. Second, data for the negative set used in the model optimization are limited because experimental validation for all of the peptides against the targeted strains which do not have antimicrobial potency is practically impossible to uncover. So, most methods use randomly selected samples whose antimicrobial activity has not been proved experimentally. The susceptibility test results show that even small changes in envelope composition can dramatically change the antimicrobial potency of a peptide. For example, peptide Hp1404 [E8K] has different values of minimum inhibitory concentration (MIC) for two different strains of *Escherichia coli* (25 μg/mL for AB94012 and 100 μg/mL for ATCC 25922).[35] So the development of a model for the specific bacterial strains is very important. But there is a problem to solve this task. The number of strains for which there are enough data for statistical analysis is small. Nevertheless, some methods[36,37] have tried to use a multi-target approach and discriminant analysis to predict antimicrobial activity against different bacterial strains. These methods are based on the Box-Jenkins moving average approach[38−43] and use the mean values of descriptors for all peptides tested against the same bacterial strain. Such an approach increases statistical power and provides results for different bacterial strains. Another problem concerns the differences in modes of action. Although according to a widely accepted supposition, physical-chemical properties are responsible for the modes of action,[44−46] to our knowledge, classification of AMPs according to physical-chemical properties has not yet been done. So the development of a predictive model that is based only on particular strains for which there are enough data and which takes into account the variety of the mechanisms of actions still remains a challenge. Here we are trying to group the AMPs that are active against *E. coli* relying on the similarity of their physical-chemical features.

Supposing the existence of different mechanisms of action, we assume that the distribution of the values of physical-chemical characteristics in the descriptors' space is discontinuous. The distribution of physical-chemical descriptors can be discontinuous also due to insufficient data. The source of discontinuity on the activity landscape can be "activity cliffs".[47,48] The impact of discontinuities on the classification by convenient machine-learning approaches is negative.

So we decided to develop a clustering-based predictive algorithm, taking into consideration the fact that AMPs' mechanisms of action against different species can vary and training sets can involve "activity cliffs". Previously reported methods that used a clustering approach for the development of a predictive model[14,15] relied on a *k*-means clustering algorithm to assign AMPs to existing families. We do not know any other approach that has used a clustering algorithm to classify peptides according to their physical-chemical characteristics and takes into account different AMPs' modes of action. Considering that all possible mechanisms of action are unknown, we developed a predictive method relying on a density-based clustering algorithm. This approach, along with the development of a new rational predictive model, can predict antimicrobial potency based on multiple mechanisms of action. Differences in the mechanisms of action can be translated at the level of physical-chemical characteristics of peptides. A multiplicity of mechanisms can be linked to the existence of different groups of peptides, grouped (clustered) in the different spaces of physical-chemical characteristics, or discontinuously distributed in the certain space and grouped into non-overlapped clusters within the space.

Here, we describe a simple predictive method based on clustering of AMPs according to their physical-chemical data instances. Predictive models were developed relying on the peptides for which there are abundant data on their antimicrobial activity against specific strains. Table 1 presents the most highly tested strains from DBAASP. The most complete data are available for Gram-negative bacteria (*E. coli* ATCC 25922, *Pseudomonas aeruginosa* ATCC 27853). Taking this into consideration, we performed a QSAR study and developed new predictive models for linear AMPs being active against these Gram-negative bacterial strains.

Table 1. Most Highly Tested Strains from DBAASP

target organism	no. of peptides
Escherichia coli ATCC 25922	2719
Staphylococcus aureus ATCC 25923	1829
Pseudomonas aeruginosa ATCC 27853	1523
Bacillus subtilis	569
Candida albicans ATCC 90028	507
Klebsiella pneumonia	355

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

METHODS

Physical-Chemical Features. Both AMPs and the lipid bilayer are amphipathic. So, the main elements guiding the physical-chemical properties of AMPs and consequently, responsible for the interaction with the membrane, must be ionic charges and hydrophobicity of the side chains of amino acid residues.[21]

Different combinations (mainly linear) of these elements allow us to evaluate physical-chemical properties of the peptides such as hydrophobicity, charge, isoelectric point, propensity to disordering, a linear hydrophobic moment, etc. The following nine features were used in the QSAR study: normalized hydrophobic moment (M), normalized hydrophobicity (H), charge (C), isoelectric point (I), penetration depth (D), orientation of peptides relative to the surface of membrane (tilt angle) (O), propensity to disordering (R), linear moment (L), and *in vitro* aggregation (A). Hydrophobic scale, described by Moon and Fleming,[49] was used for the definition of hydrophobic moment, hydrophobicity, and linear moment. Penetration depth and tilt angle were defined relying on the depth-dependent potentials developed by Senese and co-workers.[50] Propensity to disordering was calculated by Uversky's formula.[51] The detailed definitions of these characteristics can be found in a paper by Vishnepolsky and Pirtskhalava.[21]

Author: Hydrogen bond formation is energetically favorable in the membrane environment. So, it is reasonable to assume that the membrane will force the linear peptide to adopt a regular conformation, mainly α-helical.[45,52] This assumption is supported by the fact that all transmembrane domains of membrane proteins consist mainly of regular secondary structure. Thus, in the case of linear peptides, the properties, requiring information about 3D structure such as a helical hydrophobic moment, orientation of the peptides with respect to the surface of the membrane, the penetration depth, etc. can be evaluated using sequence information only.

As such, nine features, nine attributes of data instance will present each peptide in the data set. Values of the attributes are normalized by z-scores. Mean values and standard deviation for the definition of z-scores were established on the values of attributes of positive training set.

As we do not have a comprehensive knowledge about the modes of actions used by AMPs, and especially which ones from these nine attributes are the main determinants of the particular mode of action, we explored each of the possible k-dimensional spaces (k = 1, ..., 9). The possible number of k-dimensional (kD) spaces created by nine features equals the number of k combinations from a given set of nine elements $C(k,9)$. The total number of explored spaces is equal to 511. In each kD space, clustering of the instances was performed by relying on the density-based algorithm DBSCAN.[53] Clustering was performed on the basis of Euclidian distances. In each kD space clustering of the positive training set was performed for various values of DBSCAN parameters ε and *minpts*. ε varied from 0.2 to 2.0 by steps of 0.2, while *minpts* varied from 3 to 15 by steps of 1. Optimization of clustering was performed based on the criteria described below.

Construction of the Data Sets. The predictive model was developed based on the information on *E. coli* ATCC 25922, which contains sufficient experimentally validated data of antimicrobial activity. It is also assumed that peptide length is a major characteristic determining AMP potency.[45] The lengths of most peptides fall into intervals 10−16 and 18−27 aa (Figure 2). Some descriptors which are used in the paper (hydrophobic moments, penetration depth, and tilt angle) depend on peptide 3D structure. As it was mentioned above, these descriptors were calculated under condition that the peptides have α-helical conformation. For small peptides it is more likely that the whole peptide will have α-helical conformation than for longer peptides. We have decided to develop a predictive model for peptides with length of more than 17 separately to have possibility to observe the impact of supposition concerning acceptability of straight helical conformation for the peptides of this length. As such, data sets from DBAASP were created according to the following restrictions: sequence lengths 10−16 and 18−27 aa, without intrachain bonds, and without non-standard amino acids.

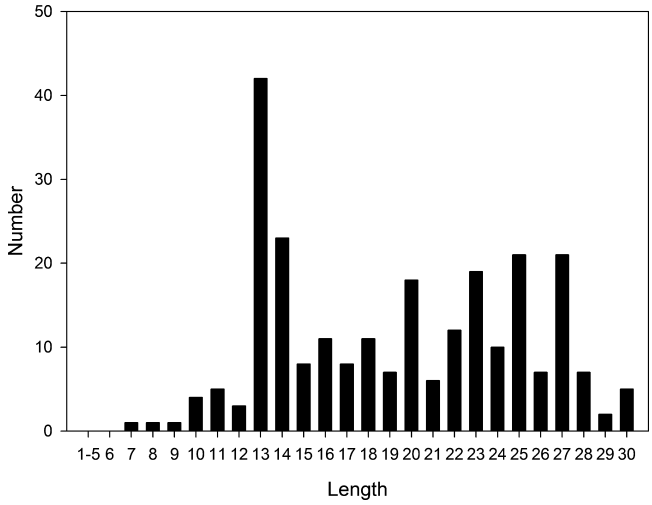


Figure 2. Distribution of peptide lengths of ribosomal AMPs active against *E. coli* ATCC 25922.

Training and test data sets were created using the following thresholds of antimicrobial activities: MIC < 25 μg/mL for the positive set and MIC > 100 μg/mL for the negative set. Threshold values were chosen based on the following considerations. Standardization of MIC evaluations is problematic. Data from DBAASP correspond to different conditions of *in vitro* susceptibility testing (different methods, i.e., agar dilution, broth dilution; and different conditions, i.e., Luria−Bertani (LB) broth, Mueller−Hinton broth (MHB), etc.). Here, we have used susceptibility data from testing in two most frequently used broths, LB and MHB. Different conditions could be the reason for differences between estimated MICs. According to the accepted practice, if MIC is within ±2 doubling dilutions for ≥95% of the compared test result sets, the matching of the results is defined as excellent.[54] Therefore, rather large interval between positive and negative sets allows us to take into consideration experimental errors which can be caused by variation in the experimental conditions. To make the data set non-redundant, the Euclidian distance between instances in the nine-dimensional space (in the general space) was set to be more than 0.2.

Description of Algorithm. A semi-supervised machine-learning approach relying on the density-based clustering algorithm DBSCAN was developed to optimize the predictive model. Learning is based on the following full sets of peptides: full positive set (FPS), 174 and 118 peptides of 10−16 and 18−

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

27 aa length, respectively; and full negative set (FNS), 174 and 118 peptides of 10−16 and 18−27 aa length, respectively. For validation purposes, five training sets (positive and negative), RTS_k (k = 1, ..., 5), were randomly created from full sets. Positive set was created on the basis of FPS and it consists of 140 and 100 peptides of 10−16 and 18−27 aa length, respectively. Negative sets were created from FNS and they consist of the same amount of peptides as positive sets.

Learning process was performed in three steps of optimization:

Initially, clusters of active peptides for different combinations of DBSCAN parameters ε and *minpts* were generated in each mth space (m = 1, ..., 511) and filtered by number of peptides in the cluster $M_j > 10$. Next, non-active peptides (from negative set) were added to the remaining clusters, and additional filtering by positive prediction value $PPV_j \geq 0.75$ (j = 1, ..., k; k is the number of clusters (for definition of PPV see below)) was performed. Schematic representation of the first step of the algorithm is shown in the Figure 3.

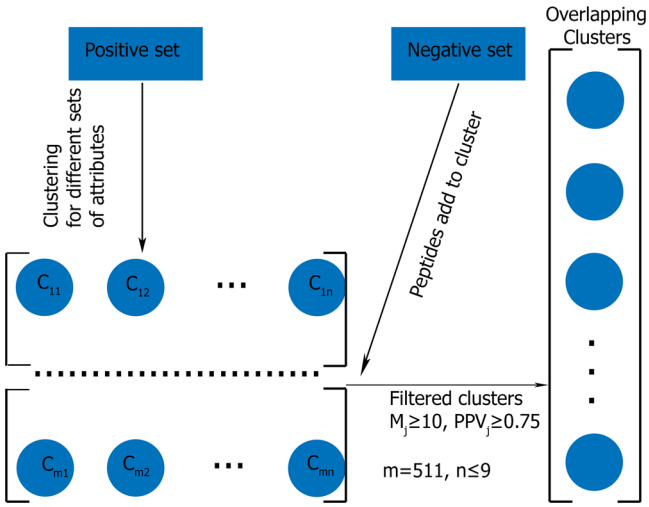


Figure 3. Schematic presentation of the optimization of clustering.

The first step of the procedure, that is, clustering optimization, was performed for each training set RTS_k (k = 1, ..., 5) for validation purposes. On the second step, the stable clusters (training set independent) were selected from the list of clusters obtained by cross-validation. Cluster is stable if similar clusters are remained in all RTS_k. We defined C_{ik} as similar to C_{jl} if $S_{ij} > 0.85$, where $S_{ij} = N_{ij}/\max(N_i,N_j)$. Here C_{ik} and C_{jl} are the ith and jth clusters appearing by the training on the kth RTS_k and lth RTS_l sets, respectively; N_{ij} is the number of peptides that fall in both C_{jl}, and C_{ik}; N_i and N_j are the the number of peptides of C_{ik} and C_{jl}, respectively (k = 1, ..., 5; l = 1, ..., 5; $k \neq l$).

The clusters obtained on the first two steps are overlapped because they were created with different combination of DBSCAN parameters (ε, *minpts*). So, on the third step, from overlapped sets of clusters were generated combinations of non-overlapping stable clusters (CSC) of instances and ranked according to the values of P_i:

$$P_i = \sum_{j=1}^{j=l_i} (SN_{ij}PPV_{ij})$$

where SN_{ij} and PPV_{ij} are sensitivity and positive predictive value for jth cluster of the ith CSC (i = 1, n; j = 1, ..., n; j = 1, ..., l_i; n is the number of CSC and l_i is the number of clusters in the ith CSC). CSC with the highest P_i was selected for AMP prediction.

Evaluation of the Quality of the Prediction. The following equations were used to evaluate the quality of the prediction:

$$SN = TP/(TP + FN)$$

$$SP = TN/(TN + FP)$$

$$AC = (TP + TN)/(TP + FN + TN + FP)$$

$$BAC = (SP + SN)/2$$

$$PPV = TP/(TP + FP)$$

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

where SN is sensitivity, SP is specificity, AC is accuracy, BAC is balance accuracy, PPV is positive predictive value, MCC is Matthews correlation coefficient, TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Availability. The prediction tool based on the suggested algorithm is available at https://dbaasp.org

RESULTS

Clustering Optimization Results. *Clustering of Instances of Peptides of 10−16 aa Length.* Three clusters of data instances were selected by the presented algorithm (Table 2, Figure 4). Cluster 1 is created in space of seven attributes: hydrophobic moment, hydrophobicity, isoelectric point, tilt angle, linear moment, propensity to disordering and propensity to aggregation, Cluster 2 is created in space of three attributes: isoelectric point, penetration depth and tilt angle. Cluster 3 is created in space of seven attributes: hydrophobic moment, hydrophobicity, isoelectric point, penetration depth, tilt angle, propensity to disordering and propensity to aggregation. Only two properties, isoelectric point and tilt angle, are attributes which determine space of all three clusters. As we can see from

Table 2. Mean Values and Standard Deviation of Nine Physical-Chemical Characteristics for the Optimized Clusters for Peptides of 10−16 aa Length

	mean values ± SD of attributes		
	Cluster 1 (MHIORLA[a])	Cluster 2 (IDO[a])	Cluster 3 (MHIDORA[a])
M ± σ	**1.5 ± 0.6**	0.6 ± 0.2	**1.4 ± 0.2**
H ± σ	**−0.5 ± 0.7**	−0.4 ± 0.5	**−0.6 ± 0.4**
C ± σ	5.4 ± 2.1	4.9 ± 1.5	3.2 ± 1.7
I ± σ	**13.8 ± 0.5**	**12.7 ± 0.8**	**10.9 ± 0.4**
D ± σ	14.2 ± 4.5	**17.6 ± 2.1**	**13.8 ± 1.2**
O ± σ	**90.9 ± 15.6**	**162.6 ± 12.9**	**94.0 ± 9.8**
R ± σ	**−0.1 ± 0.5**	−0.5 ± 0.3	**0.3 ± 0.2**
L ± σ	**0.3 ± 0.1**	0.3 ± 0.1	0.3 ± 0.1
A ± σ	**2.1 ± 8.8**	64.4 ± 157.8	**5.7 ± 10.3**

[a]Attributes that characterize space where cluster was formed. Mean and SD of these attributes are marked in bold.

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

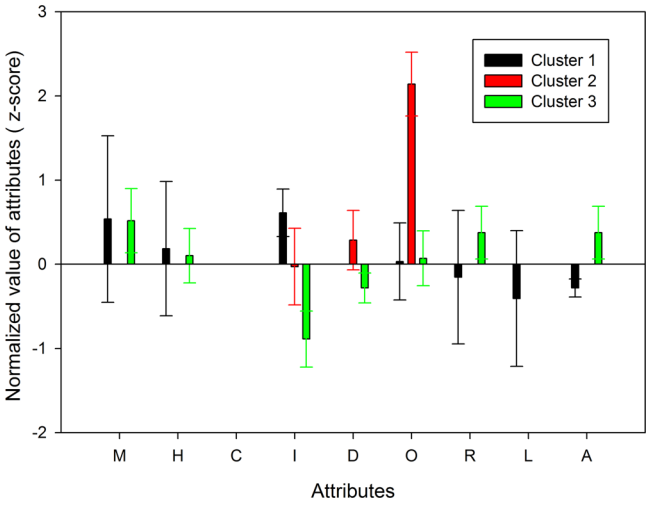


Figure 4. Mean values and standard deviation for normalized attributes that characterize clusters of instances for peptides of 10−16 aa length.

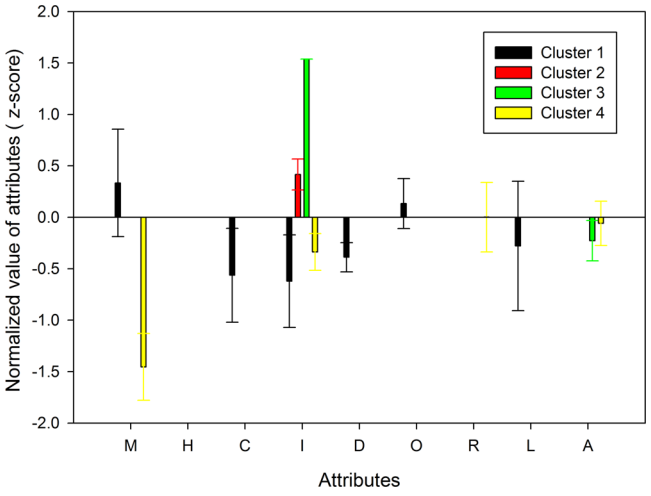


Figure 5. Mean values and standard deviation for normalized attributes that characterize clusters of instances for peptides of 18−27 aa length.

Figure 4, peptides from different clusters can have similar values for most properties and clearly differ by only several characteristics. All three clusters most strongly differ from each other by isoelectric point: peptides from Cluster 1 have the largest value of I. Then follow peptides from Clusters 2 and 3, correspondently. The peptides from Clusters 2 also differ from other two Clusters by tilt angle. Peptides from this Cluster are located almost perpendicular to the surface of the membrane, while peptides from other clusters are almost parallel. We can also note that most peptides from this cluster have lower hydrophobic moment and higher linear moment than peptides from other clusters (see Table 2 and Figure 4).

Cluster 1 is the largest cluster, and most of the peptides considered in the training sets fall into this cluster.

Clustering of Instances of Peptides of 18−27 aa Length. Four clusters were uncovered after optimization. Their physical-chemical characteristics are given in Table 3 and Figure 5. Cluster 1 is created in the space of five attributes: hydrophobic moment, charge, isoelectric point, penetration depth, tilt angle, and linear moment. Cluster 2 is created in space of one attribute: isoelectric point. Cluster 3 is created in space of two attributes: isoelectric point and propensity to aggregation. Cluster 4 is created in space of four attributes: hydrophobic moment, isoelectric point, propensity to disordering, and propensity to aggregation. As well as for peptides of 10−16 aa length, isoelectric point is the only attribute in space of which all four clusters were determined, and the majority of the clusters most strongly differ from each other by isoelectric point: peptides from Clusters 1 and 4 have similar and the lowest values of I, then follow peptides from Cluster 2, and peptides from Cluster 3 have the largest value of I. Clusters 1 and 4 differ from each other by hydrophobic moment (see Table 3 and Figure 5).

***In Silico* Testing of the Predictive Model.** *Selection of Strains for in Silico Testing.* Common features of AMPs make them effective against a broad range of microbial species and provide selectivity to prokaryotic membranes. At the same time, particular peptides may have different activities against different pathogenic bacteria. DBAASP allows us to estimate similarity of behavior of different microbial cells against a particular set of peptides. The corresponding estimations let us define if the method developed for *E. coli* ATCC 25922 can be applicable to other Gram-negative bacteria as well. These evaluations can define pairs of species which behaved similarly for the same sets of AMPs. If such pairs appeared, the predictive model developed on the basis of the data of one of them may be used for the other also.

We have analyzed the behavior of different species (Gram-positive, Gram-negative, fungi) to assess similarity between

Table 3. Mean Values and Standard Deviation of Nine Physical-Chemical Characteristics for the Optimized Clusters for Peptides of 18−27 aa Length

	mean values ± SD of attributes			
	Cluster 1 (MCIDOL[a])	Cluster 2 (I[a])	Cluster 3 (IA[a])	Cluster 4 (MIRA[a])
M ± σ	**1.2 ± 0.2**	0.9 ± 0.4	1.2 ± 0.5	**0.5 ± 0.1**
H ± σ	−0.4 ± 0.4	−0.4 ± 0.6	−0.3 ± 0.5	−0.7 ± 0.3
C ± σ	**4.2 ± 1.2**	7.9 ± 1.6	7.2 ± 2.6	6.1 ± 1.4
I ± σ	**11.4 ± 0.5**	**12.6 ± 0.2**	**14.0 ± 0.0**	**11.7 ± 0.2**
D ± σ	**13.8 ± 0.8**	20.4 ± 7.3	15.4 ± 6.1	13.6 ± 4.6
O ± σ	**88.3 ± 5.8**	91.3 ± 30.3	88.5 ± 22.6	75.1 ± 22.7
R ± σ	0.1 ± 0.2	−0.4 ± 0.3	−0.1 ± 0.4	**0.0 ± 0.1**
L ± σ	**0.2 ± 0.1**	0.3 ± 0.1	0.2 ± 0.1	0.4 ± 0.1
A ± σ	40.1 ± 161.2	23.5 ± 46.9	**10.1 ± 24.4**	**31.3 ± 26.9**

[a]Attributes that characterize space where cluster was formed. Mean and SD of these attributes are marked in bold.

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

Table 4. Coefficients Characterize the Similarity of Behavior of Microbes against the Particular Set of Peptides

	target organism				
	Klebsiella pneumoniae	*Pseudomonas aeruginosa* ATCC 27853	*Candida albicans* ATCC 90028	*Bacillus subtilis*	*Staphylococcus aureus* ATCC 25923
E. coli ATCC 25922	78.20	72.64	1.28	57.4	53.8
K. pneumoniae		69.70	−12.5	ND[a]	12.8
P. aeruginosa ATCC 27853			15.8	47.9	47.0
C. albicans ATCC 90028				ND[a]	35.2
B. subtilis					33.2
S. aureus ATCC 25923					

[a]ND, not determined due to insufficient data.

Table 5. Results of Prediction on the Training and Test Sets of Peptides of 10−16 aa Length

	TP	TP + FN	FP	TN + FP	SN	SP	AC	PPV	MCC
				Training Set T16					
Cluster 1	85	140	23	140					
Cluster 2	15	140	3	140					
Cluster 3	17	140	4	140					
all clusters	**117**	**140**	**30**	**140**	**0.84**	**0.79**	**0.81**	**0.80**	**0.62**
				Test Set E16					
Cluster 1	23	34	3	34					
Cluster 2	1	34	0	34					
Cluster 3	1	34	2	34					
all clusters	**25**	34	**5**	34	**0.74**	**0.85**	**0.79**	**0.83**	**0.59**
				Test Set P16					
Cluster 1	29	40	14	40					
Cluster 2	2	40	1	40					
Cluster 3	0	40	0	40					
all clusters	**31**	**40**	**15**	**40**	**0.78**	**0.63**	**0.70**	**0.67**	**0.40**

Table 6. Results of Prediction on Training and Test Sets of Peptides of 18−27 aa Length

	TP	TP + FN	FP	TN + FP	SN	SP	AC	PPV	MCC
				Training Set T27					
Cluster 1	32	100	9	100					
Cluster 2	21	100	5	100					
Cluster 3	16	100	5	100					
Cluster 4	10	100	0	100					
all clusters	**79**	**100**	**19**	**100**	**0.79**	**0.81**	**0.80**	**0.81**	**0.60**
				Test Set E27					
Cluster 1	5	18	2	18					
Cluster 2	2	18	0	18					
Cluster 3	7	18	2	18					
Cluster 4	0	18	0	18					
all clusters	**14**	18	**4**	18	**0.78**	**0.82**	**0.80**	**0.78**	**0.56**
				Test Set P27					
Cluster 1	5	32	4	32					
Cluster 2	4	32	0	32					
Cluster 3	15	32	7	32					
Cluster 4	0	32	0	32					
all clusters	**24**	**32**	**11**	**32**	**0.75**	**0.66**	**0.70**	**0.69**	**0.41**

them. For the assessment of similarity, pairs of species were screened, and peptides were selected from DBAASP in such a way that susceptibilities for both similarities were known. For a selected set of peptides, the ranking of susceptibilities (MIC) for each microbe was performed, and the Spearman's rank correlation coefficient was calculated. Spearman's rank correlation coefficients for particular pairs of target organisms are presented in Table 4.

The upper left elements of Table 4 describe the correlations between Gram-negative bacteria, demonstrating obvious like-

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

Table 7. Prediction Metrics Assessed on the Basis of Unification of Validation Sets of *E. coli* ATCC 25922

	SN	SP	AC	PPV	MCC
training set	0.80 ± 0.01	0.82 ± 0.004	0.81 ± 0.004	0.81 ± 0.01	0.62 ± 0.01
test set	0.77 ± 0.06	0.84 ± 0.03	0.80 ± 0.02	0.83 ± 0.02	0.61 ± 0.04

Table 8. Comparison of Predictions of Different Methods on the Test Sets U, Composed of Peptides of 10-27 aa Length and Constructed on the Basis of Data of *E. coli* ATCC 25922

	method of prediction									
	CAMP-SVM[56]	CAMP-RF[56]	CAMP-ANN[56]	CAMP-DA[56]	ADAM[57]	MLAMP[58]	DBASSP[21]	AMPA[30]	AMpred[12]	this study
SN	0.87	0.98	0.96	0.96	1.00	0.90	0.90	0.58	0.98	0.75
SP	0.27	0.25	0.23	0.21	0.04	0.21	0.33	0.87	0.13	0.83
BAC	0.57	0.62	0.60	0.59	0.52	0.62	0.62	0.73	0.56	**0.79**

ness between them. Although relatively low, some correlation is also seen for two Gram-positive bacteria (lower right part of the table) and between a few mixed (Gram-negative and Gram-positive) pairs as well. For instance, although *Bacillus subtilis* is a Gram-positive bacterium, the correlation with *Staphylococcus aureus* ATCC 25923, another Gram-positive bacterium, is less than the correlation with *E. coli* ATCC 25922, a Gram-negative bacterium. At the same time, the fungus, *Candida abligans* ATCC 90028, does not correlate with any bacterium but has some correlation with *S. aureus* ATCC 25923.

On the basis of these results, we can conclude that some common features responsible for antimicrobial activity exist, especially for Gram-negative bacteria, and so, the method developed for prediction of AMPs against *E. coli* ATCC 25922 can also be used for the prediction of AMPs active against other Gram-negative bacteria. Based on this observation, in the testing sets we used, along with the peptides that are active against *E. coli* ATCC 25922 and are not included into the training set, the peptides containing antimicrobial activity data against *P. aeruginosa* ATCC 27853 for which there are enough experimental data.

In Silico Testing. The following test sets were used for peptides of 10−16 aa length: Set E16, on the basis of data of *E. coli* ATCC 25922, with 34 peptides in the positive set and 34 peptides in the negative set, and Set P16, on the basis of data of *P. aeruginosa* ATCC 27853, with 40 peptides in the positive set and 40 peptides in the negative set.

Predictions for the AMPs of 10−16 aa length are shown in Table 5. Prediction accuracy for training and test sets for *E. coli* ATCC 25922 is 0.81 and 0.79, correspondingly. For test sets of *P. aeruginosa* ATCC 27853 prediction accuracy is lower, 0.70. Some differences in compositions of cell envelopes can be one of the causes of the lowering of the prediction accuracy.

The following test sets were used for peptides of 18−27 aa length: Set E27 was constructed on the basis of data of *E. coli* ATCC 25922, with 18 peptides in the positive set and 18 peptides in the negative set; and Set P27 was constructed on the basis of data of *P. aeruginosa* ATCC 27853, with 32 peptides in the positive set and 32 peptides in the negative set.

Prediction results for the peptides of 18−27 aa length have been presented in Table 6.

For the additional validation, the prediction metrics for the united sets of peptides of 10−16 and 18−27 aa length, were calculated. Using the method of separation of the available data into multiple randomly generated training and test sets,[55] cross validation has been performed. Combination of five training and test sets of 10−16 aa ($T16_i$, $E16_i$, i = 1, ..., 5) and five sets of 18−27 aa ($T27_j$, $E27_j$, j = 1, ..., 5) peptide length, gives 25 sets, which are used to assess united prediction metrics. The obtained results are presented in Table 7.

Comparison with Other AMP Prediction Tools. Traditional machine-learning approaches are developed to discriminate AMP from non-AMP (randomly selected peptide fragments or small proteins[21,56−58]). Our goal was to discriminate highly active AMP (MIC < 25 μg/mL) from AMP with low activity (MIC > 100 μg/mL). The solution of the last task was complicated, given the small number of data on susceptibilities for particular strains and the low standardization of *in vitro* susceptibility testing experiments. Thus, it is difficult to perform a comprehensive comparison of our tool with other AMP prediction tools because they have different goals. However, in order to show the necessity of the prediction tool developed in this study, we compared the results of AMP prediction performed by popular predictive tools with the results obtained by our tool on the test set U, constructed by unification of test sets of peptides of 10−16 and 18−27 aa length created on the basis of data of *E. coli* ATCC 25922 (see Tables S1 and S2). Table 8 presents the results of comparison of predictions performed by different predictive tools. As we can see, although sensitivities of the most tools are high, they show very low specificity values. The highest value of balance accuracy was obtained by our tool. AMPA also shows high values of specificities and balance accuracies. The last results can be explained by the fact that this tool is based on the data for particular strain of Gram-negative *P. aeruginosa*. The results of comparison of the same predictive tools on the test sets constructed on the basis of data of *P. aeruginosa* ATCC 27853 are similar.[59] So, the results of comparison clearly show the necessity in the demands of predictive models which would be able to uncover the potency of peptides against particular strain.

Thus, our algorithm can well distinguish peptides active against particular strains from others, which may also be active but against other strains. The other AMP prediction tools cannot carry out this task.

DISCUSSION AND CONCLUSIONS

Gram-negative bacteria cause various infections in healthcare settings, including pneumonia, bloodstream infections, wound or surgical site infections, and meningitis. Multidrug-resistant (MDR) bacterial infections, especially those caused by Gram-negative pathogens, are an emerging threat for public health all over the world. Infections have typically been treated with a broad spectrum of antibiotics, such as β-lactams followed by administration of carbapenems. However, even these drugs have become ineffective against certain bacterial strains.[60]

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

Consequently, new drugs are needed to combat Gram-negative bacterial infections.

Formulation of a task-oriented approach to design desired AMPs against drug-resistant strains requires the development of a method that can predict a peptide being active against particular strains. To the best of our knowledge, the majority of the tools[21,56−58] are developed to predict the activities of peptides against bacteria and can predict the antimicrobial potency without specification of the bacterial targets against which a peptide can be active.

A semi-supervised machine-learning approach relying on a density-based clustering algorithm has been used to develop a predictive model of linear AMP activity against Gram-negative bacteria. Training sets of peptides were made from data of AMPs having potency against *E. coli* ATCC 25922. For peptides of 10−16 aa length, the predictive model suggests the availability of three types of peptides distinguished by their physical-chemical features. For the sets of peptides of 18−27 in aa length, four groups of peptides were identified.

Accuracies of *in silico* testing for test sets having potency against *E. coli* ATCC 25922 were equal to 0.79 and 0.80, and PPV = 0.83 and 0.78 for peptides of 10−16 and 18−27 aa length, respectively. For the united test sets, AC = 0.80 ± 0.02 and PPV = 0.83 ± 0.02. The quality of prediction for *P. aeruginosa* is lower: AC = 0.7. Sensitivities of the prediction for *P. aeruginosa* have close values to the prediction for *E. coli* ATCC 25922, but specificities for test sets of *P. aeruginosa* have rather lower values compared to those for test sets of *E. coli* ATCC 25922. This means that some of the peptides that are predicted as active by the proposed algorithm will be active against *E. coli* ATCC 25922 but not active against *P. aeruginosa*. In general, we can conclude that, although both strains belong to Gram-negative bacteria, there are some differences in the cell envelope organization, and so in order to increase the efficiency of prediction, additional data are required to develop a predictive model just for *P. aeruginosa*.

The predictive model takes into consideration similarities and differences between the groups of peptides revealed in this study. The results show that peptides from different clusters can resemble each other in most properties. The major properties that allow us to separate peptides into different groups are isoelectric point, tilt angle, and hydrophobic moment. Peptides from different clusters are mainly distinguished by the values of isoelectric points (peptides of Cluster 1 from peptides of Cluster 3 in the case of peptides of 10−16 aa length, and peptides of Clusters 1 and 4 from peptides of Clusters 2 and 3 in the case of peptides of 28−27 aa length). Another property that separates peptides is tilt angle. Major peptides (from Clusters 1 and 3 of peptides of 10−16 aa length, and Clusters 1, 2, 3, and 4 of peptides of 18−27 aa length) are oriented parallel to the surface of the membrane (angle relative to the normal of the membrane equals 90°). Only members of Cluster 2 of peptides of 10−16 aa length have a tendency toward non-parallel orientation. Peptides of Cluster 4 differ from peptides of Cluster 1 of 18−27 aa length by hydrophobic moment. We can also note that peptides from this cluster have rather high values of linear moment. Therefore, we can suggest that amphipathicities of these peptides can be connected to the distribution of hydrophobicity along the peptide sequence (linear separation of polar and apolar parts). So our study points to the existence of different groups of peptides characterized by a low or high hydrophobic moment, parallel or non-parallel orientation to the surface of membrane, and a low or high isoelectric point. These differences can be the reason for the variations in the modes of interaction with the membrane.

The peptide−membrane interactions result from charge, hydrophobicity, hydrophobic moment, orientation relative to the membrane surface, and propensity to self-aggregation of the peptide. These parameters contribute to how strongly the peptides interact, how deeply they insert, and how much disordering and curvature strain they exert on the membrane. It means that these parameters determine modes of action of peptides. For instance, if peptides are characterized by a high hydrophobicity and only modest hydrophobic moments, they insert more easily into the membrane interior.[61] At the same time, the partitioning of liquid-disordered membranes by cationic sequences results in the well-known decrease in the order of parameters of membrane, while the changes due to a large number of hydrophobic peptides (peptaibol and alamethicin, for example) are much more modest.[62] The results of clustering show that isoelectric point is the only property which characterizes all clusters for different peptide lengths. Most clusters are distinguished from each other also by this property. It means that there are groups of peptides that have differently charged state and various affinity with the same anionic membrane and different propensities toward aggregation. Peptides with low charge can form ensembles of structurally diverse channel-like pores in the membrane,[63] whereas AMPs with high charge do not aggregate at all and behave more like cell-penetrating sequences that appear to be able to stealthily translocate across the membrane without large pore formation.[64] It means that there are groups of peptides that have different affinity with the same anionic membrane.

Under physiological conditions, a biological membrane is a stressed, non-homogenous structure because it consists of lipids of different spontaneous curvature and phase transition temperature. Such circumstances force the gathering of preferentially interacting lipids into particular domains with the accompanying phase transitions, which leads to phase separation events.[65] The clustering of anionic lipids to a region of the bacterial membrane will concentrate negative charge in a domain to which cationic peptides will congregate. A local domain of a membrane is distinguished by surface curvature that influences the mode of interaction of peptides with the membrane. Amphipathic peptides sense membrane curvature and lipid composition.[46] At the same time, cationic amphipathic peptides can change membrane curvature and, depending on the membrane curvature, can change orientation relative to the membrane.[66] An intrinsic propensity of a peptide to particular orientation can determine the site of interaction with the membrane and consequently the mode of action. The set of peptides considered in this paper is clustered by orientations relative to membrane surface. For instance, peptides from Cluster 2 of 10−16 aa length differ from the others by tilt angle. They prefer an orientation which is nearly perpendicular to the membrane surface, unlike the peptides from other clusters, and so the peptides can sense particular domains of the membrane. There is a widespread view that the shape of the bilayer can play a key role in the orientation of peptides, and for the domains of phospholipids with positive spontaneous curvature, peptides can be located at high tilt angle to the membrane surface.[46] Lipid domains which can recruit certain AMPs are the important structural components of a membrane, playing a key role in many cellular processes. So, along with the traditional membrane disrupting modes of action, other modes

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

can be considered that entail cell death by modulating key components of the cellular machinery[67] without, for example, creation of pores.

The clustering data can be used as an additional argument to support inferences about different modes of action of antimicrobial peptides. Additional data are needed to better understand and predict possible mechanisms. As the number of experimentally proven AMPs increases, we anticipate a greater understanding of AMPs' mechanisms of action in the near future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00118.

Tables S1 and S2, showing sequences included in test set U, composed of peptides 10−27 aa length, constructed on the basis of data of *E. coli* ATCC 25922 and used for the comparison of predictions of different methods (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: b.vishnepolsky@lifescience.org.ge.
*E-mail: m.pirtskhalava@lifescience.org.ge.

ORCID

Boris Vishnepolsky: 0000-0001-5770-9481
George I. Makhatadze: 0000-0003-4565-1264

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the International Science and Technology Center (Grant No. G-2102) and Shota Rustaveli National Science Foundation (Grant Nos. FR/397/7-180/14 and DI-2016-9).

■ REFERENCES

(1) Almeida, P. F.; Pokorny, A. Mechanisms of Antimicrobial, Cytolytic, end Cell-Penetrating Peptides: From Kinetics to Thermodynamics. *Biochemistry* **2009**, *48*, 8083−8093.

(2) Paul, M.; Carmeli, Y.; Durante-Mangoni, E.; Mouton, J. W.; Tacconelli, E.; Theuretzbacher, U.; Mussini, C.; Leibovici, L. Combination Therapy for Carbapenem-Resistant Gram-Negative Bacteria. *J. Antimicrob. Chemother.* **2014**, *69*, 2305−2309.

(3) Huwaitat, R.; McCloskey, A. P.; Gilmore, B. F.; Laverty, G. Potential Strategies for the Eradication Of Multidrug-Resistant Gram-Negative Bacterial Infections. *Future Microbiol.* **2016**, *11*, 955−72.

(4) Jenssen, H.; Fjell, C. D.; Cherkasov, A.; Hancock, R. E. QSAR Modeling and Computer-Aided Design of Antimicrobial Peptides. *J. Pept. Sci.* **2008**, *14*, 110−114.

(5) Fjell, C. D.; Jenssen, H.; Hilpert, K.; Cheung, W. A.; Pante, N.; Hancock, R. E.; Cherkasov, A. Identification of Novel Antibacterial Peptides by Chemoinformatics and Machine Learning. *J. Med. Chem.* **2009**, *52*, 2006−2015.

(6) Cherkasov, A.; Hilpert, K.; Jenssen, H.; Fjell, C. D.; Waldbrook, M.; Mullaly, S. C.; Volkmer, R.; Hancock, R. E. Use of Artificial Intelligence in the Design of Small Peptide Antibiotics Effective against a Broad Spectrum of Highly Antibiotic-Resistant Superbugs. *ACS Chem. Biol.* **2009**, *4*, 65−74.

(7) Torrent, M.; Andreu, D.; Nogues, V. M.; Boix, E. Connecting Peptide Physicochemical and Antimicrobial Properties by a Rational Prediction Model. *PLoS One* **2011**, *6*, e16968.

(8) Mooney, C.; Haslam, N. J.; Holton, T. A.; Pollastri, G.; Shields, D. C. Peptidelocator: Prediction of Bioactive Peptides in Protein Sequences. *Bioinformatics* **2013**, *29*, 1120−1126.

(9) Porto, W. F.; Pires, A. S.; Franco, O. L. Cs-Amppred: An Updated SVM Model for Antimicrobial Activity Prediction in Cysteine-Stabilized Peptides. *PLoS One* **2012**, *7*, e51444.

(10) Ng, X. Y.; Rosdi, B. A.; Shahrudin, S. Prediction of Antimicrobial Peptides Based on Sequence Alignment and Support Vector Machine-Pairwise Algorithm Utilizing LZ-Complexity. *BioMed Res. Int.* **2015**, *2015*, 212715.

(11) Khosravian, M.; Faramarzi, F. K.; Beigi, M. M.; Behbahani, M.; Mohabatkar, H. Predicting Antibacterial Peptides by the Concept of Chou's Pseudo-Amino Acid Composition and Machine Learning Methods. *Protein Pept. Lett.* **2013**, *20*, 180−186.

(12) Meher, P. K.; Sahu, T. K.; Saini, V.; Rao, A. R. Predicting Antimicrobial Peptides with Improved Accuracy by Incorporating the Compositional, Physico-Chemical and Structural Features into Chou's General PseAAC. *Sci. Rep.* **2017**, *7*, 42362.

(13) Lira, F.; Perez, P. S.; Baranauskas, J. A.; Nozawa, S. R. Prediction of Antimicrobial Activity of Synthetic Peptides by a Decision Tree Model. *Appl. Environ. Microbiol.* **2013**, *79*, 3156−3159.

(14) Khamis, A. M.; Essack, M.; Gao, X.; Bajic, V. B. Distinct Profiling of Antimicrobial Peptide Families. *Bioinformatics* **2015**, *31*, 849−856.

(15) Xiao, X.; Wang, P.; Lin, W. Z.; Jia, J. H.; Chou, K. C. Iamp-2l: A Two-Level Multi-Label Classifier for Identifying Antimicrobial Peptides and Their Functional Types. *Anal. Biochem.* **2013**, *436*, 168−177.

(16) Maccari, G.; Di Luca, M.; Nifosí, R.; Cardarelli, F.; Signore, G.; Boccardi, C.; Bifone, A. Antimicrobial Peptides Design by Evolutionary Multiobjective Optimization. *PLoS Comput. Biol.* **2013**, *9*, e1003212.

(17) Bhadra, P.; Yan, J.; Li, J.; Fong, S.; Siu, S. W. I. AmPEP: Sequence-based Prediction of Antimicrobial Peptides using Distribution Patterns of Amino Acid Properties and Random Forest. *Sci. Rep.* **2018**, *8*, 1697.

(18) Youmans, M.; Spainhour, C.; Qiu, P. Long Short-Term Memory Recurrent Neural Networks for Antibacterial Peptide Identification. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017; pp 498−502.

(19) Juretic, D.; Vukicevic, D.; Ilic, N.; Antcheva, N.; Tossi, A. Computational Design of Highly Selective Antimicrobial Peptides. *J. Chem. Inf. Model.* **2009**, *49*, 2873−2882.

(20) Wang, P.; Hu, L.; Liu, G.; Jiang, N.; Chen, X.; Xu, J.; Zheng, W.; Li, L.; Tan, M.; Chen, Z.; Song, H.; Cai, Y. D.; Chou, K. C. Prediction of Antimicrobial Peptides Based on Sequence Alignment and Feature Selection Methods. *PLoS One* **2011**, *6*, e18476.

(21) Vishnepolsky, B.; Pirtskhalava, M. Prediction of Linear Cationic Antimicrobial Peptides Based on Characteristics Responsible for Their Interaction with the Membranes. *J. Chem. Inf. Model.* **2014**, *54*, 1512−1523.

(22) Melo, M. N.; Ferre, R.; Feliu, L.; Bardaji, E.; Planas, M.; Castanho, M. A. Prediction of Antibacterial Activity from Physicochemical Properties of Antimicrobial Peptides. *PLoS One* **2011**, *6*, e28549.

(23) Freire, J. M.; Almeida Dias, S.; Flores, L.; Veiga, A. S.; Castanho, M. A. Mining Viral Proteins for Antimicrobial and Cell-Penetrating Drug Delivery Peptides. *Bioinformatics* **2015**, *31*, 2252−2256.

(24) Chang, K. Y.; Lin, T. P.; Shih, L. Y.; Wang, C. K. Analysis and Prediction of the Critical Regions of Antimicrobial Peptides Based on Conditional Random Fields. *PLoS One* **2015**, *10*, e0119490.

(25) Toropova, M. A.; Veselinovic, A. M.; Veselinovic, J. B.; Stojanovic, D. B.; Toropov, A. A. QSAR Modeling of the Antimicrobial Activity of Peptides as a Mathematical Function of a Sequence of Amino Acids. *Comput. Biol. Chem.* **2015**, *59*, 126−130.

(26) Toropov, A. A.; Toropova, A. P.; Raska, I., Jr.; Benfenati, E.; Gini, G. QSAR Modeling of Endpoints for Peptides Which Is Based on Representation of the Molecular Structure by a Sequence of Amino Acids. *Struct. Chem.* **2012**, *23*, 1891−1904.

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

(27) Lata, S.; Sharma, B. K.; Raghava, G. P. Analysis and Prediction of Antibacterial Peptides. *BMC Bioinf.* **2007**, *8*, 263.
(28) Lata, S.; Mishra, N. K.; Raghava, G. P. AntiBP2: Improved Version of Antibacterial Peptide Prediction. *BMC Bioinf.* **2010**, *11*, S19.
(29) Fjell, C. D.; Hiss, J. A.; Hancock, R. E.; Schneider, G. Designing Antimicrobial Peptides: Form Follows Function. *Nat. Rev. Drug Discovery* **2012**, *11*, 37−51.
(30) Torrent, M.; Di Tommaso, P.; Pulido, D.; Nogués, M. V.; Notredame, C.; Boix, E.; Andreu, D. AMPA: An Automated Web Server for Prediction of Protein Antimicrobial Regions. *Bioinformatics* **2012**, *28*, 130−131.
(31) Jenssen, H.; Lejon, T.; Hilpert, K.; Fjell, C. D.; Cherkasov, A.; Hancock, R. E. W. Evaluating Different Descriptors for Model Design of Antimicrobial Peptides with Enhanced Activity toward P. aeruginosa. *Chem. Biol. Drug Des.* **2007**, *70*, 134−142.
(32) Taboureau, O.; Olsen, O. H.; Nielsen, J. D.; Raventos, D.; Mygind, P. H.; Kristensen, H. H. Design of Novispirin Antimicrobial Peptides by Quantitative Structure-Activity Relationship. *Chem. Biol. Drug Des.* **2006**, *68*, 48−57.
(33) Kessel, A.; Ben-Tal, N. *Introduction to Proteins: Structure, Function and Motion*, 1st ed.; Chapman & Hall/CRC Mathematical and Computational Biology; CRC Press, Taylor & Francis Group: London, UK, 2011.
(34) Pirtskhalava, M.; Gabrielian, A.; Cruz, P.; Griggs, H. L.; Squires, R. B.; Hurt, D. E.; Grigolava, M.; Chubinidze, M.; Gogoladze, G.; Vishnepolsky, B.; Alekseev, V.; Rosenthal, A.; Tartakovsky, M. DBAASP v.2: an Enhanced Database of Structure and Antimicrobial/Cytotoxic Activity of Natural and Synthetic Peptides. *Nucleic Acids Res.* **2016**, *44*, D1104−D1112.
(35) Li, Z.; Liu, G.; Meng, L.; Yu, W.; Xu, X.; Li, W.; Wu, Y.; Cao, Z. K1K8: An Hp1404-Derived Antibacterial Peptide. *Appl. Microbiol. Biotechnol.* **2016**, *100* (11), 5069−77.
(36) Speck-Planche, A.; Kleandrova, V. V.; Ruso, J. M.; Cordeiro, M. N. D. S. First Multitarget Chemo-Bioinformatic Model To Enable the Discovery of Antibacterial Peptides against Multiple Gram-Positive Pathogens. *J. Chem. Inf. Model.* **2016**, *56*, 588−598.
(37) Kleandrova, V. V.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N. D. S. Enabling the Discovery and Virtual Screening of Potent and Safe Antimicrobial Peptides. Simultaneous Prediction of Antibacterial Activity and Cytotoxicity. *ACS Comb. Sci.* **2016**, *18*, 490−498.
(38) Speck-Planche, A.; Dias Soeiro Cordeiro, M. N. Speeding up Early Drug Discovery in Antiviral Research: A Fragment-Based in Silico Approach for the Design of Virtual Anti-Hepatitis C Leads. *ACS Comb. Sci.* **2017**, *19*, 501−512.
(39) Romero-Duran, F. J.; Alonso, N.; Yanez, M.; Caamano, O.; Garcia-Mera, X.; Gonzalez-Diaz, H. Brain-Inspired Cheminformatics of Drug-Target Brain Interactome, Synthesis, and Assay of Tvp1022 Derivatives. *Neuropharmacology* **2016**, *103*, 270−278.
(40) Speck-Planche, A.; Cordeiro, M. N. D. S. Simultaneous Virtual Prediction of Anti-Escherichia coli Activities and ADMET Profiles: A Chemoinformatic Complementary Approach for High-Throughput Screening. *ACS Comb. Sci.* **2014**, *16*, 78−84.
(41) Speck-Planche, A.; Cordeiro, M. N. D. S. Chemoinformatics for Medicinal Chemistry: In Silico Model to Enable the Discovery of Potent and Safer Anti-Cocci Agents. *Future Med. Chem.* **2014**, *6*, 2013−2028.
(42) Speck-Planche, A.; Kleandrova, V. V.; Scotti, M. T. Fragment-Based Approach for the in Silico Discovery of Multi-Target Insecticides. *Chemom. Intell. Lab. Syst.* **2012**, *111*, 39−45.
(43) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Unified Multi-Target Approach for the Rational in Silico Design of Anti-Bladder Cancer Agents. *Anti-Cancer Agents Med. Chem.* **2013**, *13*, 791−800.
(44) Cutrona, K. J.; Kaufman, B. A.; Figueroa, D. M.; Elmore, D. E. Role of Arginine and Lysine in the Antimicrobial Mechanism of Histone-derived Antimicrobial Peptides. *FEBS Lett.* **2015**, *589*, 3915−3920.
(45) Zelezetsky, O.; Tossi, A. Alpha-helical antimicrobial peptides—Using a sequence template to guide structure−activity relationship studies. *Biochim. Biophys. Acta, Biomembr.* **2006**, *1758*, 1436−1449.
(46) Strandberg, E.; Tiltak, D.; Ehni, S.; Wadhwani, P.; Ulrich, A. S. Lipid shape is a key factor for membrane interactions of amphipathic helical peptides. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818* (7), 1764−1776.
(47) Cruz-Monteagudo, M.; Medina-Franco, J. L.; Pérez-Castillo, Y.; Nicolotti, O.; Cordeiro, M. N.; Borges, F. Activity Cliffs in Drug Discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today* **2014**, *19*, 1069−1080.
(48) Cruz-Monteagudo, M.; Medina-Franco, J. L.; Perera-Sardiña, Y.; Borges, F.; Tejera, E.; Paz-Y-Miño, C.; Pérez-Castillo, Y.; Sánchez-Rodríguez, A.; Contreras-Posada, Z.; Cordeiro, M. N. Probing the Hypothesis of SAR Continuity Restoration by the Removal of Activity Cliffs Generators in QSAR. *Curr. Pharm. Des.* **2016**, *22*, 5043−5056.
(49) Moon, C. P.; Fleming, K. G. Side-Chain Hydrophobicity Scale Derived From Transmembrane Protein Folding into Lipid Bilayers. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10174−10177.
(50) Senese, A.; Chadi, D. C.; Law, P. B.; Walters, R. F. S.; Nanda, V.; DeGrado, W. F. E(z), a Depth-dependent Potential for Assessing the Energies of Insertion of Amino Acid Side-chains into Membranes: Derivation and Applications to Determining the Orientation of Transmembrane and Interfacial Helices. *J. Mol. Biol.* **2007**, *366*, 436−448.
(51) Uversky, V.; Gillespie, J.; Fink, A. Why are "Natively Unfolded" Proteins Unstructured Under Physiological Conditions? *Proteins: Struct., Funct., Genet.* **2000**, *41* (3), 415−427.
(52) Kessel, A.; Ben-Tal, N. Free Energy Determinants of Peptide Association with Lipid Bilayers. *Curr. Top. Membr.* **2002**, *52*, 205−253.
(53) Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*; AAAI Press, 1996; pp 226−231.
(54) Reynolds, R.; Shackcloth, J.; Felmingham, D.; MacGowan, A. Comparison of BSAC Agar Dilution and NCCLS Broth Microdilution MIC Methods for in vitro Susceptibility Testing of Streptococcus Pneumoniae, Haemophilus Influenzae and Moraxella Catarrhalis: the BSAC Respiratory Resistance Surveillance Programme. *J. Antimicrob. Chemother.* **2003**, *52*, 925−930.
(55) Martin, T. M.; Harten, P.; Young, D. M.; Muratov, E. N.; Golbraikh, A.; Zhu, H.; Tropsha, A. Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *J. Chem. Inf. Model.* **2012**, *52*, 2570−2578.
(56) Waghu, F. H.; Barai, R. S.; Gurung, P.; Idicula-Thomas, S. CAMPR3: a Database on Sequences, Structures and Signatures of Antimicrobial Peptides. *Nucleic Acids Res.* **2016**, *44*, D1094−D1097.
(57) Lee, H.-T.; Lee, C.-C.; Yang, J.-R.; Lai, J. Z. C.; Chang, K. Y. A Large-Scale Structural Classification of Antimicrobial Peptides. *BioMed Res. Int.* **2015**, *2015*, 475062.
(58) Lin, W.; Xu, D. Imbalanced Multi-Label Learning for Identifying Antimicrobial Peptides and their Functional Types. *Bioinformatics* **2016**, *32*, 3745−3752.
(59) Vishnepolsky, B.; Pirtskhalava, M. Comment on: 'Empirical Comparison of Web-Based Antimicrobial Peptide Prediction Tools'. *Bioinformatics* **2018**, in press.
(60) Versporten, A. Burden and Current Global Management on GNB Infections: Lessons from the Global-PPS of Antimicrobial Consumption and Resistance in 335 Hospitals Worldwide. *BSAC Spring Conference: The Global Challenge of Multi-drug Resistant Gram Negative Bacterial Infections*, London, March 14, 2017.
(61) Salnikov, E. S.; Mason, A. J.; Bechinger, B. Membrane Order Perturbation in the Presence of Antimicrobial Peptides by (2)H Solid-State NMR Spectroscopy. *Biochimie* **2009**, *91*, 734−43.
(62) Salnikov, E. S.; Friedrich, H.; Li, X.; Bertani, P.; Reissmann, S.; Hertweck, C.; O'Neil, J. D. J.; Raap, J.; Bechinger, B. Structure and Alignment of the Membrane-Associated Peptaibols Ampullosporin A and Alamethicin by Oriented ^{15}N and ^{31}P Solid-State NMR Spectroscopy. *Biophys. J.* **2009**, *96*, 86−100.

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151

(63) Wang, Y.; Chen, C. H.; Hu, D.; Ulmschneider, M. B.; Ulmschneider, J. P. Spontaneous Formation of Structurally Diverse Membrane Channel Architectures From a Single Antimicrobial Peptide. *Nat. Commun.* **2016**, *7*, 13535.
(64) Ulmschneider, J. P. Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores. *Biophys. J.* **2017**, *113*, 73−81.
(65) Feigenson, G. W. Phase Behavior of Lipid Mixtures. *Nat. Chem. Biol.* **2006**, *2*, 560−563.
(66) Vanni, S.; Hirose, H.; Barelli, H.; Antonny, B.; Gautier, R. A Sub-Nanometre View of how Membrane Curvature and Composition Modulate Lipid Packing and Protein Recruitment. *Nat. Commun.* **2014**, *5*, 4916.
(67) Sochacki, K. A.; Barns, K. J.; Bucki, R.; Weisshaar, J. C. Real-time attack on Single Escherichia Coli Cells By the Human Antimicrobial Peptide LL-37. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E77−E81.

DOI: 10.1021/acs.jcim.8b00118
J. Chem. Inf. Model. 2018, 58, 1141−1151