



5th International Electronic Conference on Medicinal Chemistry

1-30 November 2019

chaired by Dr. Jean Jacques Vanden Eynde

sponsored by



pharmaceuticals

Development of the model of in silico design of AMP sequences active against *Staphylococcus aureus* 25923

Boris Vishnepolsky^{1,*}, **George Zaalishvili**², **Margarita Karapetian**², **Andrei Gabrielian**³, **Alex Rosenthal**³, **Darrell E. Hurt**³, **Michael Tartakovsky**³, **Maya Grigolava**¹, and **Malak Pirtskhalava**^{1,*}

¹ I.Beritashvili Center of Experimental Biomedicine, Gotua str. 14, Tbilisi 0160 , Georgia

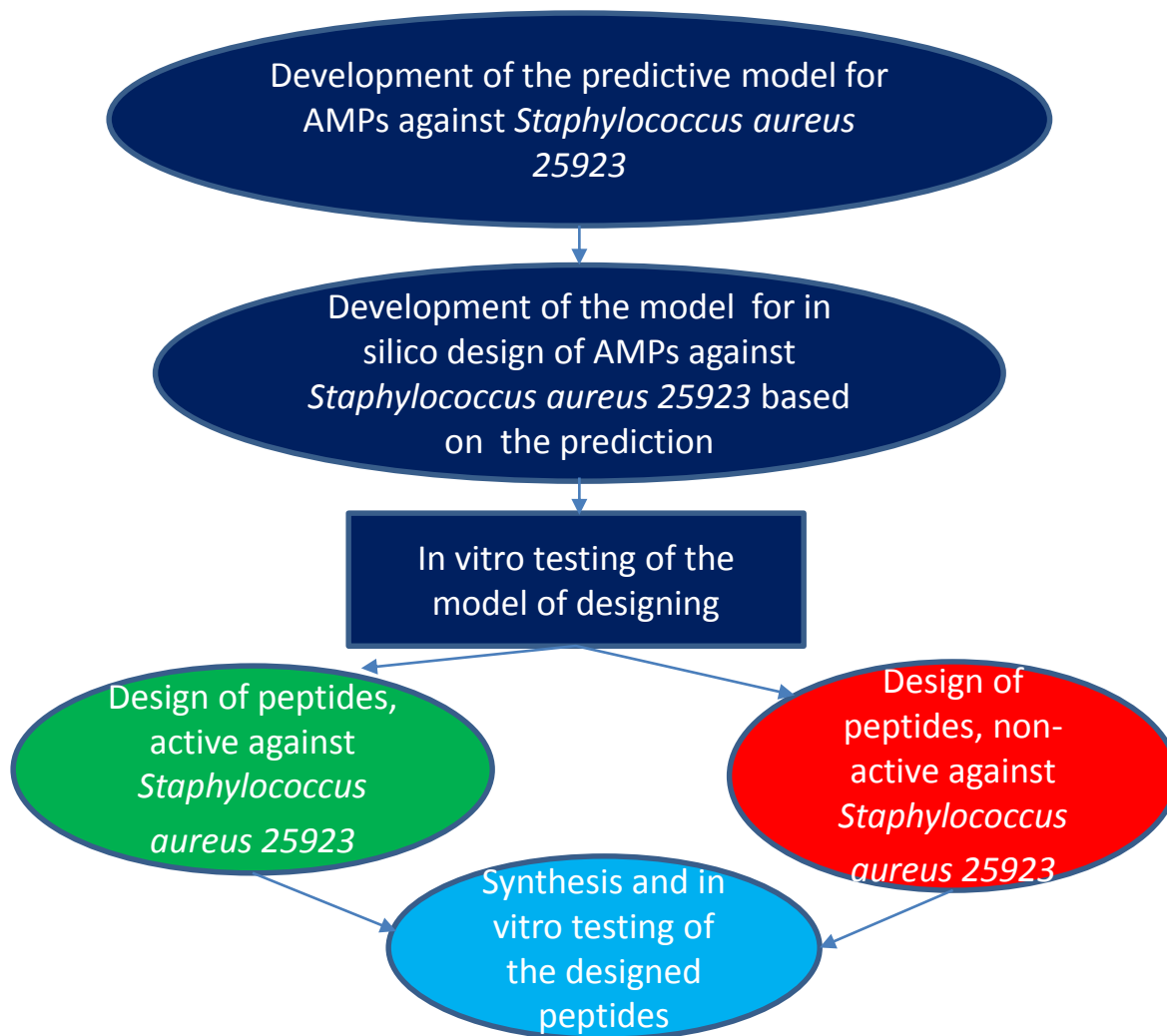
² Agricultural University of Georgia, 240 David Aghmashenebeli Alley, Tbilisi 0159, Georgia

³ Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

* Corresponding authors: b.vishnepolsky@lifescience.org.ge;
m.pirtskhalava@lifescience.org.ge



Graphical Abstract



Abstract

Emerging bacterial resistance to the existing antibiotics makes the development of new types of antibiotics an increasingly important challenge. Antimicrobial peptides (AMPs) can be considered as novel and efficient type of antibiotics that are hard to acquire resistance against. We have developed an algorithm to design peptides that are active against certain species. The prediction is based on clusterization of peptides with known biological activities by physicochemical properties. The Database of Antimicrobial Activity and Structure of Peptides (DBAASP, <https://dbaasp.org>) now includes Special Prediction (SP) tool, which allows to apply this algorithm to any amino acid sequence to predict whether this peptide is active against particular microbes. To verify the efficiency of the algorithm, we designed several variants of active peptides and tested them in vitro against two strains *Escherichia coli* ATCC 25922 and *Staphylococcus aureus* 25923. Prediction precision for the designed peptides against *Escherichia coli* ATCC was 95% and against *Staphylococcus aureus* was 68%. To improve prediction precision against *Staphylococcus aureus*, we applied the linear regression analysis based on binary classification. This approach allows us to improve the prediction precision of the peptides designed for *Staphylococcus aureus* 25923 up to 92%.



Introduction

- The problem of bacterial resistance to antibiotics is one of the important tasks in microbiology.
- Antimicrobial peptides (AMPs), also called host defense peptides (HDPs), are part of the innate immune response found among all classes of life.
- The efficacy of AMP over evolutionary time has been largely attributed to their mechanisms of action.
- AMPs are considered as an appropriate basis to develop new antibiotics against drug-resistant strains
- The demand for efficient tools for de novo designing of AMP against particular strains is valid again.



Introduction

- Recently prediction models against some microbial strains (Escherichia coli ATCC 25922, Staphylococcus aureus 25923, Bacillus subtilis) have been developed. The predictive model was based on clusterization peptides by physicochemical properties
- Models developed relied on the supposition that there exist several groups of peptides acting according to different mechanisms and so having different physicochemical properties.
- Optimization of the models is performed on the training and test sets of peptides selected from the Database of Antimicrobial Activity and Structure of Peptides(DBAASP, <https://dbaasp.org> [1]).



Introduction

- The set of peptides active against particular strain was divided into several clusters with different physicochemical properties [2]. But it turned out that only the volume of one cluster allows performing a statistically reliable prediction of sequences being active against the strain. So only data of statistically reliable clusters have been used for the in silico designing.
- Based on the statistically reliable clusters, some peptides active against *Escherichia coli* ATCC 25922 [3] and *Staphylococcus aureus* 25923 (unpublished) were designed, synthesized, and tested in vitro.
- Prediction precision for the designed peptides against *Escherichia coli* ATCC was 95% and against *Staphylococcus aureus* was 68%.



Introduction

- Not quite good precision in case of *Staphylococcus aureus* to develop an efficient predictive model to each group of peptides can be explained by insufficient data about each group.
- Consequently, In the current conditions, we think that it's reasonable to perform an in silico design of new sequences relying just on the approximation of the binary classification.
- To perform binary classification, we decided to rely on the regression model which permits to optimize the border between active and non-active instances to get an optimal threshold for efficient designing.



Results and discussion

Development of the Predictive model

The combinations of the following 12 physicochemical characteristics were used to present the sequences of AMP as n-mer vectors (instances), $n=1,2,\dots,12$

Hydrophobic moment (M)

Hydrophobicity (H)

Charge (C)

Isoelectric Point (I)

Penetration Depth (D)

Orientation of Peptides relative to the surface of membrane (O)

Propensity to Disordering (R)

Linear Moment (L)

***In vitro* aggregation (Tango) (T)**

Angle Subtended by the Hydrophobic Residue (S)

Amphiphilicity Index (A)

Propensity to Coil Conformation (P)

Peptide sequences active and non-active against *Staphylococcus aureus* 25923 were retrieved from DBAASP



Results and discussion

Development of the Predictive model

- For each i -th instance, activity value (AV_i) was defined. Instances corresponding to AMP with $MIC < 25 \mu\text{g/ml}$ were forming a positive training set with $AV_i = 1$. Instances corresponding to AMP with $MIC > 100 \mu\text{g/ml}$ were forming a negative training set with $AV_i = -1$. ($i=1, N$, N - number of the instances in the set)
- Both positive and negative training sets consist of 149 instances.
- Positive and negative test sets with 37 instances in each were formed by analogy with training sets
- The number of combinations of characteristics equals 4095. So the number of considered training and test sets of instances also equals 4095.

Results and discussion

Development of the Predictive model

- For each training set of instances, a standard linear model of regression has been used to optimize regression coefficients on the particular training set and to get optimal linear dependence between characteristics and PV in the form of the following equation

$$PV = b_0 + b_m M + b_h H + b_c C + b_i I + b_d D + b_{oi} + b_r R + b_l L + b_a T + b_s S + b_a A + b_p P$$

where PV corresponds to the predictive values of activity and b_0, b_m, \dots, b_p correspond to regression coefficients obtained by least squares optimization.

- For each optimal linear dependence, from 4095 built, threshold value of PV, p_i has been chosen as a value corresponded to maximal accuracy ($i=1, 4095$). Among optimal linear dependences as a predictive model one with maximal accuracy has been chosen. p_i that corresponded to the optimal linear dependence with maximal accuracy (p_a) was used to perform prediction on the test set. (Definition of accuracy and other prediction measures can be seen on the next slide).
- The model has been additionally optimized on hydrophobicity scales



Results and discussion

Definition of prediction measures

The following equations were used to evaluate the quality of the prediction:

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

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

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

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

$$NPV = TN / (TN + FN)$$

where *SN* is sensitivity, *SP* is specificity, *AC* is accuracy, *PPV* is prediction precision or positive predictive value, *NPV* is negative predictive value, *TP* is true positive, *TN* is true negative, *FP* is false positive, and *FN* is false negative.

For the selected threshold p , the sequence is predicted as positive if $PV \geq p$ and negative if $PV < p$.



Results and discussion

Description of predictive model

The optimization reveals the training set with maximum value of accuracy (optimal set). Optimal set corresponds to the combination of the following characteristics:

- **Hydrophobicity**
- **Isoelectric Point,**
- **Penetration Depth,**
- **Propensity to Disordering,**
- **Linear Moment,**
- **Angle Subtended by the Hydrophobic Residue,**
- **Amphiphilicity Index,**
- **Propensity to Coil Conformation**

Optimal values for other parameters are:

- Threshold p_a for predictive model = **0.05**
- Hydrophobic scale = **Hessa and White [4]**



Results and discussion

Regression coefficients and prediction measures for optimal training set

Table 1. Regression coefficients

	b_0	b_M	b_H	b_C	b_I	b_D	b_O	b_R	b_L	b_A	b_s	b_A	b_P
St. aureus ATCC 25923	-1.27	0	-1.02	0	0.07	-0.01	0	-0.45	-0.91	0	0.004	0.45	0.90

Table 2. Prediction measures

	SN	SP	AC	PPV
Training set	0.87	0.74	0.80	0.76
Test set	0.84	0.84	0.84	0.84

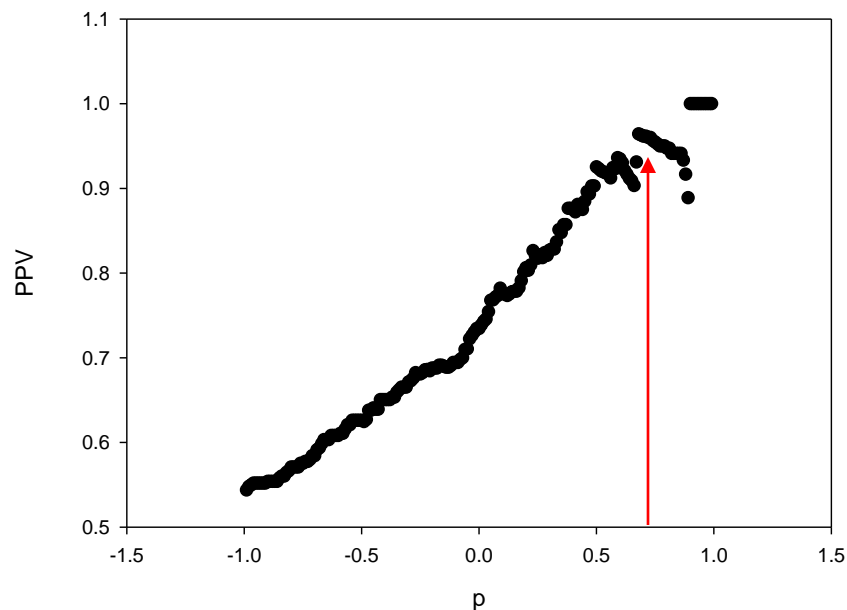
Based on the developed model of prediction the method of de novo design of AMP has been created



Results and discussion

Description of the model of design

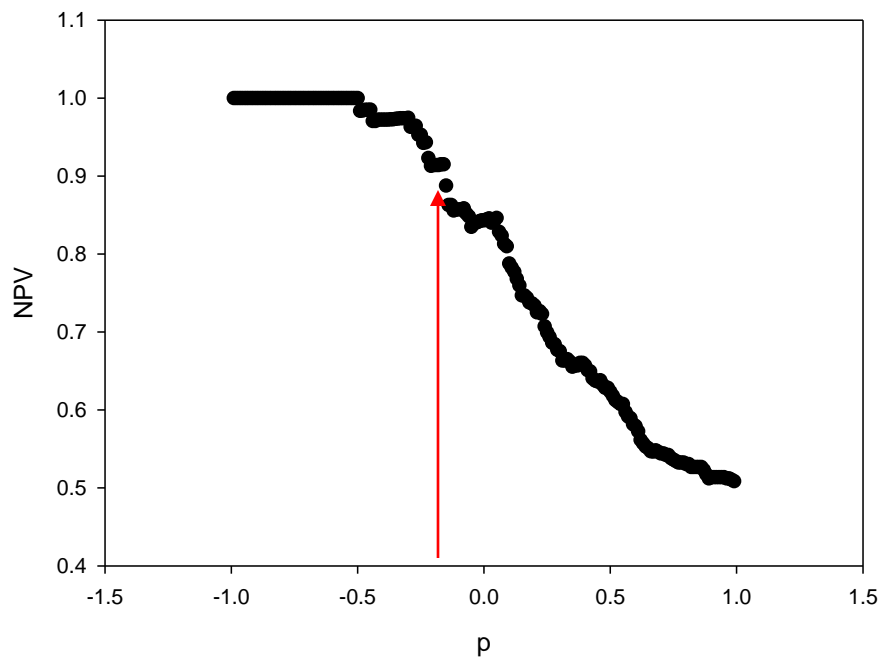
- PPV is the most valuable parameter for choosing the model of the design with high performance.
- To look for the area with high value of PPV, the dependence of PPV vs p (p varied from -1 to +1) has been plotted using data of training set, and the optimal value of $p_p = 0.52$ is chosen based on the requirement $PPV > 0.9$.



Results and discussion

Description of design model

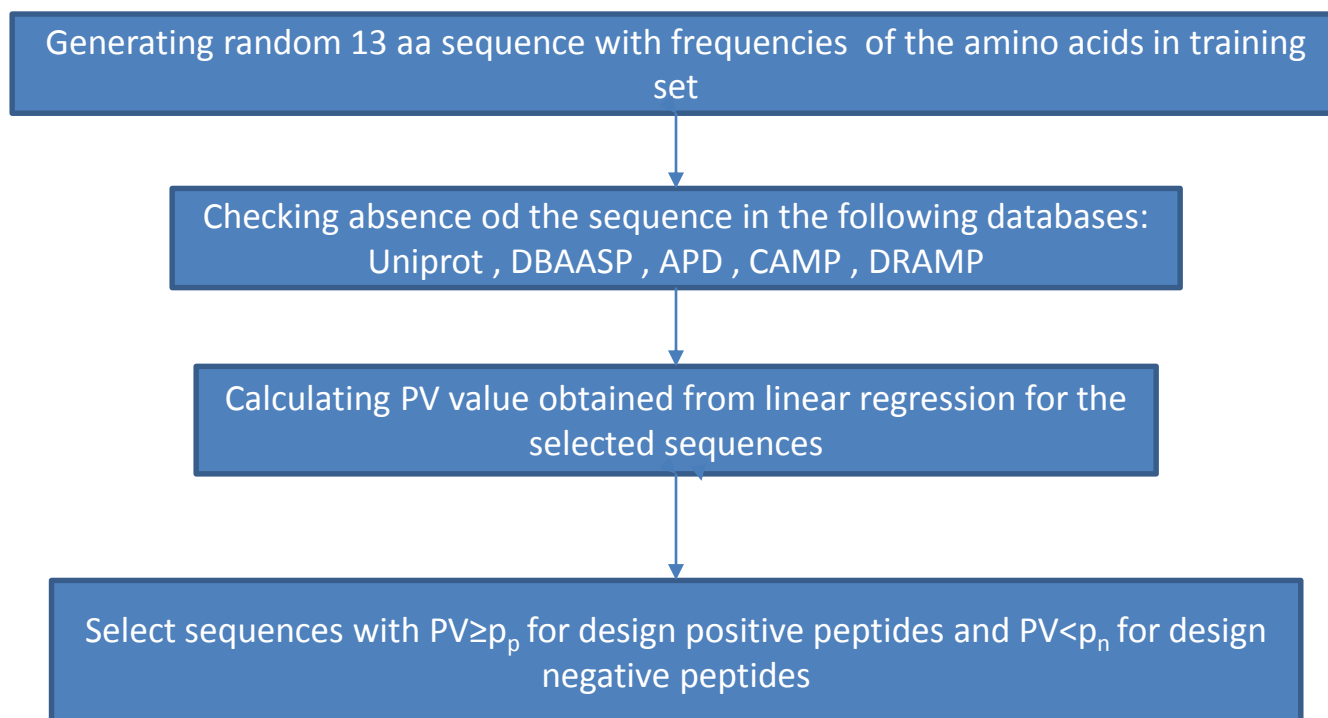
- For additional assessment of the model, we tried to design peptides non-active against *Staphylococcus aureus* 25923 also. For this purpose, the dependence of NPV vs p (p varied from -1 to +1) has been plotted. The optimal value of $p_n = -0.20$ is chosen, based on the requirement $NPV > 0.9$



Results and discussion

Design algorithm

After selecting optimal values of p_p and p_n , the following algorithm was used to design peptides



Results and discussion

Results of in vitro testing of the designed peptides

- 13 peptides predicted as active against *Staphylococcus aureus* 25923 were designed
- These peptides were synthesized and tested for antimicrobial activity in vitro
- 12 from these 13 have high antimicrobial activity, so prediction precision of the model equals to 92%



Results and discussion

Results of in vitro testing of the designed peptides

- 5 peptides predicted as non-active against *Staphylococcus aureus* 25923 were designed
- These peptides were synthesized and tested for antimicrobial activity in vitro
- All 5 peptides are non-active against *Staphylococcus aureus* ATCC 25923



Results and discussion

Future assessments and improvement

- The model of designing requires additional in vitro tests on the de novo designed peptides
- The model may be improved by adding new physicochemical characteristics and using other machine learning approaches



Conclusions

- A new model of *in silico* design of AMPs active against *Staphylococcus aureus* ATCC 25923 was developed
- The model of designing is based on the prediction using a linear model of the regression.
- *In vitro* test of the model of designing on the 13 *de novo* designed peptides has shown that the prediction precision equals 92 %.



References

1. Pirtskhalava M, Gabrielian A, Cruz P, Griggs HL, Squires RB, Hurt DE, Grigolava M, Chubinidze M, Gogoladze G, Vishnepolsky B, Alekseev V, Rosenthal A, and Tartakovsky M. DBAASP v.2: an Enhanced Database of Structure and Antimicrobial/Cytotoxic Activity of Natural and Synthetic Peptides. *Nucl. Acids Res.*, 2016, 44
2. Vishnepolsky B, Gabrielian A, Rosenthal A, Darrell EH, Tartakovsky M, Managadze G, Grigolava M, Makhatadze GI, and Pirtskhalava M. Predictive Model of Linear Antimicrobial Peptides Active against Gram-Negative Bacteria. *J. Chem. Inf. Model.* 2018, 58, 1141-1151.
3. Vishnepolsky B., Zaalishvili G., Karapetian M., Nasrashvili T., Kuljanishvili N., Gabrielian A., Rosenthal A., Hurt D.E., Tartakovsky M., Grigolava M. and Pirtskhalava M. De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria. *Pharmaceuticals* **2019**, 12(2), 82
4. Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 2007, 450, U1026–U1032.



Acknowledgments

This work was supported by International Science and Technology Center provided through National Institute of Allergy and Infectious Diseases / National Institutes of Health (G-2102)



5th International Electronic Conference
on Medicinal Chemistry
1-30 November 2019

sponsors:



pharmaceuticals