



2nd International Electronic Conference on Medicinal Chemistry

1-30 November 2016

chaired by Dr. Jean Jacques Vanden Eynde

sponsored by



pharmaceuticals

Design and in-vitro testing of new antimicrobial peptides based on QSAR modelling

**Boris Vishnepolsky^{1,*}, Maya Grigolava¹, George Zaalishvili², Margarita Karapetian²
and Malak Pirtskhalava^{1,*}**

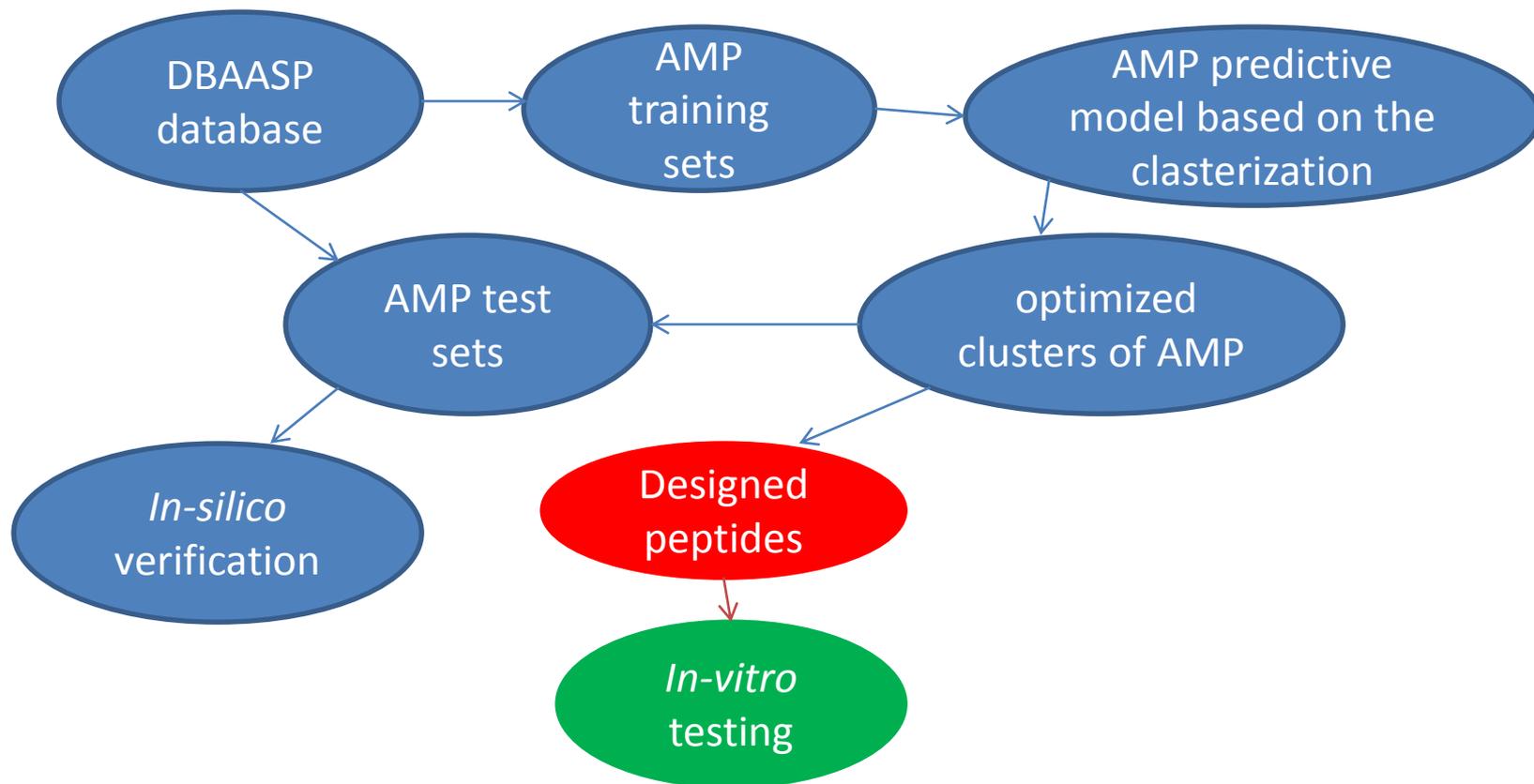
¹ I.Beritashvili Center of Experimental Biomedicine, Gotua str. 14, Tbilisi 0160 , Georgia

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

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

Design and in-vitro testing of new antimicrobial peptides based on QSAR modelling

Graphical Abstract



Abstract: Antimicrobial peptides (AMPs) are anti-infectives that may represent a novel and untapped class of biotherapeutics. In the lab of bioinformatics of IBCEB the Database of Antimicrobial Activity and Structure of Peptides (<http://dbaasp.org>) has been developed.

Contrary to available approaches, we think that strategy of AMP prediction should be based on the fact that there are at least four kinds of AMPs for which four independent algorithms of prediction have to be developed in order to get high efficacy. We can distinguish linear cationic antimicrobial peptides (LCAP), cationic peptides stabilizing structure by intra-chain covalent bonds, proline and arginine-rich peptides, and anionic antimicrobial peptides.

Simple predictive model which can discriminate AMP from non-AMP has been developed for LCAP. As descriptors the sequence-based physical-chemical characteristics responsible for capability of the peptide to interact with an anionic membrane were considered. The algorithm was based on the clusterization of AMPs by their physicochemical properties. The results show that descriptors relied mainly on hydrophobic and hydrophilic features allow us to predict AMP with the high accuracy.

The developed predictive model was used to design new peptides. Antimicrobial potency of these peptides have been evaluated by *in vitro* testing of peptides' activity against different bacteria (including drug resistant strains). *In-vitro* estimation shows high accuracy of the developed predictive model.

Keywords: Antimicrobial peptides; AMP prediction; Design of AMPs



Introduction

- Interest for AMP has increased and the rate of discovering new peptides (natural and synthetic) is very high
- The majority of the new peptides are artificial and have been created in the studies of structure-activity relationships
- Task-oriented rational *ab initio* design, which means sequence-based computational prediction of antimicrobial properties, with the subsequent experimental synthesis and testing of peptides for antimicrobial activity, is the most cost-effective way of developing novel antibiotics against drug-resistant bacteria



Introduction (continue)

- Most of the algorithms for AMP prediction do not take into account variation in mechanisms of action, structure, mode of interaction with membrane and other differences
- There are at least four kinds of AMPs: linear cationic antimicrobial peptides (LCAP), cationic peptides stabilizing structure by intra-chain covalent bonds (CICP), proline and arginine-rich peptides (PAP), and anionic antimicrobial peptides (AP)
- Consequently, our approach suggests that the development of four independent algorithms of prediction is necessary to get high efficacy



Introduction (continue)

- Most predictive methods do not distinguish target species during the model development
- Such approach does not consider several issues, which can influence the accuracy of prediction. These issues are the following:
- Antimicrobial potency of AMPs strongly depends on bacterial membrane types and thus on particular targets
- There are limited data for negative set, because there are practically no data for peptides, which have not antimicrobial potency for all target species. So they use randomly selected set
- Our approach is the use of antimicrobial potency for specific targets for which there is a large number of data.



Introduction (continue)

- There have been described several different mechanisms of action of AMP. Consequently, it's reasonable to assume the existence of different types of peptides with different physical-chemical features.
- We had to take into account all these issues and decided to develop predictive model relying on machine learning approach using clustering algorithm at the optimization of model.

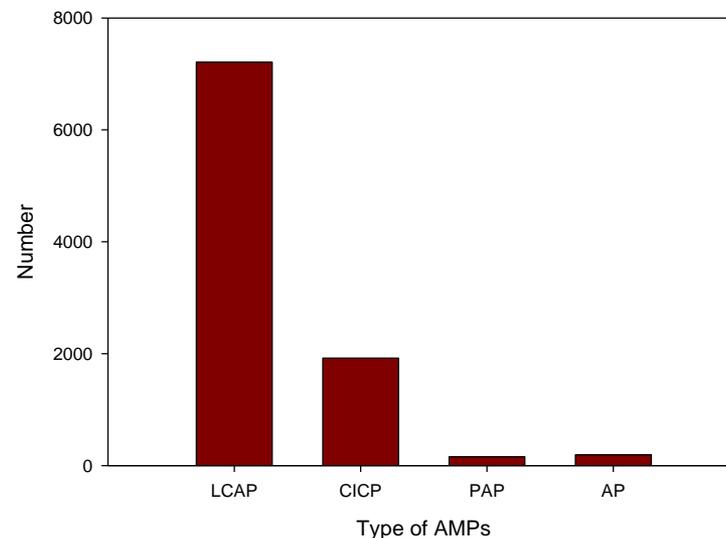


Introduction (continue)

Data from DBAASP database was used for developing predictive model

- Database of Antimicrobial Activity and Structure of Peptides (**DBAASP**) is the manually-curated database.
- DBAASP has been developed to provide the information and analytical resources to the scientific community in order to develop antimicrobial compounds.
- DBAASP is accessible at: [http:// dbaasp.org](http://dbaasp.org) [1]

Distribution of four types of AMPs in DBAASP



Linear peptides are the larger class of AMPs for which there are the most data in DBAASP and we have developed predictive algorithm for this AMP class



Introduction (continue)

Number of peptides against frequently occurring target species from DBAASP

Target organism	Number of Peptides
<i>Klebsiella pneumoniae</i>	266
<i>Candida albicans</i> ATCC 90028	469
<i>Bacillus subtilis</i>	544
<i>Pseudomonas aeruginosa</i> ATCC 27853	1173
<i>Staphylococcus aureus</i> ATCC 25923	1455
<i>Escherichia coli</i> ATCC 25922	2141



Introduction (continue)

- Stabilization of peptide structure in the membrane environment is mainly driven by intramolecular hydrogen bonding
- So it's reasonable to assume that the membrane will impel the linear peptide to a regular α -helical conformation. This assumption is supported by the fact that all transmembrane domains of membrane proteins are mainly helical.
- Therefore the properties of peptides related to the three-dimensional structure, such as a helical hydrophobic moment, an orientation of peptides relative to the surface of membrane, the penetration depth, etc can also be evaluated using only sequence information



Introduction (continue)

- The main objectives of the work were:
 - **Development of a new predictive model for linear AMP being active against certain species relying on QSAR (Quantitative Structure-Activity Relationship) study**
 - **Design new antimicrobial peptides based on the proposed predictive model**
 - ***In-vitro* test of the de novo designed peptides against bacterial species**



Results and discussion

Physico-chemical properties selection for predictive model

- Both AMP and lipid bilayer are amphipathic
 - So the main parameters which should construct the physical-chemical characteristics of AMPs, responsible for the interaction with the membrane have to be ionic charges and hydrophobic features of the side chains of amino acid residues and also a depth dependent potential of insertion of amino acids into membrane
 - 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



Physico-chemical properties selection for predictive model (continue)

The following 8 properties were used in QSAR study:

- **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)**

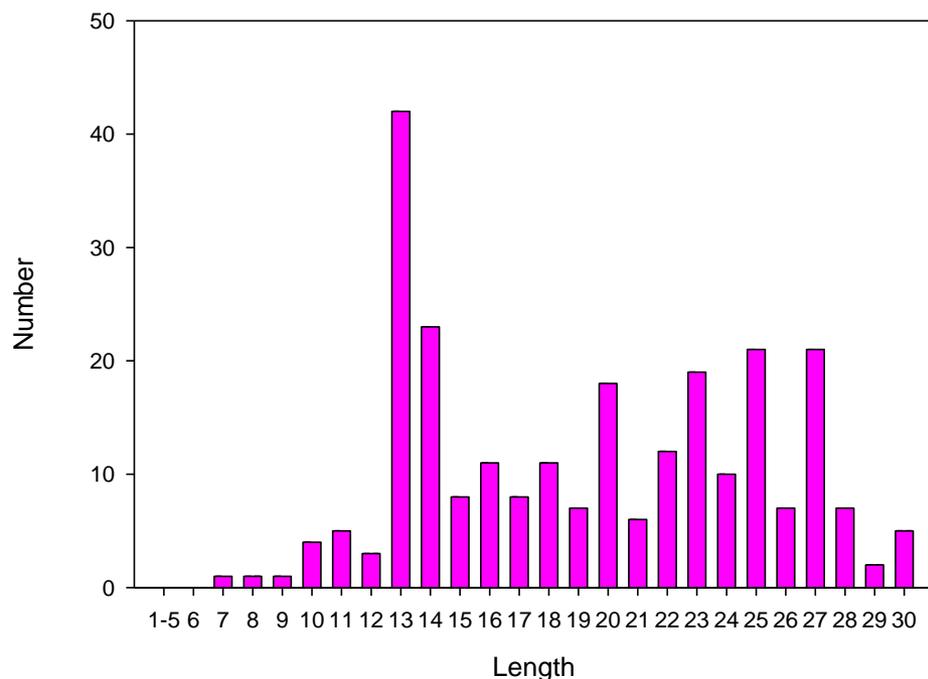
Definition of the characteristics can be found here [2]



Conditions of creation of the Training and test sets

- *Escherichia coli* ATCC 25922 has the most data in DBAASP
- Our approach is based on the data for this species

Distribution of peptides length of ribosomal AMP active against *Escherichia coli* ATCC 25922



Majority of peptides have length within intervals 10-16



Conditions of creation training and test sets

- Positive set is formed on the basis of condition MIC <25 $\mu\text{g/ml}$
- Negative set is formed on the basis of condition >100 $\mu\text{g/ml}$
- Sets were performed with the following restrictions:
 - Sequence length 10-16 amino acids
 - Without intra-chain bonds
 - Without unusual amino acids



Algorithm description (continue)

Machine-learning approach relied on density-based clustering algorithm DBSCAN [3] was used

Training and test sets were performed from DBAASP.

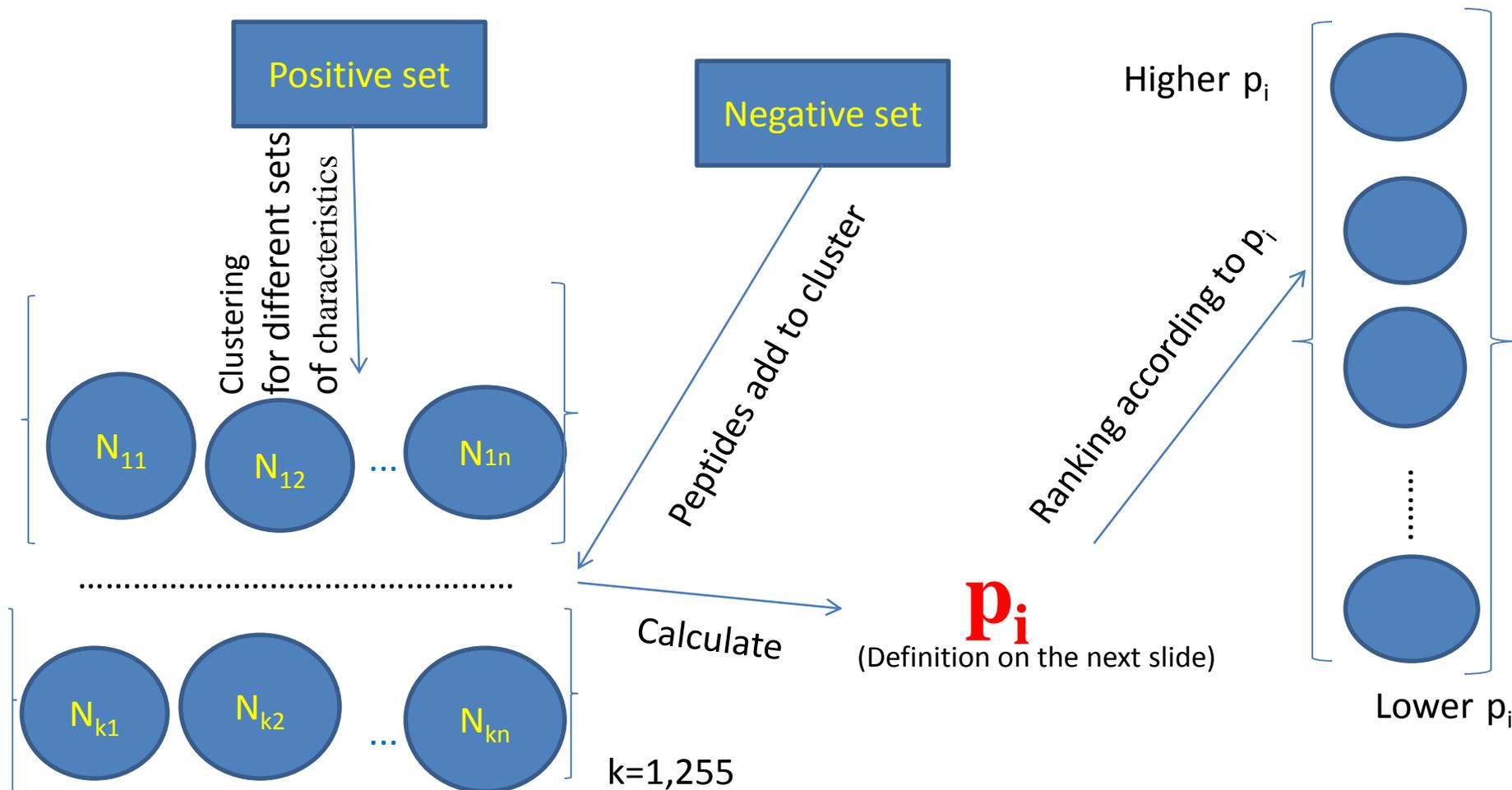
Algorithm is based on the following full sets of peptides

- Full Positive Set (FPS) - 160 peptides
- Full Negative Set (FNS) -146 peptides
- For validation purposes 5 training sets (positive and negative) were created from full sets. Each positive and negative set consists of 125 randomly selected peptides from FPS and FNS respectively.



Algorithm description (continue)

Clustering of positive sets has been performed in the 255 different space of characteristics. (Number of combinations of 8 characteristics = 255).

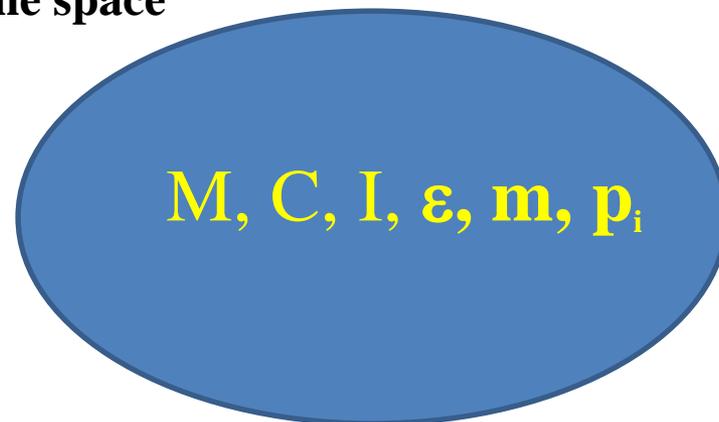


Algorithm description (continue)

Parameters of optimization : **255 sets of characteristics;**
Parameter of clusterization $\varepsilon, m;$
 P_i

Optimal clusters are defined by the following parameters:

- **Number of characteristics** (dimensionality of the space where cluster has been determined.
For the example in the figure = 3)
- **Types of characteristics** (Type of space.
For the example in the figure M, C
and I characteristics determine the space)
- **DBSCAN clusterization parameters $\varepsilon, \text{minpts} (m)$**
- **$p_i = (N_i^+ / N^+) (N_i^+ / N_i^f)$** , where N_i^+ – number of peptides from positive set in the i th cluster, N^+ – full number peptides in the positive set, N_i^f – number of peptides from positive and negative sets in the i th cluster

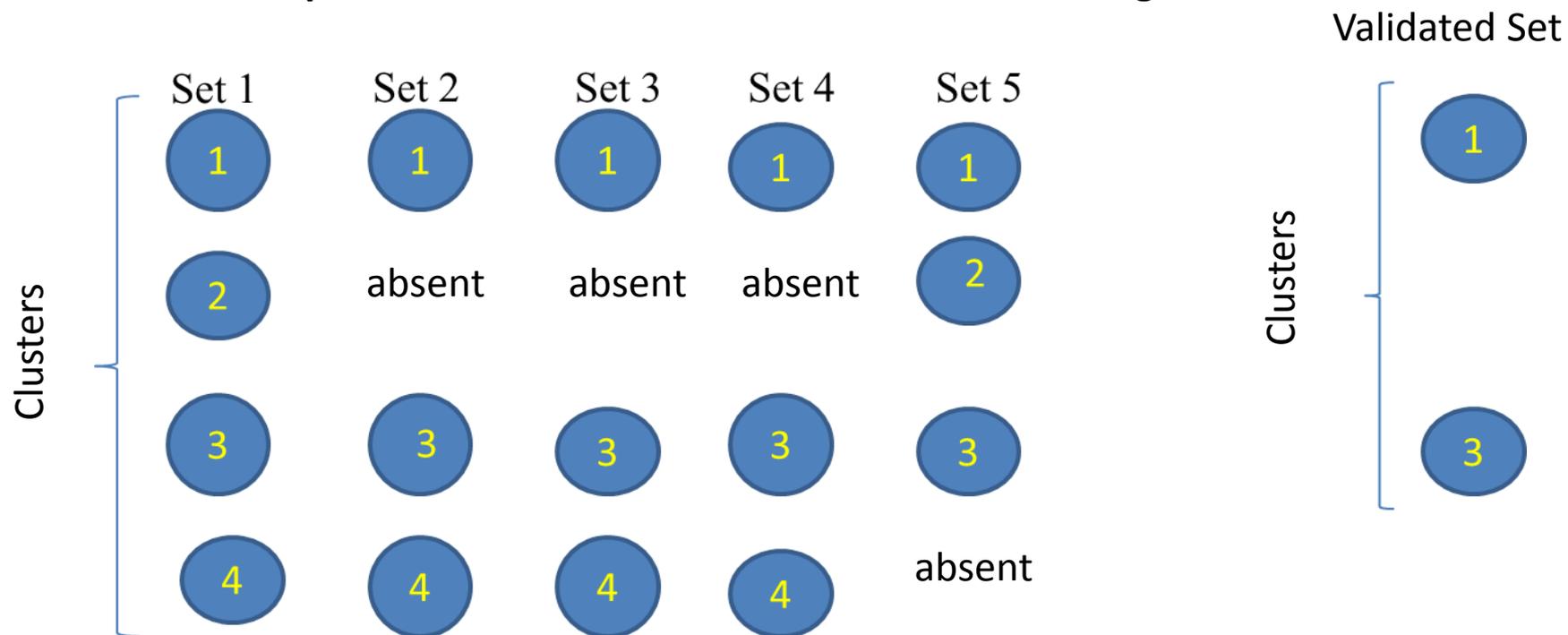


Algorithm description (continue)

Validation of the results of optimization.

Optimization by clustering took place for each of 5 pairs of randomly selected training sets. Clusters which were continuously appeared during optimization of different training pairs are chosen for the further processing

Schematic representation of validation on 5 random training sets



Algorithm description (continue)

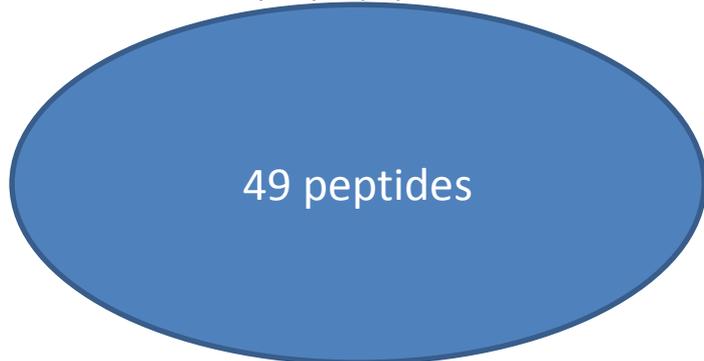
- In each training set non-overlapping cluster combinations (without common peptides) were generated
- The best non-overlapping cluster combination with maximum p_i was selected for AMP prediction



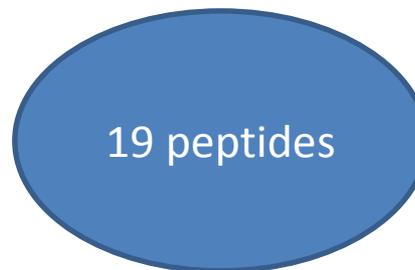
Schematic representation of the results of optimization on the training set of peptides

Optimization reveals 4 non-overlapping clusters of 4 different N-mer spaces.

Cluster I has been formed in the space of characteristics M, H, C, I, D



Cluster II has been formed in the space of characteristic C



Cluster III has been formed in the space of characteristics M, H, C, I, R



Cluster IV has been formed in the space of characteristics M, H, I, D



Testing of the predictive model

For *in silico* test the following 3 sets have been used:

- Set 1 – *Escherichia coli* ATCC 25922 – 21 peptides in positive set and 21 peptides in negative set
- Set 2 – *Pseudomonas aeruginosa* ATCC 27853 – 109 peptides in positive set and 109 peptides in negative set
- Set 3 – *Klebsiella pneumoniae* – 29 peptides in positive set and 29 peptides in negative set

Pseudomonas aeruginosa ATCC 27853 and *Klebsiella pneumoniae* were used for creation test sets, because according to DBAASP data, AMPs behave similarly against the *Escherichia coli*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*



Assessment of likeness of behavior of a particular set of peptides against different microbial cells on the basis of DBAASP data

- For the assessment, pairs of microbes were screened and peptides were selected so that susceptibilities for both of the microbes are known for each. For a selected set of peptides the ranking of susceptibilities (MIC) for each microbe was performed and Spearman's rank correlation coefficient was calculated. Spearman's rank correlation coefficients for particular pairs of target organisms are presented in Table

Target organisms	<i>E-coli</i> ATCC 25922	<i>Klebsiella_pneumoniae</i>	<i>Pseudomonas aeruginosa</i> ATCC 27853	<i>Candida albicans</i> ATCC 90028	<i>Bacillus subtilis</i>	<i>Staphylococcus aureus</i> ATCC 25923
<i>E-coli</i> ATCC 25922		78.20	72.64	1.28	57.4	53.8
<i>Klebsiella_pneumoniae</i>			69.70	-12.5	ND*	12.8
<i>Pseudomonas aeruginosa</i> ATCC 27853				15.8	47.9	47.0
<i>Candida albicans</i> ATCC 90028					ND*	35.2
<i>Bacillus subtilis.</i>						33.2
<i>Staphylococcus aureus</i> ATCC 25923						

*ND - not determined because there is not enough data



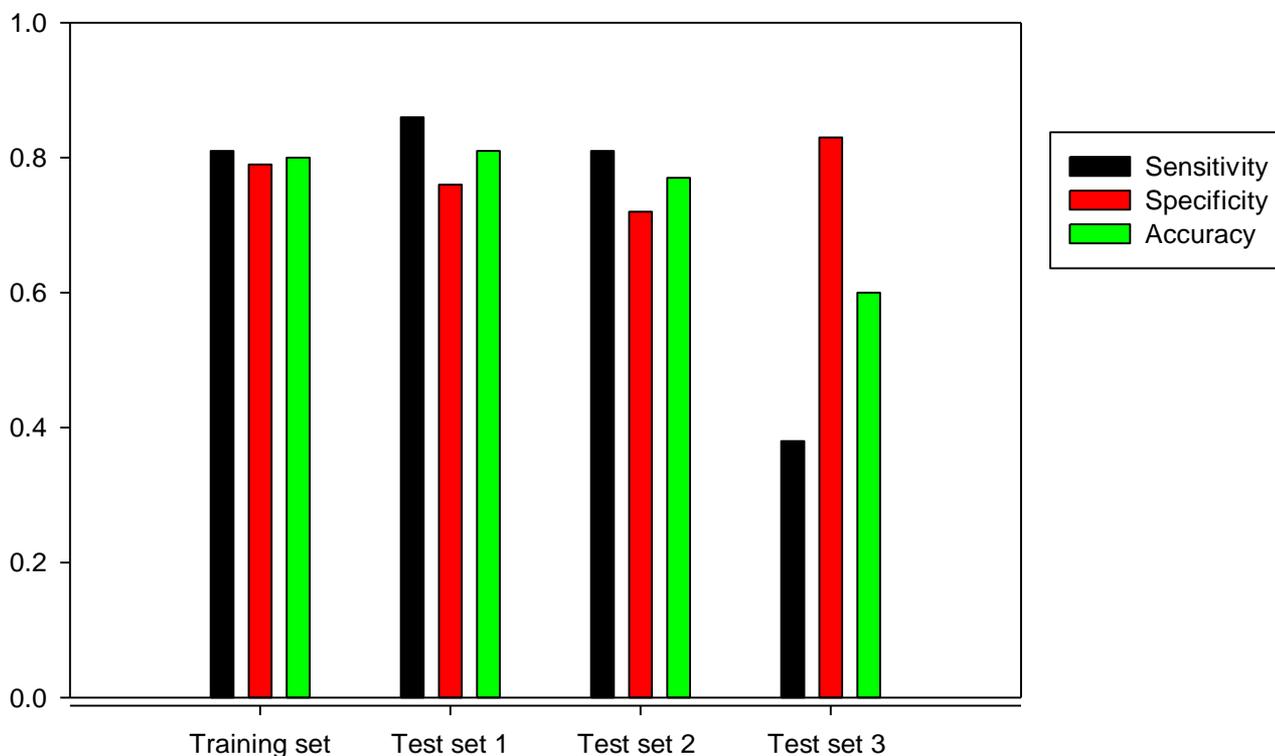
Sensitivity, Specificity and Accuracy for training and test sets

Training set - *Escherichia coli* ATCC 25922 -125 peptides in positive set and 125 peptides in negative set

Test set 1 - *Escherichia coli* ATCC 25922 -21 peptides in positive set and 21 peptides in negative set

Test set 2 *Pseudomonas aeruginosa* ATCC 27853 -109 peptides in positive set and 109 peptides in negative set

Test set 3 - *Klebsiella pneumoniae* -29 peptides in positive set and 29 peptides in negative set



Peptide design

- The most objective data from training and test sets correspond to cluster I (the highest number of peptides in it).
- So we have decided to design peptides with physical-chemical characteristics corresponding to cluster I
- Regression analysis was used for sequences from cluster 1
- Regression analysis was carried out for dependence susceptibilities (MIC in μM) from the presented characteristics based on peptides from training set within cluster I
- The obtained dependence is the following:

$$\text{MIC} = -60.23 + 7.61M + 0.36C + 1.31D + 0.19O + 1.4H + 42.2L + 4.19R + 0.59I$$



Peptide design (continue)

- Peptides were designed by the following principles:
 - Length of peptides are 13, based on the data of most frequent peptide lengths in DBAASP
 - Sequences were selected from randomly generated amino acid sequences whose physical-chemical characteristics satisfy cluster I and equation obtained by regression analysis (see previous slide)
 - 11 sequences were selected



Peptide synthesis and *in-vitro* testing

- 11 designed peptides were synthesized
- *In-vitro* tests of these peptides against Escherichia coli ATCC 25922 were carried out
- 10 from 11 peptides have MIC < 25 µg/ml
- Preliminary data show that these peptides are active against some pathogenic gram-negative strains of *P. aeruginosa*, *A. baumannii*, *E. cloacae* (including FQR, NDM, Meropenem resistant)



Conclusions

- Predictive model for AMP based on the clusterization of their physicochemical properties was developed
- Four non-overlapping clusters with different sets of physicochemical properties were obtained for the prediction of AMP against - *Escherichia coli* ATCC 25922.
- *In-silico* test of predictive model shows accuracy about 0.8 for *Escherichia coli* ATCC 25922 and *Pseudomonas aeruginosa* ATCC 27853. Accuracy in case of *Klebsiella pneumoniae* is lower (~0.6).
- 11 peptides were designed based on the data of physicochemical properties from cluster I.
- *In-vitro* tests of the 11 designed peptides were performed against *Escherichia coli* ATCC 25922
- The obtained results show that 10 from 11 peptides have high antimicrobial potency.



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) and Shota Rustaveli National Science Foundation (FR/397/7-180/14)



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 (D1), D1104-D1112.
2. Vishnepolsky B and 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.
3. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. 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.

