Models for the Prediction of Antimicrobial Peptides Activity

Rosaura Parisi[1], Ida Moccia[1], Lucia Sessa[1], Luigi Di Biasi[1,2], Simona Concilio[3], and Stefano Piotto[1(✉)]

[1] Department of Pharmacy, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
piotto@unisa.it
[2] Department of Informatics, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
[3] Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

Abstract. Antimicrobial peptides AMP are small proteins produced by the innate immune system in multicellular microorganisms. The mechanism of action of AMP on target membranes can be divided in two main categories: pore forming and non-pore forming mechanisms. We applied a computational approach to design novel linear peptides having high specificity and low toxicity against common pathogens. We built up QSAR models using the data present in a database of antimicrobial peptides. Here, we present new models of activities obtained by the use of evolutionary methods and the relative statistical validation.

1 Introduction

The drug resistance is a limit to the choice of an efficient antibiotic therapy. The reason is that any microorganisms, through different strategies, can cancel out the action of antibiotics. Unfortunately, the indiscriminate use of antibiotics accelerated this phenomenon. A classic example of antibiotic resistance is represented by the strain methicillin resistant *Staphylococcus aureus* (MRSA) [1]. Consequently, there is the need for new drugs active against pathogens. One of the most promising strategy against various pathogenic microbes is represented by antimicrobial peptides (AMP). They are small proteins produced by multicellular organisms that inhibit or kill some microorganisms (bacteria, fungi, enveloped viruses, protozoans and parasites). AMP are produced in the innate immune response [2]. These peptides, often small and cationic, are secreted into the aqueous phase where they are generally in an unfolded state, but they fold in the proximity of the target membrane [3]. Most antimicrobial peptides act on the bacterial cell membrane without specific receptors. How AMP kill bacteria interacting with the cell membrane is not yet completely understood. In fact, AMP utilize a wide variety of mechanisms, such as altering the membrane equilibrium, creating pores, disrupting the membrane, altering the membrane fluidity or docking a protein receptor [4, 5]. Consequently, their membrane interaction and broad activity spectra are becoming an ideal target to overcome the resistance resulting from bacterial

F. Rossi et al. (Eds.): WIVACE 2015, CCIS 587, pp. 83–91, 2016.
DOI: 10.1007/978-3-319-32695-5_8

mutations [6]. They are classified, according to their secondary structure, into four categories [7]: α-helical, β-sheet peptides, linear extended antibacterial peptides and the loop antibacterial peptides. To date, more than two thousands natural AMP have been isolated and characterized from different sources and several thousands of synthetic variants have been developed. For example, the most studied family of peptides extracted from mammalians is the family of β-defensins. Some researchers developed an approach to identify conserved motifs in these peptides through a computational tool based on hidden Markov models (HMMs) and a basic local alignment search tool [8]. Sequence analysis of these peptides showed low sequence homology [9] precluding the possibility to create easily a model of activity [10]. For this reason, it became important to try different computational approaches for predicting the activity of antibacterial peptides. Several computational studies permitted to develop algorithms to predict antibacterial peptides with a high accuracy. For example, some researchers using Artificial Neural Network (ANN) and Support Vector Machine (SVM) suggested that *N*- and *C*-terminals of the AMP sequence might play an important role in the activity: *C*-terminal is involved in the interaction with the membrane and in the pore formation, while the *N*-terminal helps in bacteria specific interaction process [10]. The starting point of this work was the selection of sets of homogenous AMP in terms of chemical-physical properties. This step was essential to cluster peptides acting with similar mechanisms. On these sets, we performed a QSAR analysis to determine the relationship between the structural properties of AMP, such as charge, Boman index, or flexibility, with the antimicrobial activity of these molecules (MIC, minimum inhibitory concentration). These sets were analyzed by artificial neural networks and genetic algorithms. In quantitative structure - activity relationships (QSAR) we correlate the biological activity of a class of compounds with the chemical - physical characteristics or structural properties of the compounds themselves. The main limitation of the QSAR studies is the complexity of a biological system. Genetic Algorithms (GA) are heuristic search methods based on the Darwinian theory of natural selection [11]. The artificial neural network (ANN) have been developed and designed to mimic the information processing and learning in the brain of living organisms. The ANN offer satisfactory accuracy in most cases but tend to over fit the training data. Here we present activity models on a gram positive bacterium: *Staphylococcus aureus*.

2 Materials and Methods

The working hypothesis is that peptides with similar features can share the same mechanism of action. We have chosen the parameters present in the database Yadamp [12] to create uniform subsets. We have selected 6 parameters (charge at pH 7, length, CPP index, flexibility, ΔG, helicity as listed in the server Yadamp [12]), and we generated 62 different peptide sets homogeneous in one or two parameters (for example, one set was constituted by the 173 peptides shorter than 30 residues and with a charge at pH 7 between 2 and 7).

On the 62 peptide sets, we applied two kind of mathematical methods.

Genetic algorithms are stochastic optimization techniques that mimic selection in nature that proved to be a very effective tool in QSAR studies. A genetic algorithm

chooses a suitable set of descriptors, and the selected descriptors are utilized to build a nonlinear QSAR regression equation. Nonlinear correlations in the data are explicitly dealt with by use of the descriptors in spline, quadratic, offset quadratic, and quadratic spline functions. The method has been implemented in the Material Studio 7.0 [13] package, and it was used here without modification. The smoothness parameter was kept at the default value of 1.0, and the length of an individual was let vary between 2 and 5 descriptors. A total of 500 individuals were let evolve over 5000 new generations.

ANN analysis was performed with the software Matlab 2013 [14]. The multilayers network used have two layers: the output and the hidden layer. The hidden layer consisting of ten artificial neurons, the output layer of a single neuron. The training function of the network is the algorithm based on the Levenberg-Marquardt minimization method (trainlm). This function is very fast and performs better on function fitting (nonlinear regression) problems. The adaption learning function is learngdm, that corresponds to the momentum variant of back propagation. The two different transfer functions used for the neurons are: tan–sigmoid transfer function (tansig) for the hidden layer, that returns values between −1 and 1, and linear transfer function (pureline) for the output layer. The performance function for the network is mean square error (mse).

3 Results

3.1 QSAR Analysis - GA

On each peptide set, we applied the same GA protocol. We identified two equations describing biocidal activity. The R^2 was of 0.92 and 0.81 respectively. Equation 1 was obtained from a dataset of peptides having a length between 7 and 11 amino acids (55 peptides). Equation 2 was obtained using peptides shorter than 30 amino acids and a Boman index between 1 and 2 kcal/mol for a total of 92 peptides. In Eq. 1 the critical parameters for antimicrobial activity are the peptide charge in acid and neutral solution and the number of polar amino acids in the sequence. Equation 2 is similar to Eq. 1 and gives similar importance to peptide charge.

$$MIC = 8.16\,POLAR\,AA - 2571(-0.72 - Ch5)^2 + 9963(-0.90 - Ch7)^2 + 11 \quad (1)$$

$$MIC = -\frac{(MW - 881)^2}{250000} + 122(D - 1.7)^2 + 3134(1.07 - Ch5)^2 - 3340(0.79 - Ch7)^2 + 22 \quad (2)$$

The parameter function returns the value of the argument, if it is positive, and zero otherwise.

D: Number of residues of Aspartic acid
Ch5: peptide charge at pH5
Ch7: peptide charge at pH7
POLAR AA: number of polar residues
MW: Molecular weight

Both equations confirm that AMP belonging to that set, act through electrostatic interactions with bacterial membrane [15]. However, a good R^2 cannot capture the quality of an activity model because the intrinsic experimental error in microbiological tests, due to serial dilutions, is not considered. It is more correct to talk about activity classes, and the goodness of a QSAR model must be judged in terms of its ability to discriminate among very active, active and non-active peptides. For this reason, MIC (minimum inhibitory concentration expressed in μM) values of 0.3 and 1.8 must be considered as peptides with the same activity. To evaluate the models, we divided the peptides in classes of MIC as shown in Table 1. The 5 classes have similar dimension.

Peptides of classes A, B, C, D are considered active, whereas class E corresponds to inactive peptides.

Table 1. Division of antimicrobial peptides into five classes based on the values of MIC in μmol/mL.

A	B	C	D	E
$0 \leq MIC \leq 2$	$2 < MIC \leq 5$	$5 < MIC \leq 10$	$10 < MIC \leq 30$	$MIC > 30$

The MICs have been calculated for all peptides active against *S. aureus* present in the database. We calculated the precision (PPV), the accuracy (ACC), the sensitivity (TPR)and the specificity (SPC) as defined in Eqs. 3–6.

$$PPV = \frac{TP}{TP + FP} \tag{3}$$

$$ACC = \frac{TP + TN}{total\ population} \tag{4}$$

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$SPC = \frac{TN}{TN + FP} \tag{6}$$

Whereas TP, FP, TN, and FN stand for True positives, False positives, True negatives and False negatives respectively.

The calculation of these indexes requires an arbitrary definition of what is considered *active* and *inactive*. We followed a common view in the pharma industry to consider *inactive* those peptides with a MIC higher than 30 μM. Therefore, active peptides are those belonging to classes A, B, C and D.

In Fig. 1 we plotted the precision, accuracy, sensitivity and specificity for models obtained by GA analysis. For both models, the behavior is acceptable only for three indexes. Specificity (black lines in figure) is the exception, with values that drop to 25 % for Eq. 2 for peptide longer than 40 amino acids. This is not surprising, since the model was obtained from a dataset of shorter peptides.

Low specificity indicates that models displays many false positives. However, a good R^2 and high precision, accuracy and sensitivity, cannot capture the quality of an

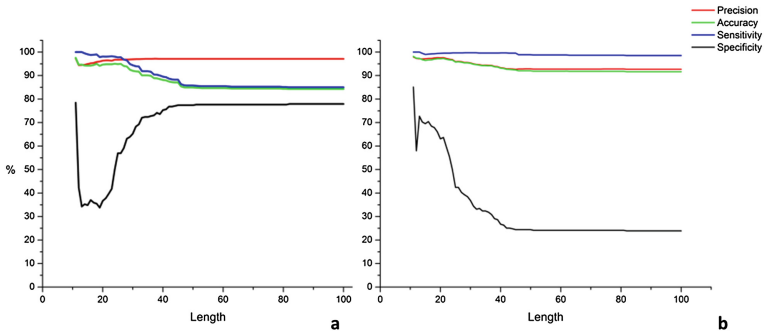

Fig. 1. Evaluation of precision, accuracy, sensitivity and specificity of Eqs. 1(a) and 2(b)

activity model because the intrinsic experimental error in microbiological tests, due to serial dilutions, is not considered. It is more correct to talk about activity classes, and the goodness of a QSAR model must be judged in terms of its ability to discriminate among very active, active and non-active peptides. The overall quality of the model (score) is calculated comparing MIC predictions with the experimental data according to Eq. (7). The scores are indicated in Table 2.

Table 2. Matrix for the computation of the overall model quality

Predicted \ Observed	A	B	C	D	E
A	2	1	0	-1	-2
B	1	2	1	0	-1
C	0	1	1	0	-1
D	-1	0	0	1	0
E	-2	-1	-1	0	2

$$Score = \sum_{i=1}^{n} Matrix[Class_{observed} - Class_{predicted}] \tag{7}$$

The scoring matrix in Table 2 attributes a reward each time the model correctly predicts the MIC. If the class is not predicted correctly, there is a penalty (negative values). The quality of the model is well represented in Fig. 2. Each point in the figure corresponds to a set of peptides of length between Length_start and Length_stop. The overall quality, calculated with Eq. (7), is rescaled between 0 (blue, unreliable) and 100 (red, reliable), and color mapped.

For example, the point 20, 50 of Fig. 3a indicates that the sum of the scores on all peptides with length between 20 and 50 is lower the 10 %. This diagram permits to easily evaluate the domain of applicability of the model.

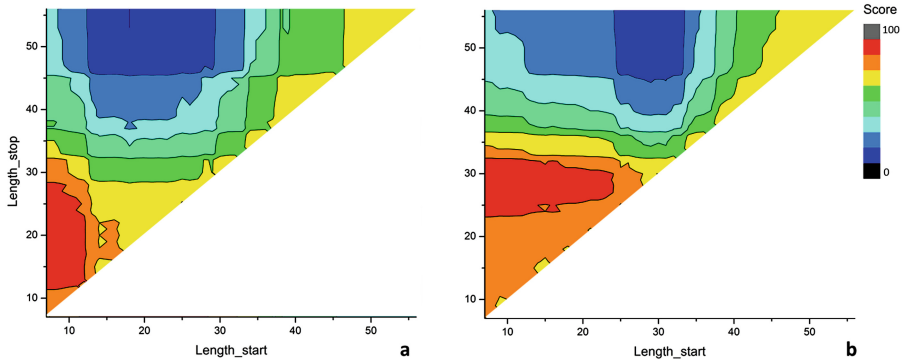

Fig. 2. Results of statistical validation of the Eqs. 1(a) and 2(b) obtained for *S. aureus* (Color figure online)

Figure 2a is relative to Eq. 1. As clearly shown in the diagram, the reliable region (red) is larger than the subset where the model was calculated. For longer peptides, the prediction capability of the model quickly degrade. The Eq. 2 (Fig. 2b) shows a wide reliable region, even larger than the original set of peptides.

3.2 QSAR Analysis – ANN

On the same data sets, we have applied ANN. The neural network used consisted of 2 layers with 10 neurons in the hidden layer. In the first dataset of 55 peptides, the neural

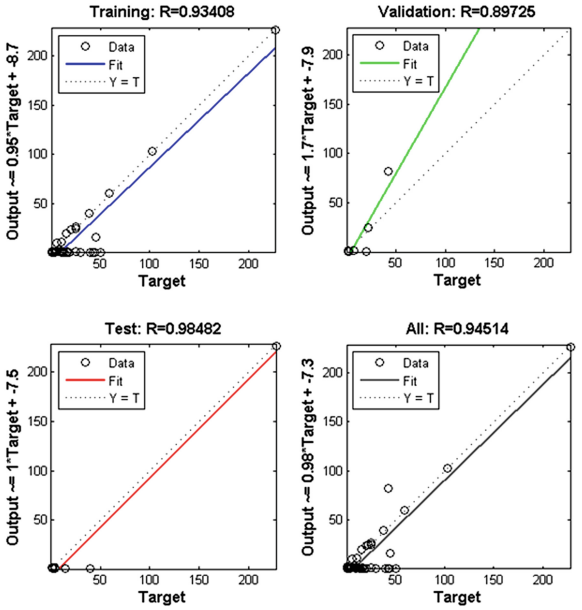

Fig. 3. Results of the application of ANN for peptides with a length between 7 and 11

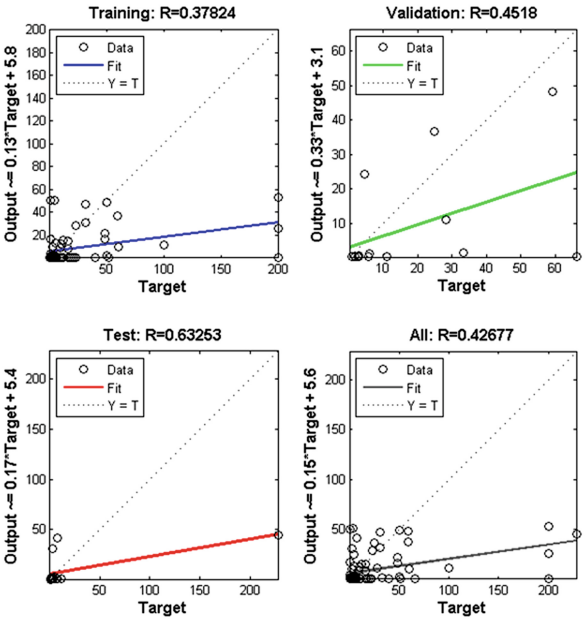

Fig. 4. Results of the application of ANN for peptides shorter than 30 amino acids and a Boman index between 1 and 2 kcal/mol

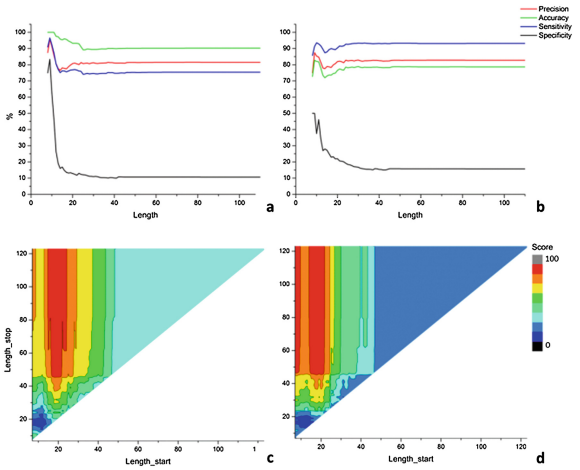

Fig. 5. Result of statistical validation of the two ANN analysis on peptides. The model (a, c) was created from peptides with a length between 7 and 11 amino acids; the model (b, d) was created from peptides shorter than 30 amino acids (b, d)

network found a good correlation between molecular descriptors and the antimicrobial activity.

The overall performance was a R^2 of 0.945, as shown in Fig. 3, whereas on the second data set, peptides shorter than 30 amino acids and a Boman index between 1 and 2 kcal/mol, the overall R^2 was of 0.427 (see Fig. 4).

The evaluation of the applicability of the neural network models were made in the same fashion of GA models. Unsurprisingly, the model is reliable only for the interval between 7 and 11 amino acids. In Fig. 5 we reported the trend of sensitivity, specificity, accuracy and precision for *active* and *inactive* peptides (Fig. 5a and b) for the two models. The more accurate evaluation using the quality matrix (Table 2) assigning peptides to 5 classes of activity is shown in Fig. 5c and d.

As shown in the diagrams, the ANN models are applicable in a range of peptides narrower than ranges obtained for GA models. Peptides longer than 40 cannot be calculated with both models.

4 Conclusion

We conducted a QSAR analysis on the activity of a large set of antimicrobial peptides. The creation of sets of peptides homogeneous in chemical-physical characteristics is indispensable for any statistical analysis. In this work, we performed GA and ANN studies on homogeneous sets of AMP extracted from the peptide database Yadamp. The GA analysis underlined the importance of peptide charge and polarity. This finding support one of most accepted models of activity, that the peptide-membrane interaction is mediated by electrostatic interactions. The artificial neural networks analysis is a complementary approach to GA. We observed a satisfactory fitting of antimicrobial activity only in one model. In that case, though with an $R^2 = 0.945$, the performance score of ANN models resulted lower than GA models, but it can be used for a peptide design based on consensus among different models. In conclusion, the models obtained by GA and ANN analysis, can be efficiently applied to peptides with length between 7 and 20. The number of sequences of peptides shorter than 20, is about 10^{26} that is an extraordinary large pool for novel antimicrobial mining.

The models presented here can be of high importance in designing novel antimicrobial peptides and all models will be offered as web service within the database Yadamp.

References

1. Liu, C., et al.: Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant Staphylococcus aureus infections in adults and children. Clin. Infect. Dis. **52**(3), e18–55 (2011) (ciq146)
2. Cruz, J., et al.: Antimicrobial peptides: promising compounds against pathogenic microorganisms. Curr. Med. Chem. **21**(20), 2299–2321 (2014)
3. Cirac, A.D., et al.: The molecular basis for antimicrobial activity of pore-forming cyclic peptides. Biophys. J. **100**(10), 2422–2431 (2011)

4. Török, Z., et al.: Plasma membranes as heat stress sensors: from lipid-controlled molecular switches to therapeutic applications. Biochim. Biophys. Acta (BBA)-Biomembr. **1838**(6), 1594–1618 (2014)
5. Scrima, M., et al.: Structural features of the C8 antiviral peptide in a membrane-mimicking environment. Biochim. Biophys. (BBA)-Biomembr. **1838**(3), 1010–1018 (2014)
6. Marr, A.K., Gooderham, W.J., Hancock, R.E.: Antibacterial peptides for therapeutic use: obstacles and realistic outlook. Curr. Opin. Pharmacol. **6**(5), 468–472 (2006)
7. Wang, G.: Human antimicrobial peptides and proteins. Pharmaceuticals **7**(5), 545–594 (2014)
8. Scheetz, T., et al.: Genomics-based approaches to gene discovery in innate immunity. Immunol. Rev. **190**(1), 137–145 (2002)
9. Hancock, R.E., Chapple, D.S.: Peptide antibiotics. Antimicrob. Agents Chemother. **43**(6), 1317–1323 (1999)
10. Lata, S., Sharma, B., Raghava, G.: Analysis and prediction of antibacterial peptides. BMC Bioinform. **8**(1), 263 (2007)
11. Holland, J.H.: Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press, Cambridge (1992)
12. Piotto, S.P., et al.: YADAMP: yet another database of antimicrobial peptides. Int. J. Antimicrob. Agents **39**(4), 346–351 (2012)
13. Accelrys, Accelrys Materials Studio. Accelrys Inc., San Diego, California (2014)
14. MATLAB, R.: Version 8.1. 0.604 (R2013a). The MathWorks Inc., Natrick, Massachusetts (2013)
15. Chen, L., et al.: How the antimicrobial peptides kill bacteria: computational physics insights. Commun. Comput. Phys. **11**(3), 709 (2012)