Prediction of Antimicrobial Activity of Synthetic Peptides by a Decision Tree Model

Felipe Lira, Pedro S. Perez, José A. Baranauskas and Sérgio R. Nozawa
Appl. Environ. Microbiol. **2013, 79(10):3156. DOI: 10.1128/AEM.02804-12.**
Published Ahead of Print 1 March 2013.

AEM
Journals.ASM.org

Prediction of Antimicrobial Activity of Synthetic Peptides by a Decision Tree Model

Felipe Lira,[a] Pedro S. Perez,[b] José A. Baranauskas,[b] Sérgio R. Nozawa[a]

Laboratório de Expressão Gênica, Universidade Nilton Lins, Manaus, Amazonas, Brazil[a]; Departamento de Computação e Matemática, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil[b]

Antimicrobial resistance is a persistent problem in the public health sphere. However, recent attempts to find effective substitutes to combat infections have been directed at identifying natural antimicrobial peptides in order to circumvent resistance to commercial antibiotics. This study describes the development of synthetic peptides with antimicrobial activity, created *in silico* by site-directed mutation modeling using wild-type peptides as scaffolds for these mutations. Fragments of antimicrobial peptides were used for modeling with molecular modeling computational tools. To analyze these peptides, a decision tree model, which indicated the action range of peptides on the types of microorganisms on which they can exercise biological activity, was created. The decision tree model was processed using physicochemistry properties from known antimicrobial peptides available at the Antimicrobial Peptide Database (APD). The two most promising peptides were synthesized, and antimicrobial assays showed inhibitory activity against Gram-positive and Gram-negative bacteria. Colossomin C and colossomin D were the most inhibitory peptides at 5 μg/ml against *Staphylococcus aureus* and *Escherichia coli*. The methods described in this work and the results obtained are useful for the identification and development of new compounds with antimicrobial activity through the use of computational tools.

The increase in microbial resistance to commercial antibiotics and the need for protection against pathogenic agents have led to the development of rapid and efficient defense mechanisms. In this context, antimicrobial peptides represent a primitive defense mechanism that is present in all organisms from invertebrates to higher organisms, including humans (1).

In recent decades, various species of antimicrobial peptides have been isolated and found to exhibit a wide spectrum of activity against Gram-positive and Gram-negative bacteria (2). The production of these peptides in higher organisms plays an important role in the adaptive defense system and the regulation of various biological systems (3). This production is beneficial, as it occurs at low metabolic cost; the peptides are easily stored in large quantities and are rapidly made available to neutralize infections caused by microorganisms.

Due to their importance for the organism's immune system, antimicrobial peptides have become the object of much interest as a source of inspiration for the development of new drugs, based on changes to known molecules (4). The manipulation of these structures is a promising source of new antimicrobial peptides capable of blocking or inhibiting the growth of bacteria, fungi, parasites, tumor cells, and even encapsulated viruses like HIV (3, 5). Different methods are used to develop these artificial peptides, in particular, the synthesis of analogous peptides, which differ from natural peptides at one or more positions of the amino acid chain by substitution, deletion, or insertion of residues (6). This enables the residues crucial for antimicrobial activity to be determined and the desired effects to be modulated, with the purpose of making the analogous peptides more effective than the parental peptide (7, 8). However, these processes are costly and time-consuming. Various pharmaceutical companies have therefore encouraged the use of bioinformatics to investigate bioactive peptides as part of the search for new drugs, where the use of computer tools is complemented by genome, transcriptome, and proteome studies (9).

Based on the accumulation of information on the mechanism of action of antimicrobial peptides, various databases with detailed information about these peptides have been created (10). The diversity of forms and characteristics makes it difficult to develop methods capable of predicting the antimicrobial activity of peptides based on the similarity of their sequences alone. Therefore, there is a need for computational tools capable of minimizing the costs of predicting the antibacterial activity of peptides and planning peptides that are more effective against pathogens. To this end, methods such as the quantitative structure-activity relationship (QSAR), structure-activity relationship (SAR), and decision tree (DT) methods were developed to look for similar sequences and predict activity based on numerical data relating to structure and antimicrobial activity (10).

Based on studies to discover new therapeutic agents through peptide modeling using known antimicrobial peptides as a backbone, this study describes the induction of a decision tree model to predict the antimicrobial activity of synthetic peptides created by substitutions of amino acid residues in the parental peptide, which was obtained from the cDNA library of *Colossoma macropomum* (tambaqui), an Amazonian neotropical teleostean with high commercial value representing an economically relevant fish species from the Amazon basin (11).

MATERIALS AND METHODS

Identification of potential antimicrobial peptides. Coding sequences for antimicrobial peptides were identified by constructing the cDNA

Received 11 September 2012 Accepted 6 February 2013
Published ahead of print 1 March 2013
Address correspondence to Sergio R. Nozawa, srnozawa@gmail.com.

doi:10.1128/AEM.02804-12

TABLE 1 Peptides selected for Fmoc solid-phase synthesis and microbiological tests[a]

Peptide	Amino acid sequence	Molecular formula	ΔAA (%)	Mol wt	Charge	HR (%)	BI (Kcal/mol)	AP
Parental peptide	C-V**I**V**V****L****M**AQPGEC**F****L**G**L****I**FH-N	$C_{99}H_{156}N_{22}O_{23}S_2$		2,086.59	0	68	−1.73	+
Colossomin C	C-L**I**I**I****L****M**KKPGEC**F****L**S**L****I**YH-N	$C_{107}H_{175}N_{23}O_{24}S_2$	37	2,231.84	+2	57	−1.09	+
Colossomin D	C-L**I**V**V****L****M**KKPGEC**F****L**S**L****I**YH-N	$C_{105}H_{171}N_{23}O_{24}S_2$	26	2,203.78	+2	57	−1.00	+

[a] Peptides were selected after verification of antimicrobial activity through APD2. Underlined residues are hydrophobic; underlined residues in bold are both hydrophobic and located on the same peptide surface. ΔAA, amino acid substitutions; charge, peptide charge; HR, hydrophobic residues; BI, Boman index; AP, antimicrobial prediction (+, the peptide is predicted to have antimicrobial activity).

library of *Colossoma macropomum*, using the SMART cDNA library construction kit (Clontech), and sequencing more than 300 clones. A BLASTX search was performed in the GenBank database on a local server (www.ncbi.nlm.nih.gov), and a cDNA encoding a potential antimicrobial peptide with 19 residues was found. This peptide was initially named colossomin.

Synthetic peptide modeling. Based on the sequence of 19 amino acids of the colossomin peptide, we created five analogous peptides using a diagram proposed by Bordo and Argos to guide the substitutions of amino acid residues, increasing or maintaining the antimicrobial activity demonstrated by the parental peptide (12). The substitutions were based on the circumstances of net charge, total hydrophobic ratio (%), positive charge distribution to arrange the hydrophobic residues on the same surface, amphiphilic character, and the protein-binding potential (also known as the Boman index) (13). The sequences obtained after the residue substitutions were submitted to the predictive tool available at the Antimicrobial Peptide Database v2.34 (APD2; http://aps.unmc.edu/AP/main.php) (14) to verify their antimicrobial potential (15).

Peptide analog synthesis. To verify if the antimicrobial activity of the analog peptides could equal or exceed the effects observed for the parental peptide, analog peptides were constructed using the 9-fluorenylmethoxy carbonyl (Fmoc) solid-phase peptide synthesis strategy (16). The C-terminal amino acid of the native peptide was maintained in some analogs, and the resulting peptides were named colossomin C and colossomin D.

Decision tree experimental setup. Induction of decision trees is a machine learning approach that has been applied to several tasks. Decision trees (DT) are well-suited for large, real-world tasks, as they scale well and can represent complex concepts by constructing simple yet robust logic-based classifiers amenable to direct expert interpretation (15). Top-down inductions of decision tree algorithms generally choose a feature that partitions the training data according to some evaluation functions (17). The partitions are then recursively split until some stopping criterion is reached. After that, the decision tree is pruned in order to avoid over-fitting (18). In our experiments, we used the algorithm J48 from Weka (19), a library of several machine learning algorithms. J48 is a Java implementation of the well-known C4.5 algorithm (17).

The training data are composed of 60 antimicrobials, each described by 53 molecular descriptors. These descriptors include structure, net charge, hydrophobic residues, and Boman index, among others, and were obtained using the program package Marvin Beans (www.chemaxon.com/download/marvin). Peptides were divided into four classes, according to their microbial activity (none, low, medium, and high) as follows. A specific peptide was classified as "none" if no activity was found in any of the cell types, "low" if the activity occurred in only one organism, "medium" if the activity occurred in exactly two organisms, and "high" if it occurred in three or more organisms. According to this procedure, the distribution of peptides into the classes none, low, medium, and high in the training data was 3 (5%), 17 (28%), 20 (33%), and 20 (33%), respectively.

In order to select the most predictive attributes and find the best configuration of parameters for J48, we used a technique called "windowing" (20), in which the decision tree model begins to learn with only a fraction (window) of the examples in the data set. A classifier is induced using the initial window, and it is tested using the examples not present in the window. A fraction of the examples outside the window, which were misclassified, is added to the window. A new classifier is induced and tested, and the process is repeated until there are no misclassifications. Windowing can be repeated many times (trials), starting with a different initial window each time. We used the windowing provided by C4.5. After applying the technique with different configurations of C4.5 and windowing itself, we chose the tree with the best test error. We then built a new data set, composed of only those attributes found in the unpruned version of the best tree: net charge, hydrogen, oxygen, isoelectric point, peptide accessible surface area (ASA_P), Balaban index, Dreiding energy, minimal projection radius, and the logarithm ratio of the partition coefficient [log(P)]. Using Weka, we ran J48 over the data set with the filtered attributes. The default parameters for the inducer were used, except for the parameter *M*, which determines the minimum number of examples that a leaf must contain. We used a value for *M* of 3, with which the best tree found before was set.

Spectrum of activity prediction using decision trees. After the detection of the peptides' antimicrobial potential using the Antimicrobial Peptide Database, they were classified by the decision tree, and the activity spectrum (none, low, medium, or high) was inferred. This was done by considering the criteria adopted to determine the peptide activity according to the types of organisms on which it will act, inhibiting, or extinguishing their growth.

Antimicrobial tests. Microbiological assays were carried out using *Staphylococcus aureus* (Gram positive) and *Escherichia coli* (Gram negative). Inoculated petri dishes were analyzed by the disk agar-diffusion method with 10 μl of each synthetic peptide diluted with water to 5 μg/ml. Four paper disks 5 mm in diameter were placed in each petri dish with solid LB (Luria Bertani) culture medium and impregnated with diluted peptides. The petri dishes were incubated for 20 h at 37°C to determine the formation of growth inhibition zones.

Statistical analysis. The data (i.e., the diameters of the zones of inhibition formed by each synthetic peptide) were analyzed using the Wilcoxon–Mann-Whitney test in the R program (http://www.r-project.org/) to compare antimicrobial activities. The antimicrobial activities of each peptide against different bacteria were analyzed and compared in order to detect the most efficient antimicrobial peptide.

RESULTS

Synthetic peptide modeling. After computational modeling and prior analysis of the antimicrobial potential, five peptides were designed using the parental peptide as the scaffold, and only two were selected for Fmoc solid-phase synthesis and microbiological tests (Table 1).

Decision tree experimental setup. The decision tree induced is composed of nine decision nodes containing the eight attributes (net charge, hydrogen, oxygen, isoelectric Point, log(P) of non-ionic species, ASA_P, Balaban index, and Dreiding energy) and 10 leaves indicating the level of activity of the synthetic peptides (high, medium, low, or no activity) (Fig. 1).

The decision tree model was validated using a leave-one-out

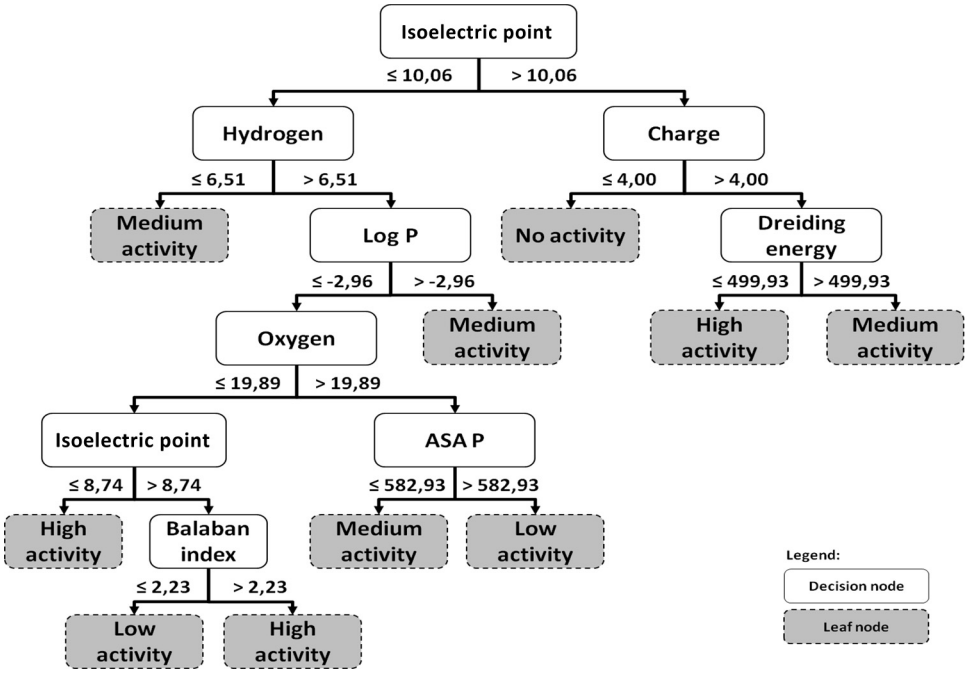

FIG 1 Decision tree model created by the algorithm J48 using the physicochemical properties of the peptides descriptors.

method. The values of accuracy, area under the ROC (receiver operating characteristic) curve, true-positive rate, true-negative rate, precision, and *F* measure are shown in Table 2. Each line of the table shows the measures for each of the four possible class values, and the last line shows the weighted average of the measures, representing the overall values for the classification task.

Antimicrobial tests. After an incubation period, an inhibition area surrounding the paper disks containing colossomin C and colossomin D on *Staphylococcus aureus* and *Escherichia coli* cultures was visible (Fig. 2). The average sizes of inhibition zones formed by colossomin C and colossomin D on these cultures were 2.9 ± 0.1 cm and 2.25 ± 0.02 cm (*S. aureus*) and 1.37 ± 0.08 cm and 0.625 ± 0.04 cm (*E. coli*), respectively. The parental peptide did not form zones of inhibition on any bacterial culture.

Statistical analysis. Statistical analyses showed that the activity of colossomin C was significantly different ($P = 0.0147$) from that of colossomin D for *E. coli* and *S. aureus*. In both cases, it was not possible to reject the null hypothesis for any significance level above 1.5%.

TABLE 2 Performance measurements for the classification of peptide activity[a]

Class value	Accuracy	AUC	TP	TN	Precision	*F* measure
None		0.97	1.00	0.97	0.60	0.75
Low		0.81	0.65	0.81	0.58	0.61
Medium		0.82	0.70	0.85	0.70	0.70
High		0.87	0.70	0.95	0.88	0.78
Total	0.70	0.84	0.70	0.88	0.72	0.70

[a] AUC, area under the ROC curve; TP, true-positive rate; TN, true-negative rate.

DISCUSSION

The results of this study show that the use of decision trees to evaluate the antimicrobial activity of synthetic peptides enables the creation of more effective models for use in the development of new drugs, using known peptides as scaffolds for designing new compounds, and reducing the cost and time required for research. As demonstrated in this study and shown in previous works (21, 22), the development of algorithms for decision tree models is an efficient tool for predicting the antimicrobial activity and construction of peptides for various therapeutic uses.

The inhibitory activity shown by colossomin C and colossomin D, which were subjected to *in vitro* tests with *S. aureus* and *E. coli*, provides the basis for a series of possible studies on the importance of these antimicrobial peptides and their mechanism of action. It also increase the possibility of using these synthetic antimicrobial peptides as important biotechnological products for treating multidrug-resistant pathogens. The antimicrobial activity demonstrated by colossomin C and colossomin D against *S. aureus* shows that these peptides are more efficient in inhibiting the growth of Gram-positive bacteria than that of Gram-negative bacteria (*E. coli*).

The results indicate that the induction of the amphiphilic behavior and the positive charge are responsible for destabilization of the membrane surface and may induct pores by the partial or total insertion of the hydrophobic portion (20). Based on the method used in this study and the possibility of predicting the antimicrobial activities of synthetic peptides created by site-targeted mutations, this methodological pipeline has great value for the discovery and development of new natural antibiotics using known peptides.

In addition, this work may highlight the efficiency of two synthetic peptides as promising antimicrobial agents for treating in-

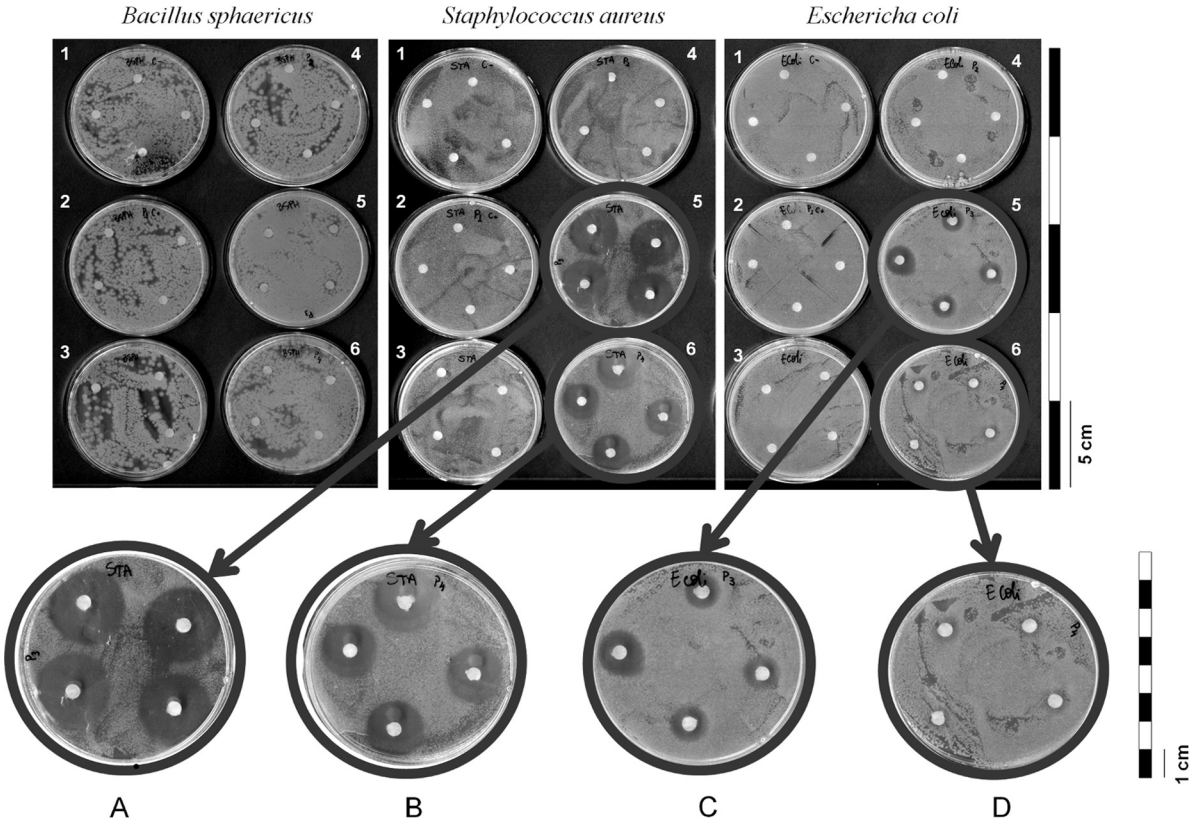

FIG 2 Antibiogram tests. Plates are labeled as follows: 1, brain natriuretic peptide (−); 2, colossomin (+); 3, no peptide; 4, colossomin B; 5, colossomin C; 6, colossomin D; A, *S. aureus* versus colossomin C; B, *S. aureus* versus colossomin D; C, *E. coli* versus colossomin C; D, *E. coli* versus colossomin D.

fections caused by Gram-positive bacteria, drawing more attention to these new methods for the discovery of new drugs.

ACKNOWLEDGMENTS

We thank Adriano Monteiro de Castro Pimenta and Thiago Verano-Braga for help with the peptide synthesis at Laboratorio de Venenos e Toxinas Animais—LVTA, UFMG.

We thank CNPq, CAPES, and FAPEAM for financial support.

REFERENCES

1. **Deslouches B, Phadke SM, Lazarevic V, Cascio M, Islam K, Montelaro RC, Mietzner TA.** 2005. De novo generation of cationic antimicrobial peptides: influence of length and tryptophan substitution on antimicrobial activity. Antimicrob. Agents Chemother. **49**:316–322.
2. **Yin Z, He W, Chen W, Yan J, Yang J, Chan S, He J.** 2006. Cloning, expression and antimicrobial activity of an antimicrobial peptide, epinecidin-1, from the orange-spotted grouper, Epinephelus coioides. Aquaculture **253**:204–211.
3. **Hancock REW, Scott MG.** 2000. The role of antimicrobial peptides in animal defenses. Proc. Natl. Acad. Sci. U. S. A. **97**:8856–8861.
4. **Landon C, Barbault F, Legrain M, Menin L, Guenneugues M, Schott V, Vovelle F, Dimarcq JL.** 2004. Lead optimization of antifungal peptides with 3D NMR structures analysis. Protein Sci. **13**:703–713.
5. **Gordon YJ, Romanowski EG, McDermott AM.** 2005. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. Curr. Eye Res. **30**:505–515.
6. **Boman H, Wade D, Boman I, Wahlin B, Merrifield R.** 1989. Antibacterial and antimalarial properties of peptides that are cecropin-melittin hybrids. FEBS Lett. **259**:103–106.
7. **Marshall SH, Arenas G.** 2003. Antimicrobial peptides: a natural alternative to chemical antibiotics and a potential for applied biotechnology. Electronic J. Biotechnol. **6**:271–284.
8. **Tossi A, Tarantino C, Romeo D.** 1997. Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. Eur. J. Biochem. **250**:549–558.
9. **Quinlan JR.** 1993. C4.5: programs for machine learning. Morgan Kaufmann Publishers, San Francisco, CA.
10. **Lata S, Sharma BK, Raghava GPS.** 2007. Analysis and prediction of antibacterial peptides. BMC Bioinformatics **8**:263.
11. **Araújo-Lima C, Goulding M.** 1998. Os frutos do tambaqui: ecologia, conservação e cultivo na Amazônia. Estudos do Mamirauá, vol IV. Sociedade Civil Mamirauá, Tefé, Brazil.
12. **Bordo D, Argos P.** 1991. Suggestions for safe residue substitutions in site-directed mutagenesis. J. Mol. Biol. **217**:721–729.
13. **Kourie JI, Shorthouse AA.** 2000. Properties of cytotoxic peptide-formed ion channels. Am. J. Physiol. Cell Physiol. **278**:C1063–C1087.
14. **Wang G, Li X, Wang Z.** 2009. APD2: the updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res. **37**:D933–D937.
15. **Patrzykat A, Gallant J, Seo J, Pytyck J, Douglas SE.** 2003. Novel antimicrobial peptides derived from flatfish genes. Antimicrob. Agents Chemother. **47**:2464–2470.
16. **Chan W, White PD.** 2000. Fmoc solid phase peptide synthesis: a practical approach. Oxford University Press, Oxford, United Kingdom.
17. **Rezende SO.** 2003. Sistemas inteligentes: fundamentos e aplicações. Manole, Barueri, São Paulo, Brazil.
18. **Kingsford C, Salzberg SL.** 2008. What are decision trees? Nat. Biotechnol. **26**:1011–1013.
19. **Witten IH, Frank E.** 2005. Data mining: practical machine learning tools and techniques. Morgan Kaufman, San Francisco, CA.
20. **Shai Y.** 2002. Mode of action of membrane active antimicrobial peptides. Peptide Sci. **66**:236–248.
21. **Dathe M, Wieprecht T.** 1999. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. Biochim. Biophys. Acta **1462**:71–87.
22. **Lee SY, Kim S, Kim SS, Cha SJ, Kwon YK, Moon BR, Lee BJ.** 2004. Application of decision tree for the classification of antimicrobial peptide. Genomics Informatics **2**:121–125.