Molecular Biology Research Communications 2013;2(1-2):39-45 MBRC

Original Article Open Access

In silico prediction of anticancer peptides by TRAINER tool

Zohre Hajisharifi[1], Hassan Mohabatkar*

Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran

A B S T R A C T

Cancer is one of the causes of death in the world. Several treatment methods exist against cancer cells such as radiotherapy and chemotherapy. Since traditional methods have side effects on normal cells and are expensive, identification and developing a new method to cancer therapy is very important. Antimicrobial peptides, present in a wide variety of organisms, such as plants, amphibians and mammals, are newly discovered agents. These peptides have various structures, sizes and molecular compositions; hence developing a computational method to predict these anticancer peptides is useful. In the present study, first, 2 databases with 138 and 206 anticancer and non-anticancer peptides were introduced, classified by TRAINER. TRAINER (http://www.baskent.edu.tr/~hogul/TRAINER/) is a new online tool designed for classification of any alphabet of sequences. TRAINER allows users to select from among several feature representation schemes and supervised machine learning methods with relevant parameters. In this study, Naive Bayes and radial basis were used in a support vector machine. The accuracy and specificity in combination of features by Naive Bayes were 83% and by radial basis 87% and 92% respectively. The results demonstrate that two methods are useful for classification of these peptides; however, the accuracy of Radial Basis is higher than Naive Bayes.

Key words: Cancer, Anticancer peptides, TRAIER, Naive Bayes, Radial Basis

INTRODUCTION

Cancer is a disease of uncontrolled cell division and one of the leading causes of death in the developed world [1]. At first the human body uses its immune system to recognize and destroy cancer cells but usually cancer cells evade immune surveillance [2, 3]. Moreover, current cancer therapy is limited by the extreme toxicity of chemotherapeutic drugs and radiotherapy, which can also cause widespread damage to non-cancerous tissues [4]. Therefore, it is necessary to develop new anticancer agents without any side effect on normal cells. Innate immunity peptides seem to overcome these limitations via distribution

* Address for correspondence: Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran. ; Tel: +98-311-7934391; Fax: +98-311-7932342; E-mail: h.mohabatkar@ast.ui.ac.ir

ISSN 2322-181X

of the membranes [5-7]. The "biologics" treatment option against cancer includes the use of proteins, monoclonal antibodies, and peptides. These peptides are mostly cationic and adopt an amphipathic structure [8]; they are found in most living species and are released in response to bacterial infection by a different regulatory process [9]. Antimicrobial peptides are generally secretory in nature [10]. These peptides appear to act via a specific, but not receptor-mediated, permeabilization of microbial and cancer membranes. Host defense membrane-active peptides can be divided into two major groups, the first including peptides that act against cancer cells and bacteria but not against normal cells, and the second including peptides that act against normal and cancer cells and bacteria [11]. The anticancer effect of peptides is based on the disrupting of cytoplasmic and mitochondrial membranes [12-15]. Membranes of cancer cells with negative charge attach to the peptides with positive charge. The amphipathic structure of peptides is a feature that allows them to insert themselves in the membrane and permeate it via the formation of pores [16]. However these peptides have small size and variety in secondary structure and composition but they share the following features: positive charge, high content of hydrophobic residues, amphipathic fold, ease of synthesis and modification and tumor penetrating ability [17-19].

To reduce costs and time, there is a need to develop a computational method for the prediction of anticancer peptides. A number of databases such as the Antimicrobial Peptide Database (APD2) [20] and the Collection of Antimicrobial Peptides (CAMP) have identified anticancer peptides [21]. APD2 is available at (http://aps.unmc.edu/AP/main.php).

MATERIALS AND METHODS

Datasets: 192 experimentally validated ACPs were extracted from the literature as well as the antimicrobial database (APD2) as positive database. A non-anticancer database was introduced manually from the Universal Protein Resource (UniProt), at http://www.uniprot.org. Generally, non-secretory proteins and randomly cut out peptides with the same length range were selected to introduce the negative database. Finally, negative database was used containing 215 non-anticancer peptides. Since there are the same sequences in each database, the CD-HIT tool (http://cd-hit.org) was used to eliminate the duplicates [22]. In the end, positive and negative databases were used, containing 138 and 206 peptides respectively.

Radial basis function networks: Like neural networks, radial basis function (RBF) networks [23] have also been proven to be universal approximators [24]. RBF networks are much easier to designe and train than neural networks and have strong tolerance to input noise, which enhances the stability of the designed systems. Therefore, it is reasonable to consider an RBF network as a competitive method for nonlinear controller designs. The basic computations of the RBF network consist of three steps:

1) Input Layer Computation: At the input layer, each input ($x_{p,i}$) is scaled by the input weights ($u_{i,h}$) presenting the weight connection between the i^{th} input and RBF unit h.

(1) $y_{p,h,i} = x_{p,i}\, u_{i,h}$

where vector $y_{p,h} = \{y_{p,h,1}, y_{p,h,2} \ldots y_{p,h,i} \ldots y_{p,h,I}\}$ is the scaled inputs, h is the index of RBF units from 1 to H, i is the index of inputs from 1 to I, and p is the index of training patterns from 1 to P.

2) Hidden Layer Computation: The output of the RBF unit h is calculated by:

(2) $\phi_h(x_p) = \exp\left(\frac{||y_{p,h} - c_h||2}{\sigma_h}\right)$

Where $\phi_h(\bullet)$ is the activation function of the RBF unit h,. c_h and σ_h are the center and width, the key properties to describe the RBF unit h,, and $\| \bullet \|$ represents the computation of Euclidean norm of two vectors.

3) Output Layer Computation: The network output for pattern x_p is calculated as the sum of weighted outputs from RBF units.

(3) $o_p = \sum_{h=1}^{H} \phi_h(\mathbf{x}_p) w_h + w_0$

Where w_h represents the weight value on the connection between the RBF unit h and network output. W_0 is the bias weight.

Naive Bayes: A naive Bayes classifier is a simple probabilistic classifier based on the application of Bayes' theorem with strong (naive) independence assumptions. Considering features as independent makes their learning simple and computationally efficient [25]. Because of these advantages, this method is widely used. Parameter estimation for Naive Bayes models is done using maximum likelihood. The method is suited when the dimensionality of the input is high.

In the present study, Radial Basis and Naive Bayes were used for classification and evaluating the performance of classifiers.

Performance of the classifier is measured in terms of sensitivity (SEN), specificity (SPEC), accuracy (ACC) and Matthew's correlation coefficients (MCC) given by equations (4-7),

(4) SEN=TP/(TP+FN)

(5) SPEC=TN/(TN+FP)

(6) ACC=(TN+TP)/(TN+FP+TP+FN)

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

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively.

The performances of classifiers are presented in tables (1-4).

Trainable Short Sequence Classifier: TRAINER is a new online and flexible tool for biosequence analysis. TRAINER interface can provide a user-friendly environment for any basic internet user. This system can respond to thousands of short sequences in seconds and has been shown to produce accurate results for all types of biological sequences [26]. TRAINER is available at (http://www.baskent.edu.tr/~hogul/TRAINER/). Biological sequences are of varying lengths, cannot be directly fed in to a classifier and need to be represented by a number of numerical features. In this tool, there are three different composition representations as 1-mer, 2-mer, 3-mer and their combinations. In the present study, three different compositions and the combination of all vectors was used.

RESULTS

In this study, two machine learning methods with three different compositions of features and their combinations were used for the classification and evaluation of the performance of the classifier. The results demonstrated that these methods were useful for predicting anticancer peptides. Both methods are useful for classifying anticancer and non-anticancer peptides but accuracy, sensitivity, specificity and Matthews' correlation coefficients of radial basis were found to be better than Naive Bayes', and in this analysis, the combination of all vectors yielded the best results.

Table 1. Matthew's correlation coefficients of two machine learning methods in three different compositional representations

	Mathew's correlation coefficients			
Methods	**1-mer**	**2-mer**	**3-mer**	**all**
Naive Bayes	0.67	0.58	0.54	0.65
Radial Basis	0.63	0.62	0.59	0.74

Table 2. Sensitivity and specificity of two machine learning methods in three different compositional representations

	Sensitivity				Specificity			
Methods	**1-mer**	**2-mer**	**3-mer**	**all**	**1-mer**	**2-mer**	**3-mer**	**all**
Naive Bayes	81%	82%	67%	83%	86%	79%	85%	83%
Radial Basis	87%	86%	98%	92%	84%	80%	74%	86%

Table 3. Accuracy of two machine learning methods in three different compositional representations

Methods	Accuracy			
	1-mer	2-mer	3-mer	all
Naive Bayes	84%	80%	76%	83%
Radial Basis	80%	82%	79%	87%

DISCUSSION

Cancer therapy is an attractive topic in the world. There are many traditional methods for treating cancer, but these methods have limitations. Therefore, we focused on a more specific method. Anticancer peptides of human, plant, and animal origin are shown to be new agents against cancerous cells. The oncolytic effect of peptides depends on their cationic and amphipathic structure [8]. Since gangliosides are present in both normal and cancer cells (although by different proportions), the function of the peptides is not specific, but they may be attracted to cancer cells more frequently than normal cells. The positive charge of the peptides is proposed to initiate electrostatic interaction with the negatively charged membranes of tumor cells. These features could lead to the permeation of peptides into the membrane and a subsequent complete membrane disruption. In addition, they depolarize the transmembrane potential of cancer cells and kill the cells by perturbation in the plasma membrane [11, 27]. Since the examination of peptides in vitro and in vivo is time consuming and expensive, providing a computational method for the prediction of anticancer peptides can be useful. There are several methods to predict the different aspects of proteins, based on amino acid sequence, template and amino acid composition (AAC). PseAAC concept has also been typically used to predict several aspects of proteins, including cyclins [28], risk type of human papillomaviruses [29], GABAA receptors [30], metalloproteinase family [31], antibacterial peptides [32] and allergenic Proteins [33].

Therefore, we collected anticancer and non-anticancer peptides from databases and articles and classified them using two machine learning methods in TRAINER tool. TRAINER is a new online tool acceptable for the classification of biosequences. In the present study, ACC, SEN and SPEC of these methods demonstrated that TRAINER is a useful tool for predicting anticancer and non-anticancer peptides. MCC is a measure for assessing the quality of binary classifications. This measure has a value between -1 and +1. When MCC is higher than 0.7, it is acceptable for predictors. In the present study, the best results for MCC were found to be in combinations of all vectors, MCC of Radial Basis being higher than 0.7, and therefore, acceptable. The results show that for predicting the mentioned peptides, the accuracy of Radial Basis was more than that of Naive Bayes.

Acknowledgments

Support of this study by the University of Isfahan in acknowledged.

Conflict of Interest: Author has no financial or any non-financial competing interests.

REFERENCE

1. Suarez-Jimenez GM, Burgos-Hernandez A, Ezquerra-Brauer JM. Bioactive peptides and depsipeptides with anticancer potential: Sources from marine animals. Mar Drugs 2012;10:963-986.
2. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002;3:991-998.
3. Mocellin S, Rossi CR, Nitti D. Cancer vaccine development: on the way to break immune tolerance to malignant cells. Exp Cell Res 2004;299:267-278.
4. Massodi I, Moktan S, Rawat A, Bidwell GL, Raucher D. Inhibition of ovarian cancer cell proliferation by a cell cycle inhibitory peptide fused to a thermally responsive polypeptide carrier. Int J Cancer 2010;126:533-544.
5. Shadidi M, Sioud M. Selective targeting of cancer cells using synthetic peptides. Drug Resist Update 2003;6:363-371.
6. Leuschner C, Hansel W. Membrane disrupting lytic peptides for cancer treatments. Curr Pharm Design 2004;10:2299-2310.
7. Papo N, Shahar M, Eisenbach L, Shai Y. A novel lytic peptide composed of D, L amino acids selectively kills cancer cells in culture and in mice. J Biol Chem 2003;278: 21018-21023.
8. Tossi A, Sandri L, Giangaspero A. Amphipathic, α-helical antimicrobial peptides. Pept Sci 2000;55:4-30.
9. Diamond G, Beckloff N, Weinberg A, Kisich KO. The roles of antimicrobial peptides in innate host defense. Curr Pharm Design 2009;15:2377.
10. Bals R. Epithelial antimicrobial peptides in host defense against infection. Respir Res 2000;1:141-150.
11. Papo N, Shai Y. Host defense peptides as new weapons in cancer treatment. Cell Mol Life Sci 2005;62:784-790.
12. Mai JC, Mi Z, Kim SH, Ng B, Robbins PD. A proapoptotic peptide for the treatment of solid tumors. Cancer Res 2001;61:7709-7712.
13. Ellerby HM, Arap W, Ellerby LM, Kain R, Andrusiak R, Rio GD, Krajewski S, Lombardo CR, Rao R, Ruoslahti E, Bredesen DE, Pasqualini R. Anti-cancer activity of targeted pro-apoptotic peptides. Nat Med 1999;5:1032-1038.
14. Shai Y. Mode of action of membrane active antimicrobial peptides. Biopolymers 2002; 66:236-248.
15. Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RA. Phylogenetic perspectives in innate immunity. Science 1999;284:1313-1318.
16. Shai Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. Biochim Biophys Acta1999;1462:55-70.
17. Nijnik A, Hancock R. Host defense peptides: antimicrobial and immunomodulatory activity and potential applications for tackling antibiotic-resistant infections. Emerg Health Threats J 2009;2:e1.
18. Thayer AM. Improving peptides. Chem Eng News 2011;89:13-20.

19. Borghouts C, Kunz C, Groner B. Current strategies for the development of peptide-based anti-cancer therapeutics. J Pept Sci 2005;11:713-726.
20. Wang G, Li X, Wang Z. APD2: The updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res 2009;37:D933-D937.
21. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S. CAMP: a useful resource for research on antimicrobial peptides. Nucleic acids Res 2010;38:D774-D780.
22. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658-1659.
23. Moody J, Darken CJ. Fast learning in networks of locally-tuned processing units. Neural Comput 1989;1:281-294.
24. Park J, Sandberg IW. Universal approximation using radial basis-function networks. Neural Comput 1991;3:246-257.
25. Duda RO, Hart PE, Stork DG. Pattern Classification, 2nd edition. John Wiley and Sons, 2000.
26. Oqul H, Kalkan AT, Umu SU, Akkaya MS. Trainer: A general-purpose trainable short biosequence classifier. Protein Pept Lett 2013 [Epub ahead of print].
27. Sundelacruz S, Levin M, Kaplan D L. Role of membrane potential in the regulation of cell proliferation and differentiation. Stem Cell Rev Rep 2009;5:231-246.
28. Mohabatkar H. Prediction of cyclin proteins using Chou's pseudo amino acid composition. Protein Pept Lett 2010;17:1207-1214.
29. Esmaeili M, Mohabatkar H, Mohsenzadeh S. Using the concept of Chou's pseudo amino acid composition for risk type predictionof human papillomaviruses. J Theor Biol 2010;263:203-209.
30. Mohabatkar H, Mohammad-Beigi M, Esmaeili A. Prediction of GABAA receptor proteins using the concept of Chou's pseudoaminoacid composition and support vector machine. J Theor Biol 2011;281:18-23.
31. Mohammad-Beigi M, Behjati M, Mohabatkar H. Prediction of metalloproteinase family based on the concept of Chou's pseudoamino acid composition using a machine learning approach. J Struct Funct Genomics 2011;12:191-197.
32. Khosravian M, Faramarzi FK, Mohammad-Beigi 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;20:1-7.
33. Mohabatkar H, Mohammad-Beigi M, Abdolahi K, Mohsenzadeh S. Prediction of allergenic proteins by means of the concept of Chous pseudo amino acid composition and a machine learning approach. Med Chem 2013;9;133-137.