ISSN : 0974 - 7435 Volume 10 Issue 18

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(18), 2014 [10480-10484]

A prediction model for research on structure-activity relationship of antimicrobial peptides

Quhuan Li
School of Bioscience and Engineering, South China University of Technology, Guangzhou, 510006, (CHINA)
E-mail : liqh@scut.edu.cn

ABSTRACT

Gene-encoded antimicrobial peptides (AMPs) are widely distributing in all classes of life ranging from plants to animals, including amphibians, birds, fish, and mammals etc. The characteristics of wide-spectrum, rapidness and specificity, and the activities against several bacteria, fungi, viruses, protozoa and cancerous cells, allow the development of numerous antimicrobial peptides with potentially useful properties as therapeutic agents. Development of activity predicted tools based on understanding the role of relationship of structure-activity is inevitable for their drug designs. In this study, we manually collected 1162 antimicrobial sequences firstly, then builded a comprehensive design platform for studying structure-activity relationship of antimicrobial peptides, finally constructed a new activity prediction model with stepwise discriminant analysis. This model integrated seven physicochemical parameters, including length, molecular weight, theoretical pI, primary amino acid composition, Instability index, aliphatic index and grand average of hydropathicity, charge. The prediction model correctly classified 70% of the known activity peptides in the database, and predicted 77.8% of the new unknown activity peptides. The results indicated that the model consisting of four discriminant functions can recognize and classify activity for antimicrobial peptides effectively.

KEYWORDS

Antimicrobial peptides; Structure activity relationship; Step discriminant analysis; Prediction model.

INTRODUCTION

Gene-encoded antimicrobial peptides (AMPs) are an evolutionarily conserved component of the innate immune response. They widely distribute in all classes of life ranging from plants to animals, including amphibians, birds, fish, and mammals etc[1,2]. So far, More than 900 such peptides have been discovered. They fit into at least four structural classes, namely β-sheet, amphipathic α-helices, extended structures, and loop structures[3]. Generally the net charge is positive[4]. Many studies have shown these peptides are potent, broad spectrum antibiotics, since they have the ability of killing Gram negative and Gram positive bacteria, mycobacteria, enveloped viruses, fungi and even transformed or cancerous cells[5].

As the time of extensive use of the conventional antibiotic and the continuous emerging of the bacterial resistance, the work of finding and developing new molecules that can overcome the limitations of the present drugs is urgent[2,6]. Naturally antimicrobial peptides (AMPs) have been considered as potentially useful properties as therapeutic agents because of their unique mechanism to combat antibiotic resistance by microorganisms[2,4]. However, in clinical applications, there are still many challenges, such as (i) Toxicity, some kinds of antimicrobial peptides have not only the ability to inhibit and kill bacteria, but also the hemolytic to mammalian cell. The reasons maybe come from the low selectivity for bacterial cell membrane[7,8]. (ii) The short half-life in vivo due to rapid proteolytic cleavage. (iii) Stimulation of an immune response[9].

As a good antimicrobial peptide must be selective to bacterial membranes while maintaining low mammalian cell cytotoxicity[10]. To this point, the important work is to reduce the toxicity and improve the maximize activity as far as possible. Molecular design based on the relationship of structure and activity would be the good way to solve this problem. Early studies have indicated that structure-activity relationship studies were related with at least six interrelated physicochemical parameters that modulate their activity and specificity, such as sequence, size, charge, amphipathicity, and hydrophobicity[11]. These parameters are intimately related, so that modifications aimed at altering one can also result in significant changes to one or more of the others. Understanding these interrelationships is the key to designing novel peptides with increased potency and directed activity[12-15]. Simultaneously, activity prediction for a new AMP can be achieved with structure-activity relationship.

How to determine the structure-activity relationship despite that the physicochemical parameters were found to be play an important role. And then, how to make use of this relationship to design peptides with high activity? Herein, we constructed a new activity prediction model with step discriminant analysis, in order to provide an integrated useful tool for activity determination of antimicrobial peptides and facilitate further investigation of de novo antimicrobial peptide design and partial modifications.

MATERRIAL AND METHOD

Data collection

A SARP (structure-activity relationship of peptides) database was built up for the effective organization and administration of antimicrobial peptides. With keywords searching, including antibacterial peptides, antifungal peptides, anticancer peptides, and antiviral peptides, the data were collected from UniProt Knowledgebase (UniProtKB, http://www.ebi.uniprot.org/). UniProtKB is a protein database curated by experts, consisting of two sections. UniProtKB/Swiss-Prot (containing reviewed, manually annotated entries) and UniProtKB/TrEMBL (containing unreviewed, automatically annotated entries). All entries in SARP database were come from UniProtKB/Swiss-Prot. 1162 queries were obtained after discarding false positive query with manual evaluation. And then, physicochemical parameters were calculated with ProtParam software http://us.expasy.org/tools/prot-param.html. Here, ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein. After that, all parameters were put into database.

Data organization in database

Each query in database has two portions of information. The first portion is the concise description of the sequence, such as ID, uniprot accession, name, origin, taxonomy, reference, PDB, family, domain, function, mature sequence. A unique accession number 'ID' that defines each record in the SARP database. A 'uniprot accession' number is same with the one in UniProtKB/Swiss-Prot, which provides hyperlinks to the UniProtKB. The 'name', 'origin', 'taxonomy', 'reference' fields contain the name of the AMP as used in the literature, source organism, taxonomic classification of the source organism, and bibliographic references. If crystal structure of record stored in SARP database is available, the field 'PDB' will contains an accession number that can hyperlink to the Protein Data Bank (PDB). 'family' and 'Domain' fields defines function cluster according to sequence identity. The field 'mature sequence' is the extent of a polypeptide chain in the mature protein discarding signal peptide and propeptide from primary translated protein. This mature peptide is enough as a complete function unit. The second portion was physicochemical parameters, including 'length (LEN)', 'molecular weight (MW)', 'theoretical pI (pI)', 'primary amino acid composition (AA)', 'Instability index (II)', 'aliphatic index and grand average of hydropathicity (GRAVY)', and 'charge (CH)'. Besides ID, the former ten items information were from UniProtKB, the latter eight items computed with ProtParam tool for each mature peptide.

Stepwise discriminant analysis

Representative 119 entries were picked up from 1162 entries stored in SARP database. A stepwise discriminant analysis was performed to investigate the activities of the known groups. This statistical analysis builds a predictive model of group membership based on observed physical and chemical parameters of each sample. If a stepwise method is used to estimate the discriminant function, the Mahalanobis measure is one of the most appropriate methods. So each individual was allocated to the group with classified algorithm of Mahalanobis distance. The criteria for entry and removal (F value) were set to be 0.02 and 0.01. The statistical treatment was carried out with SPSS 17.0 for Windows.

RESULT AND DISSCUSS

Activity statistics of antimicrobial peptides

A high-quality, manually annotated, non-redundant protein sequence database was built, which combined information extracted from UniProtKB/Swiss-Prot and activity annotation from scientific literature. Among 1162 entries in SARP database, antibacterial and antifungal, anticancer, antiviral peptides process 580, 185, 135, 262 entries respectively (Figure 1). Based on dominating activity of every mature AMP, typical 119 entries were picked up. Among these, 43, 36, 11, 29 were come from four kinds of AMP above (TABLE 1).

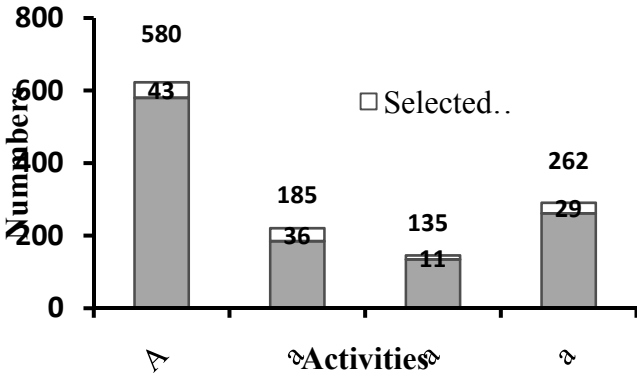

Figure 1: Four activity statistics of antimicrobial peptides. Gray columns() and numbers on top of them were mumbers for four kinds antimicrobial peptides in SARP database. Blank columns (□) and numbers enbedded in them were numbers of selected peptides from database.

TABLE 1 : Eight variables, known and predicted activities of selected AMPs

ID	Independents*								KF	PF
	LEN	MW	pI	AA	II	AI	GRAVY	CH		
293	23	2667	8.75	4	-14.16	101.3	0.778	1	1	1
897	19	1927	9.99	8	-3.32	138.42	0.774	2	1	1
389	30	2997	10	20	0.38	117.33	0.47	3	1	1
326	23	2343	8.5	20	1.4	114.78	0.252	1	1	1
707	30	3109	10.3	20	9.26	94.67	0.027	5	1	1
610	26	2868	7.97	5	9.4	63.85	0.112	1	1	2
158	67	7118	6.73	8	-5.13	17.46	-1.791	-2	2	2
520	41	4186	8.62	5	3.37	31.22	-0.066	4	2	2
519	41	4229	8.62	5	3.37	28.78	-0.195	4	2	2
910	125	13666	6.97	20	12.95	60.24	-0.458	0	2	4
511	46	5196	6.42	5	17.03	38.26	-0.539	-1	2	2
637	112	10749	4.86	20	4.8	80.09	0.341	-3	3	3
825	66	7281	7.64	8	20.24	44.39	-0.629	1	3	2
176	13	1488	6.07	1	20.48	127.69	0.669	0	3	3
165	15	1657	6.41	1	26.89	104	-0.053	0	3	3
672	113	11096	3.99	20	28.72	70.88	0.211	-6	3	3
103	13	1265	7.1	8	-22.68	44.62	-0.823	0	4	2
932	285	31996	8.76	19	21.62	70.14	-0.584	5	4	4
940	261	29200	8.86	3	23.3	88.54	-0.382	5	4	4
794	270	30119	9.1	3	25.89	89.59	-0.269	5	4	4
290	19	2097	7.96	5	77.73	15.26	-0.221	1	4	2

***Abbreviations: length (LEN), molecular weight(MW), theoretical pI (pI), primary amino acid composition(AA), Instability index(II), aliphatic index and grand average of hydropathicity(GRAVY), charge (CH),known function(KF), predicted function(PF).**

Construction of discriminant functions

Eight variables (i.e. length(x1), molecular weight(x2), pI(x3), primary amino acid composition (x4), instability index(x5), aliphatic index(x6), aliphatic index and grand average of hydropathicity(x7) and charge(x8)) had been entered,

when F value were taken 0.02 and 0.01 as entry and removal criterion. After step discriminant analysis, a set of Fisher's function coefficients were generated and were constructed into four discriminant functions (1)-(4). y1, y2, y3, y4 refers to antibacterial, antifungal, anticancer and antiviral activities. In turn, every entry was re-classified a prediction activity using these four discriminant functions (TABLE 1). All predicted results were sum into TABLE 2.

$$y1=52.875-0.0026x1+0.0002x2+11.622x3+0.2376x4+0.1029x5+0.0669x6-0.891x7-3.7329x8 \quad (1)$$

$$y2=41.907+0.0833x1-0.0003x2+10.798x3+0.1174x4+0.0995x5+0.0162x6-0.457x7-3.5604x8 \quad (2)$$

$$y3=40.405+0.1108x1-0.0006x2+9.4126x3+0.2043x4+0.1201x5+0.0963x6-2.075x7-3.5649x8 \quad (3)$$

$$y4=45.627-0.1515x1+0.0018x2+10.402x3+0.1498x4+0.1215x5+0.0774x6-2.1489x7-3.6963x8 \quad (4)$$

The results showed 69.7% of selected original groped cases correctly classified (TABLE 2). The discriminant analysis correctly classified 31 of the 43 antibacterial peptides (72.1%), 29 of the 36 antifungal peptides (80.6%), 6 of the 11 anticancer peptides (54.5%), 17 of the 29 antiviral peptides (58.6%). The correct rate was obviously higher than prior probabilities 36.1%, 30.3%, 9.2%, and 24.4%. These means the model had ability to recognize and classify correct function groups.

TABLE 2 : Classification results for the discriminant analysis

Activity		Predicted Group Membership*				Total
		B	F	C	T	
Original count (%)	B	31(72.1)	10(23.3)	1(2.3)	1(2.3)	43(100)
	F	5(13.9)	29(80.6)	1(2.8)	1(2.8)	36(100)
	C	2(18.2)	1(9.1)	6(54.5)	2(18.2)	11(100)
	T	4(13.8)	6(20.7)	2(6.9)	17(58.6)	29(100)

*** B, antibacterial activity. F, antifungal activity. C, anticancer (tumor) activity. T, antiviral activity**

Activity prediction for new AMPs using discriminant functions

For a new peptide, there is 25% probability to discriminate activity correctly. If activity prediction is carried out using discriminant functions, the validity of classification will be improved greatly. Take 18 new antimicrobial peptides as the random samples. They were predicted activities according to their physicochemical features computed with ProtParam software (TABLE 3). Meanwhile, to verify the validity of the prediction model, these sequences were predicted with the method of similar alignment. The results shown 77.8% (14 in 18) was consistent with two different methods. This meant the model consisting of discrimination functions had some kind of ability of prediction.

TABLE 3 : Prediction for unknown activity peptides

ID	Independents*								PF
	LEN	MW	pI	AA	II	AI	GRAVY	CH	
10	38	4278	8.98	5	42.89	61.58	-0.042	4	2
37	24	2294.3	10.3	20	-21.07	106.25	0.358	3	1
95	39	4395	10	13	57.7	27.44	-1.026	3	2
108	27	2925.2	8.53	16	19.59	43.33	-0.815	2	2
127	34	3635	7.78	5	30.82	25.88	-0.706	1	2
163	20	2236.6	5.96	12	66.76	78	0.37	0	3
188	17	1769.1	8.6	1	42.16	143.53	0.912	1	1
226	40	4257.9	11.07	20	17.1	90.5	-0.333	5	1
294	23	2695.1	9.75	4	-6.77	101.3	0.752	1	1
299	25	2585.1	7.02	2	29.7	171.2	1.188	0	3
391	33	3152.6	10	20	12.42	92.42	0.191	4	1
396	48	5196	8.52	5	36.09	58.96	0.008	3	2
454	43	4453.1	8.68	5	36.01	77.21	0.37	3	2
566	10	1171.2	4.03	18	0.51	49	-1.18	-2	3
673	113	11096.1	3.99	20	28.72	70.88	0.211	-6	3
676	43	4312.1	7.92	20	27.38	81.86	0.319	1	1
752	98	10189.4	4.43	20	30.3	87.55	0.098	-7	3
771	47	5404.1	8.52	5	12.59	33.19	-1	3	2

***Abbreviations descriptions were same with these in table 1**

CONCLUSIONS

In this study, stepwise discriminant analysis was utilized to derive four quantitative functions for classifying potential activities for antimicrobial peptides. The model correctly classified 70% of the known activity peptides in the database, and predicted 77.8% of the new unknown activity peptides. The model constituted of four discriminant functions obtained in this study maybe allowed the design of highly active shortened AMPs and may be generally useful in the development of this type of peptides as anti-infective agents.

ACKNOWLEDGEMENT

We thank Dr Zhang (Sun Yat-sen University) for the help of statistical analysis. The research work was supported by Natural Science Foundation of China Grants (31200705) and Fundamental Research Funds for the Central Universities (2014ZZ0051).To whom correspondence may be addressed: School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China. Tel.: 86-20-39380608; E-mail: liqh@scut.edu.cn.

REFERENCES

[1] M.Zasloff; Antimicrobial peptides of multicellular organisms, Nature, **415(6870)**, 389-95 **(2002)**.

[2] R.E.Hancock, R.Lehrer; Cationic peptides: a new source of antibiotics, Trends Biotechnol, **16(2)**, 82-8 **(1998)**.

[3] A.Tossi, L.Sandri, A.Giangaspero; Amphipathic, alpha-helical antimicrobial peptides, Biopolymers, **55(1)**, 4-30 **(2000)**.

[4] H.G.Boman; Antibacterial peptides: basic facts and emerging concepts, J Intern Med, **254(3)**, 197-215 **(2003)**.

[5] R.E.Hancock, M.G.Scott; The role of antimicrobial peptides in animal defenses, Proc Natl Acad Sci U S A, **97(16)**, 8856-61 **(2000)**.

[6] S.Meng, H.Xu, F.Wang; Research advances of antimicrobial peptides and applications in food industry and agriculture, Curr Protein Pept Sci, **11(4)**, 264-73 **(2010)**.

[7] G.Wu, H.Wu, X.Fan, et al; Selective toxicity of antimicrobial peptide S-thanatin on bacteria, Peptides, **31(9)**, 1669-73 **(2010)**.

[8] M.M.Javadpour, M.M.Juban, W.C.Lo, et al; De novo antimicrobial peptides with low mammalian cell toxicity, J Med Chem, **39(16)**, 3107-13 **(1996)**.

[9] S.P.Liu, L.Zhou, R.Lakshminarayanan, R.W.Beuerman; Multivalent Antimicrobial Peptides as Therapeutics: Design Principles and Structural Diversities, Int J Pept Res Ther, **16(3)**, 199-213 **(2010)**.

[10] R.M.Martins, M.L.Sforca, R.Amino, et al; Lytic activity and structural differences of amphipathic peptides derived from trialysin, Biochemistry, **45(6)**, 1765-74 **(2006)**.

[11] I.Zelezetsky, A.Tossi; Alpha-helical antimicrobial peptides--using a sequence template to guide structure-activity relationship studies, Biochim Biophys Acta, **1758(9)**, 1436-49 **(2006)**.

[12] E.F.Palermo, K.Kuroda; Structural determinants of antimicrobial activity in polymers which mimic host defense peptides, Appl Microbiol Biotechnol, **87(5)**, 1605-15 **(2010)**.

[13] Q.Wang, G.Hong, G.R.Johnson, R.Pachter, M.S.Cheung; Biophysical Properties of Membrane-Active Peptides Based on Micelle Modeling: A Case Study of Cell-Penetrating and Antimicrobial Peptides, J Phys Chem B, **114(43)**, 13726-35 **(2010)**.

[14] D.S.Bolintineanu, Y.N.Kaznessis; Computational studies of protegrin antimicrobial peptides: a review, Peptides, **32(1)**, 188-201 **(2010)**.

[15] A.Giangaspero, L.Sandri, A.Tossi; Amphipathic alpha helical antimicrobial peptides, Eur J Biochem, **268(21)**, 5589-600 **(2001)**.