QSAR Modeling and Design of Cationic Antimicrobial Peptides Based on Structural Properties of Amino Acids

Yuanqiang Wang[1,2,3], Yuan Ding[2], Haixia Wen[2], Yong Lin[2], Yong Hu[2], Ya Zhang[2], Qingyou Xia[*,1,3] and Zhihua Lin[*,2]

[1]*Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400716, P.R. China*

[2]*College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400054, P.R. China*

[3] *The State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400716, P.R. China*

Abstract: Drug resistance to existing antibiotics poses alarming threats to global public health, which inspires heightened interests in searching for new antibiotics, including antimicrobial peptides (AMPs). Accurate prediction of antibacterial activities of AMPs may expedite novel AMP design and reduce the costs and efforts involved in laboratory screening. In the present study, a novel quantitative prediction method of AMP was established by quantitative structure-activity relationship (QSAR) modeling based on the physicochemical properties of amino acids. The indices of these physicochemical properties were used to define AMP. The structural variables were optimized by stepwise regression (STR). Three series of AMPs from the QSAR model were constructed by multiple linear regressions (MLR). These QSAR models showed good performance in reliability and predictability. The normalized regression coefficients of the QSAR model and the contribution of amino acids at each position of AMP may determine the suitableness of a particular residue at any given position. QSAR models constructed by STR-MLR should prove to be useful tools in peptide design with respect to the calculation, explanation, good and reliable performance, and definition of physiochemical properties.

Keywords: Antimicrobial peptides (AMPs), descriptor, normalized regression coefficients (NRCs), peptide design, quantitative structural activity relationship (QSAR), structural characterization.

1. INTRODUCTION

The abuse of antibiotics in health care practice has generated a large number of bacterial strains to adapt or become drug-resistant, such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Staphylococcus epidermidis*, vancomycin-resistant *enterococci* (VRE), and ampicillin-resistant *Escherichia coli*. The rapid generation of antibiotics designed to treat disease is posing serious threats to human health [1-3]. The widespread resistance of bacterial pathogens to conventional antibiotics has prompted renewed interests in the use of natural microbial inhibitors, such as antimicrobial peptides (AMPs), as alternatives. With different antibacterial mechanisms, AMPs may have better potential as novel antimicrobial agents.

In nature, AMPs are important components of the immune system that present the first line of defense against a variety of pathogens that widely range from protozoa to vertebrates [4, 5]. AMPs are typically small, cationic, and amphiphilic molecules spanning several structural classes. Most of them can be categorized to one of the three substructural classes as follows [6]: (1) linear α-helical peptides free of cysteine residues, (2) peptides adopting a β-sheet globular structure stabilized by intra-molecular disulfide bridges, and (3) peptides with unusual bias in certain amino acids, such as histidine, glycine, proline, or tryptophan. Their similar structural patterns probably enable them to share common mechanisms of anti-bactericidal action. Although the precise mechanism of their action is not clear, it is accepted that they target the bacterial membrane. Cationic peptides may bind to negatively charged lipopolysaccharides (LPS) or lipid A (LA) of Gram-negative bacteria initially, and then lead to membrane permeation through the following steps [7-9]: (1) minor disruption of the phospholipid chain order and packing in the outer membrane, which is also called "self-promoted uptake"; (2) transmembrane channel formation *via* a "barrel-stave" or toroidal pore mechanism; and (3) membrane destruction *via* a carpet-like mechanism, which may ultimately result in the death of bacteria. An increasing number of studies suggest that the key to AMP antimicrobial activity is AMP aggregation on the bacterial membrane surface.

In the past decade, many databases have emerged and some predictive models have been built based on rapidly discovered AMPs. These databases include (1) databases containing structural and sequential information on naturally occurring AMPs, such as AMSDb [10], ANTIMIC [11], APD [12, 13], Peptaibol Database [14, 15], PenBase [16], Seebah Defensins knowledgebase [17], AMPer [18], CAMP [19], and PhytAMP [20]; (2) synthetic antibiotic peptides databases, such as SAPD [21] and Peptidomimetics [22]; and (3) secondary databases on natural antimicrobial peptides, such as RAPD [23] and BACTIBASE [24]. Some databases provide theoretical identification, quantitative prediction, or design tool of novel antimicrobial peptides. For example, the APD database can predict new peptides by: (1) templates derived from

*Address correspondence to these authors at the (Qingyou Xia) Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400716, P.R. China; Tel: 86(23) 68250099; Fax: 86(23) 68250099;
E-mail: xiaqy@swu.edu.cn and (Zhihua Lin) College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400050, P.R. China; Tel: 86(23)62563181; Fax: 86(23)62563181;
E-mail: zhlin@cqut.edu.cn

1875-5402/12 $58.00+.00

naturally occurring AMPs, (2) combined database-derived peptide motifs, or (3) percentages of each type of amino acid obtained from statistical analysis of antimicrobial peptides. The Anti-BP2 is a predictive tool constructed using the Support Vector Machine (SVM) based on APD, which has been utilized to predict and classify antibacterial peptides [23]. ANTIMIC and AMPer serve as classification tools based on HMMER software package for searching additional members of AMP groups not presented originally.

The quantitative prediction of AMPs is very important for the design or modification of peptides with higher antibacterial activity. However, most current predictive tools are based on sequential and structural features, such as AntiBP2 [25] and APD2. Fortunately, there are a few quantitative predictive models that provide more clues for designing novel AMPs. Cherkasov *et al*. [26] constructed a predictive model using an artificial neural network (ANN) based on 'inductive' descriptors for the series of synthetic cationic polypeptides. Taboureau *et al*. [27] built a QSAR model with high performance for novispirin AMPs based on 3D descriptors generated by GRID. Fjell *et al*. [28] established a high-throughput ANN model using physicochemical descriptors and successfully screened potential AMPs.

In the present study, 89 indices of amino acids from the AA index database were used to characterize 3 classes of AMPs (101 synthetic cationic polypeptides, surface-tethered cationic peptides, and novispirin AMPs). Stepwise regression (STR) was used to screen the structural variables and the QSAR models were constructed by multiple linear regressions (MLR). The contributions of the physicochemical parameters for any position of AMPs could guide peptide modification accordingly. All the models built showed good performance for the prediction of antimicrobial activity.

2. MATERIALS AND METHODS

2.1. Data Source

Three datasets of AMPs were retrieved from the literature containing 101 synthetic cationic polypeptides (CAMEL-s) with mean antibiotic potency for 24 G^- and G^+ bacteria [26], 74 novispirin AMPs with normalized growth inhibition [27], and 86 surface-tethered cationic peptides with percentage inhibition for luminescence [29]. The AMP sequences and corresponding antimicrobial activities are presented in the supplementary file. Each dataset was divided into two groups: the training set, which is used to construct a predictive model validated by the leave-one-out (LOO) validation, and the test set, which was used to validate the reliability of the model.

2.2. Characterization of AMPs

The antibacterial activities of peptides rely on their physicochemical properties, such as cationicity, amphiphilicity, spatial structure, and hydrophobicity. Thus, 89 structural parameters, including 58 hydrophobicities, 22 physiochemical properties, and 9 electrical properties, were selected for the characterization of the AMPs collected from the AA Index database (http://www.genome.jp/aaindex).

Peptide activity results from the contributions of all amino acids. The vector of structural features, which are constituted with that of every amino acid, reflects the structural information of peptides very well. For example, if the peptide consists of n amino acids and each amino acid has m parameters, the vector of parameter for every peptide has $n \times m$ elements, expressed as V_{jk}, where j (j=1…n) is the position of amino acid and k (k=1…m) is the order of parameter.

2.3. Method of Modeling

In practice, a good MLR requires the number of samples to be five times more than that of variables, and there is none or low correlation between variables. Several structural variables have been used to parameterize AMPs because they consist of at least 12 amino acids and every amino acid shows 89 physicochemical properties. Thus, STR was used to screen the critical non-collinear variables to construct QSAR models. To obtain suitable QSAR models represented with Q^2 of LOO and R^2 of test dataset, the variables were optimized by adjusting the threshold of introducing (F_1) and eliminating (F_2) to improve Q^2 and R^2. The modeling variables screened by STR have significant impact on antibacterial activities and could guide the design and modification of AMPs.

2.4. Contribution Analysis of Amino Acids

The normalized regression coefficients (NRCs) of the MLR model are the bases of calculating or predicting the activities of peptides because they reflect the contribution of amino acids at any position. If there is more than one variable selected to describe a particular position, the respective summation of positive and negative NRCs is an effective and direct method which shows comprehensive contribution for activity at the position. Similarly, the contribution of amino acids at any position could be calculated with the following formula:

$$C_{ij} = \sum_k NRC_{ik} \times para_k \quad (1)$$

where C_{ij} is the contribution of amino acid i (i=1…20) at position j (j=1…9); NRC_{jk} is the NRC of variable k (k is the screened variable and NRC is not zero) at position j; and $para_k$ is the normalized physicochemical properties of the amino acid. The contributions could facilitate the identification of dominant and deleterious amino acids at any position when NRCs are not zero, thus providing guidance for the design and modification of novel peptides efficiently.

Conservative residue analysis could be performed and visualized through a sequence logo [30], which shows how well residues are conserved and provides insight on alternative schemes for any position. Unfortunately, sequence logos are based on alignments and frequency calculations of multiple peptides. Thus, identifying a novel peptide with higher activity is difficult, and this has been a challenge in designing novel peptides. Therefore, conservative residue analysis was used in the current investigation as a supplementary method for the QSAR model to find the most suitable residues for promoting activity at positions when NRCs are zero. WebLogo is a free website for generating sequence logos [31].

3. RESULTS

3.1. CAMEL-s

The 101 CAMEL-s were split into training and testing datasets [26] to facilitate comparisons with other modeling methods in the literature. The samples of training and testing

datasets are presented in the supplementary file. Nine parameters were screened to build a QSAR model which showed good reliability and predictability (R^2=0.793, Q^2=0.751). The statistical parameters of the QSAR model and that of the literature are listed in Table **1**. The QSAR model (R^2=0.860) constructed by ANN in the literature showed high performance for predicting the AMPs, probably because ANN described nonlinear correlation between parameters and activity. However, the predictability of the STR-MLR model was not suitable for the testing dataset. There were two outliers (CAMEL9 and CAMEL46) that could be attributed to model discrepancies, non-precise characterization of experimental potency, and unsuited characterization of peptides. Fig. (**2A**) shows correlation between experimental and predictive powers, which indicates the predictability of the STR-MLR model, making it well-suited for the training dataset. Although the correlation of STR-MLR model showed no superiority to ANN model, it can be used to predict novel AMPs accurately, especially for those with high antibacterial activities.

3.2. Novispirin AMPs

The dataset of 74 novispirin AMPs were also split into training and testing sub-datasets. The reliability and predictability of the STR-MLR model (R^2=0.970, Q^2=0.872) was far more superior to that of the PLS model (R^2=0.730, Q^2=0.610) from the literature. The statistical parameters of the QSAR model are listed in Table **1**. The correlation plots (Fig. **2B**) between observed and predicted inhibitions indicate that the QSAR model has good reliability.

3.3. Surface-Tethered AMPs

All 86 surface-tethered AMPs were divided into two groups randomly. The STR-MLR model constructed with the training dataset showed good predictability (R^2=0.686, Q^2=0.639), and its reliability was well-suited for the testing dataset (R^2=0.663). The statistical parameters are listed in Table **1**. The correlation plots (Fig. **2C**) between observed and predicted inhibitions of luminescence showed that the model has good reliability and predictability and could be used to predict activity for novel AMPs accurately.

4. DISCUSSIONS

4.1. CAMEL-s

The sequence logo of CAMEL-s generated by WebLogo (Fig. **1A**) demonstrated that the sequence constituted by conservative residues possessed typical characteristics, such as cationicity and amphiphilicity. For example, the hydrophilic (KWK) and hydrophobic (VL) residues lie in the N- and C-terminals, respectively. The conservative sequence was composed of five amino acids with positive charges and lysine (K), which facilitated the binding of the peptide to the bacterial membrane.

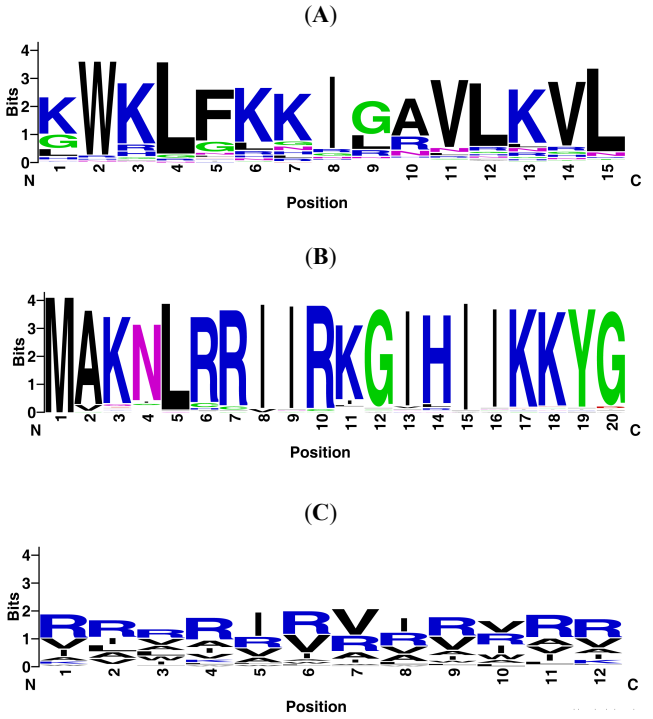

Fig. (1). Sequence logo of AMPs. (**A**) CAMEL-s, (**B**) Novispirin AMPs, (**C**) Surface-Tethered AMPs.

The contributions of amino acids were calculated by Eq. 1, and the dominant and deteriorative residues had been determined by maximal and minimal contribution values. NRCs, parameters, and dominant and deleterious amino acids are listed in Table **2**. The NRC values showed the contributions of amino acid with specific parameters, which can provide useful clues for AMP modification. For example, the NRC value (0.496) at position 13, characterized with 'partial specific volume', was greater than values at other positions. Therefore, the amino acid with larger 'partial specific volume' at position 13 could improve the activity of AMP. The calculation of amino acid contribution forecasted the significance of amino acids at a specific position. For example, arginine (R) found at position 6 increased activity, but alanine (A) at that same position showed the opposite effect.

Table 1. The Statistical Parameters of QSAR Models Comparing with Those from Literatures

Name	Method	*n*	Training					LOO		Testing		Ref.
			n	*m*	R^2	*F*	*sd*	*PRESS*	Q^2	*n*	R^2	
CAMEL-s	ANN	101	91	20	0.860	/	0.691	/	/	10	/	[26]
	STR-MLR	101	91	9	0.793	34.552	0.479	0.496	0.751	10	0.326	/
Novispirin AMPs	PLS	74	58	69	0.730	/	/	/	0.610	/	/	[27]
	STR-MLR	74	58	12	0.927	47.777	0.304	0.355	0.872	16	0.589	/
Surface-Tethere AMPs	STR-MLR	86	76	6	0.686	25.175	0.584	0.597	0.639	10	0.663	/

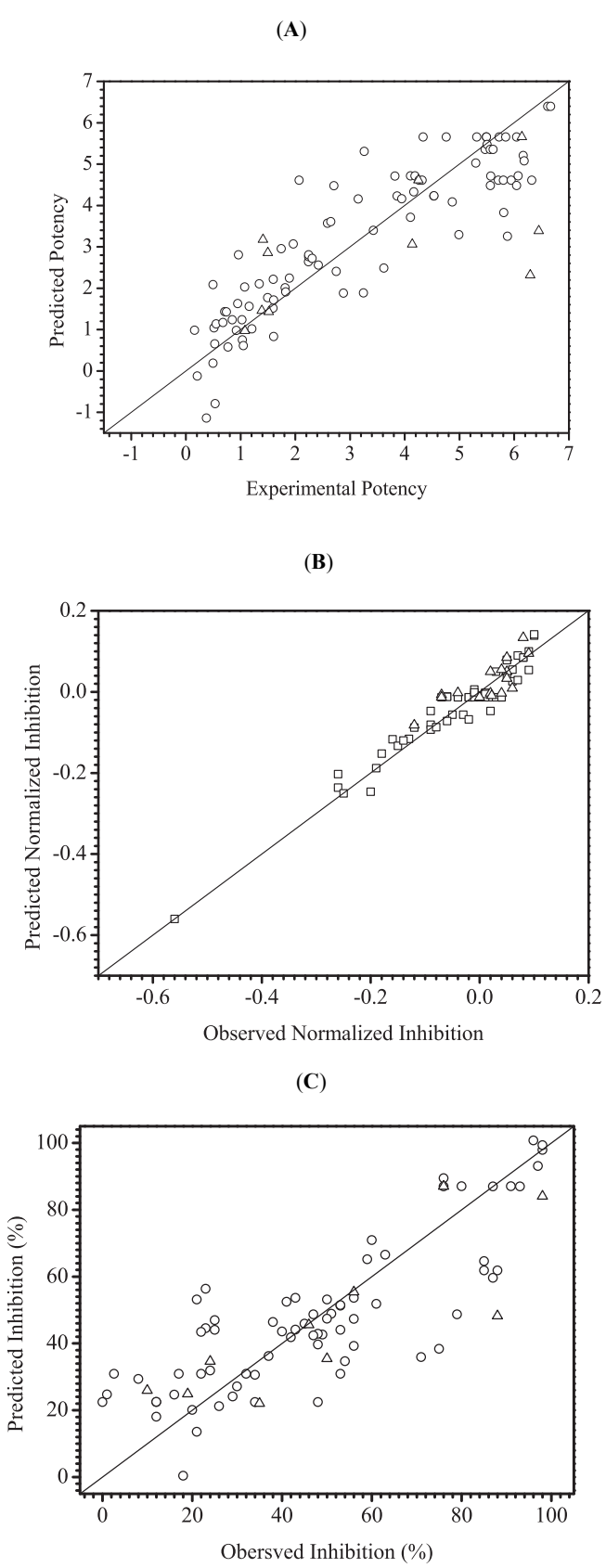

Fig. (2). Correlative plots of experimental and predictive activity (○ training dataset, Δ test dataset). (**A**) CAMEL-s, (**B**) Novispirin AMPs, (**C**) Surface-Tethered AMPs.

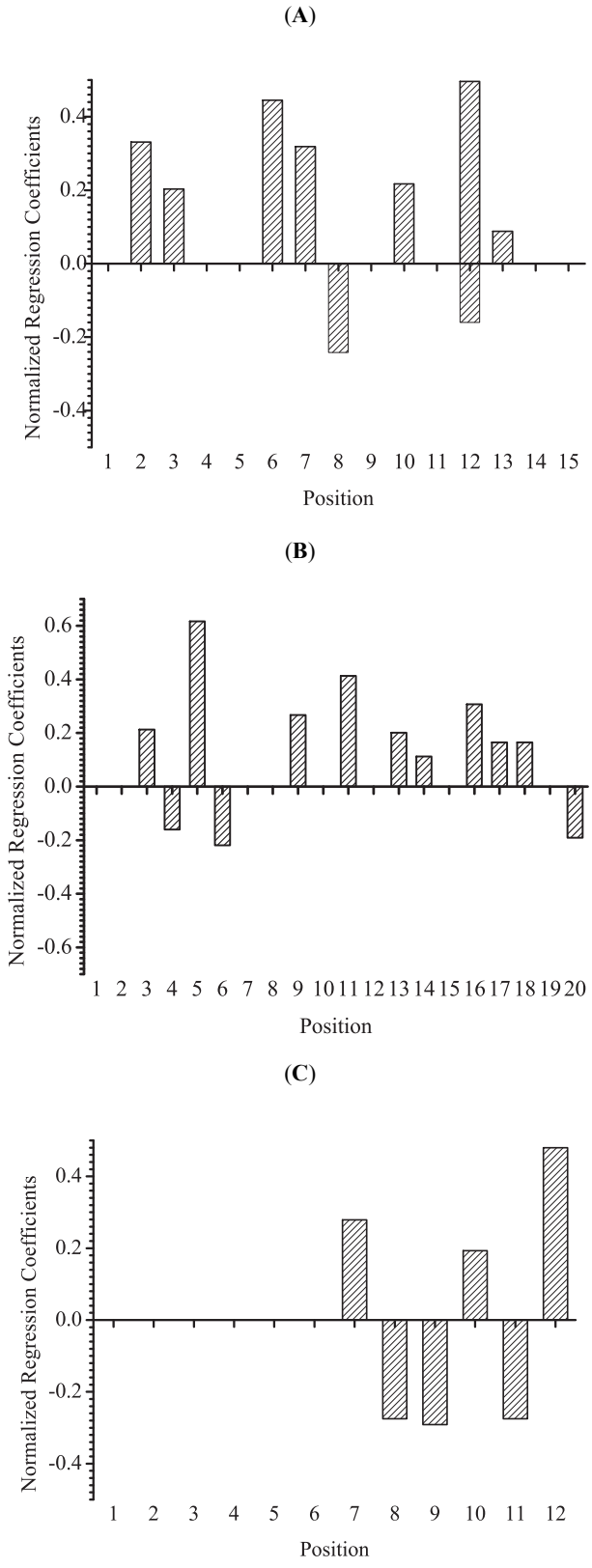

Fig. (3). Histogram of normalized regression coefficients. (**A**) CAMEL-s, (**B**) Novispirin AMPs, (**C**) Surface-Tethered AMPs.

Compared with the sequence logo, isoleucine (I) was not conservative at position 12 for the current sequence. However, when evaluated by NRCs, it had the maximal

Table 2. The Dominant and Deleterious Amino Acids, Parameters At Significant Positions for CAMEL-s

Position	Dominant	Deleterious	NRCs	Parameters
2	Cys (C)	Lys (K)	0.331	Average gain ratio in surrounding hydrophobicity
3	Asp (D)	Cys (C)	0.203	Hydration number
6	Arg (R)	Ala (A)	0.445	Polarity
7	Trp (W)	Glu (E)	0.319	Dependence of partition coefficient on ionic strength
8	Cys (C)	Asp (D)	-0.242	Hydration number
10	Arg (R)	Asp (D)	0.217	Isoelectric point
12	Trp (W)	Asn (N)	-0.160	Free energy change of aRi to aRh; Partition coefficient
13	Ile (I)	Asp (D)	0.496	Partial specific volume

Table 3. The Dominant and Deleterious Amino Acids, Parameters at Significant Positions for Novispirin AMPs

Position	Dominant	Deleterious	NRCs	Parameters
3	Arg (R)	Asp (D)	0.213	Isoelectric point
4	Gly (G)	Arg (R)	-0.160	Polarity
5	Cys (C)	Lys (K)	0.617	Transfer free energy
6	Gly (G)	Trp (W)	-0.219	Polarizability parameter
9	Trp (W)	Gly (G)	0.267	van der Waals parameter R0
11	Trp (W)	Glu (E)	0.413	Retention coefficient in HPLC, pH7.4
13	Trp (W)	Gly (G)	0.201	Mean area buried on transfer
14	Arg (R)	Asp (D)	0.112	Isoelectric point
16	Phe (F)	Asp (D)	0.307	Optimal matching hydrophobicity
17	Trp (W)	Glu (E)	0.165	Retention coefficient in HPLC, pH7.4
18	Arg (R)	Asp (D)	0.165	Isoelectric point
20	Trp (W)	Asp (D); Glu (E); Arg (R)	-0.191	Hydrophilicity value

Table 4. The Dominant and Deleterious Amino Acids, Parameters at Significant Positions for Surface-Tethered AMPs

Position	Dominant	Deleterious	NRCs	Parameters
7	Trp (W)	Gly (G)	0.279	Principal property value z2
8	Arg (R)	Phe (F)	-0.275	Free energy change of aRi to aRh
9	Gln (Q)	Ile (I)	-0.291	Hydrophobicity
10	Trp (W)	Gly (G)	0.193	Principal property value z2
11	Lys (K)	Ala (A)	-0.275	Average interactions per side chain atom
12	Asp (D)	Cys (C)	0.480	Hydration number

contribution to activity at this position (Fig. **3A**). Thus, the STR-MLR model with high performance not only predicted the antibacterial activity but also identified the most suitable amino acid in sequence for the design of novel AMPs. For positions that had no contributions (NRC is zero) to activity, the conservative amino acids were the favored choices for design and modification of AMP. Combination of QSAR models and conservative residues analysis may facilitate the development of ideal AMPs.

4.2. Novispirin AMPs

The conservative residue analyses of novispirin AMPs are shown in Fig. (**1B**). The sequence composed of conservative amino acids had eight positive charges, which enhanced the binding of peptide to bacterial membrane. Malmstel *et al.* [32] reported that the hydrophobic C-terminal end-tagging of AMPs with hydrophobic amino acid stretches is an interesting way to improve bactericidal potency of AMPs. Fig. (**1B**) shows the conservative amino acids are hydrophilic residues (YG), which

is not favorable for promoting activity. The sequence information alone was not enough for the design of novel AMP, and the quantitative model can provide more useful clues. Thus, the method integrating conservative sequence information and a quantitative model can improve the efficiency of design or modification.

The histogram of NRCs (Fig. **3B**) shows that some positions (1, 2, 7, 8, 10, 12, 15, and 19) had no contribution to antibacterial activity. NRCs, parameters, and dominant and deleterious amino acids determined by contribution are listed in Table **3**. The NRC indicating a 'free energy transfer' of amino acids at position 5 was important for activity. The amino acid with larger 'free energy transfer' at this position improved the activity to a greater extent. The contribution of amino acid computations showed cysteine (C) at position 5 was a dominant residue, whereas lysine (K) at the same site did not promote the activity. The sequence composed of amino acids with maximal contribution suggested that it met the required cationic and amphiphilic molecules. However, the positively charged sequence was relatively small, which may weaken the binding between AMP and bacterial membrane. Therefore, according to conservative residue analysis, the cationic amino acids (H) at positions 7 and 10 should be the preferred choice for promoting antibacterial activity.

4.3. Surface-Tethered AMPs

The sequence logo (Fig. **1C**) of the surface-tethered AMPs indicated that these AMPs showed subtle conservation, which made it difficult to design an ideal peptide with higher activity. The current AMP sequence is composed of numerous cationic amino acids, such as arginine (R). Hydrophobic residues are suggested for novel AMPs.

The contributions of each position are shown in Fig. (**3C**), which indicates the positions from 1 to 6 with no contribution for activity. The dominant and deleterious amino acids, NRCs, and parameters are listed in Table **4**. NRC showed that a larger 'hydration number' of amino acid at position 12 and 'principal property value z2' of amino acid at position 7 or 10 were favorable for increasing activity. Thus, in terms of contribution, aspartic acid (D), tryptophan (W), and tryptophan (W) were suggested as dominant amino acids for positions 12, 7, and 10, respectively. NRC could screen the important structural parameters and facilitate the finding of dominant amino acids. For example, at position 9, NRC showed that the 'hydrophobicity' of amino acids had a negative influence on activity. Hydrophilic residues presented here, such as glutamine (Q), might increase the activity of AMPs. The sequence with favorable amino acids as determined by their contributions was rich in hydrophilic residues. Thus, the hydrophobic and cationic residues were preferred for positions 1 to 6 of the novel AMP.

CONCLUSION

Eighty nine indices were selected from the AA Index database to characterize the AMPs. The quantitative models were constructed by STR-MLR for three series of AMPs. These quantitative models showed good reliability and predictability, and may be used to predict a novel peptide quantitatively. The contribution of amino acid equation from the quantitative model may also be used to determine the important positions of the AMPs and their suitableness for antibacterial affinity. Moreover, they may be used to guide the design and modification of novel AMPs. For the positions with no contribution for activity (NRC is zero), conservative residue analysis may facilitate the identification of the preferred amino acids. Therefore, combining the QSAR model with conservative residue analysis could result in finding the most suitable amino acids for novel AMPs and in predicting their antibacterial activity accurately.

ACKNOWLEDGEMENTS

Funding: This study was supported by grant from National Natural Science Foundation of China (No. 81171508, 60873103), Key Project of Natural Science Foundation of China (No. 30830090), Innovative New Drugs National Program (No. 2009ZX09503-005), the Program for Innovative Research Team in University of ChongQing (KJTD201039), Scientific Research Foundation of Education Committee of Chongqing (No. KJ100801), and Natural Science Foundation Project of CQ CSTC (No. CSTC, 2010BB5306).

SUPPLEMENTARY FILE

Parameter: the selected 89 properties of amino acids;

Parameter value: the value of selected parameters used to characterize amino acids;

CAMEL-s: training and test dataset, statistical parameters and NRC for CAMEL-s;

Novispirin AMPs: training and test dataset, statistical parameters and NRC for Novispirin AMPs;

Surface-Tethered AMPs: training and test dataset, statistical parameters and NRC for Surface-Tethered AMPs.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Hancock, R.E. Cationic peptides: effectors in innate immunity and novel antimicrobials. *Lancet Infect Dis.,* **2001**, *1*(3), 156-164.

[2] Levy, S.B.; Marshall, B. Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.,* **2004**, *10*(12 Suppl), S122-129.

[3] Choi, S.; Isaacs, A.; Clements, D.; Liu, D.; Kim, H.; Scott, R.W.; Winkler, J.D.; DeGrado, W.F. De novo design and *in vivo* activity of conformationally restrained antimicrobial arylamide foldamers. *Proc. Natl. Acad. Sci.,* **2009**, *106*(17), 6968-6973.

[4] Bucki, R.; Pastore, J.J.; Randhawa, P.; Vegners, R.; Weiner, D.J.; Janmey, P.A. Antibacterial activities of rhodamine B-conjugated gelsolin-derived peptides compared to those of the antimicrobial peptides cathelicidin LL37, magainin II, and melittin. *Antimicrob. Agents Chemother.,* **2004**, 48(5), 1526-1533.

[5] Hancock, R.E.; Diamond, G. The role of cationic antimicrobial peptides in innate host defences. *Trends Microbiol.,* **2000**, 8(9). 402-410.

[6] Wiesner, J.; Vilcinskas, A. Antimicrobial peptides: the ancient arm of the human immune system. *Virulence,* **2010**, 1(5): 440-464.

[7] Oren, Z.; Shai, Y. Mode of action of linear amphipathic alpha-helical antimicrobial peptides. *Biopolymers,* **1998**, 47(6), 451-463.

[8] Hwang, P.M.; Vogel, H.J. Structure-function relationships of antimicrobial peptides. *Biochem. Cell Biol.,* **1998**, 76(2-3), 235-246.

[9] Frecer, V.; Ho, B.; Ding, J.L. De novo design of potent antimicrobial peptides. *Antimicrob. Agents Chemother.,* **2004**, 48(9), 3349-3357.

[10] Tossi, A.; Sandri, L. Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. *Curr. Pharm. Des.,* **2002**, 8(9), 743-761.

[11] Brahmachary, M.; Krishnan, S.P.; Koh, J.L.; Khan, A.M.; Seah, S.H.; Tan, T.W; Brusic, V.; Bajic, V.B. ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Res.,* **2004**, *32*(Database issue), D586-589.

[12] Wang, Z.; Wang, G. APD: the antimicrobial peptide database. *Nucleic Acids Res.,* **2004**, *32*(Database issue), D590-D592.

[13] Wang, G.; Li, X.; Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.,* **2009**, *37*(Database), D933-937.

[14] Whitmore, L.; Wallace, B.A. The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res.,* **2004**, *32*(Database issue), D593-594.

[15] Whitmore, L.; Chugh, J.K.; Snook, C.F.; Wallace, BA. The peptaibol database: a sequence and structure resource. *J. Pept. Sci.,* **2003**, *9*(11-12), 663-665.

[16] Gueguen, Y.; Garnier, J.; Robert, L.; Lefranc, M.P.; Mougenot, I.; de Lorgeril, J.; Janech, M.; Gross, P.S.; Warr, G.W.; Cuthbertson, B.; Barracco, M.A.; Bulet, P.; Aumelas, A.; Yang, Y.; Bo, D.; Xiang, J.; Tassanakajon, A.; Piquemal, D.; Bachère, E. PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. *Dev. Comp. Immunol.,* **2006**, *30*(3), 283-288.

[17] Seebah S, Suresh A, Zhuo S, Choong YH, Chua H, Chuon D, Beuerman R, Verma C. Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res* 2007; *35*(Database issue): D265-268.

[18] Fjell, C.D.; Hancock, R.E; Cherkasov, A. AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics,* **2007**, *23*(9), 1148-55.

[19] 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*(Database): D774-D780.

[20] Hammami R, Ben Hamida J, Vergoten G, Fliss I: PhytAMP: a database dedicated to antimicrobial plant peptides. Nucleic Acids Res 2009, *37*(Database), D963-968.

[21] Wade, D.; Englund, J. Synthetic antibiotic peptides database. *Protein Pept. Lett.,* **2002**, 9(1), 53-57.

[22] Vagner, J.; Qu, H.; Hruby, V.J. Peptidomimetics, a synthetic tool of drug discovery. *Curr. Opin. Chem. Biol.,* **2008**, *12*(3), 292-296.

[23] Li, Y.; Chen, Z. RAPD: a database of recombinantly-produced antimicrobial peptides. *FEMS Microbiol. Lett.,* **2008**, *289*(2), 126-129.

[24] Hammami, R.; Zouhir, A.; Le Lay, C.; Ben Hamida, J.; Fliss, I. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol.,* **2010**, *10*, 22.

[25] Lata, S.; Mishra, N.K.; Raghava, G.P. AntiBP2: improved version of antibacterial peptide prediction. *BMC Bioinformatics,* **2010**; *11*(Suppl 1): S19.

[26] Cherkasov, A.; Jankovic, B. Application of 'inductive' QSAR descriptors for quantification of antibacterial activity of cationic polypeptides. *Molecules,* **2004**, *9*(12), 1034-1052.

[27] Taboureau, O.; Olsen, O.H.; Nielsen, J.D.; Raventos, D.; Mygind, P.H.; Kristensen, H.H. Design of novispirin antimicrobial peptides by quantitative structure-activity relationship. *Chem. Biol. Drug Des.,* **2006**, *68*(1), 48-57.

[28] Fjell, C.D.; Jenssen, H.; Hilpert, K.; Cheung, W.A.; Panté, N.; Hancock, R.E.; Cherkasov, A. Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.,* **2009**, *52*(7), 2006-2015.

[29] Hilpert, K.; Elliott, M.; Jenssen, H.; Kindrachuk, J.; Fjell, C.D.; Körner, J.; Winkler, D.F.; Weaver, L.L.; Henklein, P.; Ulrich, A.S.; Chiang, S.H.; Farmer, S.W.; Pante, N.; Volkmer, R.; Hancock, R.E. Screening and characterization of surface-tethered cationic peptides for antimicrobial activity. *Chem. Biol.,* **2009**, *16*(1), 58-69.

[30] Schneider, T.D.; Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.,* **1990**, *18*(20), 6097-6100.

[31] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* **2004**; *14*(6): 1188-90.

[32] Malmsten, M.; Kasetty, G.; Pasupuleti, M.; Alenfall, J.; Schmidtchen, A. Highly selective end-tagged antimicrobial peptides derived from PRELP. *PLoS ONE,* **2011**, *6*(1), e16400.

Received: October 17, 2011 **Revised: November 3, 2011** **Accepted: December 7, 2011**