JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

# Predicting the Minimal Inhibitory Concentration for Antimicrobial Peptides with Rana-Box Domain

Mara Kozić,[†] Damir Vukičević,[‡] Juraj Simunić,[§] Tomislav Rončević,[‡] Nikolinka Antcheva,[⊥] Alessandro Tossi,[⊥] and Davor Juretić*[,‡]

[†]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, U.K.

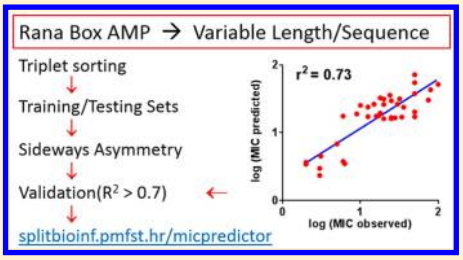
[‡]Faculty of Science, University of Split, 21000 Split, Croatia

[§]Mediterranean Institute for Life Sciences, 21000 Split, Croatia

[⊥]Department of Life Sciences, University of Trieste, 34127 Trieste, Italy

Supporting Information

**ABSTRACT:** The global spreading of multidrug resistance has motivated the search for new antibiotic classes including different types of antimicrobial peptides (AMPs). Computational methods for predicting activity in terms of the minimal inhibitory concentration (MIC) of AMPs can facilitate "*in silico*" design and reduce the cost of synthesis and testing. We have used an original method for separating training and test data sets, both of which contain the sequences and measured MIC values of non-homologous anuran peptides having the Rana-box disulfide motif at their C-terminus. Using a more flexible profiling methodology (sideways asymmetry moment, SAM) than the standard hydrophobic moment, we have developed a two-descriptor model to predict the bacteriostatic activity of Rana-box peptides against Gram-negative bacteria—the first multilinear quantitative structure−activity relationship model capable of predicting MIC values for AMPs of widely different lengths and low identity using such a small number of descriptors. Maximal values for SAMs, as defined and calculated in our method, furthermore offer new structural insight into how different segments of a peptide contribute to its bacteriostatic activity, and this work lays the foundations for the design of active artificial AMPs with this type of disulfide bridge.




## INTRODUCTION

With the worldwide rise in bacterial resistance to antibiotics,[1−3] the search for alternatives has become of utmost importance. Antimicrobial peptides (AMPs) are used by almost all organisms as part of their host defense systems to combat infections.[4−6] They are ancient innate weapons which our and other species employ as a first line of defense to fend off microbes, in tight cooperation with other sophisticated cellular or humoral effectors of immunity. Throughout their evolution, they have avoided becoming obsolete by inducing bacterial resistance, coevolving with microbes in specific niches. Under "*in vitro*" testing conditions, AMPs rapidly and efficiently kill a broad range of microbial species. They also have a remarkable potential for targeted rational modifications, allowing for the creation of a very large number of scaffolds for new classes of antibiotics. Despite this, only a limited number of AMPs have been examined in clinical settings.[7,8] A couple of bacteriocins entered into clinical practice almost 60 years ago, but they are used cautiously due to toxic side effects[9] and the desire to avoid induction of resistance mechanisms in Gram-negative pathogens.[10] The search space for novel AMPs with the desired properties is so vast that dedicated computational methods must be found for predicting their antimicrobial activity and specificity/selectivity.[11] Furthermore, modifications to natural AMPs, required to counter well-known problems with their therapeutic use, such as inadequate toxicology profiles, host cell toxicity, degradation by proteases, rapid turnover, suboptimal bactericidal activity, and salt sensitivity, are further increasing this search space.[12]

Initially, these modifications availed of template-based or residue-based modeling methods, so that synthetic analogues of natural peptides were designed with substitutions introduced according to some expert rules.[8,13,14] Peptide modeling using molecular descriptors and the building of quantitative structure−activity relationship (QSAR) models was the next logical step, but new levels of complexity were required with respect to the widely practiced drug design for small molecules.[15] Predicting the minimal inhibitory concentration (MIC, one of the most commonly used parameters to measure activity) of AMPs, when only their sequences are used as input, can easily run into overtraining problems. This is connected with the use of too many descriptors in the predictive models, combined with the generally modest size of validation data sets.[16] A compromise is therefore required between the number of descriptors used in the QSAR model and the number of peptide sequences available to compose training and validation data sets. Such a compromise usually involves using homologous peptides of the same or similar lengths, derived from a common precursor, such as magainin, lactoferricin,

**Received:** March 25, 2015
**Published:** September 2, 2015





DOI: 10.1021/acs.jcim.5b00161
*J. Chem. Inf. Model.* 2015, 55, 2275−2287

protegrin, bactenecin, brevinin, indolicidin, cecropin, and cathelicidins, among others. The "leave one out" validation procedure (a.k.a. jack-knife method)[17] was another commonly used option, removing the need to gather sequences for the test set. The drawback of such QSAR models is the resulting high dependence of specificity on the parent peptide used, and consequently a high likelihood of prediction failure for sequences with a low identity and/or different length distribution compared to the parent one.

Various nonlinear techniques, such as artificial intelligence methods, are capable of connecting sequences with antimicrobial activity even in the case of non-homologous peptides,[18] but they require many parameters to be used during the training process, so that overtraining can be a problem. In addition, these models are rather opaque, and it is not clear what (if any) general structure−activity relationships have been recognized by using these programs, or which amino acid attributes are important for a peptide's antimicrobial activity.[19]

In attempts to find structure−activity relationships, linear QSAR models are the easiest to use for selecting the most relevant descriptors, for interpreting results, and for examining the underlying mechanisms. Furthermore, being restricted to primary structures as the only input removes the need for predicting the three-dimensional (3D) peptide structure, and the consequent need to verify predictions by means of NMR experiments. This being said, it is however important to extract information as general as possible from the peptides contained within a training set. In practice, this means facing the challenge of constructing QSAR models for a set of non-homologous peptides, which in addition are of different lengths.

We have decided to explore sequence−activity relationships for a set of AMPs that are likely to have a common evolutionary origin, although today they are widely divergent with respect to pairwise identity, sequence length, and host species. This is the case for anuran AMPs bearing the "Rana-box" motif, a characteristic cyclic structure, at their C-terminus.[20−22] To the best of our knowledge, it is not known why this disulfide-bridged structure is so widespread among anuran AMPs. Some reports indicate that the deletion of the carboxy-terminal cysteine impairs antibiotic activity of Rana-box peptides,[23] while others dismiss its importance, as activity is unaffected when cysteine residues are replaced with serine to eliminate the cyclic structure.[24,25] The role of the Rana-box in pore formations has only rarely been explored.[26] In agreement with previous reports, the reduction of the disulfide bond did not alter peptide antimicrobial activity, while omitting the Rana-box does result in a decrease of ionophoric and antimicrobial activity.[26,27]

In the present paper we have assumed *ab initio* that the Rana-box loop structure is as important for some aspects of the peptides' activity as the preceding linear part of the sequence, which may be mostly or partially in a helical conformation when associated with biological membranes. Furthermore, we assumed that sideways asymmetry of the secondary structure (amphipathicity in a generally used sense, quantifiable as a hydrophobic moment) can promote a deeper burial of the peptide and favor destabilization of the membrane bilayer structure. In our QSAR method, we have employed as general an approach as possible to the structure−activity problem of Rana-box AMPs, so that we considered all 544 amino acid attributes scales listed in the AAindex database [http://www.genome.jp/aaindex/] of Kawashima et al.[28] but limited the collection of measured MIC values to *Escherichia coli* strains.

## MATERIALS AND METHODS

**Selection of Training and Test Sets.** Anuran AMPs, with disulfide-linked Rana-box at their C-terminal, were selected as an ordered set of peptides $p_1, ..., p_n$. Our goal was to obtain non-homologous training and test sets, with reported experimentally measured bacteriostatic activity against *E. coli*. All of the MIC values collected from the literature had been obtained using microdilution method, as far as possible with $5 \times 10^5$ CFU/mL inocula by incubating *E. coli* (mostly the reference strain ATCC25726 or ATCC25922) overnight at 37 °C with different peptide concentrations. In rare cases, MIC obtained using $2.5 \times 10^5$/mL CFU inocula were also allowed.

Dissimilarity or taxonomic distance between sequences was calculated using the MEGA software for finding pairwise distances (p-distances).[29] All sequences were aligned using the MUSCLE algorithm and the $p$-distances between each pair of peptides were computed using the pairwise deletion method.[30] The output was saved in the form of a lower-left matrix, and the identity (similarity) for each pair was defined as the $(1 - p)$-distance value. The similarity matrix $\boldsymbol{M}$ containing all $(1 - p)$-distance values was used as numerical input for subsequent calculations.

Peptides were divided in training and test sets with the following conditions:

(1) no two peptides in the training set have a greater similarity than a prescribed value (we used $t_1 = 0.9$), and
(2) no two peptides in the test set have a greater similarity than a prescribed value (again $t_2 = 0.9$).

We used a greedy algorithm that found the most similar triplets placing two of its elements in the training set and the third in the test set. We started with empty training and test sets and iterated the following procedure as long as a new triplet could be found. A triplet $(p_1, p_2, p_3)$ with the largest value of similarity ($sym$) was found:

$$sym(p_1, p_2) + sym(p_1, p_3) + sym(p_2, p_3)$$

such that $p_1$ and $p_2$ can be stored in the training set without violating condition (1) and that $p_3$ can be stored in test set without violating condition (2). Moreover, we applied the following two additional criteria:

(i) if $p_1$ and $p_3$ can be interchanged in the training and test sets without violating conditions (1) and (2), then it must hold that $sym(p_1, p_2) \leq sym(p_3, p_2)$, and
(ii) if $p_2$ and $p_3$ can be interchanged in the training and test sets without violating conditions (1) and (2), then it must hold that $sym(p_1, p_2) \leq sym(p_1, p_3)$.

These two requirements add more diversity to the training set. In the case of a tie in the previous maximization, the algorithm takes the triplet with the smallest indices according to lexicographical order of peptide indices in the database. Let us illustrate this by one example, suppose that triplets $(p_7, p_{11}, p_{43})$ and $(p_7, p_{14}, p_{21})$ have the same value. Then, we choose the first one, since (7, 11, 43) is smaller than (7, 14, 21) according to lexicographical order.

**Sideways Asymmetry Moment (SAM).** In order to calculate the sideways asymmetry moment (SAM) as a simple extension of Eisenberg's hydrophobic moment (HM),[31,32] the 544 scales in the AAindex database of amino acid attributes were used.[28] Scales were normalized in such a way that the 20-tuple vector obtained for each scale has length one; i.e., it holds that


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

$$n_X = s_X / \sqrt{\sum_{T \in A} s_T^{\,2}}$$

where $A$ is the set of all 20 amino acids and $s_X$ is the original index value (attribute) for each amino acid (vector normalization). Each normalized scale is proportional to the original one, so it is also perfectly correlated with it. The SAM definition follows that of Eisenberg's HM:

$$\langle \mu H \rangle = \frac{1}{N} \left\langle \left[\sum_{n=1}^{N} H_n \sin(n\delta)\right]^2 + \left[\sum_{n=1}^{N} H_n \cos(n\delta)\right]^2 \right\rangle^{1/2}$$

For the $n$th amino acid in the sequence containing $N$ residues, $n\delta$ is the angle separating side chains along the backbone ($\delta$ = 100° for an $\alpha$ helix), while $H_n$ is its hydrophobicity. Optimizations allowing for specific applications are described below.

**Definition of the Pre-Rana-Box Descriptor.** All the anuran peptides considered had a linear sequence of varying lengths prior to the Rana-box, segments of which adopt a helical conformation when in contact with biological membranes. Let us define the pre-Rana-box descriptor $P_{xy}$ that corresponds to attribute scales $x$ and $y$. Let $p$ be a peptide consisting of amino acids $a_1$, $a_2$, ..., $a_u$ $b_1$, $b_2$, ..., $b_v$, such that amino acids $b_1$, $b_2$, ..., $b_v$ constitute the Rana-box:

$$P_{xy}(a_1 a_2 \ldots a_u b_1 b_2 \ldots b_v) = \max \left\{ \begin{array}{l} \max\limits_{\substack{s=1,\ldots,u-4 \\ e=s+4,\ldots,u \\ \phi=\{80^0,81^0,\ldots,120^0\}}} \dfrac{\sqrt{\left(\sum_{i=s}^{e} x(a_i)\cdot\cos(i\cdot\phi)\right)^2 + \left(\sum_{i=s}^{e} x(a_i)\cdot\sin(i\cdot\phi)\right)^2}}{(e-s+1)^{1/3}}, \\ \max\limits_{\substack{s=1,\ldots,u-4 \\ e=s+4,\ldots,u \\ \phi=\{80^0,81^0,\ldots,120^0\}}} \dfrac{\sqrt{\left(\sum_{i=s}^{e} y(a_i)\cdot\cos(i\cdot\phi)\right)^2 + \left(\sum_{i=s}^{e} y(a_i)\cdot\sin(i\cdot\phi)\right)^2}}{(e-s+1)^{1/3}} \end{array} \right\}$$

where $x(a_i)$ and $y(a_i)$ are normalized amino acid attribute values, $\phi$ is the twist angle of successive side chains, $s$ is the start position of a maximally asymmetric segment and $e$ is the end position, while its length has the lower limit of five amino acids and the upper limit is the length $u$ of the whole pre-Rana-box sequence. Note that function $P$ does not depend on $b_1$, ..., $b_v$; hence, one could substitute $P_{xy}(a_1,a_2,\ldots,a_u)$ for $P_{xy}(a_1,a_2,\ldots,a_u b_1, b_2,\ldots,b_v)$. However, $P_{xy}(p)$ indicates the whole Rana-box peptide. Note, too, that in normalizing, we divide by the third root of an optimal segment length $e - s + 1$; the reasons for such a choice are explained below.

Let us first explain our motivation for this definition of a double maximization procedure. We start from the assumption that asymmetry in the profile of amino acid attributes along the peptide sequence is important for its antibacterial activity. When the peptide backbone in the segment under consideration is folded in accordance with a specific secondary structure, asymmetry arises if amino acid residues with a high value of certain attribute cluster mainly on one side of the segment. This is a simple extension of Eisenberg's HM concept that we term "sideways asymmetry moment", and we seek to find such SAM descriptors that can predict for the activity of the peptide in terms of its reported MIC (in this case, log(MIC)).

There is, however, good evidence in published research that there is more than one possible mechanism for the membrane attachment/insertion step leading to membrane disruption or pore formation,[33] underlying the antimicrobial activity. Furthermore, Rana-box motif peptides like gaegurins, esculentins, and ranatuerins are well known to contain deformations separating $\alpha$ helical segments in membrane-mimicking solvents.[34−36] In fact, bipartite helical structures with loops or kinks separating two amphipathic helices are in general quite common in helical AMPs.[37,38] By using two different scales of amino acid attributes we hoped to better capture segments acting via two different mechanisms, or having different functional roles for different helical segments in the pre-Rana-box domain. Considering that a peptide's ability to penetrate into the membrane should correlate more to the better of these two mechanisms, or to the more important of two helical segments, we used the maximum of the two values obtained.

Given that calculation of $x$ and $y$ is completely analogous, we can limit our explanation to the calculation for $x$. We assume that there is a most active part of the pre-Rana-box sequence $a_1$, $a_2$,...,$a_u$ starting from residue $s$ and ending in residue $e$ (i.e., the subsequence $a_s,a_{s+1},\ldots,a_e$) that is responsible for the peptide's activity. Moreover, we assume to begin with that this sequence has a conformation close to an $\alpha$ helix, with the twist angle $\phi$, between successive side chains in the range $80° \le \phi \le 120°$ (the canonical angle being 100°). We can then carry out maximization in terms of $s$, $e$, and $\phi$ and assume that the contribution of the sequence to activity is measured by its normalized SAM value. We have set the lower limit of the length to five amino acids and the upper limit to $u$, encompassing all of the pre-Rana-box segment. Furthermore, the SAM value is calculated assuming that the analyzed segment adopts a helical conformation with axis perpendicular to the plane $\pi$ in which we assign the twist angle $\phi$. We then assign to each amino acid attribute vector $x(a_i)$ so that

(1) it is parallel to plane $\pi$;
(2) it points outward from the helix axis if $x(a_i) > 0$, and inward if $x(a_i) < 0$; and
(3) its length is $|x(a_i)|$.

We obtain the SAM value by summing up all the vectors $x(a_i)$. The pre-Rana-box descriptor $P_{xy}$ is therefore denoted as $\mathrm{SAM}_{\text{pre-Rana-box}}$ and is defined as

$$\mathrm{SAM}_{x,\phi}(a_s a_{s+1} \ldots a_e) = \frac{\sqrt{\left(\sum_{i=s}^{e} x(a_i)\cos(i\phi)\right)^2 + \left(\sum_{i=s}^{e} x(a_i)\sin(i\phi)\right)^2}}{(e-s+1)^{1/3}}$$

which is $P_{xy}$ limited to the most asymmetric subsequence in the pre-Rana-box segment.

In normalizing, we divide by the third root of the segment length $e - s + 1$, in order to compromise between two extremes:

(1) the actual strength of the property throughout the observed segment (that would mean no normalization at all, i.e., division by 1), and
(2) the density of the strength of the property throughout the observed segment, i.e., the strength of property per amino acid (that would mean division by the length of the segment).

We consider condition (1) to be more important than (2), so we choose not to use the square root (that would correspond


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

to the geometrical mean of the two divisors), but rather the third root, to increase the weight of property (1), i.e., to increase the weight of a segment length, $e - s + 1$, in calculating SAM values. The alternative extremes (1) and (2) respectively neglect the strength of the property per amino acid and the actual strength of the property for observed segment.

**Definition of the Rana-Box Descriptor.** Let us now define the Rana-box descriptor, denoted by $R_x$, in similar terms to the pre-Rana-box descriptor. In this case, however, we consider only one scale at a time and analyze the whole sequence rather than a most active part, as it is very short and we assume that all residues are important to peptide activity. The descriptor $R_x$ is given by

$$R_x(a_1a_2...a_ub_1b_2...b_v) = \max_{\phi=\{170°,171°,...,180°\}} \left\{ \frac{\sqrt{(\sum_{i=1}^{v} x(b_i)\cos(i\phi))^2 + (\sum_{i=1}^{v} x(b_i)\sin(i\phi))^2}}{v^{1/3}} \right\}$$

where $x(b_i)$ is the normalized amino acid attribute value, $\phi$ is the twist angle of successive side chains, and $v$ is the length of the Rana-box sequence.

The result of the $R_x$ calculation is denoted as $\mathrm{SAM}_{\text{Rana-box}}$ and is defined as

$$\mathrm{SAM}_{x,\phi}(b_1b_2...b_v) = \frac{\sqrt{(\sum_{i=1}^{v} x(b_i)\cos(i\phi))^2 + (\sum_{i=1}^{v} x(b_i)\sin(i\phi))^2}}{v^{1/3}}$$

The main problem we addressed here is how to obtain appropriate SAM for the Rana-box segment. Rana-box residues form different types of 3D structures,[39,40,35] having in common the cystine motif. An approximate center position can be found for the cycle, although the 3D structure is not a regular flat circle, but rather a combination of loop, helix turn and disordered structure. The cycle would be asymmetric if residues with mostly high values of a given attribute point away from cycle center or toward it. Alternatively, the side chains of these residues could extend above or below the cycle plane. In both cases, one could model the sequence as being an extended one, so that a form of sideways asymmetry would be generated (resulting in high SAM values) by linearizing the cycle and determining the SAM value using twist angles from 170° to 180°. Note that this is simply a formalism and we are not implying, indeed we can safely assume, that most Rana-box segments are not in a $\beta$ strand conformation.

**Obtaining Models.** We first searched for a high but reasonable number of one-parameter models for each of two peptide sections, using two pre-Rana-box SAM descriptors and one Rana-box SAM descriptor, and the set $S$ of 544 amino acid attribute scales in the AAindex database,[28,41] to which we added extended consensus scale determined on a subset of these scales, as described by Tossi et al.[42] Hence, there are $\binom{545}{2}$ = 148 240 possible pre-Rana-box descriptors.

We then restricted our attention to 66 809 descriptors that resulted in correlation coefficients with log(MIC) of $r^2 > 0.2$, i.e., that in some way reflect the relevance of helical conformation to activity. Analogous criteria were used for the 545 Rana-box descriptors, and only 186 of these satisfy $r^2 > 0.2$ for correlation with log(MIC). Hence, we collected 66 809 and 186 one-parameter linear models from pre-Rana-box and Rana-box sections, respectively. Subsequently, we analyzed 66 809 × 186 = 12 426 474 two-parameter linear models in search for those for which it holds that

(a) the model improves $r^2$ with respect to a one-parameter correlation by more than 0.2 (this incorporates the expectation that both parts of peptide play a role in its biological activity);
(b) the overall correlation coefficient is $r^2 > 0.72$; and
(c) the "leave one out" determination coefficient is $r^2 > 0.67$.

The value 0.72 was selected simply to allow us to obtain a sufficient number of models (not just one or two), while that of 0.67 was chosen so as to be smaller by 0.05 (which we considered to be an acceptable difference). There are no other special justifications for our choices of thresholds. We simply selected the numbers in such a way that $r^2$ could be as high as possible and still to allow that some models survive. This permitted us to obtain eight models, which we then used to analyze the test set of peptides, two of which showed particularly interesting properties.

**Visualization Software.** Peptides for which Protein Data Bank 3D structures were available (Ranatuerin-2CSa, PDB code: 2K10; Gaegurin 4, PDB code: 2G9L) were visualized using the Discovery Studio version 3.5. 2012. (Accelrys Software Inc., San Diego, California).

**Peptide Synthesis.** Ranaboxin-1 (FRGAFRGVLRGG-RVAGRLCKLKLQC) was synthesized in the solid phase using a Biotage Altra instrument and standard Fmoc chemistry on a Fmoc-Cys(Trt)-NovaSyn Tentagel amide resin (substitution 0.22 mmol/g). The peptide was cleaved from the resin using 3% $H_2O$, 3% triisopropylsilane, 8% 3,6-dioxa-1,8-octanedithiol, and 86% trifluoroacetic acid (TFA) for 3.5 h at room temperature and precipitated with methyl *tert*-butyl ether. The molecular weight of the peptide was confirmed using ESI-MS (Bruker Daltonics, Esquire 4000) as 2762.1 Da (calculated mass 2762.4). The peptide was then resuspended in 10 mL of 5 M guanidine HCl and added to 1 L of folding buffer (0.1 M ammonium acetate, 2 mM ethylenediaminetetraacetic acid, pH 7.5) with final peptide concentration of 80 $\mu$g/mL. Cystine and cysteine were also added at a peptide:cystine:cysteine ratio of 1:5:50. At different times, aliquots were analyzed by RP-HPLC on a Waters Symmetry analytical column ($C_{18}$, 3.5 $\mu$m, 100 Å, 4.6 × 75 mm) using a gradient of 0−90% solvent B (0.05% TFA v/v in acetonitrile) in 30 min with flow of 0.8 mL/min. A shift in the main band from a retention time of 15 to 15.5 min (more hydrophobic) was consistent with disulfide bond formation, and was effectively complete by 24 h, as also confirmed by a shift in mass to 2760.9 Da (calculated molecular weight of oxidized peptide is 2760.4 Da). The folded peptide was purified by preparative RP-HPLC using a Phenomenex Jupiter column ($C_{18}$, 10 $\mu$m, 90 Å, 250 × 21.20 mm) and a gradient of 10−30% B in 60 min with flow of 8 mL/min. Fractions containing the folded peptide were confirmed by ESI-MS, pooled, dried and relyophilized from 10 mM HCl to remove trifluoroacetate. The absence of free cysteine was confirmed using Elman's reagent. The peptide was determined to be >95% pure by analytical RP-HPLC using a gradient of 0−40% B in 40 min at 0.8 mL/min.

## RESULTS AND DISCUSSION

**Data Sets of Antimicrobial Peptides.** Our initial aim of predicting MICs for non-homologous peptides of different length, with a small number of transparent and meaningful


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

**Table 1. Training Set Peptides with Measured and Predicted MIC against *E. coli***

| Peptide | Sequence[a] | b.c. | MIC (μM) (measured)[b] | MIC (μM) (predicted) |
|---|---|---|---|---|
| Brevinin-2Ej[48] | GIFLDKLKNFAKGVAQSLLNKASC**KLSGQC** | 10101 | 2.00 | 3.42 |
| Esculentin-1ARa[49] | GIFSKINKKKAKTGLFNIIKTVGKEAGMDVIRAGIDTISC**KIKGEC** | 10101 | 2.00 | 3.63 |
| Brevinin-2SKb[50] | GLFNVFKKVGKNVLKNVAGSLMDNLKC**KVSGEC** | 10101 | 3.00 | 2.96 |
| Brevinin-2PRc[51] | GLMSVLKGVLKTAGKHIFKNVGGSLLDQAKC**KISGQC** | 10101 | 3.00 | 2.33 |
| Esculentin-1ISb[52] | RIFSKIGGKAIKNLILKGIKNIGKEVGMDVIRTGIDVAGC**KIKGEC** | 10101 | 3.10 | 4.49 |
| Ranatuerin-2CSa[35] | GILSSFKGVAKGVAKDLAGKLLETLKC**KITG-C** | 1010 | 5.00 | 6.78 |
| Ranatuerin-2PRd[53] | GILSSIKGVAKGVAKNVAAQLLDTLKC**KITG-C** | 1010 | 6.00 | 17.56 |
| Brevinin-2PRa[51] | GLMSLFKGVLKTAGKHIFKNVGGSLLDQAKC**KITGEC** | 1010 | 6.00 | 3.74 |
| Esculentin-1ISa[52] | GIFSKFAGKGIKNLLVKGVKNIGKEVGMDVIRTGIDIAGC**KIKGEC** | 10101 | 6.30 | 3.48 |
| Brevinin-2PTb[54] | GFKGAFKNVMFGIAKSAGKSALNALACKIDKSC | 10111 | 9.00 | 18.90 |
| Brevinin-1CHa[55] | FLPIIAGVAAKVLPKLFCAITKKC | 10111 | 10.00 | 25.33 |
| Ranalexin-1Vb[56] | FLGGLFKLVPSVICAVTKKC | 00111 | 12.50 | 23.95 |
| Brevinin-2ISb[52] | SFLTTFKDLAIKAAKSAGQSVLSTLSCKLSNTC | 10111 | 12.50 | 17.22 |
| Nigrocin-2JDb[57] | GIFGKILGVGKKVLCGLSGMC | 00101 | 15.00 | 26.30 |
| Esculentin-2HSa[54] | GIFSLIKGAAQLIGKTVAKEAGKTGLELMACKVTKQC | 10111 | 16.00 | 17.43 |
| Ranatuerin-2BYb[58] | GIMDSVKGLAKNLAGKLLDSLKC**KITG-C** | 1010 | 17.00 | 17.70 |
| Brevinin-2PTa[54] | GAIKDALKGAAKTVAVELLKKAQCKLEKTC | 10111 | 18.00 | 32.55 |
| Brevinin-2PTe[54] | GFLDSFKNAMIGVAKSVGKTALSTLACKIDKSC | 10111 | 18.00 | 19.44 |
| Brevinin-1BYa[59] | FLPILASLAAKFGPKLFCLVTKKC | 00101 | 20.00 | 16.05 |
| Ranatuerin-2CHb[55] | GLMDTIKGVAKNVAASLLEKLKC**KVTG-C** | 1010 | 20.00 | 26.11 |
| Ranatuerin-1[60] | SMLSVLKNLGKVGLGFVACKINKQC | 10111 | 20.00 | 31.26 |
| Brevinin-1HSa[54] | FLPAVLRVAAKIVPTVFCAISKKC | 00111 | 24.00 | 35.64 |
| Brevinin-1Sb[61] | FLPAIVGAAAKFLPKIFCAISKKC | 00111 | 25.00 | 16.68 |
| Brevinin-1AUb[62] | FLPILAGLAANILPKVFCSITKKC | 10111 | 25.00 | 16.09 |
| Esculentin-2PRb[53] | GIFSALAAGVKLLGNTLFKMAGKAGAEHLACKATNQC | 10111 | 25.00 | 29.63 |
| Nigrocin-2HSa[54] | GLLGSLFGAGKKVACALSGLC | 00101 | 28.00 | 16.56 |
| Ranatuerin-2ARa[49] | GLMDTVKNAAKNLAGQLLDTIKC**KMTG-C** | 1010 | 30.00 | 26.34 |
| Brevinin-1CSa[63] | FLPILAGLAAKIVPKLFCLATKKC | 00101 | 32.00 | 31.51 |
| Esculentin-2JDa[57] | GLFTLIKGAAKLIGKTVAKEAGKTGLELMACKITNQC | 10111 | 34.00 | 17.02 |
| Brevinin-2HSa[54] | GLLDSLKNLAINAAKGAGQSVLNTLSCKLSKTC | 10111 | 36.00 | 23.78 |
| Brevinin-1PRc[53] | FFPMLAGVAARVVPKVICLITKKC | 00111 | 50.00 | 54.65 |
| Brevinin-1JDa[57] | FLPAVIRVAANVLPTVFCAISKKC | 00101 | 50.00 | 37.74 |
| Ranatuerin-2PRb[53] | GILDTFKGVAKGVAKDLAVHMLENLKC**KMTG-C** | 1010 | 50.00 | 20.23 |
| Nigrocin-2ISb[52] | GILGTVFKAGKGIVCGLTGLC | 00100 | 50.00 | 71.66 |
| Brevinin-2ISc[52] | SVLGTVKDLLIGAGKSAAQSVLTTLSCKLSNSC | 10111 | 50.00 | 17.18 |
| Ranatuerin-2VLb[64] | GIMDTIKGAAKDLAGQLLDKLKCKITK-C | 1011 | 75.00 | 30.25 |
| Brevinin-1CHc[55] | FFPTIAGLTKLFCAITKKC | 00111 | 80.00 | 43.01 |
| Palustrin-2ISa[52] | GFMDTAKNVAKNVAVTLLDKLKCKITGGC | 10100 | 100.00 | 51.70 |

[a]Sequences are aligned on the Rana-box (shaded); segments in bold have a repetitive binary code (b.c.) 1010 frequently associated with a high activity (low MIC): 1 (hydrophilic), 0 (hydrophobic). We used Eisenberg's scale[31] to assign I, F, V, L, W, M, A, and G as hydrophobic residues and all the others as hydrophilic. [b]Measured against initial inocula of $5 \times 10^5$ CFU/mL except for Brevinin-2Ej, Esculentin-1ARa, Ranatuerin-2BYb, and Ranatuerin-2ARa, which had initial inocula of $2.5 \times 10^5$ CFU/mL.

descriptors (two or three), was achieved when we used AMPs with common evolutionary origin in both the training and test data sets. Our choice fell on anuran AMPs with a Rana-box at their C-terminus. This segment has two flanking cysteine residues separated by four or five other residues forming a disulfide bridge, the second cysteine usually being at the C-terminus (other rarer variations of the Rana-box theme were not considered here). The Rana-box motif is restricted to the Neobatrachia suborder of Anura[43] and is always associated with a very well conserved class-1 signal peptide in precursor sequence,[44] indicating a common evolutionary origin for this disulfide motif. Some Rana-box peptides have been reported to have other activities in addition to antimicrobial activity, such as anticancer, antifungal, serine protease inhibition, mast cell degranulation, histamine release, insulin release, inhibition of nitric oxide release, and stimulation of splenocyte proliferation (riparin-1.2).[44−46] The most prominent AMP families bearing the Rana-Box motif are esculentins, brevinins, ranatuerins, ranalexins, palustrins, and nigrocins.

MIC values for Rana-box peptides have been measured in many different laboratories, and often under different conditions. When considering MIC values obtained by the microdilution method, we concluded that the inoculum size (number of CFU/mL) is the most important feature in need of standardization. Low MIC values (apparently stronger bacteriostatic activity) are obtained when inocula significantly lower than $5 \times 10^5$ CFU are used,[47] and this can be misleading. Therefore, we constructed the database of 59 natural Rana-box peptides with known sequences and reliable MIC values (mostly obtained using $5 \times 10^5$ CFU inocula). The criteria used to fill the training and test sets of peptide sequences (see Materials and Methods) assigned these 38 and 19 peptides, respectively (see Tables 1 and 2), while two peptides were left unassigned. Measured[48−64] and predicted MIC values are also included in the tables for comparison.

As far as we can tell, we have used an original method for constructing these sets of peptides, which can be provisionally termed the triplet sorting method. Namely, we wanted to obtain triplets of peptides within a certain range of sequence


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

Table 2. Test Set Peptides with Measured and Predicted MIC against *E. coli*

| Peptide | Sequence[a] | b.c. | MIC (μM) (measured)[b] | MIC (μM) (predicted) |
|---|---|---|---|---|
| Brevinin-2Ei[48] | GILSTIKDFAIKAGKGAAKGLLEMASC**KLSGQC** | 10101 | 3.00 | 5.61 |
| Brevinin-2SKa[50] | GLFSAFKKVGKNVLKNVAGSLMDNLKC**KVSGEC** | 10101 | 3.00 | 3.15 |
| Brevinin-2PRb[51] | GLMSLFRGVLKTAGKHIFKNVGGSLLDQAKC**KITGEC** | 10101 | 3.00 | 4.40 |
| Brevinin-2PRd[51] | GLMSVLKGVLKTAGKHIFKNVGGSLLDQAKC**KITGQC** | 10101 | 3.00 | 2.84 |
| Ranatuerin-2AUa[62] | GILSSFKGVAKGVAKNLAGKLLDELKC**KITG**-C | 1010 | 5.00 | 7.15 |
| Ranatuerin-2BYa[58] | GILSTFKGLAKGVAKDLAGNLLDKFKC**KITG**-C | 1010 | 7.00 | 5.85 |
| Brevinin-1CHb[55] | FLPVIAGLAAKVLPKLFCAITKKC | 00111 | 10.00 | 24.72 |
| Esculentin-1HSa[54] | GIFSKFGGKAIKNLFIKGAKNIGKEVGMDVIRTGIDVAGC**KIKGEC** | 10101 | 12.00 | 4.40 |
| Esculentin-2ISa[52] | GIFSLIKGAAKLITKTVAKEAGKTGLELMACKVTNQC | 10111 | 12.50 | 16.40 |
| Esculentin-2PRa[53] | GVFSFLKTGAKLLGSTLLKMAGKAGAEHLACKATNQC | 10111 | 12.50 | 27.19 |
| Brevinin-1AUa[62] | FLPILAGLAAKLVPKVFCSITKKC | 10111 | 13.00 | 15.45 |
| Nigrocin-2HSb[54] | GLLGSIFGAGKKIACALSGLC | 00101 | 14.00 | 16.56 |
| Brevinin-2PTc[54] | GLLDSFKNAMIGIAKSAGKTALNKIACKIDKTC | 10111 | 18.00 | 36.61 |
| Ranatuerin-2CHa[55] | GLMDTVKNAAKNLAGQLLDRLKC**KITG**-C | 1010 | 20.00 | 25.08 |
| Brevinin-1PTa[54] | FMGGLIKAATKIVPAAYCAITKKC | 00111 | 24.00 | 36.74 |
| Nigrocin-2ISa[52] | GIFSTVFKAGKGIVCGLTGLC | 00100 | 25.00 | 62.26 |
| Brevinin-1JDc[57] | FLPAVLRVAAKVVPTVFCLISKKC | 00111 | 49.00 | 50.26 |
| Brevinin-1PRd[53] | FLPMLAGLAASMVPKLVCLITKKC | 00111 | 50.00 | 44.86 |
| Brevinin-2ISa[52] | SLLDTFKNLAVNAAKSAGVSVLNALSCKISRTC | 10111 | 50.00 | 51.33 |

[a]Sequences are aligned on the Rana-box (shaded); segments in bold have a repetitive binary code (b.c.) 1010 frequently associated with a high activity (low MIC): 1 (hydrophilic), 0 (hydrophobic). We used Eisenberg's scale[31] to assign I, F, V, L, W, M, A, and G as hydrophobic residues and all the others as hydrophilic. [b]Measured against initial inocula of $5 \times 10^5$ CFU/mL except for Brevinin-2Ei and Ranatuerin-2BYa,which had initial inocula of $2.5 \times 10^5$ CFU/mL.

similarity (not too low and not too high) and place two in the training and one in the testing set, ensuring at the same time the desired degree of similarity between the two sets and desired degree of dissimilarity or taxonomic distance within each set. The training set contains sequences with identity as low as 12.5% (13 pairs) and only one pair with an identity as high as 89.2% (Brevinin-2PRc and Brevinin-2PRa). The test set contains two pairs with identity as low as 8.3% and one with identity as high as 89.2% (Brevinin-2PRd and Brevinin-2PRb). On the other hand, several peptide pairs with only one different residue at a corresponding sequence position are present when considering different sets.

Highly antimicrobially active peptides listed in Tables 1 and 2 (MIC values $\leq 10$ $\mu$M) are not necessarily also highly toxic to animal cells. Measured $HC_{50}$ values (peptide concentration that lyses 50% of human red blood cells) have been reported in literature for about 40% of the peptides included in Tables 1 and 2. Only brevinins-1 exhibit low $HC_{50}$ values (i.e., have a high hemolytic activity), but none of these have MIC < 10 $\mu$M.

Rana-box peptides are only a small subset of AMPs with disulfide bridges, with defensins being the best known examples. Defensins have three or four disulfide bridges but, like Rana-box peptides, also often form membrane attached helices with a helical N-terminal domain of an amphipathic nature (with one side of hydrophobic and another of hydrophilic nature).[65] By finding relevant and meaningful descriptors for predicting antibacterial activity of Rana-box peptides we hoped to bridge the world of helical AMPs[66−68] and the world of cysteine-containing AMPs with several intramolecular disulfide bonded AMPs.[69]

**QSAR Multilinear Models with Two Descriptors.** In the current study we intended to use a limited number of descriptors in the QSAR model. We have previously shown that only one descriptor (the *D*-descriptor, defined as the cosine of the angle between two sequence moments) was enough for predicting the therapeutic index of linear anuran AMPs that adopt a helical conformation in the presence of membranes.[11] In the case of Rana-box peptides the most productive idea for MIC prediction was to consider all possible SAMs (see Materials and Methods). We used all 544 published amino acid scales in the AAindex database[28] plus the extended consensus scale of Tossi et al.,[42] a range of twist angles, and different segment lengths in each of the examined peptide sequences, to see if regular asymmetry deriving from certain amino acid attributes in these segments could be related to measured MIC values.

Two SAMs were used as descriptors. The first descriptor was applied to all amino acids in the segment before the Rana box. It was the higher from among two SAMs, calculated using side-chain twist angles between 80° and 120° to account for different helical conformations. The second SAM was applied only to the Rana-box segment with twist angles set to between 170° and 180°, accounting for different extended structures that could effectively represent sideways asymmetry when they are looped into the cyclic Rana-box structure.

This approach, taking into account two different regions requiring two different descriptors, produced a large number of multiple linear regression (MLR) models (a total of 12 426 474) when a large number of amino acid scales (545) and all combinations of scale pairs (148 240) were used; hence, we needed a rapid and economical estimate of model quality. We therefore used correlation with the MIC, $r^2$, although it is not the necessarily the best measure. Eight MLR models were found with correlation coefficient $r^2 > 0.72$ (in the training set) and with determination coefficient (against the test set) ranging from $R^2 = 0.52$ to $R^2 = 0.70$. For all of these models, the "leave one out" cross-validation (the jack-knife resampling technique[17]) produced $r^2 = 0.67-0.69$. Figures 1 and 2 illustrate the distribution of measured and predicted log(MIC) values for the best model having correlation coefficient $r^2 = 0.726$ among measured and predicted log(MIC) values for the training set of peptides (Table 1), and a determination coefficient $R^2 = 0.703$ when validated against the test set of peptides (Table 2). We also tested this model by making one random permutation of


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

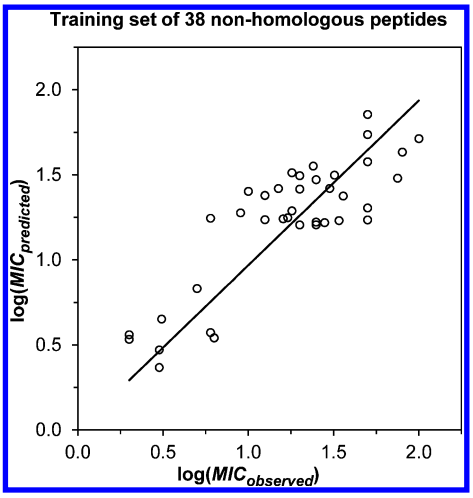



**Figure 1.** Best QSAR model ($r^2$ = 0.726) for correlating measured and predicted MIC of non-homologous anuran peptides with Rana-box at their C-terminal: $\log(MIC_{predicted}) = 4.93 - 6.75\ SAM_{pre\text{-}Rana\text{-}box} - 7.87\ SAM_{Rana\text{-}box}$.

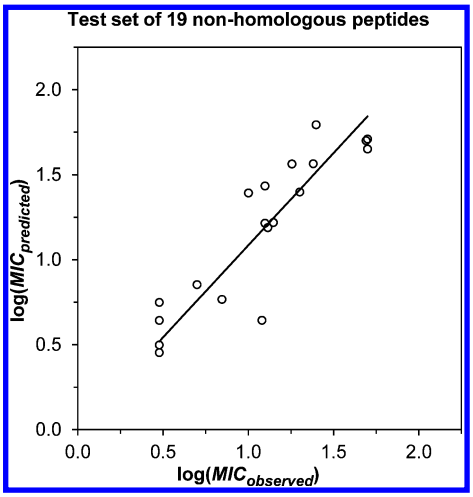



**Figure 2.** Validation of the best QSAR model at test set of non-homologous anuran peptides with Rana-box at their C-terminal (determination coefficient $R^2$ = 0.703).

measured log(MIC) to show that this high $R^2$ value is not random. We obtained a value $R^2$ = 0.268, which is significantly lower. The second best model had slightly lower correlation and determination coefficients, $r^2$ = 0.721 and $R^2$ = 0.693, respectively, while the other six models had a considerably weaker performance.

While these models are far from being perfect MIC predictors, they can help pinpoint sequences that will likely display good (low MIC) or poor (high MIC) antibacterial activity.

From Figures 1 and 2, it is obvious that our best model has its up and down sides. The latter is the apparent separation of peptides in two distinct groups, those with good activity (MIC $\leq$ 10 $\mu$M) and those with mediocre/weak activity (MIC > 10 $\mu$M), so that the high values of $r^2$ and $R^2$ are partially due to such separation. On the upside, the possibility of computational identification or vetting of peptides prior to synthesis and activity testing saves both time and resources.

In any case, an error by a factor of 2 is implicit in measured MIC values due to the nature of the standard microdilution measuring method,[70,71] making it impossible to achieve an exact prediction. The more modest goal of predicting whether a novel sequence, with features common with those of Rana-box peptides, will be more likely in the high or low MIC range is instead perfectly feasible.

**Best QSAR Models with Corresponding SAM Descriptors and Amino Acid Scales.** The QSAR models we have identified perform an automatic choice between SAM descriptors for volume and buriability for a peptide domain of optimal length, to produce maximal SAM in the pre-Rana-box segment, where domains are assumed to adopt some form of helical conformation with a twist angle in the 80−120° range. The best model uses the Grantham scale of side-chain volumes[72] (AAindex code:[28] GRAR740103) and the Zhou and Zhou buriability scale[73] (AAindex code:[28] ZHOH040103), derived by considering the effect of burial of amino acid residues on protein stability. The second best model uses the Krigbaum and Komoriya scale of side-chain volumes[74] (AAindex code:[28] KRIW790103) and again the Zhou and Zhou buriability scale.[73] Note that the Grantham and Krigbaum-Komoriya scales are highly correlated ($r$ = 0.989). The buriability SAM descriptor wins in 32 out of 38 pre-Rana-box segments from training set peptides over volume SAM descriptor in the case of our best model (see Supporting Information, Table S1).

For the Rana-box domain, the SAM descriptor used in both of the two best models is the Fasman's melting point scale[75] (AAindex code:[28] FASG760102), which is clustered together with hydrophobicity scales reported by Nakai et al.[41]

Various buriability descriptors are an automatic choice of our algorithm (see Materials and Methods) present in all of eight selected MLR models. In addition to ZHOH040103 from Zhou and Zhou,[73] these models are using the Wertz and Scheraga scale[76] (AAindex code:[28] WERD780101) for propensities to be buried inside proteins, and another buriability scale by Zhou and Zhou[73] (AAindex code:[28] ZHOH040101), a stability scale from the knowledge-based atom−atom potential.

When winning segments identified with the buriability descriptor in the best model are examined in more detail, all of such segments in peptides with MIC $\leq$ 15 $\mu$M are associated with twist angles of 93 $\pm$ 3°, a little less than the canonical 100° of an ideal $\alpha$-helix (Table S1). This might suggest the possibility of an intrahelical $\pi$ bulge,[77,78] although we should keep in mind that the MLR model does not predict secondary structure but only finds descriptors to maximize MIC prediction. The NMR structures of Rana-box peptides available in the Protein Data Bank (PDB codes: 2K10 and 2G9L) do not in fact exhibit any $\pi$-helical segments. Furthermore, segments associated with winning SAM having twist angles less than 100°, and MIC < 15 $\mu$M are often too long (8 out of 12 such segments are longer than 20 amino acids, see Table S1) for the $\pi$-helix assignment to be realistic.[79]

An additional caveat is that only MLR models can bring buriability SAM descriptors to the forefront of best attributes for relating sideways asymmetry and bacteriostatic activity. The best one-parameter model for flexible helix parameters of $SAM_{pre\text{-}rana\text{-}box}$ was obtained using Fasman's hydrophobicity index[80] (AAindex code:[28] FASG890101) (see Supporting Information, Table S2).

**Relationship between the MLR Model and 3D Structure.** The NMR structure of Ranatuerin-2CSa (PDB code: 2K10)[35] was visualized with color rendering using the buriability scale of amino acid attributes derived by considering


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

the effect of burial of amino acid residues on protein stability (Figure 3A) or Eisenberg's consensus hydrophobicity scale (Figure 3B), to get some insight about possible structure-based reasons why the former descriptor is better at predicting activity.

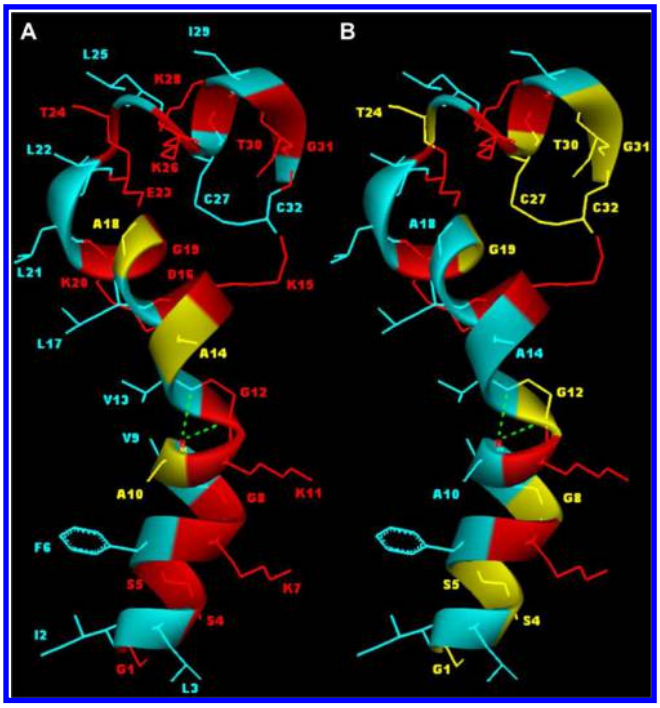



**Figure 3.** Ranatuerin-2CSa NMR structure (PDB code: 2K10) with two different color-coding schemes. AAs are sorted into three groups—red representing the lowest third, yellow the middle third, and cyan the highest third of scale values: (A) buriability scale ZHOH040103 [red AAs with low buriability parameter values (K, G, E, N, D, S, R, Q, P, T, H), yellow AAs with medium buriability parameter values (A, M), cyan AAs with high buriability parameter values (Y, V, I, L, C, F, W)] and (B) hydrophobicity scale EISD840101 [red AAs with low hydrophobicity values (R, K, D, Q, N, E, H), yellow AAs with medium hydrophobicity values (S, T, P, Y, C, G), cyan AAs with high hydrophobicity values (A, M, W, L, V, F, I)]. Visualization using Discovery Studio.

The buriability scale appears to be at least as good as Eisenberg's hydrophobicity scale in separating hydrophobic from hydrophilic helix faces. It defines groupings of hydrophobic residues with a high propensity to become buried at the concave side of the bent pre-Rana-box domain and groupings of hydrophilic residues (including glycine) at the convex side (Figure 1A). Calculated SAM values for the optimal segment $I^{(2)}$LSSFKGVAKGVAKDLAGKLLETL$^{(25)}$ having maximal SAM are 0.545 for buriability (95° twist angle) and 0.462 for the hydrophobicity scale (fixed canonical 100° twist angle). For a fair comparison we used for both evaluations the scales normalization procedure and weighted moments (the third root of the segment length) as defined in the Materials and Methods section. A helical wheel projection for Ranatuerin-2CSa segment 2−25 in the case (A) for twist angle 95 achieves perfect separation of low buriability residues K, G, E, D, S, T from all other residues (not shown).

An interesting feature observed in Figure 3 is a Möbius-like strip connecting A10 to G12 and V13 through hydrogen bonds, corresponding to a kink annotated as a $3_{10}$ helical structure by PDB entry 2K10. The intrahelical deformation is particularly evident between K11 and G12, and is likely to be both a structurally and functionally important feature, analogously to bends in regular $\alpha$ helices caused by the presence of a proline residue.[24,37] In this case, a non-proline kink caused by the sequential repeat motif GVAKGVAK is able to reposition the second helix in the Pre-Rana-box domain in such a way as to maximize the amphipathic nature of the whole segment. It also aligns residues spaced $i$ and $i$+4 apart so that glycines (G8, G12) and lysines (K7, K11, K15) are on one side of the helix while alanines (A10, A14, A18), valines (V9, V13), and leucines (L17, L21, L25) are on the opposite side. Beside increasing amphipathicity, a regular arrangement of small residues such as Ala, Gly, and Ser in motifs of the type [G,A,S]-X-X-X-[G,A,S] is important for dimerization of helices in the membrane environment.[81] The difference in color coding scheme for A, G and S (Figure 3B), due to differences between buriability and hydrophobicity scale, helps in recognition of their regular positioning along helix sides.

**Performance of the SAM Method Compared against the "Standard" HM Method.** The standard way of calculating the HM for complete pre-Rana-box segments, using Eisenberg's scale and canonical 100° twist angles, produces an unimportant correlation $r^2$ = 0.22 with bacteriostatic activity. Furthermore, Esculentin-1ISb had the lowest calculated HM among all training set peptides (HM = 0.034), despite being very active (MIC = 3.1 $\mu$M, Table 1). Using Eisenberg' scale in the SPLIT algorithm[82] (http://splitbioinf.pmfst.hr/split/), however, returns a very good HM profile along the sequence, and this is even more evident using the HeliQuest algorithm,[83] with the central 12 AA segment $I^{(15)}$LKGIKNIGKE$^{(26)}$, having HM = 0.77. This algorithm uses sliding windows of a desired length and Fauchere and Pliska's hydrophobicity scale[84] based on octanol/water partitioning of AA side chains.

The assumption that the whole pre-Rana-box segment of 39 amino acids, RIFSKIGGKAIKNLILKGIKNIGKEVGMDVIR-TGIDVAG adopts an ideal, unbent $\alpha$-helical conformation with canonical 100° twist angle is clearly unwarranted, and most likely killing the goal of standard HM calculation to determine if the peptide is amphipathic or not. Calculating the mean HM using sliding segments of 11−12 amino acids has a much better chance of identifying highly amphipathic peptides, or segments therein, compatible with a fairly regular secondary conformation. This is a good rule that is often not applied when using "standard" HM calculations for peptides longer than 15−20 residues.

Accordingly, we compared the performance of the "standard" HM method against the SAM method by finding the higher of maximal HM for all 11 or 12 AA segments in each peptide from our training set (limited to the pre-Rana-box region). Within this linear one descriptor model HMs were calculated using Fasman's hydrophobicity index[80] (FASG890101), and resulted in an $r^2$ = 0.31 when correlated with log($MIC_{experimental}$). The performance was considerably weaker than the $r^2$ = 0.51 obtained using SAM calculations for the same scale (see Table S2). The determination coefficient using test set peptides ($R^2$ = 0.29) was also weaker than that obtained with SAM calculations ($R^2$ = 0.33). Thus, one must take into account the need for some flexibility in twist angle and an optimal segment length in secondary structure elements before one can expect descriptors derived from these to exhibit a useful correlation with antibacterial activity.


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

**Rana-Box Sideways Asymmetry.** Sideways asymmetry could also be observed in Rana-box residues from peptides with better antibacterial activity (Figure 4), if the QSAR method

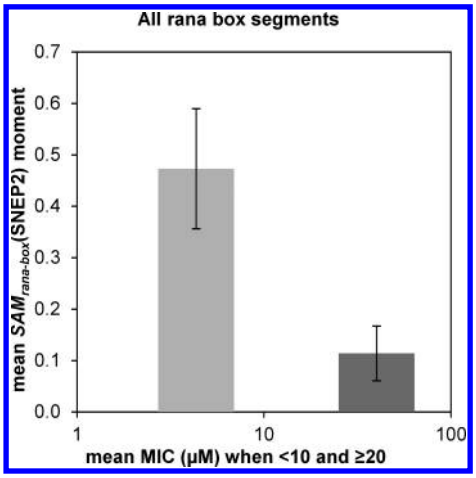



**Figure 4.** Correlation of mean Rana-box $SAM_{Rana\text{-}box}$(SNEP2) and log(MIC) for peptides showing high or low activity. The SNEP660102 amino acid attributes of Sneath[85] were used to calculate $SAM_{Rana\text{-}box}$(SNEP2) for all Rana-box segments within the full set of 57 peptides. Peptides were then selected with MIC ≤ 10 μM or ≥20 μM, and the mean MIC values for each set, on a logarithmic *x*-axis set against the corresponding mean $SAM_{Rana\text{-}box}$(SNEP2). A mean SAM of 0.47 corresponded to a mean MIC of 4.3 μM, and a mean SAM of 0.11 to a mean MIC = 40.1 μM. Error bars are set up for three standard deviations above and below the mean SAM values. SAM calculations were performed with twist angle on the 170−180° range.

considered the primary structure as in an extended strand conformation, with twist angles ranging from 170° to 180°. The SAM corresponding to the best correlation with activity, within the simplest linear one descriptor model, was found in this case using the scale derived by Sneath,[85] the second principal component vector calculated in his study of the relation between chemical structure and biological activity of peptides (AAindex code:[28] SNEP660102). It achieves the correlation coefficient $r^2$ = 0.65 for the same training set of 38 peptides (see Table S2), even though limited to only the Rana-box residues (we named this the $SAM_{Rana\text{-}box}$(SNEP2) model). The determination coefficient using the test set was $R^2$ = 0.47, still quite significant ($p$ = 0.00017), while the "leave one out" determination coefficient was 0.61. This is clearly seen in Figure 4, comparing peptides with MIC ≤ 10 to those with MIC ≥ 20 μM. Those with intermediate MIC were not considered, as considering the experimental error inherent in serial dilutions (doubling or halving of the value[70,71]), 10 ≤ MIC ≤ 20 μM values could actually fall within either the low or high MIC sets.

$SAM_{Rana\text{-}box}$(SNEP2) values ≥0.45 for Rana-box segments were all associated with peptides in the training set showing low MIC values (seven peptides with measured MIC values from 2 to 6.3 μM). Out of five such SAM values in the testing set, four were associated with the MIC = 3 μM, and one with MIC = 12.5 μM. The lower activity of this outlier is due to the pre-Rana-box segment, as the CKIKGEC Rana-box motif is otherwise found in peptides from the same family (Esculentin-1) with MIC values of 2−6.3 μM (see Tables 1 and 2).

When assembling our tables and correlating measured and predicted MIC, we noticed an interesting pattern. If one substitutes the residues bracketed by cysteines in the Rana-box with a binary code for their general hydrophobicity (1 = polar hydrophilic, 0 = neutral or hydrophobic), all the peptides with Rana-box correlating with good activity show a 1010 or 10101 pattern. This justifies our treatment of an alternating residue pattern and can thus be used in identifying/designing active Rana-box peptides.

**Public Domain Server for Predicting Bacteriostatic Activity of Rana-Box Peptides.** We used our best QSAR model to construct a public domain online server for MIC prediction: http://splitbioinf.pmfst.hr/micpredictor/. This MIC predictor accepts all sequences based on natural anuran peptides having the C-terminal Rana-box motif CXXXXC or CXXXXXC. It can be used to introduce amino acid substitutions and "*in silico*" construction of Rana-box-like artificial AMPs, allowing the user to identify those predicted to have activity against *E. coli* in particular and Gram-negative bacteria in general. The predicted MIC is indicative and is not limited to the training set range of 2 < MIC < 100 μM. It can therefore be described as a MIC estimate rather than a MIC prediction, as reconciling predicted and measured activity values for AMPs is not facile.[86]

**Ranaboxin-1 Design and Experimental Testing.** We explored the limits of using our MIC predictor on its own, with amino acid substitutions guided only by volume SAM for the pre-Rana-box domain and by the binary code pattern 10101 (Table 1) for the Rana-box domain. A 25-residue peptide designed on this basis, FRGAFRGVLRGGRVAGRLCK-LKLQC, provisionally named ranaboxin-1, has little resemblance to Ranatuerin-2AMa,[87] which served as the parent compound (common in traditional Chinese medicine *lin wa pi*). It had a predicted MIC = 0.1 μM, meaning it is predicted to have an activity at the low end of the 2−100 μM natural peptides range. The peptide was synthesized and the correct molecular weight was confirmed by ESI-MS to ensure disulfide bond formation (see Supporting Information, Figure S1). The peptide purity was verified by RP-HPLC (see Supporting Information, Figure S2). MIC against *E. coli* (ATCC 25922), *Acinetobacter baumannii* (ATCC 19606), and *Pseudomonas aeruginosa* (ATCC 27853) was determined and effectively found to be 2 μM for each of these Gram-negative strains in 20% Müler−Hinton growth medium (microdilution method as described previously[88]). This result therefore confirms ranaboxin-1 among Rana-box AMPs with good antibacterial activity against Gram-negatives. The peptide's activity was, however, found to be medium-sensitive, so that it decreased several-fold in full MH medium. This underlines the difficulty in deviating too much from naturally honed AMPs, in the absence of a suitable designer algorithm to help one. Like our previously published therapeutic index predictor[11] and Mutator algorithm[89] the MIC-predictor is likely to have a better chance for producing peptides with an improved and also robust activity against Gram-negatives, starting from natural Rana-box peptides, if only one or two amino acid substitutions are effected at each step and identifying those features determining robustness as one goes.

## ■ CONCLUSION

Success in predicting bacteriostatic activity of AMPs usually comes at a heavy price, as an excessive number of descriptors is required, while artificial intelligence methods tend not to be transparent about physicochemical factors underlying good activity. In this work we have introduced the sideways asymmetry moment (SAM) concept as a more versatile tool


DOI: 10.1021/acs.jcim.5b00161
J. Chem. Inf. Model. 2015, 55, 2275−2287

than the well-known and much used hydrophobic moment (HM). We have shown that for AMPs with a common evolutionary origin, widely different sequence lengths and low identities are not obstacles in developing multilinear QSAR models with only two SAM descriptors, one extracted from pre-Rana-box amphipathic segments and another extracted from Rana-box amphipathic segment. Our results justify the initial assumption about relevance of amphipathic secondary structure segments for peptide bacteriostatic activity, and also confirm the observation that amphiphilicity is more important than simple hydrophobicity.[90]

A quite surprising and unexpected result was the importance of the primary structure within the Rana-box motif in determining bacteriostatic activity. A binary pattern within Rana-box structures of highly active peptides (Tables 1 and 2) may allow its looplike structure to form hydrogen bonds on one side and hydrophobic interactions on the other, even though it likely does not have a $\beta$-strand conformation. To the best of our knowledge, the connection of amphipathic character of this motif (which we modeled as the HM for an imagined short $\beta$-strand-like structure) with a peptide's antibacterial activity has not been noticed before. This opens the possibility of computational construction of artificial Rana-box peptides with desired selectivity/activity properties. In a publication 20 years ago,[23] the authors had already noticed some structural analogy of Rana-box peptide with polymyxin (colistin), today a drug of last resort[91] for treating infections with the multidrug-resistant Gram-negatives *Pseudomonas* spp. and *Acinetobacter* spp. However, it too is compromised by the emergence of resistance[92] and agricultural use.[93] Peptides with good activity against Gram-negative multiresistant pathogens are likely to be more difficult to find or construct "*in silico*" than AMPs against Gram-positive pathogens,[10,94] so that our approach of a computational search for QSAR descriptors for peptides directed against Gram-negatives may prove to be useful.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00161.

Best model details for pre-Rana-box sections in the case of training set peptides (Table S1), amino acid scales ranking using SAM and correlation among predicted and observed MIC (Table S2), ESI-MS characterization of the synthesized peptide (Figure S1), and peptide purity confirmation by RP-HPLC (Figure S2) (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: juretic@pmfst.hr. Tel.: +385-21-395133.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Authors acknowledge funding from the Croatian Ministry of Science and Croatian Science Foundation project 8481 BioAmpMode. The Peptide Synthesis Laboratory, DLS at UNITS, acknowledges support from Beneficentia Stiftung, Lichtenstein, and Fondazione Stock-Weinberg, Italy.

## ■ ABBREVIATIONS

AMP, antimicrobial peptide; MIC, minimal inhibitory concentration; QSAR, quantitative structure−activity relationship; SAM, sideways asymmetry moment; HM, hydrophobic moment

## ■ REFERENCES


(1) Arias, C. A.; Murray, B. E. Antibiotic-Resistant Bugs in the 21st Century—A Clinical Super-Challenge. *N. Engl. J. Med.* **2009**, *360*, 439−443.

(2) Uchil, R. R.; Kohli, G. S.; Katekhaye, V. M.; Swami, O. C. Strategies to Combat Antimicrobial Resistance. *J. Clin. Diagn. Res.* **2014**, *8*, ME01−ME04.

(3) Spellberg, B.; Shlaes, D. Prioritized Current Unmet Needs for Antibacterial Therapies. *Clin. Pharmacol. Ther.* **2014**, *96*, 151−153.

(4) Zasloff, M. Antimicrobial Peptides of Multicellular Organisms. *Nature* **2002**, *415*, 389−395.

(5) Stotz, H. U.; Thomson, J. G.; Wang, Y. Plant Defensins: Defense, Development and Application. *Plant Signaling Behav.* **2009**, *4*, 1010−1012.

(6) Simunić, J.; Petrov, D.; Bouceba, T.; Kamech, N.; Benincasa, M.; Juretić, D. Trichoplaxin—A New Membrane-Active Antimicrobial Peptide from Placozoan cDNA. *Biochim. Biophys. Acta, Biomembr.* **2014**, *1838*, 1430−1438.

(7) Gottler, L. M.; Ramamoorthy, A. Structure, Membrane Orientation, Mechanism, and Function of Pexiganan—A Highly Potent Antimicrobial Peptide Designed from Magainin. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 1680−1686.

(8) Fjell, C. D.; Hiss, J. A.; Hancock, R. E. W.; Schneider, G. Designing Antimicrobial peptides: Form Follows Function. *Nat. Rev. Drug Discovery* **2012**, *11*, 37−51.

(9) Stauss-Grabo, M.; Atiye, S.; Le, T.; Kretschmaher, M. Decade-Long Use of the Antimicrobial Peptide Combination Tyrothricin Does Not Pose a Major Risk of Acquired Resistance with Gram-Positive Bacteria and *Candida* spp. *Pharmazie* **2014**, *69*, 838−841.

(10) Band, V. I.; Weiss, D. S. Mechanisms of Antimicrobial Peptide Resistance in Gram-negative Bacteria. *Antibiotics* **2015**, *4*, 18−41.

(11) Juretić, D.; Vukičević, D.; Ilić, N.; Antcheva, N.; Tossi, A. Computational Design of Highly Selective Antimicrobial Peptides. *J. Chem. Inf. Model.* **2009**, *49*, 2873−2882.

(12) Aoki, W.; Ueda, M. Characterization of Antimicrobial Peptides Toward the Development of Novel Antibiotics. *Pharmaceuticals* **2013**, *6*, 1055−1081.

(13) Maloy, W. L.; Kari, U. P. Structure - Activity Studies on Magainins and other Host Defense Peptides. *Biopolymers* **1995**, *37*, 105−122.

(14) Tossi, A.; Tarantino, C.; Romeo, D. Design of Synthetic Antimicrobial Peptides Based on Sequence Analogy and Amphipathicity. *Eur. J. Biochem.* **1997**, *250*, 549−558.

(15) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977−5010.

(16) Taboureau, O. Methods for Building Quantitative Structure-Activity Relationship (QSAR) Descriptors and Predictive Models for Computer-Aided Design of Antimicrobial Peptides. In *Antimicrobial Peptides: Methods and Protocols*; Giuliani, A., Rinaldi, A. C., Eds.; Methods in Molecular Biology *618*; Springer: New York, 2010; pp 77−86.

(17) Tukey, J. W. Bias and Confidence in Not Quite Large Samples. *Ann. Math. Stat.* **1958**, *29*, 614.

(18) Fjell, C. D.; Jenssen, H.; Hilpert, K.; Warren, A.; 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*, 2006−2015.




DOI: 10.1021/acs.jcim.5b00161
*J. Chem. Inf. Model.* 2015, 55, 2275−2287

(19) Hilpert, K.; Fjell, C. D.; Cherkasov, A. Short Linear Cationic Antimicrobial Peptides: Screening, Optimizing, and Prediction. In *Peptide-Based Drug Design*; Otvos, L., Ed.; Methods in Molecular Biology *494*; Humana Press: New York, 2008; pp 127−159.
(20) Morikawa, N.; Hagiwara, K.; Nakajima, T. Brevinin-1 and -2, Unique Antimicrobial Peptides from the Skin of the Frog Rana brevipoda porsa. *Biochem. Biophys. Res. Commun.* **1992**, *189*, 184−190.
(21) Simmaco, M.; Mignogna, G.; Barra, D.; Bossa, F. Novel Antimicrobial Peptides from Skin Secretion of the European Frog Rana esculenta. *FEBS Lett.* **1993**, *324*, 159−161.
(22) Park, J. M.; Jung, J. E.; Lee, B. J. Antimicrobial Peptides from the Skin of a Korean Frog: Rana rugosa. *Biochem. Biophys. Res. Commun.* **1994**, *205*, 948−954.
(23) Clark, D. P.; Durell, S.; Maloy, W. L.; Zasloff, M. A Novel Antimicrobial Peptide from Bullfrog (*Rana catesbiana*) Skin, Structurally Related to the Bacterial Antibiotic, Polymyxin. *J. Biol. Chem.* **1994**, *269*, 10849−10855.
(24) Vignal, E.; Chavanieu, A.; Roch, P.; Chiche, L.; Grassy, G.; Calas, B.; Aumelas, A. Solution Structure of the Antimicrobial Peptide Ranalexin and a Study of its Interaction with Perdeuterated Dodecylphosphocholine Micelles. *Eur. J. Biochem.* **1998**, *253*, 221−228.
(25) Kwon, M. Y.; Hong, S. Y.; Lee, K. H. Structure−Activity Analysis of Brevinin 1E Amide, an Antimicrobial Peptide from Rana esculenta. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1998**, *1387*, 239−248.
(26) Won, H.-S.; Kang, S.-J.; Lee, B.-J. Action Mechanism and Structural Requirements of the Antimicrobial Peptides Gaegurins. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 1620−1629.
(27) Abraham, P.; Sundaram, A.; R, A.; V, R.; George, S.; Kumar, K. S. Structure-Activity Relationship and Mode of Action of a Frog Secreted Antibacterial Peptide B1CTcu5 Using Synthetically and Modularly Modified or Deleted (SMMD) Peptides. *PLoS One* **2015**, *10*, e0124210.
(28) Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; Kanehisa, M. AAindex: Amino Acid Index Database, Progress Report 2008. *Nucleic Acids Res.* **2008**, *36*, 202−205.
(29) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725−2729.
(30) Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **2011**, *28*, 2731−2739.
(31) Eisenberg, D.; Weiss, R. M.; Terwilliger, C. T.; Wilcox, W. Hydrophobic Moments and Protein Structure. *Faraday Symp. Chem. Soc.* **1982**, *17*, 109−120.
(32) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* **1984**, *179*, 125−142.
(33) Nguyen, L. T.; Haney, E. F.; Vogel, H. J. The Expanding Scope of Antimicrobial Peptide Structures and their Modes of Action. *Trends Biotechnol.* **2011**, *29*, 464−472.
(34) Han, X.; Kang, W. Sequence Analysis and Membrane Partitioning Energies of $\alpha$-Helical Antimicrobial Peptides. *Bioinformatics* **2004**, *20*, 970−973.
(35) Subasinghage, A. P.; Conlon, J. M.; Hewage, C. M. Conformational Analysis of the Broad-Spectrum Antibacterial Peptide, Ranatuerin-2CSa: Identification of a Full Length Helix-Turn-Helix Motif. *Biochim. Biophys. Acta, Proteins Proteomics* **2008**, *1784*, 924−929.
(36) Chi, S.-W.; Kim, J.-S.; Kim, D.-H.; Lee, S.-H.; Park, Y.-H.; Han, K.-H. Solution Structure and Membrane Interaction Mode of an Antimicrobial Peptide Gaegurin 4. *Biochem. Biophys. Res. Commun.* **2007**, *352*, 592−597.
(37) Haney, E. F.; Hunter, H. H.; Matsuzaki, K.; Vogel, H. J. Solution NMR Studies of Amphibian Antimicrobial Peptides: Linking Structure to Function? *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 1639−1655.
(38) Dennison, S. R.; Morton, L. H. G.; Phoenix, D. A. Role of Molecular Architecture on the Relative Efficacy of Aurein 2.5 and Modelin. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 2094−2102.
(39) Park, S.-H.; Kim, H.-E.; Kim, C.-M.; Yun, H.-J.; Choi, E.-C.; Lee, B.-J. Role of Proline, Cysteine and a Disulphide Bridge in the Structure and Activity of the Anti-Microbial Peptide Gaegurin 5. *Biochem. J.* **2002**, *368*, 171−182.
(40) Park, S.-H; Son, W.-S.; Kim, Y.-J.; Kwon, A.-R.; Lee, B.-J. NMR Spectroscopic Assessment of the Structure and Dynamic Properties of an Amphibian Antimicrobial Peptide (Gaegurin 4) Bound to SDS Micelles. *J. Biochem. Mol. Biol.* **2007**, *40*, 261−269.
(41) Nakai, K.; Kidera, A.; Kanehisa, M. Cluster Analysis of Amino Acid Indices for Prediction of Protein Structure and Function. *Protein Eng.* **1988**, *2*, 93−100.
(42) Tossi, A.; Sandri, L.; Giangaspero, A. New Consensus Hydrophobicity Scale Extended to Non-Proteinogenic Amino Acids. In *Proceedings of the 27th European Peptide Symposium, Peptides 2002*, Sorrento, Aug 31−Sept 6, 2002; Benedetti E., Pedone C. Eds.; Edizioni Ziino: Napoli, Italy, 2002; pp 416−417.
(43) Irisarri, I.; San Mauro, D.; Abascal, F.; Ohler, A.; Vences, M.; Zardoya, R. The Origin of Modern Frogs (Neobatrachia) was Accompanied by Acceleration in Mitochondrial and Nuclear Substitution Rates. *BMC Genomics* **2012**, *13*, 626.
(44) Novković, M.; Simunić, J.; Bojović, V.; Tossi, A.; Juretić, D. DADP: The Database of Anuran Defense Peptides. *Bioinformatics* **2012**, *28*, 1406−1407.
(45) Riedl, S.; Zweytick, D.; Lohner, K. Membrane-Active Host Defense Peptides—Challenges and Perspectives for the Development of Novel Anticancer Drugs. *Chem. Phys. Lipids* **2011**, *164*, 766−781.
(46) Bowie, J. H.; Separovic, F.; Tyler, M. J. Host-Defense Peptides of Australian Anurans. Part 2. Structure, Mechanism of Action, and Evolutionary Significance. *Peptides* **2012**, *37*, 174−188.
(47) Resende, J. M.; Moraes, C. M.; Prates, M. V.; Cesar, A.; Almeida, F. C.; Mundim, N. C.; Valente, A. P.; Bemquerer, M. P.; Piló-Veloso, D.; Bechinger, B. Solution NMR Structures of the Antimicrobial Peptides Phylloseptin-1, -2, and -3 and Biological Activity: The Role of Charges and Hydrogen Bonding Interactions in Stabilizing Helix Conformations. *Peptides* **2008**, *29*, 1633−1644.
(48) Ali, M. F.; Knoop, F. C.; Vaudry, H.; Conlon, J. M. Characterization of Novel Antimicrobial Peptides from the Skins of Frogs of the *Rana esculenta* Complex. *Peptides* **2003**, *24*, 955−961.
(49) Ali, M. F.; Lips, K. R.; Knoop, F. C.; Fritzsch, B.; Miller, C.; Conlon, J. M. Antimicrobial Peptides and Protease Inhibitors in the Skin Secretions of the Crawfish Frog *Rana areolata*. *Biochim. Biophys. Acta, Proteins Proteomics* **2002**, *1601*, 55−63.
(50) Suzuki, H.; Iwamuro, S.; Ohnuma, A.; Coquet, L.; Leprince, J.; Jouenne, T.; Vaudry, H.; Taylor, C. K.; Abel, P. W.; Conlon, J. M. Expression of Genes Encoding Antimicrobial and Bradykinin-Related Peptides in Skin of the Stream Brown Frog *Rana sakuraii*. *Peptides* **2007**, *28*, 505−514.
(51) Conlon, J. M.; Sonnevend, A.; Patel, M.; Al-Dhaheri, K.; Nielsen, P. F.; Kolodziejek, J.; Nowotny, N.; Iwamuro, S.; Pál, T. A Family of Brevinin-2 Peptides with Potent Activity Against *Pseudomonas aeruginosa* from the Skin of the Hokkaido Frog, *Rana pirica*. *Regul. Pept.* **2004**, *118*, 135−141.
(52) Iwakoshi-Ukena, E.; Ukena, K.; Okimoto, A.; Soga, M.; Okada, G.; Sano, N.; Fujii, T.; Sugawara, Y.; Sumida, M. Identification and Characterization of Antimicrobial Peptides from the Skin of the Endangered Frog *Odorrana ishikawae*. *Peptides* **2011**, *32*, 670−676.
(53) Conlon, J. M.; Mechkarska, M.; Ahmed, E.; Coquet, L.; Jouenne, T.; Leprince, J.; Vaudry, H.; Hayes, M. P.; Padgett-Flohr, G. Host Defense Peptides in Skin Secretions of the Oregon Spotted Frog *Rana pretiosa*: Implications for Species Resistance to Chytridiomycosis. *Dev. Comp. Immunol.* **2011**, *35*, 644−649.
(54) Conlon, J. M.; Kolodziejek, J.; Nowotny, N.; Leprince, J.; Vaudry, H.; Coquet, L.; Jouenne, T.; King, J. D. Characterization of Antimicrobial Peptides from the Skin Secretions of the Malaysian Frogs, *Odorrana hosii* and *Hylarana picturata* (Anura:Ranidae). *Toxicon* **2008**, *52*, 465−473.


DOI: 10.1021/acs.jcim.5b00161
*J. Chem. Inf. Model.* 2015, 55, 2275−2287

(55) Conlon, J. M.; Mechkarska, M.; Coquet, L.; Jouenne, T.; Leprince, J.; Vaudry, H.; Kolodziejek, J.; Nowotny, N.; King, J. D. Characterization of Antimicrobial Peptides in Skin Secretions from Discrete Populations of *Lithobates chiricahuensis* (Ranidae) from Central and Southern Arizona. *Peptides* **2011**, *32*, 664−669.
(56) Conlon, J. M.; Abraham, B.; Sonnevend, A.; Jouenne, T.; Cosette, P.; Leprince, J.; Vaudry, H.; Bevier, C. R. Purification and Characterization of Antimicrobial Peptides from the Skin Secretions of the Carpenter Frog *Rana virgatipes* (Ranidae, Aquarana). *Regul. Pept.* **2005**, *131*, 38−45.
(57) Liu, J.; Jiang, J.; Wu, Z.; Xie, F. Antimicrobial Peptides from the Skin of the Asian Frog, *Odorrana jingdongensis*: De novo Sequencing and Analysis of Tandem Mass Spectrometry Data. *J. Proteomics* **2012**, *75*, 5807−5821.
(58) Conlon, J. M.; Sonnevend, A.; Patel, M.; Davidson, C.; Nielsen, P. F.; Pál, T.; Rollins-Smith, L. A. Isolation of Peptides of the Brevinin-1 Family with Potent Candidacidal Activity from the Skin Secretions of the Frog *Rana boylii*. *J. Pept. Res.* **2003**, *62*, 207−213.
(59) Pál, T.; Abraham, B.; Sonnevend, A.; Jumaa, P.; Conlon, J. M. Brevinin-1BYa: A Naturally Occurring Peptide from Frog Skin with Broad-Spectrum Antibacterial and Antifungal Properties. *Int. J. Antimicrob. Agents* **2006**, *27*, 525−529.
(60) Sonnevend, A.; Knoop, F. C.; Patel, M.; Pál, T.; Soto, A. M.; Conlon, J. M. Antimicrobial Properties of the Frog Skin Peptide, Ranatuerin-1 and its [Lys-8]-Substituted Analog. *Peptides* **2004**, *25*, 29−36.
(61) Conlon, J. M.; Ahmed, E.; Coquet, L.; Jouenne, T.; Leprince, J.; Vaudry, H.; King, J. D. Peptides with Potent Cytolytic Activity from the Skin Secretions of the North American Leopard Frogs, *Lithobates blairi* and *Lithobates yavapaiensis*. *Toxicon* **2009**, *53*, 699−705.
(62) Conlon, J. M.; Sonnevend, A.; Davidson, C.; Demandt, A.; Jouenne, T. Host-Defense Peptides Isolated from the Skin Secretions of the Northern Red-Legged Frog *Rana aurora aurora*. *Dev. Comp. Immunol.* **2005**, *29*, 83−90.
(63) Conlon, J. M.; al-Dhaheri, A.; al-Mutawa, E.; al-Kharrge, R.; Ahmed, E.; Kolodziejek, J.; Nowotny, N.; Nielsen, P. F.; Davidson, C. Peptide Defenses of the Cascades Frog *Rana cascadae*: Implications for the Evolutionary History of Frogs of the *Amerana* Species Group. *Peptides* **2007**, *28*, 1268−1274.
(64) Conlon, J. M.; Raza, H.; Coquet, L.; Jouenne, T.; Leprince, J.; Vaudry, H.; King, J. D. Purification of Peptides with Differential Cytolytic Activities from the Skin Secretions of the Central American Frog, *Lithobates vaillanti* (Ranidae). *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.* **2009**, *150*, 150−154.
(65) Morgera, F.; Antcheva, N.; Pacor, S.; Quaroni, L.; Berti, F.; Vaccari, L.; Tossi, A. Structuring and Interactions of Human Defensins 2 and 3 with Model Membranes. *J. Pept. Sci.* **2008**, *14*, 518−523.
(66) Giangaspero, A.; Sandri, L.; Tossi, A. Amphipathic $\alpha$-Helical Antimicrobial Peptides. A Systematic Study of the Effects of Structural and Physical Properties on Biological Activity. *Eur. J. Biochem.* **2001**, *268*, 5589−5600.
(67) Dennison, S. R.; Wallace, J.; Harris, F.; Phoenix, D. A. Amphiphilic $\alpha$-Helical Antimicrobial Peptides and their Structure/Function Relationships. *Protein Pept. Lett.* **2005**, *12*, 31−39.
(68) Huang, Y.; Huang, J.; Chen, Y. Alpha-Helical Cationic Antimicrobial Peptides: Relationships of Structure and Function. *Protein Cell* **2010**, *1*, 143−152.
(69) Lehrer, R. I. Evolution of Antimicrobial Peptides: A View from the Cystine Chapel. In *Antimicrobial Peptides and Innate Immunity, Progress in Inflamation Research*; Hiemstra, P. S., Zaat, S. A. J., Eds.; Springer: Basel, Switzerland, 2013; pp 1−27.
(70) Giacometti, A.; Cirioni, O.; Barchiesi, F.; Del Prete, M. S.; Fortuna, M.; Caselli, F.; Scalise, G. In Vitro Susceptibility Tests for Cationic Peptides: Comparison of Broth Microdilution Methods for Bacteria that Grow Aerobically. *Antimicrob. Agents Chemother.* **2000**, *44*, 1694−1696.
(71) Uggerhøj, L. E.; Poulsen, T. J.; Munk, J. K.; Fredborg, M.; Sondergaard, T. E.; Frimodt-Müller, N.; Hansen, P. R.; Wimmer, R. Rational Design of alpha-Helical Antimicrobial Peptides: Do's and Don'ts. *ChemBioChem* **2015**, *16*, 242−253.
(72) Grantham, R. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science* **1974**, *185*, 862−864.
(73) Zhou, H.; Zhou, Y. Quantifying the Effect of Burial of Amino Acid Residues on Protein Stability. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 315−322.
(74) Krigbaum, W. R.; Komoriya, A. Local Interactions as a Structure Determinant for Protein Molecules: II. *Biochim. Biophys. Acta, Protein Struct.* **1979**, *576*, 204−228.
(75) *Handbook of Biochemistry and Molecular Biology—Proteins*, 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1976; Vol. *1*, pp 111−174.
(76) Wertz, D. H.; Scheraga, H. A. Influence of Water on Protein Structure. An Analysis of the Preferences of Amino Acid Residues for the Inside or Outside and for Specific Conformations in a Protein Molecule. *Macromolecules* **1978**, *11*, 9−15.
(77) Fodje, M. N.; al-Karadaghi, S. Occurrence, Conformational Features and Amino Acid Propensities for the $\pi$-Helix. *Protein Eng.* **2002**, *15*, 353−358.
(78) Cartailler, J.-P.; Luecke, H. Structural and Functional Characterization of $\pi$-Bulges and Other Short Intrahelical Deformations. *Structure* **2004**, *12*, 133−144.
(79) Kim, S.; Cross, T. A. 2D Solid State NMR Spectral Simulation of $3_{10}$, $\alpha$ and $\pi$-Helices. *J. Magn. Reson.* **2004**, *168*, 187−193.
(80) *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Press: New York, 1989; p 457.
(81) Walters, R. F.; DeGrado, W. F. Helix-Packing Motifs in Membrane Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13658−13663.
(82) Juretić, D.; Zoranić, L.; Zucić, D. Basic Charge Clusters and Predictions of Membrane Protein Topology. *J. Chem. Inf. Model.* **2002**, *42*, 620−632.
(83) Gautier, R.; Douguet, D.; Antonny, B.; Drin, G. HELIQUEST: A Web Server to Screen Sequences with Specific $\alpha$-Helical Properties. *Bioinformatics* **2008**, *24*, 2101−2102.
(84) Fauchère, J.-L.; Pliska, V. Hydrophobic Parameters $\pi$ of Amino Acid Side Chains from the Partitioning of N-acetyl-amino-acid Amides. *Eur. J. Med. Chem.* **1983**, *18*, 369−375.
(85) Sneath, P. H. Relations between Chemical Structure and Biological Activity in Peptides. *J. Theor. Biol.* **1966**, *12*, 157−195.
(86) He, Y.; Lazaridis, T. Activity Determinants of Helical Antimicrobial peptides. A Large Scale Computational Study. *PLoS One* **2013**, *8*, e66440.
(87) Zhou, M.; Liu, Y.; Chen, T.; Fang, X.; Walker, B.; Shaw, C. Components of the Peptidome and Transcriptome Persist in *li wan pi*: The Dried Skin of the Heilongjiang Brown Frog (*Rana amurensis*) as Used in Traditional Chinese Medicine. *Peptides* **2006**, *27*, 2688−2694.
(88) Ilić, N.; Novković, M.; Guida, F.; Xhindoli, D.; Benincasa, M.; Tossi, A.; Juretić, D. Selective Antimicrobial Activity and Mode of Action of Adepantins, Glycine-Rich Peptide Antibiotics Based on Anuran Antimicrobial Peptide Sequences. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828*, 1004−1012.
(89) Kamech, N.; Vukičević, D.; Ladram, A.; Piesse, C.; Vasseur, J.; Bojović, V.; Simunić, J.; Juretić, D. Improving the Selectivity of Antimicrobial Peptides from Anuran Skin. *J. Chem. Inf. Model.* **2012**, *52*, 3341−3351.
(90) Fernández-Vidal, M.; Jayasinghe, S.; Ladokhin, A. S.; White, S. H. Folding Amphipathic Helices into Membranes: Amphiphilicity Trumps Hydrophobicity. *J. Mol. Biol.* **2007**, *370*, 459−470.
(91) Bergen, P. J.; Landersdorfer, C. B.; Lee, H. J.; Li, J.; Nation, R. L. 'Old' Antibiotics for Emerging Multidrug-Resistant Bacteria. *Curr. Opin. Infect. Dis.* **2012**, *25*, 626−633.
(92) Lim, L. M.; Ly, N.; Anderson, D.; Yang, J. C.; Macander, L.; Jarkowski, A., III; Forrest, A.; Bulitta, J. B.; Tsuji, B. T. Resurgence of Colistin: A Review of Resistance, Toxicity, Pharmacodynamics, and Dosing. *Pharmacotherapy* **2010**, *30*, 1279−1291.


DOI: 10.1021/acs.jcim.5b00161
*J. Chem. Inf. Model.* 2015, 55, 2275−2287

(93) Wertheim, H.; Nguyen, K. V.; Hara, G. L.; Gelband, H.; Laxminarayan, R.; Mouton, J.; Cars, O. Global Survey of Polymyxin Use: A Call for International Guidelines. *J. Glob. Antimicrob. Resist.* **2013**, *1*, 131−134.
(94) Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. A New Antibiotic Kills Pathogens Without Detectable Resistance. *Nature* **2015**, *517*, 455−459.



DOI: 10.1021/acs.jcim.5b00161
*J. Chem. Inf. Model.* 2015, 55, 2275−2287