



# Protegrin structure–activity relationships: using homology models of synthetic sequences to determine structural characteristics important for activity

Nathan Ostberg, Yiannis Kaznessis\*

*Department of Chemical Engineering and Materials Science, and Digital Technology Center, University of Minnesota,  
421 Washington Ave SE, Minneapolis, MN 55455, USA*

Received 30 July 2004; received in revised form 24 September 2004; accepted 28 September 2004

## Abstract

The protegrin family of antimicrobial peptides is among the shortest in sequence length while remaining very active against a variety of microorganisms. The major goal of this study is to characterize easily calculated molecular properties, which quantitatively show high correlation with antibacterial activity. The peptides studied have high sequence similarity but vary in activity over more than an order of magnitude. Hence, sequence analysis alone cannot be used to predict activity for these peptides. We calculate structural properties of 62 protegrin and protegrin-analogue peptides and correlate them to experimental activities against six microbe species, as well as hemolytic and cytotoxic activities. Natural protegrins structures were compared with synthetic derivatives using homology modeling, and property descriptors were calculated to determine the characteristics that confer their antimicrobial activity. A structure–activity relationship study of all these peptides provides information about the structural properties that affect activity against different microbial species.

© 2004 Elsevier Inc. All rights reserved.

*Keywords:* Antimicrobial peptide; Homology modeling; Structure–activity relationships

## 1. Introduction

### 1.1. Antimicrobial peptides

For the last two decades, antimicrobial peptides have been gaining recognition as highly valuable therapeutic agents in fighting a wide range of microorganisms. Some antimicrobial peptides (AMPs) have been found to be antibacterial, antiviral, antifungal, anti-cancer, or promoters of wound healing [19]. AMPs have been found in organisms as simple as bacteria to those as complicated as humans including nearly everything in between. These naturally occurring peptides have been preserved through evolution with many characteristics of individual classes of AMPs consistent across differ-

ent species. They have been favored through evolution for their small size and their high potency [7,18].

In most cases an AMP first locally disrupts a target cell's outer membrane. The anionic lipopolysaccharide phospholipids that make up the outer membrane are locally in equilibrium with divalent calcium and magnesium cations. The predominantly cationic AMPs have been shown [8] to have a higher affinity for LPS, displacing  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions. This destabilization of the outer membrane by AMPs has been termed self-promoted uptake [9]. From there, the peptide may create pores in the cytoplasmic membrane and cause disruption resulting in cell lyses, or the peptide may actually move into the interior of the cell and cause disruption there [19,4].

Antimicrobial peptides are especially valuable because they have been shown in many cases to be unaffected by the antibiotic resistance that renders small drug molecules

\* Corresponding author. Tel.: +1 612 624 4197; fax: +1 612 626 7246.  
E-mail address: [yiannis@cems.umn.edu](mailto:yiannis@cems.umn.edu) (Y. Kaznessis).

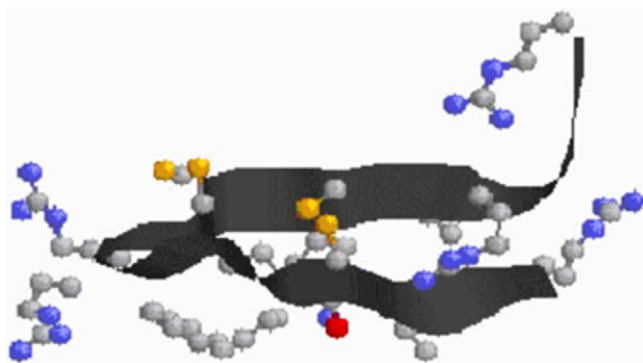


Fig. 1. Structure of protegrin-1. Backbone in dark grey and side chains (excluding hydrogen) in CPK.

useless after some time; this was shown for protegrins in [2]. AMPs do not interact with receptors or other specific sites on their target. Indeed it is highly unlikely, if impossible, for any organism to develop resistance to AMPs, because it is an electrostatic interaction that first brings an AMP to the surface of a cell membrane. Evolutionary time scales would be needed for microbial species to completely alter the LPS membrane composition. Moreover AMPs are extremely fast acting [7], hence antimicrobial peptides are an excellent candidate for therapeutic purposes.

Finally antimicrobial peptides are attractive for study with computational methods because they are small, most less than 50 residues long. This allows for quicker simulations and minimizations. The confidence in homology modeling increases with fewer residues to fit. Also experimental verification of results is faster and less costly than a larger protein, simply because of its size.

### 1.2. Protegrins

Protegrins are a family of five antimicrobial peptides naturally found in porcine neutrophils, which have cathelicidin-like precursors. Protegrins PG-1, PG-2, and PG-3 were purified initially [13], with PG-4 predicted from a cDNA clone [25] and PG-5 from a genomic clone [26]. The peptides are translated as inactive pro-peptides; extra cellular processing removes the pro-region, resulting in the release of active peptides. Protegrins are similar to other defensins with two disulfide bonds and a  $\beta$ -sheet structure, and they are between sixteen and eighteen residues long. The structure of protegrin-1, available publicly in the protein databank [1], is shown in Fig. 1.

Protegrins have been shown to act independently of any stereo-specific interaction, as all D-amino acid isomers are as active as the natural peptides. These peptides have shown widespread activity against Gram-positive and Gram-negative bacteria and yeast. Protegrins typically kill bacteria on the order of minutes, again making them attractive therapeutic targets.

The natural protegrins and over 50 synthetic sequences have been tested for activity against many microbial species

(previously unpublished data from Dr. R. Lehrer at UCLA, Table 1). Among all of these sequences, the minimum sequence identity between two peptides is 22%, and on average the sequences vary by 68% of their residues.

For the microbial species, the minimum inhibitory concentration was determined for these peptides (not all peptides tested against all species). In MIC ( $\mu\text{g}/\text{mL}$ ) the lower values in the concentration indicate a more active peptide. Studies have shown that a minimum peptide to membrane lipid ratio must be achieved before a protegrin can destroy a cell [10,20].

For cytotoxicity and hemolysis, the degree of activity was determined for a given concentration; a lower number here means the peptides kill fewer human cells. Therefore, the ideal peptide will have both a low value for MIC against a microorganism, as well as a low value for the amount of host cells that can be expected to suffer.

While the synthetic sequences differ in only a few residue mutations at a time, the peptides have widely varying structural properties enabling the set to cover a range of activities. Previous quantitative structure–activity relationship (QSAR) studies [2] qualitatively described characteristics of the protegrin sequences that made them more or less active (such as: those with four cysteines and two disulfide bonds are typically more active than the sequences with only one of these bonds; higher positive charge imparts greater activity to a point). However, we wish to quantitatively determine what properties influence protegrin activity and develop models for structure–activity relationships for protegrins against the wide spectrum of microorganisms studied.

## 2. Materials and methods

### 2.1. Protegrin sequences

Numerous experiments have been conducted on the activity of natural and synthetic protegrins by Lehrer and coworkers [11,13,15,22,24]. Other research efforts have gone into similar studies of the protegrins and related peptides [2,23,14]. For a description of all the experimental methods the reader is referred to [22]. In this study, we combine the results of activity studies on the five natural protegrins and over 50 synthetic peptides summarized in Table 1. These peptides were tested against Gram-negative bacteria (*Escherichia coli*, *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae*—two strains), Gram-positive bacteria (*Listeria monocytogenes*), and the yeast *Candida albicans* (not every peptide was tested against every organism). The peptides were also tested against cervical epithelial cells for cytotoxicity and red blood cells for hemolytic activity.

### 2.2. Homology modeling

The crystal structure of protegrin-1, PG-1, has been resolved and is publicly available in the protein databank [1], [www.rcsb.org/pdb](http://www.rcsb.org/pdb), under the PDB identification 1PG-1. The

Table 1  
Peptide activities (MIC in  $\mu\text{g/mL}$ )

| Name   | Sequence         | <i>E. coli</i> | <i>N. gonorrhoeae</i><br>(F-62) | <i>N. gonorrhoeae</i><br>(FA-19) | <i>L. monocytogenes</i> | <i>C. albicans</i> | <i>P. aeruginosa</i> | Cytotoxicity | Hemolysis  |
|--------|------------------|----------------|---------------------------------|----------------------------------|-------------------------|--------------------|----------------------|--------------|------------|
| PC001  | rggrlcyrrfvevgr  | 0.9            | 1.2                             | 1.7                              | 0.9                     | 5.3                | 0.9                  | 4            | 5          |
| PC003  | rggglcyrrfvevgr  | 2.3            | 1.4                             | 1.8                              | 2.8                     | 9.6                | Not tested           | 4            | 5          |
| PC004  | rggrlcyrgwicfvgr | 2.1            | 1.2                             | 3.5                              | 1.9                     | 30.4               | Not tested           | Not tested   | 5          |
| PC005  | rggrlcyrrfvevgr  | 2.9            | 1.4                             | 1.6                              | 2.5                     | 15.6               | Not tested           | 3            | 5          |
| PC006  | rggrlayrrfvevgr  | 2.1            | 1.1                             | 5.5                              | 2.6                     | 16                 | Not tested           | 4            | 2          |
| PC007  | rggrlcyarrfvevgr | Not tested     | 5.9                             | 32.7                             | Not tested              | 18.5               | Not tested           | 2            | 1          |
| PC009  | leyrrfvevgr      | 2.2            | 1                               | 6.3                              | 23.3                    | 61.9               | Not tested           | 2            | 0          |
| PC010  | rcyrrfvevgr      | 1.1            | 0.8                             | 1.8                              | 7.9                     | 31.4               | Not tested           | 4            | 0          |
| PC011  | rggrlcyrrfvev    | 1.3            | 1.5                             | 3.8                              | 3.7                     | 11.6               | Not tested           | 5            | 5          |
| PC012  | rggrlcyrrfvev    | 1.9            | 1.3                             | 4.3                              | 5.5                     | 20.5               | 3.1                  | 4            | 3          |
| PC013  | rggrlcyrrfvev    | 2.3            | 1.2                             | 1.9                              | 2.7                     | 10.2               | Not tested           | 5            | 5          |
| PC014  | rcyrrfvev        | 2.9            | 1.1                             | 1.4                              | 3.3                     | 10.1               | Not tested           | 4            | 5          |
| PC015  | leyrrfvev        | 1.5            | 2.3                             | 9.8                              | 2.4                     | 8.8                | 3                    | 3            | 1          |
| PC016  | lcyrarrfvev      | 0.6            | 13.6                            | 48.4                             | 1.5                     | 25.5               | 0.8                  | 2            | 0          |
| PC017  | rcyarrfvev       | 1              | 1.2                             | 2.4                              | 0.7                     | 10.5               | 0.6                  | 3            | 5          |
| PC018  | layrrfvev        | 1.2            | 1.6                             | 11.8                             | 1.4                     | 80–250             | Not tested           | 2            | 1          |
| PC019  | rayrrfvev        | 3              | 137.3                           | >500                             | 29                      | 29.4               | Not tested           | 0            | 0          |
| PC020  | cycrrfvevgr      | 1.9            | 3.5                             | 52.8                             | 21.7                    | 67.5               | Not tested           | 1            | 0          |
| PC021  | rggrlcyrrfvev    | 1.8            | >500                            | >500                             | 36.3                    | >250               | Not tested           | 0.5          | 0          |
| PC037  | lcyrtrfvev       | 1.4            | 1.8                             | 10.1                             | 4.2                     | 27.5               | 1.4                  | 3            | 1          |
| PC045  | ltyrrfvev        | 0.5            | 2                               | 10.1                             | 0.3                     | 2.7                | 0.5                  | 3            | 1          |
| PC064  | lcyrtrfvev       | 1.4            | 1.3                             | 13.8                             | 2                       | 28.4               | Not tested           | 3            | 1          |
| PC064a | lcyrtrfvev       | 0.6            | 0.7                             | 5.3                              | 3.9                     | 12.2               | Not tested           | 2            | 0          |
| PC065  | leytrfvev        | 3.1            | 4.1                             | 30                               | 10.5                    | >250               | Not tested           | 0.5          | 0          |
| PC066  | leytrfvev        | Not tested     | 59.5                            | 158                              | Not tested              | Not tested         | Not tested           | 2            | 0          |
| PC069  | cycrrfvev        | >80            | 38.1                            | 127.6                            | 80–250                  | >250               | Not tested           | 0.5          | 0          |
| PC070  | leyrrfvev        | 2.7            | 3.2                             | 32.3                             | 26.1                    | 80–250             | Not tested           | 1            | 0          |
| PC071  | cycrrfvev        | 3.2            | 4.2                             | 13                               | 4                       | 31.9               | 9.4                  | 3            | 0          |
| PC072  | cycrrfvev        | 0.6            | 1.8                             | 8.4                              | 0.7                     | 28.4               | 1.2                  | 3            | 2          |
| PC073  | leyrrfvev        | 14.1           | 31.7                            | 96.9                             | Not tested              | >79.1              | 7.1                  | 1            | 0          |
| PC074  | leyrrfvev        | 8.1            | 1.5                             | 11.2                             | Not tested              | 27.9               | 9.7                  | 2            | 0          |
| PC077  | leyrrfvev        | 1              | Not tested                      | Not tested                       | 3.5                     | 34.9               | 3.1                  | 3            | 1          |
| PC078  | leyrrfvev        | 1.8            | Not tested                      | Not tested                       | 1.8                     | >79.1              | 3.7                  | 0.5          | 3          |
| PC079  | ycyrrfvevgr      | 3.5            | Not tested                      | Not tested                       | 8.8                     | >79.1              | 16                   | 0.5          | 1          |
| PC080  | tcyrrfvevgr      | 1.1            | Not tested                      | Not tested                       | 3.6                     | >79.1              | 4.1                  | 0.5          | 3          |
| PC091  | acyrrfvevgr      | 0.7            | Not tested                      | Not tested                       | 1.5                     | 8.9                | Not tested           | 4            | 4          |
| PC092  | rggrlcyrrfvevgr  | 0.7            | Not tested                      | Not tested                       | 1.4                     | 11.2               | Not tested           | 3            | 3          |
| PC093  | icyrrfvevgr      | 0.7            | Not tested                      | Not tested                       | 1.4                     | 11.7               | Not tested           | 4            | 4          |
| PC094  | fcyrrfvevgr      | 0.6            | Not tested                      | Not tested                       | 1.3                     | 8.7                | Not tested           | 4            | 5          |
| PC095  | wcyrrfvevgr      | 0.5            | Not tested                      | Not tested                       | 1.4                     | 9.9                | Not tested           | 4            | 5          |
| PC096  | ecyrrfvevgr      | 0.6            | Not tested                      | Not tested                       | 1.2                     | 9.3                | Not tested           | 4            | 5          |
| PC097  | rggrlcyrrfvev    | 0.6            | Not tested                      | Not tested                       | 1.1                     | 10.7               | Not tested           | 4            | 5          |
| PC098  | rggrlcyrrfvev    | 0.7            | Not tested                      | Not tested                       | Not tested              | 22                 | 4.3                  | 3            | 1          |
| PC100  | rggrlcyrrfvev    | 0.8            | Not tested                      | Not tested                       | Not tested              | 3.9                | 1.2                  | 4            | 5          |
| PC101  | rggrlcyrrfvev    | 0.7            | Not tested                      | Not tested                       | Not tested              | 4.4                | 1.4                  | 2            | 3          |
| PC102  | rggrlcyrrfvev    | 0.8            | Not tested                      | Not tested                       | Not tested              | 3.6                | 1.2                  | 3            | 4          |
| PC103  | rggrlcyrrfvev    | 0.7            | Not tested                      | Not tested                       | Not tested              | 1.5                | 0.7                  | 4            | 5          |
| PC104  | rggrlcyrrfvev    | 0.6            | Not tested                      | Not tested                       | Not tested              | 3.6                | 1                    | 4            | 5          |
| PC105  | rggrlcyrrfvev    | 1              | Not tested                      | Not tested                       | Not tested              | 4.4                | 0.7                  | 3            | 5          |
| PC106  | rggrlcyrrfvev    | 0.8            | Not tested                      | Not tested                       | Not tested              | 7.5                | 1.4                  | 3            | 5          |
| PC107  | rlcytrgrfvev     | 0.6            | Not tested                      | Not tested                       | Not tested              | 31.8               | 0.5                  | 2            | 0          |
| PC108  | lcyrtrgrfvev     | 0.3            | Not tested                      | Not tested                       | Not tested              | 3                  | 0.4                  | Not tested   | Not tested |
| PC109  | rlcytrgrfvev     | Not tested     | Not tested                      | Not tested                       | Not tested              | Not tested         | Not tested           | 4            | 3          |
| PC110  | lcychhhfvev      | Not tested     | Not tested                      | Not tested                       | Not tested              | Not tested         | Not tested           | 3            | 0          |
| PC111  | lcythhhfvev      | Not tested     | Not tested                      | Not tested                       | Not tested              | Not tested         | Not tested           | 4            | 3          |
| PC112  | leyrrfvev        | 5.5            | Not tested                      | Not tested                       | Not tested              | >250               | 6                    | 0            | 2          |
| PC113  | leyrrfvev        | 10.8           | Not tested                      | Not tested                       | Not tested              | >250               | 23.7                 | 0            | 0          |
| PC146  | lcyrtrfvev       | 1              | Not tested                      | Not tested                       | Not tested              | 11                 | 1.4                  | Not tested   | 2          |
| PC147  | leyrrfvev        | 0.7            | Not tested                      | Not tested                       | Not tested              | 9.6                | 0.9                  | Not tested   | 5          |
| PC148  | lcyrtrfvev       | 1.3            | Not tested                      | Not tested                       | Not tested              | 25                 | 2.6                  | Not tested   | 0          |
| PC149  | rggrlcyrrfvevgr  | 1.2            | Not tested                      | Not tested                       | Not tested              | 28.4               | 1                    | Not tested   | 0          |
| PC150  | rggglcyrrfvevgr  | 0.8            | Not tested                      | Not tested                       | Not tested              | 10.2               | 0.9                  | Not tested   | 4          |

sequence of this peptide is RGGRLCYCRRRFCVGVGR\* with the asterisk denoting a C-terminal NH<sub>2</sub> which is found on all of the protegrins studied.

The peptides were studied using the Molecular Operating Environment (MOE) software [16]. As PG-1 is the only structure publicly available, this file was used as the template for all of the other sequences for homology modeling. Each synthetic sequence was aligned to PG-1 (in the PDB file for PG-1, the C-terminal amine is counted as a nineteenth residue; the residue was deleted and then the NH<sub>2</sub> group was added onto the final arginine residue). The target sequences have an average sequence identity of 67% with the template, and the minimum sequence identity was 39%. Sequences this similar typically are acceptable for homology modeling. Importantly, the location of cysteine residues (two or four, with one or two disulfide bonds) is absolutely conserved. There was one sequence (PC-8) tested with alanine replacing all four cysteine residues; while expecting a nearly linear structure, homology modeling with MOE did not allow the synthetic structure to relax away from the U-shaped beta sheet of PG-1. Having low confidence in this structure, it was discarded in the rest of the analysis.

Using PG-1 as a template and the synthetic sequences fit with the corresponding residue substitutions, MOE creates homology models that find the lowest energy structure for the new sequence, allowing the synthetic structures to relax to some energy minimum. The default MOE homology model settings were used with an AMBER89 force field for the potentials. In this method for each peptide, MOE creates ten models placing the substituted residues at random orientations. Each of these models is minimized and the lowest energy structure is then taken to be the best. As each of the sequences tested had a C-terminal NH<sub>2</sub> group, this was added manually for each structure. An additional minimization step was run, and the relaxed structure that resulted was used in further calculations for each peptide.

### 2.3. Structural properties

Using MOE a descriptor database was developed which characterized the natural and novel structures. After creating each structure using homology modeling, the peptide was loaded into a MOE database. Twenty six properties were calculated with the methods MOE uses which are described below. Pairwise correlation coefficients were calculated and eight of the properties were eliminated as they behaved covariantly with other properties. A general cutoff of 0.95 was used, although some properties with higher covariance were kept because they provide different levels of insight. For example the principal moment of inertia is highly correlated with size and weight of the molecule, but since it varies by the geometry of the peptide, a property not included in these descriptors, we leave the moment of inertia in the analysis. Table 2 lists the properties from MOE that were calculated along with the abbreviations used later in the models.

Table 2

Properties calculated in MOE

|   |      |
|---|------|
| Molecular weight                          | MWEI |
| Formal charge                             | CHRG |
| Solvent accessible surface area           | SASA |
| Hydrophobic component of SASA             | FOSA |
| Negative component of SASA                | NESA |
| Positive component of SASA                | POSA |
| Electrostatic portion of potential energy | EELE |
| Energy change of solvation in water       | ESOL |
| van de Waals portion of potential energy  | EVDW |
| KierFlex; flexibility of the peptide      | FLEX |
| Dipole moment of peptide                  | DIPO |
| Principle moment of inertia               | PMOI |
| Number of hydrogen bond acceptors         | HBAC |
| Number of hydrogen bond donors            | HBDN |
| Molecular volume                          | MVOL |
| Density of the peptide                    | DENS |
| Globularity of peptide                    | GLOB |
| Octanol/water partition coefficient       | LOGP |

Molecular weight is calculated based on the atoms that MOE sees, not from a reference value for the residues.

Formal charge is determined for the specific force field (here AMBER89), summing up each 'q' charge assigned to each atom in the structure.

Solvent accessible surface area is calculated using water as a solvent with a radius of 1.4 Å for the water molecule. Hydrophobic ( $|q| < 0.2$ ), negative ( $q < -0.2$ ), and positive ( $q > 0.2$ ) surface areas are calculated with the force field assigned charges and the van der Waals surface area for each atom. The accessible portion of the surface area is determined using a connection table approximation; MOE implicitly records what is bound to each atom and has stored values for the combinations of possible neighbors.

The electrostatic and van der Waals portions of the potential energy and the dipole moment of the molecule rely on the force field chosen to determine partial charges. The solvation energy is for water.

The KierFlex number is a measure of flexibility of the molecule as described in [5].

The principal moment of inertia can be calculated in the x–y–z (external) coordinates of the peptide, but the descriptor we used is calculated in the peptide's frame of reference.

Hydrogen bond acceptors and donors do not include acidic and basic atoms, but will count atoms which are both acceptors and donors, such as an –OH group.

Molecular volume is calculated from the van der Waals radii, using a 0.75 Å grid approximation. Density is calculated from the molecular weight divided by this van der Waals calculated molecular volume.

Globularity, or inverse condition number, is a measure of how spherical (value of 1.0) or flat or linear (value of 0.0) a molecule is. It is calculated as the smallest eigenvalue divided by the largest eigenvalue of the covariance matrix of atomic coordinates of the molecule.

In addition to these descriptors, the number of disulfide bonds (1 or 2) was included (descriptor: BOND). Structures

with more disulfide bonds help to stabilize the  $\beta$ -sheet motif and maintain activity [2].

Similarly to previous QSAR studies [22], we have included the term  $\text{HBAC}^*(\text{HBDN}^{1/2})/\text{SASA}$  as a gauge of the electrostatic surface tension of a molecule [3,6].

Knowing that amphiphilicity is assumed to influence AMP activity, and using the coordinates of the peptide atoms, we calculated an amphiphilic moment according to Silverman [21]. There are two methods included in this reference, henceforth called moment 1 and moment 2 (MOM\_1 and MOM\_2 respectively). For these calculations, the center of geometry for each residue's side chain (non-hydrogen) and  $\alpha$ -carbon atoms is used as the location of each residue,  $r_i$ . For the peptide, the center of geometry of these residue centers is used as the reference center,  $r_c$ . The values used for residue hydrophobicity were those used by Silverman and come from reference [17], with the signs reversed. Hydrophilic residues have  $h_i$  negative, hydrophobic are positive. For MOM\_1, the contribution of each residue is the residue hydrophobicity multiplied by the moment arm ( $r_i - r_c$ ) for the residue. Summing over all residues and dividing by the number of residues gives a linear first order amphiphilic moment. This calculation gives high weight to the hydrophobic/hydrophilic character of residues far from the center of the peptide.

The more rigorous calculation of peptide hydrophobicity, MOM\_2, comes from scaling the coordinates of the peptide from an elliptical representation to a spherical one, where each residue will equally influence the hydrophobic moment. This involves first rotating the peptide into a principle coordinate orientation and then calculating scaling factors for the minor axes. The actual moment calculation is similar to MOM\_1. These methods provide both a magnitude and a direction for the hydrophobic moment, but we only include the magnitude of MOM\_1 and MOM\_2 for our calculations.

Also using the residue hydrophobicity scale, the mean hydrophobicity of each peptide was included as a descriptor, MEAN\_HYD, using the scale from [17]. This descriptor was favored over the MOE descriptor LOGP because it is easier to calculate directly from sequence data.

Through the studies, globularity and amphiphilic characteristics continued to arise as significant in building models of protegrin activity. We then included four descriptors, which used sequence information to indicate the actual count of (small/big) and (hydrophobic/hydrophilic) residues, descriptors SMFO, SMFI, BIFO, BIFI.

Example values for properties of 12 peptides used in this study are presented in Table 3.

#### 2.4. Correlating properties with activity

The software program JMP (SAS Institute Inc. 1999) was used as a regression tool (as in [12]) to determine least-squares-error structure–activity models. We seek to correlate structural properties with the activities of each peptide against each microorganism, the microbe activity measurements being available in minimum inhibitory concentrations (MIC,  $\mu\text{g}/\text{mL}$ ) and independent scales used for cytotoxicity and hemolysis.

A stepwise model building method was employed. For the selected properties and organism, the amount of error is reported that would result from not using a property in the linear model. The most influential property is determined and included in the model. The errors for the remaining properties are recalculated, factoring in the properties included in the model. The model fit ( $r^2$ ) is reported in each step, and this process is repeated, adding the most statistically significant descriptors until a certain, high confidence is reached.

#### 2.5. Cytotoxicity versus hemolysis

We also compare models between cytotoxicity and hemolysis. A peptide that lyses red blood cells but does not kill skin cells has different therapeutic value than one that is cytotoxic but has no activity on red blood cells.

The data for cytotoxicity and hemolysis are ranked on a 0–5 scale; for cytotoxicity, the scale corresponds to a concentration range for  $\text{EC}_{50}$ . For hemolysis, the scale corresponds to the percent of red blood cells killed at 80  $\mu\text{g}/\text{mL}$  peptide. Table 4 quantifies this scale.

Table 3  
Example values for calculated structural properties of 12 peptides

| Peptide | MWEI    | CHRG | FOSA    | EVDW    | PMOI     | HBAC | GLOB  | Mean Hydro |
|---------|---------|------|---------|---------|----------|------|-------|------------|
| PC001   | 2162.70 | 7    | 1463.17 | −24.912 | 177900.6 | 19   | 0.130 | 0.15       |
| PC005   | 2091.64 | 6    | 1458.40 | −17.949 | 158684.6 | 19   | 0.159 | 0.5        |
| PC009   | 2089.60 | 7    | 1475.13 | −21.387 | 173520.7 | 19   | 0.139 | −1.09      |
| PC016   | 1611.07 | 6    | 1119.7  | 42.150  | 90106.92 | 13   | 0.157 | 0.33       |
| PC020   | 1446.85 | 4    | 1064.02 | −7.9588 | 67008.62 | 13   | 0.166 | 2.27       |
| PC045   | 1838.33 | 6    | 1281.07 | 1.251   | 140109.6 | 16   | 0.099 | 0.52       |
| PC071   | 1395.81 | 4    | 989.774 | −13.631 | 60961.5  | 12   | 0.171 | 2.68       |
| PC073   | 1296.68 | 4    | 906.391 | −10.741 | 50049.7  | 11   | 0.217 | 1.93       |
| PC074   | 1286.66 | 3    | 917.77  | −8.2776 | 50442.9  | 11   | 0.195 | 3.72       |
| PC095   | 1723.22 | 5    | 1227.42 | −12.164 | 107436.5 | 15   | 0.123 | 2.56       |
| PC105   | 1985.51 | 6    | 1399.15 | −19.186 | 149171.4 | 17   | 0.107 | 1.51       |
| PC112   | 1451.83 | 4    | 972.19  | −16.156 | 62045.6  | 13   | 0.182 | 5.46       |

FOSA in  $\text{\AA}^2$ , EVDW in kcal/mol, PMOI in amu, Mean\_Hydro scale from [26].

Table 4  
Cytotoxicity and hemolysis scales

| Reported | Cytotoxicity (EC <sub>50</sub> ) (μg/mL) | Hemolysis (%RBC killed at 80 μg/mL of peptide) (%) |
|----------|--|--|
| 0        | >400                                     | 0–3  |
| 1        | 200–400                                  | 3–6  |
| 2        | 100–200                                  | 6–12   |
| 3        | 50–100                                   | 12–25  |
| 4        | 25–50                                    | 25–50  |
| 5        | <25                                      | >50  |

While the magnitude of these effects will vary between peptides, they do change in the same direction; a peptide with a reported value of five for both cytotoxicity and hemolysis is very toxic to both types of cells.

We looked at correlations of activity between the microbe species. We anticipated that species with high correlation between their activities will be modeled best by similar structural properties.

### 3. Results

#### 3.1. Activity correlations between species

Table 5 displays the correlation of activity between the species tested. The correlation between species of *N. gonorrhoeae* is very high (0.96); although the magnitude of the activities for strain FA-19 is generally higher than F-62, they move in a covariant way. We expect these strains to incorporate the same properties in a quantitative structure–activity relationship, albeit they will have different magnitudes for the coefficients for each property. Conversely, the correlation between *C. albicans* and *N. gonorrhoeae* F-62 is low (0.10) and these will likely be modeled best by different properties.

#### 3.2. Property–activity pairwise correlations

Looking at the pairwise correlations between activities and properties (Table 6), it is interesting to note that there are generally minimal relationships between the individual properties and activity; the correlation coefficients are greater than 0.5 for only a very small number of properties and activities. It is interesting that individual properties broadly regarded

as key for antimicrobial activity, such as positive charge and amphiphilicity, do not appear to correlate highly with antimicrobial activity or with hemolytic activity or cytotoxicity. The number of disulfide bonds has the best correlation with activity against microbes,  $|r_{cc}| > 0.58$ , but this property is also highly correlated with hemolytic and cytotoxic activity. This is consistent with previous qualitative studies [2] suggesting that the molecular topology as dictated by the disulfide bond constraints is important for activity. However, in most cases the correlation coefficients are between 0.1 and 0.3. In terms of physical understanding, this may simply mean that broad conclusions drawn about the role of single properties such as the amphiphilicity of the molecules do not provide useful insight.

#### 3.3. Property–activity models

The basis for constructing a QSAR model of activity is that although any number of properties may not correlate with activity individually, together a linear combination may fit well with the data. For example, while the easy hydrophobic moment calculation alone does not correlate perfectly with *E. coli* activity, it might add valuable depth to the data when considered with globularity. Finding the most statistically significant properties one at a time results in good models for some organisms but not all of them.

The simplest model for *E. coli*, Model\_E1, is:

$$\text{Activity (MIC, } \mu\text{g/mL)} = -0.044 (\pm 0.006552) \times \text{FOSA} + 0.000208 (\pm 0.00003) \times \text{PMOI} + 58.65 (\pm 9.34) \times \text{GLOB} - 1.500 (\pm 1.114) \times \text{MOM}_1 + 0.500 (\pm 0.227) \times \text{MEAN\_HYD}$$

$$n = 55, r^2 = 0.69$$

The *F* ratio statistic is smaller than 0.001 for FOSA, PMOI and GLOB. A scatter plot of the predicted activity against *E. coli* versus the actual is shown in Fig. 2. An almost identical result comes from using μM concentrations instead of μg/mL for MIC. MIC was reported in μg/mL in the majority of the literature surveyed, so this will be used throughout. The average hydrophobicity and the amphiphilic moment only marginally contribute to the accuracy of the model. Model\_E1 shows definite trends in the data, as discussed in Section 4. Briefly, we

Table 5  
Above the diagonal are the correlation coefficients between species; below the diagonal are the number of peptides tested in both of the respective species

|                               | <i>E. coli</i> | <i>N. gonorrhoeae</i> (F-62) | <i>N. gonorrhoeae</i> (FA-19) | <i>L. monocytogenes</i> | <i>C. albicans</i> | <i>P. aeruginosa</i> | Cytotoxicity | HEMOLYSIS |
|-------------------------------|----------------|------------------------------|-------------------------------|-------------------------|--------------------|----------------------|--------------|-----------|
| <i>E. coli</i>                |                | 0.200                        | 0.654                         | 0.491                   | 0.327              | 0.715                | −0.484       | −0.359    |
| <i>N. gonorrhoeae</i> (F-62)  | 28             |                              | 0.964                         | 0.509                   | 0.103              | 0.281                | −0.484       | −0.294    |
| <i>N. gonorrhoeae</i> (FA-19) | 26             | 29                           |                               | 0.494                   | 0.491              | 0.285                | −0.594       | −0.463    |
| <i>L. monocytogenes</i>       | 38             | 26                           | 24                            |                         | 0.714              | 0.826                | −0.634       | −0.557    |
| <i>C. albicans</i>            | 46             | 24                           | 23                            | 30                      |                    | 0.488                | −0.544       | −0.669    |
| <i>P. aeruginosa</i>          | 32             | 11                           | 11                            | 13                      | 26                 |                      | −0.619       | −0.448    |
| Cytotoxicity                  | 50             | 30                           | 28                            | 37                      | 40                 | 26                   |              | 0.660     |
| Hemolysis                     | 56             | 31                           | 29                            | 38                      | 46                 | 31                   | 56           |           |

Table 6  
Activity–property correlations

|                                    | <i>E. coli</i> | <i>N. gonorrhoeae</i><br>(F-62) | <i>N. gonorrhoeae</i><br>(FA-19) | <i>L. monocytogenes</i> | <i>C. albicans</i> | <i>P. aeruginosa</i> | Cytotoxicity | hemolysis |
|------------------------------------|----------------|---------------------------------|----------------------------------|-------------------------|--------------------|----------------------|--------------|-----------|
| MWEI                               | −0.400         | −0.262                          | −0.476                           | −0.195                  | −0.325             | −0.475               | 0.572        | 0.559     |
| CHRG                               | −0.312         | −0.063                          | −0.506                           | 0.035                   | −0.239             | −0.436               | 0.460        | 0.378     |
| FOSA                               | −0.446         | −0.274                          | −0.472                           | −0.175                  | −0.308             | −0.521               | 0.584        | 0.574     |
| NESA                               | −0.422         | −0.212                          | −0.453                           | −0.110                  | −0.232             | −0.496               | 0.570        | 0.434     |
| POSA                               | −0.288         | −0.178                          | −0.466                           | −0.079                  | −0.254             | −0.316               | 0.484        | 0.421     |
| EELE                               | 0.342          | 0.367                           | 0.117                            | 0.245                   | 0.031              | 0.277                | −0.348       | −0.251    |
| ESOL                               | 0.109          | −0.079                          | 0.442                            | −0.159                  | 0.187              | 0.283                | −0.159       | −0.200    |
| EVDW                               | −0.069         | 0.262                           | 0.285                            | 0.156                   | −0.033             | −0.109               | −0.056       | −0.270    |
| FLEX                               | −0.460         | −0.270                          | −0.499                           | −0.211                  | −0.301             | −0.544               | 0.616        | 0.516     |
| DIPO                               | −0.030         | −0.237                          | −0.481                           | −0.258                  | 0.003              | −0.070               | 0.409        | 0.422     |
| PMOI                               | −0.367         | −0.217                          | −0.494                           | −0.110                  | −0.282             | −0.470               | 0.548        | 0.500     |
| HBAC                               | −0.319         | −0.261                          | −0.345                           | −0.165                  | −0.266             | −0.323               | 0.517        | 0.427     |
| HBDN                               | −0.333         | −0.265                          | −0.377                           | −0.180                  | −0.275             | −0.331               | 0.548        | 0.466     |
| SASA                               | −0.438         | −0.145                          | −0.503                           | 0.016                   | −0.270             | −0.506               | 0.527        | 0.477     |
| DENS                               | 0.091          | −0.141                          | −0.230                           | −0.539                  | −0.382             | −0.161               | 0.239        | 0.286     |
| GLOB                               | 0.568          | 0.090                           | 0.583                            | 0.232                   | 0.376              | 0.563                | −0.604       | −0.445    |
| MVOL                               | −0.408         | −0.255                          | −0.467                           | −0.158                  | −0.299             | −0.471               | 0.565        | 0.546     |
| LOGP                               | 0.114          | 0.010                           | 0.336                            | −0.052                  | 0.055              | 0.195                | −0.359       | −0.045    |
| BOND                               | −0.156         | −0.330                          | −0.372                           | −0.690                  | −0.579             | −0.731               | 0.386        | 0.549     |
| ACC*(DON <sup>1/2</sup> )/<br>SASA | −0.105         | −0.262                          | −0.035                           | −0.261                  | −0.225             | −0.009               | 0.283        | 0.204     |
| SMFI                               | −0.127         | −0.030                          | −0.095                           | 0.254                   | 0.103              | −0.190               | 0.184        | 0.021     |
| BIFI                               | −0.338         | −0.063                          | −0.506                           | 0.035                   | −0.178             | −0.442               | 0.439        | 0.312     |
| SMFO                               | −0.234         | −0.610                          | −0.493                           | −0.726                  | −0.454             | −0.235               | 0.442        | 0.660     |
| BIFO                               | −0.041         | −0.007                          | 0.199                            | 0.025                   | −0.251             | −0.112               | 0.041        | 0.279     |
| MEAN_HYD                           | 0.272          | −0.202                          | 0.286                            | −0.345                  | −0.073             | 0.439                | −0.289       | 0.070     |
| MOM_1                              | −0.190         | 0.266                           | −0.042                           | 0.170                   | 0.203              | 0.076                | 0.038        | −0.077    |
| MOM_2                              | −0.013         | −0.008                          | 0.135                            | −0.394                  | 0.163              | 0.028                | 0.031        | 0.071     |

note here that the activity is inversely proportionate to the hydrophobic component of the surface area and the measure MOM\_1 of hydrophobicity of the molecules. Interestingly, activity against *E. coli* appears to be linearly related to the peptides' amphiphilic moment, although this descriptor does not enter the model with high statistical significance. The hydrophobic component of the solvent accessible surface area is highly correlated with the principle moment of inertia and it would be assumed that only one or the other would be required to model activity; however, excluding either of these model effects prevents any reasonable fit for the data.

Less statistically significant models can be built to model activity against the other species. In every case, it is possible to find descriptors that provide a better fitting model (see Appendix A).

By considering just the most active or least active peptides as a subset, interesting results emerge. Looking at the least active peptides, the six with MIC > 3.1  $\mu\text{g/mL}$ , we get Model\_E2 (scatter plot in Fig. 3):

$$\text{Activity (MIC, } \mu\text{g/mL)} = 254.8 (\pm 31.88) \times \text{GLOB} - 41.45 (\pm 6.147)$$

$$n = 6, r^2 = 0.941, F < 0.0001 \text{ for GLOB}$$

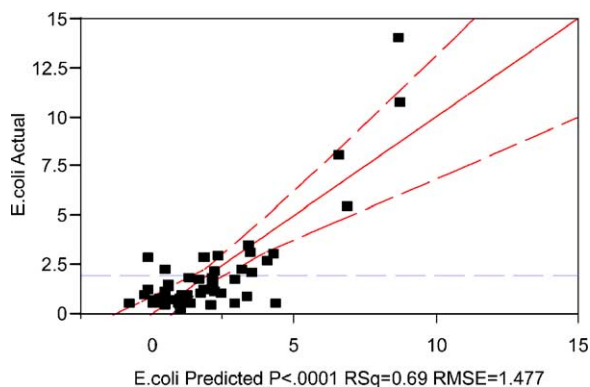


Fig. 2. Actual activity vs. predicted with Model\_E1.

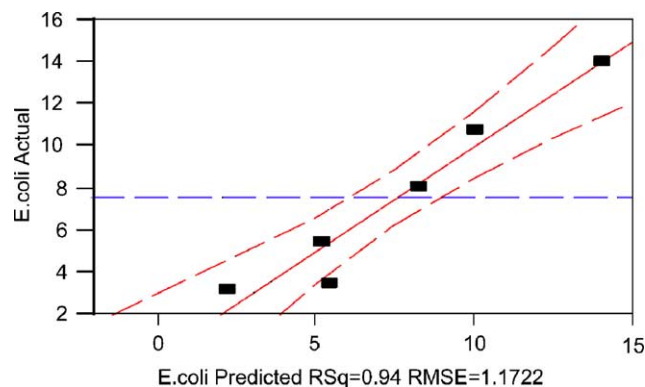


Fig. 3. Actual activity vs. predicted with Model\_E2.

When comparing the most active peptides one recurring interesting result is that the van der Waals energy term is one of the most significant components of the model.

For reference, Model\_E3-based on peptides with MIC < 0.7  $\mu\text{g}/\text{mL}$  is:

$$\text{Activity (MIC, } \mu\text{g}/\text{mL}) = -0.001997 (\pm 0.000386) \times \text{EVDW} - 0.01906 (\pm 0.0051) \times \text{DIPO} + 0.7631 (\pm 0.05618)$$

$n = 11$ ,  $r^2 = 0.857$ ,  $F < 0.006$  for EVDW, DIPO

### 3.4. Cytotoxicity versus hemolysis

Modeling how the protegrins act on skin cells compared to red blood cells aims to give insight into how to design peptides that are inactive to at least one of these cell types. The best model for cytotoxicity, Model\_Cyto:

$$\text{Cytotoxicity} = 0.07323 (\pm 0.02491) \times \text{EELE} + 0.02016 (\pm 0.005268) \times \text{ESOL} + 0.2493 (\pm 0.05782) \times \text{FLEX} - 23.42 (\pm 6.055) \times \text{GLOB} + 4.350 (\pm 1.985)$$

$n = 53$ ,  $r^2 = 0.58$ ,  $F < 0.005$  for each descriptor

These properties for cytotoxicity cannot be transferred to hemolysis. Doing so degrades the model severely; for the 59 peptides tested against red blood cells,  $r^2 = 0.308$ . However, finding the best properties for hemolysis, Model\_Hemo:

$$\text{Hemolysis} = 0.02584 (\pm 0.004333) \times \text{FOSEA} - 0.07892 (\pm 0.01653) \times \text{NESA} - 16.25 (\pm 7.760) \times \text{GLOB} + 1.252 (\pm 0.4378) \times \text{BOND} - 8.362 (\pm 2.601)$$

$n = 59$ ,  $r^2 = 0.651$ ,  $F < 0.04$  for GLOB,  $F < 0.006$  for others

If these properties are used with cytotoxicity, a model only gives an accuracy of  $r^2 = 0.443$ .

It appears that energy terms play a larger role in cytotoxicity and size terms play a role in hemolysis (considering that the properties cross over well in modeling these activities). Also where increased flexibility increases cytotoxicity, increasing numbers of disulfide bonds (and thereby implied rigidity) increase hemolytic activity. These models try to capture the properties that influence hemolysis or cytotoxicity and, can therefore, be used to describe novel peptides that are inactive against at least one of the cell types (Fig. 4).

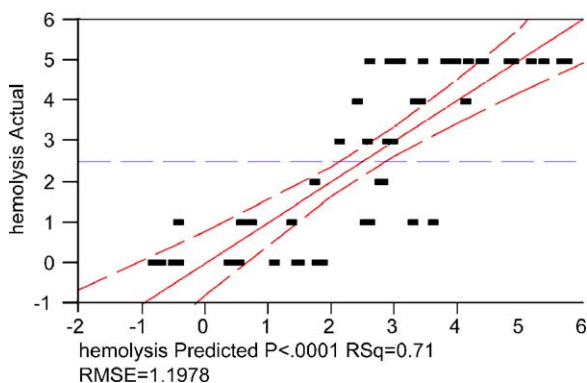


Fig. 4. Actual hemolytic activity vs. predicted with Model\_Hemo.

## 4. Discussion

### 4.1. *E. coli* QSARs

All models contain the globularity property and indicate that it is statistically very significant. In each of the *E. coli* models, the GLOB coefficient is positive; therefore, less globular molecules will have a higher activity. Linear or flat 2D molecules have a GLOB rating of 0, while a perfect sphere has GLOB = 1. The values of GLOB observed for the set of protegrins range from 0.083 to 0.218. All of the peptides maintain the  $\beta$ -sheet motif and thus are elongated, thin molecules. The actual values for GLOB can be nearly quantified in the length of the sequence (although simply using the number of residues as a descriptor does not model activity nearly as well as globularity); those peptides with GLOB less than  $\sim 0.15$  have 13–18 residues, and those with higher GLOB have 10–12 residues and are almost all less active than protegrin-1. Since all of the peptides are of the same general shape, GLOB serves as an indicator of the inverse of the length of the peptide. This implies that peptides with high GLOB (having fewer residues) are simply shorter in overall length and cannot span the lipid bilayer well enough to be fully active. The number of residues alone cannot describe this because when specific residues (those at the ends of the peptide or at the turn) are mutated to Arg, activity stays high; it is possible that these side chains can stretch out and interact with the necessary polar head groups of the bilayer lipids.

When we consider just the most active peptides, we see for the first time explicit energy terms in the QSAR (charge and the sum of the van der Waals interactions). These models compared 23 peptides with MIC ranging from 0.3 to 0.9  $\mu\text{g}/\text{mL}$ . Modeling peptides that have such similar activity requires introducing a more sensitive measure (EVDW) that was not needed in Model\_E1 to separate the most active peptides. EVDW values cover a much wider range, relative to GLOB for example, providing a term to differentiate these otherwise similar peptides.

When we model the least active peptides, we achieve an almost perfect fit,  $r^2 = 0.941$ , with only one descriptor. While Model\_E2 contains just six peptides, they span a range of activities from MIC = 3.2 to 14.1  $\mu\text{g}/\text{mL}$ , and only the descriptor GLOB is needed to separate them.

Taking the results from the least and most active sets leads us to conclude that globularity is important for spatial packing of peptides together on the surface of the target cell. There are large differences in globularity in the least active peptides that lead to a large range of minimum inhibitory concentrations high above the norm. Considering the most active peptides, globularity was nearly insignificant in establishing small differences in MIC. All of these peptides have the required geometry to be active. The terms differentiating activity must be more sensitive and van der Waals energy terms become important. This is not to suggest that pure van der Waals interactions are determining the activity, but simply that it is no longer physical geometry of the molecule that confers activ-

ity, but a more sensitive measure is required to differentiate the protegrins.

#### 4.2. Which hydrophobic moment to use?

Comparing protegrin-1 (MIC: 0.9  $\mu\text{g}/\text{mL}$ ) with PC-108 (MIC: 0.3  $\mu\text{g}/\text{mL}$ ) and PC-73 (MIC: 14.1  $\mu\text{g}/\text{mL}$ ) provides some insight as to which hydrophobic moment to use. The first method described in Methods emphasizes the residues away from the center of the peptide—the residues at the turn and those at the ends of the sequence. As shown below, all three of these sequences have the exact same amino acids from positions six through fifteen. These are the residues, which contain the four cysteines, which maintain the  $\beta$ -sheet structure of each peptide and everything in between. Beyond these residues, the ends of the peptide can freely rotate and perhaps play a role in aggregation of peptides.

The formal charge on the two more active peptides is +7 while the inactive is +3. Positive charge may influence activity, or it may be a byproduct simply from the size of the individual peptides (the longer sequences simply have more arginine residues).

Instead it is interesting to note that the more robust hydrophobic moment, MOM\_2 (which treats each residue the same) calculated for PC-73 is greater in magnitude than the other two, with the magnitude of the moment for PC-108 about half that of PC-73. However, using MOM\_1, the moments of PC-108 and PG-1 are more than double that of PC-73. This supports using the hydrophobic moment that gives greater weight to the residues far from the center.

#### 4.3. Concluding remarks

Most of the models constructed contained some terms describing the size, shape, and energy of the peptide. At times when just one or two of these properties can be used to create a good fit with the data, the addition of other structural descriptors does allow better prediction.

In the other models where globularity was a factor (see Appendix A), the coefficient is positive as for *E. coli*, indicating for those species flatter molecules again have higher activity.

In those models where the amphiphilic moment was significant, we find the coefficient is often greater than zero. This is unexpected, as we believe peptides with hydrophobic and hydrophilic components well separated would be increasingly active.

Over a wide variety of sequences, all based on the same structural motif, this study reveals a strong correlation between peptide activity and structural properties of the peptide monomers. Emphasis was put on properties that can be easily calculated and then used to reliably model activity in a variety of microbe species.

We have completed a quantitative structure activity relationship study on a large set of peptide sequences and structures. Although the dataset occasionally does not allow for

particularly accurate quantitative structure–activity relationships, the methods described show that models of peptide activity can be effectively developed, providing insight into peptide activity.

#### Acknowledgments

We are very grateful of Alan Waring and Robert Lehrer at UCLA for supplying the protegrin activity data. We also would like to acknowledge the University of Minnesota Supercomputing Institute, the University of Minnesota Biotechnology Institute, 3M and IBM for computational resources and funding for this work.

#### Appendix A. QSAR models for each species

##### *E. coli*:

$$\begin{aligned} \text{Activity (MIC, } \mu\text{g}/\text{mL}) = & -0.04327 (\pm 0.006413) \times \text{FOSA} \\ & + 0.0002007 (\pm 0.000032) \times \text{PMOI} + 56.84 \\ & (\pm 9.480) \times \text{GLOB} + 56.84 (\pm 9.480) \times \text{GLOB} + \\ & 56.84 (\pm 9.480) \times \text{GLOB} - 1.411 (\pm 1.140) \times \\ & \text{MOM}_1 + 5.752 (\pm 2.972) \times \text{MEAN\_HYD} + 24.35 \\ & (\pm 4.696) \end{aligned}$$

$$n = 55, r^2 = 0.68, F < 0.0001 \text{ for FOSA, PMOI, GLOB, } F < 0.23 \text{ for others}$$

##### *N. gonorrhoeae* (F-62):

$$\begin{aligned} \text{Activity (MIC, } \mu\text{g}/\text{mL}) = & 1.523 (\pm 0.6840) \times \text{EELE} \\ & 0.2759 (\pm 0.1577) \times \text{ESOL} - 6.925 (\pm 6.253) \times \\ & \text{SMFO} - 190.4 (\pm 87.00) \times \text{MEAN\_HYD} + 190.8 \\ & (\pm 45.48) \end{aligned}$$

$$n = 28, r^2 = 0.5148, F < 0.28 \text{ for SMFO, } F < 0.1 \text{ for others}$$

##### *N. gonorrhoeae* (FA-19):

$$\begin{aligned} \text{Activity (MIC, } \mu\text{g}/\text{mL}) = & -1.738 (\pm 1.162) \times \text{DIPO} \\ & + 550.0 (\pm 197.4) \times \text{GLOB} + 0.8161 (\pm 0.6253) \times \\ & \text{EVDW} + 2365 (\pm 1563) \times \text{HBAC} \times \\ & \text{HBDN}^{1/2}/\text{SASA} - 111.3 (\pm 59.16) \end{aligned}$$

$$n = 27, r^2 = 0.4796, F < 0.15 \text{ for all but EVDW, } F_{\text{EVDW}} = 0.2$$

##### *L. monocytogenes*:

$$\begin{aligned} \text{Activity (MIC, } \mu\text{g}/\text{mL}) = & -0.03460 (\pm 0.01613) \times \\ & \text{ESOL} - 476.8 (\pm 236.2) \times \text{HBAC} \times \\ & \text{HBDN}^{1/2}/\text{SASA} - 15.40 (\pm 2.194) \times \text{BOND} + \\ & 42.11 (\pm 9.737) \end{aligned}$$

$$n = 36, r^2 = 0.6347, F < 0.0001 \text{ for BOND, } F < 0.5 \text{ for other two}$$

*C. albicans:*

$$\text{Activity (MIC, } \mu\text{g/mL)} = 0.2682 (\pm 0.1244) \times \text{NESA} \\ + 0.9898 (\pm 0.4449) \times \text{DIPO} - 12.21 (\pm 2.962) \times \\ \text{HBAC} + 7.727 (\pm 3.546) \times \text{SMFI} - 15.80 (\pm 6.852) \\ \times \text{BOND} + 139.8 (\pm 18.21)$$

$n = 45$ ,  $r^2 = 0.5999$ ,  $F < 0.0002$  for HBAC,  $F < 0.04$  for others

*P. aeruginosa:*

$$\text{Activity (MIC, } \mu\text{g/mL)} = 53.49 (\pm 15.48) \times \text{GLOB} - \\ 17.65 (\pm 3.169) \times \text{BOND} + 30.54 (\pm 7.233)$$

$n = 32$ ,  $r^2 = 0.6698$ ,  $F < 0.0001$  for BOND,  $F_{\text{GLOB}} = 0.0017$

$$\text{Cytotoxicity} = 0.07323 (\pm 0.02491) \times \text{EELE} + \\ 0.02016 (\pm 0.005268) \times \text{ESOL} + 0.02016 \\ (\pm 0.005268) \times \text{ESOL} + 0.2493 (\pm 0.05782) \times \\ \text{FLEX} - 23.42 (\pm 6.055) \times \text{GLOB} + 4.350 (\pm 1.985)$$

$n = 53$ ,  $r^2 = 0.573$ ,  $F < 0.005$  for each descriptor

$$\text{Hemolysis} = 0.02584 (\pm 0.004333) \times \text{FOSA} - \\ 0.07892 (\pm 0.01653) \times \text{NESA} - 16.25 (\pm 7.760) \times \\ \text{GLOB} + 1.252 (\pm 0.4378) \times \text{BOND} - 8.362 \\ (\pm 2.601)$$

$n = 59$ ,  $r^2 = 0.651$ ,  $F < 0.04$  for GLOB,  $F < 0.006$  for others

## References

- [1] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucl Acids Res* 2000;28:235–42.
- [2] Chen J, Falla TJ, Liu H, Hurst MA, Fujii CA, Mosca DA, et al. Development of protegrins for the treatment and prevention of oral mucositis: structure–activity relationships of synthetic protegrin analogues. *Biopolymers (Pept Sci)* 2000;55:88–98.
- [3] Duffy EM, Jorgensen WL. Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. *J Am Chem Soc* 2000;122:2878–88.
- [4] Eband RM, Vogel HJ. Diversity of antimicrobial peptides and their mechanisms of action. *Biochim Biophys Acta* 1999;1462:11–28.
- [5] Hall LH, Kier LB. The molecular connectivity chi indices and kappa shape indices in structure–property modeling. *Rev Comp Chem* 1991;2:367–422.
- [6] Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett* 2000;10:1155–8.
- [7] Hancock REW, Lehrer R. Cationic peptides: a new source of antibiotics. *Trends Biotechnol* 1998;16:82–8.
- [8] Hancock REW, Chapple DS. Peptide antibiotics. *Antimicrob Agents Chemother* 1999;43:1317–23.
- [9] Hancock REW. Peptide antibiotics. *Lancet* 1997;349:418–22.
- [10] Heller WT, Waring AJ, Lehrer RI, Huang HW. Multiple states of beta-sheet peptide protegrin in lipid bilayers. *Biochemistry* 1998;37:17331–8.
- [11] Heller WT, Waring AJ, Lehrer RI, Harroun TA, Weiss TM, Yang L, et al., Huang HW. Membrane thinning effect of the beta-sheet antimicrobial protegrin. *Biochemistry* 2000;39:139–45.
- [12] Kaznessis YN, Snow ME, Blankley CJ. Prediction of blood–brain partitioning using Monte Carlo simulations of molecules in water. *J Comp Aided Mol Des* 2001;15:697–708.
- [13] Kokryakov VN, Harwig SSL, Panyutich EA, Shevchenko AA, Aleshina GM, Shamova OV, et al. Protegrins: leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesins. *FEBS Lett* 1993;327:231–6.
- [14] Lai JR, Huck BR, Weisblum B, Gellman SH. Design of non-cysteine-containing antimicrobial beta-hairpins: structure–activity relationship studies with linear protegrin-1 analogues. *Biochemistry* 2002;41:12835–42.
- [15] Miyasaki KT, Lehrer RI. Beta-sheet antibiotic peptides as potential dental therapeutics. *Int J Antimicrob Agents* 1998;9:269–80.
- [16] MOE (Molecular Operating Environment) modeling software. Montreal Canada: Chemical Computing Group, Inc.
- [17] Neumaier A, Huyer W, Bornberg-Bauer E. WWW document: [www.mat.univie.ac.at/~neum/software/protein/aminoacids.html](http://www.mat.univie.ac.at/~neum/software/protein/aminoacids.html). The sign of the values of amino acid hydrophobicity have been reversed in the calculations; 1998.
- [18] Nicolas P, Vanhoye D, Amiche M. Molecular strategies in biological evolution of antimicrobial peptides. *Peptides* 2003;24:1669–80.
- [19] Powers JPS, Hancock REW. The relationship between peptide structure and antibacterial activity. *Peptides* 2003;24:1681–91.
- [20] Roumestand C, Louis V, Aumelas A, Grassy G, Calas B, Chavanieu A. Oligomerization of protegrin-1 in the presence of DPC micelles: A proton high-resolution NMR study. *FEBS Lett* 1998;421:263–7.
- [21] Silverman BD. Hydrophobic moments of tertiary protein structures. *Proteins: Struct Funct and Genet* 2003;53:880–8.
- [22] Steinberg D, Lehrer RI. Designer assays for antimicrobial peptides. *Methods Mol Biol* 1997;78:169–86.
- [23] Tam JP, Wu C, Yang JL. Membranolytic selectivity of cystine-stabilized cyclic protegrins. *Eur J Biochem* 2000;267:3289–300.
- [24] Yasin B, Pang M, Wagar EA, Lehrer RI. Examination of *Chlamydia trachomatis* infection in environments mimicking normal and abnormal vaginal pH. *Sex Transm Dis* 2002;29:514–9.
- [25] Zhao C, Liu L, Lehrer RI. Identification of a new member of the protegrin family by cDNA cloning. *FEBS Lett* 1994;346:285–8.
- [26] Zhao C, Ganz T, Lehrer RI. The structure of porcine protegrin genes. *FEBS Lett* 1995;368:197–202.