Design of selective peptide antibiotics by using the sequence moment concept

Davor Juretić[1,*], Damir Vukičević[1], Nada Ilić[1], Nikolinka Antcheva[2] and Alessandoro Tossi[2]

[1]*University of Split, Faculty of Science, 21000 Split, Croatia*
[2]*Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, 34127 Trieste, Italy*
[*]*Corresponding author*

New antibiotics against multidrug-resistant bacteria are urgently needed, but rapid acquisition of resistance limits their usefulness. Endogenous antimicrobial peptides (AMPs) with moderate selectivity, but multimodal mechanism of action, have remained effective against bacteria for millions of years. Their therapeutic application, however, requires optimizing the balance between antibacterial activity and selectivity, so that rational design methods for increasing selectivity are highly desirable. We have created training (n=36) and testing (n=37) sets from frog-derived AMPs with determined therapeutic index (TI). The 'sequence moments' concept then enabled us to find a one-parameter linear model resulting in a good correlation between measured and predicted TI (r^2=0.83 and 0.64 for each set, respectively). The concept was then used in the AMP-Designer algorithm to propose primary structures for highly selective AMPs against Gram-negative bacteria. Testing the activity of one such peptide produced a TI>200 as compared to the best AMP in the data-base, with TI=125.

All higher organisms produce cationic antimicrobial peptides (AMPs)[1] whose main targets of action are membranes and associated biological functions[2]. Bacteria cannot easily reorganize, nor can they markedly alter, their lipid composition or their requirement for a strong, inside-directed electric field (attracting cations). For these reasons, AMPs have a low tendency to elicit resistance and show a moderate selectivity for bacterial cells.

A large group of AMPs acquire an amphipathic α-helical secondary structure in a low dielectric constant environment, such as when they interact with the cytoplasmic membrane[3], but in most cases their primary structure has not been related with quantitative parameters expressing their activity against either bacterial or host cells. Even when structural parameters affecting activity are known, it is a challenge to maintain or increase antibacterial activity while simultaneously decreasing hemolytic activity, since some optimal combination of specific attributes must be reached[4]. Increasing peptide selectivity requires a better definition of those structural features of AMPs which have evolved to ensure a selective action in fending off microbes without endangering host cells.

We have created an AMP activity database (AMPad) of 73 frog-derived AMPs with propensity to form amphipathic helices in membrane-mimetic solvents, which connects measured peptide selectivity with associated primary structures (**Supplementary Tables S1 and S2**). Published results were selected only where antimicrobial activity was determined by the serial dilution method, in terms of the minimal inhibitory concentration (MIC) towards *E. coli*, and cytotoxicity expressed in terms of 50% hemolysis of fresh human blood cells (HC_{50}). The therapeutic index (TI), a measure of peptide selectivity, is then defined as: TI = HC_{50}/MIC. To extract rules as general as possible we have then divided the AMPad database in a training data set of 36 non-homologous (less than 70% pair-wise identity) peptides and a test data set of the 37 remaining peptides. All but two peptides (pexiganan and gaegurin 4 W16) in the training set were of natural origin.

Our first goal was to find a one-parameter linear model for correlating measured and predicted TI values. To the best of our knowledge, non-homologous antimicrobial peptides have never been used before in quantitative structure-activity studies of TI. Furthermore, most structure-activity studies of AMPs used multi-linear models for predicting peptide activity[5]. Models with only one peptide descriptor were not expected to work, due to the well know necessity to use several interrelated structural and physicochemical parameters that all modulate activity and specificity, such as charge, secondary structure, amphipathicity, and hydrophobicity[3].

A different approach is however possible, which does not use peptide mean properties as given descriptors, but instead exploits the observation[4,6] that position-dependent physicochemical properties also determine peptide activity and selectivity. Position-dependent amino acid attributes are usually represented as smoothed, two-dimensional sequence profiles, such as the Kyte-Doolittle hydrophobicity profile. To simplify analyses, the whole sequence profile for chosen amino acid attributes can be converted into one vector with two components (x and y). This can be easily done by bending the sequence in, for instance, a right angle arc, and then performing vector summation of attribute vectors associated with all amino acids in the sequence (**Figure 1**). A vector sum that replaces the sequence profile will be named the *sequence moment*. Unlike a peptide's amphipathicity, which is quantified by Eisenberg's hydrophobic moment[7], the sequence moment does not depend on knowledge of the secondary structure, but rather is an indicator of peptide lengthwise asymmetry.

Amino acid attribute profiles were obtained by selecting 144 amino acid scales used to study integral membrane proteins and protein folding [8-10]. The simplest methods for finding sequence profiles were then: a) direct usage of amino acid attributes, or b) smoothed amino acid attributes associated with a central amino acid in a sliding window either considering all residues in the window (mean values) or without taking into account the attribute of the

central amino acid itself (sequence environment values). Using any one of the selected 144 scales to derive scalars from the corresponding sequence moments, we could not however achieve the desired high correlation among these descriptors and measured TI. The next simplest choice was to use sequence moments for two different amino acid scales and to examine for which pair any given scalar descriptor, derived from the two sequence moments, will give the desired high correlation.

The simplest parameter thus obtained, providing a correlation $r^2 = 0.83$ with measured TIs from the training data set, was named the D descriptor (**Methods**). It is the cosine of the angle between sequence moments calculated from sequence environments when Janin's[11] and Guy's hydrophobicity scales[12] are used. These scales are somewhat correlated ($r^2 = 0.64$). Janin's scale was obtained by calculating the ratio of buried to accessible molar fractions of each amino acid in soluble proteins, while Guy's scale followed from calculating residue transfer free-energy from the surface to the center of soluble proteins. After calculating the D descriptor values for training set peptides the predicted TI is found as the best linear fit: $TI = 50.126 - 44.803\,D$ (**Supplementary Figure S1**). The correlation between measured and predicted TI so obtained for the test set of peptides is then expressed as a determination coefficient, $R^2 = 0.64$, which is still an acceptable indication that a correlation exists (**Supplementary Figure** S2). **Figure 1** illustrates that angular separation of sequence moments is quite different for a highly selective (PGLa)(**Supplementary Method 1**) and a mediocre peptide antibiotic (pseudin 2). Notice that peptide lengthwise asymmetry is additionally strengthened by modeling the observation that amino acids closer to N-terminus are more relevant for peptide activity[3,13].

The D descriptor then helped us in our second goal; to create an algorithm capable of proposing novel peptides with a potentially high therapeutic index. The number of possible peptides, even when we restrict ourselves to a 23 amino acid length, is however extremely

large. We used a recursive algorithm and imposed various rational restrictions to reduce these to a workable number of potential AMP candidates with a high TI (see **Methods, Supplementary Method 2** and **Supplementary Tables S3-S5**).

General requirements included i) mean hydrophobicity values according to the CCS scale within an allowed range,[3, 14] ii) predicted TI values of at least 85, and iii) limiting design to amphipathic helix-forming peptides. A statistical analysis of the most common motifs and amino acids associated with the set of best peptides led to the remaining set of restrictions out of which the *motif regularity index* is probably the most important innovation.

As a result, the AMP-Designer algorithm produced the 7 peptides (out of $\approx 8 \cdot 10^{29}$ possibilities, see *Methods*). Six of these end with a cysteine residue, which may be a vestige from the so called "Rana-box", a cysteine-bridged stretch present at the C-terminus of many frog-derived helical antimicrobial peptides[3,15], which is not correlated with increased activity or selectivity[15]. Hence, we considered the best candidate for synthesis and characterization to be the seventh: **GI<u>GKHVG</u>KALKGLK<u>GLLKGLGES</u> (DESC 1)**, with a predicted TI = 86, mean hydrophobicity[14] H = -1.0, and a relative amphipathicity[16] of 0.66 (66% that of a perfectly amphipathic peptide). The underlined "small" motifs [A,G,S]XXX[A,G,S] promote helix-helix interaction in the membrane[17]. Neglecting the requirement that the predicted TI be higher than 85 (ie. D descriptor omission), would have lead to a total of 37 designed peptides, indicating that using the D descriptor was useful in eliminating 81% of these peptides.

The suggested peptide was synthesized in good purity (**Supplementary Figure S3**), together with two known frog peptides from the training database. Ascaphin 1 and pseudin 2 were selected as controls because we hoped that observed structural and functional properties of DESC1 would span the range of corresponding peptide attributes for good and mediocre peptide antibiotics. This was indeed the case for most attributes (see **Supplementary Figure S4** and **Supplementary Table S6**). For instance, despite the presence of numerous glycine

residues, DESC1 has an estimated helical content of about 70% in the presence of the helix promoting solvent TFE.

Antimicrobial activity assays confirmed that the DESC 1 is selectively active against Gram-negative bacteria (**Table 1**). Its C-terminally amidated form had a MIC value of 2-4 μM against *E. coli*, against which it was designed, and a moderate activity also towards *P. aeruginosa*, whereas its activity is poor against the Gram-positive bacterium *S. aureus*. The designed peptide did not quite reach 50% lysis of red blood cells even at the highest concentration tested; 800 μM (**Table 1** and **Figure 2**). By comparison, pseudin showed a HC_{50} of 50 μM under the same conditions, and ascaphin of 80 μM. Using these data to determine experimental TI values, and comparing them with predicted TI, we found values for control peptides of 6 against the predicted 8 for pseudin, 39 against the predicted 55 for ascaphin (**Table 1 and Supplementary Table S1**). These data, together with those for DESC 1, with a measured TI value of over 200 against the predicted 86, more than confirm the design method.

The high selectivity of the designed peptide DESC1 may in part derive from its unusual primary structure, which contains seven glycine residues, and which is ≤ 50% identical to any known helical AMP, being closest to bombinin (BMN_BOMVA, accession P01505, 50% identity). Based on a statistical analysis of natural helical AMPs[3], glycine residues have an increased frequency at position 7 or 14. It was shown that placing a Gly residue at position 7 in a 19mer artificial helical AMP[18], somewhat increased selectivity, while placing Gly in position 14, or 7 and 14 considerably reduced antimicrobial potency, so there seemed to be positional requirements for its use in increasing selectivity. Using the D descriptor in connection with the AMP-Designer algorithm it was possible to simultaneously introduce 6 internal glycine residues without markedly reducing activity towards Gram-negative microorganisms, while dramatically increasing selectivity.

The flexibility of the DESC1 peptide and its Gly distribution pattern may have the dual role of facilitating unstructured monomers to pass easily through cell wall components[19] and of increasing the probability of transient monomer aggregation and pore formation at the level of the cytoplasmic membrane. Being different from any known AMP, the DESC1 peptide can be considered as the first member of new glycine-rich class of helical AMP lead compounds for potential medical applications.

Our method offers rational peptide design for maximizing target selectivity (against Gram-negative bacteria) that does not depend on homology to known structures as a main bottleneck for designing new structures. Its strength is that it depends on multiple flexible screening criteria. For example, a D-descriptor close to -1.0 is only one indicator that predicts that a TI can be high. However, an all-Gly peptide would also have maximal predicted TI by just this criterion, because with the chosen pair of amino acid scales (Guy's and Janin's) all sequence environments are then of opposite sign. This follows from the classification of Gly as a buried residue in the Janin's scale, while Gly partitions as if it is polar according to Guy's scale. The AMP-Designer algorithm, with a flexible choice of screening criteria, will be useful help in exploring a broad range of helix-type peptide antibiotics, which biological evolution has left unexplored.

The concept and the algorithm used to design DESC1 are also likely to be useful tools for redesign and selectivity optimization of existing helix-type AMPs. The separation of sequence moments is such a sensitive descriptor that it can predict large TI increase even after a single point mutation. One example is the correct prediction of the TI value increase from a measured value of 20 (predicted value of 19.6) for magainin 2, to a measured value of 125 (predicted value of 94) for magainin 2 F5W[6] (**Supplementary Figure S5).** This confirms the potential usefulness of the D descriptor, when associated with appropriate constraints, for

developing and refining new lead compounds with increased selectivity against Gram-negative bacteria.

Methods

Statistical methods

We used the correlation between a given descriptor, derived from sequence moments, and the measured TI values in the training data set of 36 non-homologous peptides as selection criterion. To calculate sequence moments (see **Figure 1** legend) we used 144 different scales of amino acid attributes, several different procedures for creating sequence profiles s (see main text), 15 peptide bending angles $\alpha = \pi/15, \pi/14, ..., \pi$ and five pondering values $q = 0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$. From the so obtained sequence moment vectors we have used only those pairs of sequence profiles (s_1, s_2) that corresponded to the same angle and same pondering value. One-parameter models were obtained by transforming each (s_1, s_2) pair into 36 numbers $d_{i,j,k}(s_1, s_2)$ defined by:

$$d_{i,j,k}(s_1,s_2) = \begin{Bmatrix} 1, & i=0 \\ |s_1|, & i=1 \\ 1/|s_1|, & i=2 \text{ and } |s_1| \neq 0 \\ 0, & i=2 \text{ and } |s_1| = 0 \end{Bmatrix} \cdot \begin{Bmatrix} 1, & j=0 \\ |s_2|, & j=1 \\ 1/|s_2|, & j=2 \text{ and } |s_2| \neq 0 \\ 0, & j=2 \text{ and } |s_2| = 0 \end{Bmatrix} \cdot \begin{Bmatrix} 1, & k=0 \\ \cos(\square(s_1,s_2)), & k=1 \\ \square((s_1,s_2)), & k=2 \\ \sin(\square(s_1,s_2)), & k=3 \end{Bmatrix}.$$

Note that some of these numbers correspond to one scale only, but with such a choice of scalar descriptors the highest correlation with measured TI was $r^2 = 0.52$. Much better results are obtained with a triplet defined by choosing two different amino acid scales and one smoothing technique. In this way we have assigned 2700 scalar descriptors to each triplet. The best correlation between descriptor and observed TI in the training data set is assigned to the chosen triplet. Out of thousands of triplets, only 3 corresponded to correlation coefficients $r^2 \geq 0.81$ in the training data set. The sequence moments for these best descriptors were obtained by summing all sequence environment vectors i.e. by using the smoothing process which excludes the central amino acid in the sliding window (**Figure 1**).

The simplest descriptor (among 3 best descriptors corresponding to these triplets) was the cosine of the angle between sequence moments associated with Janin's[11] and Guy's[12] hydrophobicity scale, a peptide sequence bending arc of $\alpha = \pi/2$ and pondering factor $q = 1/3$. It was named the D descriptor. It is associated with a correlation $r^2 = 0.829$ when the training data set is used (**Supplementary Figure S1**). The remaining two descriptors were significantly more complex and only one of them was associated with a slightly higher correlation coefficient $r^2 = 0.838$. Hence, our choice fell on the D descriptor**.** (see **Supplementary Method S1** for one example of the D descriptor calculation)**.** It provides a simple method of measuring lengthwise sequence asymmetry in bent peptide sequences. The determination coefficient for the test set of peptides was $R^2 = 0.64$. This is necessarily lower than the correlation between measured TI and D descriptor for those peptides ($R^2 = 0.66$) (**Supplementary Figure S2**).

We have somewhat misused statistical methods here. Namely, the data in both the training set and in the testing set are not normally distributed. Hence, standard statistical practice would be not to use Pearson r correlation, but to use a non-parametric methods instead. However, we are not interested in an index that correctly orders peptides according to their TI, but to find an index that well predicts which peptides have a very high value of TI. It can be easily shown that a given percent error in predicting peptides with high values of TI much more influences the Pearson correlation coefficient then the same percent error in predicting peptides with the low values of TI. Therefore, r extracts information which is significant to our study.

We addressed also the concern that a small number of peptides with high measured TI in the training data set (11 peptides with $TI \geq 20$) are examined by selecting one among millions of descriptors. In order to verify that our predictions are not purely casual, we performed the same computation, but for each of the amino acid scales we have randomly permuted the data. In three such experiments the best values of r^2 have been: 0.71, 0.63 and 0.65, and only very few descriptors had comparable values. These are significantly lower than in our calculations with correct distribution of amino acid attribute values for each amino acid scale, where we furthermore had over a hundred triplets and more than 200 000 descriptors with $r^2 \geq 0.71$.

The AMP-Designer algorithm and associated restrictions

The basic idea of the algorithm is to support the design of highly selective frog-type helical peptide antibiotics by combining D descriptor based TI prediction with rational physicochemical and statistical restrictions based on the structural characteristics of such AMPs.

A first set of applied restrictions are very simple general requirements, such as:

1) D descriptor predicted TI > than 85 (the maximum predicted TI is 96 in the D descriptor model);
2) peptide net positive charge = 4 or 5 (as most frequent in natural helical AMPs[3]);
3) mean hydrophobicity (using the CCS scale, see below[14]) = -0.5 to -1.5 (as most often found for helical AMPs[3]).
4) non-polar residues A, L, M, V, I, F, W are copmletely separate from polar residues E, D, Q, N, G, K, R in a helical wheel projection (limiting design to helix-forming amphipathic peptides that have glycine residues in the polar helix face);
5) no more than 2 identical neighboring amino acids (to avoid trivial cases with long segments of identical amino acids that do not occur in natural AMPs);
6) the three C-terminal amino acids should be the C-terminal residues in at least one good natural peptide antibiotic.

The remaining restrictions have been obtained by statistical analyses of the set of best peptides (SBP) ($TI \geq 20$, the first 11 and the first 15 peptides respectively from the **AMPad database** training and testing data sets), and by statistical analyses of the full set of peptides (SAP) (all AMPad peptides, **Supplementary tables S1 and S2)**. These restrictions are:

7) the first amino acid (AA) should be *G*, because most peptides in SBP start by *G*;
8) the second AA should be either L or I, since most of peptides starting with G in SBP have L or I on the second position;
9) residues with the highest amino acid selectivity index associated with SBP peptides are E, Q, H, G and D (**Supplementary Table S3**), hence at least 9 such AAs should be present in a peptide;
10) the successor of each AA should be one of the five most common successors of that AA in the SBP data set (**Supplementary Table S4**);

11) the AA placed at the $i+4$ position with respect to any AA (spatially close in an α-helical conformation) should be one of the five most common successors at distance $i+4$ of that AA in the SBP data set (**Supplementary Table S5**)
12) the *motif regularity index* should be less then 2. This index determines how well the designed peptide p incorporates motifs that are the most common in the structure of antimicrobial peptides in the SBP data set (see **Supplementary Method S2** and **Supplementary Tables S4** and **S5**).

Based on these restrictions, we selected the following peptides out of $8x10^{29}$ possible ones: GLKGLLGKALKGIGKHIGKAQGC; GLKGLLGKALGEAKGLLGKHKGC; GITQGVLKGIGKHVGKALKGIGC; GIGKHVGKALKGLKGLLKGLGEC; GIGKHVGKALKGVKGLLKGLGEC; GIGKHVGKALGELKGLLKGLKGC; GIGKHVGKALKGLKGLLKGLGES; The last peptide, DESC 1, was selected for synthesis and characterization. It is unlike any known helical peptide. A suitable name for this novel peptide would be adepantin 1, an abbreviation for **A**utomatically **De**signed **P**eptide **Anti**biotic **N**o. **1**.

Mean hydrophobicity and relative amphipathicity

The global hydrophobicity and the amphipathicity for each synthesized peptide sequence were calculated using a hydrophobicity index (Hi) scale derived from the normalized and filtered consensus of 163 published scales, that is arbitrarily ranged between maximum values of +10 for Phe, and -10 for Arg[3, 14]. The hydrophobicity is given as the mean value ($H = (\sum Hi)/ l$), where l = peptide length. The mean hydrophobic moment (μHmax) was calculated as described by Eisenberg et al.[7]. The relative amphipathicity (μH/μHmax) for each peptides was then determined with respect to the value of the maximum hydrophobic moment for a perfectly amphipathic, 18-residue peptide composed only of Phe and Arg (μHmax = 6.4 with our scale). This relative measure of amphipathicity is less likely to vary according to the scale used, than an absolute value.

Peptide synthesis, purification and characterization

DESC 1 amide, ascaphin-1 amide (ASC 1) and pseudin-2 amide (PSEU 2) were synthesized by Fmoc-solid phase peptide synthesis on a Microwave-enhanced CEM Liberty synthesizer using Fmoc-PAL-PEG-PS resin (substitution 0.34 mmol/g). The cleaved/deprotection cocktail was trifluoroacetic acid, water and triisopropylsilane (95:2.5:2.5). The crude peptides were purified on a preparative Waters XTerra® C18, 7 μm, 300 Å, 19x300 mm column using a 25-45% CH_3CN in 60 min gradient with a 5 ml/min flow. Peptides quality and purity were verified by mass spectrometry (Esquire 4000 Ion Trap, Bruker Daltonics Inc, Germany), while peptide concentrations were determined by mass and verified by the Waddell method[20]. The capacity of each synthesized peptide to assume a helical conformation was probed by CD spectroscopy on a Jasco J-715 spectropolarimeter (Jasco, Japan), using 2 mm path length quartz cells and peptide concentrations of 20 μM, in 5 mM sodium phosphate buffer pH7, in the absence and presence of 50 % (v/v) TFE. All spectra are the mean of at least two trials, each with the accumulation of three scans. Helical content (% Helix) was estimated from the molar ellipticity at 222 nm, according to the method of Chen[21].

Hemolytic activity (HC_{50})

The hemolytic activity of peptides was assessed using freshly isolated human erythrocytes from healthy donors, by monitoring the release of hemoglobin at 405 nm. Aliquots of cell suspension (0.5% erythrocytes) were incubated in triplicate with different peptide concentrations (from 10 μM to 800 μM) in PBS at 37^{o} C for 30 minutes. Total lysis (100% hemolysis) was determined by addition of 0.2% Triton X-100. The HC_{50} value was taken as the mean concentration of the peptide producing 50% hemolysis.

Antimicrobial activity and therapeutic index determination

The bacteriostatic activity of the peptides was determined against *Escherichia coli* ATCC 25922, *Pseudomonas aeruginosa* ATCC 27853 and *Staphylococcus aureus* ATCC (10^5 cfu/ml bacteria), as minimum inhibitory concentrations (MIC), using the microdilution susceptibility test in MH broth as described previously[13], or as the effect on bacterial growth kinetics (10^6 cfu/ml bacteria in PBS) in the presence of different peptide concentrations (2, 4 and 8 μM), monitoring the optical density at 600 nm for 4h. The therapeutic index (TI) is defined as the ratio of HC50/MIC.

ACKNOWLEDGMENTS

The authors are grateful to Viktor Bojovic for the SPLIT server maintenance and to Bono Lucic who provided us with some amino acid scales. The work was supported in part by Croatian Ministry of Science, Education and Sport (Grant numbers 177-1770495-0476 (D.J. and N.I.), 177-0000000-0884 (D.V.), and 037-0000000-2779 (D.V.)) and by a Friuli Venezia Giulia LR26 grant for the R^3A^2 network project.

AUTHOR CONTRIBUTIONS

D.J., AMPad database and corresponding tables for predicting TI, *sequence moments* concept, some ideas incorporated in the AMP-Designer algorithm and writing most of manuscript, D.V., statistical analysis, predicting TI by using sequence moments, AMP-Designer algorithm, writing theoretical part of methods, Supplementary Figures S1 and S2, Supplementary Methods, Supplementary Tables S3, S4 and S5, A.T. directing experiments, Figure 2, Table 1, Supplementary Figures S3 and S4, Supplementary Table S6, writing part of manuscript and methods dealing with experiments, N.A., peptide synthesis, purification and characterization; hemolytic and antimicrobial activities, extracting primary data; N.I. experiments on antibacterial activity, homology analysis, Figure 1 and Supplementary Figure S5.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Literature:

1. Zasloff, M. Antimicrobial peptides of multicelullar origin. *Nature* **415**, 389–395 (2002).
2. Westerhoff, H.V., Juretic, D., Hendler, R.W. & Zasloff, M. Magainins and the disruption of membrane-linked free-energy transduction. *Proc. Natl. Acad. Sci. USA* **86**, 6597-6601 (1989).
3. Tossi, A., Sandri, L. & Giangaspero, A. Amphipathic, α-Helical Antimicrobial Peptides, *Biopolymers* **55**, 4–30 (2000).
4. Dathe, M., Nikolenko, H., Meyer, J., Beyermann, M. & Bienert, M. Optimization of the antimicrobial activity of magainin peptides by modification of charge. *FEBS Lett* **501** 146-150 (2001).
5. Langhama, A.A. *et al*. Correlation between simulated physicochemical properties and hemolycity of protegrin-like antimicrobial peptides: Predicting experimental toxicity. *Peptides* **29**, 1085–1093 (2008).
6. Tachi, T., Epand, R.F., Epand, R.M. & Matsuzaki, K. Position-dependent hydrophobicity of the antimicrobial magainin peptide affects the mode of peptide-lipid interactions and selective toxicity. *Biochemistry* **41**, 10723-10731 (2002)**.**
7. Eisenberg D., Weiss R.M., Terwilliger C.T. & Wilcox W. Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* **17**, 109-120 (1982).
8. Juretic, D., Lucic, B., Zucic, D. & Trinajstic, N. Protein transmembrane structure: recognition and prediction by using hydrophobicity scales through preference functions. *Theoretical and Computational Chemistry* **5**, 405-445 (1998).
9. Gromiha, M.M. A statistical model for predicting protein folding rates from amino acid sequence with structural class information. *J. Chem. Inf. Model.* **45**, 494-501 (2005).
10. Yuan, Z. *et al*. Predicting the solvent accessibility of transmembrane residues from protein sequence, *J. Proteome Res.* **5**, 1063-1070 (2006).
11. Janin, J. DeltaG-transfer from buried interior to solvent accessible surface. *Nature* **277**, 491-492 (1979).
12. Guy, H.R. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys. J.* **47,** 61-70 (1985).
13. Tossi, A., Tarantino, C. & Romeo, D. Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. *Eur. J. Biochem*. **250**, 549-558 (1997).
14. Tossi, A., Sandri, L. & Giangaspero, A. New consensus hydrophobicity scale extended to non-proteinogenic amino acids. PEPTIDES 2002, Ettore Benedetti and Carlo Pedone (Eds.), Edizioni Ziino, Napoli, Italy. pp. 416-417.
15. Simmaco, M., Mignogna, G. & Barra, D. Antimicrobial peptides from amphibian skin: what do they tell us? *Biopolymers*. **47**, 435-450 (1998).
16. Zelezetsky, I. *et al.* Controlled alteration of the shape and conformational stability of α-helical cell-lytic peptides: effect on mode of action and cell specificity. *Biochem. J.* **390**, 177-188 (2005).
17. Senes, A., Gerstein, M. & Engelman, D.M. Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. *J. Mol. Biol.* **296**, 921-936 (2000).
18. Zelezetsky, I., Pag U., Sahl, H-G. & Tossi, A. Tuning the biological properties of amphipathic alpha-helical antimicrobial peptides: rational use of minimal amino acid substitutions. *Peptides* **26**, 2368-2376 (2005).

19. Jiang, Z. *et al*. Effects of net charge and the number of positively charged residues on the biological activity of amphipathic α-helical cationic antimicrobial peptides. *Biopolymers* **90**, 369-383 (2007).
20. Waddell, W. J. A simple ultraviolet spectrophotometric method for the determination protein concentration. *J. Lab. Clin. Med.* **48**, 311-314 (1956).
21. Chen, Y. H., Yang, J. T., Chau, K.H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
22. Conlon, J.M., Sonnevend, A., Davidson, C, Smith, D. D. & Nielsen, P.F., The ascaphins: a family of antimicrobial peptides from the skin secretions of the most primitive extant frog, Ascaphus truei. *Biochem. Biophys. Res. Commun*. **320**, 170-175 (2004).
23. Olson, L., Soto, A., Knoop, F. C. & Conlon, J. M. Pseudin-2: an antimicrobial peptide with low hemolytic activity from the skin of the paradoxical frog. *Biochem. Biophys. Res. Commun.* **288**, 1001–1005 (2001).

Figure 1. **Sequence moments for PGLa (left panel) and pseudin 2 (right panel).** To take into account positional-dependent properties[4,6] p, we introduce the peptide sequence moment $v(p)$ corresponding to sequence profile $p=(p_1,...,p_l)$. $v(p)$ is defined as the weighted vector sum of p_i-vectors associated with each amino acid in a sequence of length l when the peptide sequence is bent in an arc α. Pondering of p_i-vectors by $(l+1-i)^q$, where q is a constant in the range 0 to 1, models the observation that amino acids closer to the N-terminus are more relevant for peptide's activity[3,13]. After performing vector summation of $\vec{p}_i(l+1-i)^q$ terms, the resulting sequence moment *x* and *y* coordinates are given by:

$$v(p)=v_{\alpha,q}(p)=\left(\sum_{i=1}^{l}\cos\left(\frac{l-i}{l-1}\cdot\alpha\right)\cdot p_i\cdot(l+1-i)^q,\sum_{i=1}^{l}\sin\left(\frac{l-i}{l-1}\cdot\alpha\right)\cdot p_i\cdot(l+1-i)^q\right).$$

In this example, we used optimal values $\alpha = \pi / 2$ and q = 1/3 (see **Statistical Methods**). Calculation details can be found in the **Supplementary method S1** for the case of PGLa**.** For each residue, small blue arrows are calculated by using Guy's amino acid index scale[12], while red arrows are determined by using Janin's amino acid index scale[11]. The D descriptor is the cosine of the angle between the resulting vector sums of the sequence moments (large arrows) for a chosen peptide. Since this descriptor does not take into account the lengths of sequence moments it is not important that our pondering function enhances sequence moment length for longer peptides. Predicted TI values for PGLa and pseudin 2 are 95 and 8 respectively.

Figure 2. Hemolytic activity of DESC 1 (-●-), ASC 1 (-▲-) and PSEU 2 (-■-). The hemolytic activity of the peptides was calculated as percent lysis of human erythrocytes (0.5% suspension) with increasing peptide concentrations (from 10 - 800 μM) in PBS, after incubation at 37 ^{o}C for 30 minutes, and are the mean of two experiments performed in triplicate. Total lysis (100% hemolysis) was determined by addition of 0.2% Triton X-100.

Table 1. Biological activities and therapeutic index determination

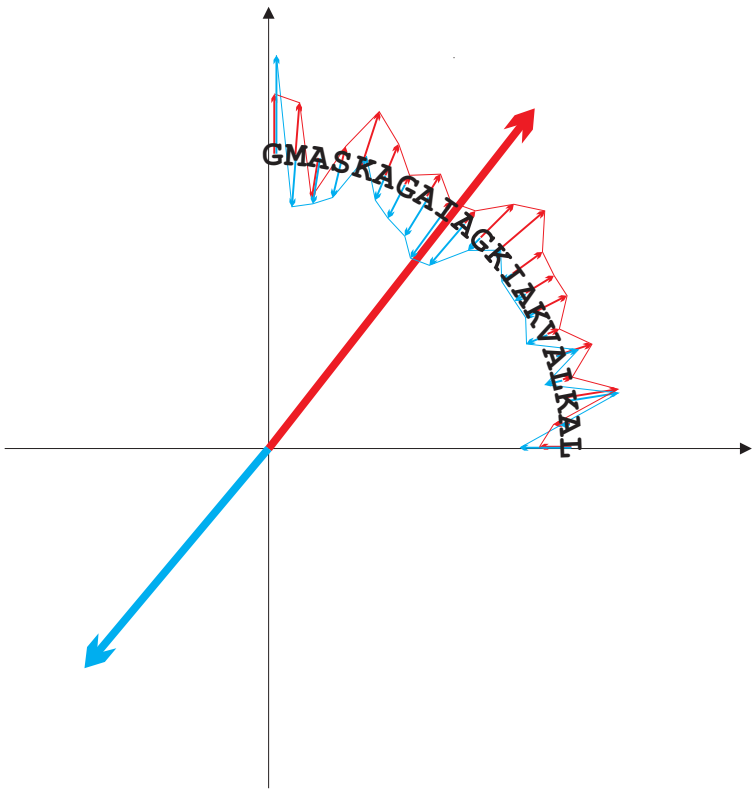
[1] DESC1 = GIGKHVGKALKGLKGLLKGLGES-NH_2 , ASC 1 = ascaphin 1[22] , PSEU 2 = pseudin 2[23].

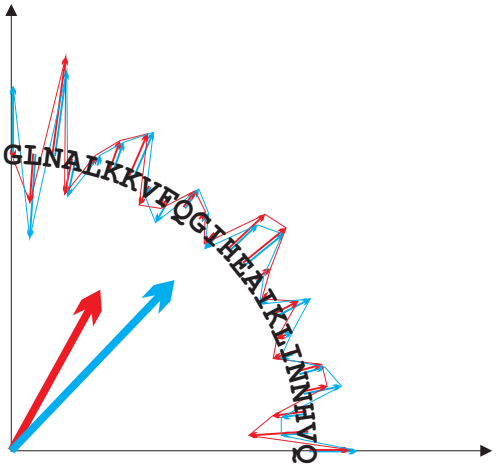
[2] Carried out in 100 % (v/v) MH broth with 10^5 CFU/ml bacteria in the logarithmic phase - mean of at least three experiments performed in duplicate.

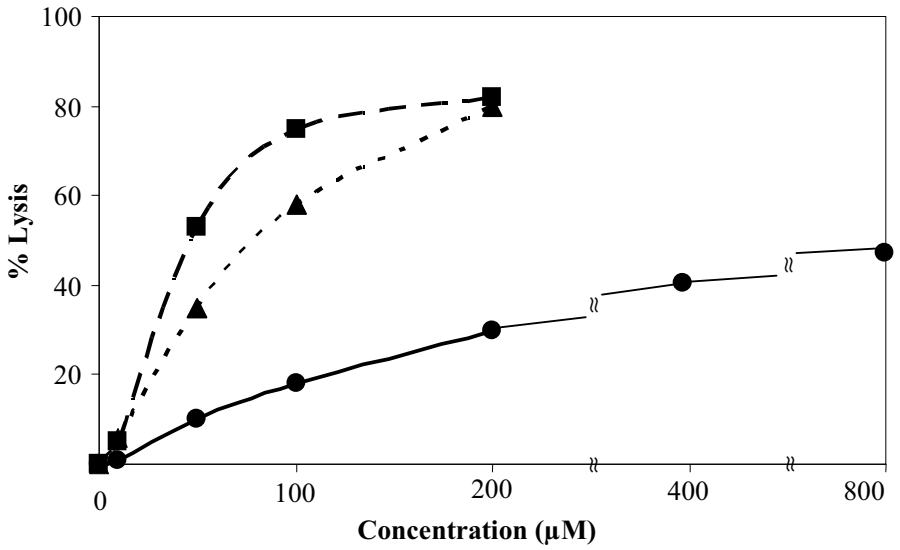
[3] Calculated from % lysis vs concentration plots, see Fig. 2 - the mean of two experiments performed in triplicate.

[4] TI, therapeutic index

[5] Calculated for the indicated concentration of peptide, at the 4^{th} hour, from bacterial growth kinetics curves, monitoring OD at 600nm (100% v/v MH broth, 10^6 CFU/ml logarithmic phase bacteria) - mean of two experiments performed in triplicate).


GMASKAGAIAGKIAKVALKAL


GLNALKKVFQGIHEAIKLINNHVQ


100
80
60
40
20
% Lysis
0
100
200
400
800
Concentration (µM)

[1]Peptide	[2]MIC (μM)			[3]HC_{50}	[4]TI	[5] **% bacterial growth inhibition** (*E. coli*)		
	E. coli	***P. aeruginosa***	***S. aureus***	(μM)	**(HC_{50}/MIC)**	2μM	4μM	8μM
DESC 1	2-4	16	>128	>800	200	10	75	100
ASC 1	2	16	32	80	39	65	100	100
PSEU 2	8	>64	32	50	6	0	25	100