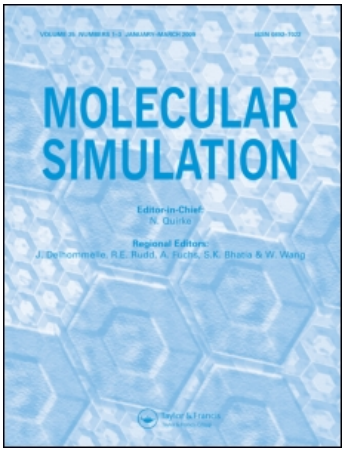

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Analysis of protegrin structure-activity relationships: the structural characteristics important for antimicrobial activity using smoothed amino acid sequence descriptors

M. Fernández[a]; J. Caballero[a]
[a] Molecular Modeling Group, Center for Biotechnological Studies, University of Matanzas, Matanzas, C.P., Cuba

To cite this Article Fernández, M. and Caballero, J.(2007) 'Analysis of protegrin structure-activity relationships: the structural characteristics important for antimicrobial activity using smoothed amino acid sequence descriptors', Molecular Simulation, 33: 8, 689 — 702
To link to this Article: DOI: 10.1080/08927020701236771
URL: http://dx.doi.org/10.1080/08927020701236771

Molecular Simulation, Vol. 33, No. 8, July 2007, 689–702

Taylor & Francis
Taylor & Francis Group

Analysis of protegrin structure–activity relationships: the structural characteristics important for antimicrobial activity using smoothed amino acid sequence descriptors

M. FERNÁNDEZ† and J. CABALLERO†‡*

†Molecular Modeling Group, Center for Biotechnological Studies, University of Matanzas, Matanzas, C.P. 44740, Cuba
‡Centro de Bioinformática y Simulación Molecular, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile

(Received January 2007; in final form January 2007)

Protegrin antimicrobial peptides (AMP) possess a high activity against a variety of microorganisms. In the present contribution, we analyse the structural requirements of protegrin analogues reported by Ostberg and Kaznessis (*Peptides* 2005; **26**: 197) for having antimicrobial activity against several microbial species by using interpretable QSAR models. Models were carried out using multiple linear regression (MLR) combined with genetic algorithm (GA) and smoothed amino acid sequence properties were employed for characterizing the peptide dataset. The main advantage of smoothing process is the alteration of local amino acid properties by the properties of the amino acids in the closer neighbourhood. We report models encompassing different characteristics for describing the activities against different microbial species. Our results suggest the existence of specific mechanisms of action for protegrin analogues against different microbial species.

Keywords: Protegrin; Antimicrobial peptides; Structure–activity relationships; Smoothed amino acid sequence properties

1. Introduction

Cationic antimicrobial peptides (AMPs) have been gaining recognition as highly valuable therapeutic agents for their roles in innate immunity and for their multiple biological effects [1]. They can possess antimicrobial activity against Gram-positive and Gram-negative bacteria, fungi and protozoa. Certain AMPs have been shown to inhibit the replication of enveloped viruses such as influenza A virus [2], vesicular stomatitis virus and human immunodeficiency virus (HIV-1) [3]; and may also possess anticancer activity [4] or promote wound healing [5]. Since the action of AMPs involves disruption of microbial membranes, they limit the ability of bacteria to develop resistance [6]. In this sense, AMPs can be considered as new antibiotics with a more efficient antimicrobial action.

AMPs are defined as peptides of less than 50 amino acid residues with an overall positive charge, imparted by the portion of hydrophobic residues. AMPs can be found presence of multiple lysine and arginine residues, and a substantial in simple and complicated organisms including humans. They have been preserved through evolution for their small size and their high potency [7]. Protegrins are short natural AMPs found in porcine neutrophils, similar to other defensins, which display a broad spectrum of antimicrobial effects against Gram-negative, Gram-positive bacteria, yeasts and fungi [8–11]. Protegrin-1 (PG-1) is 18 residues long and forms a rigid antiparallel two-stranded β-sheet linked by a short two-residue loop segment and stabilised by two disulfide bridges. Analogues retaining the crosslinked β-hairpin secondary structure conserve their potency in spite of various substitutions, while the linear ones exhibit reduced antimicrobial activity [9,12]. In addition, PG-1 has a C-terminal NH_2 group which has been kept on reported analogues. PG-1 structure can be modified by amino acid substitutions without an impairment of the antimicrobial activity; therefore overall structural features such as amphipathicity, charge and shape are more important to activity than the presence of specific amino acids. The total number of cationic residues and the amphipathicity of the β-hairpin determine the activity; however, the analogues displayed relatively high toxicity to mammalian cells [9,11].

*Corresponding author. Tel.: + 56-71-201-662. Fax: + 56-71-201-561. Email: jmcr77@yahoo.com, jcaballero@utalca.cl

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020701236771

Predicting protein physico-chemical properties and biological activities is a fundamental goal in molecular biology. The profitableness of sequence information has been highly confirmed for modeling large proteins including conformational stability [13], secondary and tertiary structures [14] or protein–ligand interactions [15], etc. The same applications have emerged for modeling biological functions of small peptides in recent years. Peptides have attracted considerable pharmacological interest due to they role as hormones, enzyme inhibitors, antibodies, olfaction and taste receptors, antimicrobial compounds or agents, and other biological functions.

The most common structural information employed for quantitative structure–activity relationships (QSAR) of peptides is provided from their macroscopic properties, e.g. molecular weight, lipophilicity, formal charge, solvent accessible surface area, etc. However, it is difficult to predict the effect of amino acid substitution on peptide descriptors derived from macroscopic properties. More explicit use can be found for a QSAR model if it indicates how the change in peptide sequence is correlated with the variation in biological activity and how to modify the sequence to achieve the improved activity. In this sense, molecular descriptors describing amino acid properties have been proposed. The main advantage of QSAR models using amino acid descriptors is their easy interpretability, since the information can be used directly to design new compounds.

In the early 1960s, Sneath [16] derived amino acid descriptor variables from physico-chemical semiqualitative data for the 20 coded amino acids and used them in a quantitative sequence-activity model analysis of oxytocin–vasopressine analogues. Afterwards, various approaches for encoding amino acid sequence information have been explored [17–23]. Different methods to transform the amino acid sequence of proteins and peptides into multivariate data for QSAR modelling are discussed in Ref. [24]. Hereafter, some recent approaches are discussed. A valuable contribution was presented by Hellberg *et al.* [19], who derived the *z*-descriptors, which were obtained by applying principal component analysis to groups of physico-chemical variables describing lipophilicity, hydrophilicity, size and charge-related properties of the amino acids. Recent applications of *z*-descriptors include predicting the activity of β-lactam antibiotics [25], modeling the antibacterial and anti-HSV activity of lactoferricin analogues [26,27], and describing the binding of peptides to the human class I major histocompatibility complex (MHC) allele HLA-A*0201 [28]. Very recently, Mei *et al.* [21] derived the VHSE scales from principal components analysis (PCA) on hydrophobic, steric and electronic properties of the 20 coded amino acids. Authors applied their descriptors to QSAR modeling of bitter-tasting dipeptides (BTD), angiotensin-converting enzyme (ACE) inhibitors, and bradykinin-potentiating pentapeptides (BPP). An adequate use of sequence information was developed by Pripp [29] who modeled inhibition of prolyl oligopeptidase by peptides derived from β-casein using simple amino acid descriptors encoding hydrophobicity, molecular bulkiness and isoelectric charge. The author assumed that prolyl oligopeptidase inhibition was related to proline residues in the endopositions of inhibitory peptides and considered properties of residues adjacent to the prolines in N- and C-terminal directions, i.e. positions P3, P2, P1′ and P2′ according to the proteolytic activity on peptides, for QSAR modeling.

The use of amino acid descriptors only requires the primary amino acid sequence, and replaces the measuring of the properties of the peptide. However, the consideration of amino acids as isolate points in the sequence is far from reality. Physico-chemical properties of amino acids inside a peptide are influenced by the solvent media and the peptidic surroundings [30]. In order to consider the peptide effects, we propose a local regression smoothing process for assigning an influence to a local amino acid descriptor from their neighbored amino acids. This transformation attempts simulating a pseudo-global effect. Smoothed amino acid sequence descriptors were tested in sequences of AMP.

The activity of natural and synthetic protegrins analogues against several microbial species have been recently reported [6,9,31,32]. In addition, these datasets have been studied by computational methods. In a recent paper, Ostberg and Kaznessis [31] used descriptors from three-dimensional homology models of PG-1 for carrying out QSAR models describing antimicrobial activity against six microbial species, and hemolytic and cytotoxic effects of protegrin analogues. They find that some terms describing the size, shape and energy of the peptide can be correlated with peptide activities. Another approach was carried out by Frecer [33] on cyclic AMPs derived from PG-1 with simple additive molecular properties related to the general mechanism of cell membrane disruption. By means of QSAR models, author proposed site-directed residue substitutions leading to simultaneous optimization of the antimicrobial and hemolytic potencies. The need of the development of computational models is clear: the number of AMPs that have been chemically and biologically characterized continues to grow, but the number of those with available high-resolution structures remains relatively small [34]. Our current contribution utilizes the simple primary structure of protegrin analogues for carrying out interpretable structure–activity relationships.

2. Materials and methods

2.1 Protegrin sequences and biological activities

The primary structure of 56 peptides analogues of PG-1 was compiled from the literature [31]. The minimal inhibitory concentrations (MICs: μg/ml) against Gram-negative

bacteria (*Escherichia coli*, *Psendomonas aeruginosa*, *Neisseria gonorrhoeae*—two strains) and Gram-positive bacteria (*Listeria monocytogenes*), and the yeast *Candida albicans* are collected and transformed in $\log(10^3/\mathrm{MIC})$ values. Peptide sequences and biological activities used in this study are summarized in table 1. Inactive peptides were not considered for QSAR modeling.

2.2 Amino acid sequence derived properties. Local regression smoothing process

The matrix of structural descriptors was generated using properties of the individual amino acids (table 2). Eight physico-chemical properties of 20 coded amino acids were employed: overall amino acid composition percentages (AAC) [35], average flexibility indexes (AF) [36], bulkiness (*B*) [37], hydrophobicities (*H*) as motilities of amino acids on chromatography paper [38], molecular weights (MW), polarities (*P*) [39], refractivities (*R*) [40] and relative mutabilities (RM) [41]. The combination of sequences with properties leads to property-sequence representations (PSRs). In a PSR, the presence of an amino acid in the sequence is described by the value of the physico-chemical property.

For considering the effects of properties of vicinal amino acids in a punctual amino acid, PSRs were modified by smoothing the physico-chemical properties using the Lowess method [42]. The name "Lowess" is derived from the term "locally weighted scatter plot smooth", the method uses locally weighted linear regression to smooth data, giving more weight to points near the point whose response is being estimated and less weight to points further away. For the smoothing process each amino acid is considered as a central point. Each smoothed value is determined for neighboring amino acids defined within a span. The span is a percentage of the total number of amino acids in the sequence; it dictates the number of neighboring amino acids influenced by the regression weight function.

The local regression smoothing process is developed for each amino acid by calculation of the regression weights for each amino acid in the span. The weights (w_i) are given by the tricube function shown below.

$$w_i = \left(1 - \left|\frac{x - x_i}{d(x)}\right|^3\right)^3 \qquad (1)$$

where x is the predictor value associated with the response value to be smoothed, x_i are the nearest neighbors of x as defined by the span, and $d(x)$ is the distance along the abscissa from x to the most distant predictor value within the span. A weighted linear least squares regression is performed using a first degree polynomial. The smoothed value is given by the weighted regression at the predictor value of interest. If the smooth calculation involves the same number of neighboring data points on either side of the smoothed central point, the weight function is symmetric. However, if the number of neighboring points is not symmetric about the smoothed central point, then the weight function is not symmetric. The weight function for an end point and for an interior point of peptide sequence is shown in figure 1.

Lowess smoothing was carried out using Proteinplot tool implemented in Bioinformatics toolbox of MATLAB [43].

2.3 Pool of descriptors and modeling procedure

Descriptors modified by Lowess smoothing are the primary structural information available for estimating models describing the relationship between activities and peptide structures. A consideration must be done for generating descriptor matrix, because protegrin analogues contain from 10 to 18 amino acids. Sequences with different numbers of amino acids produce a data set with different numbers of descriptors, causing problems during multivariate regression analysis. QSAR can be limited to equal number of amino acid derivatives if a minor fragment of the sequence is responsible for activity. Since protegrin analogues containing ten amino acids display antimicrobial activity, each sequence was aligned to PG-1 and coincident residue positions were determined (alignment was carried out using Bioinformatics toolbox of MATLAB [43]). In consequence, the pool of descriptors was generated including properties of amino acids located inside a reduced sequence: from position 6 to 15. In all, 80 descriptors were calculated, where each descriptor in the data set expresses the sequence position and an amino acid property. For each activity, descriptors that stayed constant or almost constant were eliminated, and pairs of variables with a correlation coefficient greater than 0.7 were classified as intercorrelated, and only the most correlated with modeled activity was included in the model.

After descriptor matrix has been generated, a key issue is to choose the most appropriate descriptors for modeling the antimicrobial activities. Relevant amino acid descriptors were chosen by using multiple linear regression analysis (MLR) with genetic algorithm (GA). Then, outliers were established and models were redone. Finally, models were validated by cross-validation to examine their predictive abilities.

3. Results and discussion

3.1 Analysis of smoothed amino acid sequence descriptors

We first illustrate how the smoothed amino acid sequence descriptors represent certain primary structural characteristics of a peptide and how the smoothing process influences the sequence representations. Traditionally, the difference between two sequences containing one different amino acid is only expressed by physico-chemical properties of the unequal amino acids. However,

Table 1. Experimental and predicted antimicrobial activities of PG-1 analogues ($\log(10^3/\text{MIC})$)[†].

		EC[‡]		*NG*[¶] *(F-62)*		*NG*[¶] *(FA19)*		*LM*[§]		*CA*[‖]		*PA*[#]	
Name	*Sequence*	*Exp*	*Pred*	*Exp*	*Pred*	*Exp*	*Pred*	*Exp*	*Pred*	*Exp*	*Pred*	*Exp*	*Pred*
PC001	rggrlcycrrrfcvcvgr	3.05	2.98	2.92	2.89	2.77	2.74	3.05	outl	2.28	1.97	3.05	2.84
PC003	rggglcycrrrfcvcvgr	2.64	2.91	2.85	2.88	2.74	2.74	2.55	2.63	2.02	1.98	–	–
PC004	rggrlcycrgwicfcvgr	2.68	2.87	2.92	2.97	2.46	2.54	2.72	2.60	1.52	1.52	–	–
PC005	rggrlcycrPrfcvcvgr	2.54	2.88	2.85	3.20	2.80	2.71	2.60	2.63	1.81	1.93	–	–
PC006	rggrlaycrrrfcvavgr	2.68	2.70	2.96	2.98	2.26	2.31	2.59	2.81	1.80	1.80	–	–
PC007	rggrlcyarrrfavcvgr	–	–	2.23	2.28	1.49	1.81	–	–	1.73	1.71	–	–
PC009	lcycrrrfcvcvgr	2.66	2.81	3.00	2.68	2.20	2.27	1.63	outl	1.21	outl	–	–
PC010	rcycrrrfcvcvgr	2.96	2.91	3.10	3.07	2.74	2.70	2.10	2.43	1.50	1.75	–	–
PC011	rggrlcycrrrfcvcv	2.89	3.07	2.82	2.89	2.42	2.39	2.43	2.80	1.94	1.96	–	–
PC012	rggrlcycrrrfcicv	2.72	2.85	2.89	2.49	2.37	2.12	2.26	2.38	1.69	1.81	2.51	2.78
PC013	rggrlcycrrrfcvcr	2.64	2.79	2.92	2.89	2.72	2.88	2.57	2.74	1.99	2.19	–	–
PC014	rcycrrrfcvcr	2.54	2.72	2.96	3.07	2.85	2.84	2.48	2.64	2.00	1.97	–	–
PC015	lcycrrrfcvcv	2.82	2.90	2.64	2.68	2.01	1.92	2.62	2.61	2.06	2.07	2.52	2.65
PC016	lcyarrrfavcv	3.22	3.39	1.87	2.29	1.32	1.35	2.82	2.93	1.59	1.89	3.10	3.23
PC017	rcyarrrfavcr	3.00	3.21	2.92	2.68	2.62	2.27	3.15	2.96	1.98	1.79	3.22	3.19
PC018	laycrrrfcvav	2.92	2.82	2.80	2.72	1.93	1.51	2.85	3.11	I	–	–	–
PC019	raycrrrfcvar	2.52	2.64	0.86	outl	I	–	1.54	outl	1.53	1.78	–	–
PC020	cycrrrfcvcvgr	2.72	2.60	2.46	2.48	1.28	1.28	1.66	outl	1.17	1.38	–	–
PC021	rggrlcycrrrfcvc	2.74	2.84	I	–	I	–	1.44	1.61	I	–	–	–
PC037	lcytrrrftvcv	2.85	2.80	2.74	2.64	2.00	2.18	2.38	2.36	1.56	1.56	2.85	2.68
PC045	ltycrrrfcvtv	3.30	3.17	2.70	2.99	2.00	2.24	3.52	3.13	2.57	outl	3.30	3.14
PC064	lcytrPrftvcv	2.85	2.70	2.89	2.97	1.86	2.16	2.70	2.37	1.55	1.50	–	–
PC064a	lcytrgrftvcv	3.22	3.31	3.15	2.79	2.28	1.81	2.41	2.28	1.91	1.73	–	–
PC065	lcytfrPrfvcv	2.51	2.53	2.39	2.33	1.52	1.40	1.98	2.14	I	–	–	–
PC066	lcytfrgrfvcv	–	–	1.23	outl	0.80	1.12	–	–	–	–	–	–
PC069	cycrrrfcvcv	I	–	1.42	outl	0.89	0.93	I	–	I	–	–	–
PC070	Lcycrrrfcvc	2.57	2.67	2.49	2.54	1.49	1.65	1.58	1.42	I	–	–	–
PC071	Cycrrrfcvc	2.49	2.45	2.38	2.34	1.89	outl	2.40	2.36	1.50	1.55	2.03	1.94
PC072	Cycfrrfcvc	3.22	3.09	2.74	2.69	2.08	2.07	3.15	3.06	1.55	1.55	2.92	2.87
PC073	lcycrrrrcvcv	1.85	1.87	1.50	1.73	1.01	1.05	–	–	I	–	2.15	outl
PC074	lcycfrrrcvcv	2.09	2.50	2.82	2.73	1.95	1.82	–	–	1.55	1.48	2.01	1.95
PC077	lcycrfrrcvcv	3.00	2.75	–	–	–	–	2.46	2.46	1.46	1.37	2.51	2.58
PC078	lcycrrfrcvcv	2.74	2.79	–	–	–	–	2.74	2.65	I	–	2.43	2.74
PC079	ycycrrrfcvcvgr	2.46	2.93	–	–	–	–	2.06	outl	I	–	1.80	1.92
PC080	tcycrrrfcvcvgr	2.96	3.00	–	–	–	–	2.44	2.83	I	–	2.39	2.52
PC091	acycrrrfcvcvgr	3.15	2.68	–	–	–	–	2.82	2.74	2.05	1.88	–	–
PC092	vcycrrrfcvcvgr	3.15	3.10	–	–	–	–	2.85	2.74	1.95	2.02	–	–
PC093	icycrrrfcvcvgr	3.15	3.45	–	–	–	–	2.85	2.82	1.93	2.05	–	–
PC094	fcycrrrfcvcvgr	3.22	2.97	–	–	–	–	2.89	2.72	2.06	2.06	–	–
PC095	wcycrrrfcvcvgr	3.30	3.09	–	–	–	–	2.85	2.76	2.00	2.04	–	–
PC096	ecycrrrfcvcvgr	3.22	3.10	–	–	–	–	2.92	2.57	2.03	1.74	–	–
PC097	rggrlcycrrrfcvcy	3.22	3.05	–	–	–	–	2.96	2.84	1.97	2.28	–	–
PC098	rggrlcycrrrfcvct	3.15	2.85	–	–	–	–	–	–	1.66	1.91	2.37	2.84
PC100	rggrlcycrrrfcvca	3.10	2.92	–	–	–	–	–	–	2.41	outl	2.92	2.84
PC101	rggrlcycrrrfcvcl	3.15	3.13	–	–	–	–	–	–	2.36	2.06	2.85	2.84
PC102	rggrlcycrrrfcvci	3.10	3.10	–	–	–	–	–	–	2.44	2.05	2.92	2.84
PC103	rggrlcycrrrfcvcf	3.15	3.11	–	–	–	–	–	–	2.82	outl	3.15	2.84
PC104	rggrlcycrrrfcvcw	3.22	3.10	–	–	–	–	–	–	2.44	2.48	3.00	2.84
PC105	rggrlcycrrrfcvce	3.00	2.78	–	–	–	–	–	–	2.36	2.03	3.15	2.84
PC107	rlcytrgrftvcv	3.22	3.28	–	–	–	–	–	–	1.50	1.63	3.30	3.59
PC108	lcytrgrftvcvr	3.52	3.20	–	–	–	–	–	–	2.52	outl	3.40	3.26
PC112	lcycrrrfcycv	2.26	2.30	–	–	–	–	–	–	I	–	2.22	2.49
PC113	lcycrrrfcfcv	1.97	outl	–	–	–	–	–	–	I	–	1.63	outl
PC146	lcycrrrfctcv	3.00	2.82	–	–	–	–	–	–	1.96	1.90	2.85	2.79
PC147	lcycrrrfcgcv	3.15	3.07	–	–	–	–	–	–	2.02	1.96	3.05	2.83
PC148	lcycrrrfcwcv	2.89	outl	–	–	–	–	–	–	1.60	1.64	2.59	2.52

[†] C-terminal NH_2 is found on all the protegrins represented. Inactive compounds and outliers are indicated. I: inactive; outl, outlier.
[‡] EC: *Escherichia coli*.
[¶] NG: *Neisseria gonorrhoeae*.
[§] LM: *Listeria monocytogenes*.
[‖] CA: *Candida albicans*.
[#] PA: *Psendomonas aeruginosa*.

the smoothing process modifies the local properties of more than one amino acid. In consequence, short sequence portions containing identical amino acids can be described by different values of the same descriptor. Figure 2(a) shows the behavior of the AAC descriptor profile when an amino acid is changed inside the fragment of sequence considered for QSAR modeling (between residues 6 and 15). When the Lowess smoothing is applied, the modification of amino acid content at position 14 influences descriptor values from position 11 to 15.

Table 2. Property matrix.

Amino acid code	*AAC*[†]	*AF*[‡]	*B*[¶]	*H*[§]	*MW*[‖]	*P*[#]	*R***	*RM*[††]
A	6.6	0.36	11.50	5.1	89	8.1	4.34	100
C	0.9	0.35	13.46	0.0	121	5.5	35.77	20
D	7.7	0.51	11.68	0.7	133	13.0	13.28	106
E	5.7	0.50	13.57	1.8	147	12.3	17.56	102
F	2.4	0.31	19.80	9.6	165	5.2	29.40	41
G	6.7	0.54	3.40	4.1	75	9.0	0.00	49
H	2.5	0.32	13.69	1.6	155	10.4	21.81	66
I	2.8	0.46	21.40	9.3	131	5.2	18.78	96
K	10.3	0.47	15.71	1.3	146	11.3	21.29	56
L	4.8	0.37	21.40	10.0	131	4.9	19.06	40
M	1.0	0.30	16.25	8.7	149	5.7	21.64	94
N	6.7	0.46	12.82	0.6	132	11.6	12.00	134
P	4.8	0.51	17.43	4.9	115	8.0	10.93	56
Q	5.2	0.49	14.45	1.4	146	10.5	17.26	93
R	4.5	0.53	14.28	2.0	174	10.5	26.66	65
S	9.4	0.51	9.47	3.1	105	9.2	6.35	120
T	7.0	0.44	15.77	3.5	119	8.6	11.01	97
V	4.5	0.39	21.57	8.5	117	5.9	13.92	74
W	1.4	0.31	21.67	9.2	204	5.4	42.53	18
Y	5.1	0.42	18.03	8.0	181	6.2	31.53	41

[†] AAC: overall amino acid composition percentages.
[‡] AF: average flexibility indexes.
[¶] *B*: bulkiness.
[§] *H*: hydrophobicities.
[‖] MW: molecular weights.
[#] *P*: polarities.
[**] *R*: refractivities.
[††] RM: relative mutabilities.

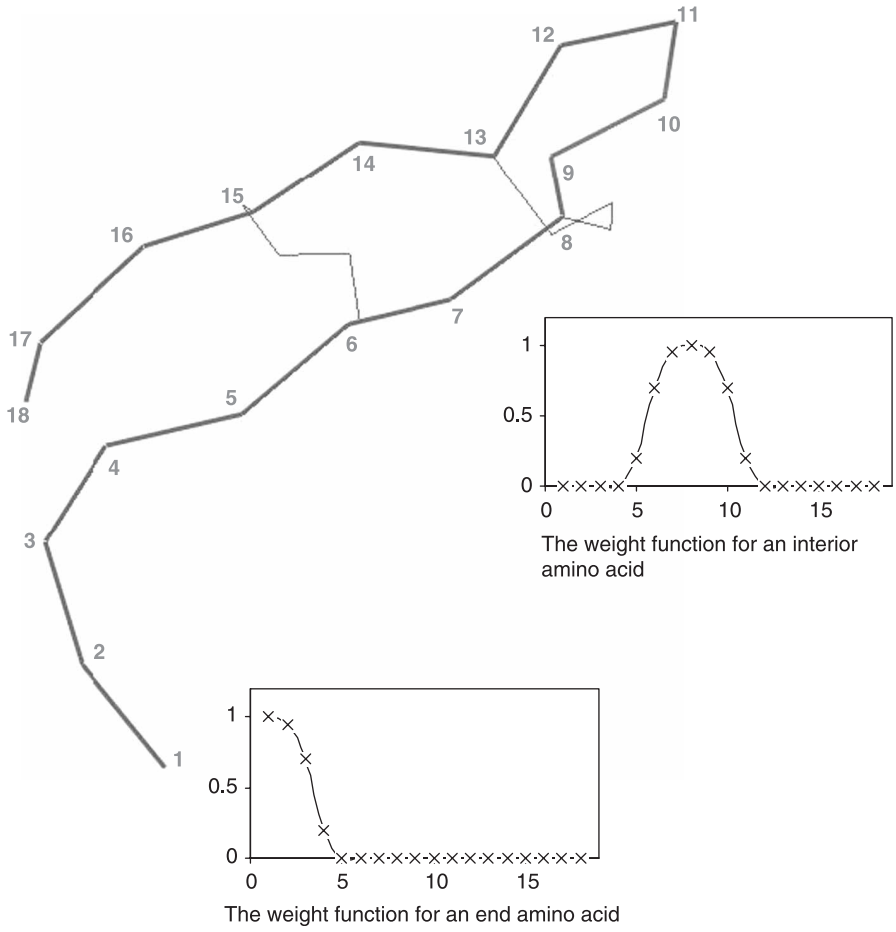

Figure 1. Lowess process over the sequence of PG-1. The weight function for an end point and for an interior point of peptide sequence.

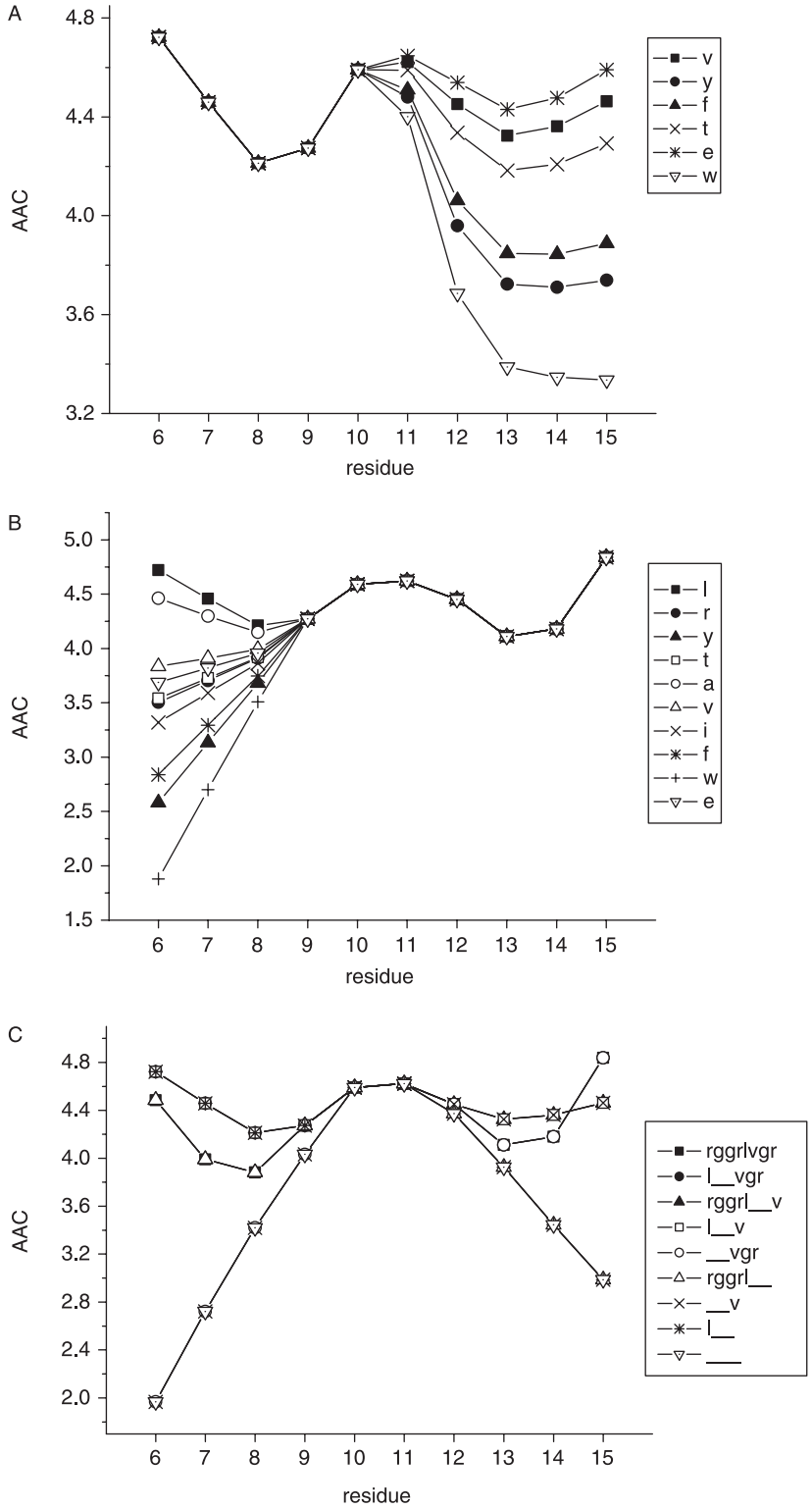

Figure 2. Behaviour of smoothed AAC descriptors inside the fragment of sequence considered for QSAR modelling (between residues 6 and 15). (a) AAC descriptors for homologous peptides (5 . . . lcycrrrfcxcv . . . 16) varying amino acid *x* at position 14, amino acids *x* are indicated in the legend at right, (b) AAC descriptors for homologous peptides (5 . . . xcycrrrfcvcvgr . . . 18) varying amino acid *x* at position 5, amino acids *x* are indicated in the legend at right, (c) AAC descriptors for peptides with the same primary structure 6 . . . cycrrrfcvc . . . 15, but varying the length of the sequences, amino acids before position 6 and after position 15 are indicated in the legend at right.

The reduction of the sequences can lead to several equal sequences; however, the smoothing process allows valuating the effects of the rest of the sequence. Peptides represented in figure 2(b) have the same sequence between residues 6 and 15, but they differ in residue at position 5. In this case, the smoothing process introduces the desired variability which was lost when the length of the sequence was reduced in order to equal the size of the

sequences for QSAR modeling. A similar effect is observed when peptides sharing the primary structure 6 . . . cycrrrfcvc . . . 15 are compared (figure 2(c)). When the sequence before 6 is rggrl, AAC values are 4.49, 3.99 and 3.88 for positions 6, 7 and 8 respectively; if the sequence is reduced maintaining only Leu-5, ACC values increase to 4.72, 4.46 and 4.21 for positions 6, 7 and 8, respectively. However, if the sequence begins with Cys-6, ACC values decrease to 1.97, 2.72 and 3.42 for positions 6, 7 and 8, respectively. On the other, when the sequence after 15 is vgr, AAC values are 4.11, 4.18 and 4.84 for positions 13, 14 and 15 respectively; if the sequence is reduced maintaining only Val-16, ACC values change to 4.32, 4.36 and 4.46 for positions 13, 14 and 15, respectively. However, if the sequence ends with Cys-15, ACC values decrease to 3.93, 3.45 and 2.99 for positions 13, 14 and 15, respectively.

3.2 QSAR models against different microbial species

Models relating smoothed amino acid sequence descriptors with several antimicrobial activities $\log(10^3/\mathrm{MIC})$ of protegrin analogues were carried out. Linear equations are reported as the linear combination which fit well with the data. The GA allows finding the most statistically significant properties which provide the simplest model for each organism. The key criterion for selecting the best models is the leave one out (LOO) cross-validation correlation coefficient (Q^2) above 0.5, according to Wold [44]. To gain a deeper insight into the impact of each descriptor in the linear models, we evaluated their relevance. For this, we chose to estimate the relative contributions of each descriptor on the antimicrobial activity. The descriptor under study was removed from the model and mean of the absolute deviation values Δmi between the observed and estimated value for all compounds were calculated. Finally, the contribution Ci [45] of descriptor i is given by:

$$Ci = \frac{100 \times \Delta mi}{\sum \Delta mi} \tag{2}$$

Since each descriptor corresponds to a physico-chemical property and an amino acid position, we can readily interpret the QSAR models based on the sign of descriptors in MLR equations and Ci values. However, it must be considered the effect of smoothing: each descriptor value is affected by its surroundings. Therefore, interpretations must be carried out by considering a pseudo-global effect.

The activity against *E. coli* was reported for 54 peptides. Before selecting the relevant descriptors, the inactive compound PC069 was excluded. The optimum QSAR model for 51 analogues is

$$\begin{aligned}\log(10^3/\mathrm{MIC}) = {} & -26.283 \times \mathrm{AF11} + 0.546 \times B8 \\ & - 0.635 \times B9 + 0.115 \times H15 \\ & + 0.241 \times R8 - 0.220 \times R12 \\ & + 0.031 \times \mathrm{RM6} + 12.889 \end{aligned} \tag{3}$$

$$N = 51 \quad R^2 = 0.681 \quad s = 0.207 \quad F = 13.05$$

$$Q^2 = 0.563 \quad s_{\mathrm{CV}} = 0.224$$

when N is the number of compounds included in the model, R^2 is the square correlation coefficient, s is the standard deviation of the regression and F is the Fischer ratio. Q^2 and s_{CV} are the correlation coefficient and standard deviation of the LOO cross-validation, respectively. Compounds PC113

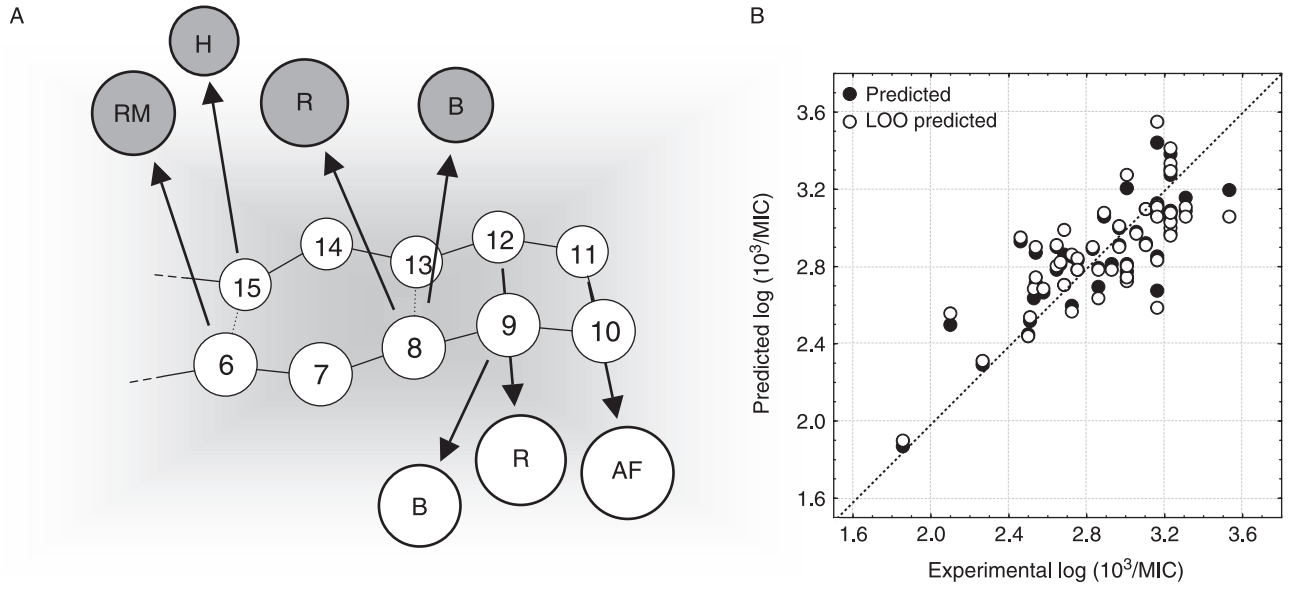

Figure 3. Model relating relevant physico-chemical properties with antimicrobial activity against *E. coli*. (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

and PC148 were identified as significant outliers, since they have unexpected biological activity and are unable to fit in the previous QSAR equation. MLR analysis gave, using seven variables in regression analysis, a model with $R^2 = 0.681$ and the predictability with full cross-validation was acceptable ($Q^2 = 0.563$). The model includes the averaged flexibility at position 11 ($Ci = 15.97$), bulkiness at positions 8 ($Ci = 12.94$) and 9 ($Ci = 14.17$), hydrophobicity at position 15 ($Ci = 11.90$), refractivities at positions 8 ($Ci = 15.06$) and 12 ($Ci = 15.58$), and relative mutability at position 6 ($Ci = 14.39$). Predictions are reported in table 1. Additional information about this model is represented in figure 3, including the scatter plot of the experimental activities versus predicted activities. Figure 3(a) shows that all descriptors display similar contributions. Due to the use of smoothing process, the relevant amino acid properties must be analyzed by considering the influence of proximate amino acids. For instance, the flexibility at position 11 increases when Phe at position 12 in PC015 is replaced by Arg (PC073) which causes a decrease of antimicrobial activity. $R12$ increases when Val at position 14 in PC015 is replaced by Tyr (PC112) which also causes a decrease of antimicrobial activity. The relative mutability at position 6 and hydrophobity at position 15 increase when Cys residues at the disulfide bond 8–13 in PC015 are replaced by Ala residues (PC016), which causes an increase of antimicrobial activity. As seen earlier (figure 2(c)), the value of descriptors of amino acids located at positions 6, 8 and 15 can be manipulated varying the length of the sequence instead of replacing Cys residues. In fact, the most active peptides against *E. coli* have a reduced sequence (from PG-1 initial sequence). In agreement with this, the bulkiness at position 8 increases for peptides in which sequence begins with Leu-5 instead of the larger sequence 1 . . . rggrl . . . 5 of PG-1.

The activity against *N. gonorrhoeae* (F-62) was reported for 31 peptides. Before selecting the relevant descriptors, the inactive compound PC021 was excluded. The equation (4) was the best QSAR model that was selected by GA for 27 derivatives, if the less active compounds PC019, PC066 and PC069 are considered outliers.

$$\log(10^3/\mathrm{MIC}) = -6.964 \times \mathrm{AF6} - 1.098 \times H8 + 2.124 \times P9 - 8.028 \times P11 + 6.277 \times P12 + 10.717 \quad (4)$$

$$N = 27 \quad R^2 = 0.733 \quad s = 0.214 \quad F = 11.53$$

$$Q^2 = 0.558 \quad s_{\mathrm{CV}} = 0.244$$

The model includes five variables and describes about 73% of the activity variance, while the predictability with cross-validation was acceptable ($Q^2 = 0.558$). The model includes the averaged flexibility at position 6 ($Ci = 15.33$), hydrophobicity at position 8 ($Ci = 20.41$), and polarities at positions 9 ($Ci = 20.97$), 11 ($Ci = 21.95$) and 12 ($Ci = 21.34$). Predictions are reported in table 1. Additional information about this model is represented in figure 4, including the scatter plot of the experimental activities versus predicted activities. Figure 4(a) shows that polarity terms appear to be providing a large portion of the structural information

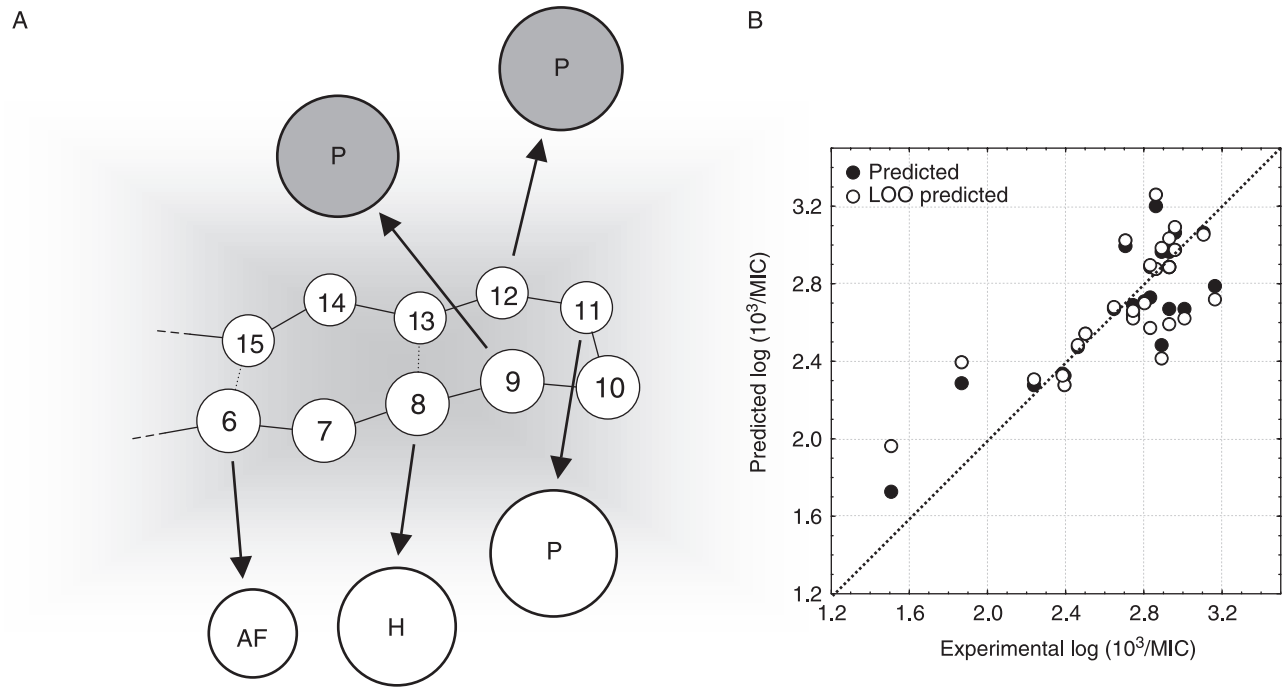

Figure 4. Model relating relevant physico-chemical properties with antimicrobial activity against *N. gonorrhoeae* (F-62). (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

needed to explain the activity against *N. gonorrhoeae* (F-62). In general, the positive global effect of polaribity terms and negative effect of hydrophobicity indicate the importance of the presence of Arg residues for this activity. The hydrophobicity at position 8 can be diminished by modifications in proximate amino acids. When Leu at position 5 in PC009 and PC015 is replaced by Arg (PC010 and PC014), *H*8 is reduced, increasing antimicrobial activity.

The activity against *N. gonorrhoeae* (FA-19) was reported for 31 peptides. Before selecting the relevant descriptors, PC019 and PC021 were excluded because they are inactive. From the data, we derived the following QSAR equation including 28 derivatives (compound PC071 is an outlier and was removed).

$$\log(10^3/\mathrm{MIC}) = -1.411 \times \mathrm{AAC8} + 4.982 \times P9 - 5.979 \times P10 + 2.260 \times P12 + 0.291 \times P15 - 0.961 \quad (5)$$

$$N = 28 \quad R^2 = 0.891 \quad s = 0.223 \quad F = 35.98$$

$$Q^2 = 0.834 \quad s_{\mathrm{CV}} = 0.245$$

The model includes five variables and describes about 89% of the activity variance, while the predictability with cross-validation was very satisfactory ($Q^2 = 0.834$). The model includes the overall amino acid composition percentages at position 8 ($Ci = 22.05$) and polarities at positions 9 ($Ci = 23.38$), 10 ($Ci = 22.44$), 12 ($Ci = 18.12$) and 15 ($Ci = 14.01$). Predictions are reported in table 1. Additional information about this model is represented in figure 5, including the scatter plot of the experimental activities versus predicted activities. Since the correlation between activity against species of *N. gonorrhoeae* is very high ($R^2 = 0.96$; table 5 in Ref. [31]), similitude among models is expected. As it was evidenced for strain F-62, the main contributions for antimicrobial activity against strain FA-19 is also provided by polarity terms (figure 5(a)). The requirement of Arg residues for increasing antimicrobial activity is also evidenced for strain FA-19: the most active peptides with $\log(10^3/\mathrm{MIC}) > 2.7$ (PC001, PC003, PC005, PC010, PC013, PC014) have Arg residues at terminal positions; in contrast, peptides with the lowest activity ($\log(10^3/\mathrm{MIC}) < 1.4$: PC066, PC069, PC073) contain less polar residues at these positions.

The activity against *L. monocytogenes* was reported for 38 peptides. Before selecting the relevant descriptors, the inactive compound PC069 was excluded. The equation (6) was the best QSAR model that was derived for 32 derivatives, if the compounds PC001, PC009, PC019, PC020 and PC079 are considered outliers.

$$\log(10^3/\mathrm{MIC}) = -0.550 \times \mathrm{AAC7} - 44.586 \times \mathrm{AF12} + 7.564 \times \mathrm{AF15} + 0.399 \times H15 - 0.033 \times \mathrm{MW7} + 1.204 \times P13 + 15.485 \quad (6)$$

$$N = 32 \quad R^2 = 0.780 \quad s = 0.221 \quad F = 14.79$$

$$Q^2 = 0.648 \quad s_{\mathrm{CV}} = 0.250$$

The model includes six variables and describes 78% of the activity variance, while the predictability with cross-validation was satisfactory ($Q^2 = 0.648$). The model includes the overall amino acid composition percentages at position 7 ($Ci = 15.97$), averaged flexibilities at

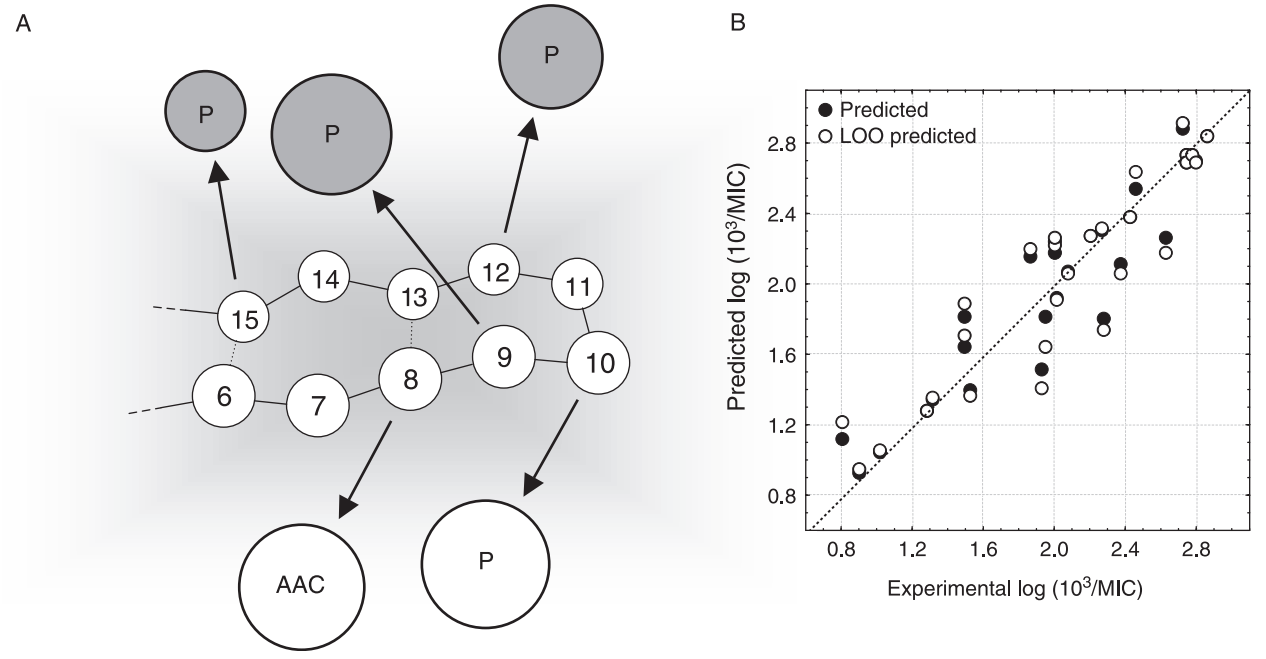

Figure 5. Model relating relevant physico-chemical properties with antimicrobial activity against *N. gonorrhoeae* (FA-19). (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

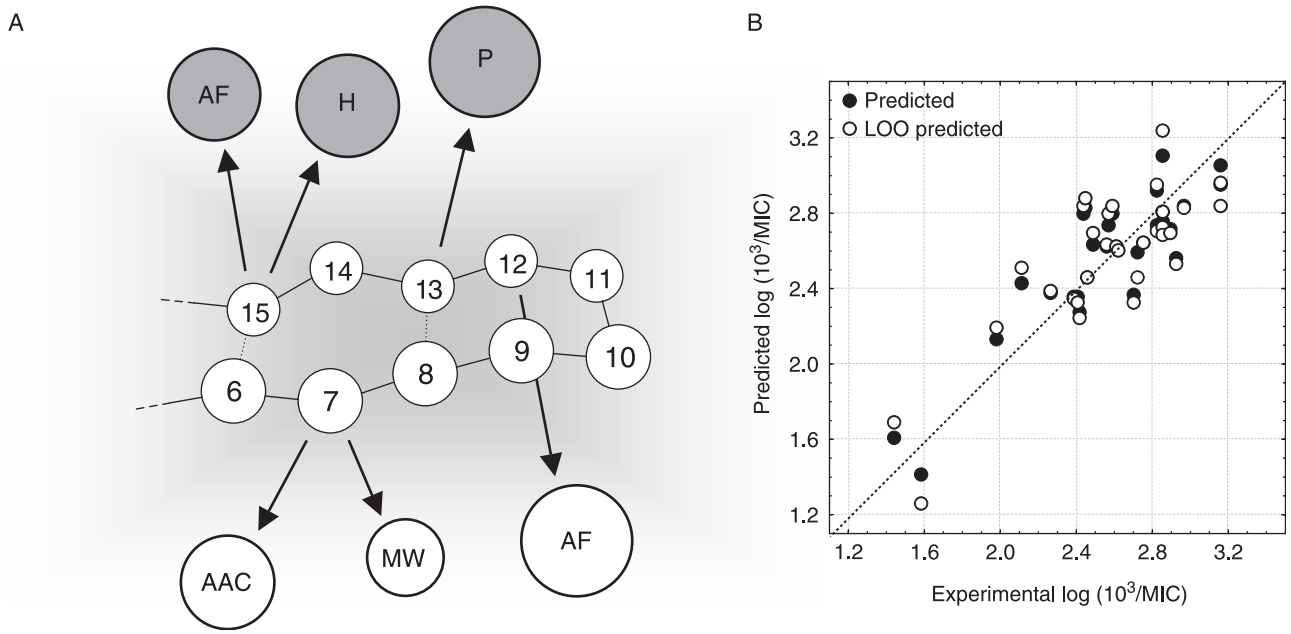

Figure 6. Model relating relevant physico-chemical properties with antimicrobial activity against *L. monocytogenes*. (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

positions 12 ($Ci = 18.99$) and 15 ($Ci = 15.65$), hydrophobicity at position 15 ($Ci = 17.46$), molecular weight at position 7 ($Ci = 13.10$) and polarity at position 13 ($Ci = 18.84$). Predictions are reported in table 1. Additional information about this model is represented in figure 6, including the scatter plot of the experimental activities versus predicted activities. Figure 6(a) shows that all descriptors display similar contributions. Due to the use of smoothing process, the positive or negative effects can be modulated by modifying proximate amino acids. For instance, AAC7 and AF12 decrease when Arg at position 9 in PC071 is replaced by Phe (PC072), which causes an increase of antimicrobial activity. The molecular weight at position 7 decreases when Cys residues at disulfide bonds 6–15 and 8–13 in PC015 are replaced by Ala (PC016 and PC018), which causes an increase of antimicrobial activity. This substitution also increases *H*15 and P13 descriptors. AF15, *H*15 and P13 decrease for peptides in which Cys15 is the C-terminal amino acid like PC021 in comparison with peptides in which C-terminal

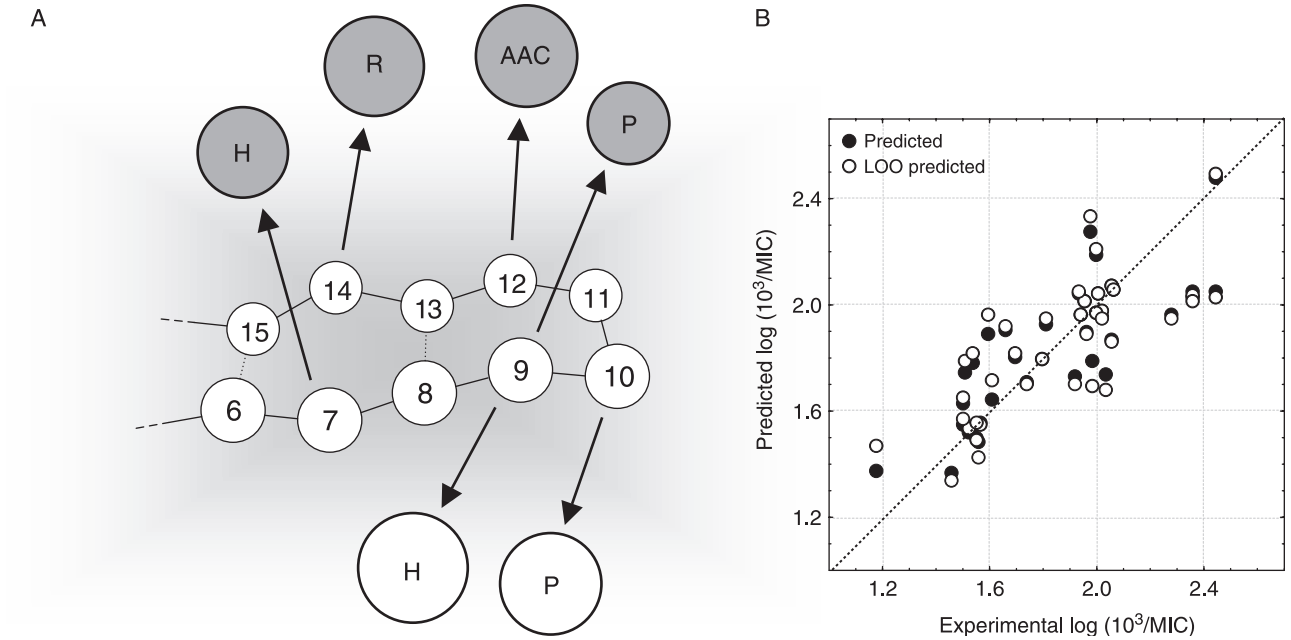

Figure 7. Model relating relevant physico-chemical properties with antimicrobial activity against *C. albicans*. (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

amino acid is Val16 (PC011); such modification causes a decrease of antimicrobial activity.

The activity against *C. albicans* was reported for 55 peptides. Before selecting the relevant descriptors, compounds PC018, PC021, PC065, PC069, PC070, PC073, PC078, PC079, PC080, PC112 and PC113 were excluded because they are inactive. The QSAR model is presented in equation (7) (compounds PC009, PC045, PC100, PC103 and PC108 were identified as outliers).

$$\log(10^3/\mathrm{MIC}) = 1.126 \times \mathrm{AAC12} + 0.180 \times H7 - 0.917 \times H9 + 0.344 \times P9 - 1.416 \times P10 + 0.079 \times R14 + 6.102 \quad (7)$$

$$N = 39 \quad R^2 = 0.664 \quad s = 0.192 \quad F = 10.53$$

$$Q^2 = 0.558 \quad s_{\mathrm{CV}} = 0.201$$

MLR analysis gave, using six variables in regression analysis, a model with $R^2 = 0.664$ and the predictability with full cross-validation was acceptable ($Q^2 = 0.558$). The model includes the overall amino acid composition percentages at position 12 ($Ci = 17.42$), hydrophobicities at positions 7 ($Ci = 15.47$) and 9 ($Ci = 18.78$), polarities at positions 9 ($Ci = 13.96$) and 10 ($Ci = 17.35$), and refractivity at position 14 ($Ci = 17.02$). Predictions are reported in table 1. Additional information about this model is represented in figure 7, including the scatter plot of the experimental activities versus predicted activities. Figure 7(a) shows that all descriptors display similar contributions. Due to the use of smoothing process, the positive or negative effects can be modulated by modifying proximate amino acids. For instance, *R*14 increases when Val residue at position 16 in PC011 is replaced by the most refractive residue Trp (PC104), which causes an increase of antimicrobial activity. AAC12 decrease when Val at position 14 in PC015 is replaced by Trp (PC148), which causes a decrease of antimicrobial activity. The requirements involving descriptors at position 9 for increasing activity suggested by equation (7) (higher P9 and lowest *H*9) are maintained in peptides containing 9 . . . rrrf . . . 12 subsequence. Changes in this subsequence (9 . . . frrr . . . 12 in PC074 or 9 . . . rfrr . . . 12 in PC077) decrease P9 and increase *H*9 leading to the deterioration of the antimicrobial activity.

The activity against *P. aeruginosa* was reported for 29 peptides. The equation (8) was the best QSAR model that was selected by GA for 27 derivatives (compounds PC073 and PC113 were identified as outliers).

$$\log(10^3/\mathrm{MIC}) = 0.451 \times \mathrm{AAC6} - 0.923 \times \mathrm{AAC12} - 31.365 \times \mathrm{AF11} - 0.193 \times B6 - 0.227 \times R12 + 28.227 \quad (8)$$

$$N = 27 \quad R^2 = 0.787 \quad s = 0.223 \quad F = 15.49$$

$$Q^2 = 0.672 \quad s_{\mathrm{CV}} = 0.249$$

MLR analysis gave, using five variables in regression analysis, a model with $R^2 = 0.787$ and the predictability with full cross-validation was satisfactory ($Q^2 = 0.672$). The model includes the overall amino acid composition percentages at positions 6 ($Ci = 24.25$) and 12 ($Ci = 17.49$), flexibility at position 11 ($Ci = 20.11$), bulkiness at position 6 ($Ci = 17.45$), and refractivity at position 12 ($Ci = 20.70$). Predictions are reported in table 1. Additional information about this model is represented in figure 8, including the scatter plot of the

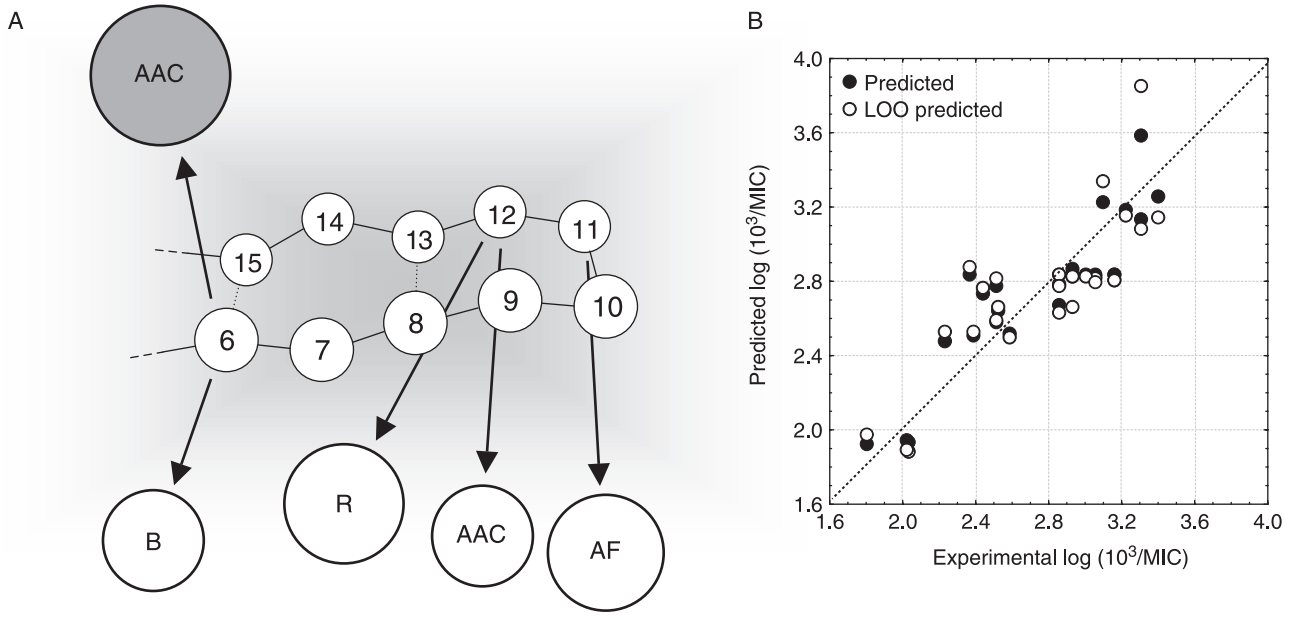

Figure 8. Model relating relevant physico-chemical properties with antimicrobial activity against *P. aeruginosa*. (a) Influence of descriptors: positive are gray and negative are white. Diameters of the balls indicate the contribution (C_i). (b) Scatter plot of the experimental activities versus predicted activities: (●) training predictions and (○) LOO cross-validated predictions.

experimental activities versus predicted activities. Figure 8(a) shows that all descriptors display similar contributions. Since there are not relevant properties involving amino acids above position 12, the model does not consider changes above position 15 as relevant for antimicrobial activity. The analysis of amino acid properties considering their close surroundings is dictated by the smoothing process. For instance, the overall amino acid composition percentage at position 6 increases and refractivity at position 12 decreases when Cys residues at the disulfide bond 8–13 in PC015 are replaced by Ala (PC016); the same effect is observed when Cys residues at the disulfide bond 6–15 in PC015 are replaced by Thr (PC045). Such modifications cause an increase of antimicrobial activity. *R*12 also decreases when Arg-10 in PC037 is replaced by Gly (PC108), which causes an increase of antimicrobial activity. *B*6 decreases when Tyr residue at position 5 in PC079 is replaced by Thr (PC080), which causes an increase of antimicrobial activity. AF11 increases when Ala-8 in PC016 is replaced by Thr (PC037) and when Arg-9 in PC015 is replaced by Phe (PC074), such modification decrease the antimicrobial activity.

3.3 Comparison with previous models

Previous QSAR studies have identified the main characteristics of the protegrin sequences that made them more active as the presence of disulfide bonds and higher positive charges [6,31]. These characteristics have been described by QSAR models reported by Ostberg and Kaznessis [31]. However, the analysis of the data reported by these authors suggests that different characteristics correspond to activities against different microbial species. The structure–activity trends captured by our models reflect these differences. In general, the presence or absence of disulfide bonds highly influences the activities against *E. coli*, *L. monocytogenes* and *P. aeruginosa*. The length of the sequences is another factor influencing the activities against *E. coli* and *L. monocytogenes*. In the other hand, the presence of higher positive charges is the main feature in the activity against *N. gonorrhoeae*. Meanwhile, the activity against *C. albicans* is sensible to modifications in the inner subsequence.

The above-mentioned differences must be traduced as differences in protegrin–microbial specie interactions. The antimicrobial activity is evidenced as cell lysis process, where the cytoplasmic membrane have been proposed as the ultimate target of AMP. Cell lysis is proposed to proceed via numerous mechanisms involving membrane perforation, disruption and solubilization [46]. Being cationic, the peptides displace the natural Ca^{2+} and Mg^{2+} ions and interact electrostatically with the negatively charged phospholipid headgroups; thus, they insert into the membrane bilayer in a manner that leads to its destabilization. Numerous studies have demonstrated that the peptides' physico-chemical properties, i.e. secondary structure with amphipathic properties, positive charge content, and hydrophobicity, are the main factors affecting membrane lysis activity. However, their precise mechanism of action is not fully understood. Our results suggest that these above-mentioned factors induce membrane lysis in some specific way for different microorganisms.

Shown in table 3 is the comparison between the statistics of our models and the previous ones reported by Ostberg and Kaznessis [31]. All the previous models have similar or lower R^2 value. We cannot compare the predictive capacity because referred authors did not assess the predictive power of their models. As Ostberg and Kaznessis did, our objective is to describe structure–activity trends instead to assess the predictive capacity of our models. In this sense, a higher Q^2 is a criterion of reliability of our models. The local character of the descriptors brings some advantages. Current models allow easy interpretation and offer clear guidelines for peptide optimization or design. The amino acid properties can be easily derived from peptide sequences and available data on amino acids. It is possible to predict the effect of a substitution using the property matrix (table 2).

4. Conclusions

A new method was applied to model the antimicrobial activity of recently reported protegrin analogues. In the application of the MLR procedure, we reported predictive equations for studying the effects of varying the PG-1 sequence. We found that different characteristics guide the activities against different microbial species. This aspect indicates the existence of specific protegrin–microbial

Table 3. Statistics of the QSAR models and comparison with previous ones in Ref. [31].

	Current models				*Models from Ref. [31]*		
Microbial specie	*N*	*vars*[†]	R^2	Q^2	*N*	*vars*[†]	R^2
Escherichia coli	51	7	0.681	0.563	55	5	0.680
Neisseria gonorrhoeae (F62)	27	5	0.733	0.558	28	4	0.515
Neisseria gonorrhoeae (FA-19)	28	5	0.891	0.834	27	4	0.480
Listeria monocytogenes	32	6	0.780	0.648	36	3	0.635
Candida albicans	39	6	0.664	0.558	45	5	0.600
Psendomonas aeruginosa	27	5	0.787	0.672	32	2	0.670

[†] vars: number of variables included.

specie interactions, suggesting specific mechanisms for different microorganisms.

The obtained results clearly pointed out that the smoothed amino acid sequence descriptors may be considered useful for describing peptides. These descriptors have been found to provide unique information regarding molecular structure and have been found to make significant contributions to resulting equations. The smoothing process introduces the effects of the closer members of the sequence. Because of this, the gain of using the smoothing process is the ability of effectively code the biological content in sequences considering the chemical neighborhood. Our current work presents a new method on the use of amino acid descriptors for QSAR modeling in peptide research. This allows the use of simple amino acid descriptors for peptides that possess different lengths. The main advantage is that obtained models can be interpreted in order to guide peptide optimization or design.

References

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

[2] T. Murakami, M. Niwa, F. Tokunaga, T. Miyata, S. Iwanaga. Direct virus inactivation of tachyplesin I and its isopeptides from horseshoe crab hemocytes. *Chemotherapy*, **37**, 327 (1991).

[3] M. Masuda, H. Nakashima, T. Ueda, H. Naba, R. Ikoma, A. Otaka, Y. Terekawa, H. Tamamura, T. Ibuka, T. Murakami, Y. Koyanagi, M. Waki, A. Matsumoto, N. Yamamoto, S. Funakoshi, N. Fujii. A novel anti-HIV synthetic peptide, T-22 ([Tyr5,12,Lys7]-polyphemusin II). *Biochem. Biophys. Res. Commun.*, **189**, 845 (1992).

[4] M.A. Baker, W.L. Maloy, M. Zasloff, L.S. Jacob. Anticancer efficacy of magainin 2 and analogue peptides. *Cancer Res.*, **53**, 3052 (1993).

[5] R.L. Gallo, M. Ono, T. Povsic, C. Page, E. Eriksson, M. Klagsbrun, M. Bernfield. Syndecans, cell surface heparan sulfate proteoglycans, are induced by a proline-rich antimicrobial peptide from wounds. *Proc. Natl. Acad. Sci. USA*, **91**, 11035 (1994).

[6] J. Chen, T.J. Falla, H. Liu, M.A. Hurst, C.A. Fujii, D.A. Mosca, J.R. Embree, D.J. Loury, P.A. Rabel, C.C. Chang, L. Gu, J.C. Fiddes. Development of protegrins for the treatment and prevention of oral mucositis: structure–activity relationships of synthetic protegrin analogues. *Biopolymers*, **55**, 88 (2000).

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

[8] V.N. Kokryakov, S.S. Harwig, E.A. Panyutich, A.A. Shevchenko, G.M. Aleshina, O.V. Shamova, O.V. Korneva, R.I. Lehrer. Protegrins: leukocyte antimicrobial peptides that combine features of corticostatic defensins and tachyplesins. *FEBS Lett.*, **327**, 231 (1993).

[9] J.R. Lai, B.R. Huck, B. Weisblum, S.H. Gelman. Design of noncysteine-containing antimicrobial beta-hairpins: structure–activity relationship studies with linear protegrin-1 analogues. *Biochemistry*, **41**, 12835 (2002).

[10] K.T. Miyasaki, R.I. Lehrer. Beta-sheet antibiotic peptides as potential dental therapeutics. *Int. J. Antimicrob. Agents*, **9**, 269 (1998).

[11] J.P. Tam, C. Wu, J.L. Yang. Membranolytic selectivity of cystine-stabilized cyclic protegrins. *Eur. J. Biochem.*, **267**, 3289 (2000).

[12] R. Mani, A.J. Waring, R.I. Lehrer, M. Hong. Membrane-disruptive abilities of β-hairpin antimicrobial peptides correlate with conformation and activity: A ^{31}P and ^{1}H NMR study. *Biochim. Biophys. Acta*, **1716**, 11 (2005).

[13] J. Caballero, L. Fernández, J.I. Pérez, M. Fernández. Amino acid sequence autocorrelation vectors and ensembles of Bayesian-regularized genetic neural networks for prediction of conformational stability of human lysozyme mutants. *J. Chem. Inf. Model.*, **46**, 1255 (2006).

[14] C. Pasquier, V.J. Promponas, S.J. Hamodrakas. PRED-CLASS: cascading neural networks for generalized protein classification and genome-wide applications. *Proteins*, **44**, 361 (2001).

[15] J.R. Bock, D.A. Gough. Virtual screen for ligands of orphan G protein-coupled receptors. *J. Chem. Inf. Model.*, **45**, 1402 (2005).

[16] P.H. Sneath. Relations between chemical structure and biological activity in peptides. *J. Theor. Biol.*, **12**, 157 (1966).

[17] E.R. Collantes, W.J. Dunn. Amino acid side chain descriptors for quantitative structure–activity relationship studies of peptide analogues. *J. Med. Chem.*, **38**, 2705 (1995).

[18] J.L. Fauchère, M. Charton, L.B. Kier, A. Verloop, V. Pliska. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.*, **32**, 269 (1988).

[19] S. Hellberg, M. Sjöström, B. Skagerberg, S. Wold. Peptide quantitative structure–activity relationships, a multivariate approach. *J. Med. Chem.*, **30**, 1126 (1987).

[20] A. Kidera, Y. Konishi, M. Poka, T. Ooi, H.A. Scheraga. Statistical analysis of the physical properties of the 10 naturally occurring amino acids. *J. Protein Chem.*, **4**, 23 (1985).

[21] H. Mei, Z.H. Liao, Y. Zhou, S.Z. Li. A new set of amino acid descriptors and its application in peptide QSARs. *Biopolymers*, **80**, 775 (2005).

[22] M. Sandberg, L. Eriksson, J. Jonsson, M. Sjöström, S. Wold. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.*, **41**, 2481 (1998).

[23] A. Zaliani, E. Gancia. MS-WHIM scores for amino acids: a new 3D-description for peptide QSAR and QSPR studies. *J. Chem. Inf. Comput. Sci.*, **39**, 525 (1999).

[24] A.H. Pripp, T. Isaksson, L. Stepaniak, T. Sørhaug, Y. Ardö. Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science. *Trends Food Sci. Technol.*, **16**, 484 (2005).

[25] S. Wanchana, F. Yamashita, H. Hara, S. Fujiwara, M. Akamatsu, M. Hashida. Two- and three-dimensional QSAR of carrier-mediated transport of beta-lactam antibiotics in Caco-2 cells. *J. Pharm. Sci.*, **93**, 3057 (2004).

[26] H. Jenssen, T.J. Gutteberg, T. Lejon. Modelling of anti-HSV activity of lactoferricin analogues using amino acid descriptors. *J. Pept. Sci.*, **11**, 97 (2005).

[27] T. Lejon, T. Stiberg, M.B. Strøm, J.S. Svendsen. Prediction of antibiotic activity and synthesis of new pentadecapeptides based on lactoferricins. *J. Pept. Sci.*, **10**, 329 (2004).

[28] P. Guan, I.A. Doytchinova, V.A. Walshe, P. Borrow, D.R. Flower. Analysis of peptide–protein binding using amino acid descriptors: prediction and experimental verification for human histocompatibility complex HLA-A*0201. *J. Med. Chem.*, **48**, 7418 (2005).

[29] A.H. Pripp. Quantitative structure–activity relationship of prolyl oligopeptidase inhibitory peptides derived from β-casein using simple amino acid descriptors. *J. Agric Food Chem.*, **54**, 224 (2006).

[30] N. Gulyaeva, A. Zaslavsky, A. Chait, B. Zaslavsky. Relative hydrophobicity of di- to hexapeptides as measured by aqueous two-phase partitioning. *J. Peptide Res.*, **61**, 129 (0000).

[31] N. Ostberg, Y. Kaznessis. Protegrin structure–activity relationships: using homology models of synthetic sequences to determine structural characteristics important for activity. *Peptides*, **26**, 197 (2005).

[32] J.A. Robinson, S.C. Shankaramma, P. Jetter, U. Kienzl, R.A. Schwenderer, J.W. Vribloed, D. Obrecht. Properties and structure–activity studies of cyclic β-hairpin peptidomimetics based on the cationic antimicrobial peptide protegrin I. *Bioorg. Med. Chem.*, **13**, 2055 (2005).

[33] V. Frecer. QSAR analysis of antimicrobial and haemolytic effects of cyclic cationic antimicrobial peptides derived from protegrin-1. *Bioorg. Med. Chem.*, **14**, 6065 (2006).

[34] J.P.S. Powers, R.E.W. Hancock. The relationship between peptide structure and antibacterial activity. *Peptides*, **24**, 1681 (2003).

[35] P. McCaldon, P. Argos. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins*, **4**, 99 (1988).

[36] R. Bhaskaran, P.K. Ponnuswamy. Positional flexibilities of amino acid residues in globular proteins. *Int. J. Pept. Protein Res.*, **32**, 242 (1988).

[37] J.M. Zimmerman, N. Eliezer, R. Simha. The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.*, **21**, 170 (1968).

[38] A.A. Aboderin. An empirical hydrophobicity scale for α-amino-acids and some of its applications. *Int. J. Biochem.*, **2**, 537 (1971).
[39] R. Grantham. Amino acid difference formula to help explain protein evolution. *Science*, **185**, 862 (1974).
[40] D.D. Jones. Amino acid properties and side-chain orientation in proteins: a cross correlation approach. *J. Theor. Biol.*, **50**, 167 (1975).
[41] M.O. Dayhoff, R.M. Schwartz, B.C. Orcutt. A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure*, Vol 5, (Suppl 3), pp. 345–352, Natl Biomed Res Found, Washington DC (1978).
[42] W.S. Cleveland. Robust locally weighted regression and smoothing scatter plots. *J. Am. Stat. Assoc.*, **74**, 829 (1979).
[43] MATLAB, version 7.0; The Mathworks Inc. Natick, MA. http://www.mathworks.com.
[44] S. Wold. Validation of QSAR's. *Quant. Struct. Act Relat.*, **10**, 191 (1991).
[45] D. Cherqaoui, M. Esseffar, D. Villemin, J.M. Cence, M. Chastrette, D. Zakarya. Structure–musk odour relationships studies of tetralin and indan compounds using neural networks. *New J. Chem.*, **22**, 839 (1998).
[46] R.M. Epand, H.J. Vogel. Diversity of antimicrobial peptides and their mechanisms of action. *Biochim. Biophys. Acta.*, **1462**, 11 (1999).