Detection of selective antibacterial peptides by the Polarity Profile method

Carlos Polanco[1✉], Thomas Buhse[2], José Lino Samaniego[1] and Jorge Alberto Castañón-González[1]

[1]Facultad de Ciencias de la Salud, Universidad Anáhuac, Huixquilucan Estado de México, México; [2]Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Cuernavaca Morelos, México

Antimicrobial peptides occupy a prominent place in the production of pharmaceuticals, because of their effective contribution to the protection of the immune system against almost all types of pathogens. These peptides are thoroughly studied by computational methods designed to shed light on their main functions. In this paper, we propose a computational approach, named the Polarity Profile method that represents an improvement to the former Polarity Index method. The Polarity Profile method is very effective in detecting the subgroup of antibacterial peptides called selective cationic amphipathic antibacterial peptides (SCAAP) that show high toxicity towards bacterial membranes and exhibit almost zero toxicity towards mammalian cells. Our study was restricted to the peptides listed in the antimicrobial peptides database (APD2) of December 19, 2012. Performance of the Polarity Profile method is demonstrated through a comparison to the former Polarity Index method by using the same sets of peptides. The efficiency of the Polarity Profile method exceeds 85% taking into account the false positive and/or false negative peptides.

INTRODUCTION

The re-emergence of multi-drug-resistant organisms (Barie PS, 2012) has focused part of the research in bioinformatics on the search for fast and reliable computational procedures (Blueggel *et al.*, 2004) that allow for the detection or prediction of the primary pathogenic action of a peptide (most of the peptides found experimentally show non-specific pathogenic action). These procedures allow for checking peptides of short or medium lengths in fractions of a second, thus avoiding the high cost of an experimental analysis. However, the mathematical abstraction of physicochemical properties is not an easy task, especially when it comes to the discrimination between primary properties of the peptides and those derived from them.

The bioinformatics mathematical algorithms related to the analysis of biological structures are classified as "supervised learning" and "non-supervised learning" (Zhao *et al.*, 2011). The supervised algorithms require "training data" that are characteristic of the population that was intended to identify. The representatives of these algorithms are Quantitative Structure Activity Relationships (QSAR) (Gonzalez-Diaz H, 2012),

Hidden Markov models (Polanco & Samaniego, 2009), Monte Carlo methods (Perez-Riverol *et al.*, 2012), Support Vector Machines (Han *et al.*, 2012), and Fourier Transforms (Silverman BD, 2005). On the other hand, non-supervised algorithms do not use "training data". One of these algorithms is called clustering (Li *et al.*, 2008). Both groups offer varying degrees of difficulty in their computational implementation, however, the supervised algorithms are less difficult in their implementation.

In this manuscript we present a QSAR algorithm called the Polarity Profile method, which represents an improvement or modification to another approach that was formerly introduced by us (Polanco *et al.* 2012). Three relevant aspects have been improved.

(1) The Polarity Profile method is highly efficient in identifying the subgroup of antibacterial peptides named Selective Cationic Amphipathic Antibacterial Peptides (SCAAP) that are characterized by highly selective toxicity towards bacteria, not adopting the alpha-helicoidal structure in neutral aqueous solution, and showing the therapeutic index higher than 75 (del Rio *et al.*, 2001). The therapeutic index of a peptide is determined by the ratio between the minimum inhibitory concentration observed against mammalian and bacterial cells, i.e. the higher the value, the more specific the peptide for bacterial-like membranes. Hence, SCAAP display strong lytic activity against bacteria but no toxicity against normal eukaryotic cells such as erythrocytes.

(2) The algebraic structure of the Polarity Profile method (matrix with 4x4 interactions formed by the polar groups P, P-, N and NP) offers extensive polarity information about the peptide characteristics. Most of the supervised algorithms measure the polarity property as one number while the Polarity Profile method considers 16 numbers that correspond to the 16 possible polar interactions. Hence, major information about the dynamics of the phenomenon is obtained.

(3) The similarity found when comparing the polarity matrices of the bacteria group and a subset of this so-called SCAAP shows that the physicochemical property called polarity constitutes an important measurement.

The Polarity Profile method uses the classification of 20 proteic amino acids, differentiated by their side chain R and by dividing them into four different categories according to their polar profile (Australian, 2012) (Table 1).

Table 1. Classification of amino acids.

Symbol	Index	Category	1-letter code amino acid
P+	1	**Basic.** Polar amino acids with positive charges have more amino groups as compared to carboxyl groups making them basic. The amino acids, which have positive charges on the R group are placed in this category.	H, K, R
P-	2	**Acidic.** Polar amino acids with negative charges have more carboxyl groups than amino groups making them acidic. The amino acids, which have negative charges on their R group are placed in this category. They are called dicarboxylic mono-amino acids.	D, E
N	3	**Neutral.** These amino acids do not have any charges on their R group. These amino acids participate in hydrogen bonding of protein structure.	C, G, N, Q, S, T, Y
NP	4	**Non-Polar.** These amino acids have equal numbers of amino and carboxyl groups and are neutral. These amino acids are hydrophobic and have no charges on the R group.	A, F, I, L, M, P, V, W

Proteinogenic amino acids classification differentiated by their side-chain according to their polarity into four categories (Australian, 2012). Index: Numeric identity assigned to polarity group.

The method was verified with all peptides listed in the antimicrobial peptide database (APD2) from December 19, 2012 (Wang *et al.*, 2009) and showed a high discriminating capacity.

MATERIAL AND METHODS

The Polarity Profile method requires building different matrices in which the rows and columns represent four polarity groups in the order [P+] polar, [P-] acidic, [N] neutral, and [NP] non-polar residues (Table 1). The elements (i,j) represent 16 possible polar interactions. For example, the definition of the $\mathbf{A}[i,j]$ polarity matrix is represented as shown in Table 2 where the value 5 is located in row 2 and column 3, so that the element (2,3) of the $\mathbf{A}$ matrix is 5, i.e. $\mathbf{A}[2,3] = 5$.

We built the polarity matrix by adding up the number of incidences moving the peptide sequence in amino acid pairs from left to right, one at the time to the end. Each amino acid pair was related to its polarity group, associating the i-th row with j-th colum, and adding 1 to the matrix element, i.e. $\mathbf{A}[i,j] = \mathbf{A}[i,j] + 1$, thus obtaining the incidence matrix $\mathbf{A}[i,j]$. This means that the amino acids of the studied peptides, are converted to numbers {P, P-, N, NP} = {1,2,3,4} (Table 1, columns #1 and #2), then the sequence is read from left to right, where each pair of numbers indicate the coordinates (i,j) in the matrix.

To illustrate the use of this rule by considering the protein sequence **QIINNPITCMTNGAIC WGP-CPTAFRQIGNCGHFKVRCCKIR** (Table 9, entry #35), where the amino acid **Q** is assigned to the polarity equivalence 3, the amino acid **I** to the equivalence 4, etc., until the complete sequence is translated to **34433443343334434 34343441343331414133141**. Once the protein sequence has been converted into its equivalent series, the polarity matrix $\mathbf{O}[i,j]$ is built. Each element (i,j) accumulates the occurrence that is obtained by reading the series **344334433433443434343441 3433331414133141** from left to right and taking as element (i,j) the pair of numbers found by moving one digit at the time through the series. As an example, the polarity matrix $\mathbf{O}[i,j]$ corresponding to the series **344334 433433443434343441343333114 14133141** is expressed in Table 3.

Polarity Profile Method Description

The method considers comparing two matrices and determines a specific profile.

The steps are:

Table 2. Polarity matrix A[i,j].

	P+	P-	N	NP
P+	0	0	0	0
P-	0	0	5	0
N	0	0	0	0
NP	0	0	0	0

The Polarity matrix $\mathbf{A}[i,j]$ uses 20 amino acid classifications differentiated by their side chains that fall into four polarity groups: [P+] polar, [N] neutral, [P+] basic hydrophilic, and [NP] nonpolar residues (Table 1), where each row and column represents (i,j) 16 possible interactions between the groups.

Table 3. Polarity matrix O[i,j].

	P+	P-	N	NP
P+	0	0	2	3
P-	0	0	0	0
N	2	0	8	9
NP	4	0	8	4

Polarity $\mathbf{O}[i,j]$ matrix interaction in the polarity groups differentiated by their lateral chain to the sequence e.g. **QIINNPITCMTNGAICWGP-CPTAFRQIGNCGHFKVRCCKIR** (Section 2.1).

1. Building the $\mathbf{P}[i,j]$ matrix with the entire peptide set of SCAAP sequences with a unique pathogenic action. When the polarity matrix $\mathbf{P}[i,j]$ is concluded, it will be normalized to one. Under this rule, we consider 51 peptides of SCAAP extracted from the APD2. Under the rule described in Section 2.1 the $\mathbf{P}[i,j]$ matrix is obtained as shown in Table 4.

2. Building the $\mathbf{O}[i,j]$ matrix with the sequence-objective of the study. The $\mathbf{O}[i,j]$ matrix is not to be normalized. For instance, we took the sequence **QIINNPITC MTNGAICWGPCPTAFRQIGNCGHFKVRCCKIR** and calculated its O[i,j] matrix (Table 3).

3. Each element of the matrix $\mathbf{P}[i,j]$ is multiplied by the corresponding element in the matrix O[i,j]. The result of this operation is multiplied by the factor 0.30. The resulting matrix of this operation will be defined with the operator $(\otimes)$ as $\mathbf{P}[i,j] \otimes \mathbf{O}[i,j]$.

4. Now, the matrices $\mathbf{P}[i,j]$ and $\mathbf{P}[i,j] \otimes \mathbf{O}[i,j]$ should restate, as matrices that identify, line by line, ordered frequency positions. This means that these matrices now contain what will be the positions, and not the frequency. The new matrix $\mathbf{P}[i,j] \otimes \mathbf{O}[i,j]$, for the sequence **QIINNPITCMTNGAICWGPCPTAFRQI-**

Table 4. Polarity matrix P[*i,j*].

	P+	P-	N	NP
P+	0.0415200554	0.0112596760	0.0267417319	0.1069669276
P-	0.0168895144	0.0007037297	0.0084447572	0.0098522166
N	0.0443349741	0.0000000000	0.0682617873	0.1182266027
NP	0.1069669276	0.0246305410	0.1238564402	0.2660098374

Polarity **P**[*i,j*] matrix presenting 51 SCAAP peptides with unique pathogenic action extracted from the APD2 (Wang *et al.*, 2009).

GNCGHFKVRCCKIR, is indicated in Table 5, and the matrix **P**[i,j] for 51 SCAAP (Table 4) is expressed in Table 6.

5. The Polarity Profile method qualifies as SCAAP candidates those peptides whose number of matches when comparing both matrices **P**[*i,j*] and **P**[*i,j*] ⊗ **O**[*i,j*] is greater than 60% (60% of 16 is 10 elements). For example, the peptide with the sequence **QIINNPITC-MTNGAICWGPCPTAFRQIGNCGHFKVRCCKIR** is accepted (Table 7).

APD2 Database Trial Data Preparation

We have studied and classified all 2169 peptides in the APD2 database by their unique and multiple action against: bacteria, viruses, fungi, parasites, insects, carcinogenic cells, mammalian cells, sperms, and SCAAP.

The set of 51 SCAAP was taken directly from previous records that have already been reported (Polanco *et al.*, 2012). The majority of these peptides fall in the bacteria group of the APD2. The bacteria group that is compared to the SCAAP group (Section 2.4) was also extracted from the APD2.

All groups extracted from the APD2 were analyzed (Table 9) and classified into two single sets: action vs. multiple action. The peptide sets with a unique pathogenic action are those peptides with confirmed experimental action against a single pathogen agent, whereas multiple action peptides show pathogenic action against two or more pathogens.

As an example of this classification, let us consider the set of anti-fungi peptides with unique action. This set is composed of peptide sequences that do not appear in other peptide subgroups, while the anti-fungi multiple action peptides contain peptides with an action against fungi and additionally with a possible action against other pathogenic groups.

We have validated the unique action peptide sets by matching peptides from the APD2 database with those identified by the Polarity Profile method. All obtained information was verified comparing the identified SCAAP with all peptides from the APD2 database.

Table 5. Positions-matrix P[*i,j*] ⊗ O[*i,j*]

	P+	P-	N	NP
P+	4	3	1	2
P-	8	7	6	5
N	12	11	9	10
NP	16	15	13	14

Array of positions in matrix P[i,j] ⊗ O[i,j] corresponding to the sequence QIINNPITCMTNGAICWGPCPTAFR QIGNCGHFKVRCCKIR.

Table 6. Positions-matrix P[*i,j*]

	P+	P-	N	NP
P+	4	1	3	2
P-	5	8	7	6
N	12	11	9	10
NP	16	15	13	14

Array of positions in matrix **P**[*i,j*] corresponding to 51 identified **SCAAP**.

Table 8. Number of matches by pathogenic action.

Number of hits	Fungi	Viruses	Mammalian cells	Bacteria	Cancer cells	Insects	Parasites	Sperms	SCAAP	%
Unique action	18	0	1	237	2	0	1	0	44	
	77	22	10	743	16	2	9	0	51	85
Multiple action	218	19	71	491	41	7	16	3	0	
	638	122	205	1489	121	20	40	9	0	

Number of hits: Number of matches of the polarity profile method within two groups of peptides from the APD2 (Wang *et al.*, 2009), with unique and multiple pathogenic action against: bacteria, viruses, fungi, parasites, insects, carcinogenic cells, mammalian cells, sperms, and SCAAP (Polanco & Samaniego, 2009, Polanco *et al.*, 2012). **Unique action**: Peptides exerting pathogenic action against only one group (Section 2.3). **Multiple action**: Peptides exerting pathogenic action against two or more groups (Section 2.3). (**%**): Percentage of number of hits/total of peptides.

Table 7. Test of Polarity Profile Method.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
O[*i,j*] vector of study.	4	1	3	2	5	8	7	6	12	11	9	10	16	15	13	14
P[*i,j*] + O[*i,j*] vector of study.	4	3	1	2	8	7	6	5	12	11	9	10	16	15	13	14
Results	✓	x	x	✓	x	x	x	x	✓	✓	✓	✓	✓	✓	✓	✓

Test of peptide **QIINNPITCMTNGAICWGPCPTAFRQIGNCGHFKVRCCKIR** by the polarity profile method (Table 9, entry #35). (✓): The polar interaction is present in the position. (x): The polar interaction is not present in the position.

Table 9. Polarity profile matches by linear sequence.

No	Peptide	Sequence	#1	#2	Reference
1	Clavanin D (Sea squirt, tunicate, invertebrates, animals)	AFKLLGRIIHHVGNFVYGFSHVF			Lee *et al.*, 1997
2	Palustrin-1b (frog, amphibians, animals; XXU)	ALFSILRGLKKLGNMGQAFVNCKIYKKC			Basir *et al.*, 2000
3	Palustrin-1d (frog, amphibians, animals; XXU)	ALSILKGLEKLAKMGIALTNCKATKKC			Basir *et al.*, 2000
4	Palustrin-1c (frog, amphibians, animals; XXU)	ALSILRGLEKLAKMGIALTNCKATKKC			Basir *et al.*, 2000
5	Brevinin-1PRc (frog, amphibians, animals; XXU)	FFPMLAGVAARVVPKVICLITKKC			Conlon *et al.*, 2011
6	Brevinin-1Be (frog, amphibians, animals; XXU)	FLPAIVGAAAKFLPKIFCVISKKC			Goraya *et al.*, 2000
7	Brevinin-1HSa (frog, amphibians, animals; XXU)	FLPAVLRVAAKIVPTVFCAISKKC			Conlon *et al.*, 2008
8	Brevinin-1Ba (frog, amphibians, animals; XXU)	FLPFIAGMAAKFLPKIFCAISKKC			Goraya *et al.*, 2000
9	Brevinin-1Bc (frog, amphibians, animals; XXU)	FLPFIAGVAAKFLPKIFCAISKKC			Goraya *et al.*, 2000
10	RANATUERIN 4 (Ranatuerin-4, frog, amphibians, animals; XXU)	FLPFIARLAAKVFPSIICSVTKKC			Goraya *et al.*, 1998
11	Phylloseptin-H11 (PLS-H11, Phylloseptin-13, PS-13; frog, amphibians, animals; XXA)	FLSLIPHAINAVGVHAKHF			Thompson *et al.*, 2007
12	Phylloseptin-H5 (Phylloseptin-7, PLS-H5, PS-7, XXA, frog, amphibians, animals)	FLSLIPHAINAVSAIAKHF			Conceicao *et al.*, 2006
13	Phylloseptin-H2 (PLS-H2, Phylloseptin-2, PS-2) (XXA, frog, amphibians, animals)	FLSLIPHAINAVSTLVHHF		N	Leite *et al.*, 2005
14	Phylloseptin-B1 (PLS-B1, PBN1; frog, amphibians, animals; XXA)	FLSLIPHIVSGVAALAKHL			Vanhoye *et al.*, 2004
15	Papilosin (tunicate, ascidian, invertebrates, sea animals)	GFWKKVGSAAWGGVKAAAKGAAVGGLNAL-AKHIQ			Galinier *et al.*, 2009
16	SMAP-34 (sheep myeloid antmicrobial peptide-34; OaMAP34, ovine cathelicidin, sheep, ruminant, animals)	GLFGRLRDSLQRGGGQKILEKAERIWCKIKDIFR			Anderson RC, 2005
17	Caerin 1.17 (frog, amphibians, animals; XXA)	GLFSVLGSVAKHLLPHVAPIIAEKL			Maclean *et al.*, 2006
18	Caerin 1.18 (frog, amphibians, animals; XXA)	GLFSVLGSVAKHLLPHVVPVIAEKL			Maclean *et al.*, 2006
19	Fallaxidin 3.2 (XXA, frog, amphibians, animals)	GLLDFAKHVIGIASKL			Jackway *et al.*, 2008
20	Fallaxidin 3.1 (XXA, frog, amphibians, animals)	GLLDLAKHVIGIASKL			Jackway *et al.*, 2008
21	Dahlein 5.2 (frog, amphibians, animals)	GLLGSIGNAIGAFIANKLKPK			Wegener *et al.*, 2001
22	Caerin 1.2 (XXA, frog, amphibians, animals)	GLLGVLGSVAKHVLPHVVPVIAEHL		N	Stone *et al.*, 1993
23	Caerin 1.4 (XXA, frog, amphibians, animals)	GLLSSLSSVAKHVLPHVVPVIAEHL		N	Stone *et al.*, 1993
24	Palustrin-2SIb (frog, amphibians, animals; XXU)	GLWNSIKIAGKKLFVNVLDKIRCKVAGGCKTSP-DVE			Iwakoshi-Ukena *et al.*, 2011
25	XPF (the xenopsin precursor fragment, African clawed frog, amphibians, animals)	GWASKIGQTLGKIAKVGLKELIQPK			Moore *et al.*, 1995
26	Pleurocidin (fish, animals)	GWGSFFKKAAHVGKHVGKAALTHYL			Cole *et al.*, 1997
27	Cecropin (insects, invertebrates, animals)	GWLKKIGKKIERVGQNTRDATVKGLEVAQQAAN-VAATVR			Boulanger *et al.*, 2002
28	Pm_mastoparan PMM (insects, invertebrates, animals; XXA; Derivatives)	INWKKIASIGKEVLKAL			Cerovský *et al.*, 2008
29	Hinnavin II (Hin II, insects, invertebrates, animals; JJsn)	KWKIFKKIEHMGQNIRDGLIKAGPAVQVVGQA-ATIYKG			Yoe *et al.*, 2006
30	Ostrich AvBD2 (Ostrich avian beta defensin 2, Ostricacin-1, OSP-1, birds, animals; BBL)	LFCRKGTCHFGGCPAHLVKVGSCFGFRACCKW-PWDV		N	Yu *et al.*, 2001
31	Clavanin D (Sea squirt, tunicate, invertebrates, animals)	LFKLLGKIIHHVGNFVHGFSHVF			Zhao *et al.*, 1997
32	Enterocin Q (EntQ, class 2d bacteriocins; leaderless, i.e. no signal peptide, bacteria)	MNFLKNGIAKWMTGAELQAYKKKYGCLP-WEKISC			Cintas *et al.*, 2000
33	Temporin-1Lb (Temporin 1Lb, frog, amphibians, animals)	NFLGTLINLAKKIM			Goraya *et al.*, 2000
34	Bovine Beta-defensin 6 (bBD-6,cow, ruminant, animals)	QGVRNHVTCRIYGGFCVPIRCPGRTRQIGTCF-GRPVKCCRRW			Selsted *et al.*, 1993

35	mBD-4 (mBD4, mouse beta defensin 4, or Defb4, animals; 3S=S)	QIINNPITCMTNGAICWGPCPTAFRQIGNCGH-FKVRCCKIR	N	Jia *et al.*, 2000
36	ChBac5 (Pro-rich; Arg-rich, Goat cathelicidin, ruminant, animals)	RFRPPIRRPPIRPPFNPPFRPPVRPPFRPPFRP-PFRPPIGPFP	N	Shamova *et al.*, 1999
37	Cyclic dodecapeptide (OaDode, ovine cathelici-din, sheep, ruminant, animals)	RICRIIFLRVCR		Bagella *et al.*, 1995
38	Bactenecin (Cyclic dodecapeptide, bovine cathelicidin, cow, cattle, ruminant, animals; BBMm; JJsn; Derivatives: Bac2A)	RLCRIVVIRVCR		Romeo *et al.*, 1998
39	RL-37 (RL37, cathelicidin, Old World monkey, primates, animals)	RLGNFFRKVKEKIGGGLKKVGQKIKDFLGNLVPR-TAS		Zelezetsky *et al.*, 2006
40	BACTENECIN 7 (bac-7, bac 7; bac7; Pro-rich; cow cathelicidin, ruminant, animals; BBL, Se-qAR, BBPP)	RRIRPRPPRLPRPRPRPLPFPRPGPRPIPRPLPFPRP-GPRPIPRPLPFPRPGPRPIPRPL	N	Storici *et al.*, 1996
41	Bac4 (Pro-rich, Arg-rich; cow cathelicidin, rumi-nant, animals)	RRLHPQHQRFPRERPWPKPLSLPLPRPGPRPW-PKPL	N	Anderson RC & Yu PL, 2003
42	Hyphancin IIIE (insects, invertebrates, animals)	RWKFFKKIERVGQNVRDGLIKAGPAIQVLGAAKAL	N	Park *et al.*, 1995
43	Cecropin B (insects, invertebrates, animals)	RWKIFKKIEKMGRNIRDGIVKAGPAIEVLGSAKAI	N	Morishima *et al.*, 1990
44	Hyphancin IIID (insects, invertebrates, animals)	RWKIFKKIERVGQNVRDGIIKAGPAIQVLGTAKAL	N	Park *et al.*, 1995
45	Hyphancin IIIG (insects, invertebrates, animals)	RWKVFKKIEKVGRHIRDGVIKAGPAITVVGQATAL	N	Park *et al.*, 1995
46	Hyphancin IIIF (insects, invertebrates, animals)	RWKVFKKIEKVGRNIRDGVIKAGPAIAVVGQAKAL	N	Park *et al.*, 1995
47	Phylloseptin-H4 (Phylloseptin-6, PLS-H4, PS-6, XXA, frog, amphibians, animals)	SLIPHAINAVSAIAKHF		Leite *et al.*, 2005
48	Pep5 (Lantibiotic, type 1, class 1 bacteriocin, Gram-positive bacteria; XXT3; XXW3)	TAGPAIRASVKQCQKTLKATRLFTVSCKGKNGCK		Kaletta *et al.*, 1989
49	Clavanin C (Sea squirt, tunicate, invertebrates, animals)	VFHLLGKIIHHVGNFVYGFSHVF		Lee *et al.*, 1997
50	Andropin (insects, invertebrates, animals)	VFIDILDKVENAIHNAAQVGIGFAKPFEKLINPK	N	Samakovlis *et al.*, 1991
51	Clavanin A (urochordates, sea squirts, and sea pork, tunicate, invertebrates, animals)	VFQFLGKIIHHVGNFVHGFSHVF		Lee *et al.*, 1997

SCAAP subjects identified by polarity index method in APD2 database (Wang *et al.*, 2009) where peptides exert action only against SCAAP. **#1**: (**N**) peptides not accepted by polarity profile method (algorithm suggested here). **#2**: (**N**) peptides not accepted by polarity index method (algorithm published to detect only SCAAP Polanco *et al.*, 2012).

Catastrophic bifurcations points

Catastrophic bifurcation points are points where abrupt changes in the behavior of a function occur. These points are associated with the positions in the matrix where the maximum/minimum frequencies observed are identified; for the sake of convenience and clarity they are marked in red/blue, respectively.

RESULTS

The Polarity profile method showed an efficiency of more than 85% to detect SCAAP (44/85) in the entire APD2 database sub-classifications (Table 8). It is equally efficient in not identifying other groups such as fungi 18/77 = 23%, or bacteria 237/743 = 31%.

The method excluded 7 out of 51 peptides from SCAAP (Table 9, column #1 with N-letter) while the formerly reported algorithm (Polanco *et al.*, 2012) excluded 6 out of 51 peptides from SCAAP (Table 9, column #2 with N-letter). There are no coincidences between both excluded sets.

The maximum value points (catastrophic bifurcations points) of the SCAAP group are located in the positions 4, 5, 12, and 16 from the polarity matrix (Table 10, first column in red color), its corresponding points for the bacteria group in the positions 4, 8, 12, and 16 (Ta-ble 11, first column in red color). The match is almost 3 out of 4 total.

The minimum value points (catastrophic bifurcations points) of the SCAAP group are located in the positions 2, 6, 10, and 14 from the polarity matrix (Table 10, first column in blue color), its corresponding points for the bacteria group in the positions 2, 6, 10, and 14 (Table 11, first column in blue color). The coincidence is total between the two groups.

DISCUSSION

For decades peptides were classified according to their toxic action. However, these features were related to the space where the peptides interact with the structural membrane of the object. As discovered by the experiment, most peptides exert, in some degree, action against multiple pathogens. One could speculate that the nature attempts to avoid a large differentiation in the linear sequence of the peptide, with regard to the toxicity against a particular group of pathogen. In other words, if a peptide requires only small changes in the sequence to face another pathogen group, the amount of energy required to do so will be much smaller than in case of large changes. The above mentioned considerations led us to believe that the detection of selective antibacterial

Table 10. Catastrophic bifurcations points of the SCAAP group

	P+	P-	N	NP
P+	4	1	3	2
P-	5	8	7	6
N	12	11	9	10
NP	16	15	13	14

Set of the **SCAAP** group. The positions in red represent maximum values and positions in blue represent the minimum points (Section 2.4).

Table 11. Catastrophic bifurcations points of the bacteria group

	P+	P-	N	NP
P+	4	3	1	2
P-	8	7	5	6
N	12	11	9	10
NP	16	15	13	14

Set of the bacteria group. The positions in red represent maximum values and positions in blue represent the minimum points (Section 2.4).

peptides and their prediction is rather related to the general features of the peptides. Hence, the most effective algorithms are those evaluating fundamental characteristics of all peptides that search only for small differences.

The design of bioinformatic algorithms related to peptide detection is basically of two types. The first depends on a system of nonlinear differential equations (Janes & Lauffenburger, 2006) that characterizes the peptide properties with exponentially growing complexity. The other type allows for the inclusion of multiple peptide characteristics without affecting its complexity. Here, the efficiency depends greatly on a good peptide training set selection. The Polarity Profile method falls into the latter type that is characterized by effectively excluding multiple action peptides with a margin of error lower than 30%. Its efficiency to identify SCAAP subjects is higher than 85%. It only measures the polarity of the peptides and this information allows for efficient classification of the pathogenic action.

The catastrophic bifurcation points in the SCAAP and bacteria groups show almost total coincidence in the points 7 of 8 (Tables 10 and 11). In the two sets these points are clearly associated to the last elements of the rows of the polarity matrix. The almost total coincidence of the catastrophic bifurcation point locations between the SCAAP and the bacterial group may indicate that there is an algebraic structure associated to the polarity matrix. This would reinforce our assumption that the catastrophic bifurcation points stand for regions, in which the protein functionality definition takes place. Depending on its validation for other peptide sets exerting different pathogenic actions, the dynamic bifurcation analysis could become an important mathematical contribution to the field of proteomics.

Both the Polarity Index and the Polarity Profile methods are almost equally efficient. However, the robustness of the polarity profile method is higher due to the presentation of a more complete polarity profile represented by 16 elements. In order to improve this method, we are working on the subclassification of Gram+/- ONLY, Gram+ ONLY, and Gram- ONLY bacteria in the APD2 database, as well as in the virus group exerting action against HIV, fungi with action against protists and mammalian cells (e.g. hemolytic or cytotoxic effects on those with chemotaxis property). In addition to this, we are also determining the toxicity of all antimicrobial peptides of this database to obtain a new classification by toxicity. This work of subclassification by toxicity appears to us useful as a contribution to other researchers designing prediction algorithms.

Finally, we consider the Polarity Profile method as a simple mathematical and computational algorithm that does not demand heavy computational resources such as processing memory or speed. Therefore, it can be used to explore peptide regions. These peptide regions can be worked out by evaluating massively all possible peptide combinations with the same length (Polanco & Samaniego, 2009; Polanco et al., 2012).

CONCLUSIONS

The computational mathematical method called the Polarity Profile method is a robust and fast method that can be used as a first filter tool in the detection of selective antibacterial peptides (SCAAP). It can be also used in high-performance computing platforms for search patterns in peptide regions.

AVAILABILITY

The test files are given as "Supplementary Material"

Appendix

Source Program for the Detection of selective antibacterial peptides' group by the Polarity Profile Method see web page: www.actabp.pl

Conflict of Interests

We declare that we do not have any financial and personal interest with other people or organizations that could inappropriately influence (bias) our work.

Author Contributions

Theoretical conception and design: CP and JLS. Computational performance: CP. Mathematical analysis: CP and JLS. Experimental analysis: JACG. Chemical analysis: TB. Data analysis: CP, TB and JLS. Results discussion: CP, JLS, TB, and JACG.

REFERENCES

Anderson RC (2005) *Antimicrobial peptides isolate from ovine blood neutrophils, thesis,* Massey University, Palmerston, North New Zealand.

Anderson RC, Yu PL (2003) Isolation and characterisation of proline/arginine-rich cathelicidin peptides from ovine neutrophils. *Biochem Biophys Res Commun* **312**: 1139–1146.

Australian Naturopathic Network (http://www.ann.com.au/medsci/amino.htm) accesed in December, 2012.

Bagella L, Scocchi M, Zanetti M (1995) cDNA sequences of three sheep myeloid cathelicidins. *FEBS Lett* **376**: 225–228.

Barie PS (2012) Multidrug-resistant organisms and antibiotic management. *Surg Clin North Am* **92**: 345–391.

Basir YJ, Knoop FC, Dulka J, Conlon JM (2000) Multiple antimicrobial peptides and peptides related to bradykinin and neuromedin N

isolated from skin secretions of the pickerel frog, *Rana palustris*. *Biochim Biophys Acta* **1543**: 95–105.

Blueggel M, Chamrad D, Meyer HE (2004) Bioinformatics in proteomics. *Curr Pharm Biotechnol* **5**: 79–88.

Boulanger N, Brun R, Ehret-Sabatier L, Kunz C, Bulet P (2002) Immunopeptides in the defense reactions of *Glossina morsitans* to bacterial and *Trypanosoma brucei* infections. *Insect Biochem Mol Biol* **32**: 369–375.

Cerovský V, Slaninová J, Fucík V, Hulacová H, Borovicková L, Jezek R, Bednárová L (2008) New potent antimicrobial peptides from the venom of Polistinae wasps and their analogs. *Peptides* **29**: 992–1003.

Cintas LM, Casaus P, Herranz C, Hâvarstein LS, Holo H, Hernández PE, Nes IF (2000) Biochemical and genetic evidence that Enterococcus faecium L50 produces enterocins L50A and L50B, the sec-dependent enterocin P, and a novel bacteriocin secreted without an N-terminal extension termed enterocin Q. *J Bacteriol* **182**: 6806–6814.

Cole AM, Weis P, Diamond G (1997) Isolation and characterization of pleurocidin, an antimicrobial peptide in the skin secretions of winter flounder. *J Biol Chem* **272**: 12008–12013.

Conceição K, Konno K, Richardson M, Antoniazzi MM, Jared C, Daffre S, Camargo AC, Pimenta DC (2006) Isolation and biochemical characterization of peptides presenting antimicrobial activity from the skin of *Phyllomedusa hypochondrialis Peptides* **27**: 3092–309.

Conlon JM, Kolodziejek J, Nowotny N, Leprince J, Vaudry H, Coquet L, Jouenne T, King JD (2008) Characterization of antimicrobial peptides from the skin secretions of the *Malaysian frogs, Odorrana hosii* and *Hylarana picturata* (Anura:Ranidae). *Toxicon* **52**: 465–473.

Conlon JM, Mechkarska M, Ahmed E, Coquet L, Jouenne T, Leprince J, Vaudry H, Hayes MP, Padgett-Flohr G (2011) Host defense peptides in skin secretions of the Oregon spotted frog *Rana pretiosa*: implications for species resistance to chytridiomycosis. *Dev Comp Immunol* **35**: 644–649.

Davies J, Shaffer B, Littlewood W (1979) *Elementary Biochemistry — An Introduction to the Chemistry of Living Cells*. Prentice Hall, Upper Saddle River, NJ, USA.

del Rio G, Castro-Obregon S, Rao R, Ellerby HM, Bredesen DE (2001) APAP, a sequence-pattern recognition approach identifies substance P as a potential apoptotic peptide. *FEBS Lett* **494**: 213–219.

Galinier R, Roger E, Sautiere PE, Aumelas A, Banaigs B, Mitta G (2009) Halocyntin and papillosin, two new antimicrobial peptides isolated from hemocytes of the solitary tunicate, *Halocynthia papillosa*. *J Pept Sci* **15**: 48–55.

Gonzalez-Diaz H (2012) Editorial: QSAR/QSPR models as enabling technologies for drug & targets discovery in: medicinal chemistry, microbiology-parasitology, neurosciences, bioinformatics, proteomics and other biomedical sciences. *Curr Top Med Chem* **12**: 799–801.

Goraya J, Knoop FC, Conlon JM (1998) Ranatuerins: antimicrobial peptides isolated from the skin of the American bullfrog, *Rana catesbeiana*. *Biochem Biophys Res Commun* **250**: 589–592.

Goraya J, Wang Y, Li Z, O'Flaherty M, Knoop FC, Platz JE, Conlon JM (2000) Peptides with antimicrobial activity from four different families isolated from the skins of the North American frogs *Rana luteiventris*, *Rana berlandieri* and *Rana pipiens*. *Eur J Biochem* **267**: 894–900.

Han M, Dai J, Zhang Y, Lin Q, Jiang M, Xu X, Liu Q, Jia J (2012) Support vector machines coupled with proteomics approaches for detecting biomarkers predicting chemotherapy resistance in small cell lung cancer. *Oncol Rep* **28**: 2233–2238.

Iwakoshi-Ukena E, Okada G, Okimoto A, Fujii T, Sumida M, Ukena K (2011) Identification and structure-activity relationship of an antimicrobial peptide of the palustrin-2 family isolated from the skin of the endangered frog *Odorrana ishikawae*. *Peptides* **32**: 2052–2057.

Jackway RJ, Bowie JH, Bilusich D, Musgrave IF, Surinya-Johnson KH, Tyler MJ, Eichinger PC (2008) The fallaxidin peptides from the skin secretion of the Eastern Dwarf Tree Frog *Litoria fallax*. Sequence determination by positive and negative ion electrospray mass spectrometry: antimicrobial activity and cDNA cloning of the fallaxidins. *Rapid Commun Mass Spectrom* **22**: 3207–3216.

Janes KA, Lauffenburger DA (2006) A biological approach to computational models of proteomic networks. *Curr Opin Chem Biol* **10**: 73–80.

Jia HP, Wowk SA, Schutte BC, Lee SK, Vivado A, Tack BF, Bevins CL, McCray PB Jr (2000) A novel murine beta-defensin expressed in tongue, esophagus, and trachea. *J Biol Chem* **275**: 33314–33320.

Lee IH, Zhao C, Cho Y, Harwig SS, Cooper EL, Lehrer RI (1997) Clavanins, α-helical antimicrobial peptides from tunicate hemocytes *FEBS Lett* **400**: 158–162.

Leite JR, Silva LP, Rodrigues MI, Prates MV, Brand GD, Lacava BM, Azevedo RB, Bocca AL, Albuquerque S, Bloch C Jr (2005) Phylloseptins: a novel class of anti-bacterial and anti-protozoan peptides from the Phyllomedusa genus. *Peptides* **26**: 565–573.

Li J, Halgamuge SK, Tang SL (2008) Genome classification by gene distribution: an overlapping subspace clustering approach. *BMC Evol Biol* **8**: 116.

Maclean MJ, Brinkworth CS, Bilusich D, Bowie JH, Doyle JR, Llewellyn LE, Tyler MJ (2006) New caerin antibiotic peptides from the skin secretion of the Dainty Green Tree Frog *Litoria gracilenta*. Identification using positive and negative ion electrospray mass spectrometry. *Toxicon* **47**: 664–675.

Moore KS, Bevins CL, Brasseur MM, Tomassini N, Turner K, Eck H, Zasloff M (1991) Antimicrobial peptides in the stomach of *Xenopus laevis*. *J Biol Chem* **266**: 19851–19857.

Morishima I, Suginaka S, Ueno T, Hirano H (1990) Isolation and structure of cecropins, inducible antibacterial peptides, from the silkworm, *Bombyx mori*. *Comp Biochem Physiol B* **95**: 551–554.

Perez-Riverol Y, Vera R, Mazola Y, Musacchio A (2012) A parallel systematic-Monte Carlo algorithm for exploring conformational space. *Curr Top Med Chem* **12**: 1790–1796.

Polanco C, Samaniego JL (2009) Detection of selective cationic amphipatic antibacterial peptides by Hidden Markov models. *Acta Biochim Pol* **56**: 167–176.

Polanco C, Samaniego JL, Buhse T, Mosqueira FG, Negron-Mendoza A, Ramos-Bernal S, Castanon-Gonzalez JA (2012) Characterization of selective antibacterial peptides by polarity index. *Int J Pept* ?.

Romeo D, Skerlavaj B, Bolognesi M, Gennaro R (1988) Structure and bactericidal activity of an antibiotic dodecapeptide purified from bovine neutrophils. *J Biol Chem* **263**: 9573–9575.

Selsted ME, Tang YQ, Morris WL, McGuire PA, Novotny MJ, Smith W, Henschen AH, Cullor JS (1993) Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J Biol Chem* **268**: 6641–6648.

Shamova O, Brogden KA, Zhao C, Nguyen T, Kokryakov VN, Lehrer RI (1999) Purification and properties of proline-rich antimicrobial peptides from sheep and goat leukocytes. *Infect Immun* **67**: 4106–4111.

Silverman BD (2005) Underlying hydrophobic sequence periodicity of protein tertiary structure. *J Biomol Struct Dyn* **22**: 411–423.

Stone DJ, Waugh RJ, Bowie JH, Wallace JC, Tyler MJ (1993) Peptides from Australian frogs. Structures of the caeridins from *Litoria caerulea*. *J Chem Soc, Perkin Transactions* **1**: 573–576.

Storici P, Tossi A, Lenarcic B, Romeo D (1996) Purification and structural characterization of bovine cathelicidins, precursors of antimicrobial peptides. *Eur J Biochem* **238**: 769–776.

Thompson AH, Bjourson AJ, Orr DF, Shaw C, McClean S (2007) A combined mass spectrometric and cDNA sequencing approach to the isolation and characterization of novel antimicrobial peptides from the skin secretions of *Phyllomedusa hypochondrialis azurea*. *Peptides* **28**: 1331–1343.

Vanhoye D, Bruston F, El Amri S, Ladram A, Amiche M, Nicolas P (2004) Membrane association, electrostatic sequestration, and cytotoxicity of Gly-Leu-rich peptide orthologs with differing functions. *Biochemistry* **43**: 8391–8409.

Wabnitz PA, Bowie JH, Tyler MJ, Wallace JC, Smith BP (2000) Differences in the skin peptides of the male and female Australian tree frog *Litoria splendida*. The discovery of the aquatic male sex pheromone splendipherin, together with phe8 caerulein and a new antibiotic peptide caerin 1.10. *Eur J Biochem* **267**: 269–275.

Wang G, Li X, Wang Z (2009) APD2: the updated antimicrobial peptide database and its application in peptide design *Nucleic Acids Res* **37**: D933–D937.

Wegener KL, Brinkworth CS, Bowie JH, Wallace JC, Tyler MJ (2001) Bioactive dahlein peptides from the skin secretions of the Australian aquatic frog *Litoria dahlii*: sequence determination by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* **15**: 1726–1734.

Yoe SM, Kang CS, Han SS, Bang IS (2006) Characterization and cDNA cloning of hinnavin II, a cecropin family antibacterial peptide from the cabbage butterfly, *Artogeia rapae*. *Comp Biochem Physiol B Biochem Mol Biol* **144**: 199–205.

Yu PL, Choudhury SD, Ahrens K (2001) Purification and characterization of the antimicrobial peptide, ostricacin. *Biotechnol Lett* **23**: 207–210.

Zelezetsky I, Pontillo A, Puzzi L, Antcheva N, Segat L, Pacor S, Crovella S, Tossi A (2006) Evolution of the primate cathelicidin. Correlation between structural variations and antimicrobial activity. *J Biol Chem* **281**: 19861–19871.

Zhao C, Liaw L, Lee IH, Lehrer RI (1997) cDNA cloning of Clavanins: antimicrobial peptides of tunicate hemocytes. *FEBS Lett* **410**: 490–492.

Zhao N, Pang B, Shyu CR, Korkin D (2011) Feature-based classification of native and non-native protein-protein interactions: Comparing supervised and semi-supervised learning approaches. *Proteomics* **11**: 4321–4330.