ACTA ABP BIOCHIMICA POLONICA

Vol. 63, No 3/2016
483–491
http://dx.doi.org/10.18388/abp.2015_993

Regular paper

Identification of antimicrobial peptides by using eigenvectors

Carlos Polanco✉

Department of Mathematics, Faculty of Sciences, Universidad Nacional Autónoma de México, México

Antibacterial peptides are subject to broad research due to their potential application and the benefit they can provide for a wide range of diseases. In this work, a mathematical-computational method, called the Polarity Vector Method, is introduced that has a high discriminative level (>70%) to identify peptides associated with Gram (–) bacteria, Gram (+) bacteria, cancer cells, fungi, insects, mammalian cells, parasites, and viruses, taken from the Antimicrobial Peptides Database. This supervised method uses only eigenvectors from the incident polar matrix of the group studied. It was verified with a comparative study with another extensively verified method developed previously by our team, the Polarity Index Method. The number of positive hits of both methods was up to 98% in all the tests conducted.

Key words: antimicrobial peptides; supervised methods; structural bioinformatics

Received: 13 February, 2015; revised: 10 February, 2016; accepted: 13 June, 2016; available on-line: 23 June, 2016

INTRODUCTION

The manufacture of pharmaceutical drugs (Blundell *et al*., 2006; Ekins, 2004; Ekins *et al*., 2007; Kantardjieff & Rupp, 2004; Readhead & Dudley, 2013) from proteins made or identified by "bioinformatics methods" is strategically important mainly for two reasons: the experimental location of peptides in living organisms is less frequent every time, and the costs involved in their synthesis and trial and error assays (Adams & Brantner, 2006; Bahar & Ren, 2013; Breda *et al*., 2006; Dudley *et al*., 2011) is constantly increasing. These two factors have given an impulse to design a new generation of mathematical-computational algorithms, oriented to measure the characteristic physico-chemical profiles (Gill *et al*., 2007; Liu *et al*., 2012; Vilar *et al*., 2008) of different groups of proteins and thus "computationally" build peptides by design. Among the different bioinformatics methods, the Polarity Index Method (PIM) (Polanco & Samaniego, 2009; Polanco *et al*., 2012; 2013; 2013a; 2014; 2014a; 2014b; 2014c; 2014d) stands out for its high level of efficiency to identify the major action of proteins with antimicrobial action (>75% in a double blind test). The metric of this method is based on a polar matrix, which is built by counting the polar incidents of the amino acids forming the linear sequence of a protein. Although this polar matrix is the core of the PIM metrics, an equally effective matrix has been identified from the eigenspace of the polar matrix of the studied group (Poole, 2011; Sahai & Bist, 2002). In this work, method called "Polarity Vector Method" (PVM) is presented, which uses as a metric an eigenvector matrix. With the substitution of the polar matrix by an eigenvector matrix, this variant shows that the mathematical substitute is equally efficient. Another important aspect to highlight is that both methods, PIM and PVM, use in the metric a single physico-chemical property, i.e. polarity (Pauling, 1955). The verification of this method was done by comparing the results found in PIM and PVM, taking the main antimicrobial peptides from the APD2 Database, as accessed in December, 2012 (Wang & Wang, 2009). This public database undergoes constant maintenance and it includes the notes of the publications that support the information; these were important factors considered in the selection of the group of peptides used to train the methods mentioned.

MATERIALS AND METHODS

The eight main groups of antibacterial peptides from the APD2 Database (Dec 2012) (Wang & Wang, 2009) were tested in the automated versions of the mathematical-computational Polarity Index Method (PIM) and Polarity Vector Method (PVM). The first has already been extensively tested with different groups of peptides and proteins (Polanco & Samaniego, 2009; Polanco *et al*., 2012; 2013; 2013a; 2014; 2014a; 2014b; 2014c; 2014d), the second will be introduced in this paper.

Polarity Index Method Metrics

In summary, the metric of the supervised method called Polarity Index method (PIM) (Polanco & Samaniego, 2009; Polanco *et al*., 2012; 2013; 2013a; 2014; 2014a; 2014b; 2014c; 2014d) required a data training set formed by the amino acid sequences experimentally identified by the pattern of interest. This data training set with the protein sequences formed by the amino acids, was translated to its numerical equivalence from the rule: {P+, P-, N, NP}: P-={D, E}, P+={H, K, R}, NP={A, F, I, L, M, P, V, W}, and N={C, G, N, Q, S, T, Y} (Timberlake, 1992).

Afterward, the polar incidents obtained from reading the sequences from left to right, number by number, were registered in a polar incident matrix called the polar matrix, where (row, column)=(number equivalent A, number equivalent B). Once the incidents were recorded in the matrix, this was normalized (Table 1). Note that this matrix informs as to which polar interaction is more/less frequent from the 16 possibilities available. Then, the PIM compared the polar interaction between the polar matrix of the training set and the sequence of

✉e-mail: polanco@unam.mx

Abbreviations: PIM, Polarity Index Method; PVM, Polarity Vector Method

the group being studied, to determine whether the sequence studied was similar to that particular group.

Example

In order to evaluate if the peptide has the profile of the training set, let us take the following protein: MSLLTEVETYVLSIIPSGPLKAEIAQRLEDVFAGKNTDLEVLMEWLKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNGDPNNMDKAVKLYRKLKREITFHGAKEISLSYSAGALASCMGLIYNRMGAVTTEVAFGLVCATCEQIADSQHRSHRQMVTTTNPLIRHENRMVLASTTAKAMEQMAGSSEQAAEAMEVASQARQMVQAMRTIGTHPSSSAGLKNDLLENLQAYQKRMGVQMQRFK (taken from the Example section, Polanco *et al*., 2014d) and follow these steps:

The numerical equivalent is obtained according to the rule mentioned above

1. 434432423344344433441424431422444313324244424413144434431344344434344321343111443344333243342144143114112434134124343334344433434433143443324443443433234423311311344333344411231444433341442344333234424442443341344344134331433434313244234343311434343141.
2. This sequence (step 1) is read from left to right, the first element is (4,3), the second element is (3,4) (note this element appears when one position is run to the right). The element (4,3) is recorded in the polar matrix in row 4, column 3; the second element (3,4) is recorded in the same matrix in row 3, column 4, and so on until all incidents are recorded; this matrix will be called A [i, j].
3. The same procedure is conducted for the training set representing the characteristic sought, gathering all incidences in a matrix; this matrix will be called P [i, j].
4. Both matrices A [i, j] and P [i, j] are weighted.
5. Matrix **C**[i,j] is created; **C**[i,j]=**A**[i,j] + **P**[i,j].
6. Now the C [i, j] matrix has as elements the normalized relative frequencies of the sequence studied, and the P [i, j] matrix has as elements the normalized relative frequencies from the training set. In order to compare them, two vectors are built for each one of them, sorting their elements in an ascending order, and instead of using their relative frequencies, the position they have in each vector is used.
7. Step 6 is also applied to the P [i, j] matrix comparing both vectors. The greater the number of hits, the greater the similarity between the two sequences.

In this step a percentage of similarity is determined, if the peptide or protein has an equal or a greater percentage, the sequence is accepted.

Polar Vector Method metrics

The metrics of the supervised method called the Polarity Vector Method (PVM) had the comparison structure of PIM but replaced the polar matrix by the eigenvector matrix (Table 3). This matrix was built as follows:

1. The four eigenvectors (Hait, 2002; Poole, 2011; Sahai & Bist, 2002) of each polar matrix group were calculated (Table 1). These four eigenvectors (v_1, v_2, v_3, v_4) were then integrated into a 4×4 matrix, where v_1 was the first column, v_2 the second, v_3 the third and v_4 the last column (Table 2). The 16 elements of this matrix were placed in an ascending order by the following rule: Since $z_1=a + bi$, and $z_2=c + di$, where a, b, c and d are real numbers, and i is an imaginary identity. $z_1=z_2 \Leftrightarrow a=c$ and $b=d$; $z_1 \geq z_2 \Leftrightarrow$ *if* $a \geq c$ and $b \geq d$; $z_1<z_2 \Leftrightarrow$ *if* $a<c$ and $b \leq d$.
2. These 16 elements (step1) were grouped to form triplets in this way: (*order*, *a+bi*, *old position*); where "*order*" was the position occupied by number "*a+bi*", (*order*=1 for the smallest *a+bi* and 16 for the largest *a+bi);* and "*old position*" was the place the a+bi element had in the matrix (Table 2) (step 1). For instance, taking the matrix of the fungi group (Table 2), this procedure shows the following triplets: (1, –0.905484+0.000000i,8),(2, –0.848264+0.000000i,3),…,(16, 0.660015+0.000000i, 14).
3. Then a straight line for the group studied was analytically built, taking its maximun frequency (y_{max}) and its minimum frequency (y_{min}) from the polar matrix (Table 1), where $f(x)=m(x-x_{max}) + y_{max}$; $m=(y_{max}-y_{min})/(x_{max}-x_{min})$; $x_{max}=16$ and $x_{min}=1$; and thus obtaining 16 ordered pairs (x_i,$f(x_i)$). For instance, taking the fungi group (Table1), the corresponding equation is $f(x)=(0.143232-0.004617)/15)(x-16) + 0.143232$, resulting in 16 pairs with this order: (1, 0.004617), (2, 0.013858),…, (16, 0.143232).
4. If *"order"* (step 2)=x_i (step 3), then the pair generated was (*old position*, $f(x_i)$). These 16 ordered pairs form the eigenvector matrix (Table 3) when the values $f(x_i)$ are allocated to the "*old position"* place in the matrix.

Phase Portrait

The eigenvalues of each of the eight antimicrobial peptide groups from the APD2 database (Dec 2012) (Wang & Wang, 2009) were calculated by the Bluebit Software http://www.bluebit.gr/, accessed Nov 26, 2014 (Hait, 2002), and then were plotted with the GNU Octave http://www.gnu.org/software/octave/doc/interpreter (Eaton *et al*., 2009). The analysis of the portrait phase considered the spatial distribution of the eigenvalues for each group. An eigenvector matrix was calculated for each peptide, as well as its four complex eigenvalues. The real part of each complex eigenvalue was located in the X-axis and its imaginary part in the Y-axis (Figs. 1–8).

APD2 Database

The selected antimicrobial peptide groups from the APD2 Database (Dec 2012) (Wang & Wang, 2009), were verified to make sure if any of their representative sequences were included in another group. Avoiding duplication of peptides provided a more accurate fingerprint of the group studied, minimizing false positives and false negatives, and raising the level of efficiency of the method. It should be noted that this filter reduced the number of the peptides studied to almost 60%. The number of peptides analyzed was 1146 with the following distribution: 131 for Gram (–) bacteria, 260 for Gram (+)bacteria, 54 for cancer cells, 527 for fungi, 7 for insects, 93 for mammalian cells, 20 for parasites, and 54 for viruses. There are other groups of antimicrobial peptides in this database, however, their number and relevance to compare the efficiency of the methods were not meaningful; therefore, they were not included. There are also at least 15 other public databases, some of them are general and others specialized; however, to assess the efficiency of the PVM it was decided to analyze the antimicrobial peptides for the importance they have in the production of new pharmaceutical drugs.

Test Trial

Polarity Index Method (PIM) was calibrated with the polar matrix (Table 1) and Polarity Vector Method

Table 1. Incident polar matrix calculated by the Polarity Index Method (PIM) (see Polarity Index Method Metrics section), for the antimicrobial peptide groups from the APD2 Database (Dec 2012) (Wang & Wang, 2009).

Polar matrix				
Gram (+) bacteria	P+	P-	N	NP
P+	0.022799	0.008608	0.057462	0.043504
P-	0.005932	0.002094	0.015703	0.019193
N	0.060254	0.012911	0.176806	0.160172
NP	0.045597	0.018030	0.161103	0.159591
Gram (-) bacteria	P+	P-	N	NP
P+	0.023045	0.012187	0.056282	0.069798
P-	0.011744	0.004653	0.018170	0.022601
N	0.057833	0.016397	0.144472	0.142699
NP	0.069798	0.025039	0.138710	0.157545
Viruses	P+	P-	N	NP
P+	0.027231	0.007960	0.056975	0.059908
P-	0.005865	0.002933	0.028069	0.020528
N	0.073733	0.025136	0.205279	0.134059
NP	0.046502	0.022204	0.143276	0.117721
Parasites	P+	P-	N	NP
P+	0.033254	0.009501	0.068884	0.049881
P-	0.010689	0.005938	0.019002	0.027316
N	0.054632	0.024941	0.174584	0.137767
NP	0.064133	0.021378	0.130641	0.143705
Insects	P+	P-	N	NP
P+	0.042857	0.014286	0.078571	0.078571
P-	0.007143	0.007143	0.021429	0.025000
N	0.053571	0.014286	0.100000	0.125000
NP	0.107143	0.028571	0.092857	0.178571
Mammalian Cells	P+	P-	N	NP
P+	0.022883	0.010297	0.056064	0.076278
P-	0.007246	0.000763	0.012204	0.023265
N	0.050725	0.011060	0.086575	0.135393
NP	0.087338	0.022121	0.131579	0.230740
Fungi	P+	P-	N	NP
P+	0.035794	0.013622	0.078199	0.061898
P-	0.011684	0.004617	0.023368	0.017555
N	0.075349	0.019607	0.143232	0.134454
NP	0.065261	0.018467	0.131946	0.134910
Cancer cells	P+	P-	N	NP
P+	0.016271	0.010847	0.042712	0.061695
P-	0.012203	0.006102	0.025763	0.006102
N	0.051525	0.021017	0.195932	0.136949
NP	0.054237	0.016271	0.134915	0.170847

(PVM) with the eigenvector matrix (Table 3) for each antimicrobial group in the APD2 Database (Dec 2012) (Wang & Wang, 2009). Afterward, the percentage of efficiency for both methods was determined by comparing the target group with the group studied, and with the other groups extracted (Table 4). Additionally, the coincidence of each protein accepted or rejected by both methods was also evaluated (see Appendices A–H at www.actabp.pl); i.e. both methods were calibrated with one of the groups (Table 4) (training set) and were tested with the others (test sets), and this procedure was conducted for each group studied.

Table 2. Eigenvectors (vi) calculated by the Bluebit Software http://www.bluebit.gr/, accessed Nov 26, 2014 (Hait, 2002), from the polar matrix (Table 1) of each peptide group from the APD2 Database (Dec 2012) (Wang & Wang, 2009), (see Item 1, Polarity Vector Method Metrics section).

Eigenvectors from the polar matrix				
Gram (+) bacteria	v1	v2	v3	v4
	0.216818+0.000000i	0.598261+0.000000i	-0.618116+0.000000i	-0.556458+0.000000i
	0.073043+0.000000i	-0.167398+0.000000i	0.468525+0.000000i	-0.627495+0.000000i
	0.708651+0.000000i	0.435142+0.000000i	0.506082+0.000000i	0.479056+0.000000i
	0.667435+0.000000i	-0.651700+0.000000i	-0.377224+0.000000i	-0.259056+0.000000i
Gram (-) bacteria	v1	v2	v3	v4
	-0.288445+0.000000i	-0.336067+0.000000i	0.896464+0.000000i	0.069243+0.000000i
	-0.097335+0.000000i	-0.224416+0.000000i	-0.230048+0.000000i	-0.884079+0.000000i
	-0.656654+0.000000i	0.750342+0.000000i	0.037634+0.000000i	-0.286375+0.000000i
	-0.690022+0.000000i	-0.523147+0.000000i	-0.376849+0.000000i	0.362765+0.000000i
Viruses	v1	v2	v3	v4
	0.258017+0.000000i	0.763161+0.000000i	0.763161+0.000000i	-0.234183+0.000000i
	0.105471+0.000000i	-0.167749+0.133205i	-0.167749-0.133205i	-0.956391+0.000000i
	0.768635+0.000000i	-0.355820-0.290217i	-0.355820+0.290217i	0.113722+0.000000i
	0.575763+0.000000i	0.148042+0.372761i	0.148042-0.372761i	0.132449+0.000000i
Parasites	v1	v2	v3	v4
	0.282287+0.000000i	0.609505+0.000000i	0.609505+0.000000i	-0.010177+0.000000i
	0.106931+0.000000i	-0.287230-0.114225i	-0.287230+0.114225i	-0.993330+0.000000i
	0.712082+0.000000i	0.237653+0.341110i	0.237653-0.341110i	0.089027+0.000000i
	0.633892+0.000000i	-0.481222-0.358533i	-0.481222+0.358533i	0.072565+0.000000i
Insects	v1	v2	v3	v4
	0.385473+0.000000i	0.644822+0.000000i	0.644822+0.000000i	0.417758+0.000000i
	0.109107+0.000000i	0.069113+0.197119i	0.069113-0.197119i	-0.871145+0.000000i
	0.547585+0.000000i	0.169515+0.432305i	0.169515-0.432305i	0.161319+0.000000i
	0.734613+0.000000i	-0.473112+0.317983i	-0.473112+0.317983i	-0.201396+0.000000i
Mammalian cells	v1	v2	v3	v4
	0.284377+0.000000i	0.837276+0.000000i	0.837276+0.000000i	0.496367+0.000000i
	0.079361+0.000000i	0.347860+0.222396i	0.347860-0.222396i	-0.228493+0.000000i
	0.492869+0.000000i	0.013855+0.046164i	0.013855-0.046164i	0.636628+0.000000i
	0.818481+0.000000i	-0.351835-0.048897i	-0.351835+0.048897i	-0.544166+0.000000i
Fungi	v1	v2	v3	v4
	-0.341511+0.000000i	-0.626062+0.000000i	-0.848264+0.000000i	-0.244428+0.000000i
	-0.101370+0.000000i	-0.318251+0.000000i	-0.079870+0.000000i	-0.905484+0.000000i
	-0.678005+0.000000i	-0.266727+0.000000i	0.517107+0.000000i	0.335880+0.000000i
	-0.642964+0.000000i	0.660015+0.000000i	-0.081671+0.000000i	-0.086822+0.000000i
Cancer cells	v1	v2	v3	v4
	-0.222751+0.000000i	-0.770622+0.000000i	-0.433813+0.000000i	0.274795+0.000000i
	-0.075727+0.000000i	0.616004+0.000000i	-0.882999+0.000000i	-0.231170+0.000000i
	-0.718040+0.000000i	0.022753+0.000000i	0.121349+0.000000i	-0.669551+0.000000i
	-0.655032+0.000000i	0.161749+0.000000i	0.131887+0.000000i	0.650191+0.000000i

RESULTS

The PIM and PVM methods (Table 4) show a high affinity to identify the group studied (>70%) and are discriminative against other groups. The analysis of each sequence, individually evaluated by both methods, shows a coincidence higher than 98% (Appendices A–H at www.actabp.pl); therefore, it can be stated that both methods are equivalent.

Table 3. Eigenvectors matrix calculated (see Items 1 & 2, Polarity Vector Method Metrics section) by peptide groups from the APD2 Database (Dec 2012) (Wang & Wang, 2009).

Eigenvectors matrix				
Gram (+) bacteria	P+	P-	N	NP
P+	0.033333	0.054167	0.008333	0.012500
P-	0.029167	0.025000	0.041667	0.004167
N	0.062500	0.037500	0.050000	0.045833
NP	0.058333	0.000001	0.016667	0.020833
Gram (-) bacteria	P+	P-	N	NP
P+	0.076845	0.067878	0.157545	0.130645
P-	0.112712	0.103745	0.094778	0.023045
N	0.040978	0.148578	0.121678	0.085812
NP	0.032012	0.049945	0.058912	0.139612
Viruses	P+	P-	N	NP
P+	0.151320	0.178300	0.191789	0.043402
P-	0.083871	0.070382	0.056892	0.002933
N	0.205279	0.016423	0.029912	0.097361
NP	0.164810	0.137830	0.124341	0.110851
Parasites	P+	P-	N	
P+	0.129612	0.140855	0.152098	0.062153
P-	0.095883	0.039667	0.050910	0.005938
N	0.174584	0.118369	0.107126	0.084639
NP	0.163341	0.017181	0.028424	0.073396
Insects	P+	P-	N	NP
P+	0.121428	0.155714	0.167142	0.132857
P-	0.075714	0.064286	0.052857	0.007143
N	0.144285	0.110000	0.098571	0.087143
NP	0.178571	0.018572	0.030000	0.041429
Mammalian Cells	P+	P-	N	NP
P+	0.108086	0.215408	0.230740	0.169413
P-	0.092754	0.138749	0.123417	0.046758
N	0.154081	0.077422	0.062090	0.184745
NP	0.200076	0.016095	0.031427	0.000763
Fungi	P+	P-	N	NP
P+	0.050822	0.041581	0.013858	0.078545
P-	0.087786	0.060063	0.115509	0.004617
N	0.023099	0.069304	0.133991	0.124750
NP	0.032340	0.143232	0.097027	0.106268
Cancer cells	P+	P-	N	NP
P+	0.094689	0.018757	0.069379	0.170621
P-	0.107345	0.183277	0.006102	0.082034
N	0.031413	0.120000	0.132655	0.044068
NP	0.056723	0.157966	0.145311	0.195932

The spatial distribution of the eigenvectors of each group studied is not discriminative (Figs 1–8), as in all groups the cumulus is located in quadrants I and II, except in the insects group that is located in quadrant I (Fig. 7).

DISCUSSION

Although in practical terms the Polarity Vector Method is equally discriminative as the Polarity Index Method, it is important to note that it includes in its met-

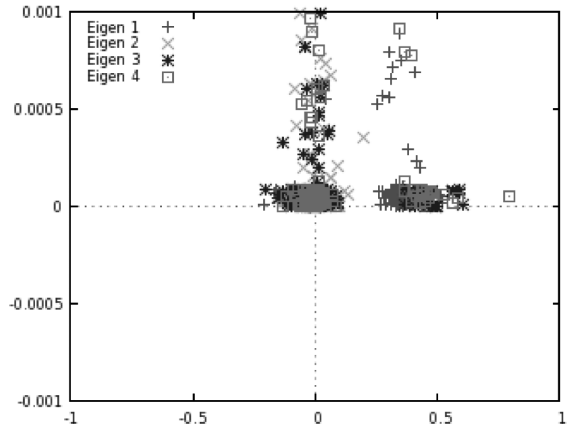

Figure 1. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix I at www.actabp.pl) of the Gram (+) bacteria group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

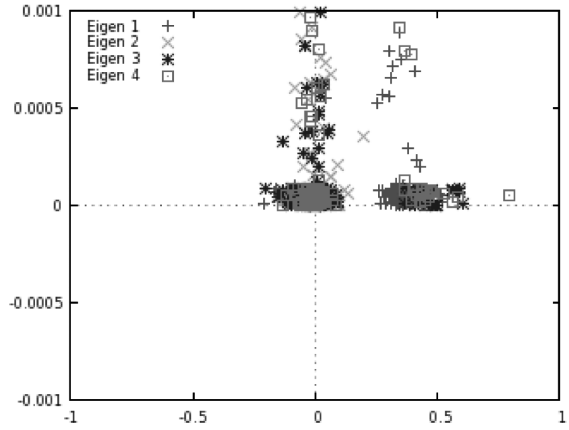

Figure 2. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix J at www.actabp.pl) of the Gram (–) bacteria group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

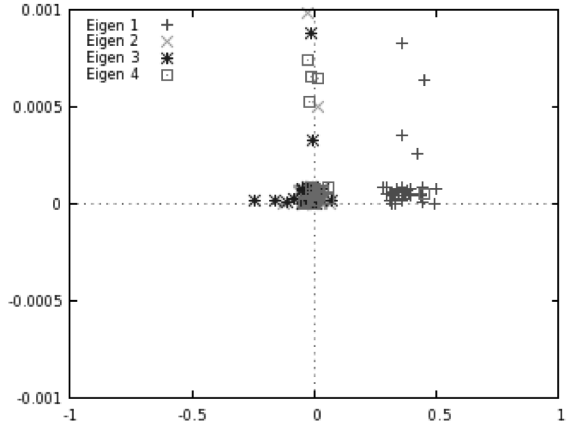

Figure 3. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix K at www.actabp.pl) of the virus group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

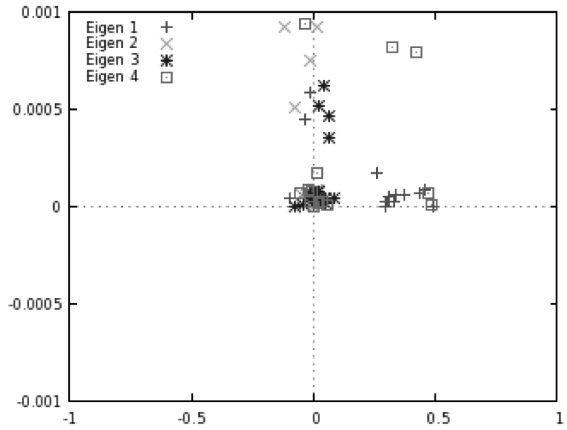

Figure 4. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix L at www.actabp.pl) of the parasite group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

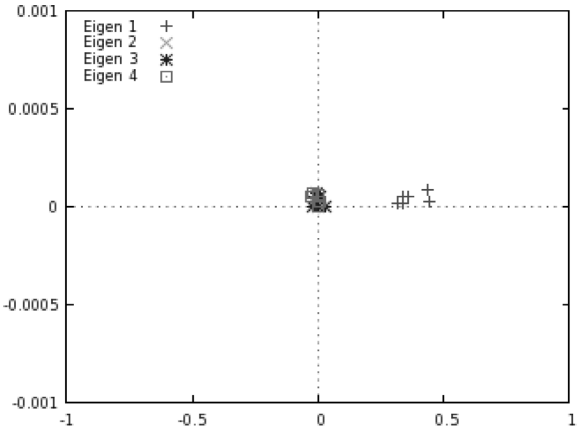

Figure 5. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix M at www.actabp.pl) of the insect group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

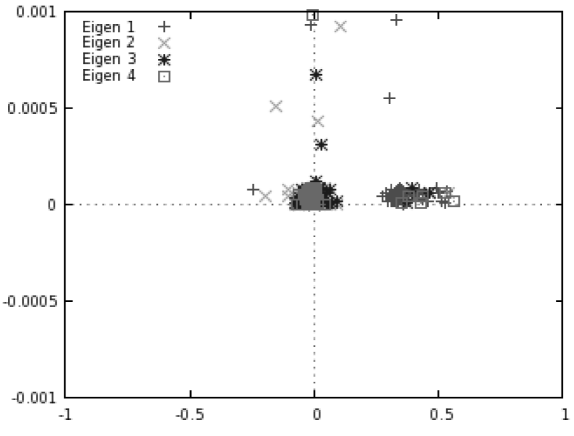

Figure 6. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix N at www.actabp.pl) of the mammalian cells group from APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

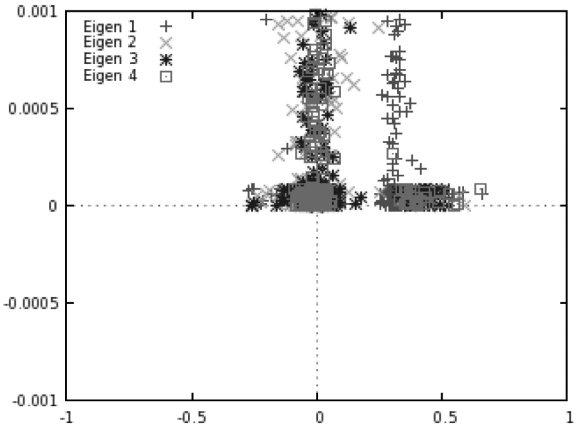

Figure 7. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix O at www.actabp.pl) of the fungi group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

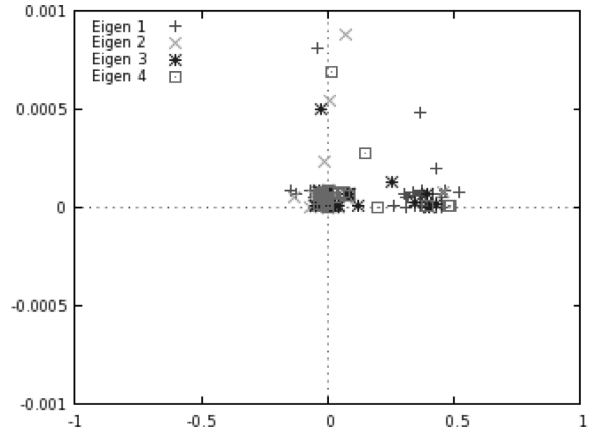

Figure 8. Spatial distribution of the eigenvalues (see Phase Portrait section, Appendix P at www.actabp.pl) of the cancer cells group from the APD2 Database accessed in December 2012 (Wang & Wang, 2009).
The X-axis corresponds to the *real part* of the eigenvalue, and the Y-axis to its *imaginary part*.

Table 4. Percentage of hits calculated by the Polarity Index Method (PIM) (see Polarity Index Method Metrics section) and Polarity Vector Method (PVM) (see Polarity Vector Method section), for the eight groups of antimicrobial peptides from the ADP2 database (Dec 2012) (Wang & Wang, 2009).

Matches by antimicrobial group								
	Gram (+) Bacteria	Gram (–) Bacteria	Viruses	Parasites	Insects	Mammalian cells	Fungi	Cancer cells
Gram (+) Bacteria								
PIM	71	47	31	40	14	44	45	41
PVM	72	37	17	30	14	35	33	31
Gram (-) Bacteria								
PIM	53	70	39	35	29	48	44	44
PVM	47	70	59	55	43	69	58	56
Viruses								
PIM	40	23	72	30	14	22	24	30
PVM	24	17	69	25	0	28	18	22
Parasites								
PIM	8	5	6	70	0	5	6	11
PVM	7	5	6	80	0	6	6	13
Insects								
PIM	0	0	0	0	86	0	0	0
PVM	0	0	0	0	86	0	0	0
Mammalian cells								
PIM	32	27	15	20	0	72	31	31
PVM	25	32	22	30	29	72	36	24
Fungi								
PIM	64	65	54	30	43	66	71	39
PVM	54	57	50	40	29	57	71	43
Cancer cells								
PIM	8	18	24	20	14	17	11	70
PVM	23	27	37	30	0	28	22	72

rics the eigenvectors of the polarity matrix studied. This made possible to study the span of eigenspace, showing that although the eigenvectors (v) and eigenvalues (λ) of the system are related from the expression $Av=\lambda v$, according to the results, the eigenvalues are not effective discriminants. This can be attributed to the fact that

each eigenvalue (λ: complex number) is associated with an eigenvector (r: vector column formed by four complex numbers); in this sense, the exhaustive character of the metric formed by the eigenvectors shows a regularity that is lost when the eigenvalues are compared. Furthermore, the metric of the Polarity Vector Method does not depend on previous calculations, therefore, its execution with high-performance computational architectures makes it possible to predict the affinity of a peptide or protein in a processing time t_p/n, where "n" is the number of computer processors.

PVM depends on a single physico-chemical property, polarity, that quantifies the electromagnetic balance of the protein. It originates from the electronegativity of the valence electrons in the constituent amino acids. Linus Pauling (Pauling, 1955) defined the "electronegativity" as: the affinity between the electrons in a covalent bond. This property has been verified as "necessary and sufficient" to efficiently identify the major association of a protein. Although there are other physico-chemical properties that have been used together as a metric i.e., hydrophobicity (Borgese & Fasana, 2011), isoelectric point (Kidman *et al*., 2004), and net charge (Shaw *et al*., 2001) among others; it was important to find a physico-chemical property capable to describe the activity of a protein by itself, which would give an impulse to basic science and will allow the cleaning of bioinformatics codes used for this purpose.

This bioinformatics product can contribute to the computational and structural proteomics research. The performance of the method is high, and using a single physico-chemical property could enable scholars to gain a deeper insight about polarity. This fundamental property of matter, strengthens the field of bioinformatics since its metric adapts smoothly to parallel and distributed processing schemes, making possible the assessment of all peptides and proteins in the public databases.

Finally, it is worth mentioning the importance of encouraging the creation and use of public databases, as a significant part of basic research is founded on the availability of free and updated information that is carefully revised. This has been mainly the reason for using the APD2 database for many years.

CONCLUSIONS

The Polarity Vector Method is a robust and highly discriminative method that can be used as a "first filter" in the identification of antimicrobial peptides. Its programming scheme also allows its execution with high-performance computing platforms for the comprehensive analysis of peptide regions.

Availability

The F77 programs, test source files and scripts for both methods (PIM and PVM) are available from the author by request (polanco@unam.mx).

Conflict of Interests

I declare that I do not have any financial or personal interest with other people or organizations that could inappropriately influence (bias) this work.

Acknowledgments

The author thanks Concepción Celis Juárez for her suggestions and proof-reading that have improved the original manuscript, he also acknowledges the Computer Science department at the Institute for Nuclear Sciences at Universidad Nacional Autónoma de México for support.

REFERENCES

Adams CP, Brantner VV (2006) Estimating the cost of new drug development: Is it really 802 million dollars? *Health Aff* **25**: 420–428.

Bahar AA, Ren D (2013) Antimicrobial Peptides. *Pharmaceuticals* **6:** 1543–1575. http://doi.org/10.3390/ph6121543.

Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Chelliah V, Worth CL, Harmer NJ, Davies O, Burke D (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos Trans R Soc Lond B Biol Sci* **361**: 413–423. DOI:10.1098/rstb.2005.1800.

Borgese N, Fasana E (2011) Targeting pathways of C-tail-anchored proteins. *Biochim Biophys Acta* **1808**: 937–946. DOI: 10.1016/j.bbamem.2010.07.010.

Breda A, Fonseca Valadares N, Norberto de Souza O, Garratt RC (2006) Protein Structure, Modelling and Applications. 2006 May 1 [Updated 2007 Sep 14]. In *Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach* [Internet]. Gruber A, Durham AM, Huynh C, del Portillo HA eds. Bethesda (MD): National Center for Biotechnology Information (US); 2008. Chapter A06. Available from: http://www.ncbi.nlm.nih.gov/books/NBK6824/.

Dudley JT, Deshpande T, Butte AJ (2011) Exploiting drug-disease relationships for computational drug repositioning. *Brief Bioinform* **12**: 303–311.

Eaton JW, Bateman D, Hauberg S (2009) GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations. *CreateSpace Independent Publishing Platform*. ISBN 1441413006, http://www.gnu.org/software/octave/doc/interpreter accessed Dec 18, 2014.

Ekins S, Mestres J, Testa B (2007). *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Brit J Pharmacol* **152**: 9–20. http://doi.org/10.1038/sj.bjp.0707305.

Ekins S (2004) Predicting undesirable drug interactions with promiscuous proteins *in silico*. *Drug Discov Today* **9**: 276–285.

Gill AL, Verdonk M, Boyle RG, Taylor R (2007) A comparison of physicochemical property profiles of marketed oral drugs and orally bioavailable anti-cancer protein kinase inhibitors in clinical development. *Curr Top Med Chem* **7**: 1408–1422.

Hait DJ (2002) *Bluebit Software*. http://www.bluebit.gr/ accessed Nov 26, 2014.

Kantardjieff K, Rupp B (2004) Structural bioinformatic approaches to the discovery of new antimycobacterial drugs. *Curr Pharm Des* **10**: 3195–3211.

Kidman G, Park H, Northrop DB (2004) Pressure stability of proteins at their isoelectric points. *Protein Pept Lett* **11**: 543–546.

Liu B, Wang X, Chen Q, Dong Q, Lan X (2012) Using amino acid physicochemical distance transformation for fast protein remote homology detection. *PLoS ONE* **7**: e46633. http://doi.org/10.1371/journal.pone.0046633.

Pauling L (1955) *General Chemistry,* 3rd edn W. H. Freeman and Company Publishers, San Francisco Calif, USA.

Polanco C, Buhse T, Castañón-González JA, Samaniego JL (2014d) Possible computational filter to detect proteins associated to influenza A subtype H1N1. *Acta Biochim Pol* **61**: 693–698.

Polanco C, Buhse T, Samaniego JL, Castañón-González JA (2013) Detection of selective antibacterial peptides by the Polarity Profile Method. *Acta Biochim Pol* **60**: 183–189.

Polanco C, Castañón-González JA, Samaniego JL (Letter to the Editor) Arabi YM, Arifi AA, Balkhy HH, Najm H, Aldawood AS, Ghabashi A, Hawa H, Alothman A, Khaldi A, Raiy B (2014) Clinical course and outcomes of critically ill patients with middle east respiratory syndrome coronavirus infection. *Ann Intern Med* DOI: 10.7326/M13-2486.

Polanco C, Castañón-González JA, Uversky VM (Letter to the Editor) Buhimschi IA, Nayeri UA, Zhao G, Shook LL, Pensalfini A, Funai EF, Bernstein IM, Glabe CG, Buhimschi CS (2014b) Protein misfolding, congophilia, oligomerization, and defective amyloid processing in preeclampsia. *Sci Transl Med* **6**: 245ra92. DOI: 10.1126/scitranslmed. 3008808.

Polanco C, Castañón-González JA, Uversky VM, Samaniego JL, Buhse T (Letter to the Editor) G Vogel G (2014d) Delays hinder Ebola genomics. *Science* **346**: 684–685. DOI: 10.1126/science.346.6210.684.

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 Peptides* **2012**: 58502. DOI:10.1155/2012/585027.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T (2014) Polar profile of antiviral peptides from AVPpred database. *Cell Biochem Biophys* **70**: 1469–1477. DOI: 10.1007/s12013-014-0084-4.

Polanco C, Samaniego JL, Castañón-González JA, Buhse T, Sordo ML (2013a) Characterization of a possible uptake mechanism of selective antibacterial peptides. *Acta Biochim Pol* **60**: 629–633.

Polanco C, Samaniego-Mendoza JL, Buhse T, Castañón-González JA, Leopold-Sordo M (2014a) Polar Characterization of antifungal peptides from APD2 database. *Cell Biochem Biophys* **70**: 1479–1488. DOI: 10.1007/s12013-014-0085-3.

Polanco C, Samaniego-Mendoza JL, Castañón-González JA, Buhse T (Letter to the Editor) Howard SJ, Hopwood S, Davies SC (2014c) Antimicrobial Resistance: A Global Challenge Sci Transl Med. DOI: 10.1126/scitranslmed. 3009315.

Readhead B., Dudley J (2013) Translational bioinformatics approaches to drug development. *Advances in Wound Care* **2**: 470–489. http://doi.org/10.1089/wound.2012.0422.

Sahai V, Bist V (2002) Linear Algebra, Alpha Science International Ltd, Pangbourne RG8 8UT, UK.

Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci* **10**: 1206–1215. DOI: 10.1110/ps.440101.

Timberlake KC (1992) The chemistry of life. In *Chemistry* 5th edn, Haper-Collins Publishers Inc, NY.

Tomlinson E (1975) Chromatographic hydrophobic parameters in correlation analysis of structure-activity relationships. *J Chromatogr* **113**: 1–45.

Vilar S, González-Díaz H, Santana L, Uriarte E (2008) QSAR model for alignment-free prediction of human breast cancer biomarkers based on electrostatic potentials of protein pseudofolding HP-lattice networks. *J Comput Chem* **29**: 2613–2622. DOI: 10.1002/jcc.21016.

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. DOI: 10.1093/nar/gkn823, accessed December 19, 2012.