Cell Biochem Biophys (2014) 70:1469–1477
DOI 10.1007/s12013-014-0084-4

ORIGINAL PAPER

Polar Profile of Antiviral Peptides from AVPpred Database

Carlos Polanco · José Lino Samaniego · Jorge Alberto Castañón-González · Thomas Buhse

Published online: 4 July 2014

Abstract Diseases of viral origin in humans are among the most serious threats to health and the global economy. As recent history has shown the virus has a high pandemic potential, among other reasons, due to its ability to spread by air, hence the identification, investigation, containment, and treatment of viral diseases should be considered of paramount importance. In this sense, the bioinformatics research has focused on finding fast and efficient algorithms that can identify highly toxic antiviral peptides and to serve as a first filter, so that trials in the laboratory are substantially reduced. The work presented here contributes to this effort through the use of an algorithm already published by this team, called polarity index method, which identifies with high efficiency antiviral peptides from the exhaustive analysis of the polar profile, using the linear sequence of the peptide. The test carried out included all peptides in APD2 Database and 60 antiviral peptides identified by Kumar and co-workers (Nucleic Acids Res 40:W199–204, 2012), to build its AVPpred algorithm. The validity of the method was focused on its discriminating capacity so we included the 15 sub-classifications of both Databases.

Keywords Polarity index method · Antiviral peptides · AVPpred algorithm

Abbreviations

SCAAP	Selective cationic amphipathic antibacterial peptides
APD2	Antimicrobial peptide database http://aps.unmc.edu/AP/ [1] accessed December 19, 2012. AVPpred, http://crdd.osdd.net/servers/avppred [2] accessed March 10, 2013
QSAR	Quantitative structure activity relationships
SVM	Support vector machine
SARS	Severe acute respiratory syndrome
HIV	Human immunodeficiency virus
AAindex	Amino acid indices, substitution matrices and pair-wise contact potentials database http://www.genome.jp/aaindex/ [3] accessed March 10, 2013

Introduction

Within the antimicrobial peptides there is a group of particular importance, the antiviral peptides. These peptides are potential drugs to face the increased incidence of chronic viral infections caused by HIV and hepatitis [4, 5]; for that reason there is a need to accelerate the development of synthetic antiviral peptides. Inclusive, the testing of vaccines against HIV and hepatitis [6, 7] that are under

Electronic supplementary material The online version of this article (doi:10.1007/s12013-014-0084-4) contains supplementary material, which is available to authorized users.

C. Polanco · J. L. Samaniego · J. A. Castañón-González
Facultad de Ciencias de la Salud, Universidad Anáhuac, Av. Universidad Anáhuac No. 46, Col. Lomas Anáhuac, 52786 Huixquilucan, Estado de México, Mexico

C. Polanco (✉) · J. L. Samaniego
Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Av. Universidad 3000, 04510 Mexico, DF, Mexico
e-mail: polanco@unam.mx

T. Buhse
Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, 62209 Cuernavaca, Morelos, Mexico

Table 1 Q[*i*,*j*] Polarity matrix

	P+	P−	N	NP
P+	0.0354861617	0.0184528027	0.0354861617	0.0681334287
P−	0.0163236335	0.0177430809	0.0298083741	0.0709723234
N	0.0404542238	0.0326472670	0.0872959569	0.0908445716
NP	0.0667139813	0.0617459193	0.0915542915	0.1937544346

Incidences of 60 AVPpred antiviral peptides extracted from Kumar [2], and one antiviral peptide from APD2 [1]

development offers no guarantee of success regardless of the millions of infected people. In counterpart, antiviral medications are not oriented to the most acute infections that cause serious diseases, such as hemorrhagic fever and cancer [8–10]; they have limited effectiveness and serious side effects. Perhaps more importantly, the antiviral chemotherapy is producing a rapid development of drug-resistant strains, as a result of the high rate of virus replication, due to the low resistance to replication [11].

In such a scenario, the work in proteomics and bioinformatics should focus on the generation of fast and robust algorithms [12, 13] identifying the antiviral action, from the linear or three-dimensional structure of the peptide. However, the simulation of the three-dimensional structure of the peptide is very complex, without taking into account other factors involved such as: The dynamics of the membrane and toxicity. In this work, we use the Polarity index method [14], already published by our team to identify SCAAP [15–17] and which only uses the linear peptide sequence to identify the same group of antiviral peptides that were identified by AVPpred algorithm [2].

This method [14] generates an exhaustive analysis of the peptide polarity through its polarity matrix. In this sense, the method apparently does not consider other factors, but indirectly it does, because it requires being calibrated by "a set of peptides" that are characteristic of the profile.

Materials and Methods

The identification of antiviral peptides performed by the polarity index method [14] requires the modifications below (Supplementary Material).

Polarity Index method. Updates

Modifications

1. Replacing the Q[*i*,*j*] matrix in the source program [14] by the Table 1, which represents the incidences of antiviral sequences with a unique pathogenic action. Table 1 considers 60 AVPpred antiviral peptides extracted from Kumar and co workers [2], and 1 antiviral peptide extracter from APD2 database [1].
2. Replacing the rule in the source program [14] by P[*i*,*j*] + Q[*i*,*j*] vector complying with the next rule: polar interactions 15 or 16 are present in the 1st position, polar interaction 14 is not present from 3th to 8th positions, polar interaction 2 is not present in the first seven positions, polar interaction 5 is not present in the first five positions, polar interactions 4 or 8 are not present from 12th to 16th positions, polar interactions 14, 15 or 16 are not present from 11th to 16th positions, polar interaction 9 is not present in the 1st, 4th, 5th or 9th positions, polar interaction 1 is not present in the 10th or 11th positions, polar interaction 11 is not present in the first position, polar interaction 1 is not present in 10th or 11st positions, polar interaction 14 is not present in the 5th position, and polar interaction 4 is not present in the 4th position (Table 2).

APD2 Database Trial Data Preparation

We use 3,636 peptides in the APD2 Database [1] classified by their *multiple* action against: 149 Gram− ONLY, 1711 Gram +/Gram− ONLY, 315 Gram+ ONLY, 141 cancer cells, 9 sperms, 88 HIV, 744 fungi, 21 insects, 244 mammalian cells, 47 parasites, 3 protists, 39 chemotaxis, 0 SCAAP, 125 virus; and also 1,059 by their *unique* action against: 111 Gram− ONLY, 213 Gram+ ONLY, 518 Gram+/Gram− ONLY, 20 cancer cells, 0 HIV, 88 fungi, 35 H1N1, 2 insects, 11 mammalian cells, 9 parasites, 1 protists, 0 chemotaxis, 30 SCAAP, 0 sperms and 21 virus.

AVPpred Algorithm Trial Data Preparation

From Kumar and co workers [2] work about antiviral peptides, we took 60 validated and experimental peptides from 1,245 antiviral peptides. His work evaluated these 60 peptides with 25 physicochemical properties (Table 3) out of 144 properties from the same database and called it AAindex database [3]. These peptides were used to build the SVM AVPpred algorithm [2], and we are using them here to validate the Polarity index method

Table 2 Polarity index method rules

Position P[*i*,*j*] + Q[*i*,*j*] vector of study.	1 (1,1)	2 (1,2)	3 (1,3)	4 (1,4)	5 (2,1)	6 (2,2)	7 (2,3)	8 (2,4)	9 (3,1)	10 (3,2)	11 (3,3)	12 (3,4)	13 (4,1)	14 (4,2)	15 (4,3)	16 (4,4)
Rule # 1	✔															
Polar interactions 15 or 16 are present in the 1st position																
Rule # 2			×	×	×	×	×	×								
Polar interaction 14 is not present from 3th to 8th positions																
Rule # 3	×	×	×	×	×	×	×									
Polar interaction 2 is not present in the first seven positions																
Rule # 4	×	×	×	×	×											
Polar interaction 5 is not present in the first five positions																
Rule # 5												×	×	×	×	×
Polar interactions 4 or 8 are not present from 12th to 16th positions																
Rule # 6											×	×	×	×	×	×
Polar interactions 14, 15 or 16 are not present from 11th to 16th positions																
Rule # 7	×			×	×					×						
Polar interaction 9 is not present in the 1st, 4th, 5th or 9th positions																
Rule # 8										×	×					
Polar interaction 1 is not present in the 10th or 11th positions																
Rule # 9	×															
Polar interaction 11 is not present in the first position																
Rule # 10										×	×					
Polar interaction 1 is not present in 10th or 11st positions																
Rule # 11					×											
Polar interaction 14 is not present in the 5th position																
Rule # 12				×												
Polar interaction 4 is not present in the 4th position																

Identification Rules in Polarity index method for antiviral peptides. (✔) The polar interaction is present in the position. (×) The polar interaction is not present in the position

Results

Polarity index method is an algorithm that determines the probable antiviral pathogen action of peptides by using the peptide polarity sequence. It was applied to the APD2 database and the experimental peptides from AVPpred algorithm [2] with the following results.

Table 3 Physicochemical properties from AVPpred

#	AAindex ID	Description	Polarity	Reference
1	BEGF750103	Conformational parameter of beta-turn	✔	[18]
2	BULH740102	Apparent partial specific volume		[19]
3	CHAM810101	Steric parameter	✔	[20]
4	CHOP780204	Normalized frequency of N-terminal helix	✔	[21]
5	CHOP780206	Normalized frequency of N-terminal non helical region	✔	[21]
6	CHOP780215	Frequency of the 4th residue in turn	✔	[21]
7	CIDH920104	Normalized hydrophobicity scales for alpha/beta-proteins	✔	[21]
8	COHE430101	Partial specific volume		[22]
9	FASG760105	PK-C	✔	[23]
10	FAUJ880104	STERIMOL length of the side chain		[24]
11	FINA770101	Helix-coil equilibrium constant	✔	[25]
12	FINA910101	Helix initiation parameter at position $i - 1$		[26]
13	GEIM800101	Alpha-helix indices	✔	[27]
14	GEIM800102	Alpha-helix indices for alpha-proteins	✔	[27]
15	GEIM800104	Alpha-helix indices for alpha/beta-proteins	✔	[27]
16	KARP850101	Flexibility parameter for no rigid neighbors	✔	[28]
17	KARP850102	Flexibility parameter for one rigid neighbor	✔	[28]
18	AURR980101	Normalized positional residue frequency at helix termini N4′	✔	[29]
19	AURR980118	Normalized positional residue frequency at helix termini C”	✔	[29]
20	AURR980120	Normalized positional residue frequency at helix termini C4′	✔	[29]
21	AVBF000107	Slopes tripeptide FDPB PARSE neutral		[30]
22	GEOR030102	Linker propensity from 1-linker dataset		[31]
23	KIDA850101	Hydrophobicity-related index	✔	[32]
24	GUYH850102	Apparent partition energies calculated from Wertz-Scheraga index	✔	[33]
25	CASG920101	Hydrophobicity scale from native protein structures	✔	[34]

Selected physicochemical properties to build the AVPpred algorithm by Kumar [2]. *AAindex* amino acid index database [3]. *Polarity* (✔) physicochemical properties are directly or indirectly related to the polarity

Table 4 Polarity index matches by linear sequence in virus

No	Code	ID PUBMED	Sequence	#1	#2	References
1	AVP_0618	6096849	GPPISLERLDVGTNLGNAIAKLEAKELLESSDQI			[35]
2	AVP_0629	3788062	KVLHLEGEVNKIALLSTNKAVVSLSNGVSVLTS		N	[36]
3	AVP_0607	2893293	DFLEENITALLEEAQIQQEKNMYELQKLNSWDVFG			[37]
4	AVP_0168	8382405	EGPTLGNWAREIWATLFGKA		N	[38]
5	AVP_0179	8382405	NWAREIWATLFKKA		N	[38]
6	AVP_0467	10390360	FAIKWEYVLLLFLL			[39]
7	AVP_0512	1848704	SWLRDIWDWKCEVLSDFK			[40]
8	AVP_0514	1848704	SWLRDIWDWLCEVLSDFK			[40]
9	AVP_0373	1848704	SWLRDIWDWICEVLSDFK			[40]
10	AVP_0372	52472831	SWLRDIWDWICEVLSD			[41]
11	AVP_0387	9223423	TWLRAIWDWVCTALTDFK			[42]
12	AVP_0323	8822631	PPVYTKDVDISSQISSMNQSLQQSKDYIKEAQKILDTVNPSL			[43]
13	AVP_0328	3012869	VANDPIDISIELNKAKSDLEESKEWIRRSNQKLDSD		N	[44]
14	AVP_0210	11118300	ANTAFVSSAHNTQKIPAGAPFNRNLRAMLADLRQNAAFAG			[45]
15	AVP_0024	1695254	CGGNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQ			[46]
16	AVP_0312	16667080	EQCREEEDDR		N	[47]
17	AVP_0358	15893660	GGTIFDCGETCFLGTCYTPGCSCGNYGFCYGTN		N	[48]

Table 4 continued

No	Code	ID PUBMED	Sequence	#1	#2	References
18	AVP_0019	7841460	GICRCICGKGICRCICGR			[49]
19	AVP_0284	21685289	GICRCICGRGICRCICGR			[50]
20	AVP_0397	7529412	GIKEWKRIVQRIKDFLRNLV		N	[51]
21	AVP_0361	16872274	GLPVCGETCVGGTCNTPGCTCSWPVCTRN		N	[52]
22	AVP_0286	10521339	GVCRCLCRRGVCRCICRR			[53]
23	AVP_0304	15949629	KTCENLADTY			[54]
24	AVP_0684	7031661	LEAIPCSIPPCFLFGKPFVF			[55]
25	AVP_0692	7031661	LEAIPISIPPELAFAKPFVF			[55]
26	AVP_0703	7031661	LEAIPMSIPPEVFFGKPFVF			[55]
27	AVP_0155	9777331	LSYRCPCR		N	[56]
28	AVP_0409	1433527	WMEWDREIEELAKKAEELAKKAEELAKKAWASLWNWF			[57]
29	AVP_0222	3031048	YALLIRMIYKNI			[58]
30	AVP_0224	19468303	YQLLARMIYKNI			[59]
31	AVP_0225	9504927	YQLLIAMIYKNI			[60]
32	AVP_0584	2578615	YTSLIHSLIEESQNQQEKNEQELLEFDKWASLWNWF		N	[61]
33	AVP_0591	3040055	YTSLIHSLIEESQNQQEKNEQELLELNKWASLWNWF		N	[62]
34	AVP_0310	9516047	GLFGVLGSIAKHVLPHVVPVIAEKL		N	[63]
35	AVP_0183	7826699	DWHLGQGVSIEWRKK			[64]
36	AVP_0444	9784389	FLFPLITSFLSKVL			[65]
37	AVP_0452	10951191	GLFDIIKKIAESW			[66]
38	AVP_0459	12089438	GWLKKIESIIDAF			[67]
39	AVP_0460	11053427	HVDKKVADKVLLLKQLRIMRLLTRL		N	[68]
40	AVP_0348	7826699	TTWEAWDRAIAEYAARIEALIRAAQEQQEKNEAILREL			[64]
41	AVP_0354	7826699	TTWEAWDRAIAEYAARIEALIRASQEQQEKNEAELREL			[64]
42	AVP_0612	3012869	ELNKAKSDLEESKEWIRRSNQKLDSIGNWHQSSTT		N	[44]
43	AVP_0623	3012869	IELNKAKSDLEESKEWIRRSNQKLDSIGNWHQSST		N	[44]
44	AVP_0390	3012465	AAHLIDALYAEFLGGRVLTT			[69]
45	AVP_0068	10077657	HRWRKRWRKHRWRKRWRK			[70]
46	AVP_0061	10077657	KRWRKRWRKWRWRKRWRK			[70]
47	AVP_0055	15095345	RTRKRGRKRTRKRGRK			[71]
48	AVP_0191	14648299	RGGKIAGKIAKIAGKIAKIAGKIA			[72]
49	AVP_0134	3012869	ISIELNKAKSDLEESKEWIRRSNQKLDSIGNWHQS		N	[44]
50	AVP_0482	2959959	RRKKAAVALLPAVLLA			[73]
51	AVP_0483	2959959	RRKKAAVALLPAVLLAL			[73]
52	AVP_0302	2661722	KDLLFK		N	[74]
53	AVP_0113	3788062	GPPISLERLDVGTNLGNAIAKLEDAKELLESSDQI			[35]
54	AVP_0114	3788062	HRIDLGPPISLERLDVGTNLGNAIAKLEDAKELLE			[35]
55	AVP_0116	3788062	ISLERLDVGTNLGNAIAKLEDAKELLESSDQILRS			[35]
56	AVP_0288	12810954	CGECGGGHIVGRFCMVVRFLRLVFI			[75]
57	AVP_0289	9634230	CRCCELKSLCPTLMRVVRLLGLVLL		N	[76]
58	AVP_0138	6096849	LVFPSDEFDASISQVNEKINQSLAFIRKSDELLHN			[77]
59	AVP_0278	2833514	KWKLFKKIGIGKFLHVAKKF			[78]
60	AVP_0335	12886019	CDVIALLCHLNT			[79]
61	APD2_01209	7744961	RICRCICGRRICRCICGR			[80]

Subject sequences identified by polarity index method in APD2 database [1] and by AVPpred [2], where peptides have action only on virus. #1: (N) rejected peptide by SVM AVPpred algorithm. #2: (N) rejected peptide by polarity index method. Source: National Center for Biotechnology Information, US National Library of Medicine http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins in database: Swiss-Prot (swissprot), accessed March 20, 2013

Table 5 Polarity index matches by pathogenic action

Number of hits	Gram+ only	Gram− only	Gram+/ Gram−	Virus	%	HIV	Fungi	Protists	Parasites	Insects	Sperms	Cancer cells	Mammalian cells	Chemotaxis	SCAAP	AVPpred	%
Unique action	38	19	134	1	100	0	6	0	1	0	0	2	2	0	13	42	70
	213	111	518	1		0	88	1	9	2	0	20	11	0	51	60	
Multiple action	66	27	426	19		16	163	0	12	3	1	28	81	8	0	0	
	315	149	1,711	125		18	744	3	47	21	9	141	244	39	0	0	

Polarity index method matches for both APD2 [1] and AVPpred antiviral peptide groups [2]. *Unique action* peptides with pathogenic action against only one group. *Multiple action* peptides with pathogenic action against two or more groups. (%): Percentage of hits/total peptides

From the 25 physicochemical properties used to design the SVM AVPpred algorithm [2], 21 are directly or indirectly (18/25 = 84 %) related to the polarity (Table 3, column Polarity with entries with figure ✔).

Polarity index method had over 43/60 = 71 % efficiency detecting the 60 validated and experimental antiviral peptides from Kumar and coworkers [2] (Table 4, entries 1–60 and Table 5 column AVPpred), and 1/1 = 100 % detecting the antiviral peptides from APD2 database (Table 4, entry 61 and Table 5, colum Virus). There are no coincidences between both excluding sets (Table 4, columns #1 and #2).

Discussion

When reviewing different databases of antimicrobial peptides, we detected a peptide with predominantly toxic action toward a pathogenic group; this allows us to assume that nature considers only small changes in the primary structure of the peptide to induce its possible pathogenic action. In that sense, the peptide linear structure plays an important role in the identification of its pathogenic action, when we use algorithms that use "training sets" with the desired profile. Although the physicochemical property called polarity is involved in most of the algorithms that predict anti-virus peptides, this method has an innovative aspect as it expresses the metric through a polarity matrix that includes 16 interactions. We see this matrix as a picture of the polar dynamics of the peptide. The polarity matrix clearly shows a pattern that led us to achieve an identification efficiency of 70 % on the AVPpred database. The same pattern also rejected other groups of peptides in APD2 database, with the exception of the anti-virus set. We assume that if we built the matrix with one digit, perhaps we did not have enough information to focus the method correctly.

We believe the effectiveness of the polarity index method in terms of the computing resources required makes it suitable candidate for a more detailed analysis related to the subdomains of peptides. In this regards, we have initiated a comprehensive classification of the APD2 database gathering, from published manuscripts, the toxicity values of antimicrobial peptides and thus explore peptide sub-domains with specific and very toxic pathogenic action. Our team is working on this as we consider it of vital importance to strongly support basic scientific research. We have also published work where the same method is used to understand the profile of the "first proteins" from 4 billion years ago. For this task, we are using clusters of GPU coprocessors, which will allow the analysis of 15 amino acids in length peptides.

Finally, within the antiviral peptide group, there are two sub-groups not approached in this work as they constitute by themselves an independent topic for the importance they

have and the degree of subject matter expertise required: the influenza A type H1N1 and HIV peptides. Given their potential to provoke a world pandemic, these two groups of peptides concentrate the efforts of several research groups, as they can undoubtedly become a problem of enormous proportions. Our team is now directing the method toward the identification of these two sub-groups.

Conclusions

In summary, we report an implementation of a polarity index method in the exhaustive prediction of antiviral peptides from AVPpred and APD2 databases, with high level of discriminative efficiency (44/61 = 72), from the reading of its linear sequence.

Availability

The test files and source code are given as "Supplementary Material"

Acknowledgments The authors acknowledge the Departamento de Cómputo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán for support, and the proof-reading of this manuscript by Concepción Celis Juárez.

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

References

1. Wang, G., Li, X., & Wang, Z. (2009). APD2: The updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Research, 37*, D933–D937.
2. Thakur, N., Qureshi, A., & Kumar, M. (40). AVPpred: Collection and prediction of highly effective antiviral peptides. *Nucleic Acids Research, 2012*, W199–W204.
3. Kawashima, S., & Kanehisa, M. (2000). AAindex: Amino acid index database. *Nucleic Acids Research, 28*, 374.
4. Haagmans, B. L., Andeweg, A. C., & Osterhaus, A. (2009). The application of genomics to emerging zoonotic viral diseases. *PLoS Pathogens, 2009*, 5.
5. De Clercq, E. (2004). Antivirals and antiviral strategies. *Nature Reviews Microbiology, 2*, 704–720.
6. Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., et al. (2009). Investigators, M.-T. vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine, 361*, 2209–2220.
7. Wedemeyer, H., Schuller, E., Schlaphoff, V., Stauber, R. E., Wiegand, J., Schiefke, I., et al. (2009). Therapeutic vaccine IC41 as late add-on to standard treatment in patients with chronic hepatitis C. *Vaccine, 27*, 5142–5151.
8. Dykxhoorn, D. M., & Lieberman, J. (2006). Silencing viral infection. *PLOS Medicine, 3*, e242.
9. Kricka, L. J., & Fortina, P. (2009). Analytical ancestry: "Firsts" in fluorescent labeling of nucleosides, nucleotides, and nucleic acids. *Clinical Chemistry, 55*, 670–683.
10. Bharti, A. C., Shukla, S., Mahata, S., Hedau, S., & Das, B. C. (2009). Anti-human papillomavirus therapeutics: Facts & future. *Indian Journal of Medical Research, 130*, 296–310.
11. McKeegan, K. S., Borges-Walmsley, M. I., & Walmsley, A. R. (2002). Microbial and viral drug resistance mechanisms. *Trends in Microbiology, 10*, S8–S14.
12. Polanco González, C., Nuño Maganda, M. A., Arias-Estrada, M., & del Rio, G. (2011). An FPGA implementation to detect selective cationic antibacterial peptides. *PLoS ONE, 6*(6), e21399.
13. Polanco, C., & Samaniego, J. L. (2009). Detection of selective cationic amphipathic antibacterial peptides by Hidden Markov models. *Acta Biochimica Polonica, 56*, 167–176.
14. Polanco, C., Samaniego, J. L., Buhse, T., Mosqueira, F. G., Negron-Mendoza, A., Ramos-Bernal, S., et al. (2012). Characterization of selective antibacterial peptides by polarity index. *International Journal of Peptide, 2012*, 585027.
15. del Rio, G., Castro-Obregon, S., Rao, R., Ellerby, H. M., & Bredesen, D. E. (2001). APAP, a sequence-pattern recognition approach identifies substance P as a potential apoptotic peptide. *FEBS Letters, 3*, 213–219.
16. Polanco, C., Buhse, T., Samaniego, J. L., & Castañón-González, J. A. (2013). A toy model of prebiotic peptide evolution: The possible role of relative amino acid abundances. *Acta Biochimica Polonica*, *60*, 175–182.
17. Polanco, C., Samaniego, J. L., Castañón-González, J. A., Buhse, T., Sordo, M. L. (2013). Characterization of a possible uptake mechanism of selective antibacterial peptides. *Acta Biochimica Polonica*, *60*, 629–633.
18. Beghin, F., & Dirkx, J. (1975). Proceedings: A simple statistical method to predict protein conformations. *International Archives of Physiology and Biochemistry, 83*, 167–168.
19. Bull, H. B., & Breese, K. (1974). Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Archives of Biochemistry and Biophysics, 161*, 665–670.
20. Charton, M. (1981). Protein folding and the genetic code: An alternative quantitative model. *Journal of Theoretical Biology, 1981*(91), 115–123.
21. Chou, P. Y., & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology, 47*, 145–148.
22. Cohn, E. J., & Edsall, J. T. (1943). *Proteins amino acids and peptides as ions and dipolar ions*. New York: Reinhold.
23. Fasman, G. D. (1976). *Handbook of biochemistry and molecular biology, section D: Physical and Chemical Data*. London: Taylor & Francis.
24. Fauchère, J. L., Charton, M., Kier, L. B., Verloop, A., & Pliska, V. (1988). Amino acid side chain parameters for correlation studies in biology and pharmacology. *International Journal of Peptide and Protein Research, 32*, 269–278.
25. Finkelstein, A. V., Ptitsyn, O. B., & Kozitsyn, S. A. (1977). Theory of protein molecule self-organization. II. A comparison of calculated thermodynamic parameters of local secondary structures with experiments. *Biopolymers, 16*, 497–524.
26. Finkelstein, A. V., Badretdinov, A. Y., & Ptitsyn, O. B. (1991). Physical reasons for secondary structure stability: Alpha-helices in short peptides. *Proteins, 10*, 287–299.
27. Geisow, M. J., & Roberts, R. D. (1980). Amino acid preferences for secondary structure vary with protein class. *International Journal of Biological Macromolecules, 2*, 387–389.
28. Karplus, P., & Schulz, G. (1985). Prediction of chain flexibility in proteins. A tool for the selection of peptide antigens. *Naturwissenschaften, 72*, 212–213.

29. Aurora, R., & Rose, G. D. (1998). Helix capping. *Protein Science, 7*, 21–38.
30. Avbelj, F. (2000). Amino acid conformational preferences and solvation of polar backbone atoms in peptides and proteins. *Journal of Molecular Biology, 300*, 1335–1359.
31. George, R. A., & Heringa, J. (2002). An analysis of protein domain linkers: Their classification and role in protein folding. *Protein Engineering, 15*, 871–879.
32. Kidera, A., Konishi, Y., Oka, M., Ooi, T., & Scheraga, A. (1985). Statistical analysis of the physical properties of the 20 naturally occuring amino acids. *Journal of Protein Chemistry, 4*, 23–55.
33. Guy, H. R. (1985). Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophysical Journal, 1985*(47), 61–70.
34. Casari, G., & Sippl, M. J. (1992). Structure-derived hydrophobic potential. Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. *Journal of Molecular Biology, 224*, 725–732.
35. Richardson, C., Hull, D., Greer, P., Hasel, K., Berkovich, A., Englund, G., et al. (1986). The nucleotide sequence of the mRNA encoding the fusion protein of measles virus (Edmonston strain): A comparison of fusion proteins from several different paramyxoviruses. *Virology, 155*, 508–523.
36. Collins, P. L., Huang, Y. T., & Wertz, G. W. (1984). Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus. *Proceedings of National Academy of Science USA, 1984*, 7683–7687.
37. Kestler, H. W, 3rd, Li, Y., Naidu, Y. M., Butler, C. V., Ochs, M. F., Jaenel, G., et al. (1988). Comparison of simian immunodeficiency virus isolates. *Nature, 1988*(331), 619–622.
38. Verschoor, E. J., Hulskotte, E. G., Ederveen, J., Koolen, M. J., Horzinek, M. C., & Rottier, P. J. (1993). Post-translational processing of the feline immunodeficiency virus envelope precursor protein. *Virology, 193*, 433–438.
39. Lohmann, V., Körner, F., Koch, J., Herian, U., Theilmann, L., & Bartenschlager, R. (1999). Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science, 285*, 110–113.
40. Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., et al. (1991). Genetic organization and diversity of the hepatitis C virus. *Proceedings of National Academy of Science USA, 88*, 2451–2455.
41. Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D., et al. (2004). Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *Journal of Biological Chemistry, 2004*(279), 40835–40843.
42. Chamberlain, R. W., Adams, N. J., Taylor, L. A., Simmonds, P., & Elliott, R. M. (1997). The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. *Biochemical and Biophysical Research Communications, 236*, 44–49.
43. Gould, A. R. (1996). Comparison of the deduced matrix and fusion protein sequences of equine morbillivirus with cognate genes of the Paramyxoviridae. *Virus Research, 1996*(43), 17–31.
44. Spriggs, M. K., Olmsted, R. A., Venkatesan, S., Coligan, J. E., & Collins, P. L. (1986). Fusion glycoprotein of human parainfluenza virus type 3: Nucleotide sequence of the gene, direct identification of the cleavage-activation site, and comparison with other paramyxoviruses. *Virology, 152*, 241–251.
45. Wang, H., & Ng, T. B. (2000). Ginkbilobin, a novel antifungal protein from Ginkgo biloba seeds with sequence similarity to embryo-abundant protein. *Biochemical and Biophysical Research Communications, 279*, 407–411.
46. Stevenson, M., Haggerty, S., Lamonica, C., Mann, A. M., Meier, C., & Wasiak, A. (1990). Cloning and characterization of human immunodeficiency virus type 1 variants diminished in the ability to induce syncytium-independent cytolysis. *Journal of Virology, 64*, 3792–3803.
47. Belanger, F. C., & Kriz, A. L. (1989). Molecular characterization of the major maize embryo globulin encoded by the glb1 gene. *Plant Physiology, 91*, 636–643.
48. Mulvenna, J. P., Sando, L., & Craik, D. J. (2005). Processing of a 22 kDa precursor protein to produce the circular protein tricyclon A. *Structure, 13*, 691–701.
49. Harris, J. D., Hibler, D. W., Fontenot, G. K., Hsu, K. T., Yurewicz, E. C., & Sacco, A. G. (1994). Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: The ZPA, ZPB and ZPC gene families. *DNA Sequence, 4*, 361–393.
50. Hauser, L. J., Land, M. L., Brown, S. D., Larimer, F., Keller, K. L., Rapp-Giles, B. J., et al. (2011). Complete genome sequence and updated annotation of *Desulfovibrio alaskensis* G20. *Journal of Bacteriology, 193*, 4268–4269.
51. Agerberth, B., Gunne, H., Odeberg, J., Kogner, P., Boman, H. G., & Gudmundsson, G. H. (1995). FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. *Proceedings of the National Academy of Sciences USA, 92*, 195–199.
52. Ireland, D. C., Colgrave, M. L., & Craik, D. J. (2006). A novel suite of cyclotides from *Viola odorata*: Sequence variation and the implications for structure, function and stability. *Biochemical Journal, 400*, 1–12.
53. Tang, Y. Q., Yuan, J., Osapay, G., Osapay, K., Tran, D., Miller, C. J., et al. (1999). A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated alpha-defensins. *Science, 286*, 498–502.
54. Wong, J. H., & Ng, T. B. (2005). Sesquin, a potent defensin-like antimicrobial peptide from ground beans with inhibitory activities toward tumor cells and HIV-1 reverse transcriptase. *Peptides, 26*, 1120–1126.
55. Kurachi, K., Chandra, T., Degen, S. J., White, T. T., Marchioro, T. L., Woo, S. L., et al. (1981). Cloning and sequence of cDNA coding for alpha 1-antitrypsin. *Proceedings of the National Academy of Sciences, 78*, 6826–6830.
56. Nishimura, Y., Miyazawa, T., Ikeda, Y., Izumiya, Y., Nakamura, K., Cai, J. S., et al. (1998). Molecular cloning and sequencing of feline stromal cell-derived factor-1 alpha and beta. *European Journal of Immunogenetics, 25*, 303–305.
57. Collman, R., Balliet, J. W., Gregory, S. A., Friedman, H., Kolson, D. L., Nathanson, N., et al. (1992). An infectious molecular clone of an unusual macrophage-tropic and highly cytopathic strain of human immunodeficiency virus type 1. *Journal of Virology, 66*, 7517–7521.
58. Taylor, G. R., Lagosky, P. A., Storms, R. K., & Haynes, R. H. (1987). Molecular characterization of the cell cycle-regulated thymidylate synthase gene of Saccharomyces cerevisiae. *Journal of Biological Chemistry, 262*, 5298–5307.
59. Church, D. M., Goodstadt, L., Hillier, L. W., Zody, M. C., Goldstein, S., She, X., et al. (2009). Mouse genome sequencing consortium. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLOS Biology, 7*, e1000112.
60. Ousley, A., Zafarullah, K., Chen, Y., Emerson, M., Hickman, L., & Sehgal, A. (1998). Conserved regions of the timeless (tim) clock gene in Drosophila analyzed through phylogenetic and functional studies. *Genetics, 148*, 815–825.
61. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., et al. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature, 313*, 277–284.
62. Ratner, L., Fisher, A., Jagodzinski, L. L., Mitsuya, H., Liou, R. S., Gallo, R. C., et al. (1987). Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Research and Human Retroviruses, 3*, 57–69.

63. Steinborner, S. T., Currie, G. J., Bowie, J. H., Wallace, J. C., & Tyler, M. J. (1998). New antibiotic caerin 1 peptides from the skin secretion of the Australian tree frog *Litoria chloris.* Comparison of the activities of the caerin 1 peptides from the genus *Litoria*. *Journal of Peptide Research, 51*, 121–126.
64. Reitz, M. S, Jr, Hall, L., Robert-Guroff, M., Lautenberger, J., Hahn, B. M., Shaw, G. M., et al. (1994). Viral variability and serum antibody response in a laboratory worker infected with HIV type 1 (HTLV type IIIB). *AIDS Research and Human Retroviruses, 10*, 1143–1155.
65. Goraya, J., Knoop, F. C., & Conlon, J. M. (1998). Ranatuerins: Antimicrobial peptides isolated from the skin of the American bullfrog, *Rana catesbeiana*. *Biochemical and Biophysical Research Communications, 250*, 589–592.
66. Rozek, T., Wegener, K. L., Bowie, J. H., Olver, I. N., Carver, J. A., Wallace, J. C., et al. (2000). The antibiotic and anticancer active aurein peptides from the Australian Bell Frogs *Litoria aurea* and *Litoria raniformis* the solution structure of aurein 1.2. *European Journal of Biochemistry, 267*, 5330–5341.
67. Tamas, I., Klasson, L., Canbäck, B., Näslund, A. K., Eriksson, A. S., Wernegreen, J. J., et al. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science, 296*, 2376–2379.
68. Lamberty, M., Zachary, D., Lanot, R., Bordereau, C., Robert, A., Hoffmann, J. A., et al. (2001). Insect immunity. Constitutive expression of a cysteine-rich antifungal and a linear antibacterial peptide in a termite insect. *Journal of Biological Chemistry, 276*, 4085–4092.
69. McGeoch, D. J., & Davison, A. J. (1986). DNA sequence of the herpes simplex virus type 1 gene encoding glycoprotein gH, and identification of homologues in the genomes of varicella-zoster virus and Epstein-Barr virus. *Nucleic Acids Research, 14*, 4281–4292.
70. Mushahwar, I. K., Erker, J. C., Muerhoff, A. S., Leary, T. P., Simons, J. N., Birkenmeyer, L. G., et al. (1999). Molecular and biophysical characterization of TT virus: Evidence for a new virus family infecting humans. *Proceedings of National Academy of Science USA, 96*, 3177–3182.
71. Giménez-Bonafé, P., Soler, F. M., Buesa, C., Sautière, P. E., Ausió, J., Kouach, M., et al. (2004). Chromatin organization during spermiogenesis in *Octopus vulgaris*. II: DNA-interacting proteins. *Molecular Reproduction and Development, 68*, 232–239.
72. Hammond, J., & Hammond, R. W. (2003). The complete nucleotide sequence of isolate BYMV-GDD of Bean yellow mosaic virus, and comparison to other potyviruses. *Archives of Virology, 148*, 2461–2470.
73. Yoshida, T., Miyagawa, K., Odagiri, H., Sakamoto, H., Little, P. F., Terada, M., et al. (1987). Genomic sequence of hst, a transforming gene encoding a protein homologous to fibroblast growth factors and the int-2-encoded protein. *Proceedings of National Academy of Science USA, 84*, 7305–7309.
74. Gombart, A. F., Blissard, G. W., & Rohrmann, G. F. (1989). Characterization of the genetic organization of the HindIII M region of the multicapsid nuclear polyhedrosis virus of *Orgyia pseudotsugata* reveals major differences among baculoviruses. *Journal of General Virology, 70*, 1815–1828.
75. Suerbaum, S., Josenhans, C., Sterzenbach, T., Drescher, B., Brandt, P., Bell, M., et al. (2003). The complete genome sequence of the carcinogenic bacterium Helicobacter hepaticus. *Proceedings National Academy of Science USA, 100*, 7901–7906.
76. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature, 393*, 537–544.
77. Collins, P. L., Huang, Y. T., & Wertz, G. W. (1984). Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus. *Proceedings of National Academy of Science USA, 81*, 7683–7687.
78. Terry, A. S., Poulter, L., Williams, D. H., Nutkins, J. C., Giovannini, M. G., Moore, C. H., et al. (1988). The cDNA sequence coding for prepro-PGS (prepro-magainins) and aspects of the processing of this prepro-polypeptide. *Journal of Biological Chemistry, 263*, 5745–5751.
79. Gil, R., Silva, F. J., Zientz, E., Delmotte, F., González-Candelas, F., Latorre, A., et al. (2003). The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proceedings of National Academy of Science USA, 100*, 9388–9393.
80. Gettner, S. N., Kenyon, C., & Reichardt, L. F. (1995). Characterization of beta pat-3 heterodimers, a family of essential integrin receptors in *C. elegans*. *Journal of Cell Biology, 129*, 1127–1141.