Regular paper

The polar profile of ancient proteins: a computational extrapolation from prebiotics to paleobiochemistry

Carlos Polanco[1✉], Thomas Buhse[2], Gloria Vizcaíno[3] and Jacobo Levy Picciotto[4]

Department of Mathematics, Faculty of Sciences, Universidad Nacional Autónoma de México, México City, 04510 México; Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, 62209 México; Department of Biochemistry, Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán", México City, 14000 México; Health Sciences Research Center, Faculty of Health Sciences, Universidad Anáhuac, Huixquilucan Edo. de México, 52786 México

This paper addresses the polar profile of ancient proteins using a comparative study of amino acids found in 25 000 000-year-old shells described in Abelson's work. We simulated the polar profile with a computer platform that represented an evolutionary computational toy model that mimicked the generation of small proteins starting from a pool of monomeric amino acids and that included several dynamic properties, such as self-replication and fragmentation-recombination of the proteins. The simulations were taken up to 15 generations and produced a considerable number of proteins of 25 amino acids in length. The computational model included the amino acids found in the ancient shells, the thermal degradation factor, and the relative abundance of the amino acids observed in the Miller-Urey experimental simulation of the prebiotic amino acid formation. We found that the amino acid polar profiles of the ancient shells and those simulated and extrapolated from the Miller-Urey abundances are coincident.

✉e-mail: polanco@unam.mx

INTRODUCTION

For decades, the analysis of functional and structural profiles of proteins has been an important research subject, and it is still a topic of utmost relevance (González-Díaz *et al.*, 2008; Maccari *et al.*, 2013). Understanding the origin of the protein structure and its function represents a fundamental contribution to science that, among other aspects, could lead to a deeper insight into the abiogenic evolution of proteins on the early Earth (Forslund & Sonnhammer, 2012), as well as to the development of new and more effective and less toxic drugs (Chen & Chen 2008; Fernald *et al.*, 2005). We think that the knowledge of the polar profile of proteins that prevailed in the ancient past constitutes a key element for understanding the specialization lines of the proteins we know today.

This paper was based on the experimental work of Abelson (Abelson, 1957; Abelson, 1959; Abelson, 1966), who evaluated the amino acid compositions of two groups of *Mercenaria mercenaria* fossil shells: one of the Miocene epoch (25 000 000 years ago) and one of recent age. Both groups of shells were subjected to similar ex-traction protocol to identify their amino acid contents by paper chromatography. The amino acids P, K, V, F, I, L, Y, A, T, G, E, S, and D were recovered from *recent shells*, whereas the amino acids A, G, E, L, P, I, and V were found in *fossil shells*. The conserved amino acids common to both shell groups were A, E, G, I, L, P, and V. Other fossils studied from the Ordovician period (430 000 000 years ago) and the late Pleistocene (5 000 000 years ago) showed the amino acids G, A, V, L, and I as the most conserved.

With the aforementioned information, we used an evolutionary computational toy model that mimics the generation of small proteins starting from a pool of amino acid monomers. The computational abstraction of Abelson's analysis was based on an approach that we had previously designed to perform an extrapolation of the computational Miller-Urey amino acid abundances (Miller, 1953; Polanco *et al.*, 2013), to simulate the amino acid profiles of so-called biological common ancestors found in *E. coli*, *M. jannaschii*, and *S. cerevisiae* (Delaye *et al.*, 2005). Our present computational approach intended to re-create possible elements of a prebiotic scenario and included a simulation of thermal amino acid degradation, which appears to be a relevant aspect in the findings of Abelson. The model design had a Markovian profile (Meyn & Tweedie, 2005) to enable the handling of several variables without increasing model complexity. This is particularly useful when the re-creation includes a large number of variables.

The molecular self-replication was also introduced into the model, as an essential attribute of life (Bag & von Kiedrowski, 1996; Issac *et al.*, 2001; Orgel, 1992; Reinhoudt *et al.*, 1996; Robertson *et al.*, 2000). Here, the molecular self-replication can be understood as the "autocatalysis by a reaction product which is able to recognize at least two individual reactants with a high degree of selectivity" (Bissette & Fletcher, 2013), like the association of a product with the reactants that leads to an acceleration of the product formation. Pioneering studies by the Ghadiri group (Lee *et al.*, 1996; Lee *et al.*, 1997; Severin *et al.*, 1997; 1998) showed that molecular self-replication can occur during amino acid peptide linkage in the case of short α-helical peptides.

Self-replication is basically the formation of a helical peptide (T) through the amide bond formation between two peptide fragments, where T acts as a template that assists the coupling of the two fragments forming a product that is identical to itself. Observed kinetic data indicates an autocatalytic pathway in which T, as a net effect, catalyzes its own formation. Such system could be

Table 1. Starting amino acids used in the simulations.

#	Amino acid	Symbol	Thermal degradation	µmol[a]	Miller-Urey equivalence[b]	Polarity[c]	Numeric equivalence[d]
1	Glycine	**G**		440.0	2	NP	3
2	Alanine	**A**		790.0	1	NP	4
3	Valine	**V**		19.5	41	NP	4
4	Leucine	**L**		11.3	71	NP	4
5	Isoleucine	**I**		4.8	166	NP	4
6	Proline	**P**		1.5	526	NP	4
7	Aspartic Acid	**D**	*	34.0	23	P-	2
8	Glutamic acid	**E**		7.7	102	P-	2
9	Serine	**S**	*	5.0	158	N	3
10	Threonine	**T**	*	0.8	987	N	3
11	Lysine	**K**	*	1.2	403	P+	1
12	Phenylalanine	**F**	*	0.3	2800	NP	4
13	Tyrosine	**Y**	*	0.3	2800	N	3

[a]yields from sparking CH_4 (336 mmoles), N_2, and H_2O with traces of NH_3 (based on the carbon added as CH_4). Glycine=0.26%; Alanine=0.71% total yield of amino acids in the table=1.90%; [b]amino acid amount in µmol compared to the Miller-Urey experiment; [c]classification of amino acids by their charge state: acidic hydrophilic (P-), basic hydrophilic (P-), neutral (N), and non-polar (NP); [d]representative numerical value for each amino acid according to polarity.

regarded as a possible model for peptide-based molecular evolution on the early Earth.

Regarding our simulation, protein segments were assumed to catalyze the formation of other protein segments with the same amino acid sequences. The so-called offspring replicas of protein segments were identical copies randomly generated from the proteins previously formed and they were sent to a so-called cutting record; i.e., they were transferred to a parallel program as a seed where they served for further protein formation. In this parallel program, the protein formation mechanism, as described above, took place using these fragments as starting material, instead of the amino acid monomers.

With these variables, the model generated a set of proteins and calculated its polar profile. This profile represented the number of polar incidents resulting from reading the linear representation of the protein (Polanco *et al.*, 2013). The tests compared the polar profile of the Miller experiment (Miller 1953; Polanco *et al.*, 2013), with the Abelson model, and included changes to the variables of polarity and abundance. We showed that both models (Abelson and Miller) have coincident polar profile representations.

MATERIALS AND METHODS

The computational model. Our computational model (see Availability Section for source code) simulates protein

formation. The simulations started from a monomer pool comprising the 13 different amino acids (P, K, V, F, I, L, Y, A, T, G, E, S, and D) that were identified in *recent* shells of *Mercenaria mercenaria*. It was assumed that the abundances of these amino acids, i.e. their relative concentrations, were equivalent to the abundances found in the Miller-Urey spark discharge experiments (Miller, 1953; Polanco *et al.*, 2013) (Table 1). For practical reasons the generated proteins were fixed at 25 amino acids in length as the protein length does not affect the polarity profile calculation (Polanco *et al.*, 2013). The specific length of 25 amino acids was chosen to compare the present simulations with those that we performed in our previous studies (Polanco *et al.*, 2014; 2014a).

The protein building and sequencing rules were based on the classification of the amino acids by their side chain into [P+] basic hydrophilic, [P-] acidic hydrophilic, [N] neutral, and [NP] non-polar residues. The simulations also included a thermal degradation process (Abelson, 1957; 1959) starting with the 13 amino acids P, K, V, F, I, L, Y, A, T, G, E, S, and D and successively decreasing the abundances of D, K, F, S, T, and Y to obtain the seven amino acids identified in the *fossil* shells, i.e. A, G, E, L, P, I, and V. The simulations were performed up to 15 protein generations, being a tractable limit in respect to computational cost, to determine the future trend of the polar profiles of the proteins.

Table 2. Charge state matrix.

	P+	P-	N	NP		P+	P-	N	NP
P+	0.01	0.79	0.15	0.05		99	21	85	95
P-	0.79	0.01	0.15	0.05	–	21	99	85	95
N	0.40	0.40	0.15	0.05		60	60	85	95
NP	0.40	0.40	0.15	0.05		60	60	85	95
	P+	**P-**	**N**	**NP**		**P+**	**P-**	**N**	**NP**

Quantitative representation of the polar interaction between amino acids, based on their charge state (Polanco *et al.*, 2013). Proposed values based on considerations of a previous work (Fig. 9, Mosqueira *et al.*, 2015).

Amino acid polymerization. For our computational model, polarity was used as the main selection criterion for an amino acid monomer to bond the growing protein chain. Polarity is a measure of the molecular bonding (Pauling, 1955) that seems to be an effective discriminant for the protein groups (Polanco *et al.*, 2015). In order to accept or reject the bond of an amino acid monomer, the model considered the polarity group of the monomer and the polarity group of the amino acid at the end of the protein chain (Table 2).

The corresponding stochastic polarity matrix P[i,j], based on the four-group [P+] [P-] [N] [NP] classification, was determined by the indexes (i,j) with the row/column relation {P+, P-, N, NP} leading to a total of 16 possible and weighted interactions. The given values expressing the probability of each of these interactions were represented in Table 2 and were chosen based on a previous study (Polanco *et al.*, 2013). The construction of the polarity matrix was based on the assumption that the polymerization of amino acids occurs partly based on the difference in their electrical charge and that it is possible to classify them into four groups.

The bonding of the amino acids was simulated by counting the number of incidents between the candidate amino acid monomer and the amino acid at the end of the protein. The bonding was accepted when the number of incidents reached the given threshold value given in Table 2, increasing the protein chain by one amino acid. For instance, if both the amino acid monomer and the amino acid at the end of the protein chain were [P+], the amino acid monomer was added in iteration 99. On the other hand, if the amino acid monomer was [P+] and the amino acid at the end of the protein was [P-], the amino acid monomer was added to the protein in iteration 21.

Protein splitting and merging. The protein splitting and merging mimicked the potential instabilities of the growing proteins and introduced into the modelling some aspects of dynamic combinatorial libraries that could be associated with the interaction between hydrolysis and condensation reactions in a variable prebiotic environment. In particular, protein splitting was simulated by randomly cutting the forming protein in two segments and sending one segment to a so-called cutting record. The cutting probability (Polanco *et al.*, 2013) of the protein was defined by:

$$C(L) = 1/e^{L} \tag{1}$$

where $e = 2.7183$ and $L =$ length of the protein. Hence, the cutting probability was assumed to change inversely to the protein length. The protein recombination was simulated by adding one segment of the cutting record to another protein segment in formation according to the polarity criteria outlined above for the protein-amino acid monomer interaction.

Autocatalysis. Autocatalysis or self-replication was simulated by taking a segment of the forming protein and using it as "seed" for a new protein building process. Both the selection of the segment and the protein formation were done at random. This procedure allowed the new generations to be formed from the seed and not from the scratch.

Thermal degradation temperature. In our simulation, the presence of amino acids D, K, F, S, T, and Y (Table 1, rows marked with (*)), was affected by the thermal degradation, according to the Abelson work (Abelson, 1957; Abelson, 1959), reducing the probability of these six amino acids to emerge by 8% each generation.

Polarity index method. The computational polarity index method (PIM) (Polanco *et al.*, 2013a), used to evaluate the simulation results, only took the linear representation of the proteins. This linear representation was composed of an orderly sequence of amino acids. The metric evaluated the polar interaction of amino acids by pairs from one end of the protein to the other (polarity/numeric equivalence, Table 1). This metric generated a matrix of polar incidents A [i,j] that represented all the polar possibilities with 16 polar interactions, i.e. (i, j) = {P +, P-, N, NP} x {P +, P-, N, NP}. When all proteins in matrix A [i,j] were registered, each (i, j) element in the matrix was divided by the *n* number of amino acids of the proteins, (1/n) A [i,j]. Finally, the 16 elements were geometrically represented as a smooth curve of relative frequencies. The smooth curve of relative frequencies allowed the algebraic interpretation of the information based on the location of the maximum, minimum and inflection points of the function.

Trial test. The simulated proteins were analyzed using the bioinformatics method PIM (Polanco *et al.*, 2013a) to obtain their polar profile. The polar profile of the proteins generated by the computational Abelson model was compared in three different ways (Polanco *et al.*, 2013): with polar bias (probability distribution from Table 2 turned on), without polar bias (probability distribution from Table 2 turned off), and with decreasing the abundance of the amino acid Alanine by 50% (probability distribution from Table 2 turned on). The graphs of the polar profiles were also compared to locate their maximum, minimum and inflection points of the function. The simulations were also performed including thermal
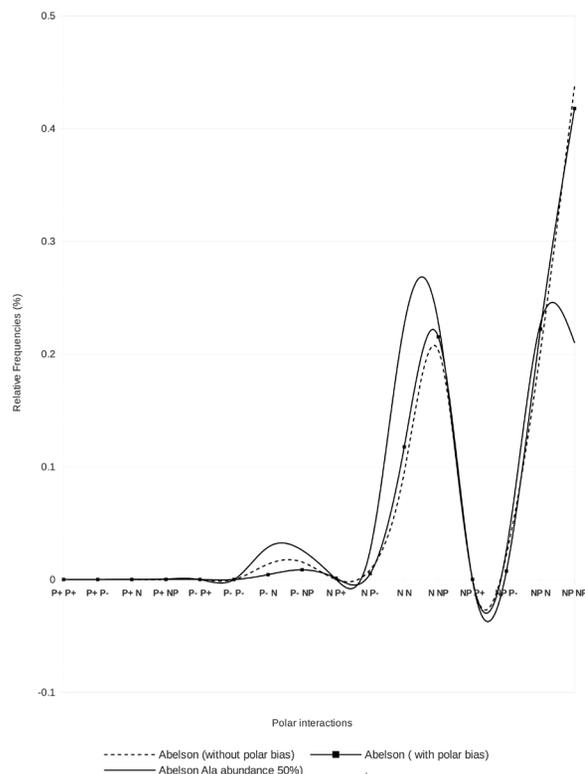

Figure 1. Polar profiles of the former Abelson model simulations (Polanco *et al.*, 2013) with and without polar bias and with variation in the Alanine abundance of 20%.
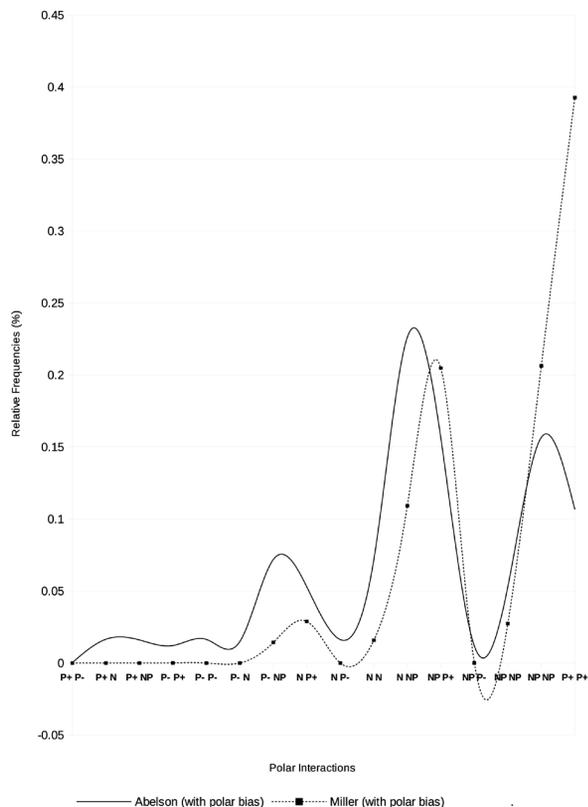The X-axis represents the 16 polar interactions.

Figure 2. Polar profile of the present Abelson model simulations compared with the polar profile of our former Miller model simulations (Polanco *et al.*, 2013) with polar bias.

Figure 3. Polar profile of the present Abelson model simulations with, and without thermal degradation bias.

degradation bias (turned on) and without thermal degradation bias (turned off) (Fig. 3).

Inflection points. The smooth curve of relative frequencies (Fig. 1) allowed the algebraic interpretation of the information based on the location of the maximum, minimum and inflection points of the function. An inflection point is a point in the domain of the function where the concavity in the graph changes. These points are important as they define the behavior of the function.

RESULTS

Figure 1 compares the profile of our model with three different polar biases: with polar bias, without polar bias, and with decreasing the abundance of the amino acid Alanine by 50% (see Trial test section). The corresponding graphs show a matching behavior, except in three polar interactions: [P-,N̲], [N,N̲], and [NP,N̲].. These three interactions shared the same "N" terminal group, which is affected by amino acids S, T, and Y (N polar group); the probability that these amino acids emerge was decreased by the thermal degradation bias. When the abundance of amino acid A (NP polar group) was decreased, the [NP, NP] interaction decreased substantially.

Figure 2 compares the polar profile of our simulation, and the simulation implemented for the amino acid abundance found in the Miller experiment (Polanco *et al.*, 2013). In both simulations, we kept the same polar bias (Table 2). In three polar interactions: [P +, NP], [P-, NP], and [NP, NP], the relative frequency was greater or equal to the Miller simulation. We think that the reason
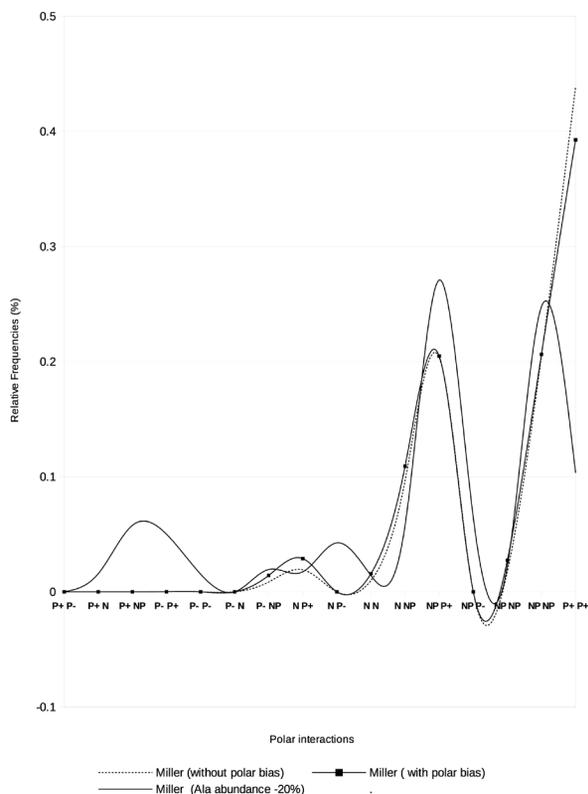
for these results is the fact that in the Miller simulation we included 10 non-proteinogenic amino acids that belong to the NP polar group.

Figure 3 compares the polarity profile of our model, with and without thermal degradation bias. In the simulation without thermal degradation bias, three of the four polar interactions [P-, N], [N, N], [NP, N], showed maximum points. These three polar interactions shared the same "N" terminal group, which was affected by the amino acids S, T, and Y (N polar group), whose probability of emergence decreased as a result of the thermal degradation bias.

DISCUSSION

In this paper, we simulated the evolutionary generation of small peptides taking into account an initial distribution of amino acid monomers as observed in the Miller-Urey electric discharge experiments of presumed amino acid formation under prebiotic conditions. Starting from this combination, the peptide sequences were successively built according to a simple criterion that refered to the amino acid charges expressed in a four-group classification: acidic hydrophilic, neutral, non-polar and basic hydrophilic amino acids. The polar interactions between these amino acid groups were weighted, resulting in amino acid sequences with limited randomness. Other simulation rules were introduced to account for the dynamic aspects of the peptide formation, namely the peptides were allowed to split and merge and the principles of peptide self-replication and thermal degradation were introduced. These dynamic rules were based on for-

mer experimental studies and permitted to render the model more realistic in terms of an evolutionary scenario.

Geometrically, this evidenced that the computational Abelson model eventually resembles the polar distribution found in our computational Miller model, due to the coincident location of the maximum, minimum and inflection points.

The thermal degradation temperature and the self-replication factors played an important role in this modeling because the thermal degradation process, as shown in Fig. 3, affected the abundance of the amino acids, leading to an alteration of the polar profile of the proteins generated by the simulation.

The self-replication was the "heritage" component from one generation of proteins to another. The model randomly formed a new generation of proteins from a fragment of the protein that is being built with the computational model i.e., the parent process sent a segment from one of its proteins as a template to the offspring process. Without taking this into account, the multifactorial nature of the polymerization phenomenon would decrease. In other words, abundance and polar interaction were intrinsic features of a protein, while "heritage" was related to its fragmentation, thermal degradation temperature, and self-replication. In our opinion, it is not possible to quantify the term "generation" as it is difficult to define the time-lapse from one generation to the next. If the polymerization process started 4 billion years ago, it must have been influenced by various factors, including cataclysms, whose complexity can only be partially re-created by computer simulations. However, we are certain that exploration of this complexity with a differential mathematical system, would not have led to definitive results; therefore, these simulations had to be based on Markovian models.

We believe that these results are worth to be reported because they *could* indicate a relationship between prebiotic amino acid abundances and early life on Earth, and in this case, the records of ancient life. We are aware that this observation does not support any deeper rationalization about the origin of life and requires further experimental and theoretical studies that are out of the scope of the present work. The time-lapse between the prebiotic world and the fossil records where life probably emerged 3.5 billion years ago is too large to be represented with a computer simulation in terms of a coarse-grained toy model. On the other hand, we believe that the striking similarity of the two polar profiles (for Miller-Urey sequences and Abelson's sequences) is not coincidental and therefore the further research on a possible reminiscence of the prebiotic world contained in protein amino acid sequences of earlier life appears rational.

Therefore, our development of computer simulations of prebiotic scenarios as in the present and past work (Polanco *et al.*, 2013; Polanco *et al.*, 2014; Polanco *et al.*, 2014a) could allow better understanding of the functional and structural profiles of today's proteins. A possible future direction in these studies would be to run these simulations more exhaustively to observe the limiting composition of the proteins. The question is if these implementations would result in a common profile. Another direction would be reviewing the most consolidated genes registered, identifying their proteins and comparing them to proteins produced by the present simulations.

CONCLUSIONS

Taking the polarity profile of the generated peptides as a reminiscence of a Miller/Urey-type of amino acid monomers, we compared this simulated polarity profile with that of proteins found in ancient shells. We found that the amino acid polar profiles of the ancient shells and those simulated and extrapolated based on the Miller-Urey abundances were coincident.

Availability

The source program can be requested from the corresponding author (polanco@unam.mx).

Software Resources

The model was written in FORTRAN 77 and executed on a Linux Fedora 23 Unix-type platform (GNU). The program run up to 15 generations on an HP Workstation Z210-CMT-4x Intel Xeon E3-1270/3.4 GHz (Quad-Core)-RAM 8GB-SSD 1x 160GB-DVD Super-Multi-Quadro 2000-Gigabit LAN, Linux Fedora 14, 64-bits. Cache Memory 8 MB. Cache Per Processor 8 MB. RAM 8 GB.

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. Computational algorithms: CP. Data analysis: CP, TB, and GV. Results discussion: CP, TB, and GV.

Acknowledgements

REFERENCES

Abelson PH (1966) Chemical events on the primitive earth. *Proc Natl Acad Sci USA* **55**: 1365–1372

Abelson PH (1959) Geochemistry and organic substances. *Res Geochem*, pp 79–103. Wiley

Abelson PH (1957) Some aspects of Paleobiochemistry. *Ann N Y Acad Sci* **69** 276–285

Bag BG, von Kiedrowski G (1996) Templates autocatalysis and molecular replication. *Pure Appl Chem* **68**: 2145–2152

Bissette AJ, Fletcher SP (2013) Mechanisms of autocatalysis. *Angew Chem Int* **52**: 2–30 doi: 10.1002/anie.201303822

Chen YP, Chen F (2008) Identifying targets for drug discovery using bioinformatics. *Expert Opin Ther Targets* **12**: 383–389. doi: 10.1517/14728222. 12.4.383.

Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* **35**: 537–554. doi: 10.1007/s11084-005-5760-3

Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB (2011) Bioinformatics challenges for personalized medicine. *Bioinformatics* **27**: 1741–1748. doi: 10.1093/bioinformatics/btr295

Forslund K, Sonnhammer EL (2012) Evolution of protein domain architectures. *Methods Mol Biol* **856**: 187–216. doi: 10.1007/978-1-61779-585-5_8

González-Díaz H, Prado-Prado F, Ubeira FM (2008) Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach. *Curr Top Med Chem* **8**: 1676–1690

Issac R, Ham YW, Chmielewski J (2001) The design of self-replicating helical peptides. *Curr Opin Struct Biol* **11**: 458–463

Lee DH, Severin K, Yokobayashi Y, Ghadiri MR (1997) Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390**: 591–594. doi: 10.1038/37569

Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. *Nature* **382**: 525–528. doi: 10.1038/382525a0

Maccari G, Di Luca M, Nifosí R, Cardarelli F, Signore G, Boccardi C, Bifone A (2013) Antimicrobial peptides design by evolutionary multiobjective optimization. *PLoS Comput Biol* **9**: e1003212. doi: 10.1371/journal. pcbi.1003212

Meyn SP, Tweedie RL (2005) *Markov Chains and Stochastic Stability*. 622 pp. University Press Cambridge

Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* **117**: 528–529. doi: 10.1126/science.117.3046.528

Mosqueira FG, Negrón-Mendoza A, Ramos-Bernal S. (2015) An interpretation of the oligomerization of amino acids under prebiotic conditions. *Boletín la Sociedad Geológica Mexicana* **67**: 421–430

Orgel LE (1992) Molecular replication. *Nature* **358**: 203–209. doi: 10.1038/358203a0

Pauling L (1955) *General Chemistry* [Química General] 3rd edn. W. H. Freeman & Company Publishers pp 227621

Polanco C, Buhse T, Samaniego JL, Castañón-González JA (2013) A toy model of prebiotic peptide evolution: the possible role of relative amino acid abundances. *Acta Biochim Pol* **60**: 175–182

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

Polanco C, Buhse T, Samaniego JL, Castañón-González JA, Estrada MA (2014) Computational model of abiogenic amino acid condensation to obtain a polar amino acid profile. *Acta Biochim Pol* **61**: 253–258

Polanco C, Samaniego JL, Buhse T, Castañón-González JA (2014a) Discrete dynamic system oriented on the formation of prebiotic dipeptides from Rode's experiment. *Acta Biochim Pol* **61**: 717–726

Polanco C, Samaniego JL, Uversky VN, Castañón-González JA, Buhse T, Leopold-Sordo M, Madero-Arteaga A, Morales-Reyes A, Tavera-Sierra L, González-Bernal JA, Arias-Estrada M (2015) Identification of proteins associated with amyloidosis by polarity index method. *Acta Biochim Pol* **62**: 41–55

Reinhoudt DN, Rudkevich DM, de Jong F (1996) Kinetic analysis of the Rebek self-replicating system: is there a controversy? *J Am Chem Soc* **118**: 6880–6889. doi: 10.1021/ja960324g

Robertson A, Sinclair AJ, Philp D (2000) Minimal self-replicating systems. *Chem Soc Rev* **29**: 141–152. doi: 10.1039/A803602K

Severin K, Lee DH, Yokobayashi Y, Ghadiri MR (1998) Dynamic error correction in autocatalytic peptide networks Angew. *Chem Int Ed Engl* **37**: 126–128

Severin K, Lee DH, Martinez JA, Dhadiri MR (1997) Peptide self-replication *via* template-directed ligation. *Chem Eur J* **3**: 1017–1024. doi: 10.1002/chem.19970030706

Zhou XX, Wang YB, Pan YJ, Li WF (2008) Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* **34**: 25–33. doi: 10.1007/s00726-007-0589-x