ACTA ABP BIOCHIMICA POLONICA

Vol. 60, No 2/2013
175–182
on-line at: www.actabp.pl

Regular paper

A toy model of prebiotic peptide evolution: the possible role of relative amino acid abundances

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

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

This paper presents a mathematical-computational toy model based on the assumed dynamic principles of prebiotic peptide evolution. Starting from a pool of amino acid monomers, the model describes in a generalized manner the generation of peptides and their sequential information. The model integrates the intrinsic and dynamic key elements of the initiation of biopolymerization, such as the relative amino acid abundances and polarities, as well as the oligomer reversibility, i.e. fragmentation and recombination, and peptide self-replication. Our modeling results suggest that the relative amino acid abundances, as indicated by Miller-Urey type electric discharge experiments, played a principal role in the early sequential information of peptide profiles. Moreover, the computed profiles display an astonishing similarity to peptide profiles observed in so-called biological common ancestors found in the following three microorganisms; *E. coli*, *M. jannaschii*, and *S. cereviasiae*. The prebiotic peptide fingerprint was obtained by the so-called polarity index method that was earlier reported as a tool for the identification of cationic amphipathic antibacterial short peptides.

Key words: amino acids, prebiotic polymerization, peptides, toy model, computer simulation, polarity index method.

Received: 27 Decemeber, 2012; revised: 25 April, 2013; accepted: 14 May, 2013; available on-line: 06 June, 2013

INTRODUCTION

This paper addresses the question of how primitive information transfer may have developed during the process of amino acid polymerization in the prebiotic world. Taking a minimalist view of such a scenario, one may sketch a starting point consisting of a mixture of relevant amino acid monomers, as well as an environment enabling chemical polymerization reactions between them with a possible support of the catalytic matter. Various evolution scenarios based on these elements have been proposed (Miller, 1953; Fox *et al.*, 1970), among them the suggestion of a sufficiently efficient co-evolution between nucleotides and amino acids to form the "first" peptides (Lambert, 2008).

If we consider the "first peptide" formation not only as an initiation of biopolymerization on the journey to the emergence of life, but also as potential information carriers, the question arises whether the sequential information found in *biotic* ancestors of our present life could be — at least partially — traced back to *prebiotic* evolution scenarios as described above.

Hence, we will focus on the dynamics of sequential information generation starting at zero in a world composed of amino acid monomers that can be assumed prototypically to have evolved under conditions suggested by the pioneering Miller experiment (Miller, 1953). We present a mathematical model that describes the evolution of peptide sequences from such a pool of amino acids taking into account a number of intrinsic and dynamic evolution rules.

The intrinsic imperatives concern, on one hand, the composition of various amino acids in the monomer pool that we assume to be related to their relative abundances according to Miller-type experiments and, on the other hand, the nature of the amino acid — amino acid interaction during polymerization that we consider to be governed by their polarities, i.e. by differences in the reaction probability between interactions of, for instance, equally charged *vs.* oppositely charged amino acids, polar/polar *vs.* polar/non-polar encounters, and so forth.

The dynamic rules of the model consist of the successive growth of the peptide chain with a certain clock frequency. We allow already formed peptides to break apart into smaller fragments, as well as to recombine or combine with other fragments or with monomers of the pool, which could be associated to the interplay between hydrolysis and condensation reactions in a variable prebiotic environment (Lahav & White, 1980).

Such kind of reversibility results not only in a growing diversity of the peptide population, but also in a higher dynamic flexibility because the fragments carry in their memory relative abundances of their building blocks, and have to obey the polarity rules when reacting with other fragments or monomers. The continuous breaking and merging of the peptide strains could lead to a higher selectivity in respect to the more probable amino acid sequences in the polymers, introducing the evolutionary terms of so-called dynamic combinatorial libraries (Cousins *et al.*, 2000) into the model.

Additionally, we included into the model the basic principles of template-directed peptide self-replication (Issac *et al.*, 2001; Paul & Joyce, 2004). This element of copying peptide fragments from a mother peptide that are subsequently subjected as daughter peptides to the evolutionary modeling process, leads to an acceleration in the peptide formation and can be regarded from a more general but also debated viewpoint as a possible

✉e-mail: polanco@unam.mx

Abbreviations: KEGG: Kyoto Encyclopedia of Genes and Genomes (ftp://ftp.genome.ad.jp/pub/kegg/) (Kanehisa *et al.*, 2000) accessed in December 2012

Table 1. Amino acids used for the simulations with their corresponding polarity groups.
Protein amino acids are symbolized with a letter and non-protein amino acids with a number.

#	Amino acid	Symbol	PRO	SPK	FTT	MET	Abundance	Polarity
1	Glycine	G	X	++++	X	++++	440.0	NP
2	Alanine	A	X	++++	X	+++	790.0	NP
3	Valine	V	X	++	?	+++	19.5	NP
4	Leucine	L	X	++			11.3	NP
5	Isoleucine	I	X	++	?		4.8	NP
6	Proline	P	X	+	?	+++	1.5	NP
7	Aspartic Acid	D	X	++	X	+++	34.0	P-
8	Glutamic acid	E	X	++	X	+++	7.7	P-
9	Serine	S	X	++			5.0	N
10	Threonine	T	X	+			0.8	N
11	Lysine	K	X				1.2	P+
12	Sarcocine	1					55.0	NP
13	N-Ethylglycine	2					33.0	NP
14	N-Methylalanine	3					15.0	NP
15	β-Alanine	4		++	X	+++	18.8	NP
16	β-Amino-η-butyric acid	5		+		++	0.3	NP
17	β-Aminoisobutyric acid	6		+	?	++	0.3	NP
18	γ -Aminobutyric acid	7					2.4	NP
19	Norvaline	8		+++	X	+++	61.0	NP
20	α-Amino-n-Butyric acid	9		++	?	+++	270.0	NP
21	acid α-Aminoisobutyric	0		++	?	+++	30.0	NP

Source: Number and correspondent polarity group. Protein amino acids can be identified as they are symbolized with a letter while non-protein amino acids are symbolized with a number. **PRO**: Found in proteins, **SPK**: Found in spark syntheses, **FTT**: Found in Fischer-Tropsch type syntheses, **MET**: Found in the Murchison meteorite, (100= ++++, 10=+++, 1=++, 0.1=+, < 0.01=? Tentative identification). Abundance: (in μmol) Yields from Sparking CH_4 (336 mmoles), N_2, and H_2O with traces of NH_3 (yield based on the carbon added as CH_4). Glycine = 0.26%; Alanine = 0.71% total yield of amino acids in the table = 1.90%. Polarity: Type of polarity: Polar amino acids with positive charge (P+), Polar amino acids with negative charge (P-), Neutral (N) and Non Polar (NP). Amino acids in blue are considered in this work as alpha-amino-acids.

support for a peptide-based molecular evolution on the early Earth. From the dynamic perspective, peptide self-replication stands for a positive feedback that is assumed to lead, particularly in combination with the reversibility notation, to non-intuitive behavior (Dadon *et al.*, 2008; Oprea *et al.*, 2007), such as perhaps dynamic error correction (Hopfield, 1974) in the course of the sequential information transfer or, in general, to complex dynamic phenomena like bifurcations or catastrophic events during the peptide evolution.

Models that pretend to shed light on prebiotic scenarios usually share the difficulty to be corroborated by experimental data. Our aim was to evaluate if the prebiotic world could have generated sequential peptide characteristics that were transferred to the first living systems. In order to compare the modeling results (prebiotic world) with available sequential records of so-called common ancestors (biological world), we made reference to our former studies on selective antibacterial short peptides, where we developed the polarity index method (Polanco *et al.*, 2012) that delivers a specific fingerprint of the peptides in form of a polarity profile, which was proved as an efficient classification tool. Defining this common point of classification, we will show that the polarity profile of our simulated peptide sequences and that of the indicated biological records display an astonishing similarity.

THE MODEL

General aspects

The computational-mathematical model builds peptides in a linear format. It considers a group of 21 amino acids (Table 1), where only 11 of them (G, A, V, L, I, P, D, E, S, T, K) are nowadays identified as basic amino acids, while the others (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) were classified as prebiotic amino acids.

As outlined in Fig. 1, the formation of peptides starts with a weighted random generation of the amino acid monomers in accordance with amino acid abundances, orientations, lateral chains, and polarities, to subsequently form the peptide chain.

Then, the model comprises the growth of the peptide chain, its breaking, re-combination, and self-replication. Breaking refers to the splitting of the peptide bond and binding of the remaining segments to other amino acid oligomers that are simultaneously built. Self-replication multiplies the evolutionary process through the heritage of segments from the same peptide into processes that generate offspring. This way, the model simulates the peptide building by taking into account specific features of the amino acid polymerization as well as the interaction with a hypothetical aqueous-lipid medium.

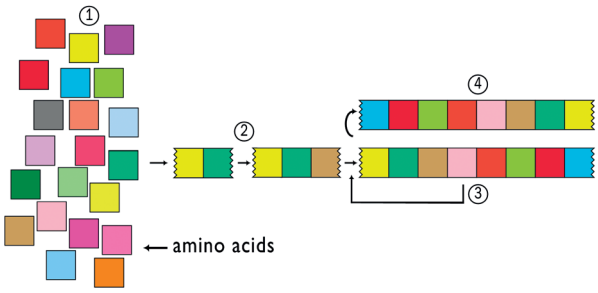

Figure 1. Sketch of the simulated peptide evolution comprising a weighted random generation of amino acid monomers (1), successive growth of the peptide chain (2), peptide fragmentation and recombination (3) and, template-directed self-replication (4).

Amino acid polymerization

The first amino acid monomer that initiates the peptide chain is chosen randomly from the 21 amino acids presented in Table 1. New amino acids are successively added on one side of the oligomer. Since all amino acids exhibit an electrical charge configuration, the model considers the amino acid polarities as selection criteria for an amino acid monomer to bind or not to bind to the peptide chain. In particular, for the acceptation or rejection of an amino acid, the model takes into account its polarity group according to a classification given in Table 2 that is not unique (Australian Naturophathic Network, 2012). By means of this classification, we take advantage of previous works related to the prebiotic molecular outline of peptides through Markov processes (Polanco & Samaniego, 2009). In these, the polarity classification is set in four groups [P-] polar, [N] neutral, [NP] non-polar, and [P+] basic hydrophilic. Thus, the stochastic polarity matrix $\mathbf{P}[i,j]$ is determined by the indexes $[i,j]$ corresponding to a specific row/column {P+, P-, N, NP}. The corresponding interaction values are stated in Table 2. The polarity matrix was outlined in a previous work (Mosqueira *et al.*, 2012) and anticipated for the possible polymerization of amino acids under anhydrous conditions in a prebiotic Earth environment.

The model simulates the polarity interaction among amino acids by counting the number of times at which this interaction occurs over the whole simulation process. The criterion to accept the binding of an amino acid is given when the interaction between the candidate amino acid from the monomer pool and the terminal amino acid of the peptide chain of a certain polarity group reaches the figure given in Table 2. For instance, if the interaction between two amino acids is [**D**] – [**S**], the corresponding polarity value is located in row 2 column 3 of Table 2, $[i,j] = (2,3)$. Therefore, the polarity interaction value [P-][N] is 0.15, which is equivalent to 85. The model recreates this interaction by adding up one unit to the interaction counter. Once the indicated value is reached, the amino acid binding [**D**] – [**S**] will be accepted.

The model takes into account the absolute amino acid abundances as a bias for the binding interaction between the candidate and terminal amino acid. For this purpose, the model uses again a counter that adds up the number of times of potential interactions (Table 3). Like in the previous case, unless this number is not reached, the binding will be rejected. For example, if the terminal amino acid of the peptide chain is **P** and the candidate amino acid to be bound is **T**, the interaction is only accepted when the abundance counter reaches the value 987, which is the number corresponding to the amino acid **T**.

The general structure of an amino acid is given by the presence of a central carbon attached to the carboxylic group, amino group, hydrogen, and the side chain. Those amino acids that are differentiated by their amino groups ($-NH_2$) located one, two, or three atoms in proximity to the carboxylic group (-COOH), are called alpha-, beta-, and gamma-amino acids, respectively (Herrera, 1993). The model establishes the absolute abundance of the beta- and gamma-amino acids 0, 4, 5, 6, and 9 at 15% (see Table 1), thereby encouraging these elements to be part of the simulated peptides.

Peptide splitting and merging

The processes of splitting and merging of the peptides, possibly due to the presence of a variable aqueous-lipid environment, constitute the dynamic aspects of the model. It comprises splitting of the peptide in construction by cutting it at a random position and adding the segment to a so-called cutting record. The splitting probability of each peptide is defined by the function $C(L)$ that decreases inversely to the length of the peptide,

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

where, $e = 2.7183$, and L is the peptide length at any particular moment.

After the random splitting, the model keeps one part of the peptide and transfers the other part to a cutting record, in which all these segments accumulate. All peptides can split again where the corresponding part is added to the same cutting record (Fig. 2a). For instance, if the peptide GVVLAAASE20T was constructed and then it splits in the position 9, the segment GVVLAAASE will be transferred to the cutting record and the construction peptide becomes 20T that is now 3 amino acids long.

The second aspect of reversibility is peptide recombination. It increases the size of the peptide in construction by adding segments from the cutting record. A peptide that was split can access the cutting record

Table 2. Stochastic matrix interaction in the polarity groups differentiated by their lateral chain: [P-] Polar, [N] Neutral, [P+] Basic hydrophilic, and [NP] Non-polar residues.

	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

Table 3. Absolute amino acid abundances.

#	Amino acid	Symbol	Abundance	Number of times
1	Glycine	G	440.0	2
2	Alanine	A	790.0	1
3	Valine	V	19.5	41
4	Leucine	L	11.3	71
5	Isoleucine	I	4.8	166
6	Proline	P	1.5	526
7	Aspartic Acid	D	34.0	23
8	Glutamic acid	E	7.7	102
9	Serine	S	5.0	158
10	Threonine	T	0.8	987
11	Lysine	K	1.2	403
12	Sarcocine	1	55.0	14
13	N-Ethylglycine	2	33.0	26
14	N-Methylalanine	3	15.0	52
15	β-Alanine	4	18.8	42
16	β-Amino-η-butyric acid	5	0.3	2653
17	β-Aminoisobutyric acid	6	0.3	2653
18	γ -Aminobutyric acid	7	2.4	329
19	Norvaline	8	61.0	13
20	α-Amino-n-Butyric acid	9	270.0	3
21	acid α-Aminoisobutyric	0	30.0	24

Abundance: (in μmol). Yields from sparking CH4 (336 mmol), N2, and H2O with traces of NH3 (yield based on the carbon added as CH4. Glycine = 0.26%; Alanine = 0.71% total yield of amino acids in the table = 1.90%. **Number of times**: modeling equivalence with respect to the amino acid amount in μmol (see text).

anytime, and can add one of these segments. However, this proceeds only as long as the amino acid at the right side of the segment in the cutting record and the left side of the amino acid from the peptide in construction match the amino acid polymerization criteria described above. In case these criteria are met, the segment is withdrawn from the cutting record and added to the left-side ending of the peptide in construction (Fig. 2b).

For example, if the model has constructed the peptide LTKSAGVA7, it searches for a segment in the cutting record between for instance GTK, 0AGAASV, and 4CS56G. Each of these segments will be read and analyzed forward and backwards consecutively. Then, the model verifies the amino acid interaction of each segment in the cutting record with the peptide in construction GTK-LTKSAGVA7, GTK-0AGAASV, and GTK-0AGAASV. In case the first segment is approved, according to the selection criteria, it will be added to the left side of the peptide in construction. If the first segment does not satisfy the criteria mentioned above, the next segment will be chosen and analyzed accordingly. If all segments are rejected or the cutting record has no elements, the polymerization procedure restarts by adding one amino acid monomer at the time.

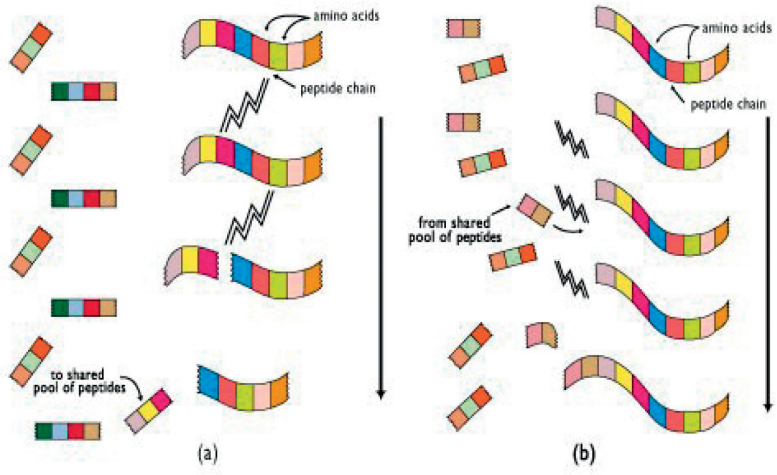

Figure 2. Illustration of peptide fragmentation and recombination.
(**a**) After random splitting, peptide fragments are added to a shared pool (cutting record), while the other part remains subject to further chain growth; (**b**) Peptides under construction combine with fragments of the shared pool and increase their chain lengths.

Peptide self-replication

Self-replication represents the evolutionary aspect of the modeling and signifies a very high demand for computational resources. As illustrat-

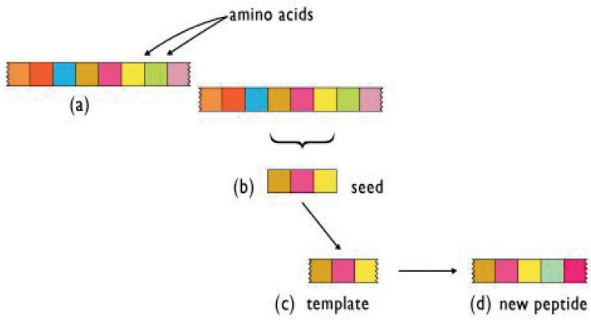

Figure 3. Peptide self-replication as considered in the model. Starting with a mother peptide (**a**), oligomer segments (**b**) are randomly created that are then subjected to the amino acid polymerization mechanism (**c–d**).

ed in Fig. 3, replicates of oligomer segments are created and are then subjected to the amino acid polymerization mechanism. The replicates are generated by randomly selecting a segment from the "mother" peptide in construction. Both the "mother" segment and its replicate share the cutting record. For instance, if the peptide A432PNDCEG34GSLD6LKEEPS77YV is constructed, a segment copy is generated, e.g. G34GSLD6LKE, and transferred as a seed to a parallel program. Then in this parallel program the peptide building mechanism starts with the seed G34GSLD6LKE, and not with the amino acid monomers.

The introduced segment replication mimics the self-replication of peptides. As a consequence, the sequential information of a "mother" peptide that has evolved under certain building criteria, such as amino acid abundances and polarity interactions, is recycled and provides a memory of these criteria for the newly formed oligomers. Such mechanism could lead to an evolutionary advantage of certain amino acid sequences through amplification.

Prebiotic code composition

Any comparison between the sequential information generated by the simulation and that of experimental data, such as the sequential records of the so-called common ancestors, tackles with non-equivalence amongst the amino acids. Today's peptides do not exhibit the prebiotic code composition anymore. It is unknown so far, how the 21 constitutive amino acids from the late prebiotic world evolved to the 20 amino acids related to the genetic code of the present times. Taking into account this non-equivalence, a conversion (Table 4) was used to match the corresponding polarity groups of the prebiotic and biotic amino acids.

To illustrate the use of this conversion, we need to consider the peptide sequence G34GSLD6LKE, where the amino acid G is assigned to the polarity equivalence 3, the amino acid 3 to the equivalence 4, etc., until the complete sequence is translated to 34433424412. Once the peptide sequence has been converted into its equivalent series, the polarity matrix $\mathbf{P}[i,j]$ is built. Each element $[i,j]$ accumulates the occurrence that is obtained by reading the series 34433424412 from left to right, and taking as element $[i,j]$ the pair of numbers found by moving one digit at the time through the series. As an

Table 4. Amino acid polarity group classification.

#	Amino acid	Symbol	Polarity	Numeric equivalence
1	Glycine	G	NP	4
2	Alanine	A	NP	4
3	Valine	V	NP	4
4	Leucine	L	NP	4
5	Isoleucine	I	NP	4
6	Proline	P	NP	4
7	Aspartic Acid	D	P-	2
8	Glutamic acid	E	P-	2
9	Serine	S	N	3
10	Threonine	T	N	3
11	Lysine	K	P+	1
12	Sarcocine	1	NP	4
13	N-Ethylglycine	2	NP	4
14	N-Methylalanine	3	NP	4
15	β-Alanine	4	NP	4
16	β-Amino-η-butyric acid	5	NP	4
17	β-Aminoisobutyric acid	6	NP	4
18	γ -Aminobutyric acid	7	NP	4
19	Norvaline	8	NP	4
20	α-Amino-n-Butyric acid	9	NP	4
21	acid α-Aminoisobutyric	0	NP	4

Source: Number and correspondent polarity group. Protein amino acids can be identified as they are symbolized with a letter, while non-protein amino acids are symbolized with a number. **Polarity**: Type of polarity: Polar amino acids with positive charge (P+), Polar amino acids with negative charge (P-), Neutral (N) and Non Polar (NP). **Numeric equivalence**: representative numeric value for each amino acid according to its polarity.

Table 5. Polarity matrix P[*i,j*] built by adding the [*i,j*] sorted pair occurrences formed when reading the series (34433424412) from the left to the right one digit at a time.

P+ (1)	P- (2)	N (3)	NP (4)
0	1	0	0
0	0	0	1
0	0	1	2
1	1	1	2

example, the polarity matrix **P**[*i,j*] corresponding to the series 34433424412 is expressed in Table 5.

When the polarity matrix **P**[*i,j*] is concluded, it will be normalized to one and stated in its linear form to express the relative frequencies of the 16 elements in 16 columns giving rise to a normalized relative frequency histogram.

Since there is no available data of prebiotic amino acid sequences, we made reference to best preserved genes from the following three microorganisms: *E. coli*, *M. jannaschii,* and *S. cereviasiae,* considered as the so-called common ancestors of approximately 2.5 billion years ago (Delaye *et al.*, 2005). This set was used to compare the peptide tendency generated by our simulations.

Polarity index method

The polarity index method (Polanco *et al.*, 2012) uses the polarity matrix **P**[*i,j*] of the peptides generated by the model simulations. By reading each row from left to right, it searches the positions where maximum (M_i) and minimum (m_i) incidences occur in the matrix. These are grouped into two sets. For instance, the polarity matrix of the peptide G34GSLD6LKE, as given in Table 5, shows the maximum positions of incidences for each row at $\{M_1, M_2, M_3, M_4\} = \{2, 8, 12, 16\}$ and the minimum positions at $\{m_1, m_2, m_3, m_4\} = \{1, 5, 9, 13\}$. After grouping them together, we obtain $\{M_1, M_2, M_3, M_4\} - \{m_1, m_2, m_3, m_4\} = \{2, 8, 12, 16\} - \{1, 5, 9, 13\}$.

RESULTS AND DISCUSSION

Simulations of the first six peptide generations showed that the obtained polarity index corresponded to {4, 9, 12, 16} – {1, 6, 10, 14}. However, as shown in Fig. 4, a successive decrease of the maximum for the polarity interaction [N] – [P+] in position 9 of the polarity matrix that goes along with a corresponding growth in the maximum of the interaction [P-] – [NP] in position 8 can be observed. This tendency was confirmed by additional simulations reaching 15 peptide generations. Hence, it can be assumed that the maximum in the polarity profiles of higher-order generations switches from the interaction 9 to 8 when the number of generations *n* tends to infinite. Accordingly, for a higher-order generation that could reflect a prebiotic scenario the polarity index method should result in {4, 8, 12, 16} – {1, 6, 10, 14}.

As a demonstration of the main features of our simulations, Fig. 5 shows the obtained relative frequency histogram of prebiotic peptides by simulation in comparison to that of best-preserved genes from *E. coli*, *M. jannaschii,* and *S. cereviasiae* reported by Delaye *et al.* (Delaye *et al.*, 2005). The data set corresponds to 20 completely sequenced cellular genomes from eubacteria, archaebacteria, and eukaryotic nucleocytoplasm that are now referred to as new taxonomic categories as *Bacteria*, *Archaea,* and *Eucarya*. The data was obtained by twice one-way BLAST searches in order to define the set of the most conserved proteins encoding sequences to characterize the gene complement of the last common ancestor of extant life. Our simulation results show an astonishingly good agreement with this reported data.

The polarity index identified for common ancestors is {4, 8, 12, 16} – {1, 6, 10, 14}. The difference between this data and the simulation results for the occurrence frequency from 16 polarity interactions did not exceed 3% in 14 out of 16 polarity interactions. The only exceptions were the interactions 8 and 9 that can be understood by the already described successive maximum shift between the polarity matrix positions 8 and 9. We also observed that the positions 4, 8, and 12 of the polarity matrix correspond to sudden changes in the tendencies of the polarity graphic. This means that if the tendency was ascendant before any of these positions, it became descendant after them and *vice versa*. Such effect did not occur in any other position of the polarity matrix where the before-after tendencies remain the same. Those points located in the polarity matrix can be characterized as catastrophic bifurcation points.

To evaluate the robustness of the simulations, a number of variations in the initial conditions and parameter values were conducted. In particular, we tested variations in the amino acid abundances. These refer to the initial conditions of the simulations taking into account a possible abundance divergence in the prebiotic world. Changing the preference for the amino acid side chain by setting the absolute abundances of these amino acids to different values between 3% and 25% altered the polarity profile of less than 3% in 16 interactions. Moving glycine (G), which is placed second in the absolute

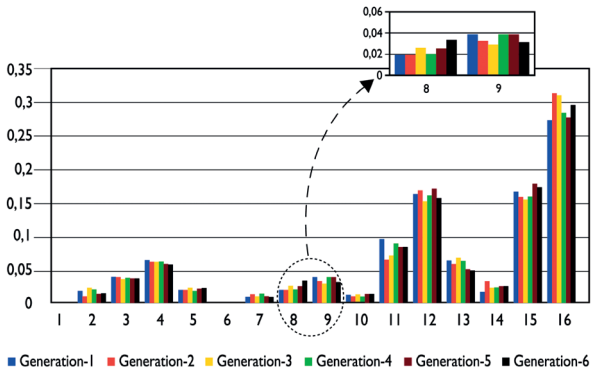

Figure 4. Comparison of polarity profiles between the first and sixth protein generation obtained by simulation.

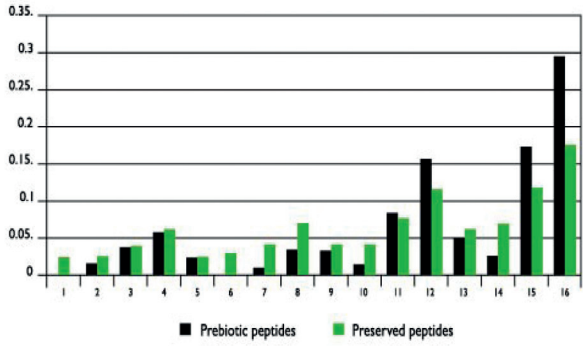

Figure 5. Comparison of polarity profiles between simulated peptides and those of best-preserved genes from three microorganisms (Delaye *et al.*, 2005).

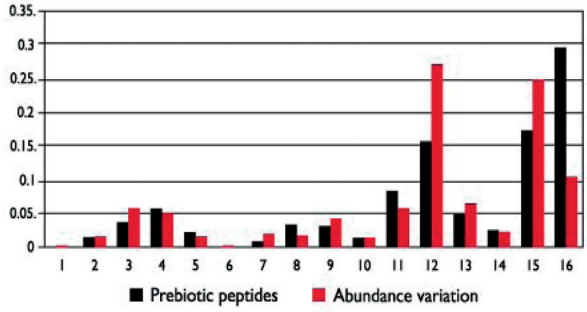

Figure 6. Effect of the alteration in the absolute abundance by 20% in the alanine polarity profile in simulation.

amino acid abundances with 440 μmol, from the neutral (N) to the non-polar (NP) polarity group, caused a similarly small change. In contrast, significant changes in the polarity profile of up to 30% were observed when the assumed abundances for alanine (A, 790 μmol), glycine (G, 440 μmol), and α-amino-n-butyric acid (9, 270 μmol) were varied by up to 20%. Fig. 6 shows the performance of alanine that is similar to that of glycine and α-amino-n-butyric acid. Equivalent procedures for the amino acids with low abundances resulted in insignificant changes of the polarity profile.

The importance of the amino acid abundances in the peptide composition may imply that the prebiotic composition constituted the main element in the generation of sequential information. It also reinforces the idea that biochemical processes produce changes in the polarity interactions that induce the specialization of peptides. Consequently, the vast majority of peptides act as templates, which could explain why their toxicity is generalized and not specialized when analyzing peptides from antimicrobial databases (Wang *et al.*, 2009).

The predicted by simulation polarity profile, and that of the best-preserved genes, is almost identical. Based on this correspondence, one may speculate that the "first peptides", i.e. those that were generated during the course of the chemical evolution about 1.5 billion years *before* the best-preserved genes have evolved, had already a defined structure. In consequence, this thought could support the so-called panspermia scenario for the origins of life on Earth where already more complex peptides came from outwards.

The dynamical features of the peptide generation are reflected in the polarity profiles by the presence of catastrophic bifurcation points (Arnold, 1974; Thom, 1975). We approached the polarity profiles by seven-degree continuous function graphics and identified the turning points where the graphics change their ascendant or descendent tendency as well as their concavity. These changes are distinctive dynamic features and allow for the recognition of the catastrophic bifurcation points. Such dynamic bifurcation analysis applied to the polarity matrix of the simulated peptides and of best preserved genes identified the catastrophic bifurcation points {4,8,12} in two sets (Fig. 7), which are clearly associated to the last elements of the first three rows of the polarity matrix. This coincidence in the location of the bifurcation points may indicate the existence of an algebraic structure associated to the polarity matrix and could support the assumption that the catastrophic bifurcation points stand for regions, in which the definition of the peptide functionality takes place. Depending on its validation for other peptide sets, for instance those with pathogenic actions, the dynamic bifurcation analysis could become an unprecedented mathematical contribution to the field of proteomics.

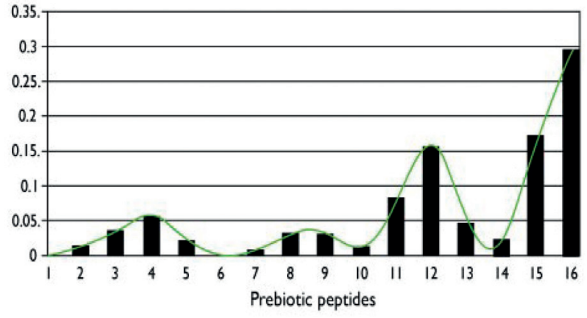

Figure 7. Catastrophic bifurcation points 4, 8, 12, and 16 that coincide with the change in the polarity matrix row P[*i,j*].

CONCLUSIONS

In this paper, we have presented a toy model that takes into account the assumed dynamic elements of prebiotic peptide evolution is able to generate sequential information that is close to that of best-preserved genes. The main impact for this sequential information appears to originate from the prebiotic amino acid abundances. A more detailed modeling of a prebiotic scenario to recreate the peptide generation certainly requires a thorough study of a possible peptide/nucleotide co-evolution and the biochemical factors regulating the aqueous-lipid medium. However, our results suggest that even a minimalist model that includes the elements of self-replication and peptide splitting/recombination can possibly predict the main factors responsible for the generation of peptide sequential information.

In addition to the mathematical-computational modeling, it was required to design a method to equip the peptide profiles with a quantitative parameter by considering their components without using the specific prebiotic elements or amino acid identification. During our former research, we designed the so-called polarity index method that uses only the constitutive amino acids of the peptides and their electrical charges. This method required only the translation of the linear sequence into its electrical charge or polarity equivalence. This way, the peptide sequences were evaluated based on their polarity, which allowed for the comparison of peptides with different components.

We are aware that each modeling approach remains restricted by the consideration of a limited amount of variables. Hence, our model does not account for more microscopic details of sequential information generation. However, we believe that our method and the obtained results are useful as a contribution to the fundamental question about the origins of life, as well as for the developing field of proteomics. The programming structure of the model allows for the addition of other important prebiotic aspects such as the origin of homochirality of the biopolymers that are planned for future studies.

COMPUTATIONAL RESOURCES

The computer program was written in Fortran 77 and executed on a Fedora 14 Unix-type platform (GNU). The program is freely available at request at polanco@unam.mx. Its implementation was optimal in computers of four or more processors in shared memory. The pro-

cessing time or stay was 24 h. Its implementation in uniprocessor computers is not recommended. We run the program from 1 up to 50 generations in an HP Workstation z210 — CMT — 4 x Intel Xeon E3-1270/3.4 GHz (Quad-Core) — RAM 8 GB — SSD 1 x 160 GB — DVD SuperMulti — Quadro 2000 — Gigabit LAN, Linux Fedora 14, 64-bit. Cache Memory 8 MB. Cache Per Processor 8 MB. RAM 8 GB.

Acknowledgments

The authors gratefully acknowledge the Departamento de Cómputo, Instituto de Ciencias Médicas y Nutrición Salvador Zubirán for support, and Concepcion Celis Juarez for proofreading.

Author Contributions

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

Conflict of Interests

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

REFERENCES

Arnold VI (1974) Critical point of smooth functions. *Vancouver Intern. Congr. of Math* **1**: 19–39.

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

Cousins GR, Poulsen SA, Sanders JK (2000) Molecular evolution: dynamic combinatorial libraries, autocatalytic networks and the quest for molecular function. *Curr Opin Chem Biol* **4**: 270–279.

Dadon Z, Wagner N, Ashkenasy G (2008) The road to non-enzymatic molecular networks. *Angew Chem Int Ed Engl* **47**: 6128–6136.

Delaye L, Becerra A, Lazcano A. (2005) The last common ancestor: what's in a name? *Orig Life Evol Biosph* **35**: 537–554.

Fox SW, Harada K, Krampitz G, Mueller G (1970) Chemical Origins of Cells, pp 80–94. Chemical and Engineering News, USA.

Herrera E (1993) *Elementos de Bioquímica,* pp 34–41. Interamericana, McGraw-Hill, DF, México.

Hopfield JJ (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci USA* **71**: 4135–4139.

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

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27–30.

Lahav N, White DH (1980) A possible role of fluctuating clay-water systems in the production of ordered prebiotic oligomers. *J Mol Evol* **16**: 11–21.

Lambert JF (2008) Adsorption and polymerization of amino acids on mineral surfaces: a review. *Orig Life Evol Biosph* **38**: 211–242.

Miller SL (1953) A Production of amino acids under possible primitive earth conditions. *Science* **117**: 528–529.

Mosqueira FG, Negron A, Ramos S, Polanco C (2012) Biased versus unbiased randomness in homo-polymers and copolymers of amino acids in the prebiotic world. *Acta Biochim Pol* **59**: 543–547.

Oprea TI, Tropsha A, Faulon JL, Rintoul MD (2007) Systems Chemical Biology. *Nat Chem Biol* **3**: 447–450.

Paul N, Joyce GF (2004) Minimal self-replicating systems. *Curr Opin Chem Biol* **8**: 634–639.

Polanco C, Samaniego JL (2009) Detection of selective cationic amphipatic antibacterial peptides by Hidden Markov models. *Acta Biochim Pol* **56**:167–76.

Polanco C, Samaniego JL, Buhse T, Mosqueira FG, Negron-Mendoza A, Ramos-Bernal S, Castañón-González JA (2012) Characterization of selective antibacterial peptides by polarity index. *Int J Pept* 585027. doi: 10.1155/2012/585027.

Thom R (1975) Stabilité structurelle et morphogénèse: essai d'une théorie générale des modèles, pp 348–362. Addison-Wesley, Inc, USA.

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