Regular paper

# Computational model of abiogenic amino acid condensation to obtain a polar amino acid profile

Carlos Polanco[1✉], Thomas Buhse[2], José Lino Samaniego[1], Jorge Alberto Castañón González[1], and Miguel Arias Estrada[3]

[1]Facultad de Ciencias de la Salud, Universidad Anáhuac, Col. Lomas Anáhuac C.P. 52786 Huixquilucan Estado de México, México; [2]Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, C.P. 62209 Cuernavaca, Morelos, México; [3]Departamento de Cómputo Reconfigurable y de Alto Rendimiento, Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro # 1, C.P.: 72840 Tonantzintla, Puebla, México

In accordance with the second law of thermodynamics, the Universe as a whole tends to higher entropy. However, the sequence of far-from-equilibrium events that led to the emergence of life on Earth could have imposed order and complexity during the course of chemical reactions in the so-called primordial soup of life. Hence, we may expect to find characteristic profiles or biases in the prebiotic product mixtures, as for instance among the first amino acids. Seeking to shed light on this hypothesis, we have designed a high performance computer program that simulates the spontaneous formation of the amino acid monomers in closed environments. The program was designed in reference to a prebiotic scenario proposed by Sydney W. Fox. The amino acid abundances and their polarities as the two principal biases were also taken into consideration. We regarded the computational model as exhaustive since 200 000 amino acid dimers were formed by simulation, subsequently expressed in a vector and compared with the corresponding amino acid dimers that were experimentally obtained by Fox. We found a very high similarity between the experimental results and our simulations.

Key words: amino acids; proteinoids, origins of life; polarity; computer simulations

## INTRODUCTION

Most of the assumed prebiotic processes, particularly those related to the condensation of amino acids, were revealed by synthetic attempts using complex building block mixtures. While it is true that the found synthetic products are still far from the first living systems, they may resemble in some of their properties living matter. There are many types of such product mixtures known as sulfobios and colpoides (Kanehisa & Goto, 2000), coacervates (Herrera, 1942), and proteinoids (Fox & Harada, 1960; Fox, 1960; Fox & Huyama, 1963).

Computational simulations of these experiments could reveal useful information, for instance about the formation of prebiotic pre-cellular dipeptides. However, each model faces different difficulties in its implementation because of the high number of involved molecular components as well as the associated complex atmospheric and catalytic processes.

Given these constraints, we built a simplified model of amino acid condensation. The model, a result of our computational approach, was designed and implemented by us to mimic the succession of amino acids and was also restricted by their proportions (Polanco et al., 2013a), as in the present case used in the seminal experiments of Fox (Fox & Harada, 1960). We tried to evaluate one main aspect, which we considered fundamental for the function of the amino acid succession, namely a profile or polar pattern that evolves as the result of the predominance of certain amino acids by their relative abundances.

Fox and Harada (1960) described the experimental procedures as follows: "Ten grams of L-glutamic acid was heated at 175–180° until molten (about 30 min.) after which period it had been largely converted to the lactam. At this time, 10 g of DL-aspartic acid and 5 g, of the mixture of the sixteen basic and neutral (BN) amino acids were added. The solution was then maintained at 170±2° under an atmosphere of nitrogen for varying periods of time. Within a period of a few hours considerable gas had been evolved, and the colour of the liquid changed to amber. When the inside of the tubes were observed and chromatograms were taken, it showed the presence of structures". The authors called these structures "proteinoids". The assumed structures were composed of glutamic acid, aspartic acid and other amino acids. The percentages of each amino acid suggested that the arrangement of the constituents were non-random due to the high concentration of glutamic acid and aspartic acid, compared to the rest of the amino acids found. The highest temperatures reached today on the Earth's surface is 80°C, except in volcanoes. Therefore, the thermal condensation of amino acids by direct heating could have happened most likely in the vicinity of volcanic activities (Fox & Harada, 1959; Nelsestuen, 1980).

Although the research activities in proteinoids "have gone out of fashion" (Rauchfuss, 2008) because of their supposedly limited prebiotic relevance and only marginal peptide bond formation (Jakschitz & Rode, 2012), we have taken the Fox-type experiments as a prototypical case to built our computational model. We have taken

✉e-mail: polanco@unam.mx

**Table 1. Amino acid description**

| # | Chemical name | Symbol (one-letter code) | Proportion of abundance (grams) | Inverse of abundance | Polar profile |
|---|---|---|---|---|---|
| 1 | Arginine | R | 0.31 | 90 | P+ |
| 2 | Cystine | C | 0.31 | 90 | N |
| 3 | Alanine | A | 0.31 | 90 | NP |
| 4 | Glycine | G | 0.31 | 90 | NP |
| 5 | Histidine | H | 0.31 | 90 | P+ |
| 6 | Isoleucine | I | 0.31 | 90 | NP |
| 7 | Leucine | L | 0.31 | 90 | NP |
| 8 | Lycine | K | 0.31 | 90 | P+ |
| 9 | Methionine | M | 0.31 | 90 | NP |
| 10 | Phenylalanine | F | 0.31 | 90 | NP |
| 11 | Proline | P | 0.31 | 90 | NP |
| 12 | Serine | S | 0.31 | 90 | N |
| 13 | Threonine | T | 0.31 | 90 | N |
| 14 | Tryptophan | W | 0.31 | 90 | NP |
| 15 | Tyrosine | Y | 0.31 | 90 | N |
| 16 | Valine | V | 0.31 | 90 | NP |
| 17 | Glutamic acid | E | 10.00 | 3 | P– |
| 18 | Aspartic acid | D | 10.00 | 3 | P– |

Source: Number and corresponding polarity group. Protein amino acids are symbolized by a letter and non-protein amino acids by a number; Proportion of abundance: Fox experiment, 2:2:1 (in gr) (Oparin & Gladilin, 1980); Inverse of abundance: Inverse of the abundance normalized to hundred. (g). Polar profile: Type of polarity: polar amino acids with positive charge (P+), polar amino acids with negative charge (P-), neutral (N) and non-polar (NP) (Polanco et al., 2013).

from these experiments only the proportion and number of the initial amino acids to generate amino acid successions, with the purpose to compare the initial and final relative amino acid frequencies from the Fox experiments with our modeling results. We will show that both profiles exhibit a very high similarity, which indicates that the bias in the Fox experiments was most likely due to the polarity of the involved amino acids.

The studies of presumed molecular scenarios during chemical evolution, among them the Fox experiment, are also related to the still disputed origin of biomolecular homochirality (Flügel, 2011; Meierhenrich, 2008). In the present case, the question which arises is how the homochirality evolved in today's proteins from possible racemic or nearly racemic mixtures of amino acids generated for instance by Miller-Urey-type processes on the primitive Earth (Miller, 1953) or by other settings perhaps in space (Muñoz Caro et al., 2002). Although Fox did not make explicit reference to the chirality question, a number of mathematical models were developed showing chiral symmetry breaking or amplification scenarios during generalized condensation processes that also could stand for short peptides building (Sandars, 2003; Schmidt, 2006; Wattis & Coveney, 2005). Taking into account possible diastereomeric implications in the formation of homochiral vs. heterochiral short peptides or short peptide fragments (Banik & Nandi, 2013; Jakschitz & Rode, 2012), some experimental progress has been made in respect to the preferential adsorption of dipeptide diastereomers on clay mineral surfaces (Bujdák et al., 2006) or diastereomeric discrimination due to differences in their hydrophobicities (Munegumi & Shimoyama, 2003). Even though these aspects are out of the scope of our present work. Thus, they deserve attention

for prospective modeling attempts of prebiotic short peptides building.

## MATERIAL AND METHODS

**Description of the model.** The computer simulations of the Fox experiment were based on the relative abundances of the intial amino acids and their polar characteristics. Sequences of two amino acids generated by the model were taken as equivalent to experimentally described dipeptides (nowadays perhaps better characterized as amino acid dimers). The model generates amino acids randomly and creates not further specified bonds between two of them. Successful bonding is restricted to comply with the constraints of abundance and polarity. Otherwise, the amino acids were discarded and the model repeated the generation of new starting amino acids.

The model considered the amino acid abundances and polarities by a set of 18 amino acids (Table 1). In reference to the Fox experiment (Fox & Harada, 1960), the following proportions were chosen: 10 g of aspartic acid, 10 g of glutamic acid and 5 g from of the other 16 ami-

**Table 2. Matrix incidents by polarity**

| | P+ | P– | N | NP |
|---|---|---|---|---|
| P+ | 0 | 0 | 0 | 0 |
| P– | 0 | 0 | 0 | 0 |
| N | 0 | 0 | 0 | 0 |
| NP | 0 | 0 | 1 | 0 |

An array of incidents of polarity with occurrences of the amino acid dimer sequences generated by the model. In the case of study note that this registered incidence (P, Y) = 1 (Section Description of the model)

**Table 3. Matrix incidents by abundance**

|   | R | C | A | G | H | I | L | K | M | F | P | S | T | W | Y | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

An array of incidents of abundance with occurrences of the simulated amino acid dimer sequences generated by the model. In the case of study note that this registered incidence (P, Y) = 1 (Section Description of the model).

no acids, i.e. 0.31 g from each. Once the model has produced two amino acids, the formation of a dimer may take place in accordance with the criteria of abundance and polarity. As given in Table 1, the amino acids exhibit different abundances and belong to different polar groups. The model considers these differences to allow the formation of a two-succession of amino acids.

The model keeps a detailed record of the number of every interaction incident during the formation of dimeric amino acid sequences, i.e. it checks how many times a particular bond and type of amino acid has been produced. It only allows that the amino acid dimer formation occurs until it has reached a preset number of binding attempts.

As a first aspect, we define how the model generates an interaction between the amino acid monomers. For instance, the model has produced the first amino acids P and Y, then it creates an interaction {P, Y} and registers it in the array of polarity (Table 2) and abundance (Table 3) by adding a 1 to the element (P,Y) in the array of polarities and the element (P,Y) in the array of abundances. In the Tables 2 and 3 these incidences have been highlighted in **blue** color.

The interaction processes run continuously until in both arrays of Tables 2 and 3 the incidence values of their matrices have reached pre-established values (Tables 4 and 5). Once one of this value has been reached, the model (i) accepts that interaction, (ii) the amino acid dimer sequence is added to the file of already generated amino acid dimer sequences, and (iii) the records of the incidents in Tables 2 and 3 are restored to zero.

For example, if two amino acids were generated randomly, one corresponding to the P+ group and the other to the of P– group, the number of iterations must occur before that the interaction can be accepted is 17. It means that the polar contact counter increases by one unit in each binding attempt, but the amino acid dimer formation is allowed only until it reaches the value of 17. A similar procedure is performed in respect to the criterion of abundance.

**Matrices of pre-established values.** Pre-established matrices (Tables 4 and 5) have been constructed from the default values of abundance (Table 1) and polarity (Nelsestuen, 1980; Polanco *et al.*, 2012; Wickramasinghe, 1973). To interpret them computationally, we built a function distance that acts inversely proportional to the values of abundance and polarity; that is, the higher its value the lower is the possibility of interaction. In the case of abundance, we followed the experiments by Fox. Here aspartic acid and glutamic acid were employed in much higher proportions than the rest of the other 16 amino acids. The case of polar interactions was simulated using the 16 possible polar interactions of a polarity matrix introduced by Mosqueira (Mosqueira *et al.*, 2012). Note that this matrix is not symmetric.

**Implementation of incidence vector. Computational model.** The set of all computational amino acid dimer sequences by the model are stated in an array of incidences or frequencies. For instance, if the file contains four simulated amino acid dimer sequences: AC, KP, EE, and DW, then the array of incidents (Table 6) is constructed by adding one to the incidents corresponding to each residue.

**Table 4. Matrix of pre-established values by polarity**

|     | P+ | P– | N | NP |
|-----|----|----|----|----|
| P+  | 87 | 17 | 85 | 99 |
| P–  | 17 | 87 | 85 | 99 |
| N   | 60 | 60 | 90 | 99 |
| NP  | 60 | 60 | 85 | 100 |

Matrix of pre-established values built from the polarity values of Mosqueira *et al.* (2012).

**Table 5. Matrix of pre-established values by abundance**

|   | R | C | A | G | H | I | L | K | M | F | P | S | T | W | Y | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| C | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| A | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| G | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| H | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| I | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| L | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| K | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| M | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| F | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| P | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| S | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| T | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| W | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| Y | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| V | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| E | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 3 | 3 |
| D | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 3 | 3 |

Matrix of pre-established values built from the values of abundance referred in Table 1 (Oparin & Gladilin, 1980).

**Table 6. Example of incidences vector**

| % | R | H | K | C | S | T | Y | A | G | I | L | M | F | P | W | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.0 | 0.0 | 12.5 | 12.5 | 0.0 | 0.0 | 0.0 | 12.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 12.5 | 12.5 | 0.0 | 25.0 | 12.5 |

Incidences vector for the case study of the four simulated amino acid sequences of length two forms by the computational model: AC, KP, EE, and DW (Section Implementation of incidence vector).

**Table 7. Computed frecuency vector**

| % | R | H | K | C | S | T | Y | A | G | I | L | M | F | P | W | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 8.4 | 7.4 | 5.8 | 1.3 | 2.7 | 2.8 | 2.0 | 4.6 | 2.0 | 2.8 | 4.0 | 5.4 | 1.2 | 1.2 | 3.1 | 2.6 | 13.7 | 29.0 |

Each position corresponds to the relative frequency of amino acids that form amino acid dimer sequences.

**Table 8. Experimental frecuency vector**

| % | R | H | K | C | S | T | Y | A | G | I | L | M | F | P | W | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 1.2 | 1.4 | 2.4 | 0.1 | 0.5 | 0.1 | 1.8 | 3.0 | 2.0 | 1.0 | 1.4 | 2.1 | 1.7 | 0.3 | 1.0 | 0.9 | 10.0 | 67.5 |

Distribution of relative frequencies of amino acids in the Fox experiment.

**Table 9. Bifurcation points (Fox experiment)**

| % | R | H | K | C | S | T | Y | A | G | I | L | M | F | P | W | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **100** | 1.2 | 1.4 | **2.4** | **0.1** | 0.5 | **0.1** | 1.8 | **3.0** | 2.0 | 1.0 | 1.4 | 2.1 | 1.7 | **0.3** | 1.0 | 0.9 | **10.0** | **67.5** |

Final relative frequency of amino acids from the experiment of Fox. The positions in **red** colour represent maximum values and positions, in **blue** colour represent the minimum points, and in **green** colour represent the catastrophic bifurcations points (Section Fox Experiment).

**Fox Experiment**. The incident vector from the Fox experiment was taken from the percentages of the initial amounts of amino acids expressed in their work (Fox & Harada, 1959, Table 1).

**Implementation test.** The computational model generated 200 000 simulated amino acid dimer sequences. With this set, we generated the vector of incidences, according to the procedure described in Section Implementation of incidence vector.

**Catastrophic bifurcations points.** Catastrophic bifurcation points (Arnold, 1974; Herrera, 1942) are points (in **green** color) where abrupt changes in the behavior of a function occur (Tables 9 and 10). These points are associated with the positions in the polarity matrix where the maximum/minimum (in **red**/**blue** colors respectively) frequencies are observed. Catastrophic bifurca-

tion points are the midpoints between the maximum and minimum positions, i.e. the inflection points.

**Mathematical description of the model.** The outline of the computational model was based on the Markov process (Rabiner, 1989). A Markov process is a chain of events in a random process that takes values according to certain restrictions. These systems evolve over time, generating time series randomly from state to state and emitting products in every moment at random. In this particular case, generating sequences of amino acids represent dimers. The two main restrictions used here are the polarity and the abundance. Both are represented by already described matrices (Section Matrices of pre-established values). The computational implementation of Markov processes is not affected by the number of involved variables, in contrast to differential models that
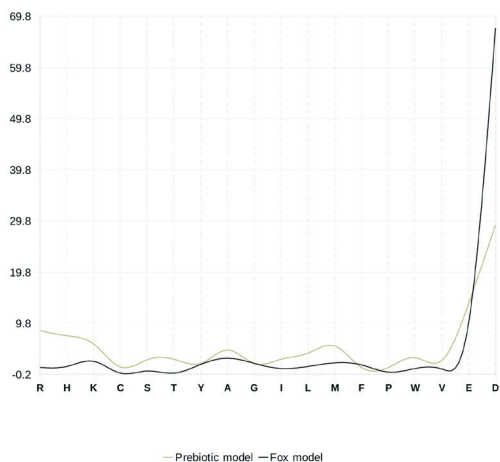
**Figure 1. Smoothed curves visualizing the relative frequencies of amino acids obtained by the computational model *vs*. by the Fox experiment.**
The 18 columns on the x-axis correspond to the 18 amino acids of the incidence vector (Section Fox Experiment).

increase exponentially in complexity with the number of variables. The Markov process used here is called "Hidden Markov model", which we have already used for the identification of antibacterial short peptides (Polanco & Samaniego 2009). It means that the model restrictions are known but the searched pattern is hidden. The pattern in the present case is the polarity profile of the set of amino acid dimer sequences.

## RESULTS

The incidence vectors from the group of simulated amino acid dimer sequences (Table 7) and the experimental results by Fox (Table 8) were represented by continuous curves (Fig. 1). Both curves match almost evenly in their points of maximum and minimum, with the exception of the first three amino acids (R, H and K), where the slope in both curves is contrary. The apparent exchange between the maximum and minimum values (Fig. 1, see M and F amino acids, or W and V amino acids) is not present in the recorded values (see Tables 9 and 10). This distortion is caused by the approximation through curve smoothing. However, the smoothed graphs are useful to visualize the coincidence between both scenarios.

The main maximum value points from the experiment of Fox are located in the positions K, A and D from the incidence vector (Table 9, numbers in **red** colour), their corresponding relative frequency of amino acids counted from the computational model is in the positions R, M Y D (Table 10, numbers in **red** colour). The match is 1 out of 3 total. The main minimum value points from the Fox experiment are located in the positions T and P from the incidence vector (Table 9, numbers in **blue** colour), their corresponding relative frequency of amino acids counted

from the computational model is in the positions G and P (Table 10, numbers in **blue** colour). The match is 1 out of 2 total. The catastrophic bifurcation points in the Fox experiment and computational model, are located at the {C and E} positions. Both sets coincide completely (Fig. 1).

## DISCUSSION

The relative amino acid frequency predicted by the computational model is in good agreement to the frequency observed in the Fox experiment. This similarity shows almost a tie at the location of the maximum, minimum, and catastrophic points in both graphics (Fig. 1). This ensures that the computer model correctly emulated the Fox experiment. The computational model is fully stochastic and it is suggested that all stochastic processes reach stability over time (Arnold, 1974; Thom, 1975). Such stability is independent of the random generators used. It means that the stochastic process results in a profile that is always the same for a sufficiently small range (neighborhoods) to the established biases and independent of the number of biases (Polanco *et al.*, 2013a). The feature of our present stochastic modeling was the generation of 200,000 amino acid dimer sequences that was sufficient to achieve stability of the system. It is worth mentioning that the predicted relative frequency of the amino acids, i.e. the obtained end profile, does not vary substantially or becomes modified in the neighborhood of the polarity or abundance biases.

Our previous studies by simulating the Miller experiment indicated that very few variables had a fundamental effect on the process of amino acids profile (Polanco *et al.*, 2013a). In the present work, we have used the same two variables of abundance and polarity and found that they were sufficient to recreate the Fox experiment. Hence, we consider computational modeling attempts like ours as useful to reveal exactly such similarities. Depending on further studies, these similarities can perhaps convert into generalities and contribute to the understanding of the evolution of amino acid sequences and abundances from the prebiotic world to today's biosphere (Jordan *et al.*, 2005).

## COMPUTING RESOURCES

The computer program was written in Fortran 77 and executed on a platform Unix-type operating system (GNU) Fedora 14. Its implementation is optimal in computers of two or more processors in shared memory. The processing time or stay is 120 h. We do not recommend its implementation in uni-processor computers or under MPICH (MPICH, 2013). The computational program should be executed in a computer 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-bits. Cache Memory 8 MB. Cache Per Processor 8 MB.

**Table 10. Bifurcations points (computational model)**

| % | R | H | K | C | S | T | Y | A | G | I | L | M | F | P | W | V | E | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | **8.4** | 7.4 | 5.8 | **1.3** | 2.7 | 2.8 | **2.0** | 4.6 | 2.0 | 2.8 | 4.0 | **5.4** | 1.2 | 1.2 | 3.1 | 2.6 | **13.7** | **29.0** |

Final relative frequency of amino acids from the computational model. The positions in **red** colour represent maximum values and positions in **blue** colour represent the minimum points, and in **green** colour represent the catastrophic bifurcations points (Section Implementation test).

RAM 8 GB, because of linux scripts inserted in the computer program.

## CONCLUSIONS

In this work, we assumed a minimum of physical-chemical components for the generation of amino acid dimer sequences, namely the abundance and polarity. Once we had constituted a set amino acid dimers, we evaluated its final relative frequency and compared it with the final relative frequency of the amino acids found in the "proteinoids" of Fox's experiments. We found a very high similarity. Hence, it is indicated that the biases of abundance and polarity are important factors for the random generation of amino acids and highlights the importance of the Fox experiments for other trials of amino acid condensation.

## Availability

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

## Conflict of Interests

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

## Author Contributions

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

## Acknowledgements

## REFERENCES

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

Banik SD, Nandi N (2013) Chirality and protein biosynthesis. *Top Curr Chem* **333**: 255–306.

Bujdák J, Remko M, Rode BM (2006) Selective adsorption and reactivity of dipeptide stereoisomers in clay mineral suspension. *J Colloid Interface Sci* **294**: 304–308.

Flügel RM (2011) *Chirality and Life*, Springer-Verlag Berlin Heidelberg (ISBN 978-3-642-16976–2).

Fox SW (1960) How did life begin. *Science* **132**: 200–208.

Fox SW, Harada K (1960) The thermal copolymerization of amino acids common to protein. *J Am Chem Soc* **82**: 3745–3751.

Fox SW, Harada K (1959) Thermal copolymerization of amino acids to a product resembling protein. *Science, New Series* **128**: 1214.

Fox SW, Yuyama S (1963) Effects of the gram stain on microspheres from thermal polyamino acid. *J Bacteriol* **85**: 279–283.

Herrera AL (1942) A new Theory of the origin and nature of Life. *Science* **96**: 14.

Jakschitz TAE, Rode BM (2012) Chemical evolution from simple inorganic compounds to chiral peptides. *Chem Soc Rev* **41**: 5484–5489.

Jordan IK, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, Kondrashov AS, Sunyaev S (2005) A universal trend of amino acid gain and loss in protein evolution. *Nature* **433**: 633–638.

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

Kenyon DH, Steinman G (1969) Biochemical Predestination, p243, McGraw-Hill, USA. Message Passing Interface chamalleon (MPICH) (http://www.mpich.org/) accessed on May 9, 2013.

Meierhenrich U (2008) *Amino Acids and the Asymmetry of Life*, Springer-Verlag Berlin Heidelberg (ISBN 978-3-540–76885–2).

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.

Munegumi T, Shimoyama A (2003) Development of Homochiral Peptides in the Chemical Evolutionary Process: Separation of Homochiral and Heterochiral Oligopeptides. *Chirality* **15**: S108–S115.

Muñoz Caro GM, Meierhenrich UJ, Schutte WA, Barbier B, Arcones-Segovia A, Rosenbauer A, Thiemann WHP, Brack A, Greenberg JM (2002) Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416**: 403–406.

Nelsestuen GL (1980) Origin of life: consideration of alternatives to proteins and nucleic acids. *J Mol Evol* **15**:59–72.

Oparin AI (1961) The origin of life. *Nord Med* **65**: 693–697.

Oparin AI, Gladilin KL (1980) Evolution of self-assembly of probionts. *Biosystems* **12**: 133–145.

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

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

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

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

Rabiner LR (1989) A tutorial on hidden markov models and selected applications in speech recognition. *Proc IEEE* **77**.

Rauchfuss H (2008) *Chemical Evolution and the Origin of Life*, Springer-Verlag Berlin Heidelberg, (ISBN: 978-3-540-78822-5) p. 139.

Rode BM, Schwendinger MG (1990) Copper-catalyzed amino acid condensation in water- a simple possible way of prebiotic peptide formation. *Origins Life Evol Biosphere* **20**: 401–410.

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

Sandars PGH (2003) A toy model for the generation of homochirality during polymerization. *Origins Life Evol Biosphere* **33**: 575–587.

Schmidt P (2006) Evolution of homochirality by epimerization of random peptide chains. A Stochastic Model. *Origins Life Evol Biosphere* **36**: 391–411.

Schwendinger MG, Rode BM (1989) Possible role of copper and sodium chloride in prebiotic evolution of peptides. *Analytical Science* **5**: 411–444.

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** (Database issue): D933–D937.

Wattis JAD, Coveney PV (2005) Symmetry breaking in chiral polymerization. *Origins Life Evol Biosphere* **35**: 243–273.

Wickramasinghe RH (1973) Iron-sulphur proteins: Their possible place in the origin of life and the development of early metabolic systems *Space Life Sciences* **4**: 341–352.