# Proteinoid microspheres - Fox & Harada model: their impact in Structural Proteomics

**C. Polanco*[,1], A. Ponce de León[1], M. Bañuelos Cedano[1], and P. Vazquez González[2]**

[1] Department of Mathematics, Faculty of Sciences, Universidad Nacional Autónoma de México, 04510 México.
[2] Department of Epidemiology, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, 14080 México.

This review discusses the characteristics of the proteinoid microspheres observed in Fox & Harada's works. How these cellular models evoke the Protocells and support the understanding of the chemical units' evolution and the current mathematical implementations oriented to Structural Proteomics.

**Keywords**: abiogenesis; amino acids; panspermia hypothesis; proteinoid microspheres; origin of life; chemical evolution; mathematical models; structural proteomics; peptides; proteins.

## 1. A controversial experiment

An aspect related to the Fox & Harada experiment [1] in the constant curvature observed in proteinoid microspheres, is assumed to be the result of the balance of electrical charge [2] in the amino acids that form a protein. This conjecture has been documented with the computational modelling of this experiment, done by the same author [3,4], that shows there is a balance in the characteristic electrical charge of the different protein groups, called polarity profile [4]. The present work provides elements to conjecture this polar profile, that starts in the polymerization, extends and it is reflected in the constant curvature [5] observed in the proteinoid microspheres.

## 2. What was first RNA or Protein World?

The debate on RNA as the first protein comes from a term used by Walter Gilbert in 1986 ([6], where he raises the conjecture of the "RNA world".). This hypothesis states that RNA has the capacity to store, transfer and reproduce the genetic material in a similar way than DNA [7].

What is RNA? It is a nucleic acid formed by a chain of ribonucleotides (adenine (A), uracil (U), guanine (G) and cytosine (C)). This acid is present in both prokaryotic and eukaryotic cells, RNA adopts a linear cell conformation of a simple strand [8]. The main difference with DNA is that it is formed by ribose as a monosaccharide and uracil as one of the nitrogenous bases, while DNA has deoxyribose and thymine [9]. DNA is responsible for storing genetic information, even if this genetic material is not used, and its strands are closed forming chromosomes; while RNA is located in other cells and it is responsible for copying and self-replicating the information contained in DNA and convert it into something functional. For instance, a DNA strand can determine a person has green eyes, this information is taken by RNA from DNA to create the necessary proteins to transfer these genes [10]. These properties of RNA suggest the possibility of the hypothesis of the "RNA world" [11], however, its acceptance as an explanation for the origin of life is not final. It is known that RNA is an efficient catalyst and as DNA it has the capacity to store information, it is also capable of self-replication conducting tasks of DNA and enzymes and therefore, it is thought to promote life independently from DNA. Although in the Miller-Urey experiment were not found nucleotides, essential for life formation [12], they were in the simulations and experiments of other researchers, see Oró J.'s work [13], who synthesized adenine from hydrocyanic acid. If the hypothesis raised in the RNA world is correct, this would have important implications for what we define as life. However, the most accepted theory corresponds to the discoverers of the DNA structure, Watson & Crick's work [14], where it is stated that life arose from DNA and extended to proteins by polymerization. DNA and proteins seemed to be the key molecules in living cells, whereas from this perspective, RNA had the sole purpose of creating proteins from a copy of DNA. This hypothesis enhanced the role of RNA in the origin of life, leading to numerous studies that have revealed interesting aspects related to the RNA function that, up to a couple of decades ago, were unknown. One of these discoveries is the formation of the peptide bond [15], which is the reaction that binds amino acids together to form proteins, catalysed by an adenine residue in the ribosomal RNA and therefore, the ribosome is a ribozyme. This discovery suggests that RNA molecules were capable of generating the first proteins.

## 3. Fox & Harada experiment

### 3.1 Computational approach

For the computational interpretation of the Fox & Harada experiment, 18 amino acids. The proportions were 10g Glu, and 10g Asp as well as 5g of the remaining 16 amino acids given in Table 1 [4; Table 10]. We took these proportions and two polarity distributions for the amino acids one of which induced a bias (Table 2-A) [4;Table 8-A], and one did not (Table 2-B) [4; Table 8-B]. 3000 peptides were generated.

**Table 1** Fox matrix of pre-established values by abundance

| | Arg | Cys | Ala | Gly | His | Ile | Leu | Lys | Met | Phe | Pro | Ser | Thr | Trp | Tyr | Val | Glu | Asp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Arg** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Cys** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Ala** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Gly** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **His** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Ile** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Leu** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Lys** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Met** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Phe** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Pro** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Ser** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Thr** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Trp** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Tyr** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Val** | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 90 | 85 | 85 |
| **Glu** | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 3 | 3 |
| **Asp** | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 85 | 3 | 3 |

Matrix of pre-established values by abundance used in Fox's hypothetical model. Table taken from [4, Table 10].

**Table 2** Fox matrix of pre-established values by abundance

| **A: Fox's model with polar bias** | | | | | | **B: Fox's model without polar bias** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **P+** | **P-** | **N** | **NP** | | | **P+** | **P-** | **N** | **NP** |
| **P+** | 99 | 21 | 85 | 95 | | **P+** | 100 | 100 | 100 | 100 |
| **P-** | 21 | 99 | 85 | 95 | | **P-** | 100 | 100 | 100 | 100 |
| **N** | 60 | 60 | 85 | 95 | | **N** | 100 | 100 | 100 | 100 |
| **NP** | 60 | 60 | 85 | 95 | | **NP** | 100 | 100 | 100 | 100 |

Inverse relative polarities by lateral chain: [P-] acidic hydrophilic, [N] neutral, [P+] basic hydrophilic and [NP] non-polar amino acids. A bias: with polar bias, B bias: without polar bias. Table taken from [4; Table 8-A and Table 8-B].

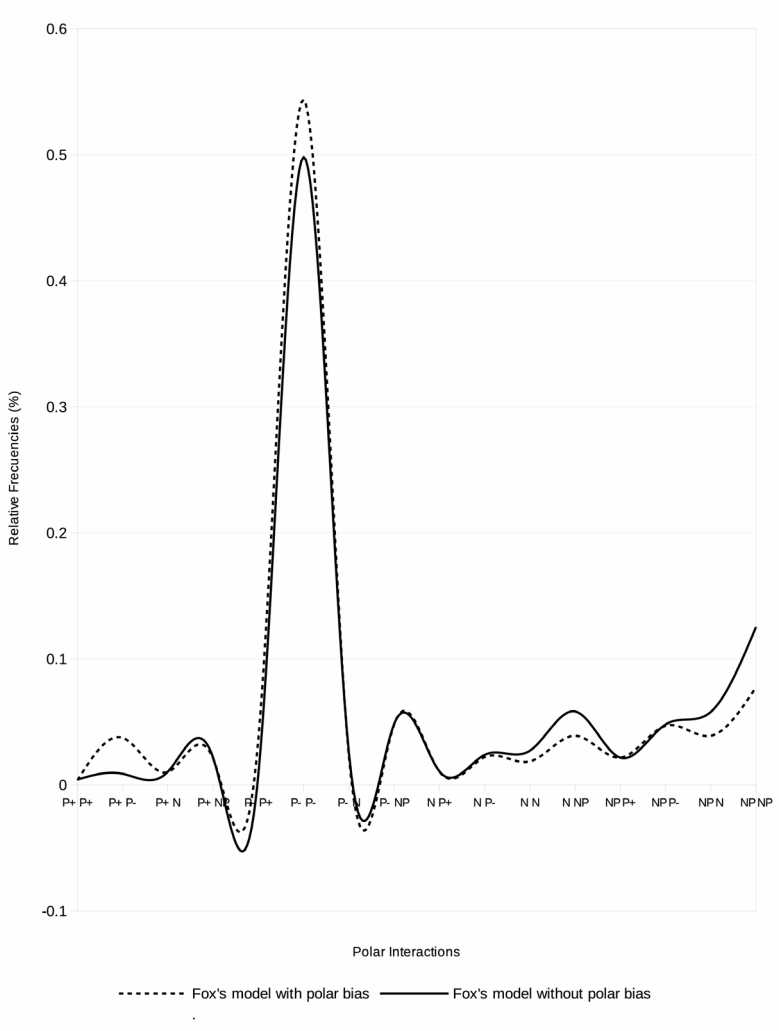


**Figure 1**. Linear polar interaction between the simulated peptides formed in Fox & Harada computational approach with and without polar bias. The 16 columns on the x-axis correspond to the 16 possible polar interactions [4].

### 3.2 Polar profile

The polarity profile was obtained from 3000 peptide sequences computationally calculated [3, 4] considering the physico-chemical restrictions of the experiment [3, 4]. The polarity profile is the cumulative relative frequency of the polarity incidents obtained when reading each peptide sequence from left to right by pairs one at a time. This simulation had two modes, with polarity bias and without polarity bias (Fig. 1). Note that in both cases the graph is very similar. This graph was compared with the corresponding polarity profiles of other prebiotic experiments already simulated by the author [16, 17] and with the different groups of peptides and proteins [18-25].

### 3.3 Curvature

Before delving on what constant curvature surfaces are, let us see what it is understood by surface [26] in mathematics. In differential and integral calculus of several variables [27], there are two types of surfaces. The first type corresponds to a function $F : D \subseteq R^2 \to R$ such that $F(x, y) = z$ .Surfaces can be obtained from this type of functions, however, they will not necessarily be differential in their entire domain. What does it mean? That the derivative is indefinite at least at one point of its domain. It could also be that the entire domain of the function is not differential, however, in order to consider a surface of constant curvature, it would be necessary to have functions that are differential in their entire domain and that will require another type of surfaces, the so-called parametric surfaces [27].

These surfaces are of the form $G : D \subseteq R^2 \to R^3$ where the surface is $S = G(D)$ thus, surface *S* is the image of function *G.* This type of parametrized functions comply with the properties required to define a function of constant curvature, provided they are differential functions [28].

The purpose of having parametric functions is to obtain surfaces of constant curvature. Why are they necessary? Because what it is required is the flow through these surfaces defined as surface integral of a vector field.

Thus, let ***F*** be a vector field defined in $D \subseteq R^3$ where the image of this function is *S*, a parametrized surface [29], then

$$\int_S F \cdot dD = \int_D F \cdot (T_u \times T_V) du dv \quad (1)$$

where $T_u$ and $T_v$ are the tangent vectors at a given point on surface *S.* This integral surface (Eq. 1) is also called integral flow as it is possible to measure the flow of vector field ***F*** with this integral.

To what extent have we explained surfaces of constant curvature? Enough has been said, however, it is necessary to measure the flow of a vector field, in its positive direction, and this is done with the so-called integral of the divergence of Gauss (Eq. 2) [29-31], which is expressed as follows:

Let *D* be a solid bounded by surface *S*, which is a close surface with positive orientation. Thus, let ***F*** be a defined vector field in *D* such that

$$\int_S F \cdot dS = \int_D (divF) dV \quad (2)$$

### 3.4 Stochastic model

Stochastic models are the representation of dynamic systems with random variables that evolve as functions from other variables that can be fully or partially known (random) [32]. Each random variable has a function of probability distribution that associates each value of the random variable with the cumulative probability up to that value [33]. These functions may or may not correlate, if they do, each pair of variables will have a joint distribution function i.e., the probability distribution function of the intersection of events in each variable [34]; and their marginal distribution functions, which are the probability distribution functions of a subset of random variables [35].

The probability distribution function can be defined for any real value of a random variable x (·) as $F_x(t) = P(\{ \xi : X(\xi) \leq t\} ) = P(X \leq t)$, for $t \in \mathbb{R}$ [36]. Whereas the joint distribution function $P_{xy}(\cdot)$ can be defined as (t, u) = Z (ξ) = [x, y] (ξ). And the marginal distribution function is defined as $P_{xy}(\cdot X R_2) = P_x(\cdot)$ and $P_{xy}(R_1 X \cdot) = P_y(\cdot)$, generated by x(·) and y(·) [37]. Thus, in order to characterize and model a stochastic process, it is necessary to know the functions mentioned above, which can be complicated. However, they can be simplified with the mean and variance of the random variables and the covariance, when available [38]. So, the mean (Eq. 3) in a set of N numbers $x_1, x_2, \cdots, x_N$ [39] would be:

$$\acute{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{\sum_{j=1}^{N} x_j}{N} = \frac{\sum x}{N} \quad (3)$$

The variance (Eq. 4), which is a measure of dispersion defined as the square of the standard deviation from its mean [40], would be:

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \acute{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \acute{x}^2 = \frac{1}{n}\sum_{i<j} (x_i - x_j)^2 \quad (4)$$

where $x_i$: each data, n: number of items, and x: arithmetic mean.

And the covariance, which is a value that indicates the degree of joint variation of two random variables i.e., the covariance of x and y is given by the relation between the variance of x and the variance of y, would be:

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \acute{x})(y_i - \acute{y}) \quad (5)$$

Thus, when $s_{xy} > 0$ (Eq. 5) the relation between variables is direct i.e., the greater values of x correspond to the greater values of y. When $s_{xy} = 0$ there is no lineal relation between variables. When $s_{xy} < 0$ the relation between variables is inverse i.e., the greater values of x correspond to the smaller values of y [41]. With this simplification, the stochastic models are widely used in different areas and disciplines such as economics, biology, physics, mathematics, anthropology, finance, actuarial science, marketing, sociology, psychology, etc. [42]. With them, it is possible to deal with dynamic models that have randomness, regardless of the size of their systems as they significantly reduce the processing time used to calculate series of large sampling. In this way, the characteristics of the underlying probabilistic structure can be inferred [43].

## 4. Stochastic simulation

Stochastic modelling is a real alternative when the number of variables of the function increases and it is not possible to fix one of them, as it makes it possible to reduce the domain without detriment on the quality of the simulation. In this sense, it is important to note that dynamical systems in the R and N fields do not provide reliable simulations. The work here presented is based on the Markov [44-46] conjecture, whose operator only takes into account the previous state of each variable and not all its behaviour and whose evaluation is indexed to its probability of occurrence. Thus, Markov systems reduce the definition of a domain in variables and enable the simulation of phenomena related to multiple variables.

## 5. Results

This interpretation of the geometric representation of the electrical charge balance (Fig. 1), leads to the assumption that this balance is reflected in the last geometric conformation of the proteinoid microspheres. Studies on the electrical charge of different peptide groups show that each group of proteins reach an electrical charge balance [44-46] and as result of this balance, they have a specific conformation in three-dimensional space.

**Acknowledgements** The authors thank Concepción Celis Juárez for proof-reading.

## References

[1] Fox SW, Harada K. Thermal copolymerization of amino acids to a product resembling protein. Science 1958.128:1214.
[2] Pauling L. General Chemistry [Química General] 3rd edition W. H. Freeman & Company Publishers, 1955, pp. 227,621.
[3] Polanco C, Buhse T, Samaniego JL, Castañón González JA, Arias Estrada M. Computational model of abiogenic amino acid condensation to obtain a polar amino acid profile. Acta Biochim Pol. 2014. 6:253-258.
[4] Polanco C, Samaniego JL, Buhse T, Castañón González JA. Discrete dynamic system oriented on the formation of prebiotic dipeptides from Rode's experiment. Acta Biochim Pol. 2014. 61:717-26.
[5] Jiang H, Si F, Margolin W, Sun SX. Mechanical Control of Bacterial Cell Shape. *Biophysical Journal*. 2011;101(2):327-335. doi: 10.1016/j.bpj.2011.06.005.
[6] Hipótesis del mundo de ARN. (2016, 14 de marzo). Wikipedia, La enciclopedia libre. Accesado en: March 14th, 2016 https://es.wikipedia.org/w/index.php?title=Hip%C3%B3tesis_del_mundo_de_ARN&oldid=90084389.
[7] Abiogénesis. (2016, 11 de marzo). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 15:15, marzo 15, 2016 desde https://es.wikipedia.org/w/index.php?title=Abiog%C3%A9nesis&oldid=89759632.
[8] Serrato MP. ¿Qué es el ADN? Universidad Nacional Autónoma Metropolitana, 2012. p. 1-4.
[9] López YJ. CÓDIGO GENÉTICO Transcripción, Traducción y Síntesis Proteica. Escuela Superior Politécnica del Litoral. 2009. p. 11-19.
[10] Burriel CV. Estructura y propiedades de los ácidos nucleicos. Universidad de Valencia, 2008. p.11-14.
[11] Watanabe F, Robles M, García JA. Redalyc. Sobre el origen del código Genético. Universidad La Salle. 2007. p. 94-105.
[12] Rojas PI. El origen de la vida sobre la tierra. Instituto de Física, Pontificia Universidad Católica de Valparaíso Chile. 2004. p. 19-22.
[13] El origen del universo y de la vida. p. 20-23.
[14] Ortiz HC. Encontramos el secreto de la vida. 50 años del descubrimiento de la estructura del ADN. Hospital ABC. 2003. p. 182-188.
[15] Enlace peptídico. (2016, 13 de febrero). Wikipedia, La enciclopedia libre. Fecha de consulta: 22:10, marzo 31, 2016 desde https://es.wikipedia.org/w/index.php?title=Enlace_pept%C3%ADdico&oldid=89127647.
[16] Polanco C, Buhse T, Samaniego JL, Castañón González JA. A toy model of prebiotic peptide evolution: the possible role of relative amino acid abundances. Acta Biochim Pol. 2013. 60:175-82.
[17] Polanco C, Buhse T, Vizcaíno G, Levy Picciotto J. The polar profile of ancient proteins: a computational extrapolation from prebiotics to paleobiochemistry. Acta Biochim Pol. 2017. (in press).
[18] Polanco C, Buhse T, Samaniego JL, Castañón González JA. Detection of selective antibacterial peptides by the Polarity Profile method. Acta Biochim. Pol. 2013. 60:183-199.

[19] Polanco C, Buhse T, Samaniego JL, Castañón González JA, Sordo ML. Characterization of a possible uptake mechanism of selective antibacterial peptides. Acta Biochim Pol. 2013. 60:629-633.
[20] Polanco C, Samaniego JL. Detection of selective cationic amphipatic antibacterial peptides by Hidden Markov models. Acta Biochim. Pol 2009. 56:167-176.
[21] Polanco C, Buhse T, Samaniego JL, Castañón González JA. Discrete dynamic system oriented to the formation of prebiotic dipeptides from Rode's experiment Acta Biochim. Pol. 2014. 56:167-176.
[22] Polanco C, Samaniego JL, Castañón González JA, Buhse T, Sordo ML. Characterization of a possible uptake mechanism of selective antibacterial peptides. Acta Biochim Pol 2013. 60:629-633.
[23] Polanco C, Samaniego JL. Detection of selective cationic amphipatic antibacterial peptides by Hidden Markov models. Acta Biochim Pol 2009. 56:167-176.
[24] Polanco C MICROBIAL PATHOGENS AND STRATEGIES FOR COMBATING THEM: SCIENCE, TECHNOLOGY AND EDUCATION: Selective antibacterial peptides: a review on their polarity. A. Méndez-Vilas. Formatex; 2013. ISBN 13 978-84-942134-0-3
[25] Polanco C, Samaniego JL, Castañón González JA, Buhse T, Sordo ML. Characterization of a possible uptake mechanism of selective antibacterial peptides. Acta Biochim Pol 2013. 60:629-633.
[26] Apostol, T.M., Calculus. Volumen I, México: Ed. Reverté, 2001.
[27] Courant, R., Differential and Integral Calculus. Vol 2. New York: J. Wiley, 1936.
[28] Courant, R., John, F., Introducción al Cálculo y al Análisis Matemático. Vol. 2. México: Limusa,1974.
[29] Lang, S., Calculus of Several Variables. New York: Springer, 1987.
[30] Marsden, J., Tromba, A., Cálculo Vectorial. México: Addison-Wesley, Pearson Educación,1998.
[31] Thomas, G.B., Finney, R.L., Cálculo: Varias Variables. México: Adisson-Wesley Longman,1999.
[32] R. R. Bush and F. Mosteller, Stochastic Models for Learning, Oxford, ed. Martinio Fine Books, 2012, 25-58, ISBN: 1614273197, 9781614273196.
[33] H. M. Taylor and S. Karlin, An Introduction to Stochastic Modeling, Revised Edition, United States of America, Academic Press Inc., 1994 (1984), p.1-49.
[34] J. E. Freund and R. E. Walpole, Estadística Matemática con Aplicaciones, 4ª ed., México, Prentice-Hall Hispanoamericana, S.A., 1990, p.109-111.
[35] M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach, Canada, ed. Dover, 1994 (1981), p.42-47.
[36] G. F. Lawler, Introduction to Stochastic Processes, 2nd ed., United States of America, Chapman & Hall, 2006, p.12.
[37] P. E. Pfeiffer, Concepts of Probability Theory, 2nd ed., New York, ed. Dover, 1978, p.132-144.
[38] D. Freedman, R. Pisani, R. Purves, Statistics, 4th ed., University of California, Berkeley, 2010, p.58-70.
[39] M. Lefebvre, Applied Stochastic Processes, Canada, Springer Science+Business Media LLC, 2007, p.16.
[40] P. W. Glynn, Introduction to Stochastic Modeling, Review of Probability Theory, Stanford University Courses Notes, 2009.
[41] M. R. Spiegel, Theory and Problems of Statistics, Schaum's Outline Series, New York, Schaum Publishing Company, 1961, p.45-73.
[42] E. Parzen, Stochastic Processes, Philadelphia: SIAM, 1999, ISBN: 9781611971125.
[43] Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume IV: Biology and Problems of Health, University of California, ed. Jerzy Neyman, 1954 and 1955, p.83-85.
[44] Polanco C, Castañón-González JA. Cross-sectional bioinformatics analysis on proteins associated to the Ebola virus. Comment on: Hantel A, Olusola-Olopade C. Drug and Vaccine Access in the Ebola Epidemic: Advising Caution in Compassionate Use.Ann Intern Med 2014;162:141-142.
[45] Polanco C, Castañón González JA, Uversky V. Profiling TSOP proteins by Polarity Index Method. Comment on: Dobson A. Mathematical models for emerging disease. Science 2015;347:551-555.
[46] Polanco C, Castañón González JA, Samaniego JL, Buhse T. A novel Mathematical-Computational approach that helps manufacturing new drugs. Comment on: Howard SJ, Hopwood S, Davies SC. Antimicrobial Resistance: A Global Challenge. Sci Transl Med 2014;6:236ed10.