International Journal of General Systems

Publication details, including instructions for authors and subscription information: http://www.informaworld.com/smpp/title~content=t713642931

Description of polypeptides as string systems

Online Publication Date: 01 June 2007
To cite this Article: Majerník, V., (2007) 'Description of polypeptides as string systems', International Journal of General Systems, 36:3, 339 - 345
To link to this article: DOI: 10.1080/03081070601176356
URL: http://dx.doi.org/10.1080/03081070601176356

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007
Description of polypeptides as string systems

V. MAJERNÍK†‡*

†Institute of Mathematics, Slovak Academy of Sciences, SK-81473 Bratislava, Štefánikova 49, Slovak Republic
‡Department of Theoretical Physics, Palacky University, Tr. 17. Listopadu 50, CZ-772 07, Olomouc, Czech Republic

(Received 12 December 2004; in final form 7 November 2006)

In this article the proteins are treated as chains of amino acids modeled by a string system. After a brief review of essential characteristics of a general string system (length, stochasticity, letter utility, etc.) we estimate them for a simple enzyme. Finally, we discuss these results from the point of view of the functional organization of living systems.

Keywords: String system; Polypeptides; Enzyme; Maxwell’s demon

1. Introduction

The biological cell is an immensely complex object, there is a vast amount of biochemical activity occurring at any one time, in general, and in addition there is a great deal of interaction between the cell and its environment, which could be either other cells or outside world. Two types of biomolecules are responsible for the most biological processes in a cell, nucleic acids and proteins both belonging to the set of biopolymers. Proteins are biopolymers constructed from 20 building blocks, the amino acids. Each amino acid consists of a backbone and a side chain. The backbone is the same in all amino acids, but the side chains are different. The arrangement of the building blocks in proteins, the primary sequence, is crucial for their functions. Nucleic acid, the information carriers, are built from four different building blocks, the nucleotides, which form the letters of the genetic alphabet. A nucleotide consists, of a base, sugar and phosphate group are the same for all nucleotides and the four letters are distinguished by four different bases. In DNA the bases are adenine, cytosine, guanine and thymine. DNA carries information (and may be considered as the legislative of a cell) while the proteins perform the activity prescribed by the information in the nucleic acids (and may be considered as the executive of a cell). Both substances take part in the biosynthesis a process in which the nucleic acids store and transport information and direct
the assembling of proteins. Proteins, assembled from amino acids, are the “performers” of live processes which can be often modeled by the cognitive robots.

An important subset of proteins are enzymes whose activity is isomorphic with that of the classical Maxwell demon well-known in thermodynamics. A realistic device performing the activity of Maxwell’s demon (Majerník, 1999) contains two essential units: an informer which gains information of the speeds of molecules and an operator which passes the marked molecules through a microscopic door in the wall dividing a gas cylinder. Although the classical Maxwell demon, decreasing entropy of gas, do not exist in living organism there are many elementary biological processes performed by Maxwell-like systems. To this class of systems belong particularly enzymes which are responsible for a remarkably wide range of biological functions (Parry and Baker 1984). Enzymes are proteins whose goal and function is to catalyze specific biological reactions connected with building of structural elements of organism and with the bioenergetic life processes. A particular enzyme binds specific substrate molecule and catalyzes a specific chemical transformation of the substrate (Monod 1971). The specificity of the activities of enzymes is largely responsible for the high functional organization of living systems. Similarly, as by the classical Maxwell demon, the activity of an enzyme is goal-directed and consists of two fundamental processes: (i) recognition of specific substrate (this corresponds to the informer of Maxwell’s demon), and (ii) catalyzing a specific reaction with it (this corresponds to the operator of Maxwell’s demon). An enzyme represents, therefore, a chemical cognitive robot consisting of a recognition unit (informer of enzyme) and a performing unit (operator of enzyme). The action of an enzyme is functionally isomorphic with that of Maxwell’s demon. This is why enzymes can be modeled by a simple cognitive robot. Since enzymes represent sequences of amino acids they can be modeled by one-dimensional strings. In what follows we briefly review the quantities which describing a general string system and then we try to determine these quantities for simply enzymes.

2. String systems

A string system is a linear one-dimensional sequence of letters ordered stochastically or according to certain rules. The set of all letters forms the alphabet of a string system. For example, the Ising system is a string system formed by a chain of spins (Kubo 1974). Its letters are spins pointing up and spins pointing down. The human language is a string system with an alphabet of about 50 letters (graphemes in the written form or phones in the spoken form). Proteins are string system formed by a chain of 20 amino acids representing their alphabet (Parry and Baker 1984).

The following quantities characterize a string system:

(i) The length of string $n$, i.e. the number of its elements.
(ii) The stochasticity $S$ of a string is given by the Watanabe measure of its configurational organization $W$ defined as (Watanabe 1969):

$$W = \left( \text{sum of entropies of parts of this system} \right) - \left( \text{entropy of whole system} \right).$$

This measure expresses the property of a string systems to have order between its elements. The probability to find a letter given the foregoing ones determines the
stochasticity of a string. If a system consists only of letters which are statistically independent then its Watanabe measure becomes zero. If a string system consists only of elements which are deterministically-dependent, then its Watanabe measure becomes maximum. Generally, the configurational organization of a string system lies between these extremes.

(iii) The number of the functional strings \(N_f\), i.e. strings which perform a function in a string system, such are, e.g. the combinations of graphemes which express words in a language.

(iv) The number of non-functional string \(N_n\), such are, e.g. the combinations of graphemes which does not express words in a language.

(v) The probability \(P\) of assembling a functional string of given length by a random combination of its letters.

(vi) The mean time \(T\) for finding a functional string among all strings. To determine \(T\) we use the urn model. We suppose that in an urn there are all strings (functional and non-functional) and one randomly withdraws a string. Let this procedure lasts a time interval \(\Delta t\). If one, after a unsuccessful attempt to withdraw a functional string, discards it from the set all strings, and then continues to withdraw the next string, then the mean time \(T\) for finding of a functional string is given by a series

\[ T = P N \Delta t + P (N-1) 2 \Delta t + \cdots + P N_f (N - N_f) \Delta t, \]

where \(P = N_f / N, P (N-1) = N_f / (N - 1), \ldots P (N - i) = N_f / (N - i)\) are the probabilities for withdrawing a functional string in \(i\) attempts. Hence, we have

\[ T = \Delta t N_f \sum_{i=0}^{i=N-N_f} \frac{i+1}{N-i} \]  

(1)

(vii) The mean time \(Q\) for finding a functional string among all strings if these strings are assembled by a set of appropriate “pre-fabricated” substrings. Consider a string which is assembled from a set of substrings of length \(x\) containing all parts of the complete string. When these substrings are assembled simultaneously then the mean time for finding a functional string is given as the sum of time for forming a substring \(T^{(s)}\) and that of their combination \(T^{(c)}\)

\[ Q = T^{(s)} + T^{(c)} = \left[ \frac{(x^{20} + 1)}{2} + \frac{(Y/x)! + 1}{2} \right] \Delta t, \]  

(2)

where \(Y\) is length of the whole string system. Taking \(x\) as continuous variable we can find the length of substrings which minimalizes \(Q\). The necessary condition for this length is

\[ \frac{dQ}{dx} = 10x^{19} - \frac{Y \text{ Gamma}(1 + (Y/x)) \text{ PolyGamma}(0, 1 + (Y/x))}{2x^{2}} = 0, \]

where Gamma[z] is the Euler gamma function and PolyGamma[n, z] gives the \(n\)-th derivation of the digamma function (Wolfram 1991).

(viii) The letter utility \(K\) expresses the effectiveness of the letter utilization in a string system. If we denote by symbol \(N^{(o)}\) and \(N^{(f)}\) the number of all possible strings of length \(n\) and number of the functional strings of length \(n\), respectively then an appropriate quantitative characteristic of the system’s effective use of its letters is
called its letter utility $K$ and given by the formula (Majerník 1998)

$$K = \frac{1}{\ln 2} \ln \left(1 + \frac{N_f(n)}{N(n)}\right).$$

(3)

$K$ assumes values from the interval $[0, 1]$. If all possible strings of length $n$ are functional, then this system has letter utility $K = 1$, if a system has no functional strings its letter utility is $K = 0$. For example, a string system whose elements are figures has utility $K = 1$, because any string of figures represents a number. The letter utility for two-letter words in English is $K \approx 0.012$.

(ix) The information content ($I$) of a string.

A typical string system with functional and non-functional string is the set of polypeptides of length $n$. Next, we will try to determine the quantitative characteristics of enzyme which represent a subset of functional polypeptides.

3. The estimation of the system properties of an enzyme

Let denote the number of possible polypeptides consisting of $n$ letters (amino amides) as $N_p$ and that of functional polypeptides (proteins) as $N_f$. As it is well-known, only some $10–20\%$ of the amino acids comprising an enzyme are immutable for enzyme activity. The other amino acids can be changed by random mutations without changing the biochemical effect of an enzyme. This means that if we take an enzyme assembled from $2 \times 10^3$ amino acids then 200 to 400 are immutable for its enzyme activity. All possible polypeptides arisen by junction of the amino acids of length $n = 200$ and $n = 400$ is

$$N_{200} = 20^{200} = 1.6 \times 10^{600}$$

and

$$N_{400} = 20^{400} = 2.56 \times 10^{520},$$

respectively.

In order to separate enzymes from a set of all polypeptides a selection process has to take place which can be realized by the Maxwell-like device where the separation, as a sorting process, is performed in two steps: (i) the recognition of the selected object, and (ii) the proper physical separation of this object. The total time for separation of enzymes from the set of all polypeptides $N$ is

$$T_s = (\Delta t_r + \Delta t_s)N,$$

where $\Delta t_r$ is time for recognition and $\Delta t_s$ that of separation. While the separation time might be unlimited, there is a principal quantum mechanical–mechanical ground for the minimal time for recognition of an enzyme macromolecule among the set of all polypeptides molecules. According to the time-energy uncertainty relation, we have (Messiah 1961)

$$\Delta t \Delta E \geq \hbar,$$

where $\Delta t$ is the time interval and $\Delta E$ is the uncertainty in energy ($\hbar$ being the Planck constant). Inserting the rest energy $E_m$ assigned to the mass of a macromolecule of an enzyme into equation (3) we get the minimal time interval $\Delta t$ for which the energy uncertainty $\Delta E$ is approximately equal to $E_m$. In such a situation no recognition is possible. Since the rest energy of an enzyme molecule is approximately $E_m \approx 10^5$ erg the minimal recognition time of a polypeptides molecule is $\Delta t \approx 10^{-30}$ s. The total recognition time $\Delta T$ of all polypeptides is $\Delta T_{200} \approx 10^{230}$ s for $n = 200$ and $\Delta T_{400} \approx 10^{490}$ s for $n = 400$, respectively.
Let us now estimate the string characteristics, presented in previous section, of a typical enzyme:

(i) The sequence of an enzyme in which amino acids are immutable forms a deterministic string with the maximal Watanabe organization $W_{200} \approx \ln 10^{200}$ and $W_{400} \approx \ln 10^{520}$, respectively.

(ii) The probability of assembling an enzyme by a random way is $P_{200} \approx 10^{-260}$ and $P_{400} \approx 10^{-520}$, respectively.

(iii) The mean time of a random finding of an enzyme in the set of all polypeptides when discarding the non-functional ones, given by equation (1), is $T_{200} \approx 10^{46} \Delta t$ and $T_{400} \approx 10^{56} \Delta t$, respectively.

(iv) The mean time $Q$ for a random finding of an enzyme among all polypeptides assembled from a set of subpeptides, forming parts of the whole enzyme, as a function of the length of a subpeptide $x$ is given in Table 1.

The smallest value of $Q$ in Table 1 is $Q_{200}^{(\text{min})} = 5.12 \times 10^{19}$ for $x = 10$ and $Q_{400}^{(\text{min})} = 6.31 \times 10^{24}$ for $x = 18$, respectively.

The dependence of $Q$ on $n$ for $n = 200$ and $n = 400$ is depicted in figure 1 (curve a–c). We see that the minimal $Q$ can be determined as the solution of equation (2). The numerically determined solutions of equation (2) gives the minimal length of a subpeptide. We get $x = 9.3$ for $Y = 200$ and $x = 16.65$ for $Y = 400$, respectively.

![Figure 1](image-url)

Figure 1. Graphical representation of the mean time for assembling $Q$ as a function of $x$ for $Y = 200$ (curve a), $Y = 400$ (curve b) and $Y = 600$ (curve c).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$n = 200$</th>
<th>$n = 400$</th>
<th>$n = 200$</th>
<th>$n = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$4.66 \times 10^{37}$</td>
<td>$3.94 \times 10^{37}$</td>
<td>$4.18 \times 10^{22}$</td>
<td>$1.041 \times 10^{30}$</td>
</tr>
<tr>
<td>4</td>
<td>$1.52 \times 10^{64}$</td>
<td>$4.66 \times 10^{57}$</td>
<td>$6.04 \times 10^{23}$</td>
<td>$8.36 \times 10^{24}$</td>
</tr>
<tr>
<td>6</td>
<td>$1.40 \times 10^{77}$</td>
<td>$4.48 \times 10^{93}$</td>
<td>$6.31 \times 10^{24}$</td>
<td>$6.37 \times 10^{25}$</td>
</tr>
<tr>
<td>8</td>
<td>$75 \times 10^{24}$</td>
<td>$1.52 \times 10^{34}$</td>
<td>$5.24 \times 10^{25}$</td>
<td>$5.24 \times 10^{25}$</td>
</tr>
<tr>
<td>10*</td>
<td>$5.12 \times 10^{40}$</td>
<td>$4.07 \times 10^{77}$</td>
<td>$2.09 \times 10^{28}$</td>
<td>$2.09 \times 10^{28}$</td>
</tr>
<tr>
<td>12</td>
<td>$1.91 \times 10^{31}$</td>
<td>$1.40 \times 10^{37}$</td>
<td>$8.38 \times 10^{30}$</td>
<td>$8.38 \times 10^{30}$</td>
</tr>
</tbody>
</table>
The dependence of the minimal number of the amino acids in a subpeptide $x_{\text{min}}$ on the whole number of polypeptides $Y$ can be represented by the linear function $x_{\text{min}} = 0.03675Y + 1.96$ shown in figure 2.

(v) The letter utility for amino acids in polypeptides (taking 10,000 functional polypeptides in the set of all polypeptides is $K_{200} \approx 10^{-256}$ is $K_{400} \approx 10^{-516}$, respectively.

All the estimated string characteristics of an enzyme are summarized in Table 2.

4. Conclusions

From what has been said as far it follows:

(i) The probability of assembling of a functional polypeptide (an enzyme) is incredible small.

(ii) The mean time for finding an enzyme among all possible polypeptides is enormous large. To make an image of it we assume that in $T$ we take the shortest astrophysical time interval—Planck time—$\Delta t = t_p = 4.310^{-35}$ s (Weinberg 1972) then the mean time reaches the values $\hat{T}_{200} = 1.85.10^{225}$ s or $\hat{T}_{400} = 6.10^{485}$ s, respectively. Thus, a time which is about $10^{212}$ or $10^{457}$ time longer than the age of the Universe.

(iii) Some orders of magnitude smaller we get values for $T_{200}$ and $T_{400}$.

Table 2. Statistical properties of enzyme.

<table>
<thead>
<tr>
<th></th>
<th>$n = 200$</th>
<th>$n = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>$\approx \ln 10^{260}$</td>
<td>$\approx \ln 10^{520}$</td>
</tr>
<tr>
<td>$P$</td>
<td>$\approx 10^{-260}$</td>
<td>$\approx 10^{-520}$</td>
</tr>
<tr>
<td>$T$</td>
<td>$\approx 10^{48} \Delta t$</td>
<td>$\approx 10^{56} \Delta t$</td>
</tr>
<tr>
<td>$x_{\text{min}}$</td>
<td>9.3</td>
<td>16.65</td>
</tr>
<tr>
<td>$Q_{\text{min}}$</td>
<td>$1.13.10^{21} \Delta t$</td>
<td>$2.00.10^{93} \Delta t$</td>
</tr>
<tr>
<td>$K$</td>
<td>$\approx 10^{-256}$</td>
<td>$\approx 10^{-516}$</td>
</tr>
<tr>
<td>$I$</td>
<td>152.8 bit</td>
<td>172.8 bit</td>
</tr>
<tr>
<td>$N$</td>
<td>$\approx 10^{260}$</td>
<td>$\approx 10^{520}$</td>
</tr>
</tbody>
</table>
(iv) Still smaller value we get for \( Q \) (see Table 1). However, here it is supposed that in the reaction environment the subpeptides, needed for the construction of a polypeptide, occur in exact given rates.

(v) The letter utility \( K \) of a set of polypeptides is enormous small. Hence, using the amino acids as letters forming a polypeptide string is extremely ineffective in comparison, e.g. with languages.

(vi) The information content of a polypeptide with 200 or 400 amino acids is \( I_{200} = 152.8 \) bit or \( I_{400} = 172.8 \) bit, respectively.

(vii) The total number of polypeptides containing \( n \) amino acids is extremely large.

The necessary condition of origin of life is the existence of reproductive cell in which biosynthesis takes place. The error-free assembling of proteins by DNA is not possible without enzymes. Without their activities the assembled proteins would quickly become a non-functional polypeptides and the living processes in cell would soon die out. If we assume that enzymes arose by a random way we soon arrive at a deadlock. It is clearly inconceivable that enzymes of a determined amino-acid sequences have been formed by mere chance; there is neither sufficient time nor matter for such a random formation. Rather, it seems very probably that the formation of enzyme molecules can only be understood within the already existing high functional organization of the whole living object.

References


**Vladimír Majerník** is a professor of mathematical physics at the Mathematical Institute of Slovak Academy of Sciences, Bratislava, Slovak Republic. In years 1960–1973 he was with the Department of Acoustics and with Department of Theoretical Physics in the Institute of Physics of the Slovak Academy of Sciences in Bratislava. He received the degree in mathematical physics from the University of Jena (Germany), a PhD in Theoretical Physics from the Comenius University of Bratislava and Dr.Sc. degree from the Czech Technical University of Prague in 1979. He taught quantum and statistical physics at the Palacky University in Olomouc (Czech Republic) and music acoustics in the College of Fine Arts of Bratislava. His main research interests are in mathematical quantum physics, mathematical theory of music and mathematical biophysics.