The set of chromosomes as potential source of information about next steps to cell of structure‘s sequence of proteins. (optimal model of picture)

S.Y. Ruderman
Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia

O.S. Nurgayanova
Computer Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
e-mail: onurgayanova@yandex.ru

Abstract[1]

The cell of any organism consist of complex system, which fulfil a lot of important functions for life. It is considered, that number chromosomes in the cell of each species is own, the number of messages in nucleic acid is 4, the number of proteins is 20. Evidently that these constants are organized, but their nature and depending factors are unknown. One of the most important factor is variety of opportunities. It’s rather good that the variety of protein’s sequences which is able to be built was the better the more but “energy” costs the better the less in this process and I’d like the reliability of the process is as much as better. In this article it is built and analyzed the model of cell’s structure to the proper protein sequence. It is considered that the process of constructing next protein consists of two phases:
1) the searching for chromosomes and the code of protein;
2) the process of searching the code in the found chromosome (and in the constructer of the protein).

Specially putting the problem let us asking the question about the optimum number of chromosomes is twenty three (23). It is observed that in the cell of human only 23 pairs of chromosomes, let us believe that this construction consists of main features of complex observed reality help to find its physic-chemical compositions.
Keywords: cell, protein, protein sequences, chromosome, random choice, DNA (deoxyribonucleic acid), searching, variety, reliability, source.

1. Introduction

Searching for chromosome

Let me sign via “n” the quantity of chromosome’s pair in the cell. Each of them has DNA, which contains its own number of protein’s codes. We should find one of the “n” chromosomes which has the needed protein code. Seeking needed chromosome we’ll write as random choice without recurrence from possible ‘*n’ to finding necessary*. On each step of this choice which is 1/*n* will be got necessary chromosome (it is known that in case of random choice from ‘*n’ parts* unconditional probability of taking the needed element from the i-th extraction not depending on number of ‘i’ is 1/*n*). The number of steps in following search is random quantity with 1,2,3,…,n sequence. The special probability is belonged to all numbers of sequence which is 1/n, that’s why the mathematical expectation of steps’ number for seeking needed chromosome is:

$$1\cdot\frac{1}{n}+2\cdot\frac{1}{n}+3\cdot\frac{1}{n}+...+n\frac{1}{n}=\frac{1}{n}(1+2+...+n)=\frac{n+1}{2}\cdot \quad (1)$$

Each act of chromosome’s random choice and definition of found one required special energy of any resource. We’ll say about physical illustration of this resource later but now we’ll guess that each step of described procedure about random choice, consisted of the unit of this resource and then the expectation of needed resource of searched chromosome will have the form (1).

Seeking protein code in chromosome

After found chromosome which kept own code it’s necessary to seek this code. It is known that each chromosome has own DNA cell which keep a lot of protein codes as a rule, seeking needed code required energy. We guess that the energy, necessary for seeking needed chromosome and seeking protein code, spent this way: after needed chromosome is found (it’s necessary the resource, which expectation has form (1)), the process, called as multistep game for surviving in the theory of chances [1], is begun that’s on each step the resource decreasing on one unit with probability is ½ or increasing on one unit with same probability. The process of this random changing goes on until the resource either grows up to the value after work with chromosome is ended (it means “the victory” in this game that’s chromosome is found),or it’s 0 (it means “losing” in this game that’s the code of needed protein is not found and we’ll guess the whole needed chain of proteins will not built, it means that the length of protein’s chain is 0).

Proceedings of the 3rd International Workshop "Intelligent Technologies for Information Processing and Management", November 10 - 12, Ufa, Russia, 2015

Next chapter has the information about seeking processes of each needed chromosome and its own protein code which will be both discussed and we'll have the opportunity to solve the problem of the optimal number of chromosomes. Optimized model of seeking chromosomes and its own protein code

Total pairs of chromosomes will sign as symbol *n*. The number of proteins which cell will synthesized sign as *L*. The variety of possible choices from *n* chromosomes necessary for creating of *L* protein sequence is n^L. The quantity $K=n^L$ should be as much as possible.

For increasing *K* for a given *L* the quantity *n should be bigger*, but if *n increase* then the expectation of resource necessary for seeking needed *chromosome having form (1)* will also increase.

We can sign general resource which the cell has for seeking the sequence of protein codes as *B*. For real *L* original resource is $x=\frac{B}{L}$ which necessary for seeking *chromosome* and work with it.

Then it means that the process of seeking each chromosome and following searching its own protein code has following form.

During process of seeking first chromosome each resource is used for all chromosomes. That's why each *L* equal resource for seeking chromosome has first protein code will be $x-Y_1=\frac{B}{L}-Y_1$, where Y_1 – random quantity with expectation (1).

After seeking chromosome which has first *L* protein code it's begun the seeking of first protein code which interprets the surviving game including the meaning of original resource $x=\frac{B}{L}$ that means the victory in game (seeking the first protein code) or it is 0 that means losing the game or in the end of the end the length of protein sequence is 0.

In case of victory in game related to the seeking of first protein code the resource which belongs to the seeking of first protein code go away from game. In further process the resources $\frac{B}{L}-Y_1$ are involved and belonged to work of seeking second, third till L chromosomes.

After seeking second needed chromosome and its special protein code (in case of the seeking is successful), the resource for the seeking of each leftover chromosome will be $\frac{B}{L}-Y_1-Y_2$, where Y_1, Y_2 are random quantity with expectation (1).

After the seeking i-st chromosome for its i-st protein code the resources for *i-st*, *i+1*-st, …, *L*-st will be $\frac{B}{L}-Y_1-Y_2-\ldots-Y_i$. The successful seeking *i*-st protein code is equal to completing of symmetric chaotic straying beginning in point $\left(x-\sum_{k=1}^{i}Y_k\right)$ with the absorbing screen $\left(x-\sum_{k=1}^{i-1}Y_k\right)$, another screen is located in point 0. Probability of successful game over is [1]

$$\frac{x-\sum_{k=1}^{i}Y_k}{x-\sum_{k=1}^{i-1}Y_k}. \quad (2)$$

It is important that in our problem the quantities Y_1, Y_2, … , Y_i are random. The successful seeking i-st protein code is the occasion which we sign as A_i. The composition of *L* protein sequence is the chain of occasions $A_1A_2\ldots A_L$.

Foregoing that if any protein code will not be found then the whole length of chain of protein codes will be 0. That's why the whole group of incompatible occasions consists of chains A_1, A_1A_2, $A_1A_2A_3$, … , $A_1A_2\ldots A_{L-1}$ A_L, A_L, $A_1A_2\ldots A_L$.

Considering that $Y_1, Y_2, \ldots, Y_L$ are random numbers, the probability of getting chain $A_1A_2\ldots A_L$ is:

$$P(A_1A_2\ldots A_L)=M\left(\frac{x-Y_1}{x}\cdot\frac{x-Y_1-Y_2}{x-Y_1}\cdot\ldots\cdot\frac{x-Y_1-Y_2-\ldots-Y_L}{x-Y_1-Y_2-\ldots-Y_{L-1}}\right)=1-\frac{n+1}{2}\cdot\frac{L}{x}, \quad (3)$$

where *M* is the symbol of expectation:

$$M\left(\frac{x-Y_1-Y_2-\ldots-Y_L}{x}\right)=M\left(1-\frac{n+1}{2}\cdot\frac{L}{x}\right)=1-M\left(\frac{n+1}{2}\cdot\frac{L}{x}\right). \quad (4)$$

As the probable meanings of found protein codes are 0 and *L* then the expectation of the difference between *L* and random chain is being the characteristic of the inaccurate of chain composition and is

$$\Delta L=L^3\cdot\left(\frac{n+1}{2B}\right)=\frac{L^3(n+1)}{2B}=\frac{n+1}{2}\cdot\frac{\ln^3 K}{\ln^3 n}. \quad (5)$$

Assuming that *Z is* the random composed chain of proteins. Possible meanings are *Z* – 0 and *L*. Their varieties are

$$P(Z=L)=P(A_1A_2\ldots A_L)=1-\left(\frac{n+1}{2}\cdot\frac{L}{x}\right).$$

$$P(Z=0)=1-P(A_1A_2\ldots A_L)=1-1+\frac{n+1}{2}\cdot\frac{L}{x}=\frac{n+1}{2}\cdot\frac{L}{x} \quad (6)$$

The expectation *Z* is:

$M(Z)=0\cdot P\left(\frac{n+1}{2}\cdot\frac{L}{x}\right)+L$

$$\left(1-\frac{n+1}{2}\cdot\frac{L}{x}\right)=L-L\left(\frac{n+1}{2}\cdot\frac{L}{x}\right)=L-\frac{n+1}{2}\cdot\frac{L^2}{x} \quad (7)$$

The necessary for us characteristic of inaccurate of composing protein codes sequence process is

$$\Delta L=M(L-Z)=L-M(Z)=L-L+\left(\frac{n+1}{2}\cdot\frac{L^2}{x}\right)=\frac{n+1}{2}\cdot\frac{L^3}{B}=\frac{(n+1)\ln^3 K}{2B\ln^3 n} \quad (8)$$

We indicate three optimized problems about quantity n having the equal decisions. To find the number of chromosomes n which:

1) for definite B and K minimizes ΔL;

2) for definite B and ΔL maximizes variety of K;

3) for definite K and ΔL minimizes B.

In each of these three problems the required number n minimizes

$$\frac{n+1}{\ln^3 n}, \quad (9).$$

where n is positive integer. The meanings of (9) for possible positive integer n are interpreted in this chart where minimum (9) is reached for n=23.

N	2	3	12	17	20	21	22	23	24	25	26
$\frac{n+1}{\ln^3 n}$	9,09	3	0,85	0,79	0,78	0,779	0,778	0,776	0,778	0,25	0,78

The fact that the human cell involves only twenty-three pairs of chromosomes let us assume that this model consists of main features of complex reality and can help us to seek its physicochemical basement.

Let take a look on the questions of realism of this model. The assumption of random selection without coming back from available chromosomes to seeking needed one which has its protein code looks physically natural. The next behavior of resource for seeking the protein code in chromosome («survival game» when this resource like zip level from vertical clasp is down on random quantity for seeking chromosome and then a little bit hesitated up to original level) can be unnatural [2]. Rather interesting for spectacular of our usual macrospace is discrete change of resource on each step for plus-minus unit (with probability ½)).

In quantum microspace the same changes are «natural». In the particular case the electron with spin ½ connected to magnetic moment is vector with projection on the external magnetic field is random quantity with two equal on absolute quantity but with different signs values. If spin is isolated from the external magnetic field then the chances of these values are equal and they are ½.

Possible interpretation of «resource» by the projection of magnetic moment with special system of spins on definite directions is on [4] for the analysis of protein formation's model. We can guess that the system of spins is complex part of any structure same as empirical ones in organic magnets [3]. The last ones have the radicals with carbon atoms including unpaired electrons, take fixed ordered orientation with connected influence on alkali metal cations between them. Unpaired electrons interact with each others with metal cations, created large total magnetic moment.

We can guess that «the resource» in this model is connected with the magnetic moment of special spin's systems then its dependence on external current magnetic field is big favor for us. The understanding of it let us quickly influence on this system the magnetic moment of which also influence on efficiency of the synthesis sequence of proteins.

We hypothesize that in the basis of positive act of appropriate current magnetic field on the stamina of organisms is belonged the process which describing was in [2] for protein synthesis but also in this article the describing of the composition by the cell of needed protein sequence.

The mechanism of «disease» and «treatment» (on cell level) is following: if cell has «filthy» substances, which we can't delete, minimizes internal magnetic moment. We can compensate lack of magnetic field's numbers with immersing the cell in current magnetic field. The reduction of its magnetic moment to needed quantity leads to normal protein synthesis and proper protein sequence. The quantity of external magnetic field, which is able to treat, is related to the quantity of the magnetic field of cell nature.

If the value of EMF is large then induced moment is too large that can lead to big deceleration of protein synthesis [4].

This fact is rather interesting for oncology [5]: if the action on cancer cells is induced large magnetic moment by strong magnetic field then synthesis is slowdown as each protein [6] as needed protein sequence.

The cell of each multicellular organisms has its own number of chromosomes, in which DNA is written protein codes. That bee and pigeon have 8 pairs, butterfly has 190, bear – 37, fly – 6, cucumber – 12, rabbit – 22, gorilla and potato – 24, horse – 32.

Trying to understand the reasons of difference of these quantities from optimal number which is 23 which is led by the model of forming protein's code sequence we need to explain the faults all variety of answers. Reflecting on this subject and wanting to preserve the most significant in the assumptions made, we came to the conclusion that the possibility of failure for different types of assumptions about random sampling without replacement from the set of chromosomes to find the desired. It is considered that this choice is «adorable» be course of the expectation of the needed steps' chromosomes isn't $\frac{n+1}{2}$ but is $\alpha n+\beta$,

where α and β describes «desire» of species for seeking needed chromosome.

If numeral characteristic of human cell is $\Delta L = \frac{n+1}{2} \cdot \frac{\ln^3 K}{B\sqrt[3]{n}}$, then now it is:

$$(\alpha n + \beta) \cdot \frac{\ln^3 K}{B\sqrt[3]{n}} . \quad (9)$$

The quantity of chromosomes n is the quantity which minimize equality (9) coefficients α и β show the degree of species' «desire» for seeking chromosome.

If we change *n in (9)* on number of considered chromosomes in cell n_0, then we almost seek dependence on α and β which minimize the value $\frac{\alpha n + \beta}{\sqrt[3]{n_0}}$. Further analysis can help us to see all processes occurring in cell.

References

1. Feller W. Introduction in the theory of variety and its features. In two books,. T. 1; English translation., – M. Mir. 1984. – 528 c.
2. S.Y Ruderman, I.A. Solomeshch. About optimal sizes of alphabets and possible reason of their observability in nature. The review of Industrial and Applied Mathematics. Book 6, 2nd issue. Moscow 1999г.
3. Dirak P.A.M. The ways of physics – M.: EnergoAtomIzdat, 1983 y. – 88p.
4. S.Y. Ruderman, I.A. Solomeshch. The problem of alphabet volume selection in biopolymers and statistical thermodynamics. Bashkir Journal of Chemistry. The issue «Reaktiv», 1996 y., b.3, №5-6, p.37–44.
5. Lawrence R., Rosh P. J., Plowden D. Magnetic therapy . Alternative method of painkilling/English translation S. Yevtushenko. – M.: KRON-PRESS, 1998. – 234 p.
6. Sei-Hum Jang, Ruth Ann Bertsch, James E. Jackson, Bart Kahr. // Mol. Crystanol Liquid Cryst. – 1992. V.211. – P.289.