

## Molecular Biocoding of Insulin – Amino Acid Ser

Lutvo Kurić

The School of Economics and Business, University of Sarajevo, kalinska 7/6, novi travnik, SBK, Bosnia and Herzegovina

### Abstract

The modern science mainly treats the biochemical basis of sequencing in bio-macromolecules and processes in medicine and biochemistry. One can ask whether the language of biochemistry is the adequate scientific language to explain the phenomenon in that science. Is there maybe some other language, out of biochemistry, that determines how the biochemical processes will function and what the structure and organization of life systems will be? The research results provide some answers to these questions. They reveal to us that the process of sequencing in bio-macromolecules is conditioned and determined not only through biochemical, but also through cybernetic and information principles. Many studies have indicated that analysis of protein sequence codes and various sequence-based prediction approaches, such as predicting drug-target interaction networks [14], predicting functions of proteins [15,18], analysis and prediction of the metabolic stability of proteins [16], predicting the network of substrate-enzyme-product triads [7], membrane protein type prediction [1,2,5], protein structural class prediction [4,12], protein secondary structure prediction [6,11], enzyme family class prediction [3,11], identifying cyclin proteins [20], protein subcellular location prediction [9,10,17,19], among many others as summarized in a recent review [15], can timely provide very useful information and insights for both basic research and drug design and hence are widely welcome by science community. The present study is attempted to develop a novel sequence-based method for studying insulin in hopes that it may become a useful tool in the relevant areas.

**Keywords:** Human insulin; Insulin model; Insulin code; Genetics code; Amino acid ser

### Introduction

The biologic role of any given protein in essential life processes, eg, insulin, depends on the positioning of its component amino acids, and is understood by the “positioning of letters forming words“. Each of these words has its biochemical base. If this base is expressed by corresponding discrete numbers, it can be seen that any given base has its own program, along with its own unique cybernetics and information characteristics.

Indeed, the sequencing of the molecule is determined not only by distinct biochemical features, but also by cybernetic and information principles. For this reason, research in this field deals more with the quantitative rather than qualitative characteristics of genetic information and its biochemical basis. For the purposes of this paper, specific physical and chemical factors have been selected in order to express the genetic information for insulin. Numerical values are then assigned to these factors, enabling them to be measured. In this way it is possible to determine if a connection really exists between the quantitative ratios in the process of transfer of genetic information and the qualitative appearance of the insulin molecule. To select these factors, preference is given to classical physical and chemical parameters, including the number of atoms in the relevant amino acids, their analog values, the position in these amino acids in the peptide chain, and their frequencies. There is a large number of these parameters, and each of them gives important genetic information. Going through this process, it becomes clear that there is a mathematical relationship between quantitative ratios and the qualitative appearance of the biochemical “genetic processes“ and that there is a measurement method that can be used to describe the biochemistry of insulin.

### Methods

The biologic role of any given protein in essential life processes, eg, insulin, depends on the positioning of its component amino acids, and is understood by the “positioning of letters forming words“. Each of these words has its biochemical base. If this base is expressed by

corresponding discrete numbers, it can be seen that any given base has its own program, along with its own unique cybernetics and information characteristics. Indeed, the sequencing of the molecule is determined not only by distinct biochemical features, but also by cybernetic and information principles. For this reason, research in this field deals more with the quantitative rather than qualitative characteristics of genetic information and its biochemical basis. For the purposes of this paper, specific physical and chemical factors have been selected in order to express the genetic information for insulin. Numerical values are then assigned to these factors, enabling them to be measured. In this way it is possible to determine if a connection really exists between the quantitative ratios in the process of transfer of genetic information and the qualitative appearance of the insulin molecule. To select these factors, preference is given to classical physical and chemical parameters, including the number of atoms in the relevant amino acids, their analog values, the position in these amino acids in the peptide chain, and their frequencies. There is a large number of these parameters, and each of them gives important genetic information. Going through this process, it becomes clear that there is a mathematical relationship between quantitative ratios and the qualitative appearance of the biochemical “genetic processes“ and that there is a measurement method that can be used to describe the biochemistry of insulin.

Insulin can be represented by two different forms, ie, a discrete form and a sequential form. In the discrete form, a molecule of insulin is represented by a set of discrete codes or a multiple dimension vector. In the sequential form, an insulin molecule is represented by a series of

**\*Corresponding author:** Lutvo Kurić, The School of Economics and Business, University of Sarajevo, kalinska 7/6, novi travnik, SBK, Bosnia and Herzegovina, Tel: 0387 61 763 917; Fax: 0387 30 525 064; E-mail: [lutvokuric@yahoo.com](mailto:lutvokuric@yahoo.com)

Received Jan 16, 2011; Accepted July 22, 2011; Published July 26, 2011

**Citation:** Kurić L (2011) Molecular Biocoding of Insulin – Amino Acid Ser. J Bioengineer & Biomedical Sci 1:102. doi:10.4172/2155-9538.1000102

**Copyright:** © 2011 Kurić L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

amino acids according to the order of their position in the sequence length 51 AA.

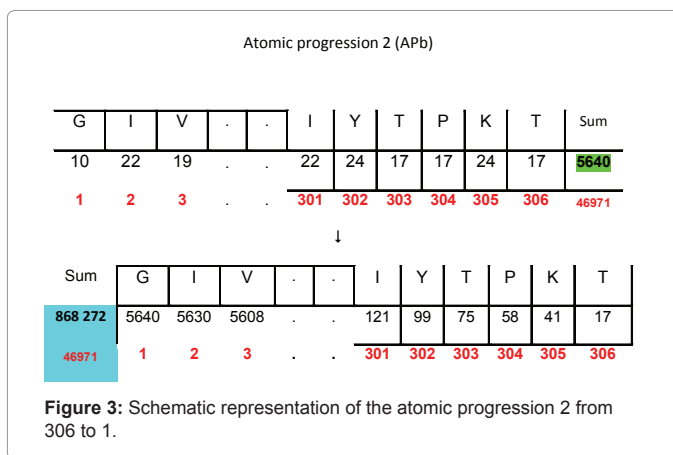
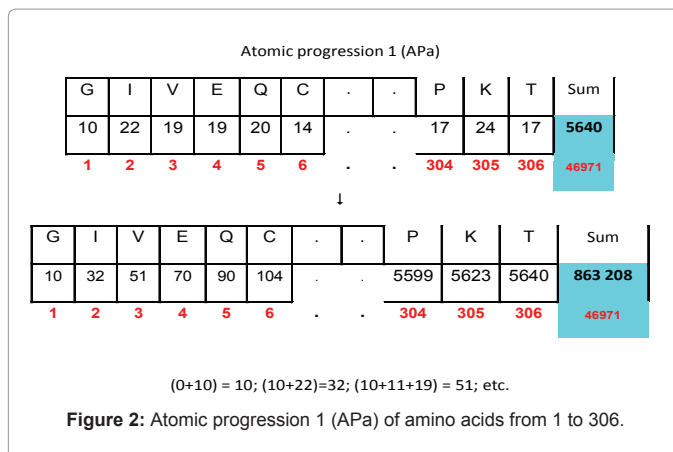
Therefore, the sequential form can naturally reflect all the information about the sequence order and length of an insulin molecule. The key issue is whether we can develop a different discrete method of representing an insulin molecule that will allow accommodation of partial, if not all sequence order information? Because a protein sequence is usually represented by a series of amino acids should be assigned to these codes in order to optimally convert the sequence order information into a series of numbers for the discrete form representation?

### Results

The matrix mechanism of Insulin, the evolution of biomacromolecules and, especially, the biochemical evolution of Insulin language, have been analyzed by the application of cybernetic methods, information theory and system theory, respectively. The primary structure of a molecule of Insulin (Figure 1) is the exact specification of its atomic composition and the chemical bonds connecting those atoms.

Notes: Aforementioned aminoacids are positioned from number 1 to 306. Numbers 1, 2, 3, n... present the position of a certain aminoacid. This positioning is of the key importance for understanding of programmatic, cybernetic and information principles in this protein. The scientific key for interpretation of bio chemical processes is the same for insulin and as well as for the other proteins and other sequences in biochemistry.

The first aminoacid in (Figure 2,3) example has 10 atoms, the second one 22, the third one 19, etc. They have exactly these numbers of atoms because there are many codes in the insulin molecule, analog codes, and other coded features. In fact, there is a cybernetic algorithm which it is „recorded“ that the first amino acid has to have 10 atoms, the second one 22, the third one 19, etc. The first amino acid has its own biochemistry, as does the second and the third, etc. The obvious conclusion is that there is a concrete relationship between quantitative ratios in the process of transfer of genetic information and qualitative



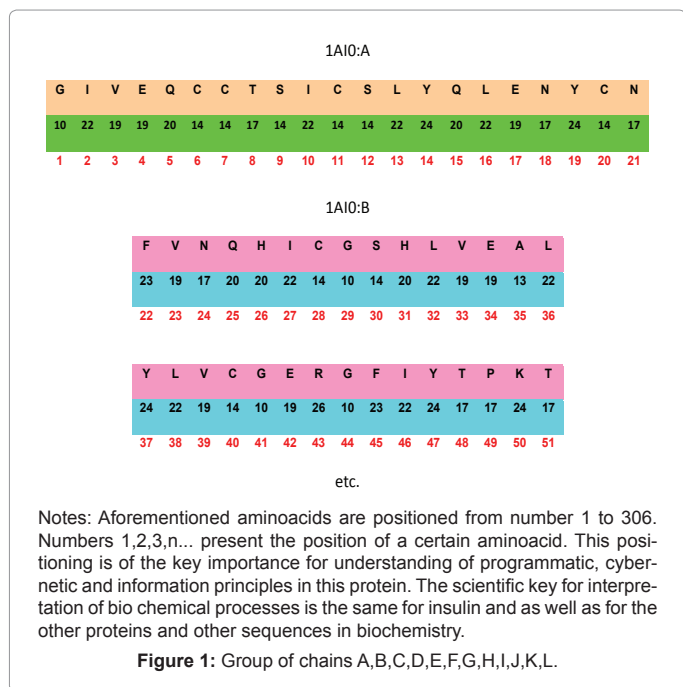
The structure 1A1O – Amino acid Ser									
	S	S	S	S	S	S	S	S	S
Number of atoms	14	14	14	14	14	14	14	14	14
Rank	9	12	30	60	63	81	111	114	132
APa	149	199	537	1089	1139	1477	2029	2079	2417
APb	5505	5455	5117	4565	4515	4177	3625	3575	3237
AP(a,b)	5654	5654	5654	5654	5654	5654	5654	5654	5654
	S	S	S	S	S	S	S	S	S
Number of atoms	14	14	14	14	14	14	14	14	14
Rank	162	165	183	213	216	234	264	267	285
APa	2969	3019	3357	3909	3959	4297	4849	4899	5237
APb	2685	2635	2297	1745	1695	1357	805	755	417
AP(a,b)	5654	5654	5654	5654	5654	5654	5654	5654	5654

**Table 1:** Schematic representation of the atomic progression APa and APb (Amino acid Ser – position from 1 to 306 AA).

appearance, ie, the characteristics of the organism.

### Algorithm

We shall now give some mathematical evidences that will prove that in the biochemistry of hemoglobin in there really is programmatic and cybernetic algorithm in which it is “recorded“, in the language of mathematics, how the molecule will be built and what will be the quantitative characteristics of the given genetic information.



**Atomic progression**

Step 1 (Amino acids from 1 to 306)

$AC_1 = 10$  atoms;  $AC_2 = 22$  atoms;  $AC_3 = 19$  atoms;...  $AC_{306} = 17$  atoms;

$$[AC_1 + (AC_1 + AC_2) + (AC_1 + AC_2 + AC_3)... + (AC_1 + AC_2 + AC_3... + AC_{147})] = S_1;$$

$$AC_1 = APa1 = 10;$$

$$(AC_1 + AC_2) = (10+22) = APa2 = 32;$$

$$(AC_1 + AC_2 + AC_3) = (10+22+19) = APa_3 = 51;$$

$$(AC_1 + AC_2 + AC_3... + AC_{306}) = AP_{a306} = 5640 \text{ atoms};$$

$AP_{a1,2,3,n}$  = Atomic progression of amino acids 1,2,3,n

$$[AP_{a1} + AP_{a2} + AP_{a3}... + AP_{a306}] = (10+32+51... + 5640) = S_1;$$

$$S_1 = 863\ 208;$$

**Notes:** By using chemical-information procedures, we calculated the arithmetic progression for the information content of aforementioned aminoacids.

Step 2 (Amino acids from 306 to 1)

$AC_{306} = 17$  atoms;  $AC_{305} = 24$  atoms;  $AC_{304} = 17$  atoms;...  $AC_1 = 10$  atoms;

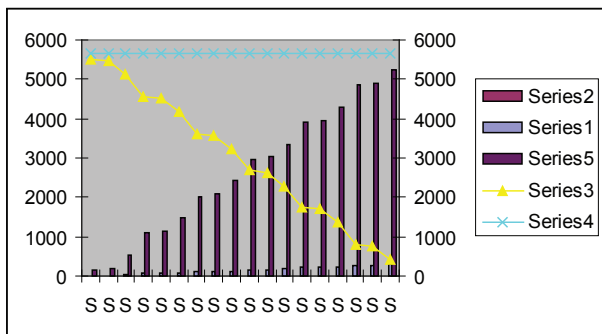
$$[AC_{306} + (AC_{306} + AC_{305}) + (AC_{306} + AC_{305} + AC_{304})... + (AC_{306} + AC_{305} + AC_{304}... + AC_1)] = S_2;$$

$$AC_{306} = APb_{306} = 17;$$

$$(AC_{306} + AC_{305}) = (17+24) = APb_{306} = 41;$$

$$(AC_{306} + AC_{305} + AC_{304}) = (17+24+17) = APb_{304} = 58;$$

$$(AC_{306} + AC_{305} + AC_{304}... + AC_1) = APb_1 = 5640 \text{ atoms};$$



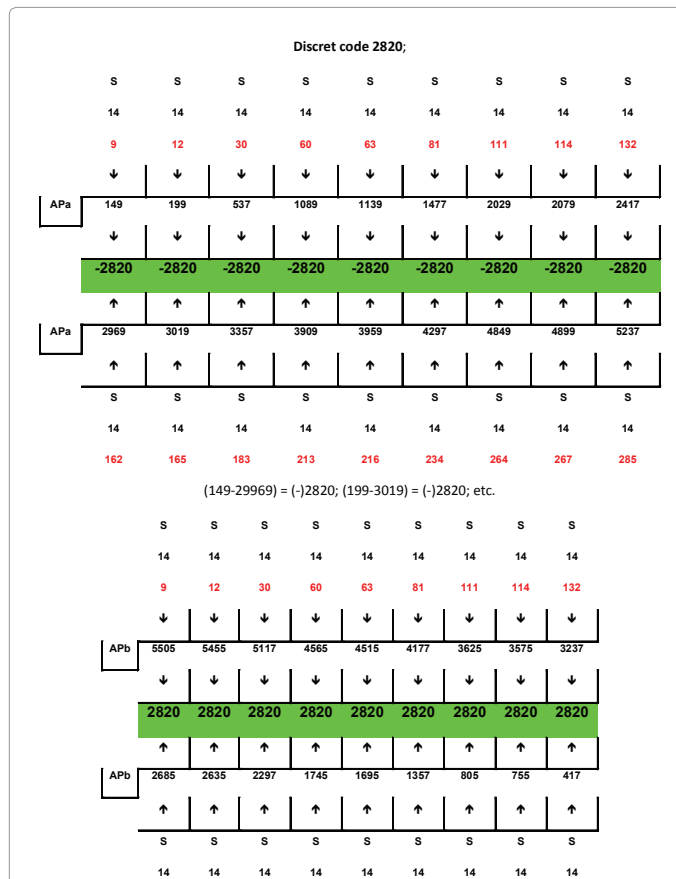
**Figure 4.5:** A schematic diagram to show of the atomic progression APa and APb (Amino acid Ser – position from 1 to 306 AA).

$APb_{306,305,304,...,1}$  = Atomic progression of amino acids 306,305,304,...,1;

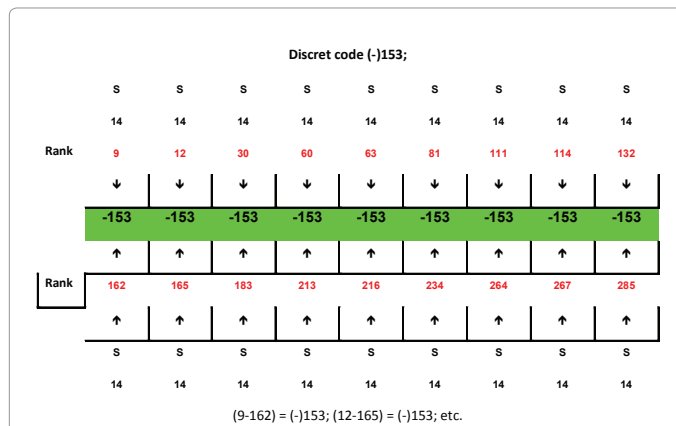
$$[APb_{306} + APb_{305} + APb_{1304})... + APb_1] = (17+41+58... + 5640) = 868\ 272;$$

$$S_2 = 868\ 272;$$

Within the digital pictures in biochemistry, the physical and chemical parameters are in a strict compliance with programmatic, cybernetic and information principles. Each bar in the protein chain



**Figure 6:** Schematic representation of the discret code 2820. (Amino acid Ser – position from 1 to 306 AA).



**Figure 7:** Schematic representation of the discret code (-)153.



**Square 1-4**

149	199	537
1089	1139	1477
2029	2079	2417
2969	3019	3357
3909	3959	4297
4849	4899	5237
5505	5455	5117
4565	4515	4177
3625	3575	3237
2685	2635	2297
1745	1695	1357
805	755	417

**33924 33924 33924**

**Square 1-4**

149	199	537	2969	3019	3357
1089	1139	1477	3909	3959	4297
2029	2079	2417	4849	4899	5237
5505	5455	5117	2685	2635	2297
4565	4515	4177	1745	1695	1357
3625	3575	3237	805	755	417

**16962 16962 16962 16962 16962 16962**

Figure 11: Schematic representation of the atomic progression in squares 1,2,3 and 4.

also by cybernetic and information principles.

Atomic progression model of insulin should, in fact, be “remodelled” into a periodic system. Schematic representation of the amino acid Ser and atomic progression we will shown in the (Figure 6-9).

In (Figure 6-9), the amino acids Ser atomic progression APa and APb as a result was given the discret codes 2820, 153, 268 and 556<sup>th</sup>.

As we see, the insulin code is itself a unique structure of program, cybernetic and informational system and law. The research we carried

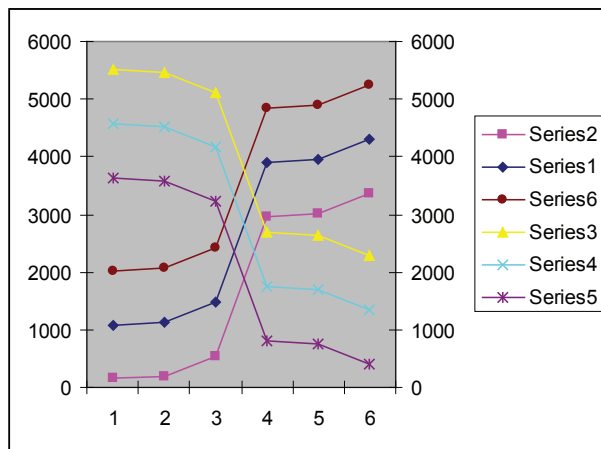


Figure 12: A schematic diagram to show of the atomic progression in square 1,2,3 and 4.

**Square 1-4 (Inner progressions)**

1139	1477	3909	3959
2079	2417	4849	4899
5455	5117	2685	2635
4515	4177	1745	1695

**13188 13188 13188 13188**

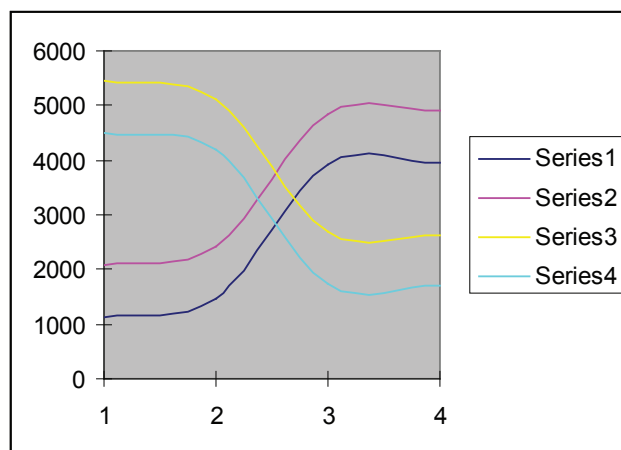
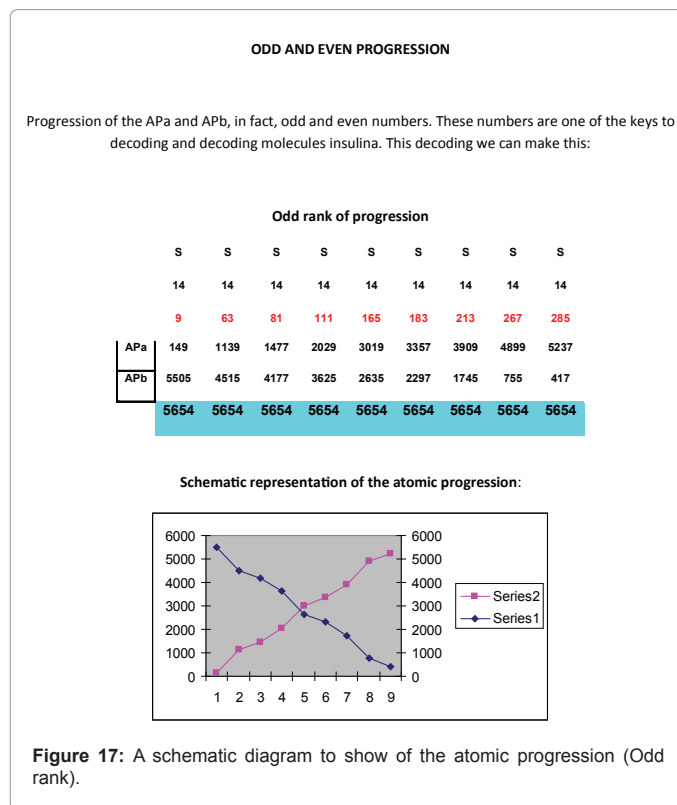
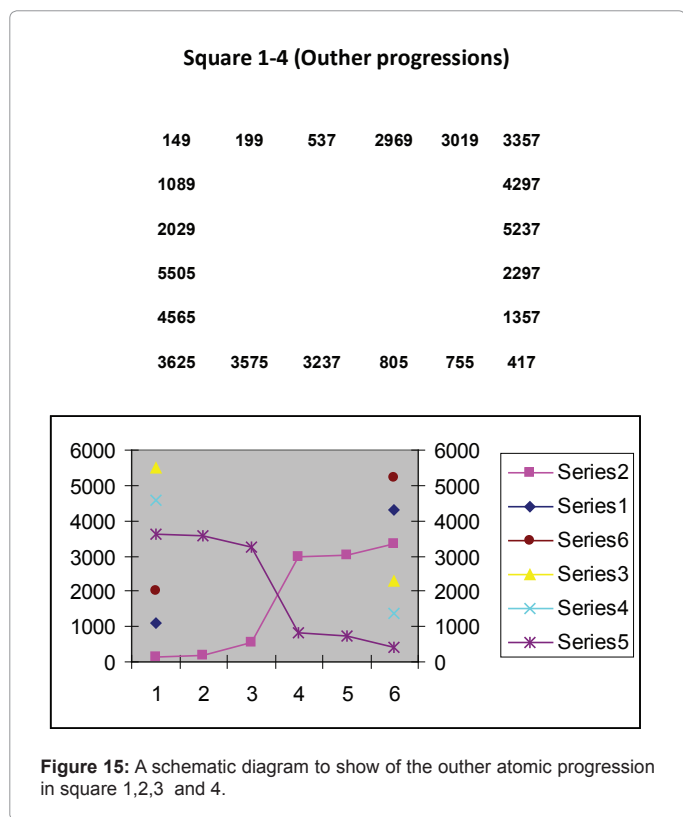
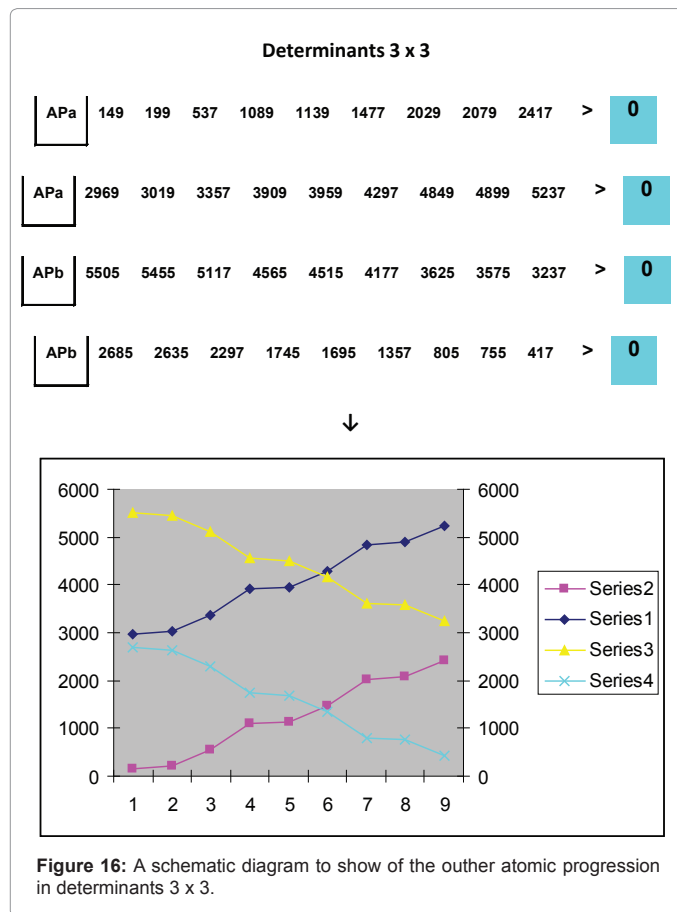
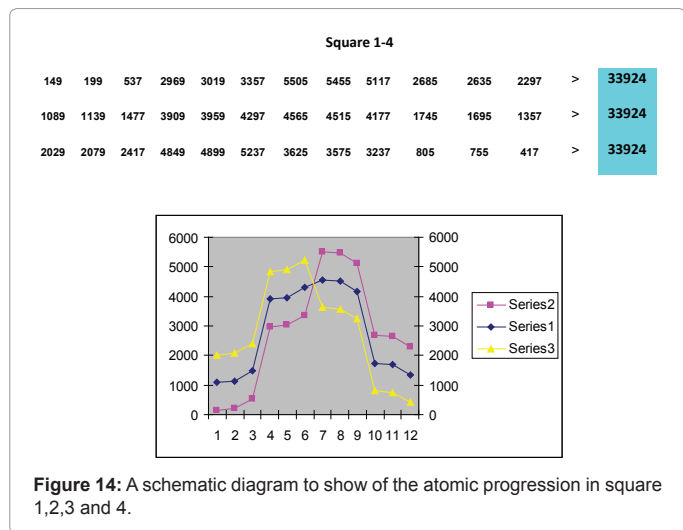


Figure 13: A schematic diagram to show of the inner atomic progression in square 1,2,3 and 4.

out have shown that atomic progression are one of quantitative characteristics in biochemistry. Atomic progression is, actually, a discrete code that protects and guards genetic information coded in bio-chemical processes. This a recently discovered code, and more detailed knowledge on it is yet to be discovered. In a similar way we shall calculate bio codes of other unions of amino acids. Once we do



this, we will find out that all these unions of amino acids are connected by various bio codes, analogue codes as well as other quantitative features.

In (Figure 10,11) there is also a mathematical balance between atomic progression APa and APb (Amino acid Ser – position from 1 to 306 AA) square 1,2,3,4.

Atomic progression presented in previous (Figure 12-15) are calculated using the relationship between corresponding groups of those progressions. These are groups with different progression. There are different ways and methods of selecting these groups of progressions, which method is most efficient some We hope that

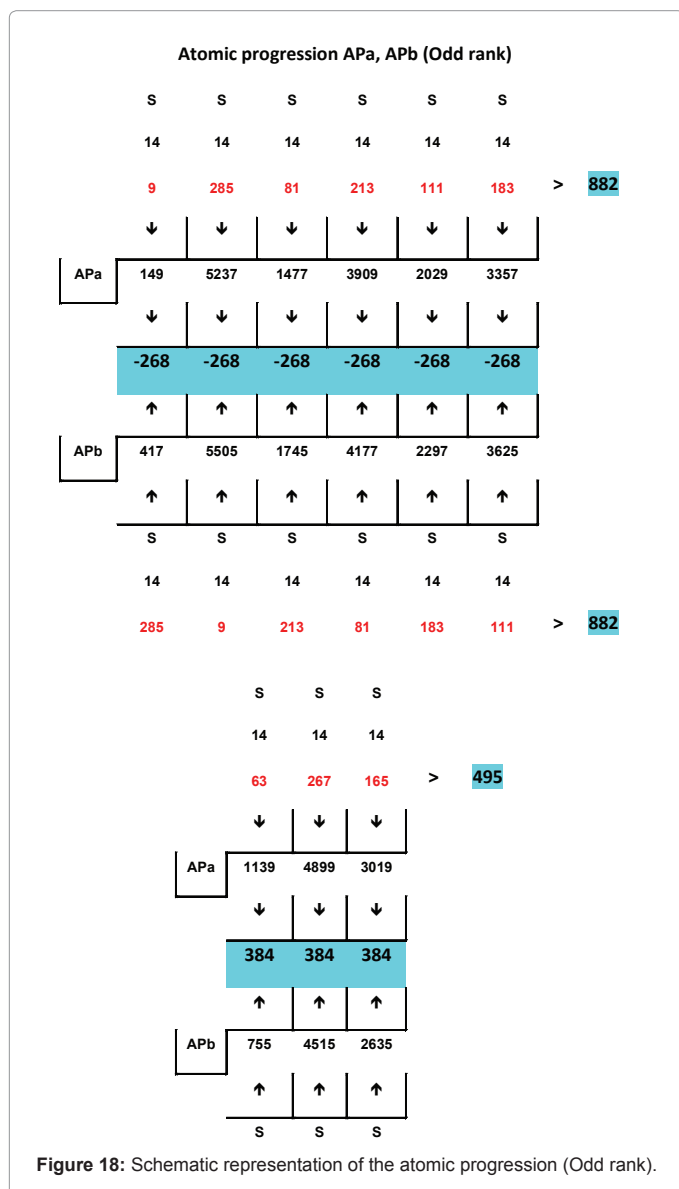


Figure 18: Schematic representation of the atomic progression (Odd rank).

science will determine which method (Figure 19) is most efficient for this selection.

Biological specificum of proteins depends on the order of amino acids in their molecules. If this order is changed, their biological specificum will change as well. The basic parameters that determine the change in the status of system of bio-synthesis matrix of macro molecules are the system entropy, volume of transferred information and degree of probability that the genetic information will be transferred in full. System means any group of objects (elements), their inter-relations (number of atoms, atomic number, atomic weight, covalent radius, molar rotation, electrical negativity, etc...), as well as the relations among its attributes. In this example we have analyzed only the number of atoms of the said amino acids. This number of atoms has been analyzed from one perspective only. A similar analyses can be done using some other parameters and observing these parameters from other perspectives. This creates the need for research of other cyber, information and system characteristics of biochemical basis of genetic processes of insulin. The results of that research will hopefully

be published in the time to come.

The result of the research that we have carried out clearly shows that there is a matrix code in insulin. It also shows that the coding system within the amino acidic language gives a full information, not only for the amino acid "record", but also for its structure, configuration and its various shapes. In the following text we shall discuss the issue of the existence of the insulin code, and also the issue of coding of individual structural levels in this protein.

In the previous (Figure 17-20) we translated the physical and chemical parameters from the language of biochemistry into the digital language of programmatic, cybernetic and information principles. This we did by using the adequate mathematical algorithms. By using chemical-information procedures, we calculated the numerical value for the information content of molecules. What we got this way is the digital picture of the phenomenon of biochemistry. These digital pictures reveal to us a whole new dimension of this science. They reveal to us that the biochemical process is strictly conditioned and determined by programmatic, cybernetic and information principles.

From the previous (Figure 17-20) we can see that this protein

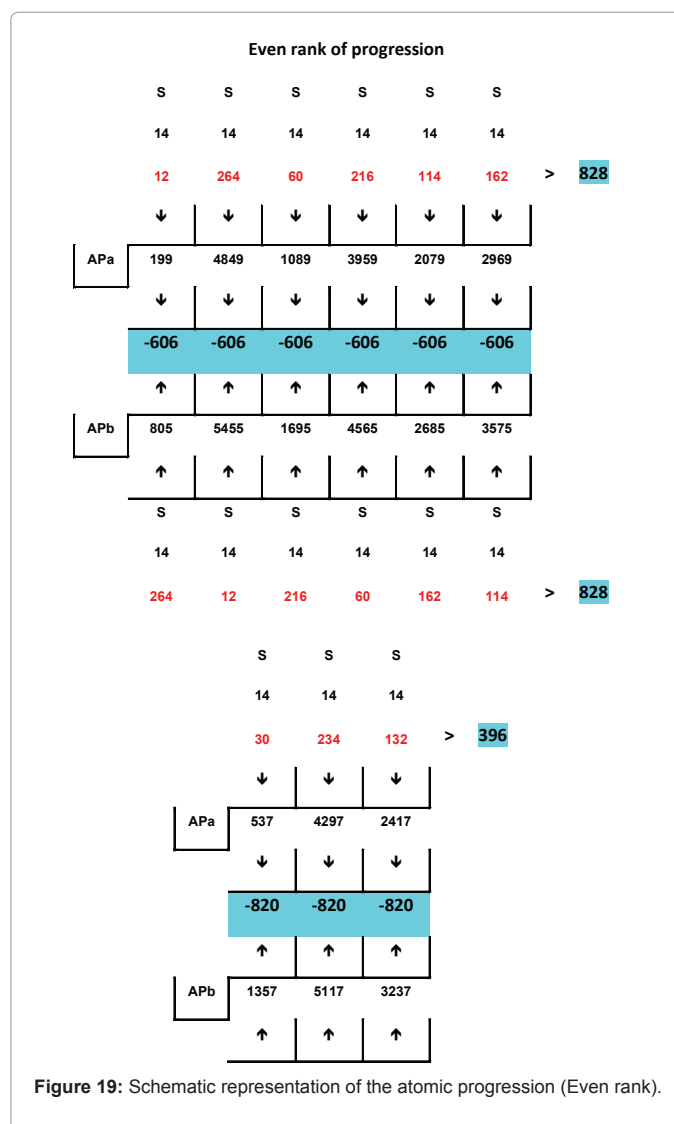
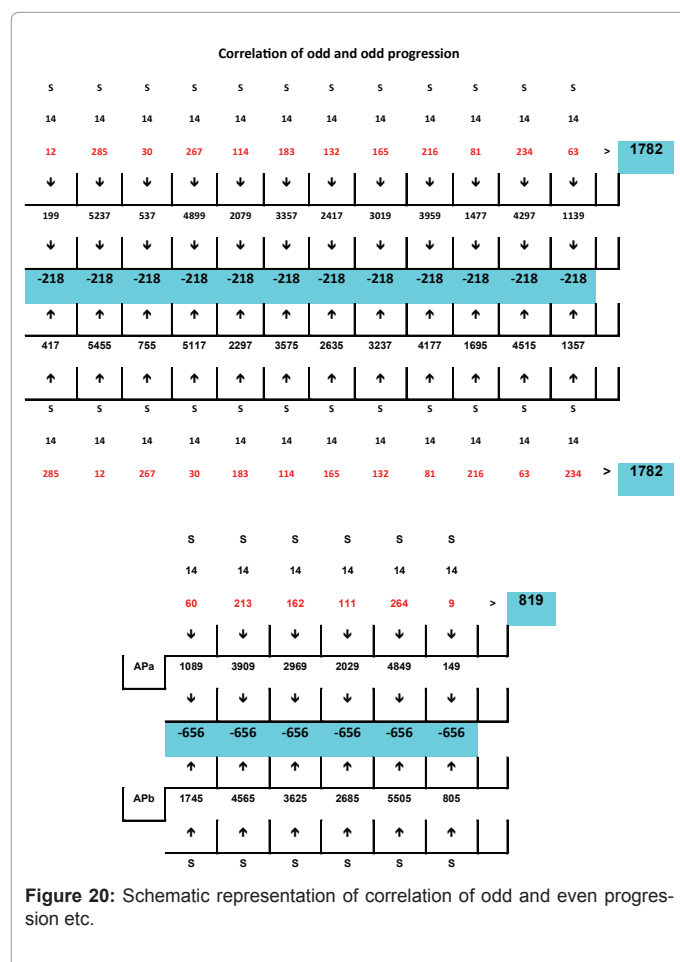


Figure 19: Schematic representation of the atomic progression (Even rank).



really has its quantitative characteristics. It can be concluded that there is a connection between quantitative characteristics in the process of transfer of genetic information and the qualitative appearance of given genetic processes.

## Discussion

The results of our research show that the processes of sequencing the molecules are conditioned and arranged not only with chemical and biochemical lawfulness, but also with program, cybernetic and informational lawfulness too. At the first stage of our research we replaced nucleotides from the Amino Acid Code Matrix with numbers of the atoms and atomic numbers in those nucleotides. Translation of the biochemical language of these amino acids into a digital language may be very useful for developing new methods of predicting protein sub-cellular localization, membrane protein type, protein structure secondary prediction or any other protein attributes.

The success of human genome project has generated deluge of sequence information. The explosion of biological data has challenged scientists to accelerate the speed for their analysis. Nowadays, protein sequences are generally stored in the computer database system in the form of long character strings. It would act like a snail's pace for human beings to read these sequences with the naked eyes [22]. Also, it is very hard to extract any key features by directly reading these long character strings. However, if they can be converted to some signal process, many important features can be automatically manifested and easily studied by means of the existing tools of information theory [22]. The novel

approach as presented here may help improve this kind of situation.

## Conclusions and Perspectives

The process of sequencing in bio-macromolecules is conditioned and determined not only through biochemical, but also through cybernetic and information principles. The digital pictures of biochemistry provide us with cybernetic and information interpretation of the scientific facts. Now we have the exact scientific proofs that there is a genetic language that can be described by the theory of systems and cybernetics, and which functions in accordance with certain principles.

## Disclosure

The author reports no conflict of interest in this research.

## Acknowledgement

I am indebted to Kuo-Chen Chou and Peer Fogarty for their help in data analysis and made valuable comments.

## References

- Cai YD, Chou K.C (2006) Predicting membrane protein type by functional Domain composition and pseudo-amino acid composition. *J Theor Biol* 238: 395-400.
- Cai YD, Zhou GP, Chou KC (2003) Support vector machines for predicting membrane protein types by using functional domain composition. *Biophys J* 84: 3257- 3263.
- Cai YD, Zhou GP, Chou KC (2005) Predicting enzyme family classes by hybridizing gene product composition and pseudo-amino acid composition. *J Theor Biol* 234: 145-149.
- Cai YD, Feng KY, Lu WC, Chou KC (2006) Using LogitBoost classifier to predict protein structural classes. *J Theor Biol* 238: 172-176.
- Cai YD, Ricardo PW, Jen CH, Chou KC (2004) Application of SVM to predict membrane protein types. *J Theory Biol* 226: 373-376.
- Chen C, Chen L, Zou X, Cai P (2009) Prediction of protein secondary structure content by using the concept of Chou's pseudo amino acid composition and support vector machine. *Protein & Peptide Letters* 16: 27-31.
- Chen L, Feng KY, Cai YD, Chou KC, Li HP (2010) Predicting the network of substrate-enzyme-product triads by combining compound similarity and functional domain composition. *BMC Bioinformatics* 11: 293.
- Chou KC (2011) Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review). *J Theor Biol* 273: 236-247.
- Chou KC, Shen HB (2010) Cell-PLoc 2.0: An improved package of web-servers for predicting subcellular localization of proteins in various organisms. *Natural Science* 2: 1090-1103.
- Chou KC, Shen HB (2010) Plant-mPLoc: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. *PLoS ONE* 5: e11335.
- Ding H, Luo L, Lin H (2009) Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. *Protein & Peptide Letters* 16: 351-355.
- Ding YS, Zhang, TL, Chou C (2007) Prediction of protein structure classes with pseudo amino acid composition and fuzzy support vector machine network. *Protein & Peptide Letters* 14: 811-815.
- Ding YS, Zhang TL, Gu Q, Zhao PY, Chou KC (2009) Using maximum entropy model to predict protein secondary structure with single sequence. *Protein & Peptide Letters* 16: 552-560.
- He Z.S, Zhang J, Shi XH, Hu LL, Kong X.G. et al. (2010) Predicting drug-target interaction networks based on functional groups and biological features. *PLoS ONE* 5: e9603.
- Hu L, Huang T, Shi X., Lu WC, Cai YD, et al. (2011) Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties *PLoS ONE* 6: e14556.



16. Huang T, Shi XH, Wang P, He Z, Feng KY, et al. (2010) Analysis and prediction of the metabolic stability of proteins based on their sequential features, subcellular locations and interaction networks PLoS ONE 5: e10972.
17. Kandaswamy KK, Pugalenth G, Moller S, Hartmann E, Kalies KU, et al. (2010) Prediction of Apoptosis Protein Locations with Genetic Algorithms and Support Vector Machines through a New Mode of Pseudo Amino Acid Composition. Protein and Peptide Letters 17: 1473-1479.
18. Kannan S, Hauth AM, Burger G (2008) Function prediction of hypothetical proteins without sequence similarity to proteins of known function. Protein & Peptide Letters 15: 1107-1116.
19. Liu T, Zheng X, Wang C, Wang J (2010) Prediction of Subcellular Location of Apoptosis Proteins using Pseudo Amino Acid Composition: An Approach from Auto Covariance Transformation. Protein & Peptide Letters 17: 1263-1269.
20. Mohabatkar H (2010) Prediction of cyclin proteins using Chou's pseudo amino acid composition. Protein & Peptide Letters 17: 1207-1214.
21. Wang YC, Wang XB, Yang ZX, Deng NY (2010) Prediction of enzyme subfamily class via pseudo amino acid composition by incorporating the conjoint triad feature. Protein & Peptide Letters 17:1441-1449.
22. Xiao X, Chou KC (2007) Digital coding of amino acids based on hydrophobic index. Protein & Peptide Letters 14: 871-875.