# Molecular Biocoding of Insulin – Amino Acid Gly

Lutvo Kurić

**Abstract** – The modern science mainly treats the biochemical basis of sequencing in bio-macromolecules and processes in medicine and biochemistry. One can ask weather 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 (He et al., 2010), predicting functions of proteins (Hu et al., 2011; Kannan et al., 2008), analysis and prediction of the metabolic stability of proteins (Huang et al., 2010), predicting the network of substrate-enzyme-product triads (Chen et al., 2010), membrane protein type prediction (Cai and Chou, 2006; Cai et al., 2003; Cai et al., 2004), protein structural class prediction (Cai et al., 2006; Ding et al., 2007), protein secondary structure prediction (Chen et al., 2009; Ding et al., 2009b), enzyme family class prediction (Cai et al., 2005; Ding et al., 2009a; Wang et al., 2010), identifying cyclin proteins (Mohabatkar, 2010), protein subcellular location prediction (Chou and Shen, 2010a; Chou and Shen, 2010b; Kandaswamy et al., 2010; Liu et al., 2010), among many others as summarized in a recent review (Chou, 2011), 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.

**Index Terms**–Amino Acid Gly, Human Insulin, Insulin Model, Insulin Code.

## 1 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 distin biochemical features, but also by cybernetic and information principles. For this reason, research in this field deals more with the quantitative rather than qualitative characteristcs 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 them assigned to these factors, enabling them to be measured. In this way it is possible to determine oif 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 frenquencies.There is a arge numbers of these parameters, and each of their 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.

## 2 METHODS

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 represent by a series of amino acids according to the order of their position in the chains 1AI0.

Therefore, the sequential form can naturally reflect all the information about the sequence order and lenght of an insulin molecule. The key issue is whether we can develop a different discrete method of representing an insulin molecule that will allow accomodation 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?

## 3 Expression of Insulin Code Matrix- 1AI0

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 is the exact specification of its atomic composition and the chemical bonds connecting those atoms.

### 3.1 Model

The structure **1AI0** has in total **12** chains: A,B,C,D,E,F,G,H,I,J,K,L.

1AI0:A

| G | I | V | E | Q | C | C | T | S | I | C | S | L | Y | Q | L | E | N | Y | C | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 22 | 19 | 19 | 20 | 14 | 14 | 17 | 14 | 22 | 14 | 14 | 22 | 24 | 20 | 22 | 19 | 17 | 24 | 14 | 17 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |

1AI0:B

| F | V | N | Q | H | I | C | G | S | H | L | V | E | A | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 19 | 17 | 20 | 20 | 22 | 14 | 10 | 14 | 20 | 22 | 19 | 19 | 13 | 22 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |

| Y | L | V | C | G | E | R | G | F | I | Y | T | P | K | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 22 | 19 | 14 | 10 | 19 | 26 | 10 | 23 | 22 | 24 | 17 | 17 | 24 | 17 |
| 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 |

etc.

Fig. 1. Group of chains A,B,C,D,E,F,G,H,I,J,K,L.

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 this 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 voded features. In fact, there is a cybernetic algorithm which it is „recorded“ that the firs 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 appearance, ie, the characteristcs of the organism.

### 3.2 Algorithm

We shall now give some mathematical evidences that will prove that in the biochemistry of hemoglob 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.

*3.2.1 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;

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

$AC_1 = APa1 = 10$;
$(AC_1 + AC_2) = (10+22) = APa2 = 32$;

$(AC_1 + AC_2 + AC_3) = (10+22+19) = APa3 = 51$;
$(AC_1 + AC_2 + AC_3..., + AC_{306}) = APa_{306} = 5640$ atoms;
$APa_{1,2,3,n}$ = Atomic progression of amino acids 1,2,3,n

$[APa_1+APa_2+APa_3)..., + APa_{306})] = (10+32+51\ldots, + 5640) = \mathbf{S_1}$;

$S_1$ = 863 208;

Example :

Atomic progression 1 (APa)

| G | I | V | E | Q | C | . | . | P | K | T | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 22 | 19 | 19 | 20 | 14 | . | . | 17 | 24 | 17 | **5640** |
| **1** | **2** | **3** | **4** | **5** | **6** | **.** | **.** | **304** | **305** | **306** | **46971** |

↓

| G | I | V | E | Q | C | . | . | P | K | T | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 32 | 51 | 70 | 90 | 104 | . | . | 5599 | 5623 | 5640 | **863 208** |
| **1** | **2** | **3** | **4** | **5** | **6** | **.** | **.** | **304** | **305** | **306** | **46971** |

(0+10) = 10; (10+22)=32; (10+11+19) = 51; etc.

Fig. 2. Atomic progression 1 (APa) of amino acids from 1 to 306.

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;

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

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

$(AC_{306}+ AC_{305}+ AC_{304}) = (17+24+17) = APb_{304} = 58;$
$(AC_{306}+ AC_{305}+ AC_{304}..., + AC_1) = APb_1 = 5640$ atoms;
$APb_{306,305,304, \ldots,1}$ = Atomic progression of amino acids 306,305,304,…1;

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

$S_2 = 868\ 272;$

Example:

Atomic progression 2 (APb)

| G | I | V | · | · | I | Y | T | P | K | T | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 22 | 19 | . | . | 22 | 24 | 17 | 17 | 24 | 17 | 5640 |
| 1 | 2 | 3 | . | . | 301 | 302 | 303 | 304 | 305 | 306 | 46971 |

↓

| Sum | G | I | V | · | · | I | Y | T | P | K | T |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 868 272 | 5640 | 5630 | 5608 | . | . | 121 | 99 | 75 | 58 | 41 | 17 |
| 46971 | 1 | 2 | 3 | . | . | 301 | 302 | 303 | 304 | 305 | 306 |

(0+17) = 17; (17+24)=41; (17+24+17)=58; etc.

Fig. 3. Schematic representation of the atomic progression 2 from 306 to 1.

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 attracts only the corresponding aminoacid, and only the relevant aminoacid can be positioned at certain place in the chain. Each peptide chain can have the exact number of aminoacids necessary to meet the strictly determined mathematical conditioning. It can have as many atoms as necessary to meet the mathematical balance of the biochemical phenomenon at certain mathematical level, etc. The digital language of biochemistry has a countless number of codes and analogue codes, as well as other information content. These pictures enable us to realize the very essence of functioning of biochemical processes. There are some examples:

Table 1. Atomic progression APa and APb (Amino acid Gly – position from 1 to 306 AA)

The structure 1AI0 – Amino acid Gly

| | G | G | G | G | G | G | G | G |
|---|---|---|---|---|---|---|---|---|
| Number of atoms | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Rank | 1 | 29 | 41 | 44 | 52 | 80 | 92 | 95 |
| APa | 10 | 523 | 741 | 796 | 950 | 1463 | 1681 | 1736 |
| APb | 5640 | 5127 | 4909 | 4854 | 4700 | 4187 | 3969 | 3914 |
| AP(a,b) | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 |

| | G | G | G | G | G | G | G | G |
|---|---|---|---|---|---|---|---|---|
| Number of atoms | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Rank | 103 | 131 | 143 | 146 | 154 | 182 | 194 | 197 |
| APa | 1890 | 2403 | 2621 | 2676 | 2830 | 3343 | 3561 | 3616 |
| APb | 3760 | 3247 | 3029 | 2974 | 2820 | 2307 | 2089 | 2034 |
| AP(a,b) | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G |
| Number of atoms | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Rank | 205 | 233 | 245 | 248 | 256 | 284 | 296 | 299 |
| APa | 3770 | 4283 | 4501 | 4556 | 4710 | 5223 | 5441 | 5496 |
| APb | 1880 | 1367 | 1149 | 1094 | 940 | 427 | 209 | 154 |
| AP(a,b) | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 |

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

Notes: Namely, having mathematically analyzed the atomic preogression model of Insulin Model (Table 1) we have found out that the protein code is based on a periodic law. This being the only to „read“ the picture, the solution of the main problem (concering an arrangement where each amino acid takes only one, precisely determined position in the code), is quite manifest:

Atomic progression model of insulin should, in fact, be „remodelled“ into a periodic system. Examples:

Atomic progression APa and APb

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G | |
| Rank | 299 | 1 | 29 | 296 | 41 | 284 | 44 | 256 | > 1250 |
| APa | 5496 | 10 | 523 | 5441 | 741 | 5223 | 796 | 4710 | > 22940 |
| R | -144 | -144 | 314 | 314 | 314 | 314 | -144 | -144 | > 680 |
| APb | 5640 | 154 | 209 | 5127 | 427 | 4909 | 940 | 4854 | > 22260 |
| | G | G | G | G | G | G | G | G | |
| Rank | 1 | 299 | 296 | 29 | 284 | 41 | 256 | 44 | > 1250 |

R > (5496-5640) = (-)144; (10-154) = (-)144; (523-209) = 314; etc.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G | |
| Rank | 52 | 248 | 80 | 245 | 92 | 233 | 95 | 205 | > 1250 |
| APa | 950 | 4556 | 1463 | 4501 | 1681 | 4283 | 1736 | 3770 | > 22940 |
| R | -144 | -144 | 314 | 314 | 314 | 314 | -144 | -144 | > 680 |
| APb | 1094 | 4700 | 1149 | 4187 | 1367 | 3969 | 1880 | 3914 | > 22260 |
| | G | G | G | G | G | G | G | G | |
| Rank | 248 | 52 | 245 | 80 | 233 | 92 | 205 | 95 | > 1250 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G | |
| Rank | 103 | 197 | 131 | 194 | 143 | 182 | 146 | 154 | > 1250 |

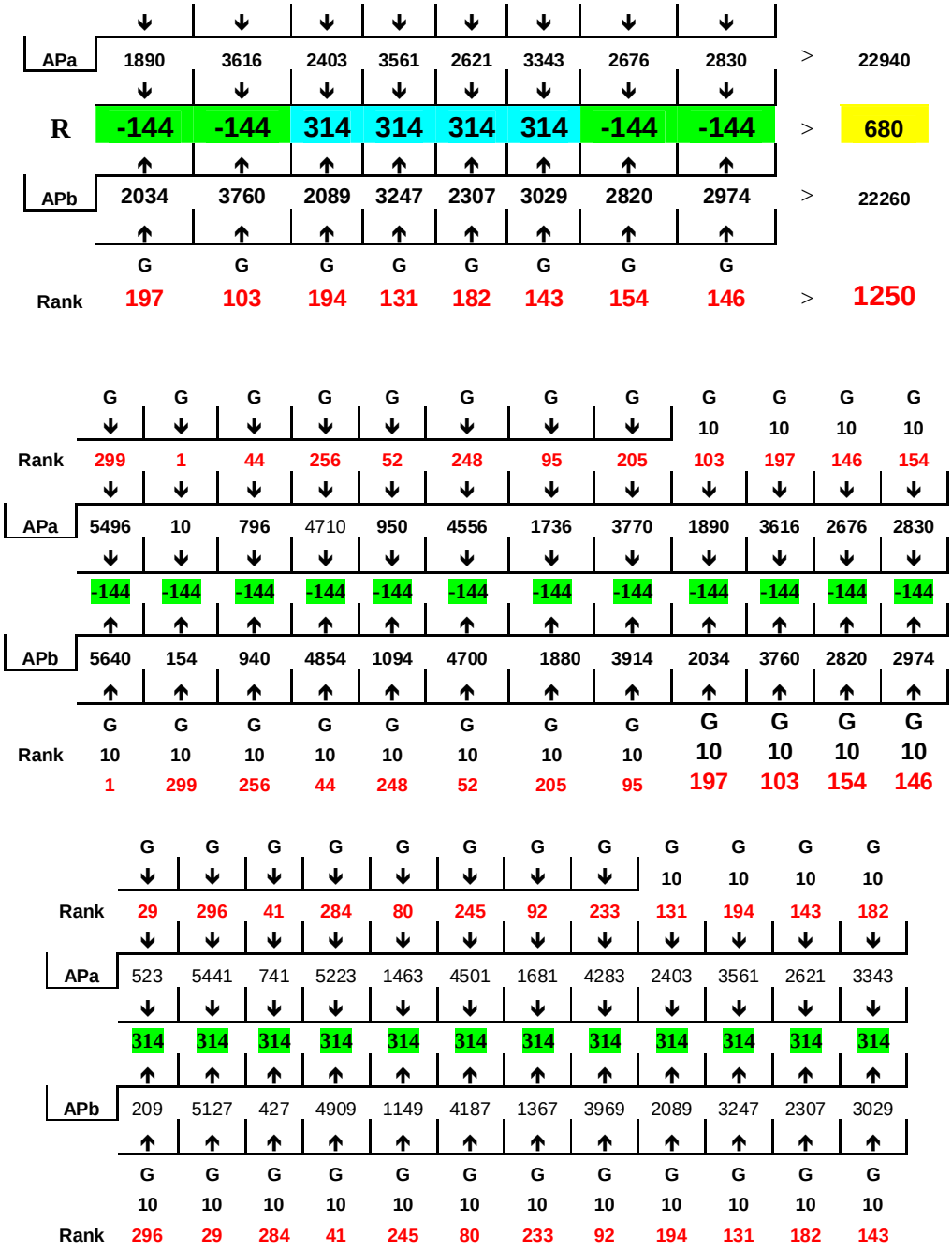


Fig. 4. Atomic progression APa and APb (Amino acid Gly – position from 1 to 306 AA).

In this example, the amino acids Gly atomic progression APa and APb as a result was given the codes 144 and 314th.

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

The research we carried 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. Examples:

Atomic progression APa and APb

| | G | G | G | G | G | G | G | G |
|---|---|---|---|---|---|---|---|---|
| APa | 5496 | 10 | 523 | 5441 | 741 | 5223 | 796 | 4710 |
| APa | 950 | 4556 | 1463 | 4501 | 1681 | 4283 | 1736 | 3770 |
| APa | 1890 | 3616 | 2403 | 3561 | 2621 | 3343 | 2676 | 2830 |
| APb | 5640 | 154 | 209 | 5127 | 427 | 4909 | 940 | 4854 |
| APb | 1094 | 4700 | 1149 | 4187 | 1367 | 3969 | 1880 | 3914 |
| APb | 2034 | 3760 | 2089 | 3247 | 2307 | 3029 | 2820 | 2974 |
| | | 67800 | | | | | 67800 | |

| | G | G | G | G | | Sum |
|---|---|---|---|---|---|---|
| APa | 5496 | 10 | 523 | 5441 | > | 11470 |
| APa | 950 | 4556 | 1463 | 4501 | > | 11470 |
| APa | 1890 | 3616 | 2403 | 3561 | > | 11470 |

| | G | G | G | G | | Sum |
|---|---|---|---|---|---|---|
| APa | 741 | 5223 | 796 | 4710 | > | 11470 |
| APa | 1681 | 4283 | 1736 | 3770 | > | 11470 |
| APa | 2621 | 3343 | 2676 | 2830 | > | 11470 |

| | G | G | G | G | | Sum |
|---|---|---|---|---|---|---|
| APb | 5640 | 154 | 209 | 5127 | > | 11130 |
| APb | 1094 | 4700 | 1149 | 4187 | > | 11130 |
| APb | 2034 | 3760 | 2089 | 3247 | > | 11130 |

| | G | G | G | G | | Sum |
|---|---|---|---|---|---|---|
| APb | 427 | 4909 | 940 | 4854 | > | 11130 |
| APb | 1367 | 3969 | 1880 | 3914 | > | 11130 |
| APb | 2307 | 3029 | 2820 | 2974 | > | 11130 |

Fig. 5. Atomic progression APa and APb (Amino acid Gly – position from 1 to 306 AA).

Atomic progression presented in figure 2 are calculated using the relationship between corresponding groups of those rogressions. 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 science will determine which method is most efficient for this selection.

3.2.2 Rank of atomic progression

| | G ◎ | G ◎ | G ◎ | G ◎ | G ◎ | G ◎ | G ◎ | G ◎ | Sum |
|---|---|---|---|---|---|---|---|---|---|
| Rank | 299 | 1 | 29 | 296 | 41 | 284 | 44 | 256 | 1250 |
| Rank | 1 | 299 | 296 | 29 | 284 | 41 | 256 | 44 | 1250 |
| Rank | 52 | 248 | 80 | 245 | 92 | 233 | 95 | 205 | 1250 |
| Rank | 248 | 52 | 245 | 80 | 233 | 92 | 205 | 95 | 1250 |
| Rank | 103 | 197 | 131 | 194 | 143 | 182 | 146 | 154 | 1250 |
| Rank | 197 | 103 | 194 | 131 | 182 | 143 | 154 | 146 | 1250 |
| Sum | 900 | 900 | 975 | 975 | 975 | 975 | 900 | 900 | |

| | G ◎ | G ◎ | G ◎ | G ◎ | | |
|---|---|---|---|---|---|---|
| Rank | 299 | 1 | 29 | 296 | > | 625 |
| Rank | 1 | 299 | 296 | 29 | > | 625 |
| Rank | 52 | 248 | 80 | 245 | > | 625 |
| Rank | 248 | 52 | 245 | 80 | > | 625 |
| Rank | 103 | 197 | 131 | 194 | > | 625 |
| Rank | 197 | 103 | 194 | 131 | > | 625 |

| | G ◎ | G ◎ | G ◎ | G ◎ | | Sum |
|---|---|---|---|---|---|---|
| Rank | 41 | 284 | 44 | 256 | > | 625 |
| Rank | 284 | 41 | 256 | 44 | > | 625 |
| Rank | 92 | 233 | 95 | 205 | > | 625 |
| Rank | 233 | 92 | 205 | 95 | > | 625 |
| Rank | 143 | 182 | 146 | 154 | > | 625 |
| Rank | 182 | 143 | 154 | 146 | > | 625 |

| | G ◎ | G ◎ | G ◎ | G ◎ | | Sum |
|---|---|---|---|---|---|---|
| Rank | 299 | 1 | 44 | 256 | > | 600 |
| Rank | 1 | 299 | 256 | 44 | > | 600 |
| Rank | 52 | 248 | 95 | 205 | > | 600 |
| Rank | 248 | 52 | 205 | 95 | > | 600 |
| Rank | 103 | 197 | 146 | 154 | > | 600 |
| Rank | 197 | 103 | 154 | 146 | > | 600 |
| Sum | 900 | 900 | 900 | 900 | | |

| | G ◎ | G ◎ | G ◎ | G ◎ | | Sum |
|---|---|---|---|---|---|---|
| Rank | 29 | 296 | 41 | 284 | > | 650 |
| Rank | 296 | 29 | 284 | 41 | > | 650 |
| Rank | 80 | 245 | 92 | 233 | > | 650 |
| Rank | 245 | 80 | 233 | 92 | > | 650 |
| Rank | 131 | 194 | 143 | 182 | > | 650 |
| Rank | 194 | 131 | 182 | 143 | > | 650 |
| Sum | 975 | 975 | 975 | 975 | | |

Fig. 6. Rank atomic progression APa and APb (Amino acid Gly – position from 1 to 306 AA).

In those examples, we have the mathematical balance in rows and columns in this figure.

In fact, we have discovered the mathematical balance in distribution of sequences in figure 6 is achieved.

3.2.3 Correlation of atomic progression

Atomic progression of this amino acid to us as a result of its correlation give a variety of codes. Here are some examples:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | G | | | | G | G | | | | G | G | | |
| | 10 | 10 | | | | 10 | 10 | | | | 10 | 10 | | |
| | 41 | 284 | > | 325 | | 80 | 245 | > | 325 | | 143 | 182 | > | 325 |
| APa | 741 | 5223 | > | 5964 | | 1463 | 4501 | > | 5964 | | 2621 | 3343 | > | 5964 |
| APb | 4909 | 427 | > | 5336 | | 4187 | 1149 | > | 5336 | | 3029 | 2307 | > | 5336 |
| | ◎ | ◎ | | ◎ | | ◎ | ◎ | | ◎ | | ◎ | ◎ | | ◎ |
| R | 314 | 314 | | 11300 | | 314 | 314 | | 11300 | | 314 | 314 | | 11300 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | G | | | | G | G | | | | G | G | | |
| | 10 | 10 | | | | 10 | 10 | | | | 10 | 10 | | |
| | 44 | 256 | > | 300 | | 52 | 248 | > | 300 | | 146 | 154 | > | 300 |
| APa | 796 | 4710 | > | 5506 | | 950 | 4556 | > | 5506 | | 2676 | 2830 | > | 5506 |
| APb | 4854 | 940 | > | 5794 | | 4700 | 1094 | > | 5794 | | 2974 | 2820 | > | 5794 |
| | ◎ | ◎ | | ◎ | | ◎ | ◎ | | ◎ | | ◎ | ◎ | | ◎ |
| R | 314 | 314 | | 11300 | | 314 | 314 | | 11300 | | 314 | 314 | | 11300 |

Fig. 7. Correlation of atomic progression and rank of APa and APb (Amino acid Gly – position from 1 to 306 AA).

In those examples we have the correlation of atomic progression and rank of APa and APb.

3.2.4 Odd and even progression

Progression of the APa and APb, in fact, odd and even numbers. These numbers are one of the keys to decoding and decoding molecules insulina. This decoding we can make this:

Steam progression

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G | G | G | G | G |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| | 1 | 44 | 52 | 95 | 103 | 146 | 154 | 197 | 205 | 248 | 256 | 299 |
| APa | 10 | 796 | 950 | 1736 | 1890 | 2676 | 2830 | 3616 | 3770 | 4556 | 4710 | 5496 |
| APb | 5640 | 4854 | 4700 | 3914 | 3760 | 2974 | 2820 | 2034 | 1880 | 1094 | 940 | 154 |
| | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 |

etc.

Steam progression are:

APa = (10, 5640); (1736, 3914); (1890, 3760); etc.

Progression are the odd and even rank.

Odd rank

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | | |
| | 10 | 10 | 10 | 10 | 10 | 10 | | |
| Rank | 1 | 95 | 103 | 197 | 205 | 299 | > | 900 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| APa | 10 | 1736 | 1890 | 3616 | 3770 | 5496 | | 16518 |
| APb | 5640 | 3914 | 3760 | 2034 | 1880 | 154 | > | 17382 |
| | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | | |

Even rank

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | | |
| | 10 | 10 | 10 | 10 | 10 | 10 | | |
| Rank | 44 | 52 | 146 | 154 | 248 | 256 | > | 900 |
| APa | 796 | 950 | 2676 | 2830 | 4556 | 4710 | | 16518 |
| APb | 4854 | 4700 | 2974 | 2820 | 1094 | 940 | > | 17382 |
| | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | | |

Odd rank = Even rank = 900;
Odd APa = Even APa = 16518;
Odd APb = Even APb = 17382;

Fig. 8. Odd and even progression of aminoacid Gly.

Notes: 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 attracts only the corresponding aminoacid, and only the relevant aminoacid can be positioned at certain place in the chain. Each peptide chain can have the exact number of aminoacids necessary to meet the strictly determined mathematical conditioning. It can have as many atoms as necessary to meet the mathematical balance of the biochemical phenomenon at certain mathematical level, etc. The digital language of biochemistry has a countless number of codes and analogue codes, as well as other information content. These pictures enable us to realize the very essence of functioning of biochemical processes.

Odd rank progression

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | | |
| Rank | 1 | 299 | 95 | 205 | 103 | 197 | > | 900 |
| APa | 10 | 5496 | 1736 | 3770 | 1890 | 3616 | > | 16518 |
| | -144 | -144 | -144 | -144 | -144 | -144 | | |
| APb | 154 | 5640 | 1880 | 3914 | 2034 | 3760 | > | 17382 |
| | G | G | G | G | G | G | | |
| Rank | 299 | 1 | 205 | 95 | 197 | 103 | > | 900 |

↘ ↘ ↘ ↙ ↙

34836

(1+10+(-)144+154+299…+ 103) = 34836;

Even rank progression

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | | |
| Rank | 44 | 256 | 52 | 248 | 146 | 154 | > | 900 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| APa | 796 | 4710 | 950 | 4556 | 2676 | 2830 | > | 16518 |
| | -144 | -144 | -144 | -144 | -144 | -144 | | |
| APb | 940 | 4854 | 1094 | 4700 | 2820 | 2974 | > | 17382 |
| | G | G | G | G | G | G | | |
| Rank | 256 | 44 | 248 | 52 | 154 | 146 | > | 900 |

34836

(44+796+ (-)144+940+256…+ 146) = 34836;

Fig. 9. Odd and even rank progression of aminoacid Gly.

Odd AP-a progression

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | G | G | G | G | G | G | G | G | G | G | G |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| | 29 | 41 | 80 | 92 | 131 | 143 | 182 | 194 | 233 | 245 | 284 | 296 |
| APa | 523 | 741 | 1463 | 1681 | 2403 | 2621 | 3343 | 3561 | 4283 | 4501 | 5223 | 5441 |
| APb | 5127 | 4909 | 4187 | 3969 | 3247 | 3029 | 2307 | 2089 | 1367 | 1149 | 427 | 209 |
| | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 | 5650 |

Odd APa progression – Even rank

| | | | | | | |
|---|---|---|---|---|---|---|
| | G | G | G | G | G | G |
| Rank | 80 | 296 | 92 | 284 | 182 | 194 |
| APa | 1463 | 5441 | 1681 | 5223 | 3343 | 3561 |
| | 1254 | 1254 | 1254 | 1254 | 1254 | 1254 |
| APb | 209 | 4187 | 427 | 3969 | 2089 | 2307 |
| | G | G | G | G | G | G |
| Rank | 296 | 80 | 284 | 92 | 194 | 182 |

Odd APa progression – Odd rank

| | | | | | | |
|---|---|---|---|---|---|---|
| | G | G | G | G | G | G |
| Rank | 29 | 245 | 41 | 233 | 131 | 143 |

| | | | | | | |
|---|---|---|---|---|---|---|
| APa | 523 | 4501 | 741 | 4283 | 2403 | 2621 |
| | -626 | -626 | -626 | -626 | -626 | -626 |
| APb | 1149 | 5127 | 1367 | 4909 | 3029 | 3247 |
| | G | G | G | G | G | G |
| Rank | 245 | 29 | 233 | 41 | 143 | 131 |

Even rank and odd rank

| | G | G | G | G | G | G |
|---|---|---|---|---|---|---|
| Rank par | 80 | 296 | 92 | 284 | 182 | 194 |
| Rank nepar | 29 | 245 | 41 | 233 | 131 | 143 |
| | 51 | 51 | 51 | 51 | 51 | 51 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Rank par | 296 | 80 | 284 | 92 | 194 | 182 |
| Rank nepar | 245 | 29 | 233 | 41 | 143 | 131 |
| | 51 | 51 | 51 | 51 | 51 | 51 |

(80-29) = 51; (296-245) = 51; (92-41) = 51;

Even APa and odd rank

| | G | G | G | G | G | G |
|---|---|---|---|---|---|---|
| APa rank even | 1463 | 5441 | 1681 | 5223 | 3343 | 3561 |
| APa rank odd | 523 | 4501 | 741 | 4283 | 2403 | 2621 |
| | 940 | 940 | 940 | 940 | 940 | 940 |

| | | | | | | |
|---|---|---|---|---|---|---|
| APb odd | 1149 | 5127 | 1367 | 4909 | 3029 | 3247 |
| APb even | 209 | 4187 | 427 | 3969 | 2089 | 2307 |
| | 940 | 940 | 940 | 940 | 940 | 940 |

(80-29) = 51; (296-245) = 51; (92-41) = 51;

Figure 10. Odd AP-a progression, odd APa progression–even rank, odd APa progression –odd rank, even rank and odd rank and even APa and odd rank

3.3 Bio frequency

Insulin is composed of aminoacids with various numerical values. This numerical values are in an irregular order. For example, the first one has 10 atoms, the second one 22. Their frequency is X. Second amino acid has 22 atoms, and the third one 19. Their frequency is Y; etc... Frequency is the measurement for establishment of intervals of numerical values of amino acids in proteins. This value can be positive, negative or a zero value. These frequencies are showing us one completely new dimension of protein sequencing. Through these frequencies we can establish which of aminoacids are of primary, and which are of secondary significance in biochemical processes of insulin. Here is a concrete example:

| | G | I | V | E | Q | C | C | T |
|---|---|---|---|---|---|---|---|---|
| | 10 | 22 | 19 | 19 | 20 | 14 | 14 | 17 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 10 | 12 | -3 | 0 | 1 | -6 | 0 | 3 | -3 |

From 0 to 10 = 10; From 10 to 22 = 12; From 22 to 19 = (-) 3; From 19 to 19 = 0; etc

Schematic representation of the amino acid and frequency we will show in the fig.11.

| G | G | G | G | G | G | G | G | G | G | G | G | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| 1 | 29 | 41 | 44 | 52 | 80 | 92 | 95 | 103 | 131 | 143 | 146 | | |
| 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | > | 114; |

| G | G | G | G | G | G | G | G | G | G | G | G | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| 154 | 182 | 194 | 197 | 205 | 233 | 245 | 248 | 256 | 284 | 296 | 299 | | |
| 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | > | 114; |

◎◎

| | G | G | G | G | G | G | G | G | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Rank | 1 | 299 | 29 | 296 | 41 | 284 | 44 | 256 | > | 1250 |
| Fr | 12 | 13 | 4 | 9 | 9 | 4 | 13 | 12 | > | 76 |
| | -1 | 1 | -5 | 5 | 5 | -5 | 1 | -1 | > | 0 |
| Fr | 13 | 12 | 9 | 4 | 4 | 9 | 12 | 13 | > | 76 |
| | G | G | G | G | G | G | G | G | | |
| Rank | 299 | 1 | 296 | 29 | 284 | 41 | 256 | 44 | > | 1250 |

| | G | G | G | G | G | G | G | G | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Rank | 52 | 248 | 80 | 245 | 92 | 233 | 95 | 205 | > | 1250 |
| APa | 12 | 13 | 4 | 9 | 9 | 4 | 13 | 12 | > | 76 |
| | -1 | 1 | -5 | 5 | 5 | -5 | 1 | -1 | > | 0 |
| APb | 13 | 12 | 9 | 4 | 4 | 9 | 12 | 13 | > | 76 |

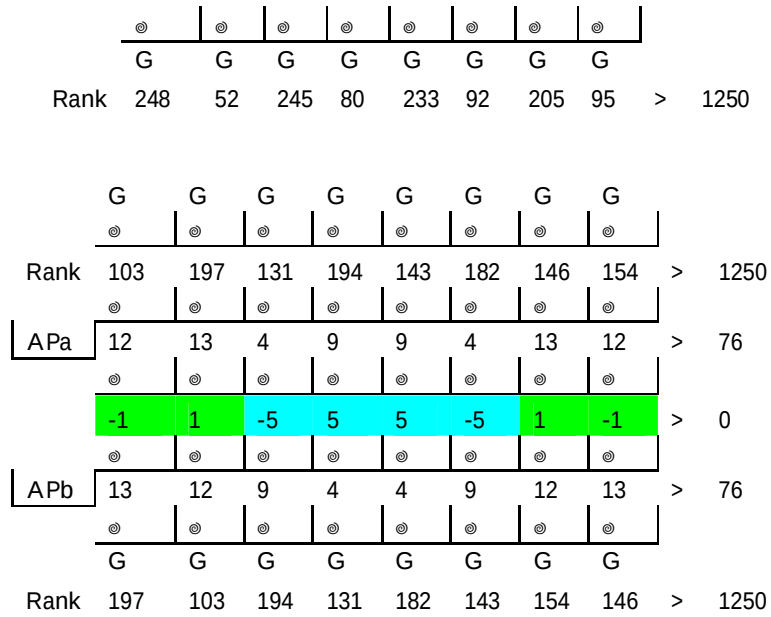


Figure 11. Schematic representation of the amino acid Gly and their frequency

Odd rank and frequency

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G | G | G | G | G | G | G | G | G | G | G | G | | |
| 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| 1 | 29 | 41 | 95 | 103 | 131 | 143 | 197 | 205 | 233 | 245 | 299 | | |
| 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | > | 114; |

Even rank and frequency

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G | G | G | G | G | G | G | G | G | G | G | G | | |
| 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| 44 | 52 | 80 | 92 | 146 | 154 | 182 | 194 | 248 | 256 | 284 | 296 | | |
| 13 | 12 | 4 | 9 | 13 | 12 | 4 | 9 | 13 | 12 | 4 | 9 | > | 114; |

Figure 12. Schematic representation of the amino acid Gly and their odd-even rank and frequency

Therefore, there is a mathematical balance between the group of aminoacids with positive frequency and those of negative frequency. Aminoacids with a positive frequency have a primary role in the mathematical picture of that protein, and the negative frequencies have a secondary role in it. We assume that aminoacids with a positive frequency have a primary role in the biochemical picture of that protein, and the negative frequencies have a secondary role in it. If this really is the case and research on an experimental level proves it, a radically new way of learning about biochemical processes will be opened.

3.4 Analog bio code

Each numerical value has its analogue expression. For example: The analogue expression for number 19 is 91.

91 || 19

In a similar way we can calculate the analogue expression for any numerical value. Our research has shown that analog codes are quantitative characteristics in biochemistry. Analogue biocode is a discrete code that protects and guards genetic information coded in biochemical processes.

This a recently discovered code, and more detailed knowledge about it is necessary.

Odd rank and analog frequency

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | G | G | G | G | G | G | G | G | G | G | G | G | | |
| | | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| | | 1 | 29 | 41 | 95 | 103 | 131 | 143 | 197 | 205 | 233 | 245 | 299 | | |
| Frequencu | > | 12 | 04 | 09 | 13 | 12 | 04 | 09 | 13 | 12 | 04 | 09 | 13 | > | 114; |
| | | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | | |
| Analog frequency | > | 21 | 40 | 90 | 31 | 21 | 40 | 90 | 31 | 21 | 40 | 90 | 31 | > | 546; |

Even rank and analog frequency

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | G | G | G | G | G | G | G | G | G | G | G | G | | |
| | | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| | | 44 | 52 | 80 | 92 | 146 | 154 | 182 | 194 | 248 | 256 | 284 | 296 | | |
| Frequency | > | 13 | 12 | 04 | 09 | 13 | 12 | 04 | 09 | 13 | 12 | 04 | 09 | > | 114; |
| | | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | ◎ | | |
| Analog Frequency | > | 31 | 21 | 40 | 90 | 31 | 21 | 40 | 90 | 31 | 21 | 40 | 90 | > | 546; |

Figure 13. Schematic representation of the amino acid Gly and their odd-even rank and analog frequency

Analogue code is , actually, a discrete code that protects and guards genetic information coded in biochemical processes.

In the previous examples 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 examples we can see that this protein 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.

## 4 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 (Xiao and Chou, 2007). 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 (Xiao and Chou, 2007). The novel approach as presented here may help improve this kind of situation.

## 5 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.

# BIBLIOGRAPHY

[1] Cai, Y.D., and Chou, K.C., 2006. Predicting membrane protein type by functional Domain composition and pseudo amino acid composition. J Theor Biol 238, 395-400.

[2] Cai, Y.D., Zhou, G.P., and Chou, K.C., 2003. Support vector machines for predicting membrane protein types by using functional domain composition. Biophys J 84, 3257-3263.

[3] Cai, Y.D., Zhou, G.P., and Chou, K.C., 2005. Predicting enzyme family classes by hybridizing gene product composition and pseudo-amino acid composition. J Theor Biol 234, 145-149.

[4] Cai, Y.D., Feng, K.Y., Lu, W.C., and Chou, K.C., 2006. Using LogitBoost classifier to predict protein structural classes. J Theor Biol 238, 172-176.

[5] Cai, Y.D., Pong-Wong, R., Feng, K., Jen, J.C.H., and Chou, K.C., 2004. Application of SVM to predict membrane protein types. J Theor Biol 226, 373-376.

[6] Chen, C., Chen, L., Zou, X., and 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.

[7] Chen, L., Feng, K.Y., Cai, Y.D., Chou, K.C., and Li, H.P., 2010. Predicting the network of substrate-enzyme-product triads by combining compound similarity and functional domain composition. BMC Bioinformatics 11, 293.

[8] Chou, K.C., 2011. Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review). J Theor Biol 273, 236-247.

[9] Chou, K.C., and Shen, H.B., 2010a. Cell-PLoc 2.0: An improved package of web-servers for predicting subcellular localization of proteins in various organisms. Natural Science 2, 1090-1103 (openly accessible at http://www.scirp.org/journal/NS/).

[10] Chou, K.C., and Shen, H.B., 2010b. Plant-mPLoc: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. PLoS ONE 5, e11335.

[11] Ding, H., Luo, L., and Lin, H., 2009a. Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. Protein & Peptide Letters 16, 351-355.

[12] Ding, Y.S., Zhang, T.L., and Chou, K.C., 2007. Prediction of protein structure classes with pseudo amino acid composition and fuzzy support vector machine network. Protein & Peptide Letters 14, 811-815.

[13] Ding, Y.S., Zhang, T.L., Gu, Q., Zhao, P.Y., and Chou, K.C., 2009b. Using maximum entropy model to predict protein secondary structure with single sequence. Protein &

Peptide Letters 16, 552-560.
[14] He, Z.S., Zhang, J., Shi, X.H., Hu, L.L., Kong, X.G., Cai, Y.D., and Chou, K.C., 2010. Predicting drug-target interaction networks based on functional groups and biological features. PLoS ONE 5, e9603.
[15] Hu, L., Huang, T., Shi, X., Lu, W.C., Cai, Y.D., and Chou, K.C., 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, X.H., Wang, P., He, Z., Feng, K.Y., Hu, L., Kong, X., Li, Y.X., Cai, Y.D., and Chou, K.C., 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, K.K., Pugalenthi, G., Moller, S., Hartmann, E., Kalies, K.U., Suganthan, P.N., and Martinetz, T., 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, A.M., and 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., and 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, 1263-9.
[20] Mohabatkar, H., 2010. Prediction of cyclin proteins using Chou's pseudo amino acid composition. Protein & Peptide Letters 17, 1207-1214.
[21] Wang, Y.C., Wang, X.B., Yang, Z.X., and Deng, N.Y., 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., and Chou, K.C., 2007. Digital coding of amino acids based on hydrophobic index. Protein & Peptide Letters 14, 871-875.