Atomic Hemoglobin Code

Lutvo Kurić[1],Bosnia and Herzegovina[2],Novi Travnik[3], Kalinska[4]

NLMC GJMR Classification
WH190

***Abstract*-This paper discusses cyberinformation studies of the amino acid composition of *Hemoglobin protein sequences* Q6WN27, in particular the identification of scientific terminology that could describe this phenomenon, ie, the study of genetic information, as well as the relationship between the genetic language of hemoglobin and theoretical aspect of this system and cybernetics. The result of this research show that there is a matrix code for hemoglobin. It also shows that the coding system within the amino acidic language gives detailed information, not only on the amino acid „record", but also on its structure, configuration and its various shapes. The issue of the existence of an hemoglobin code and coding of the individual structural elements of this protein are discussed. Answers to the following questions are sought. Does the matrix mechanism for biosynthesis of this protein function within the law of the general theory of information systems, and what is the significance of this for understanding the genetic language of hemoglobin? What is the essence of existence and functioning of this language? Is the genetic information characterized only by biochemical, or also by cyberinformation principles? The potential effects of physical and chemical, as well as cybernetic and information principles, on the biochemical basis of hemoglobin are also investigated. This paper discusses new methods for developing genetic technologies, in particular more advanced digital technology based on programming, cybernetics, and informational laws and systems, and how this new technology could be useful in medicine, bioinformatics, genetics, biochemistry, and other natural sciences.**

I. INTRODUCTION

The biologic role of any given protein in essential life processes, eg, ***Hemoglobin protein sequences* Q6WN27**, 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 ***Hemoglobin protein sequences* Q6WN27**. Numerical values are them assigned to these factors, enabling them to be measured.

About-:BSC, Senior expert collaborant to Faculty of Economics in Sarajevo E-mail lutvokuric@yahoo.com

In this way it is possible to determine oif connection really exists between the quantitative ratios in the process of transfer of genetic information and the qualitative appearance of the ***Hemoglobin protein sequences* Q6WN27** 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 ***Hemoglobin protein sequences* Q6WN27**.

II. METHODS

***Hemoglobin protein sequences* Q6WN27** can be represented by two different forms, ie, a discrete form and a sequential form. In the discrete form, a molecule of ***Hemoglobin protein sequences* Q6WN27** is represented by a set of discrete codes or a multiple dimension vector. In the sequential form, an ***Hemoglobin protein sequences* Q6WN27** molecule is represent by a series of amino acids according to the order of their position in the ***Hemoglobin protein sequences* Q6WN27.** Therefore, the sequential form can naturally reflect all the information about the sequence order and lenght of an ***Hemoglobin protein sequences* Q6WN27** molecule. The key issue is whether we can develop a different discrete method of representing an ***Hemoglobin protein sequences* Q6WN27** 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?**Expression of Hemoglobin protein sequences Q6WN27** The matrix mechanism of ***Hemoglobin protein sequences* Q6WN27**, the evolution of biomacromolecules and, especially, the biochemical evolution of ***Hemoglobin protein sequences* Q6WN27** language, have been analyzed by the application of cybernetic methods, information theory and system theory, respectively. The primary structure of a molecule of ***Hemoglobin protein sequences* Q6WN27** is the exact specification of its atomic composition and the chemical bonds connecting those atoms.

*Characteristics:*Length : **147**, molecular weight: **16043**, CRC64 check sum: **1EC2AB4138BA7899**

Sequence:>Q6WN27 Hemoglobin subunit beta (Hemoglobin beta chain) (Beta-globin.
MVHLTGEEKAAVTALWGKVNVDEVGGEALGRLLV VYPWTQRFFESFGDLSSPDAVMNNPKVKAHGKKVL GAFSDGLAHLDNLKGTFAQLSELHCDKLHVDPENFR LLGNVLVCVLAHHFGKEFTPQVQAAYQKVVAGVAN ALAHKYH.

Fragment of sequence 147 AA;

Number of atoms

M	V	H	L	T	G	E	E	K	A	.	.	H	K	Y	H	Sum
20	19	20	22	17	10	19	19	24	13	.	.	20	24	24	20	2699
1	2	3	4	5	6	7	8	9	10	.	.	144	145	146	147	10878

Amino acid 1 (AC1) = 20 atoms; AC2=19 atoms; AC3= 20 atoms, etc.

Figure 1. Group of amino acids from 1 to 147.

Notes: Aforementioned aminoacids are positioned from number 1 to 147. 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 ***Hemoglobin protein sequences*** **Q6WN27** and as well as for the other proteins and other sequences in biochemistry.

The first aminoacid in this example has 20 atoms, the second one 19, the third one 20, etc. They have exactly these numbers of atoms because there are many codes in the ***Hemoglobin protein sequences*** **Q6WN27**, 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 20 atoms, the second one 19, the third one 20, 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.

III. Algorithm

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

Arithmetic progression

Step 1 (Amino acids from 1 to 147)

$AC_1 = 20$ atoms; $AC_2 = 19$ atoms; $AC_3 = 20$ atoms;... $AC_{147} = 20$ atoms;

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

$AC_1 = APa1 = 20$;
$(AC_1 + AC_2) = (20+19) = APa2 = 39$;

$(AC_1 + AC_2 + AC_3) = (20+19+20) = APa3 = 59$;
$(AC_1 + AC_2 + AC_3..., + AC_{147}) = APa_{147} = 2699$ atoms;
$APa_{1,2,3,n}$ = Arithmetic progression of amino acids 1,2,3,n

$[APa_1 + APa_2 + APa_3)..., + APa_{147})] = (20+39+59..., + 2699) = 199\ 076$;

$S_1 = 199076$;

Example :

Arithmetic progression 1 (APa)

M	V	H	L	T	G	E	E	.	.	K	Y	H	Sum
20	19	20	22	17	10	19	19	.	.	24	24	20	2699
1	2	3	4	5	6	7	8	.	.	145	146	147	10878

M	V	H	L	T	G	E	E	.	.	K	Y	H	Sum
20	39	59	81	98	108	127	146	.	.	2631	2655	2679	199076
1	2	3	4	5	6	7	8	.	.	145	146	147	10878

(0+20) = 20; (20+19)=39; (20+19+20) = 59; etc.

Figure 1. Arithmetic progression 1 (APa) of amino acids from 1 to 147.

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

Step 2 (Amino acids from 147 to 1)

$AC_{147} = 20$ atoms; $AC_{146} = 24$ atoms; $AC_{145} = 24$ atoms;... $AC_1 = 20$ atoms;

$[AC_{147} + (AC_{147} + AC_{146}) + (AC_{147} + AC_{146} + AC_{145}) \ldots, + (AC_{147} + AC_{146} + AC_{145} \ldots, + AC_1)] = \mathbf{S_2}$;

$AC_{147} = APb_{147} = 20$;
$(AC_{147} + AC_{146}) = (20+24) = APb_{146} = 44$;
$(AC_{147} + AC_{146} + AC_{145}) = (20+24+24) = APb_{145} = 68$;
$(AC_{147} + AC_{146} + AC_{145} \ldots, + AC_1) = APb_1 = 2699$ atoms;
$APb_{147,146,145, \ldots,1}$ = Arithmetic progression of amino acids 147,146,145,...1;

$[APb_{147} + APb_{146} + APb_{145}) \ldots, + APb_1)] = (20+44+68 \ldots, + 2699) = 200\ 376$;

$\mathbf{S_2 = 200\ 376}$;

Example:

Arithmetic progression 2 (APb)

Sum	M	V	H	.	.	N	A	L	A	H	K	Y	H
2699	20	19	20			17	13	22	13	20	24	24	20
10878	1	2	3	.	.	140	141	142	143	144	145	146	147

Sum	M	G	D	.	.	D	E	E	A	G	E	G	N
200 376	2699	2679	2660			153	136	123	101	88	68	44	20
10878	1	2	3	.	.	140	141	142	143	144	145	146	147

(0+20) = 20; (20+24)=44; (20+24+24)=68; etc.

Figure 2. Schematic representation of the arithmetic progression 2 from 147 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.Bio intervals APa and APb (sequence from 1 to 147 AA)

Sequence

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	M	D	H	G	K	L	K	G	P	Y	H	Sum	
	20	16	20	10	24	22	24	10	17	24	20	207	
	1	22	64	65	66	82	83	84	125	146	147		
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		
APa	20	399	1174	1184	1208	1491	1515	1525	2300	2679	2699	>	16194
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		
	-2679	-2280	-1126	-341	-307	0	+307	+341	+1126	+2280	+2679		
	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		
APb	2699	2679	2300	1525	1515	1491	1208	1184	1174	399	20	>	16194
	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑		
	M	V	E	G	K	K	K	G	T	Q	H	Sum	
	20	19	19	10	24	24	24	10	17	20	20	207	
	1	2	23	65	66	67	83	84	85	126	147		

(20-2699) = (-)2679; (399-2679) = (-)2280; (1174-2300) = (-)1126; etc.

APa1=20; APa22=399; APa64=1174; etc.

APb1=2699; APb2=2679; APb23=2300; etc.

APa1= APb147; APa22= APb126;

APa64= APb85; APa65= APb84; APa66 = APb83

APa82 = APb67; APa83 = APb66; APa84 = APb65; APa125 = APb23;

etc.

Table 1. Schematic representation of the bio intervals APa and APb (sequence from 1 to 147 AA).

Notes: Namely, having mathematically analyzed the Bio intervals model of ***Hemoglobin protein sequences*** **Q6WN27** (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: Bio intervals model of ***Hemoglobin protein sequences*** **Q6WN27** should, in fact, be „remodelled—into a periodic system.

Table 2. Bio intervals Apa and APb (sequence from 2 to 146 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	D	H	G	K	L	K	G	P	Y	Sum	
	16	20	10	24	22	24	10	17	24	167	
	22	64	65	66	82	83	84	125	146		
	↓	↓	↓	↓	↓	↓	↓	↓	↓		
APa	379	1154	1164	1188	1471	1495	1505	2280	2659	>	13295
	↓	↓	↓	↓	↓	↓	↓	↓	↓		
	-2280	-1126	-341	-307	0	+307	+341	+1126	+2280		
	↑	↑	↑	↑	↑	↑	↑	↑	↑		
APb	2659	2280	1505	1495	1471	1188	1164	1154	379	>	13295
	↑	↑	↑	↑	↑	↑	↑	↑	↑		
	V	E	G	K	K	K	G	T	Q	Sum	
	19	19	10	24	24	24	10	17	20	167	
	2	23	65	66	67	83	84	85	126		

Table 2. Schematic representation of the bio intervals APa and APb (sequence from 2 to 146 AA).

Table 3. Bio intervals Apa and APb (sequence from 3 to 145 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	R	P	L	H	K		
	26	17	22	20	24		
	31	59	89	117	145		
	↓	↓	↓	↓	↓		
APa	508	1035	1581	2108	2616	>	7848
	↓	↓	↓	↓	↓		
	-2108	-1073	0	+1073	+2108		
	↑	↑	↑	↑	↑		
APb	2616	2108	1581	1035	508	>	7848
	↑	↑	↑	↑	↑		
	H	L	K	S	H		
	20	22	24	14	20		
	3	32	60	90	118		

Table 3. Schematic representation of the bio intervals APa and APb (sequence from 3 to 145 AA).

Table 4. Bio intervals Apa and APb (sequence from 4 to 144 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	E	G	A	A	E	N	H		
	19	10	13	13	19	17	20		
	7	25	71	77	122	140	144		
APa	68	388	1237	1335	2184	2504	2572	>	10288
	-2504	-2116	-947	0	+947	+2116	+2504		
APb	2572	2504	2184	1335	1237	388	68	>	10288
	L	E	G	F	H	F	A		
	22	19	10	23	20	23	13		
	4	8	26	72	78	123	141		

Table 4. Schematic representation of the bio intervals APa and APb (sequence from 4 to 144 AA).

Table 5. Bio intervals Apa and APb (sequence from 5 to 143 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	E	S	G	A	A		
	19	14	10	13	13		
	8	73	75	139	143		
APa	65	1252	1278	2465	2530	>	7590
	-2465	-1213	0	+1213	+2465		
APb	2530	2465	1278	1252	65	>	7590
	T	K	D	L	N		
	17	24	16	22	17		
	5	9	74	76	140		

Table 5. Schematic representation of the bio intervals APa and APb (sequence from 5 to 143 AA).

Table 6. Bio intervals Apa and APb (sequence from 6 to 142 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	T	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	L	L	W	T	D	K	A
	22	22	27	17	16	24	13
	15	32	38	39	53	60	63
APa	169	471	599	616	873	1000	1056
	-2331	-1860	-1430	-1285	-1011	-627	-444
APb	2500	2331	2029	1901	1884	1627	1500
	G	W	L	T	Q	A	V
	10	27	22	17	20	13	19
	6	16	33	39	40	54	61

T
17
85
1444
O
1444
H
20
64

Q	D	G	N	A	Q	L
20	16	10	17	13	20	22
88	95	108	109	116	132	142
1500	1627	1884	1901	2029	2331	2500
+444	+627	+1011	+1285	+1430	+1860	+2331
1056	1000	873	616	599	471	169
F	L	K	N	V	H	K
23	22	24	17	19	20	24
86	89	96	109	110	117	133

Table 6. Schematic representation of the bio intervals APa and APb (sequence from 6 to 142 AA).

Table 7. Bio intervals Apa and APb (sequence from 7 to 141 AA)

M	V	H	L	T	G	E	E	K	A	A	V	T	A	L	W	G	K	V	N	V
D	E	V	G	G	E	A	L	G	R	L	L	V	V	Y	P	W	T	Q	R	F
F	E	S	F	G	D	L	S	S	P	D	A	V	M	N	N	P	K	V	K	A
H	G	K	K	V	L	G	A	F	S	D	G	L	A	H	L	D	N	L	K	G
T	F	A	Q	L	S	E	L	H	C	D	K	L	H	V	D	P	E	N	F	R
L	L	G	N	V	L	V	C	V	L	A	H	H	F	G	K	E	F	I	P	Q
V	Q	A	A	Y	Q	K	V	V	A	G	V	A	N	A	L	A	H	K	Y	H

	K	N	E	Q	K	D	L	F	Q	G	A
	24	17	19	20	24	16	22	23	20	10	13
	9	20	23	40	67	80	106	123	126	137	141
APa	62	256	310	626	1124	1344	1842	2158	2212	2406	2468
	-2406	-2150	-1902	-1532	-718	o	718	1532	1902	2150	2406
APb	2468	2406	2212	2158	1842	1344	1124	626	310	256	62
	E	A	V	V	R	V	N	L	T	V	V
	19	13	19	19	26	19	17	22	17	19	19
	7	10	21	24	41	68	81	107	124	127	138

Table 7. Schematic representation of the bio intervals APa and APb (sequence from 4 to 144 AA).

Notes:Consequently, the periodic law allows for only one form of the amino acid bio interval (Table 1,2,3,4,5,6 and 4) which can show both the relationships among the amino acids, and their position in peptide chain. As we see, protein code can be developed into the periodic system (and the table), in which each of 20 natural amino acids has an exactly established place.etc.Biological particularity of proteins depends on the order of amino acids in their molecules. Change of that order will lead to the change of their biological particularity. The base parameters that determine the change of status of biosynthesis matrix of macromolecules are the system entropy, volume of information transferred and the level of probability that the genetic information will be transferred as a whole.Each peptide chain can have as many atoms as necessary to meet the mathematical balance of the biochemical phenomenon at certain mathematical level. By using the system and information procedures and methods in studying biochemistry, we can analyze the effects of the classical and information parameters in the protein evolution process.

We would particularly like to stress here that the genetic, as well as biochemical information in a broader sense of the word, is determined and characterized by very complex cybernetic and information principles. The constantans in those principles are: the number of atoms and molecules, atomic numbers, atomic weight, physical and chemical parameters, even and odd values, codes and analogue codes, standard deviations, frequencies, primary and secondary values, and many other things.

IV. Conclusion

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 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. Since the concept of Chou's pseudo amino acid composition was proposed [1,2], there have been many efforts to try to use various digital numbers to represent the 20 native amino acids in order to better reflect the sequence-order effects through the vehicle of pseudo amino acid composition. Some investigators used complexity measure factor [3], some used the values derived from the cellular automata [4-7], some used hydrophobic and/or hydrophilic values [8-16], some were through Fourier transform [17,18], and some used the physicochemical distance [19]. It is going to be possible to use a completely new strategy of research in genetics in the future. However, close observation of all these relationships, which are the outcomes of periodic laws (more specifically the law of binary coding), stereo-chemical and digital structure of proteins.

V. Disclosure

The author reports no conflict of interest in this research.

VI. References

1) K.C. Chou, Gene Cloning & Expression Technologies, Chapter 4 (Weinrer, P.W., and Lu, Q., Eds.), Eaton Publishing, Westborough, MA (2002), pp. 57-70.
2) K.C. Chou, Prediction of protein cellular attributes using pseudo amino acid composition PROTEINS: Structure, Function, and Genetics (Erratum: ibid., 2001, Vol.44,60) 43 (2001) 246-255.
3) X. Xiao, S. Shao, Y. Ding, Z. Huang, Y. Huang, K. C. Chou, Using complexity measure factor to predict protein subcellular location, Amino Acids 28 (2005) 57-61.
4) Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, K. C. Chou, Using cellular automata to generate Image representation for biological sequences, Amino Acids 28 (2005) 29-35.
5) X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, K. C. Chou, An Application of Gene Comparative Image for Predicting the Effect on Replication Ratio by HBV Virus Gene Missense Mutation, Journal of Theoretical Biology 235 (2005) 555-565 X. Xiao, S. H. Shao, Z. D. Huang, K. C. Chou, Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor,
6) Journal of Computational Chemistry 27 (2006) 478-482. X. Xiao, S. H. Shao, Y. S. Ding, Z. D. Huang, K. C. Chou, Using cellular automat images and pseudo amino acid composition to predict protein sub-cellular location, Amino Acids 30 (2006) 49-54.
7) K. C. Chou, Using amphiphilic pseudo amino acid composition to predict enzym subfamily classes, Bioinformatics 21 (2005) 10-19. K. C. Chou, Y. D. Cai, Prediction of membrane protein types by incorporating amphipathic effects,
8) Journal of Chemical Information and Modeling 45 (2005) 407-413. Z. P. Feng, Prediction of the subcellular location of prokaryotic proteins based on a new representation of the amino acid composition,
9) Biopolymers 58 (2001) 491-499.Z. P. Feng, An overview on predicting the subcellular location of a protein, In Silic Biol 2 (2002) 291-303 M. Wang, J. Yang, Z. J. Xu, K. C. Chou, SLLE for predicting membrane protei types, Journal of Theoretical Biology 232 (2005) 7-15.S. Q. Wang, J. Yang, K. C. Chou,
10) Using stacked generalization to predict membrane protein types based on pseudo amino acid composition, Journal of Theoretical Biology,
11) in press (2006) doi:10.1016/j.jtbi.2006.1005.1006. M. Wang, J. Yang, G. P. Liu, Z. J. Xu, K. C. Chou,

Weighted-support vector machine for predicting membrane protein types based on pseudo amino acid composition, Protein
12) Engineering, Design, and Selection 17 (2004) 509-516. 15. S. W. Zhang, Q. Pan, H. C. Zhang, Z. C. Shao, J. Y. Shi, Prediction protein homo- oligomer types by pseudo amino acid composition: Approached with an improved feature extraction and naive Bayes feature fusion,
13) Amino Acids 30 (2006) 461-468. Y. Gao, S. H. Shao, X. Xiao, Y. S. Ding, Y. S. Huang, Z. D. Huang, K. C. Chou, Using pseudo amino acid composition to predict protein subcellular location: approached with
14) Lyapunov index, Bessel function, and Chebyshev filter, Amino Acids 28 (2005) 373- 376.Y. Z. Guo, M. Li, M. Lu, Z. Wen, K. Wang, G. Li, J. Wu, Classifying G protein-coupled receptors and nuclear receptors based on protein power spectrum from fast Fourier transform, Amino Acids 30 (2006) 397-402.
15) H. Liu, M. Wang, K. C. Chou, Low-frequency Fourier spectrum for predicting membrane protein types, Biochem Biophys Res Commun 336 (2005) 737-739. K. C. Chou, Prediction of protein subcellular locations by incorporating quasi-sequence-order effect, Biochemical & Biophysical Research Communications 278 (2000) 477-483. K. C. Chou,
16) A novel approach to predicting protein structural classes in a (20-1)- amino acid composition space, Proteins: Structure, Function & Genetics 21 (1995) 319- 344. K. C. Chou, C. T. Zhang, Predicting protein folding types by distance functions that
17) make allowances for amino acid interactions, Journal of Biological Chemistry 269 (1994) 22014-22020. K. C. Chou, C. T. Zhang, Review: Prediction of protein structural classs, Critical Reviews in Biochemistry and Molecular Biology 30 (1995) 275-349.K. C. Chou, D. W. Elrod, Protein subcellular location prediction, Protein Engineering 12 (1999) 107-11
18) K. C. Chou, Review: Prediction of protein structural classes and subcellular locations Current Protein and Peptide Science 1 (2000) 171-208. K. C. Chou, D. W. Elrod, Prediction of membrane protein types and subcellular locations, PROTEINS: Structure, Function, and Genetics 34 (1999) 137-153. 26. K. C. Chou, D. W. Elrod, Prediction of enzyme family classes, Journal of Proteome Research 2 (2003) 183-190. K. C. Chou, Y. D. Cai, Predicting enzyme family class in a hybridization space Protein Science 13 (2004) 2857-2863
19) K. C. Chou, D. W. Elrod, Bioinformatical analysis of G-protein-coupled receptors,Journal of Proteome Research 1 (2002) 429-433.K. C. Chou, Prediction of G-protein-coupled receptor classes, Journal of Proteome Research 4 (2005) 1413-1418.K. C. Chou, Y. D. Cai, Prediction of protease types in a hybridization space, Biochem. Biophys. Res. Comm. 339 (2006) 1015-1020.
20) K. C. Chou, Y. D. Cai, Predicting protein-protein interactions from sequences in a hybridization space, Journal of Proteome Research 5 (2006) 316-322.
21) K. C. Chou, Y. D. Cai, W. Z. Zhong, Predicting networking couples for metabolic pathways of Arabidopsis, EXCLI Journal 5 (2006) 55-65.K. C. Chou, Y. D. Cai, Predicting protein quaternary structure by pseudo amino acid composition,
22) PROTEINS: Structure, Function, and Genetics 53 (2003) 282-289. Kurić L. The digital language of amino acids. Amino Acids (2007) 653-661. Kurić L. The Atomic Genetic Code. J. Comput Sci Biol 2 (2009) 101-116.
23) Kurić L. Mesure complexe des caracteristiques dynamiques de series temporelles–Journal de la Societe de statistique de Paris"- tome 127, No 2.1986. Kurić L. The Insulin Bio Code - Zero Frenquencies, GJMR Vol. 10 Issue 1: 15 May 2010.Kurić L.Molecular biocoding of insulin, Advances and Applications in Bioinformatics and Chemistry, Jul. 2010.p.45 – 58.
24) Kurić L. The Insulin Bio Code – Prima sequences, GJMR Vol. 1 Issue 1: 15 June 2010.