Advances and Applications in Bioinformatics and Chemistry

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Molecular biocoding of insulin

This article was published in the following Dove Press journal:
Advances and Applications in Bioinformatics and Chemistry
23 July 2010
Number of times this article has been viewed

Lutvo Kurić

Novi Travnik, Kalinska,
Bosnia and Herzegovina

Abstract: This paper discusses cyberinformation studies of the amino acid composition of insulin, 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 proteins and theoretical aspects of this system and cybernetics. The results of this research show that there is a matrix code for insulin. It also shows that the coding system within the amino acid language gives detailed information, not only on the amino acid "record", but also on its structure, configuration, and various shapes. The issue of the existence of an insulin 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 insulin? What is the essence of existence and functioning of this language? Is the genetic information characterized only by biochemical principles or it is also characterized by cyberinformation principles? The potential effects of physical and chemical, as well as cybernetic and information principles, on the biochemical basis of insulin 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.

Keywords: human insulin, insulin model, biocode, genetic code, amino acids

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 own 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 cybernetic and information characteristics. Indeed, the sequencing of the molecule is determined not only by distint 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. Numeric 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

Correspondence: Lutvo Kurić
72290 Novi Travnik, Kalinska 7,
Bosnia and Herzegovina
Tel +3 876 176 3917
Email lutvokuric@yahoo.com

© 2010 Kurić, publisher and licensee Dove Medical Press Ltd. This is an Open Access article which permits unrestricted noncommercial use, provided the original work is properly cited.

the qualitative appearance of the insulin molecule. To select these factors, preference is given to classical physical and chemical parameters, including the numbers of atoms in the relevant amino acids, their analog values, the position of 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

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 amino acids according to the order of their position in the 1AI0 chains. 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 acid codes, what kind of numerical values 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? This section discusses how biochemical function of the molecule is determined by cybernetic information principles.

Expression of insulin code matrix 1AI0

The matrix mechanism of insulin, evolution of biomacromolecules and, especially, biochemical evolution of the 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 an expression of the exact specification of its atomic composition and the chemical bonds between those atoms.

R6 insulin hexamer (d1ai02)

The structure 1AI0 has a total of 12 chains. Of these, two are sequence-unique identical chains of BFHJL (Figure 1).

>1AI0:BFHJL	FVNQHLCGSHLVEALYLVCGERGFFYTPKT

Insulin B chain
Number of atoms

F	V	N	Q	H	I	C	G	S	H
23	19	17	20	20	22	14	10	14	20
1	2	3	4	5	6	7	8	9	10

L	V	E	A	L	Y	L	V	C	G
22	19	19	13	22	24	22	19	14	10
11	12	13	14	15	16	17	18	19	20

E	R	G	F	I	Y	T	P	K	T
19	26	10	23	22	24	17	17	24	17
21	22	23	24	25	26	27	28	29	30

Rank = 1,2,3,n

Figure 1 Number of atoms in insulin B chain.
Notes: And in that dimension we can find an explanation for the given empirical reality. Aforementioned amino acids are positioned from numbers 1 to 30. Numbers 1,2,3,n... present the position of a certain amino acid.

This positioning is of key importance for understanding the programmatic, cybernetic, and information elements of this protein. The scientific key for interpretation of biochemical processes is the same for insulin as for other proteins and sequences in biochemistry. The first amino acid in this example has 23 atoms, the second 19, the third 17, etc. They have exactly these numbers of atoms because there are many codes in the insulin molecules, analog codes, and other coded features. In fact, there is a cybernetic algorithm in which it is "recorded" that the first amino acid has to have 23 atoms, the second 19, the third 17, 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 characteristics of the organism.

Bio insulin code

The bio insulin code is an area of biomacromolecular processes in biochemistry (chemical engineering, bioprocess engineering, information technology, biorobotics) that treats signals as stochastic processes, dealing with their biosignal properties (eg, frequencies, mean, and covariance). In this context, biocodes are modeled as functions consisting of both deterministic and stochastic components. A simple example and also a common model of many bio systems is a signal Bio (t) that consists of a deterministic part x (t) as a white biocode (Figure 2).

$$\text{Bio}(t) = [X1,2,3,n(t) - X1,2,3,n(t)]$$

$$X(t) = \text{Biocode } 1,2,3,n$$

(A,B,C) – (B,C,D) = biocode 1
(B,C,D – C,D,E) = biocode 2
(D,E,F – E,F,G) = biocode 3 (1) etc.

where A, B, C, and n represent the connection of group amino acid positions, 1,2,3,n

F	V	N	Q	H	I
23	19	17	20	20	22
1	2	3	4	5	6

Bio(t) = (F,V,N) – (Q,H,I) > (23,19,17) – (20,20,22)

↓

(231917 – 202022) = 29895

Figure 2 Fragment of insulin B chain.
Notes: The connection of the groups of these amino acids gives us number 29895 as a result. That was the code that represent one of the quantitative characteristics of the given genetic information.

Thus, the union of these amino acids generates the number 29895, which is a code representing one of the quantitative characteristics of the given genetic information. In a similar way we can calculate biocodes for other groups of amino acids, which are connected by various biocodes and analog codes, as well as other quantitative features. Connection is one of the numerical expressions that connect various corresponding features in biochemistry. It has a very prominent place in our mathematical understanding of all processes in biochemistry. This is a recently discovered phenomenon, the role and significance of which will hopefully be clarified in the future. Those bioprocesses are well correlated, and the autocorrelation function is a biocode:

Bio I = (X1,2,3,n(t) – X1,2,3,n(t)) = [(–)Y ↔ (+)Y]
[(–)Y = (+)Y] (2)

where Y represents the result as a functional biocode.

Insulin B chain

Mathematical evidence is provided here to prove that in the biochemistry of insulin there really is a 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.

Bio(t) = X1,2,3,n(t) – X1,2,3,n(t)

Where:
X(t) = biocodes 1,2,3,n
Bio 1 = (A,B,C) – (B,C,D) = biocodes 1

Examples are presented in Figure 3.

Biocode 1

F	V	N	Q	H	I	C
23	19	17	20	20	22	14
1	2	3	4	5	6	7
↘	↓	↙	↓	↘	↓	↙
	59		20		56	

↓

59,20,56 > 592056

Q	H	I	C	G
20	20	22	14	10
4	5	6	7	8
↓	↘	↓	↙	↓
20		56		10

↓

20,56,10 > 205610

Bio 1 = (592056 – 205610) = 386 446

Biocode 2

H	I	C	G	S	H	L
20	22	14	10	14	20	22
5	6	7	8	9	10	11
↘	↓	↙	↓	↘	↓	↙
	56		10		56	

↓

56,10,56 > 561056

G	S	H	L	V
10	14	20	22	19
8	9	10	11	12
↓	↘	↓	↙	↓
10		56		19

↓

10,56,19 > 105619

Bio 2 = (561056 – 105619) = 455 437

Biocode 3

S	H	L	V	E	A	L
14	20	22	19	19	13	22
9	10	11	12	13	14	15
↘	↓	↙	↓	↘	↓	↙
	56		19		54	

↓

56,19,54 > 561954

V	E	A	L	Y	L	V
19	19	13	22	24	22	19
12	13	14	15	16	17	18
↓	↘	↓	↙	↘	↓	↙
19		54			65	

↓

19,54,65 > 195465

Bio 3 = (561954 – 195465) = 366 489

Biocode 4

E	A	L	Y	L	V	C
19	13	22	24	22	19	14
13	14	15	16	17	18	19
↘	↓	↙	↘	↓	↙	↓
	54			65		14

↓

54,65,14 > 546514

Y	L	V	C	G	E	R
24	22	19	14	10	19	26
16	17	18	19	20	21	22
↘	↓	↙	↓	↘	↓	↙
	65		14		55	

↓

65,14,55 > 651455

Bio 4 = (546514 – 651455) = (–) 104 941

Biocode 5

C	G	E	R	G
14	10	19	26	10
19	20	21	22	23
↓	↘	↓	↙	↓
14		55		10

↓

14,55,10 > 145510

G	E	R	G	F	I	Y
10	19	26	10	23	22	24
20	21	22	23	24	25	26
↘	↓	↙	↓	↘	↓	↙
	55		10		69	

↓

55,10,69 > 551069

Bio 5 = (145510 – 551069) = (–) 405 559

Biocode 6

G	F	I	Y	T
10	23	22	24	17
23	24	25	26	27
↓	↘	↓	↙	↓
10		69		17

↓

10,69,17 > 106917

F	I	Y	T	P	K	T
23	22	24	17	17	24	17
24	25	26	27	28	29	30
↘	↓	↙	↓	↘	↓	↙
	69		17		58	

↓

69,17,58 > 691758

Bio 6 = (106917 – 691758) = (–) 681 061

Figure 3 Examples of biocodes.
Notes: The connection of the unions of these amino acids gives us number (–) 681061 as a result. That was the code that represent one of the quantitative characteristics of the given genetic information.

The biocodes presented in Figure 3 are calculated using the relationship between corresponding groups of amino acids. These are groups with different numbers of amino acids. There are different ways and methods of selecting these groups of amino acids, and it is hoped that science will soon determine which method is most efficient. Some biocodes have a positive numeric value and some have a negative one. Table 1 gives these codes (see also Figure 4).

Number of atoms

G	I	V	E	Q	C	C	T	S	I	C
10	22	19	19	20	14	14	17	14	22	14
1	2	3	4	5	6	7	8	9	10	11

S	L	Y	Q	L	E	N	Y	C	N
14	22	24	20	22	19	17	24	14	17
12	13	14	15	16	17	18	19	20	21

where amino acid G has 10 atoms, amino acids I 22 atoms, V 19 atoms, etc.

In this example there are 21 amino acids.

Rank = 1,2,3,n.

Bio(t) = [X1,2,3,n(t) – X1,2,3,n(t)] (3)

Where

X(t) = Biocodes 1,2,3,n

Figure 4 The formula for calculating of biocodes.
Notes: Aforementioned amino acids are positioned from number 1 to 21. The first amino acid in this example has 10 atoms, the second one 22, the third one 19, etc…. Why do they have exactly this many atoms? There is a program-cybernetic algorithm in which it is "recorded" that the first amino acid has to have 10 atoms, the second one 22, the third one 19, etc… Mathematical expression for this algorithm are biocodes 1,2,3,n.

Table 1 Overview of positive and negative values of biocodes for insulin chain B showing some of the quantitative characteristics of the insulin molecule and the exact mathematical balance between its components. Schematic representation of the biocode processing (1)

Biocodes 1,2,3,n	Biocodes 1,2,3,n	(1–2) (+)	(1–2) (–)
1	2	3	4
592056	205610	Bio 1 = 386446	–
205610	561056	–	Bio 2 = –355446
561056	105619	Bio 3 = 455437	–
105619	561954	–	Bio 4 = –456335
561954	195465	366489	–
195465	546514	–	–351049
546514	651455	–	–104941
651455	145510	505945	–
145510	551069	–	–405559
551069	106917	444152	–
106917	691758	–	–584841
691758	592056	99702	–
SUM		(+) 2258171	(–) 2258171

Biocode 1

G	I	V	E	Q	C
10	22	19	19	20	14
1	2	3	4	5	6
↘	↓	↙	↘	↓	↙
10,22,19			19,20,14		

GIV = 102219

EQC = 192014

Bio 1 = (GIV – EQC)

Bio 1 = (102219 – 192014) = (–) 89795

In this example we calculated the numbers of atoms in groups of three amino acids of the insulin chain.

Biocode 2

E	Q	C	C	T	S
19	20	14	14	17	14
4	5	6	7	8	9
↘	↓	↙	↘	↓	↙
19,20,14			14,17,14		

EQC = 192014

CTS = 141714

Bio 2 = (EQC – CTS)

Bio 2 = (192014 – 141714) = (+) 50300

Biocode 3

C	T	S	I	C	S
14	17	14	22	14	14
7	8	9	10	11	12
↘	↓	↙	↘	↓	↙
14,17,14			22,14,14		

CTS = 141714

ICS = 221414

Bio 3 = (CTS – ICS)

Bio 3 = (141714 – 221414) = (–) 79700 etc.

Biocodes are also shown in Table 2.

Figure 5 Examples of the connection of numbers of atoms in group of three amino acids of the insulin chain.
Notes: Biocodes 1,2,3,n, represent one of the quantitative characteristics of the given genetic information.

Analog biocode

Insulin A chain

Each numerical value has its analog expression. For example, the analog expression for number 19 is 91 (91 || 19). In a similar way we can calculate the analog expression for any numerical value (Figure 6). Our research has shown that analog codes are quantitative characteristics in biochemistry. An analog biocode is a discrete code that protects and guards genetic information coded in biochemical processes. This is a recently discovered code, and more detailed knowledge about it is necessary.

Analog biocode 1

G	I	V	E	Q	C
10	22	19	19	20	14
1	2	3	4	5	6
↘	↓	↙	↘	↓	↙
10,22,19			19,20,14		
↓			↓		
Analog code			Analog code		
↓			↓		
912201			410291		

GIV = 912201

EQC = 410291

Analog biocode 1 = (912201 – 410291) = (+) 501910

Analog code of number 102219 is number 912201 (912201 || 102219)

Analog biocode 2

E	Q	C	C	T	S
19	20	14	14	17	14
4	5	6	7	8	9
↘	↓	↙	↘	↓	↙
19,20,14			14,17,14		
↓			↓		
Analog code			Analog code		
↓			↓		
410291			417141		

EQC = 410291

CTS = 417141

Analog bio 2 = (410291 – 417141) = (–) 6850, etc.

Figure 6 Examples of the analog codes of insulin chain A.
Notes: The analog expression for number 19 20 14 is 41 02 91, and number 14 17 14 is 417 141. Analog code is, actually, a discrete code that protects and guards genetic information coded in biochemical processes.

Positive and negative analog codes of insulin chain A are shown in Table 3.

Correlation between code and analog code

Arithmetical expression for biocode = 190 001; there is a correlation between the code and analog code = 899 809; 100091 | | 190001 > 89 910; 908998 | | 899809 > 9189. This phenomenon will be investigated into more detail in the future.

Insulin model

The structure 1AI0 has in total 12 chains, ie, A,B,C,D,E,F, G,H,I,J,K,L. In this group of chains there are three unions with four chains each. Each of these three groups of chains has an identical number of amino acids, identical number of atoms, and an identical sum of position numbers for these amino acids.

Groups A,B,K,L

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:A

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

1AI0:K

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
256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276

1AI0:L

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
277	278	279	280	281	282	283	284	285	286	287	288	289	290	291

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
292	293	294	295	296	297	298	299	300	301	302	303	304	305	306

Number of atoms = 1880

RANK = 15.657

Groups C,D,I,J

1AI0:C

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
52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72

1AI0:D

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
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87

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
88	89	90	91	92	93	94	95	96	97	98	99	100	101	102

1AI0:I

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
205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225

1AI0:J

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
226	227	228	229	230	231	232	233	234	235	236	237	238	239	240

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
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Number of atoms = 1880; RANK = 15.657

Groups E,F,G,H

1AI0:E

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
103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123

1AI0:F

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
124	125	126	127	128	129	130	131	132	133	134	135	136	137	138

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
139	140	141	142	143	144	145	146	147	148	149	150	151	152	153

1AI0:G

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
154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174

1AI0:H

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
175	176	177	178	179	180	181	182	183	184	185	186	187	188	183

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
190	191	192	193	194	195	196	197	198	199	200	201	202	203	204

Figure 7 Group of chains E,F,G,H.

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

Table 2 Overview of the mathematical balance between the positive and negative values of the biocodes of the insulin A chain. Schematic representation of biocode processing (2)

Biocodes 1,2,3,n	Biocodes 1,2,3,n	(1–2) (+)	(1–2) (–)
1	2	3	4
102219	192014	–	Bio 1 = (–) 89795
192014	141714	Bio 2 = (+) 50300	–
141714	221414	–	Bio 3 = (–) 79700
221414	222420	–	Bio 4 = (–) 1006
222420	221917	Bio 5 = (+) 503	–
221917	241417	–	Bio 6 = (–) 19500
241417	102219	Bio 7 = (+) 139198	–
SUM		(+) 190001	(–) 190001

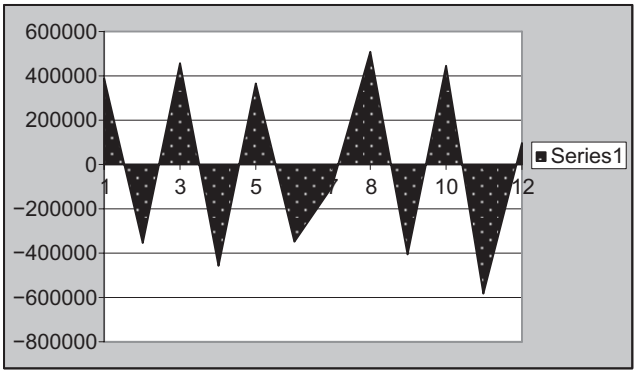

Figure 8 A schematic diagram to show the positive and negative output biocodes.

Number of atoms = 1880; RANK = 15.657

	A,B,K,L	C,D,I,J	E,F,G,H
Number of atoms	1880	1880	1880
Rank	15657	15657	15657
Sum	17537	17537	17537

Figure 9 Groups of chains (A,B,K,L), (C,D,I,J) and (E,F,G,H).
Notes: Each of these three unions of chains has an identical number of amino acids, identical number of atoms and identical sum of position numbers for the said amino acids.

Figure 8 shows the atomic structure 1A10.

Each of these three groups of chains has an identical number of amino acids, an identical number of atoms, and an identical sum of position numbers for these amino acids.

Code 6

The insulin matrix code can be developed into a periodic system in which each of 20 natural amino acids has an exactly established place. On the basis of this system, amino acids are distributed according to the criteria of mathematical similarity or difference (Figure 10).

Table 3 mathematical balance between positive and negative values of analog codes of insulin chain A. Schematic representation of biocode processing (3)

Analog bio impulse 1,2,3,n	Analog bio impulse 1,2,3,n	(1–2) (+)	(1–2) (–)
1	2	3	4
912201	410291	Analog bio (+) 501910	–
410291	417141	–	Analog bio (–) 6850
417141	414122	Analog bio (+) 3019	–
414122	24222	Analog bio (+) 389900	–
24222	719122	–	Analog bio (–) 694900
719122	714142	Analog bio (+) 4980	–
714142	912201	–	Analog bio (–) 198059
SUM		(+) 899 809	(–) 899 809

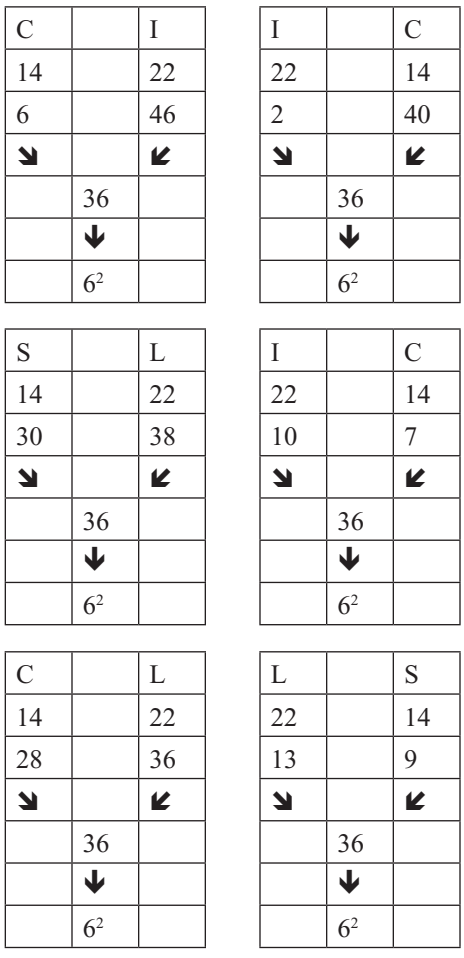

Figure 10 Distribution of amino acids according to the criteria of mathematical similarity or difference.
Notes: The sum of the atoms in these amino acids, we, as a result, are given the number 36, or 6^2. That was the code that represents one of the quantitative characteristics of the given genetic information.

In this example, the quantitative characteristics of genetic information are given in the sign of the code 6. That code links the corresponding groups of amino acids.

Rank of the biocodes

Some groups of amino acids that are discussed here are connected by the discrete code 11. Here are some examples (Figure 11)

Fragment 1

I		S
22		14
10		12
↘		↙
	22	
↙		↘
11		11

(10 + 12) = 22; 22 = (11 + 11)

Fragment 2

S		L
14		22
9		13
↘		↙
	22	
↙		↘
11		11

Fragment 3

C		L
14		22
6		16
↘		↙
	22	
↙		↘
11		11

etc.

Figure 11 Groups of amino acids connected by the discrete code 11.
Notes: In the aforementioned examples, amino acids are connected by the code 6, and their position numbers are connected with code 11. As it can be observed, quantitative characteristics of the biochemistry of insulin apply to the position number.

In the aforementioned examples, amino acids are connected by the code 6, and their position numbers are connected with code 11. As can be seen, the quantitative characteristics of the biochemistry of insulin apply to the position number.

Biocode 11

The aforementioned code 11 also connects groups of amino acids. We can see that in the following example (Figure 12).

G	V	Q	C	T	C	Y	Q	E	Y	N	SUM	Code
10	19	20	14	17	14	24	20	19	24	17	198	(198:X) = 11:
1	3	5	7	8	11	14	15	17	19	21	121	(121:11) = 11:
–	–	–	–	–	–	–	–	–	–	–	319	(319:Y) = 11;
(198 + 121 + 319) = (11 × Y); (198 121 319) = (11 × Y1); (319 121 198) = (11 × Yn);												

Figure 12 Discrete code 11 connects groups of amino acids.
Notes: These groups of amino acids are connected by the discrete code 11.

G	V	Q	C	T	C	Y	Q	E	Y	N
10	19	20	14	17	14	24	20	19	24	17
↓										
10192014171424201924 17										
↓										
(11 × B)										

We will give a few more examples of coding of amino acids of insulin with code 11: (Figure 13–15).

Biocode 11(1)

G		N
10		17
1		21
↘		↙
	121	
	↓	
	11^2	

Figure 13 Connection of the rank of amino acids G and N.
Notes: In this example we connected the position number of amino acids G and N, and got number 121 as result.

In this example we connected the position number of amino acids G and N, and got number 121 or 112 as result.

Biocode 11(2)

V		Y
19		24
3		19
↘		↙
	319	
	↓	
	(11 × Y)	

Figure 14 Connection of the rank of amino acids V and Y.
Notes: In this example we connected the position number of amino acids V and Y, and got number (11 × Y) as result.

Biocode 11(3)

Q		E
20		19
5		17
↘		↙
	517	
	↓	
	(11 × Y1)	

Figure 15 Connection of the rank of amino acids Q and E.
Notes: In this example we connected the position number of amino acids Q and E, and got number (11 x Y1) as result.

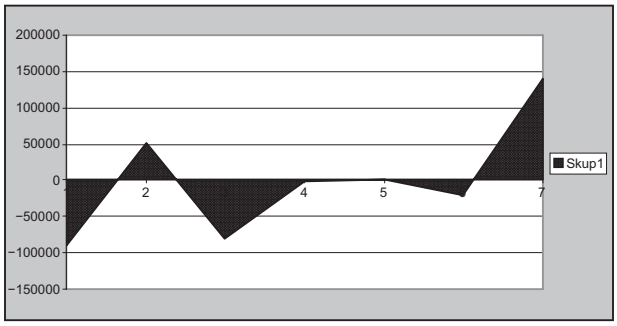

Figure 16 A schematic diagram to show the positive and negative output biocodes of the insulin A chain.

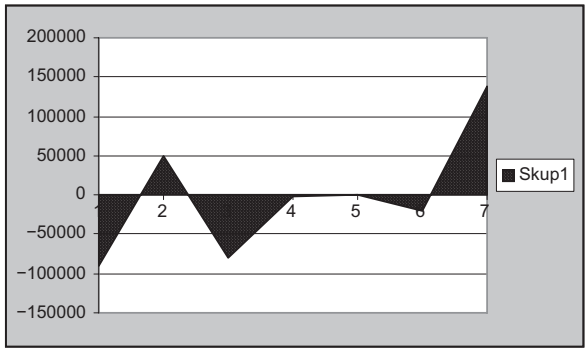

Figure 17 A schematic diagram to show the positive and negative output analog biocodes of the insulin C chain.

From the previous examples it can be seen that this protein has quantitative characteristics. It can be concluded that there is a connection between these quantitative characteristics in the process of transfer of genetic information and the qualitative appearance of the given genetic processes.

Biofrequency

r6 insulin hexamer (d1ai02)

Insulin is composed of amino acids with various numeric values. These numeric values are in an irregular order. For example, the first one has 10 atoms, the second one 22. Their frequency is X. The 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 numeric values of amino acids in proteins. This value can be positive, negative, or zero. These frequencies demonstrate a completely new dimension of protein sequencing. Using these frequencies we can establish which of the amino acids are of primary and which are of secondary significance in the biochemical processes of insulin. A practical example:

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
10	12	−3	0	1	−6	0	3	−3	8	−8	0	8	2	−4	2	−3	−2	7	−10	3

From 0 to 10 = 10; from 10 to 22 = 12; from 22 to 19 = (−) 3; from 19 to 10 = 0, etc.

A schematic representation of the amino acids and their frequency is shown in Figure 17.

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
6	−4	−2	3	0	2	−8	−4	4	6	2	−3	0	−6	9

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
2	−2	−3	−5	−4	9	7	−16	13	−1	2	−7	0	7	−7

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
52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
−7	12	−3	0	1	−6	0	3	−3	8	−8	0	8	2	−4	2	−3	−2	7	−10	3

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
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87
6	–4	–2	3	0	2	–8	–4	4	6	2	–3	0	–6	9

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
88	89	90	91	92	93	94	95	96	97	98	99	100	101	102
2	–2	–3	–5	–4	9	7	–16	13	–1	2	–7	0	7	–7

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
103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123
–7	12	–3	0	1	–6	0	3	–3	8	–8	0	8	2	–4	2	–3	–2	7	–10	3

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
124	125	126	127	128	129	130	131	132	133	134	135	136	137	138
6	–4	–2	3	0	2	–8	–4	4	6	2	–3	0	–6	9

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
139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
2	–2	–3	–5	–4	9	7	–16	13	–1	2	–7	0	7	–7

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
154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174
–7	12	–3	0	1	–6	0	3	–3	8	–8	0	8	2	–4	2	–3	–2	7	–10	3

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
175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
6	–4	–2	3	0	2	–8	–4	4	6	2	–3	0	–6	9

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
190	191	192	193	194	195	196	197	198	199	200	201	202	203	204
2	–2	–3	–5	–4	9	7	–16	13	–1	2	–7	0	7	–7

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
205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
–7	12	–3	0	1	–6	0	3	–3	8	–8	0	8	2	–4	2	–3	–2	7	–10	3

(*Continued*)

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
226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
6	−4	−2	3	0	2	−8	−4	4	6	2	−3	0	−6	9

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
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
2	−2	−3	−5	−4	9	7	−16	13	−1	2	−7	0	7	−7

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
256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276
−7	12	−3	0	1	−6	0	3	−3	8	−8	0	8	2	−4	2	−3	−2	7	−10	3

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
277	278	279	280	281	282	283	284	285	286	287	288	289	290	291
6	−4	−2	3	0	2	−8	−4	4	6	2	−3	0	−6	9

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	
292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	
2	−2	−3	−5	−4	9	7	−16	13	−1	2	−7	0	7	−7	−17

Bio frequency (+)

G	I	Q	T	I	L	Y	L	Y	N	F	Q	I	S	H	L	L	Y	E
10	22	20	17	22	22	24	22	24	17	23	20	22	14	20	22	22	24	19
1	2	5	8	10	13	14	16	19	21	22	25	27	30	31	32	36	37	42
10	12	1	3	8	8	2	2	7	3	6	3	2	4	6	2	9	2	9

Bio frequency (+) = (10 + 12 + 1 + 3 ⋯ + 7) = (+) 718

Bio frequency (zero)

E	C	S	H	E	P	E	C	S	H	E	P	E	C	S	H	E	P
19	14	14	20	19	17	19	14	14	20	19	17	19	14	14	20	19	17
4	7	12	26	34	49	55	58	63	77	85	100	106	109	114	128	136	151
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

E	C	S	H	E	P	E	C	S	H	E	P	E	C	S	H	E	P
19	14	14	20	19	17	19	14	14	20	19	17	19	14	14	20	19	17
157	160	165	179	187	202	208	211	216	230	238	253	259	262	267	281	289	304
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Bio frequency (–)

V	C	S	C	Q	E	N	C	V	N	C	G	V	A	L	V	C
19	14	14	14	20	19	17	14	19	17	14	10	19	13	22	19	14
3	6	9	11	15	17	18	20	23	24	28	29	33	35	38	39	40
–3	–6	–3	–8	–4	–3	–2	–10	–4	–2	–8	–4	–3	–6	–2	–3	–5

Figure 18 Frequency of insulin.
Notes: Insulin is composed of amino acids with various numerical values. These numerical values are in irregular order. For example, the first one has 10 atoms, the second one 22. Their 10, 12, (–)3, 0 etc… … Frequency is the measurement for establishment of intervals of numerical values of amino acids in proteins.

etc.

Bio frequency (–) = [(–3) + (–6) + (–3) ⋯ + (–17)] = (–)718

Results

Bio frequency = [(–) 718 ↔ (+) 718]

Therefore, there is a mathematical balance between the group of amino acids with a positive frequency and those with a negative frequency. Amino acids 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 amino acids 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 emerge.

Within insulin's constituents, 133 amino acids have a positive frequency. These amino acids have 2848 atoms. There are 137 amino acids with a negative frequency. These amino acids have 2174 atoms. Number 133 from the union of acids with a positive frequency is the discrete number that we can use for decoding a matrix with positive and negative frequencies. Here are some examples:

Example 1

$$[(137 + 2174) + (133 + 2848)] = [(133 + 133) + (718 \times Y)]$$

$$Y = 7;$$

$$133 = (7 + 7 + 7 \cdots, + 7)$$

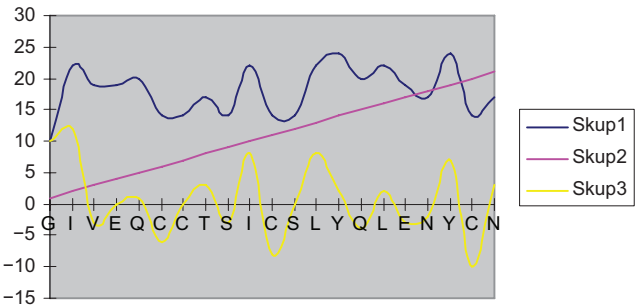

Figure 19 Bio frequency (+), (–), and zero.

Example 2

4712 || 2174

8482 || 2848

$$(4712 + 2174 + 8482 + 2848) = [(133 + 133) + (718 \times Y)]$$

$$Y = 25$$

Levels of frequencies

In the group of amino acids with a positive frequency there are two subunions. One is the subunion with positive frequencies, and the other one is a subunion with negative frequencies. Also, in the union of amino acids with negative frequency, there are subunions with positive and negative frequencies. These groups also have their subunions with positive and negative frequencies. These groups demonstrate a new dimension of the biochemistry of insulin. We expect that this discovery will unlock many of the secrets of protein biochemistry in the future.

Discussion

The results of our research show that the process of sequencing molecules is conditioned and arranged not only according to chemical and biochemical laws, but also according to program, cybernetic, and informational laws. At the first stage of our research we replaced nucleotides from the amino acid code matrix with the numbers of 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 subcellular localization, membrane protein type, secondary protein structure prediction, or indeed any other protein attribute. Since the concept of Chou's pseudo amino acid composition was proposed,[1,2] there have been many attempts to use various digital numbers to represent the 20 native amino acids in order to reflect better the sequence-order effects involved. Some investigators have used complexity measure factor,[3] whereas others have used values derived from cellular automata,[4–7] hydrophobic and/or hydrophilic values,[8–16] Fourier transform techniques,[17,18] or physicochemical distance.[19]

It is going to be possible to use a completely new strategy of research in genetics in the future. However, close investigation of all these relationships, which are the outcomes of periodic laws (more specifically the law of binary coding) is necessary to be able to decode the conformational, stereochemical, and digital structure of proteins.

Disclosure

The author reports no conflict of interest in this research.

References

1. Chou KC. A new branch of proteomics: Prediction of protein cellular attributes. In: Weinrer PW, Lu Q, editors. *Gene Cloning and Expression Technologies*. Westborough, MA: Eaton Publishing; 2002.
2. Chou KC. Prediction of protein cellular attributes using pseudo amino acid composition. *Proteins*. 2001;43:246–255.
3. Xiao X, Shao S, Ding Y, Huang Z, Huang Y, Chou KC. Using complexity measure factor to predict protein subcellular location. *Amino Acids*. 2005;28:57–61.
4. Xiao X, Shao S, Ding Y, Huang Z, Chen X, Chou CK. Using cellular automata to generate image representation for biological sequences. *Amino Acids*. 2005;28:29–35.
5. Xiao X, Shao S, Ding Y, Huang Z, Chen X, Chou KC. An application of gene comparative image for predicting the effect on replication ratio by HBV virus gene missense mutation. *J Theor Biol*. 2005;235:555–565.
6. Xiao X, Shao SH, Huang ZD, Chou KC. Using pseudo amino acid composition to predict protein structural classes: Approached with complexity measure factor. *J Comput Chem*. 2006;27:478–482.
7. Xiao X, Shao SH, Ding YS, Huang ZD, Chou KD. Using cellular automata images and pseudo amino acid composition to predict protein sub-cellular location. *Amino Acids*. 2006;30:49–54.
8. Chou KC. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics*. 2005;21:10–9.
9. Chou KC, Cai YD. Prediction of membrane protein types by incorporating amphipathic effects. *J Chem Inf Model*. 2005;45:407–413.
10. Feng ZP. Prediction of the subcellular location of prokaryotic proteins based on a new representation of the amino acid composition. *Biopolymers*. 2001;58:491–499.
11. Feng ZP. An overview on predicting the subcellular location of a protein. *In Silico Biol*. 2002;2:291–303.
12. Wang MJ, Yang J, Xu ZJ, Chou KC. SLLE for predicting membrane protein types. *J Theor Biol*. 2005;232:7–15.
13. Wang SQ, Yang J, Chou KC. Using stacked generalization to predict membrane protein types based on pseudo amino acid composition. *J Theor Biol*. 2006;242:941–946.
14. Wang M, Yang J, Liu GP, Xu ZJ, Chou KC. Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. *Protein Eng Des Sel*. 2004;17:509–516.
15. Zhang SW, Pan Q, Zhang HC, Shao ZC, Shi JY. Prediction protein homo-oligomer types by pseudo amino acid composition: Approached with an improved feature extraction and naive Bayes feature fusion. *Amino Acids*. 2006;30:461–468.
16. Gao Y, Shao SH, Xiao H, et al. Using pseudo amino acid composition to predict protein subcellular location: approached with Lyapunov index, Bessel function, and Chebyshev filter. *Amino Acids*. 2005;28:373–376.
17. Guo YZ, Li M, Lu M, et al. Classifying G protein-coupled receptors and nuclear receptors based on protein power spectrum from fast Fourier transform. *Amino Acids*. 2006;30:397–402.
18. Liu H, Wang M, Chou KC. Low-frequency Fourier spectrum for predicting membrane protein types. *Biochem Biophys Res Comm*. 2005;336:737–739.
19. Chou KC. Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. *Biochem Biophys Res Comm*. 2000;278:477–483.
20. Chou CK. A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. *Proteins*. 1995;21:319–344.
21. Chou KC, Zhang CT. Predicting protein folding types by distance functions that make allowances for amino acid interactions. *J Biol Chem*. 1994;269:22014–22020.
22. Chou KC, Zhang CT. Review: Prediction of protein structural classes. *Crit Rev Biochem Mol Biol*. 1995;30:275–349.
23. Chou KC, Elrod DW. Protein subcellular location prediction. *Protein Eng*. 1999;12:107–118.
24. Chou KC. Review: Prediction of protein structural classes and subcellular locations. *Curr Protein Pept Sci*. 2001;1:171–208.
25. Chou KC, Elrod DW. Prediction of membrane protein types and subcellular. Locations. *Proteins*. 1999;34:137–153.
26. Chou KC, Elrod DW. Prediction of enzyme family classes. *J Proteome Res*. 2003;2:183–190.
27. Chou KC, Cai YD. Predicting enzyme family class in a hybridization space. *Protein Sci*. 2004;13:2857–2863.
28. Chou KC, Elrod DW. Bioinformatical analysis of G-protein-coupled receptors. *J Proteome Res*. 2002;1:429–433.
29. Chou KC. Prediction of G-protein-coupled receptor classes. *J Proteome Res*. 2005;4:1413–1418.
30. Chou KC, Cai YD. Prediction of protease types in a hybridization space. *Biochem Biophys Res Comm*. 2006;339:1015–1020.
31. Chou KC, Cai YD. Predicting protein-protein interactions from sequences in a hybridization space. *J Proteome Res*. 2006;5:316–322.
32. Chou KC, Cai YD, Zhong WZ. Predicting networking couples for metabolic pathways of arabidopsis. *EXCLI Journal*. 2006;5:55–65.
33. Chou KC, Cai YD. Predicting protein quaternary structure by pseudo amino acid composition. *Proteins*. 2003;53:282–9.
34. Kuri L. The digital language of amino acids. *Amino Acids*. 2007;653–661.
35. Kuri L. The atomic genetic code. *J Comput Biol*. 2009;2:101–16.
36. Kuri L. Mesure complexe des caracteristiques dynamiques de series temporelles. *Journal de la Societe de statistique de Paris*. tome 127, No 2.1986.

Advances and Applications in Bioinformatics and Chemistry Dovepress

Publish your work in this journal

Advances and Applications in Bioinformatics and Chemistry is an international, peer-reviewed open-access journal that publishes articles in the following fields: Computational biomodelling; Bioinformatics; Computational genomics; Molecular modelling; Protein structure modelling and structural genomics; Systems Biology; Computational Biochemistry; Computational Biophysics; Chemoinformatics and Drug Design; In silico ADME/Tox prediction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/advances-and-applications-in-bioinformatics-and-chemistry-journal