Amino Acids (2007) 33: 653–661
DOI 10.1007/s00726-006-0476-x
Printed in The Netherlands

Amino Acids

The digital language of amino acids

L. Kurić

Economic Faculty, Sarajevo, University of Bosnia and Herzegovina, Novi Travnik, Herzegovina

Received August 30, 2006
Accepted November 5, 2006
Published online January 26, 2007;

Summary. The subject of this paper is a digital approach to the investigation of the biochemical basis of genetic processes. The digital mechanism of nucleic acid and protein bio-syntheses, the evolution of biomacromolecules and, especially, the biochemical evolution of genetic language have been analyzed by the application of cybernetic methods, information theory and system theory, respectively. This paper reports the discovery of new methods for developing the new technologies in genetics. It is about the most advanced digital technology which is based on program, cybernetics and informational systems and laws. The results in the practical application of the new technology could be useful in bioinformatics, genetics, biochemistry, medicine and other natural sciences.

Keywords: Amino acids code matrix – Digital genetic code – Mathematical evolution of genetics processes

Introduction

Scientists of new genetics and other natural sciences have been asking the same question repeatedly: Is there a biochemical language for which we can use the theory of systems and cybernetics to describe, in a conditional and specified way, the process of sequencing in bio-molecular genetics, not just with biochemical but with cybernetic and informational lawfulness?

With the goal of finding the answers to some of these questions, we decided to conduct research into whether, in genetics, there exists one unique mathematical array and one mathematical lawfulness that might be valid for all phenomena in this science. Phenomena from this science were therefore designated with specific numeric validity, such as number of atoms in those sequences, number of molecules, number of atom elements, atomic numbers, etc. We tried to discover mathematical lawfulness in sequences of the above-mentioned phenomena, and we were successful in doing that. We discovered that sequences of all molecules in genetics are connected, not just according to their biochemical characteristics, but also according to the program lawfulness, cybernetic lawfulness, and informational lawfulness. In creating the formula for this science we have often encountered problems involving the following: lawfulness of groups of numbers, analog codes and their relation, even-odd correlation of corresponding and complementary sequences, symmetry of sequences, and many others. A radical new vision in science may be developed, differing from those which only consist of ''letters'' of mathematical language in genetics. Here we present our views about the program-cybernetics lawfulness in genetics.

Materials and methods

A digital image where the Amino Acids Code Matrix is represented is in the form of numbers can be created with the help of the new scientific methods. At the first stage of our research we replaced amino acids from the Amino Acid Code Matrix with numbers of the atoms in those amino acids. By this means we acquired a digital image of the Amino Acids Code Matrix. Then we mathematically analyzed those digital images of this code matrix. After making such an analysis, we discovered the existence of digital codes in this matrix, which interconnect all amino acids and other sequences in genetics. Below is a brief introduction to the way we discovered those amino acid digital codes and how those codes interconnect all the amino acids of this matrix.

Results

The results of our research show that the processes of sequencing the molecules are conditioned and arranged

Table 1. Digital genetic code

Number of atoms

Nucleotide position in codon					
first	**Second**				**third**
	U	**C**	**A**	**G**	
U	12,12,12=36; Phe 12,12,13=37; Phe 12,12,15=39; **Leu** 12,12,16=40; **Leu** 152	12,13,12=37; **Ser** 12,13,13=38; **Ser** 12,13,15=40; **Ser** 12,13,16=41; **Ser** 156	12,15,12=39;Tyr 12,15,13=40;Tyr 12,15,15=42; X 12,15,16=43; X 164	12,16,12=40; Cys 12,16,13=41; Cys 12,16,15=43; X 12,16,16=44; Trp 168	**U C A G**
C	13,12,12=37; **Leu** 13,12,13=38; **Leu** 13,12,15=40; **Leu** 13,12,16=41; **Leu** 156	13,13,12=38; Pro 13,13,13=39; Pro 13,13,15=41; Pro 13,13,16=42; Pro 160	13,15,12=40; His 13,15,13=41; His 13,15,15=43; Gln 13,15,16=44; Gln 168	13,16,12=41; ***Arg*** 13,16,13=42; ***Arg*** 13,16,15=44; ***Arg*** 13,16,16=45; ***Arg*** 172	**U C A G**
A	15,12,12=39; Ile 15,12,13=40; Ile 15,12,15=42; Ile 15,12,16=43; Met 164	15,13,12=40; Thr 15,13,13=41; Thr 15,13,15=43; Thr 15,13,16=44; Thr 168	15,15,12=42; Asn 15,15,13=43; Asn 15,15,15=45; Lys 15,15,16=46; Lys 176	15,16,12=43; **Ser** 15,16,13=44; **Ser** 15,16,15=46; ***Arg*** 15,16,16=47; ***Arg*** 180	**U C A G**
G	16,12,12=40; Val 16,12,13=41; Val 16,12,15=43; Val 16,12,16=44; Val 168	16,13,12=41; Ala 16,13,13=42; Ala 16,13,15=44; Ala 16,13,16=45; Ala 172	16,15,12=43; Asp 16,15,13=44; Asp 16,15,15=46; Glu 16,15,16=47; Glu 180	16,16,12=44; Gly 16,16,13=45; Gly 16,16,15=47; Gly 16,16,16=48; Gly 184	**U C A G**
	640	**656**	**688**	**704**	**2688**

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 (see Table 1).

$(152 + 168) = (164 + 168); (156 + 172) = (160 + 168);$
$(164 + 180) = (168 + 176); (168 + 184) = (172 + 180);$
$(640 + 704) = (656 + 688);$

Number of atoms in nucleotides AUGC = 2688;

Connection

$36, 37, 39, 40 \rightarrow 36373940;$
$37, 38, 40, 41 \rightarrow 37384041;$ etc.

Example 1

$(37384041 - 36373940) = (38394142 - 37384041) = 1010101;$
$(40414344 - 39404243) = (41424445 - 40414344) = 1010101;$
$(39404243 - 37384041) = (40414344 - 38394142) = 2020201;$ etc.

Example 2

$(41403837 - 40393736) = (42413938 - 41403837) = 1010101;$
$(44434140 - 43424039) = (45444241 - 44434140) = 1010101;$
$(43424039 - 41403837) = (44434140 - 42413938) = 2020201;$ etc.

Example 3

$152, 156, 164, 168 \rightarrow 152156164168;$
$156, 160, 168, 172 \rightarrow 156160168172;$ etc.

$(156160168172 - 152156164168) = (168172180184 - 164168176180) = 4004004004;$

Code 336

$2688 = (\mathbf{336} + \mathbf{336} + \mathbf{336} + \mathbf{336} + \mathbf{336} + \mathbf{336} + \mathbf{336} + \mathbf{336});$

In those examples, the mathematical balance in the distribution of atoms is achieved.

Formula for decoding the digital genetic code matrix

$(\mathrm{D}1, 2, 3, \mathrm{n}, \mathrm{C}1, 2, 3, \mathrm{n}, \mathrm{B}1, 2, 3, \mathrm{n}, \mathrm{A}1, 2, 3, \mathrm{n} - \mathrm{A}1, 2, 3, \mathrm{n}, \mathrm{B}1, 2, 3, \mathrm{n}, \mathrm{C}1, 2, 3, \mathrm{n}, \mathrm{D}1, 2, 3, \mathrm{n}) = \mathrm{Y};$

ABCD = Groups of atoms in codons UCAG

Example 1

$\mathrm{UUU} = 36$ atoms; $\mathrm{UUC} = 37$ atoms;
$\mathrm{UUA} = 39$ atoms; $\mathrm{UUG} = 40$ atoms;
$\mathrm{A1} = 36$; $\mathrm{B1} = 37$; $\mathrm{C1} = 39$; $\mathrm{D1} = 40$;
$(40, 39, 37, 36 - 36, 37, 39, 40) \rightarrow \mathrm{Y}$;
$\mathrm{Y} = 40\ 197\ 96$;

Example 2

$\mathrm{UCU} = 37$ atoms; $\mathrm{UCC} = 38$ atoms;
$\mathrm{UCA} = 40$ atoms; $\mathrm{UCG} = 41$ atoms;
$\mathrm{A2} = 37$; $\mathrm{B2} = 38$; $\mathrm{C2} = 40$; $\mathrm{D2} = 41$;
$(41, 40, 38, 37 - 37, 38, 40, 41) \rightarrow \mathrm{Y}$;
$\mathrm{Y} = 40\ 197\ 96$;

Example 3

$\mathrm{UAU} = 39$ atoms; $\mathrm{UAC} = 40$ atoms;
$\mathrm{UAA} = 42$ atoms; $\mathrm{UAG} = 43$ atoms;
$\mathrm{A3} = 39$; $\mathrm{B3} = 40$; $\mathrm{C3} = 42$; $\mathrm{D3} = 43$;
$(43, 42, 40, 39 - 39, 40, 42, 43) \rightarrow \mathrm{Y}$;
$\mathrm{Y} = 40\ 197\ 96$; etc.

$$\mathbf{40\ 19796} = [\mathbf{7} + (\mathbf{19} + \mathbf{19} + \mathbf{19} + \cdots + \mathbf{19}) + \mathbf{7}]$$

Codes 7 and 19

In the groups of all natural numbers from X to Y there are two codes which interconnect all those numbers. Those are codes A and B.

Formula of codes A and B

$$\{[\mathrm{SA}(\mathrm{R}1, 2, 3, \mathrm{n}) \times \mathrm{B}] - [\mathrm{SB}(\mathrm{R}1, 2, 3, \mathrm{n}) \times \mathrm{A}] + (\mathrm{AB})\} = \mathrm{ABA};$$

$\mathbf{S}\mathrm{A}$, $\mathbf{S}\mathrm{B}$ = Groups of AB numbers in group of all natural numbers from X to Y

$\mathrm{R}1, 2, 3, \mathrm{n}$ = Natural numbers from X to Y;

$\mathbf{A = 7;\ B = 19;}$

Example

$\mathrm{R} = 35$;

$\{[\mathrm{S7}(35) \times 19] - [\mathrm{S19}(35) \times 7] + (7 \times 19)\} = (7 \times 19 \times 7)$;

$\mathrm{S7}(35) = (29 + 30 + 31 + 32 + 33 + 34 + 35) = 224$;

$\mathrm{S19}(35) = (17 + 18 + 19 + 20 + 21 + 22 + 23 + 24 + 25 + 26 + 27 + 28 + 29 + 30 + 31 + 32 + 33 + 34 + 35) = 494$;

$\{[(224 \times 19) - (494 \times 7)] + (7 \times 19)\} = (7 \times 19 \times 7)$;

Determinants in digital genetic code

$\mathrm{DET}\ 152, 156, 164, 168 = -48$;
$\mathrm{DET}\ 156, 160, 168, 172 = -48$;
$\mathrm{DET}\ 164, 168, 176, 180 = -48$;
$\mathrm{DET}\ 168, 172, 180, 184 = -48$;

$2688 = (\mathbf{48} + \mathbf{48} + \mathbf{48} + \cdots + \mathbf{48})$; Code 336
$= (\mathbf{48} + \mathbf{48} + \mathbf{48} + \mathbf{48} + \mathbf{48} + \mathbf{48} + \mathbf{48})$; etc.

Groups of atoms

152	156	164	168	640
156	160	168	172	656
164	168	176	180	688
168	172	180	184	704
640	656	688	704	2688

Column = Rows;
Column $= (640 + 656 + 688 + 704)$;
Rows $= (640 + 656 + 688 + 704)$;

Diagonals

$\mathrm{D1} = (152 + 160 + 176 + 184) = \mathbf{336}$;
$\mathrm{D2} = (168 + 168 + 168 + 168) = \mathbf{336}$;
Other numbers $= (156 + 164 + 156 + 172 + 164 + 180 + 172 + 180) = (\mathbf{336} \times \mathrm{Y})$;

Other relations

152	156	164	168
156	160	168	172
164	168	176	180
168	172	180	184

152	156	164	168
156	160	168	172
164	168	176	180
168	172	180	184

etc.

In fact, we discovered that the *mathematical balance* in the distribution of atoms in the genetic code is achieved.

Correlation of ATGC, UCAG and amino acids

Amino Acid = 384 atoms; Ac = 384;
Triplets ATGC = 2832 atoms; Tc = 2832;
Triplets UCAG = 2688 atoms; Nc = 2688;

Example 1

$\mathrm{Nc} = (\mathrm{Ac}\times\mathrm{Y});\ \mathrm{Y} = 7;$

Nc = (384 × 7) = 2 688;

Example 2

$\mathrm{Tc}, \mathrm{Nc} \rightarrow \mathrm{Ac};$

$2832, 2688 \rightarrow 2832\ 2688;$

$2832\ 2688 = (384 + 384 + 384 + \cdots + 384);$

Groups of atoms in UCAG

Number of atoms

152	156 156	160	164 164	168 168 168 168	172 172	176	180 180	184

Correlation

$(152 + 184) = \mathbf{336};$
$(156 + 180) = \mathbf{336};$
$(160 + 176) = \mathbf{336};$
$(164 + 172) = \mathbf{336};$
$(168 + 168) = \mathbf{336};$

$\mathrm{Nc} = (336 + 336 + 336 + 336 + 336 + 336 + 336 + 336)$

Digital quadrant UCAG

164	172		172	164	672
156	180		156	180	672
168	168		168	168	672
184	152		176	160	672
672	672		672	672	

Groups of atoms

$(164 + 172 + 156 + 180) = \mathbf{672} = (336 + 336);$
$(172 + 164 + 156 + 180) = \mathbf{672} = (336 + 336);$
$(168 + 168 + 184 + 152) = \mathbf{672} = (336 + 336);$
$(168 + 168 + 176 + 160) = \mathbf{672} = (336 + 336);$

Diagonals:

$(164 + 180 + 168 + 160) = \mathbf{672} = (336 + 336);$
$(164 + 156 + 168 + 184) = \mathbf{672} = (336 + 336);$

Inner numbers in quadrant:

$(180 + 156 + 168 + 168) = \mathbf{672} = (336 + 336);$

Outer numbers in quadrant:

$(164 + 172 + 172 + 164 + 156 + 180 + 168 + 168 + 184 + 152 + 176 + 160) = (\mathbf{672}\times\mathrm{Y}) = (336\times\mathrm{Y});$ etc.

Mathematical position of the nucleotides in codon

Number of atoms					
	12,15,12=39; Tyr 12,15,13=40; Tyr 12,15,15=42; X 12,15,16=43; X 164	13,16,12=41; ***Arg*** 13,16,13=42; ***Arg*** 13,16,15=44; ***Arg*** 13,16,16=45; ***Arg*** 172	16,13,12=41; Ala 16,13,13=42; Ala 16,13,15=44; Ala 16,13,16=45; Ala 172	15,12,12=39; Ile 15,12,13=40; Ile 15,12,15=42; Ile 15,12,16=43; Met 164	672
	12,13,12=37; **Ser** 12,13,13=38; **Ser** 12,13,15=40; **Ser** 12,13,16:41; **Ser** 156	15,16,12=43; **Ser** 15,16,13=44; **Ser** 15,16,15=46; ***Arg*** 15,16,16=47; ***Arg*** 180	13,12,12=37; **Leu** 13,12,13=38; **Leu** 13,12,15=40; **Leu** 13,12,16=41; **Leu** 156	16,15,12=43; Asp 16,15,13=44; Asp 16,15,15=46; Glu 16,15,16=47; Glu 180	672
	12,16,12=40; Cys 12,16,13=41; Cys 12,16,15=43; X 12,16,16=44; Trp 168	13,15,12=40; His 13,15,13=41; His 13,15,15=43; Gln 13,15,16=44: Gln 168	15,13,12=40; Thr 15,13,13=41; Thr 15,13,15=43; Thr 15,13,16=44; Thr 168	16,12,12=40; Val 16,12,13=41; Val 16,12,15=43; Val 16,12,16=44; Val 168	672
	16,16,12=44; Gly 16,16,13=45; Gly 16,16,15=47; Gly 16,16,16=48; Gly 184	12,12,12=36; Phe 12,12,13=37; Phe 12,12,15=39; **Leu** 12,12,16=40; **Leu** 152	15,15,12=42; Asn 15,15,13=43; Asn 15,15,15=45; Lys 15,15,16=46; Lys 176	13,13,12=38; Pro 13,13,13=39; Pro 13,13,15=41; Pro 13,13,16=42; Pro 160	672
	672	672	672	672	

$672 = (336 + 336)$

Mathematical position of the nucleotides in codon

Connection of numbers of atoms				
12,15,12=39; Tyr 12,15,13=40; Tyr 12,15,15=42; X 12,15,16=43; X 39404243	13,16,12=41; ***Arg*** 13,16,13=42; ***Arg*** 13,16,15=44; ***Arg*** 13,16,16=45; ***Arg*** 41424445	16,13,12=41; Ala 16,13,13=42; Ala 16,13,15=44; Ala 16,13,16=45; Ala 41424445	15,12,12=39; Ile 15,12,13=40; Ile 15,12,15=42; Ile 15,12,16=43; Met 39404243	161 657 376
12,13,12=37; **Ser** 12,13,13=38; **Ser** 12,13,15=40; **Ser** 12,13,16:41; **Ser** 37384041	15,16,12=43; **Ser** 15,16,13=44; **Ser** 15,16,15=46; ***Arg*** 15,16,16=47; ***Arg*** 43444647	13,12,12=37; **Leu** 13,12,13=38; **Leu** 13,12,15=40; **Leu** 13,12,16=41; **Leu** 37384041	16,15,12=43; Asp 16,15,13=44; Asp 16,15,15=46; Glu 16,15,16=47; Glu 43444647	161 657 376
12,16,12=40; Cys 12,16,13=41; Cys 12,16,15=43; X 12,16,16=44; Trp 40414344	13,15,12=40; His 13,15,13=41; His 13,15,15=43; Gln 13,15,16=44: Gln 40414344	15,13,12=40; Thr 15,13,13=41; Thr 15,13,15=43; Thr 15,13,16=44; Thr 40414344	16,12,12=40; Val 16,12,13=41; Val 16,12,15=43; Val 16,12,16=44; Val 40414344	161 657 376
16,16,12=44; Gly 16,16,13=45; Gly 16,16,15=47; Gly 16,16,16=48; Gly 44454748	12,12,12=36; Phe 12,12,13=37; Phe 12,12,15=39; **Leu** 12,12,16=40; **Leu** 36373940	15,15,12=42; Asn 15,15,13=43; Asn 15,15,15=45; Lys 15,15,16=46; Lys 42434546	13,13,12=38; Pro 13,13,13=39; Pro 13,13,15=41; Pro 13,13,16=42; Pro 38394142	161 657 376
161 657 376	161 657 376	161 657 376	161 657 376	

39,40,42,43 → 39404243; 41,42,44,45 → 41424445; etc.
Diagonal 1 = (39404243 + 43444647 + 40414344 + 38394142) = 161 657 376;
Diagonal 2 = (39404243 + 37384041 + 40414344 + 44454748) = 161 657 374; etc.

Digital image of UCAG and amino acids

Groups of atoms in UCAG

152	156	160	164	168	172	176	180	184
	156		164	168	172		180	
				168				
				168				

Groups of the amino acids (1)

Ala	Cys	Asp	Val	Gln	Leu	Phe	Lys	Arg
	Ser		Glu	His	Jle		Tyr	
				Met				

Groups of the amino acids (2)

Gly	Ala	Cys	Asp	Asn	Val	Gln	Leu	Phe	Lys	Arg	Trp
		Ser		Thr	Glu	His	Jle		Tyr		
				Pro		Met					

It is obvious that the digital matrix of amino acid code evolved from the digital matrix of the nucleotide code.

Code 1512

Groups of atoms in amino acids

Gly	Ala	Cys	Asp	Asn	Val	Gln	Leu	Phe	Lys	Arg	Trp
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
10	13	14	16	17	19	20	22	23	24	26	27
		↓		↓	↓	↓	↓		↓		
		Ser		Thr	Glu	His	Jle		Tyr		
		↓		↓	↓	↓	↓		↓		
		14		17	19	20	22		24		
				↓		↓					
				Pro		Met					
				↓		↓					
				17		20					

In this table, in the first row, there are 12 numbers: 10, 13, 14, 16, 17, 19, 20, 22, 23, 24, 26 and 27.

In the group of numbers from 1 to 27 there are the following numbers:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27.

Correlation:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27.

Numbers highlighted are in the tables of the amino code. Those numbers are:

$10, 13, 14, 16, 17, 19, 20, 22, 23, 24, 26$ and 27.

There are in total **12** of those numbers.

Unmarked numbers are not in the tables of genetic code. Those numbers are:

$1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 15, 18, 21$ and 25.

There are in total 15 of those numbers.

Correlation of marked and unmarked numbers

15 and 15 $\rightarrow$ 1512;

First row of nucleotide code matrix

152	156	160	164	168	172	176	180	184
	156		164	168	172		180	
				168				
				168				

In the first row of the nucleotide matrix code there are these numbers:

$$(152 + 156 + 160 + 164 + 168 + 172 + 176 + 180 + 184) = 1512;$$

$$152, 156, 160, 164, 168, 172, 176, 180, 184 \rightarrow 1512;$$

$$152156160164168172176180184 = (1512 \times \mathrm{Y});$$

$$\mathrm{Y} = 100,632,381,060,957,785,830,807.$$

Therefore, the numbers of atoms in the nucleotide matrix code are in correlation with the number of atoms in the amino acid code.

Other rows of nucleotide code

In the second, third and fourth rows of the nucleotide code there are following numbers:

$$156, 164, 168, 172, 180, 168, 168 \rightarrow (156 + 164 + 168 + 172 + 180 + 168 + 168) + (1512 \times \mathrm{Y});$$

$$\mathrm{Y} = 103,283,180,008,055,666.$$

$$[(103,283,180,008,055,666. \times 1512) + 1176] = 156, 164, 168, 172, 180, 168, 168;$$

Therefore, the second, third and fourth rows of the nucleotide code are filled with the help of code 1512.

Mathematical correlation of groups of codons and amino acids

Example 1

Number of atoms in codons

GUU	40		GCU	41
GGU	44		GAU	43
	84			84

Number of atoms in amino acids

GUU	Val=19		GCU	Ala=13
GGU	Gly=10		GAU	Asp=16
	29			29

Example 2

Number of atoms in codons

GUC	41		GCC	42
GGC	45		GAC	44
	86			86

Number of atoms in amino acids

GUC	Val=19		GCC	Ala=13
GGC	Gly=10		GAC	Asp=16
	29			29

Example 3

Number of atoms in codons

GUA	43		GCA	44
GGA	47		GAA	46
GCG	45		GUG	44
GAG	47		GGG	48
	182			182

$$(43 + 48) = (47 + 44) = (45 + 46) + (47 + 44)$$

Number of atoms in amino acids

GUA	Val=19		GCA	Ala=13
GGA	Gly=10		GAA	Glu=19
GCG	Ala=13		GUG	Val=19
GAG	Glu=19		GGG	Gly=10
	61			61

Example 4

Number of atoms in codons

CUU	37		CCU	38
CGU	41		CAU	40
CCC	39		CUC	38
CAC	41		CGC	42
	158			158

$$(37 + 42) = (41 + 38) = 39 + 40) = (41 + 38);$$

Number of atoms in amino acids

CUU	Leu=22		CCU	Pro=17
CGU	Arg=26		CAU	His=20
CCC	Pro=17		CUC	Leu=22
CAC	His=20		CGC	Arg=26
	85			85

Example 5

Number of atoms in codons

AUU	39		ACU	40
AGU	43		AAU	42
ACC	41		AUC	40
AAC	43		AGC	44
	166			166

$(39+44)=(43+40)=(41+42)=(43+40)$

Number of atoms in amino acids

AUU	Jle=22		ACU	Thr=17
AGU	Ser=14		AAU	Asn=17
ACC	Thr=17		AUC	Jle=22
AAC	Asn=17		AGC	Ser=14
	70			70

Example 6

Number of atoms in codons

UUU	36		UCU	37
UGU	40		UAU	39
UCC	38		UUC	37
UAC	40		UGC	41
	154			154

$(36+41)=(40+37)=(38+39)+(40+37)$

Number of atoms in amino acids

UUU	Phe=23		UCU	Ser=14
UGU	Cys=14		UAU	Tyr=24
UCC	Ser=14		UUC	Phe=23
UAC	Tyr=24		UGC	Cys=14
	75			75

Example 7

Number of atoms in codons

CUA	40		CCA	41
CGA	44		CAA	43
CCG	42		CUG	41
CAG	44		CGG	45
	170			170

$(40+45)=(44+41)=(42+43)=(44+41)$

Number of atoms in amino acids

CUA	Leu=22		CCA	Pro=17
CGA	Arg=26		CAA	Gln=20
CCG	Pro=17		CUG	Leu=22
CAG	Gln=20		CGG	Arg=26
	85			85

etc.

$\mathrm{DET}\ 164,172,172,164 = -2688;\ \mathrm{DET}\ 184, 152,176,160 = +2688;$

$\mathrm{DET}\ 156,180,156,180 = 0;\ \mathrm{DET}\ 168,168, 168,168 = 0;$

$(672+672+672+672)=2688;$

Table 2. Determinants in the digital genetic code

Nucleotide mathematical position in codon					
	12,15,12=39; Tyr 12,15,13=40; Tyr 12,15,15=42; X 12,15,16=43; X 164	13,16,12=41; ***Arg*** 13,16,13=42; ***Arg*** 13,16,15=44; ***Arg*** 13,16,16=45; ***Arg*** 172	16,13,12=41; Ala 16,13,13=42; Ala 16,13,15=44; Ala 16,13,16=45; Ala 172	15,12,12=39; Ile 15,12,13=40; Ile 15,12,15=42; Ile 15,12,16=43; Met 164	**-2688**
	12,13,12=37; **Ser** 12,13,13=38; **Ser** 12,13,15=40; **Ser** 12,13,16=41; **Ser** 156	15,16,12=43; **Ser** 15,16,13=44; **Ser** 15,16,15=46; ***Arg*** 15,16,16=47; ***Arg*** 180	13,12,12=37; **Leu** 13,12,13=38; **Leu** 13,12,15=40; **Leu** 13,12,16=41; **Leu** 156	16,15,12=43; Asp 16,15,13=44; Asp 16,15,15=46; Glu 16,15,16=47; Glu 180	**0**
	12,16,12=40; Cys 12,16,13=41; Cys 12,16,15=43; X 12,16,16=44; Trp 168	13,15,12=40; His 13,15,13=41; His 13,15,15=43; Gln 13,15,16=44: Gln 168	15,13,12=40; Thr 15,13,13=41; Thr 15,13,15=43; Thr 15,13,16=44; Thr 168	16,12,12=40; Val 16,12,13=41; Val 16,12,15=43; Val 16,12,16=44; Val 168	**0**
	16,16,12=44; Gly 16,16,13=45; Gly 16,16,15=47; Gly 16,16,16=48; Gly 184	12,12,12=36; Phe 12,12,13=37; Phe 12,12,15=39; **Leu** 12,12,16=40; **Leu** 152	15,15,12=42; Asn 15,15,13=43; Asn 15,15,15=45; Lys 15,15,16=46; Lys 176	13,13,12=38; Pro 13,13,13=39; Pro 13,13,15=41; Pro 13,13,16=42; Pro 160	**+2688**
	672	**672**	**672**	**672**	

Table 3. Translation table for an antecedent code in codon octads by D. Grafstein

Codon	Number of atoms	L-amino acid	Dimer D-amino acid	Codon	Number of atoms	D-amino acid	Dimer L-amino acid
CAG	44	Gln	Val	GUC	41	Glu	Val
CAU	40	His	Val	GUA	43	His	Val
GUG	44	Val	His	CAC	41	Val	His
GUU	40	Val	Glu	CAA	43	Val	Glu
	168				168		
GCG	45	Ala	Arg	CGC	42	Ala	Arg
GCU	41	Ala	Arg	CGA	44	Ala	Arg
CGG	45	Arg	Ala	GCC	42	Arg	Ala
CGU	41	Arg	Ala	GCA	44	Arg	Ala
	172				172		
AGG	47	Arg	Ser	UCC	38	Arg	Ser
AGU	43	Ser	Ser	UCA	40	Ser	Ser
UCG	41	Ser	Ser	AGC	44	Ser	Ser
UCU	37	Ser	Arg	AGA	46	Ser	Arg
	168				168		
CUG	41	Leu	ASP	GAC	44	Leu	Asp
CUU	37	Leu	Glu	GAA	46	Leu	Glu
GAG	47	Glu	Leu	CUC	38	Glu	Leu
GAU	43	Asp	Leu	CUA	40	Asp	Leu
	168				168		
AAG	46	Lys	Phe	UUC	37	Lys	Phe
AAU	42	Asn	Leu	UUA	39	Asp	Leu
UUG	40	Leu	Asp	AAC	43	Leu	Asp
UUU	36	Phe	Lys	AAA	45	Phe	Lys
	164				164		
UGG	44	Trp	Thr	ACC	41	Trp	Thr
UGU	40	Cys	Thr	ACA	43	Cys	Thr
ACG	44	Thr	Cys	UGC	41	Thr	Cys
ACU	40	Thr	X	UGA	43	Thr	X
	168				168		
AUG	43	Met	Tyr	UAC	40	Met	Tyr
AUU	39	Jle	X	UAA	42	Jle	X
UAG	43	X	Jle	AUC	40	X	Jle
UAU	39	Tyr	Jle	AUA	42	Tyr	Jle
	164				164		
GGG	48	Gly	Pro	CCC	39	Gly	Pro
GGU	44	Gly	Pro	CCA	41	Gly	Pro
CCG	42	Pro	Gly	GGC	45	Pro	Gly
CCU	38	Pro	Gly	GGA	47	Pro	Gly
	172				172		
Total	1344				1344		

There is a mathematical balance within all of the phenomena in the digital genetic code matrix (see Table 2).

$$1344 \rightarrow \text{Code } 336;\ 1344 = (336 + 336 + 336 + 336);$$

There is a mathematical balance in the translation table for an antecedent code in codon octads by D. Grafstein (see Table 3).

Conclusion

It is rewarding to translate the biochemical language of amino acids into a digital language because it may be very useful for developing new methods for predicting protein sub-cellular localization, membrane protein type, protein structure secondary prediction or any other protein attributes.

This is because ever since the concept of Chou's pseudo amino acid composition was proposed (Chou, 2001, 2002), 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 (Xiao et al., 2005a), some used the values derived from the

cellular automata (Xiao et al., 2005, 2006a, b), some used hydrophobic and/or hydrophilic values (Chou, 2005; Chou and Cai, 2005; Feng, 2001, 2002; Gao et al., 2005; Wang et al., 2004, 2005, 2006a, b; Zhang et al., 2006), some were through Fourier transform (Gao et al., 2006; Liu et al., 2005), and some used the physicochemical distance (Chou et al., 2000).

Now it is going to be possible to use a completely new strategy of research in genetics. However, observation of all these relations which are the outcome of the periodic law (actually, of the law of binary coding) is necessary, because it can be of great importance for decoding conformational forms and the stereo-chemical and digital structure of proteins.

References

Chou KC (1995) A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. Proteins Struct Funct Gen 21: 319–344

Chou KC (2000a) Review: Prediction of protein structural classes and subcellular locations. Curr Prot Peptide Sci 1: 171–208

Chou KC (2000b) Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. Biochem Biophys Res Commun 278: 477–483

Chou KC (2001) Prediction of protein cellular attributes using pseudo amino acid composition, Proteins Struct Funct Genet 43: 246–255

Chou KC (2002) In: Weinrer PW, Lu Q (eds) Gene cloning and expression technologies. Eaton Publishing, Westborough, pp 57–70

Chou KC (2005a) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics 21: 10–19

Chou KC (2005b) Prediction of G-protein-coupled receptor classes. J Proteome Res 4: 1413–1418

Chou KC, Cai YD (2003) Predicting protein quaternary structure by pseudo amino acid composition. Proteins Struct Funct Genet 53: 282–289

Chou KC, Cai YD (2004) Predicting enzyme family class in a hybridization space. Protein Sci 13: 2857–2863

Chou KC, Cai YD (2005) Prediction of membrane protein types by incorporating amphipathic effects. J Chem Inform Model 45: 5407–5413

Chou KC, Cai YD (2006a) Prediction of protease types in a hybridization space. Biochem Biophys Res Comm 339: 1015–1020

Chou KC, Cai YD (2006b) Predicting protein-protein interactions from sequences in a hybridization space. J Proteome Res 5: 316–322

Chou KC, Cai YD, Zhong WZ (2006) Predicting networking couples for metabolic pathways of Arabidopsis. EXCLI J 5: 55–65

Chou KC, Elrod DW (1999a) Protein subcellular location prediction. Protein Eng 12: 107–118

Chou KC, Elrod DW (1999b) Prediction of membrane protein types and subcellular locations. Proteins Struct Funct Genet 34: 137–153

Chou KC, Elrod DW (2002) Bioinformatical analysis of G-protein-coupled receptors. J Proteome Res 1: 429–433

Chou KC, Elrod DW (2003) Prediction of enzyme family classes. J Proteome Res 2: 183–190

Chou KC, Zhang CT (1994) Predicting protein folding types by distance functions that make allowances for amino acid interactions. J Biol Chem 269: 22014–22020

Chou KC, Zhang CT (1995) Review: Prediction of protein structural classes. Crit Rev Biochem Mol Biol 30: 275–349

Dawn JB, Jacques RF, Arthur ML, Mona S (2002) Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. Mol Biol Evol 19: 1645–1655

Feng ZP (2001) Prediction of the subcellular location of prokaryotic proteins based on a new representation of the amino acid composition. Biopolymers 58: 491–499

Feng ZP (2002) An overview on predicting the subcellular location of a protein. In Silico Biol 2: 291–303

Gao Y, Shao SH, Xiao X, Ding YS, Huang YS, Huang ZD, Chou KC (2005) Using pseudo amino acid composition to predict protein subcellular location: approached with Lyapunov index, Bessel function, and Chebyshev filter. Amino Acids 28: 373–376

Guo YZ, Li M, Lu M, Wen Z, Wang K, Li G, Wu J (2006) Classifying G protein-coupled receptors and nuclear receptors based on protein power spectrum from fast fourier transform. Amino Acids 30: 397–402

Liu H, Wang M, Chou KC (2005) Low-frequency Fourier spectrum for predicting membrane protein types. Biochem Biophys Res Commun 336: 737–739

Rakočević MM (1998) Geni molekuli jezik, drugo izdanje. Naučna knjiga Beograd: 147–216

Wang M, Yang J, Liu GP, Xu ZJ, Chou KC (2004) Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. Protein Eng Des Select 17: 509–516

Wang M, Yang J, Xu ZJ, Chou KC (2005) SLLE for predicting membrane protein types. J Theor Biol 232: 7–15

Wang SQ, Yang J, Chou KC (2006) Using stacked generalization to predict membrane protein types based on pseudo amino acid composition. J Theor Biol (DOI: 10.1016/j.jtbi.2006.1005.1006)

Xiao X, Shao S, Ding Y, Huang Z, Chen X, Chou KC (2005) An application of gene comparative image for predicting the effect on replication ratio by HBV virus gene missense mutation. J Theor Biol 235: 555–565

Xiao X, Shao S, Ding Y, Huang Z, Huang Y, Chou KC (2005a) Using complexity measure factor to predict protein subcellular location. Amino Acids 28: 57–61

Xiao X, Shao S, Ding Y, Huang Z, Chen X, Chou KC (2005b) Using cellular automata to generate Image representation for biological sequences. Amino Acids 28: 29–35

Xiao X, Shao SH, Huang ZD, Chou KC (2006a) Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor. J Comput Chem 27: 478–482

Xiao X, Shao SH, Ding YS, Huang ZD, Chou KC (2006b) Using cellular automata images and pseudo amino acid composition to predict protein sub-cellular location. Amino Acids 30: 49–54

Zhang SW, Pan Q, Zhang HC, Shao ZC, Shi JY (2006) Prediction protein homo-oligomer types by pseudo amino acid composition: Approached with an improved feature extraction and naive Bayes feature fusion. Amino Acids 30: 461–468

Authors' address: Latvo Kurić, Economic Faculty, Sarajevo, University of Bosnia and Herzegovina, Kalinska 7, 72290 Novi Travnik, Herzegovina, Fax: 4387 30 513535, E-mail: lutvokuric@yahoo.com