Vol 18 No 1, January 2009
1674-1056/2009/18(01)/0363-07

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

The structural analysis of protein sequences based on the quasi-amino acids code*

Zhu Ping(朱 平)[a)†], Tang Xu-Qing(唐旭清)[a)b)], and Xu Zhen-Yuan(徐振源)[a)]

[a)] *School of Science, Jiangnan University, Wuxi* 214122, *China*
[b)] *Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei* 230039, *China*

(Received 18 March 2008; revised manuscript received 12 April 2008)

Proteomics is the study of proteins and their interactions in a cell. With the successful completion of the Human Genome Project, it comes the postgenome era when the proteomics technology is emerging. This paper studies protein molecule from the algebraic point of view. The algebraic system $(\Sigma, +, *)$ is introduced, where Σ is the set of 64 codons. According to the characteristics of $(\Sigma, +, *)$, a novel quasi-amino acids code classification method is introduced and the corresponding algebraic operation table over the set ZU of the 16 kinds of quasi-amino acids is established. The internal relation is revealed about quasi-amino acids. The results show that there exist some very close correlations between the properties of the quasi-amino acids and the codon. All these correlation relationships may play an important part in establishing the logic relationship between codons and the quasi-amino acids during the course of life origination. According to Ma F *et al* (2003 *J. Anhui Agricultural University* **30** 439), the corresponding relation and the excellent properties about amino acids code are very difficult to observe. The present paper shows that $(ZU, \oplus, \otimes)$ is a field. Furthermore, the operational results display that the codon *tga* has different property from other stop codons. In fact, in the mitochondrion from human and ox genomic codon, *tga* is just tryptophane, is not the stop codon like in other genetic code, it is the case of the Chen W C *et al* (2002 *Acta Biophysica Sinica* **18**(1) 87). The present theory avoids some inexplicable events of the 20 kinds of amino acids code, in other words it solves the problem of 'the 64 codon assignments of mRNA to amino acids is probably completely wrong' proposed by Yang (2006 *Progress in Modern Biomedicine* **6** 3).

Keywords: algebraic operation, quasi-amino acids code, protein sequences, structural analysis
PACC: 8700, 8715, 0210

1. Introduction

The origin and evolution of the genetic code, which is one of nature's best creations, are the hot topics for life science. So far, many authors have explored and made a variety of hypotheses on the origin of the genetic code from different perspectives on this issue. The main forms are the genomic genetic codes and the mitochondrial genetic codes.[1] But there exists a mismatch between some biological phenomena and these genetic code systems. How to solve this problem is a world focus.

For all life forms, codon usage is non-random and associated with various factors, such as gene expression level, gene length and secondary protein structures. Most amino acids in all species bear a highly significant association with gene functions. Codon usage at the level of individual amino acids is closely coordinated with the gene function in general. It is shown that the corresponding relation and the excellent properties about amino acids code is very difficult to observe.[2] This explains the reason why different quantitative codon values may be expressed by codon in genes. The codon can be represented in a four-column table; hydrophobic and hydrophilic amino acids can be found in different columns.

Sánchez *et al*[3−6] established the order in the Boolean lattice of the four DNA bases. They defined a new vector space and a new Lie algebra for a better understanding of the logic underlying the genetic code. It is pointed out that the algebraic relationship, codon assignments and physicochemical properties of amino acids are not connected at random. Starting from the four DNA bases order in the Boolean lattice, Sánchez *et al* proposed a novel Lie Algebra of the genetic code.[3,6] The main partitions of the genetic code

*Project supported in part by the International Technology Collaboration Research Program of China (Grant No 2007DFA706700).
†Corresponding author. E-mail: zptl2002@yahoo.com.cn; zhuping@jiangnan.edu.cn

table were obtained as equivalent classes of quotient spaces of the genetic code vector space over the Galois field of the four DNA bases. In Ref.[4], the group obtained that the set of codon is isomorphic to the group $(Z_{64}, +)$ of the integer module 64. In Ref.[5], Sánchez *et al* proposed a genetic code Boolean structure derived from hydrogen bond numbers and chemical types of bases, purines and pyrimidines.

The protein sequence is the important object in the research of life sciences. The protein function is decided by its structure. It is easy to determine the amino acid sequence of proteins through the experimental method, but it is rather difficult to observe its spatial structure actually. A number of mathematical models have been proposed for understanding the origin and evolution of the genetic code. In this dissertation, the protein molecules is studied from the perspective of algebra. According the relative physical and chemical properties, including the chemical type of base pairs, a new hypothesis of the origin of genetic codes is introduced and some characterizations of bioinformatics is obtained. In order to explain the recombining codon better, we further compute the synonymous codon preference combined with the strength under the new classifications. The results show that the algebraic structure and the operation fully demonstrate the important features of biological significance about codons named quasi-amino acids genetic coding. This has avoided some inexplicable events of the genomic genetic codes. The algebraic system with operation based on the protein sequences coded by codon is established. It is proved a distribution lattice of rings theoretically and a genomic signature method, called algebraic genomic signature method, is introduced. The present paper will focus on the description of a novel quasi-amino acid structure of the standard genetic code and its corresponding experimental data. The aim of this work is to connect the algebraic relationships among codon with the physicochemical properties of amino acids, particularly to quasi-amino acids. It is obvious that the quasi-amino acids coded by genetic code must be strongly connected with the physicochemical properties. Here, it is possible to infer a different codon order and a different algebraic structure of the genetic code by using the same basic properties. This structure will reflect the quantitative relationship among codon in genes.

According to the protein sequence decided by codon, the protein sequence operation and the algebraic system are introduced by using the natural operation based on the 64 codons.

Resting on some well known algebraic structures, such as semigroups, semi-rings and so on,[7] the natural operation table of the codon is established. Its structure by using the algebraic theory and its corresponding physicochemical properties of amino acids are characterized. As a result, a novel quasi-amino acid code classification method is introduced and the corresponding algebraic operation table over the ZU of the 16 kinds of quasi-amino acid classes is established. The internal relation is revealed about quasi-amino acids. The results show that there exist some very close correlations among the properties of the quasi-amino acids coded by codon.

The natural operation has been established based on the $\Delta = \{g, t, a, c\}$ as in Refs.[8–10] and it forms a four elements field naturally. This structure is induced by the bijection with the polynomial set $\{0, 1, x, 1+x\}$, where the sum $+$ is defined as the sum module 2, by components, and the product $*$ is defined as the product module of the irreducible polynomial $x^2 + x + 1$.

In fact, the additive table constructs the double-stranded helix model. According to the above tables, algebraic operations about 64 codons can be introduced as follows:[8]

Let $\Sigma = \{ggg, ggt, gga, ggc, \ldots\}$. For any $x_1x_2x_3$, $y_1y_2y_3 \in \Sigma$, then $x_1x_2x_3 + y_1y_2y_3 = (x_1 + y_1)(x_2 + y_2)(x_3+y_3)$, $x_1x_2x_3 * y_1y_2y_3 = (x_1 * y_1)(x_2 * y_2)(x_3 * y_3)$ are called the algebraic operations about 64 codons over Σ.

For example, the codon tca, atc, so $tca+atc = cat$, $tca * atc = act$, which is different from Refs.[3,4,6]. In Ref.[4], the operation about codon is used by the order of codon such as follows: $tca+atc(\doteq 28+49(\mathrm{mod}64) \doteq 13) = tac$.

$(\Sigma, +, *)$ is a field naturally, where $\Sigma = \{ggg, ggt, gga, ggc, \ldots\}$. The codon symmetry is obvious in Table 2 in Ref.[8]. The codon is divided into 4 subsets by using the order of g, t, a, c, denoted by $\Sigma_i, i \in N$, where $\Sigma_1 = \{xgz \mid x, z \in \Sigma\}$; $\Sigma_2 = \{xtz \mid x, z \in \Sigma\}$; $\Sigma_3 = \{xaz \mid x, z \in \Sigma\}$; $\Sigma_4 = \{xcz \mid x, z \in \Sigma\}$. It is seen that Σ_i is lying in the ith column of Table 1 in Ref.[8].

Epstein[11] stated that amino acids cannot be randomly allocated by considering the features of the genetic code. The set of codon $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4$ is arranged as a code scale of physicochemical properties of amino acids in proteins. Codon in the second subset coded by xtz is just the most hydrophobic amino acids

which is denoted by Σ_2. Codon in the third subset coded by xaz is just the most hydrophilic amino acids which is denoted by Σ_3, where $x, z \in \Delta = \{g, t, a, c\}$. Similarly, the slight hydrophobic amino acids are denoted by Σ_1 and the slight hydrophilic amino acids by Σ_4. Therefore, there are results as follows:[9]

Let $\Sigma = \{ggg, ggt, gga, ggc, \ldots\}$, then $\Sigma = \bigcup_{i=1}^{4} \Sigma_i$ and for any $x_i \in \Sigma_i (i = 1, 2, 3, 4)$, there exists

$$x_1 + x_i \in \Sigma_i, \quad i = 1, 2, 3, 4;$$
$$x_2 + x_3 \in \Sigma_4; \quad x_2 + x_4 \in \Sigma_3; \quad x_3 + x_4 \in \Sigma_2.$$

That is, hydrophobic amino acids plus slight hydrophobic ones to hydrophilic ones and so on. Hence the operative table reflects the physicochemical properties completely.

It is shown that the hydrophilic (polar) amino acids subset Σ_3 are algebraically responding to hydrophobic (non-polar) amino acids Σ_2. Besides, the non-polar amino acids tend to buffer the effect produced by polar amino acids, such that the global effect corresponds to protein regions with middle polarity in protein structure.[2,12,13] As we expected, the above results display this characteristics.

Particularly, for stop codon and start codon, there exists $taa + atg = cca, tag + atg = ccg$, but $tga + atg = cta$, while $taa * atg = aag, tag * atg = aag$, but $tga * atg = agg$. This result just shine that the stop codon tga has the different property from other stop codon. Just right, since 1980, according to sequence of the mitochondrion from human and ox genomic genetic code, scientists have discovered that the codon tga is the amino acids Trp (tryptophane), but is not the stop codon.[1] So the above operation just shows the code mutation's case of higher mammal. The biological significance of the mitochondrion genetic codes and the possible evolution path of genetic codon needs further research. Meanwhile, in Ref.[14], since Yang proposed that the 4^3 table of triplet codes given by Crick[15,16] is a typical wrong application and a wrong calculation, the present theory avoids this problem by using the above operative tables.

2. Some important properties about quasi-amino acids

In the 20th century, one of the most valuable research results has been the proposal of the codon.[15−17] The origin and evolution of the genetic code are also the most interesting yet challenging problems for understanding life phenomena. In fact, in the mitochondrion from corn genomic codon, cgg is just Try, and is not the Arg like in other genetic code. Because there exist some internal relations,[2,13] the origin, composition, structure, and evolution of the genetic code mentioned above can be revealed.[18−29] Some excellent properties such as the square deviation (SD) of the water affinity and π-energy can be found out, while the amino acids in Ref.[2] have no obviously symmetric form. Therefore, quasi-amino acids based on logic operation of codon will be established from another viewpoint. It is very important to study the inner relationships of quasi-amino acids.

Due to some specific characteristics of the operation, codon is divided into 16 kinds of quasi-amino acids. A novel quasi-amino acids classification method is established by using logic characteristics of the codon table.

Definition 2.1. Let

$$WC \equiv \{tgg, tgt, tga, tgc\}, \quad RS \equiv \{agg, agt, aga, agc\},$$
$$LF \equiv \{ttg, ttt, tta, ttc\},$$
$$MI \equiv \{atg, att, ata, atc\}, \quad ED \equiv \{gag, gat, gaa, gac\},$$
$$-Y \equiv \{tag, tat, taa, tac\},$$
$$KN \equiv \{aag, aat, aaa, aac\}, \quad QH \equiv \{cag, cat, caa, cac\},$$
$$S' \equiv \{tcg, tct, tca, tcc\},$$
$$R' \equiv \{cgg, cgt, cga, cgc\}, L' \equiv \{ctg, ctt, cta, ctc\}.$$

Let $ZU = \{G, V, ED, A, WC, \quad LF, -Y, S', RS, MI, KN, T, R', L', QH, P\}$, then the ZU is called the quasi-amino acids set.

Where $-Y$ is a new stop codon, but $(tga \in) WC$ is not a stop codon. It is just tallied with the result from Ref.[2]. Hence we obtain Table 1.

Sánchez *et al*[4] proposed that the third base transversion changes the codon parity and order, however, this kind of mutations generally does not produce extreme hydrophobic changes in the coded amino acids. Now, the set ZU shows that the third mutation generally does not produce extreme quasi-amino acids structure changes. But the first and the second mutation is very dangerous in the quasi-amino acids structure hydrophobic changes in protein sequence. The mean and the square deviations of the water affinity, base force, molecular weight, and π-energy over the ZU are computed in Table 2, where square deviation is denoted by SD and

$$\mathrm{SD} = \frac{1}{n-1}\sum_{i=1}^{4}(x_{ki} - \bar{x}_k)^2, \quad k = 1, 2, \ldots 16.$$

The results show that there exist some very close correlations among the properties of the quasi-amino acids coded codon.

Table 1. The primal genetic code table induced by the primal order $\{g, t, a, c\}$ and the corresponding novel quasi-amino acids.

$$\begin{pmatrix}
ggg & G & gtg & V & gag & ED & gcg & A \\
ggt & G & gtt & V & gat & ED & gct & A \\
gga & G & gta & V & gaa & ED & gca & A \\
ggc & G & gtc & V & gac & ED & gcc & A \\
tgg & WC & ttg & LF & tag & -Y & tcg & S' \\
tgt & WC & ttt & LF & tat & -Y & tct & S' \\
tga & WC & tta & LF & taa & -Y & tca & S' \\
tgc & WC & ttc & LF & tac & -Y & tcc & S' \\
agg & RS & atg & MI & aag & KN & acg & T \\
agt & RS & att & MI & aat & KN & act & T \\
aga & RS & ata & MI & aaa & KN & aca & T \\
agc & RS & atc & MI & aac & KN & acc & T \\
cgg & R' & ctg & L' & cag & QH & ccg & P \\
cgt & R' & ctt & L' & cat & QH & cct & P \\
cga & R' & cta & L' & caa & QH & cca & P \\
cgc & R' & ctc & L' & cac & QH & ccc & P
\end{pmatrix}$$

Proposition 2.1. The following important properties about the quasi-amino acids from Table 2 can be found out.

(1) The variance of the water affinity is similar to 0.033 expect G;
(2) The variance of the molecular weight is similar to 373.583;
(3) The variance of the π-energy is similar to 0.003;
(4) The molecular weight and the π-energy display the close positive relationships;
(5) The molecular weight and the water affinity display the close negative relationships;
(6) The water affinity and the π-energy display the close negative relationships;
(7) The properties of the base force have no clear characteristics for all index.

As a result of the artificial function, sour degeneracy degree, the close meaning of the quasi-amino acids set ZU is very logical and interesting. So these results have fully proved that the significant method of our dividing codon into 16 kinds of quasi-amino acids really exposes the inner meaning of codon. All these correlation relationships may play an important part in establishing the logic relationship between codons and the quasi-amino acids during the course of life origination. Meanwhile, it avoids the problem of 'the 64 codon assignments of mRNA to amino acids is probably completely wrong' proposed by Yang in Ref.[14].

Table 2. Some important properties of codon about quasi-amino acids code.

codons	quasi-amino acids	water affinity	mean ± SD	base force	mean ± SD	MW	mean ± SD	π energy	mean ± SD
ggg	*G*	1.1	1.25± 0.057	144	173± 1495.333	849	825± 373.583	0.81	0.79± 0.003
ggt	*G*	1.5		184		810		0.73	
gga	*G*	1.0		141		833		0.86	
ggc	*G*	1.4		223		809		0.77	
gtg	*V*	1.5	1.6± 0.033	152	155.5± 433	810	786.25± 373.583	0.73	0.71± 0.003
gtt	*V*	1.8		152		771		0.65	
gta	*V*	1.4		134		794		0.78	
gtc	*V*	1.7		184		770		0.69	
gag	*ED*	1.1	1.2± 0.033	145	153.25± 422.917	833	809.25± 373.583	0.86	0.84± 0.003
gat	*ED*	1.4		143		794		0.78	
gaa	*ED*	1.0		141		817		0.91	
gac	*ED*	1.3		184		793		0.82	
gcg	*A*	1.4	1.5± 0.033	209	204± 199.583	809	785.25± 373.583	0.77	0.76± 0.003
gct	*A*	1.7		195		770		0.69	
gca	*A*	1.3		191		793		0.82	
gcc	*A*	1.6		222		769		0.73	
tgg	*WC*	1.5	1.6± 0.033	141	156.25± 563.583	810	786.25± 373.583	0.73	0.7125± 0.003
tgt	*WC*	1.8		152		771		0.65	
tga	*WC*	1.4		141		794		0.78	

Table 2 (*Continued*)

codons	quasi-amino acids	water affinity	mean ± SD	base force	mean ± SD	MW	mean ± SD	π energy	mean ± SD
tgc	WC	1.7		191		770		0.69	
ttg	LF	1.8	1.9± 0.033	109	112.5± 433	771	747.25± 373.58	0.65	0.6325± 0.003
ttt	LF	2.1		109		732		0.57	
tta	LF	1.7		91		755		0.7	
ttc	LF	2.0		141		731		0.61	
tag	$-Y$	1.4	1.5± 0.033	95	103.5± 416.333	794	770.25± 373.583	0.78	0.7625± 0.003
tat	$-Y$	1.7		94		755		0.7	
taa	$-Y$	1.3		91		778		0.83	
tac	$-Y$	1.6		134		754		0.74	
tcg	S'	1.7	1.8± 0.033	159	154.25± 199.583	770	746.25± 373.58	0.69	0.6725± 0.003
tct	S'	2.0		145		731		0.61	
tca	S'	1.6		141		754		0.74	
tcc	S'	1.9		172		730		0.65	
agg	RS	1.1	1.2± 0.033	144	160± 574	833	809.25± 373.583	0.86	0.8425± 0.003
agt	RS	1.4		156		794		0.78	
aga	RS	1.0		145		817		0.91	
agc	RS	1.3		195		793		0.82	
atg	MI	1.4	1.5± 0.033	112	115.5± 286.333	794	770.25± 373.583	0.78	0.7625± 0.003
att	MI	1.7		122		755		0.7	
ata	MI	1.3		94		778		0.83	
atc	MI	1.6		134		754		0.74	
aag	KN	1.0	1.1± 0.033	113	121.5± 416.333	817	793.25± 373.583	0.91	0.8925± 0.003
aat	KN	1.3		112		778		0.83	
aaa	KN	0.9		109		801		0.96	
aac	KN	1.2		152		777		0.87	
acg	T	1.3	1.4± 0.033	170	166± 211.667	793	769.25± 373.583	0.82	0.80± 0.003
act	T	1.6		156		754		0.74	
aca	T	1.2		152		777		0.87	
acc	T	1.5		184		753		0.78	
cgg	R'	1.4	1.5± 0.033	158	174.5± 569.667	809	785.25± 373.583	0.77	0.7525± 0.003
cgt	R'	1.7		172		770		0.69	
cga	R'	1.3		159		793		0.82	
cgc	R'	1.6		209		769		0.73	
ctg	L'	1.7	1.8± 0.033	113	116.5± 433	770	746.25± 373.583	0.69	0.6725± 0.003
ctt	L'	2.0		113		731		0.61	
cta	L'	1.6		95		754		0.74	
ctc	L'	1.9		145		730		0.65	
cag	QH	1.3	1.4± 0.033	113	121.5± 416.333	793	769.25± 373.583	0.82	0.8025± 0.003
cat	QH	1.6		112		754		0.74	
caa	QH	1.2		109		777		0.87	
cac	QH	1.5		152		753		0.78	
ccg	P	1.6	1.7± 0.033	159	154± 206	769	745.25± 373.583	0.73	0.71± 0.003
cct	P	1.9		144		730		0.65	
cca	P	1.5		141		753		0.78	
ccc	P	1.8		172		729		0.69	

3. The algebraic operation of the protein sequences based on quasi-amino acids

In addition, a new additional operation (Table 3) and multiple operation (Table 4) about the ZU can be established by using the algebraic operation over the Σ respectively.

Proposition 3.1. $(ZU, \oplus, \otimes)$ is a field.

Definition 3.1. $(ZU, \oplus, \otimes)$ is called the ZU protein system.

According to Tables 3 and 4, the properties of quasi-amino acids rest on the coded codon,[2] which may veritably reflet the corresponding relationships between codon and quasi-amino acids to some extent. It is clear that the comprehension of these correlations could provide a unique vantage for elucidating co-evolutionary mechanisms between genetic codon and amino acids. Further, it has displayed the characteristics that the first and second codes decided the quasi-amino acids.

Table 3. The additional operation table over $(ZU, \oplus)$.

$\oplus$	G	V	ED	A	WC	LF	$-Y$	S'	RS	MI	KN	T	R'	L'	QH	P
G	G	V	ED	A	WC	LF	$-Y$	S'	RS	MI	KN	T	R'	L'	QH	P
V	V	G	A	ED	LF	WC	S'	$-Y$	MI	RS	T	KN	L'	R'	P	QH
ED	ED	A	G	V	$-Y$	S'	WC	LF	KN	T	RS	MI	QH	P	R'	L'
A	A	ED	V	G	S'	$-Y$	LF	WC	T	KN	MI	RS	P	QH	L'	R'
WC	WC	LF	$-Y$	S'	G	V	ED	A	R'	L'	QH	P	RS	MI	KN	T
LF	LF	WC	S'	$-Y$	V	G	A	ED	L'	R'	P	QH	MI	RS	T	KN
$-Y$	$-Y$	S'	WC	LF	ED	A	G	V	QH	P	R'	L'	KN	T	RS	MI
S'	S'	$-Y$	LF	WC	A	ED	V	G	P	QH	L'	R'	T	KN	MI	RS
RS	RS	MI	KN	T	R'	L'	QH	P	G	V	ED	A	WC	LF	$-Y$	S'
MI	MI	RS	T	KN	L'	R'	P	QH	V	G	A	ED	LF	WC	S'	$-Y$
KN	KN	T	RS	M I	QH	P	R'	L'	ED	A	G	V	$-Y$	S'	WC	LF
T	T	KN	M I	RS	P	QH	L'	R'	A	ED	V	G	S'	$-Y$	LF	WC
R'	R'	L'	QH	P	RS	MI	KN	T	WC	LF	$-Y$	S	G	V	ED	A
L'	L'	R'	P	QH	MI	R S	T	KN	LF	WC	S'	$-Y$	V	G	A	ED
QH	QH	P	R'	L'	KN	T	RS	MI	$-Y$	S'	WC	LF	ED	A	G	V
P	P	QH	L'	R'	T	KN	M I	RS	S'	$-Y$	LF	WC	A	ED	V	G

Table 4. The multiple operation table over $(ZU, \otimes)$.

$\otimes$	G	V	ED	A	WC	LF	$-Y$	S'	RS	MI	KN	T	R'	L'	QH	P
G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
V	G	V	ED	A	G	V	ED	A	G	V	ED	A	G	V	ED	A
ED	G	ED	A	V	G	ED	A	V	G	ED	A	V	G	ED	A	V
A	G	A	V	ED	G	A	V	ED	G	A	V	ED	G	A	V	ED
WC	G	G	G	G	WC	WC	WC	WC	RS	RS	RS	RS	R'	R'	R'	R'
LF	G	V	ED	A	WC	LF	$-Y$	S'	RS	MI	KN	T	R'	L'	QH	P
$-Y$	G	ED	A	V	WC	$-Y$	S	LF	RS	KN	T	MI	R	QH	P	L'
S'	G	A	V	ED	WC	S'	LF	$-Y$	RS	T	MI	KN	R	P	L'	QH
RS	G	G	G	G	RS	RS	RS	RS	R'	R'	R'	R'	WC	WC	WC	WC
MI	G	V	ED	A	RS	MI	KN	T	R'	L'	QH	P	WC	LF	$-Y$	S'
KN	G	ED	A	V	RS	KN	T	MI	R'	QH	P	L'	WC	$-Y$	S'	LF
T	G	A	V	ED	RS	T	MI	KN	R'	P	L'	QH	WC	S'	LF	$-Y$
R'	G	G	G	G	R'	R'	R'	R'	WC	WC	WC	WC	RS	RS	RS	RS
L'	G	V	ED	A	R'	L'	QH	P	WC	LF	$-Y$	S'	RS	MI	KN	T
QH	G	ED	A	V	R'	QH	P	L'	WC	$-Y$	S'	LF	RS	KN	T	MI
P	G	A	V	ED	R'	P	L'	QH	WC	S'	LF	$-Y$	RS	T	MI	KN

4. Conclusion

For all life forms, codon usage is non-random and associated to various factors such as gene expression level, gene length and protein secondary structures. Moreover, most amino acids in all species bear a highly significant association with gene functions.

Now the novel ZU of the 16 kinds of quasi-amino acids code has been introduced and the corresponding algebraic operation table has been established. The internal relation has been revealed for quasi-amino acids. As a result of the artificial function, sour degeneracy degree, the excellent properties about the set ZU have been obtained. These results have fully proved that the significant method of our dividing codon into 16 kinds of quasi-amino acids really exposes the inner meaning of codon. All these correlation relationships may play an important part in establishing the logic relationships between codon and the quasi-amino acids during the course of life origination. Furthermore, it has been pointed that the set $(ZU, \oplus, \otimes)$ is a field. This result inspires us that amino acids may have many abstruse meaning corresponding to codon. It is clear that comprehension of these correlations could provide unique vantage for elucidating co-evolutionary mechanisms between genetic codon and quasi-amino acids. Meanwhile, it solves the problem of 'the 64 codon assignments of mRNA to amino acids is probably completely wrong' proposed by Yang in Ref.[14]. However, it is only a mathematical model and has no effective way to achieve inquiring emulative experiment. Otherwise, it would be another kilometre-stone in the study of bioinformatics.

References

[1] Chen W C, Chen Z H, Wang Z Q, Liu W and Zuo L 2002 *Acta Biophysica Sin.* **18**(1) 87 (in Chinese)
[2] Ma F, Wu Y T and Xu X F 2003 *J. Anhui Agricultural University* **30** 439 (in Chinese)
[3] Sánchez R, Grau R and Morgado E 2006 *Math. Biosci.* **202** 156
[4] Sánchez R, Morgado E and Grau R 2005 *J. Math. Biol.* **51** 431
[5] Sánchez R, Grau R and Morgado E 2004 *MATCH Commun. Math. Comput. Chem.* **52** 29
[6] Sánchez R, Perfetti L A, Grau R and Morgado E 2005 *MATCH Commun. Math. Comput. Chem.* **54** 3
[7] Howie J M 1976 *An Introduction to Semigroup Theory* (London: Academic)
[8] Zhu P, Tang X Q, Guo L X and Xu Z Y 2007 *Proceedings of the 4th International Conf. on Impulsive and Hybrid Dynamical Systems* (DCDIS A) p2661
[9] Zhu P, Tang X Q, Guan W H and Xu Z Y 2006 *Proceedings of the 16th International Conf. on Artificial Reality and Telexistence - Workshops* p282
[10] Li R H and Yu W 2001 *Chin. J. Computers* **24** 974 (in Chinese)
[11] Epstein C J 1966 *Nature* **210** 25
[12] Volkenshtein M V 1985 *Editorial MIR, Moscu, Capitulo* **17** 621
[13] Chen Z H, Chen W C, Qiu H X and Wang Z Q 2001 *Acta Biophysica Sin.* **17** 187 (in Chinese)
[14] Yang X Y 2006 *Progress Mod. Biomed.* **6** 3 (in Chinese)
[15] Crick F H C 1968 *J. Mol. Biol.* **38** 367
[16] Watson I D and Crick F H C 1953 *Nature* **171** 964
[17] Jiang F and Li N 2007 *Chin. Phys.* **16** 392
[18] Sánchez R, Morgado E and Grau R 2005 *Bull. Math. Biol.* **67** 1
[19] Bai Y Q, Tang A H, Wang S Q and Zhu X 2007 *Acta Phys. Sin.* **56** 3607 (in Chinese)
[20] Xu H, Guo A M and Ma S S 2007 *Acta Phys. Sin.* **56** 1208 (in Chinese)
[21] Liu X L, Xu H, Ma S S, Deng C S and Guo A M 2006 *Acta Phys. Sin.* **55** 5562 (in Chinese)
[22] Wei Z Y, Zang L H, Li M, Fan W and Xu Y J 2005 *Acta Phys. Sin.* **54** 4955 (in Chinese)
[23] Hornos J E and Hornos Y M 1993 *Phys. Rev. Lett.* **71** 4401
[24] Jijoy Joseph and Roschen Sasikumar 2006 *BMC Bioinformatics* **7** 243
[25] Holm L and Sander C 1999 *Nucleic Acids Reserach* **27** 244
[26] Liu T, Wang Y and Wang K L 2007 *Chin. Phys.* **16** 0272
[27] Yan X L and Dong R X 2007 *Chin. Phys.* **16** 2062
[28] Ma S S and Xu H 2007 *Chin. Phys.* **16** 0862
[29] Xie P, Dou S X and Wang P Y 2006 *Chin. Phys.* **15** 0536