6

И Н С Т И Т У Т
ЯДЕРНОЙ ФИЗИКИ СОАН СССР

ПРЕПРИНТ И Я Ф 75 - 12

B.G.Konopel'chenko, Yu.B.Rumer

CLASSIFICATION OF THE CODONS OF THE GENETIC CODE .I.

Новосибирск

1975

CLASSIFICATION OF THE CODONS OF THE GENETIC CODE.I.

B.G.Konopel'chenko, Yu.B.Rumer

Institute of Nuclear Physics, Novosibirsk 90, USSR

A B S T R A C T

A classification is proposed based on the break down of 16 roots into two octets - strong and weak. It is shown that the sequence of letters CGUA is the canonical one. The quantitative rules are given to define the strength of roots and letters.

CLASSIFICATION OF THE CODONS OF THE GENETIC CODE.I.

B.G.Konopel'chenko, Yu.B.Rumer

One of the authors (Yu.B.R.) in his paper[1] pointed out that it is reasonable to consider the termination /z) separately from the two-letter root (xy/. The 16 roots of 64 codons fall into octets (see Table 1). The first octet includes 8 "strong" roots, each of which codes the same amino-acid with any of the terminations C,G,U,A. The secomd octet includes 8 "weak" roots, which when terminated with a pyrimidine letter (C,U) code one amino-acide and when terminated with a purine letter (G,A) code for a different one.

Table 1

Strong roots	Weak roots
CC	AA
GC	UA
UC	GA
AC	CA
CG	AU
GG	UU
CU	AG
GU	UG

The table of roots possesses two anomalies. First, each of the three amino-acids Ser, Leu and Arg is coded by two codons with different roots. One of the roots being strong and another weak[1].

The second anomaly is that for the three codons with

roots (UA), (AU) and (UG) and terminations of purine pair the code degenracy is entirely cancelled.

In the present paper we will consider the properties of the two octets of roots leaving out the anomalites mentioned above. Based on the empirical evidence it will be shown that there exists a canonical sequence of letters namely C,G,U,A. We will also introduce the quantitative characteristics of the letter and root strengths and show that in the root (xy) the letter x is as important as y.

Consider the table of roots. Following /1/ we assume that the root strength is defined by the letter content. Therefore, it is naturally to classify the letters as strong and weak. Since we have the same number of strong and weak roots, the number of strong and weak letters is also the same i.e. there are two strong letter and two weak ones.

Then C as the most frequently encountered letter in the strong roots is the strongest one. Analogous the A is the weakest letter. From two letters G and U being of intermediate strength one should naturally consider G as the strong letter since the root (GG/ is the strong root and (UU/ is the weak. As a result we have the following sequence of letters

C	G	U	A
strongest	strong	weak	weakest

This sequence of letters enables us to formulate the following rules defining the root strength:

a) the strength of the root containing C or A as y letter is determined by the y letter strength;

b) the strength of the root containing G or U as the y

letter is determined by the x letter strength.

We see that the root strength is determined not only by the second letter (y) but the first letter (x) too. Note that the simplicity of rules is closely connected with the canonical order of letters.

Let us dispose the letters C,G,U,A in the form of two x two matrix

$$\begin{pmatrix} C & G \\ U & A \end{pmatrix}$$

The first row of the matrix contains strong letters C,G and the second row contains weak letters U,A. The first column contains pyrimidines and the second - purines. When we form the tensor square of this matrix we obtain the four x four matrix of roots (xy):

$$\begin{pmatrix} CC & GC & CG & GG \\ UC & AC & UG & AG \\ CU & GU & CA & GA \\ UU & AU & UA & AA \end{pmatrix}$$

We see that all strong roots dispose above the lateral diagonal and the weak roots dispose under the lateral diagonal of the matrix. It is reasonable to call the roots located on the diagonal as transition roots. Note that this matrix differ from that given in /1/. Hence, the transition roots differs too.

We will now quantitavely formulate the rules given above To this end, we will associate with of the letters C,G,U,A

a „charge" Q that depends of the place of the letter in the root.

Table 2

First letter of the root	Second letter of the root
$Q_1(C) = 2$	$Q_2(C) = 4$
$Q_1(G) = 2$	$Q_2(G) = 1$
$Q_1(U) = -2$	$Q_2(U) = -1$
$Q_1(A) = -2$	$Q_2(A) = -4$

Note, that under the substitution C ↔ A , G ↔ U, Q changes the sign $Q \to -Q$. The "Charge" Q is a quantitative characteristic of the letter strength. The root strength is defined by the following formulas:

$$Q(XY) = Q_1(X) + Q_2(Y)$$

For the strong roots $Q > 0$; for the weak roots $Q < 0$ (see Table 3).

Table 3

$$\begin{pmatrix} 6 & 6 & 3 & 3 \\ 2 & 2 & -1 & -1 \\ 1 & 1 & -2 & -2 \\ -3 & -3 & -6 & -6 \end{pmatrix}$$

Note that the charge values proposed above are not the only possibles. General limitations for the charges Q have the form ($|Q|$ is modulus of Q):

$$Q_{1,2}(C) = -Q_{1,2}(A), \quad Q_{1,2}(G) = -Q_{1,2}(U),$$
$$Q_2(C) > |Q_2(G,U)|, \ |Q_1(C,G,U,A)|,$$
$$|Q_2(G,U)| < |Q_1(C,G,U,A)| .$$

We chose the charge values in such a way to emphasize the falling of each octet (see Table 1) into quartet (for example CC, GC, UC, AC) and two doublets (for example CG, GG and CU, GU). The quartet in its turn falls into two doublets.

Considering the first letter charges (see Table 2) one can perceive the connection of the charge with the number of hydrogen bonds of corresponding letter (three for strong letters C,G and two for weak letters A,U /1/).

The authors are grateful to A.A.Prokopenko for numerous discussions.

R e f e r e n c e:

/1/ Yu.B.Rumer, Doklady Akademii Nauk SSSR, 183, 225 (1968).

Ответственный за выпуск Г.А.СПИРИДОНОВ
Подписано к печати 30.I-75г. МН 02643
Усл.0,25 печ.л., тираж 200 экз. Бесплатно.
Заказ № I2 ПРЕПРИНТ
Отпечатано на ротапринте в ИЯФ СО АН СССР