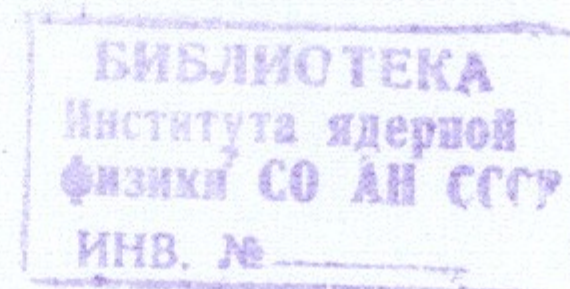# И Н С Т И Т У Т
ЯДЕРНОЙ ФИЗИКИ СОАН СССР

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

ON THE CORRELATION BETWEEN THE STRUCTURE OF AMINO ACIDS AND THE DEGENERACY OF THE GENETIC CODE

Новосибирск

1975

ON THE CORRELATION BETWEEN THE STRUCTURE OF AMINO

ACIDS AND THE DEGENERACY OF THE GENETIC CODE

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

Institute of Nuclear Physics,
Novosibirsk 90, USSR

A b s t r a c t

It is shown that there exists a correlation between the
hydrogen structure of the amino acids R-groups and the degree of
degeneracy of the genetic code.

ON THE CORRELATION BETWEEN THE STRUCTURE OF AMINO

ACIDS AND THE DEGENERACY OF THE GENETIC CODE

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

Institute of Nuclear Physics,

Novosibirsk 90, USSR

The purpose of the present communication is to show the existence of a correlation between the hydrogen structure of the amino acids R-groups and the decomposition of the codons roots into strong and weak ones.

In the papers [1] it was shown that sixteen roots (xy| of the codons (xy|z) decompose into two octets. The First octet includes eight "strong" roots, each of which codes for the same amino acid with any of the terminations C, G, U, A. The second octet includes eight "weak" roots, which when terminated with a pyrimidine nucleotide (C, U) code for one amino acid, and when terminated with a purine nucleotide (G, A) code for a different one (See table 1).

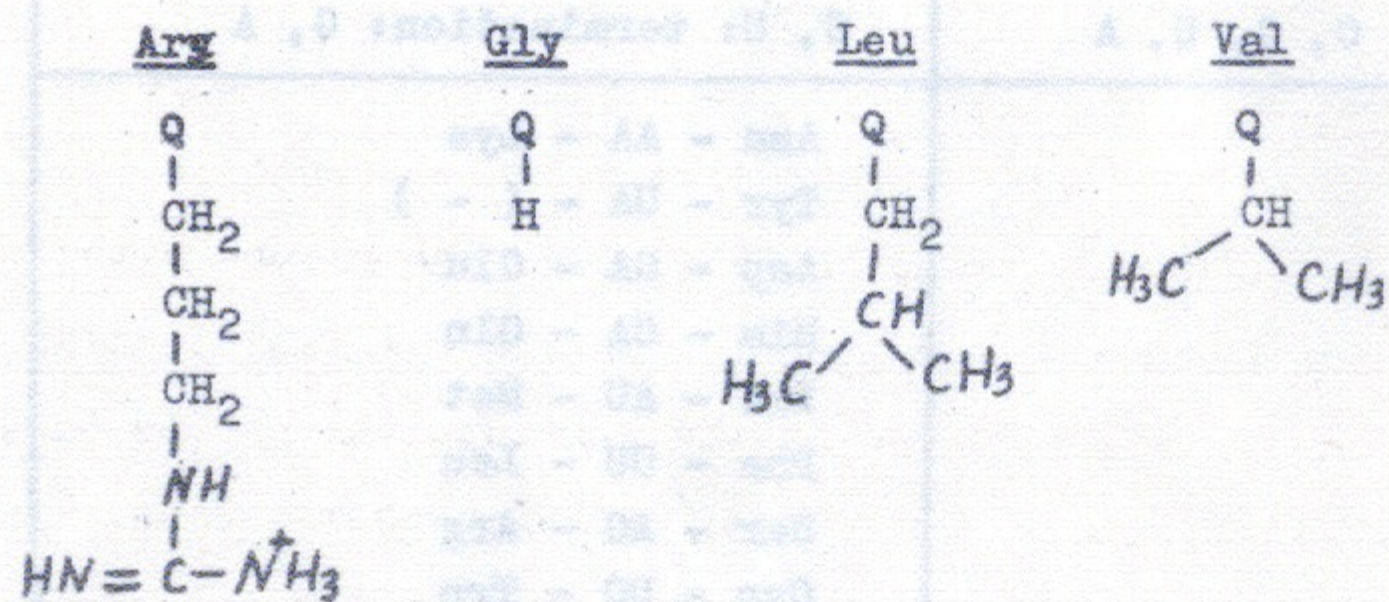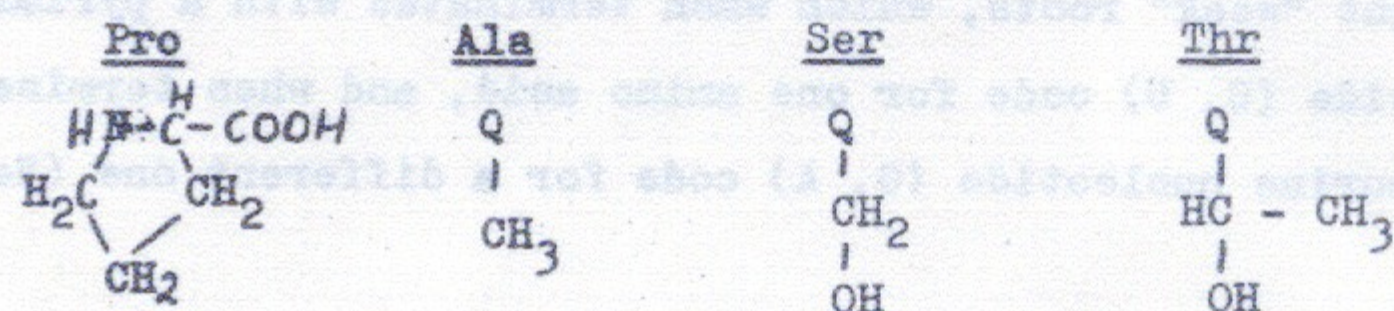| Strong roots termination: C, G, U, A	weak roots C, U: termination: G, A
CC – Pro	Asn – AA – Lys
GC – Ala	Tyr – UA – (–)
UC – Ser	Asp – GA – Glu
AC – Thr	His – CA – Gln
CG – Arg	Ile – AU – Met
GG – Gly	Phe – UU – Leu
CU – Leu	Ser – AG – Arg
GU – Val	Cys – UG – Trp

Table 1

Let us call the amino acids which are coded by the strong roots the strong amino acids and the amino acids which are coded by the weak roots the weak ones.

Table 1. contains two type of irregularity. First, three amino acids Ser, Arg, Leu are both strong and weak. Secondly, in the three cases the simple structure of the table 1 is broken, namely; α) nonsenses (UA|G) and (UA|A) are different, β) Ile is coded by three codons (AU|C), (AU|U) and (AU|A), while Met is coded by one codon (AU|G), γ)Try is coded by one codon (UG|G) and codon (UG|A) is either nonsense (E.coli) or Cys (vertebrates) [2]. Thus, Ile and Cys more resemble the strong amino acids than the weak ones.

Let us now consider the chemical structure of the amino acids. (where $Q = H_2N - \overset{H}{\underset{}{C}} - COOH$)

Strong amino acids $\left(\frac{Q}{R}\right)$

Pro

$$HN - \overset{H}{C} - COOH$$
$$H_2C \qquad CH_2$$
$$CH_2$$

Ala

$$Q$$
$$|$$
$$CH_3$$

Ser

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$OH$$

Thr

$$Q$$
$$|$$
$$HC - CH_3$$
$$|$$
$$OH$$

Arg

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$NH$$
$$|$$
$$HN = C - N^+H_3$$

Gly

$$Q$$
$$|$$
$$H$$

Leu

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH$$
$$H_3C \qquad CH_3$$

Val

$$Q$$
$$|$$
$$CH$$
$$H_3C \qquad CH_3$$

Weak amino acids (+Ser, Arg, Leu)

Asn

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$O = C - NH_2$$

Lys

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$H_3N^+ - CH_2$$

Tyr

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$\text{(benzene ring)}$$
$$|$$
$$OH$$

Asp

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$O = C - O^-$$

Glu

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$O = C - O^-$$

His

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$C - N$$
$$HC - N$$
$$|$$
$$H$$

Gln

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$O = C - NH_2$$

Ile

$$Q$$
$$|$$
$$CH$$
$$H_2C \qquad CH_3$$
$$|$$
$$H_3C$$

Met

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$CH_2$$
$$|$$
$$S - CH_3$$

Phe

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$\text{(benzene ring)}$$

Cys

$$Q$$
$$|$$
$$CH_2$$
$$|$$
$$SH$$

Trp

$$Q$$
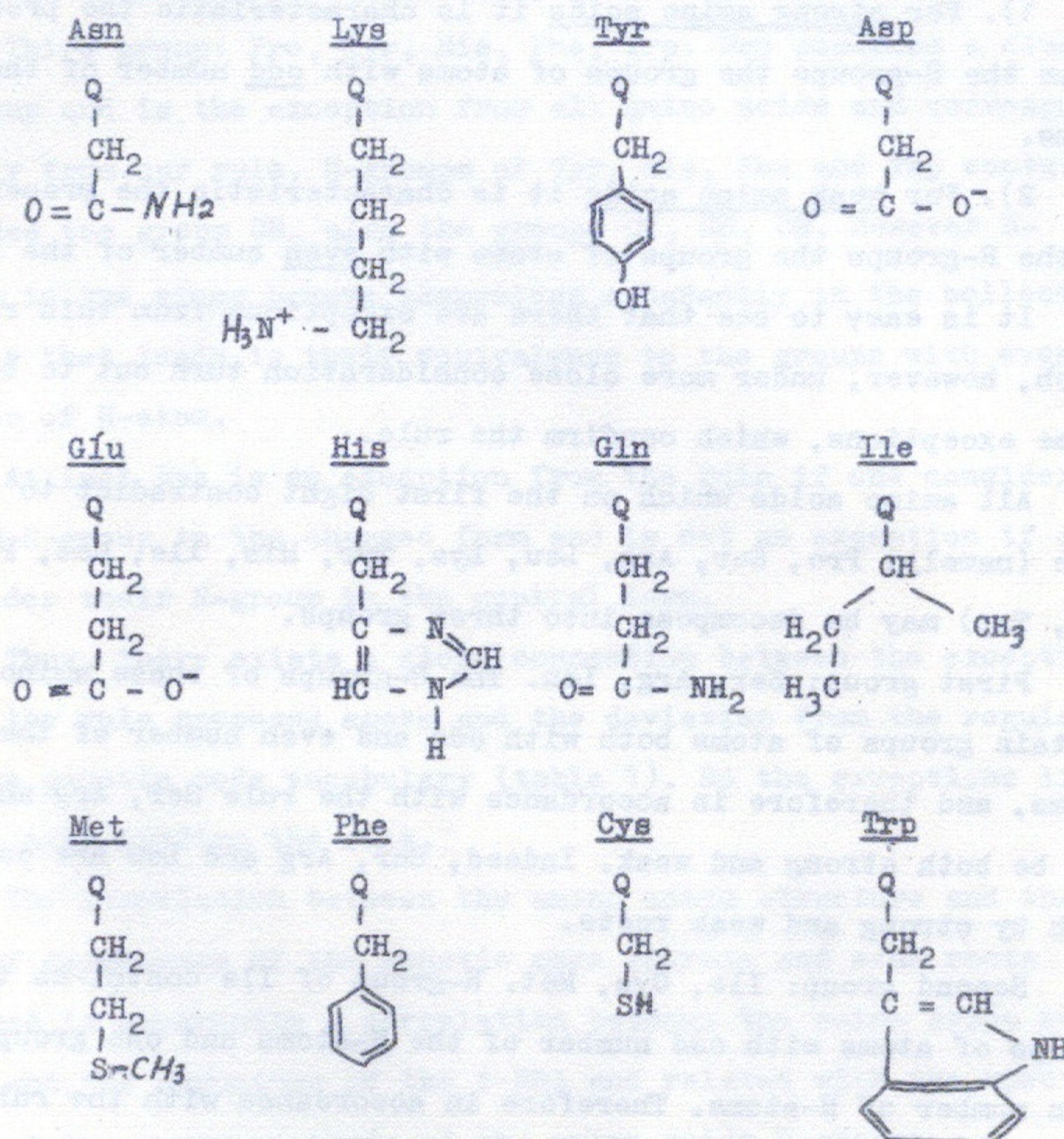$$|$$
$$CH_2$$
$$|$$
$$C = CH$$
$$\qquad NH$$

Table 2

Considering the formulas cited above we observe that for the strong amino acids it is characteristic the presence in the R-groups the groups CH, CH_3, OH, NH, NH_3, i.e. the groups with odd number of the hydrogen atoms, while for the weak amino acids it is characteristic the presence in the R-groups the groups CH_2, i.e. the groups with even number of the hydrogen atoms. Thus, there exists a correlation between the decomposition of all amino acids into strong and weak and the properties of their R-gro-

ups. This correlation may be represented by the following rule:

1). For strong amino acids it is characteristic the presence in the R-groups the groups of atoms with odd number of the H-atoms.

2). For weak amino acids it is characteristic the presence in the R-groups the groups of atoms with even number of the H-atoms.

It is easy to see that there are exceptions from this rule, which, however, under more close consideration turn out to be those exceptions, which confirm the rule.

All amino acids which on the first sight contradict to the rule (namely; Pro, Ser, Arg, Leu, Lys, Tyr, His, Ile, Met, Phe, Cys, Trp) may be decompose into three groups.

First group: Ser, Arg, Leu. The R-groups of these amino acids contain groups of atoms both with odd and even number of the H-atoms, and therefore in accordance with the rule Ser, Arg and Leu can be both strong and weak. Indeed, Ser, Arg and Leu are coded both by strong and weak roots.

Second group: Ile, Cys, Met. R-group of Ile contains three groups of atoms with odd number of the H-atoms and one group with even number of H-atoms. Therefore in accordance with the rule Ile must be more like to the strong amino acid than to the weak one. Indeed, Ile is coded as we have seen above by three codons and therefore Ile is "semi-strong".

The difference between Ser, Arg, Leu anomaly and Ile anomaly is apparently related with the fact that in the first case group of atoms nearest to Q contain even number of H-atoms while in the second case such group contains odd number of H-atoms.

Further Cys which contains groups of atoms both with odd and even number of H-atoms can be both weak amino acid (E.coli) and

"semistrong" (vertebrates). Met which is coded by the same root as Ile is apparently evolutionally a more late amino acid.

Third group: Pro, Tyr, His, Phe, Trp. Pro possesses a closed R-group and is the exception from all amino acids and correspondently from our rule. R-groups of Tyr, His, Phe and Trp contain besides the group CH_2 also the groups CH, NH, OH. However H-atoms in the rings behave themselves apparently in the collective manner that leads to their equivalence to the groups with even number of H-atom.

At last Lys is an exception from the rule if one consider their R-group in the charged form and is not an exception if one consider their R-group in the neutral form.

Thus, there exists a close connection between the exceptions from the rule proposed above and the deviation from the regularity of the genetic code vocabulary (table 1). So the exceptions listed above only confirm the rule.

The correlation between the amino acids structure and the degree of degeneracy of the genetic code (strong and weak roots) considered is apparently a correlation between the amino acids structure and the structure of the t-RNA and related with the specificity of the hydrogen bonds of the amino acids R-groups.

R e f e r e n c e s

1. Yu.B.Rumer, Doklady Akademii Nauk SSSR, 183, 225 (1968) B.G.Konopel'chenko, Yu.B.Rumer, DAN SSSR (1975).

2. M.Yčas, The Biological code. 1969.