J.Mol.Evol.5,327-332 (1975)

Letter to the Editor

Symmetries of Genetic Code-Doublets

H.-J. DANCKWERTS[1] and D. NEUBERT[2]

[1]Fachbereich Chemie der Universität Regensburg *

[2]Physikalisch-Technische Bundesanstalt, Institut Berlin

Received September 4, 1974; March 28, 1975

Summary. The fact that 64 base triplets code only about 20 essential amino acids implies a strong degeneracy of certain base doublets. It is shown that the set of degenerate base doublets and the set of non degenerate base doublets are highly structured. A mathematical formalism is introduced which allows a systematic description of the consequences of an exchange of bases in a doublet. By this formalism it is shown that the two mentioned sets have in fact the same structure.

Key words: Genetic Code, Degeneracy of ∿, Symmetry of ∿ - Base Triplets - Base Doublets, Group Property of ∿, Transformation of ∿

SYMMETRIES OF GENETIC CODE-DOUBLETS

In the genetic code, the base triplets of the m-RNA determine the amino acids of the protein. The four bases U, C, A and G of m-RNA form 64 different triplets $B_1B_2B_3$ which in turn code only 20 essential amino acids. Apparently, certain different triplets code the same amino acid and it has been noticed [1] that many amino acids are already fully determined by base-doublets B_1B_2. This degeneracy of the genetic code provides automatically for a certain stability of the genetic information against natural and induced mutations.

The purpose of this communication is to give a systematic description of the role of doublets in coding. Especially, we shall deal with the structure of two sets: The set M_1 contains as elements the doublets for which the third base in the triplet has no influence on the coded amino acid, the set M_2 contains those doublets which do not code the amino acid uniquely but require the knowledge of the third base. From the genetic

*Present address: Abt.für Biomedizinische Technik, Medizinische Hochschule Hannover

code [1] the sets M_1 and M_2 are taken to be

(1a) $M_1 = \{AC, CC, CU, CG, UC, GC, GU, GG\}$

(1b) $M_2 = \{CA, AA, AU, AG, GA, UA, UU, UG\}$

Evidently, the doublets of M_1 belong the fourfold degenerate triplets, since each doublet results in four different triplets coding the same amino acid.

In order to find out the structure of M_1 and M_2 we introduce doublet - exchange operators (σ_i, σ_j), where the operators σ_i, i = 1,2,3,4, exchange the four bases A, C, U, G in pairs as definded in Fig.1a.
Biochemically, the σ_i have the following meaning:

- 1 is the unit operator and does not produce any exchange
- α exchanges bases of purine-type to pyrimidine-type
- β exchanges bases which can undergo hydrogen bonds to complementary bases
- γ exchanges a given purine against the other purine and a given pyrimidine against the other pyrimidine.

Mathematically, the σ_i form an Abelian group (Klein's 4-group) i.e.

(2) $1^2 = \alpha^2 = \beta^2 = \gamma^2 = 1$

The doublet-exchange operators (σ_i, σ_j) are now defined by

(3) $(\sigma_i, \sigma_j)\ B_1B_2 = B_1'B_2'$

where σ_i operates on B_1 and σ_j on B_2 and where the doublets B_1B_2 and $B_1'B_2'$ are some combination of two bases from A, C, U, G. With i, j = 1,2,3,4, there are 16 different doublet-exchange operators.

The set M_1 of doublets can be generated by operating with suitable doublet-exchange operators on a special, conveniently chosen doublet, say AC. First, however, we generate a set M_o:

(4) $M_o = \{[(1,1) \cup (\alpha,1) \cup (\alpha,\beta) \cup (\alpha,\gamma)]\ AC\}$

(4a) $= \{AC, CC, CG, CU\}$

and then obtain M_1 from M_o by the operation

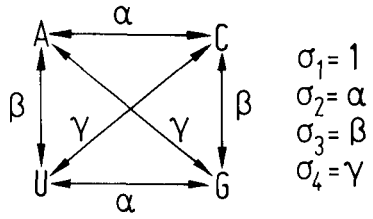

Fig.1a. Definition of the exchange operators σ_i

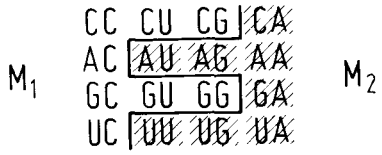

Fig.1b. Structure of the doublet set M

(5) $\quad M_1 = [(1,1) \cup (\beta,1)]\ M_o.$

From M_1 the set M_2 is obtained by the operation

(6) $\quad M_2 = (\alpha,\alpha) M_1.$

The set M of all doublets, $M = M_1 \cup M_2$, has by this construction the structure

(7) $\quad M = [(1,1) \cup (\beta,1)][(1,1) \cup (\alpha,\alpha)]\ M_o$

which can be easily verified by looking at Fig.1b in combination with Fig.1a.

From the complete transformation table (Fig.2) it is seen explicitly, which transformations play a special rôle: The operation $[(1,1) \cup (\beta,1)]$ is the only operation producing M_1 from M_o. Evidently, $(\beta,1)$ is the only operation beside the trivial operation (1,1) under which M_1 is invariant. Naturally, also M_2 is invariant under $(\beta,1)$.

Further, we notice that not only (α,α) generates M_2 from M_1, but also the operation (γ,α). However, if we write $(\alpha,\alpha) = \alpha(1,1)$ and $(\gamma,\alpha) = \alpha(\beta,1)$, we see that (α,α) is the significant operation, the possibility of using also (γ,α) is a consequence only of the invariance of M_1 and M_2 under $(\beta,1)$. Because of Eq.(2), M_1 is generated in turn from M_2 by (α,α): $M_1 = (\alpha,\alpha) M2$.

The rest of the doublet-exchange operators (σ_i, σ_j) transform always half of M_1 into M_2 and half of M_2 into M_1.

By inversion of Eq.(4) and Eq.(5) and making use of Eq.(2) one obtains a rule to determine, whether an element B_1B_2 belongs to M_1 or to M_2: first generate doublets $B'_1B'_2$ by

(8) $\quad [(1,1) \cup (\beta,1)][(1,1) \cup (\alpha,1) \cup (\alpha,\beta) \cup (\alpha,\gamma)] B_1B_2 = B'_1B'_2$

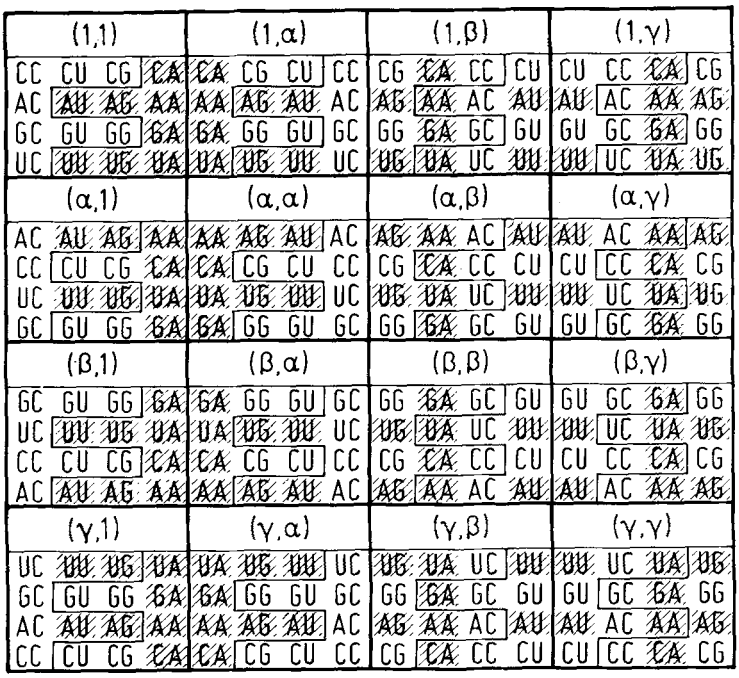

Fig.2. Transformations obtained by the operations (σ_i, σ_j) on M (unshaded doublets belong to M_1, shaded doublets belong to M_2).

If for one of the $B'_1B'_2$ holds

$B'_1B'_2$ = AC, B_1B_2 belongs to M_1

$B'_1B'_2$ = CA, B_1B_2 belongs to M_2.

In summary, we have demonstrated complete symmetry between the sets M_1 and M_2 under transformations affected by the doublet-exchange operators (σ_i, σ_j). Especially, we have shown:

a) M_1 and M_2 are invariant by operating with (β,1) on B_1, but no operation on B_2 leaves M_1 or M_2 invariant. Thus B_2 carries more information than B_1 and B_2 is therefore more important for the stability of M_1 and M_2 than B_1. A change of B_1 with respect to its hydrogen bond property does not change the resulting amino acids if all doublets of either M_1 or M_2 are affected.

Reversing supposition and conclusion, M_1 and M_2 may be defined as those doublet sets of 8 elements which are invariant under the (β,1)-transformation. Then experience shows that M_1 and M_2 are fourfold and less than fourfold degenerate respectively.

b) The operation (α,α) transforms M_1 into M_2. This indicates that purine- and pyrimidine-type bases are distributed in a well determined order onto the bases B_1 and B_2 and that this order determines, whether the third base carries information or not. By Eq.(8) it is seen that for special doublets this different order is simply the reserve order of the bases.

There are also other ways to look at M_1 and M_2. From Fig. 1b one sees e.g., that M_1 is C and G dominated and M_2 is A and U dominated in the following sense: M_1 is determined by B_1 = C or G, M_2 is determined by B_1 = A or U; B_2, however, is decisive also: B_2 = C beats B_1 = A or U, B_2 = A beats B_1 = C or G. Since C and G are equivalent with respect to the hydrogen bond property, one might say therefore, M_1 is hydrogen bond dominated and similarly, M_2 is dominated by the hydrogen bond complement property.

One may try to find simpler criteria. If one defines e.g. a set S_1 consisting of doublets composed of base pairs with three hydrogen bonds, S_1 = {CC, CG, GC, GG}, and a set S_2 consisting of doublets with two hydrogen bonds, S_2 = {UU, UA, AU, AA}, then S_1 is a subset of M_1 and S_2 a subset of M_2. However, in this case there is still a set S_3 = {all other doublets} for which there is no answer to the question, if the doublet specifies an amino acid uniquely or not.

Evidently, the transformation properties of M_1 and M_2 are independent of the arrangement of the genetic code. The table of Fig.1b has been chosen for convenience of representation, but the same results will be obtained by applying doublet-exchange operators onto the table of "best allocations" proposed by Crick [2].

It has been pointed out by Woese et al. [3] that questions related to the explanation of the genetic code "are almost certainly closely allied to the answers to questions regarding the genetic code's evolution". In fact, code generating equations like those given in Eq.(4) to (7) provide a formalism for describing the genetic code's evolution if interpreted as being executed in time. Of course, there are many types of information, e.g. statistical information like that given by Roberts [4], which have to be taken into account in the search for a consistent description of the role of doublets in the genetic code.

It may be noticed, that the number of doublets in M_1 and M_2 is both 8. Only due to this property the given symmetry relations between M_1 and M_2 are possible. Unless it can be proven that the fourfold degeneracy of *exactly 8 doublets* has no biological significance, we suggest to adopt the existence of this fact as a hypothesis in future investigations.

REFERENCES

1. Nirenberg,M.W., Matthaei,J.H. (1961). Proc.N.A.S. 47, 1588
Ochoa,S. (1962). In: Horizons in biochemistry, M.Kasha, B.Pullman eds. New York: Academic Press
Nishimura,S., Jones,S.D., Khorana,H.G. (1965). J.Mol.Biol.13, 302
2. Crick,F.H.C. (1966). Cold Spring Harbor Symposia on Quantitative Biology 31, 1
3. Woese,C.R., Dugre,D.H., Dugre,S.A., Kondo,M., Saxinger,W.C. (1966). Cold Spring Harbor Symposia on Quantitative Biology 31, 723
4. Roberts,R.B. (1962). Proc.N.A.S.48, 1245

D.Neubert, Physikalisch-Technische Bundesanstalt,
Institut Berlin, D-1000 Berlin 10, Abbestr.2-12