ORDER IN THE GENETIC CODE

By C. R. Woese

DEPARTMENT OF MICROBIOLOGY, UNIVERSITY OF ILLINOIS, URBANA

Communicated by T. M. Sonneborn, May 3, 1965

While it is important to know what the genetic code codon assignments are, it is more important to know *why* they are, i.e., to know the mechanisms giving rise to the particular assignments observed. Only when the latter question is answered can we truly claim to begin to understand the genetic code. To date, however, most scientific attention has been turned to the former question. The reasons for this are quite clear. For one, the experimental systems for answering *what* the codon assignments are have been developed and are rapidly yielding a solution, while those for elucidating *why* such codon assignments exist are either not yet discovered or are very far from yielding the sought-for answer. For another reason, the adaptor hypothesis, one of the central dogmas of the coding field today, predicts the answer to the "why" question to be entirely uninteresting and trivial. The prediction here is that the codons are assigned to amino acids merely by "historical accidents," and thus no causal relationship between an amino acid and its codon exists.[1] Another way of putting this is: were the genetic code to evolve again (under the same conditions), the codon assignments would be unrelated to those now observed. I shall not debate the validity of the adaptor hypothesis' prediction at this time, but merely state that, this prediction notwithstanding, the question of "*why* codon assignments" is currently unanswered and therefore remains of prime interest.

The main reason for determining the set of codon assignments, or "codon catalogue," is that a knowledge of it may shed some light upon the mechanisms *behind* these particular assignments. While it is unlikely that this approach will tell exactly what mechanisms are involved here, the approach should be of use in answering the initial question of whether there really *is* an order, a logic, to the codon catalogue, and then perhaps in giving some indication of where to look for the basis of such a logic. At present the codon catalogue is nearly complete, due to the recent techniques developed to determine absolute nucleotide order within codons.[2] Thus we are now in a position to investigate the constraints in the codon catalogue.

Table 1 presents the (ordered) codon catalogue as determined by Nirenberg and co-workers.[2] It is immediately apparent that the catalogue possesses a high degree of order. The most obvious order is in the grouping of codons assigned to the same amino acid. Changing one nucleotide to another—particularly U to C or A to G—in the third position of the triplet in many or possibly all cases leaves the amino acid assignment unchanged. This type of order verifies the general prediction made some time ago on theoretical grounds, that the genetic code has *specific* degeneracies (equivalences of nucleotides) confined to *particular* positions in the codon.[3] At that time it was also conjectured that such an order, logic, could result from either of two mechanisms: (1) it could be a manifestation of an interaction between an amino acid and some nucleotide grouping (related simply to the codon), either in one or more of the translation steps, or possibly existing at some earlier stage in evolution—a "codon-amino acid" logic, or (2) it could manifest

interactions in the translation process, etc., which did *not* involve the amino acid—in loose terms an "intercodon" logic.[3] Alternatively, Sonneborn[4] has recently suggested an ingenious evolutionary mechanism whereby the codon catalogue can be highly ordered, but the order not derive from any sort of molecular interactions. Ordering in this case would result from selection pressure for a code which is the least sensitive to the lethality introduced by mutation (see below). At present we cannot distinguish with certainty among the possibilities here, but right now the task is to scan the codon catalogue for constraints and, having found such, to see whither these take us.

In particular, I should like to focus on the question of related amino acids possessing related codon assignments. A suggestion that this may be the case has always lurked in the earlier data on codon compositions,[3, 7] but the point was not provable at that time, the intracodon orders being unknown. From Table 1 we see more clearly that a correlation between codons assigned to related amino acids may exist,[2] but again one is initially uncertain as to how extensive or significant this is. The point I wish to elaborate below is that this correlation is not only significant but very extensive.

The main problem encountered in correlating codons assigned to related amino acids is in defining the term "related." It will not do to refer to the usual "picture" of a molecule garnered from 2-dimensional formulas or their 3-dimensional equivalents, for this purpose, because these give a picture of only *one* type of interaction of which a molecule is capable (i.e., van der Waals repulsion). "Relatedness" should be defined in terms of a composite of *all* the interactions of which a molecule is capable *and* these in a proportion defined by the context in question. The ideal definition here is clearly unachievable now, and would undoubtedly amount to having answered the main problem posed here anyway. However, it is possible to get a more useful definition of "related" by utilizing a more "functional" approach to the amino acids.

A great deal of data exists to do with the partitioning of amino acids in multiphase systems, particularly to do with their chromatography on paper. In this instance, one would suspect that a *composite* of the interactions of which an amino

TABLE 1

CATALOGUE OF CODON ASSIGNMENTS ACCORDING TO NIRENBERG *et al.*[2]

Phe	*UUU* *UUC*	Ser	*UCU* *UCC*	Cys	*UGU* *UGC*	Tyr	*UAU* *UAC*
Leu	UUA *UUG*	Ser	UCA *UCG*	?	*UGA* UGG	Gln	*UAA* *UAG*
Leu	*CUU* *CUC*	Pro	*CCU* *CCC*	Arg	CGU *CGC*	His	*CAU* *CAC*
Leu	CUA *CUG*	Pro	*CCA* CCG	Arg	*CGA* CGG	Gln	*CAA* *CAG*
Ilu	*AUU* *AUC*	Thr	*ACU* *ACC*	Ser	*AGU* *AGC*	Asn	*AAU* *AAC*
Met	AUA *AUG*	Thr	*ACA* *ACG*	?	*AGA* AGG	Lys	*AAA* *AAG*
Val	*GUU* GUC	Ala	*GCU* *GCC*	Gly	*GGU* GGC	Asp	*GAU* *GAC*
Val	GUA GUG	Ala	GCA GCG	Gly	GGA GGG	Glu	*GAA* GAG

Italicized codons only have been tested directly. The remaining amino acid assignments are inferred from earlier codon *composition* data plus amino acid replacement data in some cases.

acid is capable is more likely to be manifested. And, by the right choice of system, a situation resembling the biological case *might* be approached. (In a crude way one can speak of "partitioning" amino acids between an aqueous phase and an "enzyme site phase," for example.)

One correlation is immediately apparent upon examining R_F values for amino acids. In practically all solvent systems phe, leu, and ilu are nearly indistinguishable. This of course is precisely what would be predicted from the codon catalogue (Table 1).

In this manner, then, let us define families of amino acids from the codon assignments of Table 1, using the rule: hold the last two nucleotides in the codon constant and vary the first nucleotide. The following families of amino acid then emerge:

1*a*	Phe,	leu,	ilu,	val
1*b*	Leu,	met,	val	
2	Ser,	pro,	thr,	ala
3*a*	Cys,	arg,	ser,	gly
3*b*	?,	arg,	?,	gly
4*a*	Tyr,	his,	asn,	asp
4*b*	Gln,	lys,	glu	

Table 2 gives the R_F values for amino acids in several solvent systems. In a pyridine-H_2O system, for example, an R_F of 0.67 ±0.02 characterizes a group of amino acids comprising phe, leu, ilu, met, val, and tyr, i.e., groups 1*a* and 1*b* plus tyr. A second group of amino acids centers about an R_F of 0.53 ± 0.04 in this solvent system, i.e., those amino acids of group 2 above. Although group 2 does not travel as a unit in *all* solvent systems, one finds a reasonable grouping of these four also in n-butanol:formic acid:water. Beyond this point the correlations, for groups 3 and 4, are not so clear, but they nevertheless do exist. With the occasional exception of the basic amino acids in some solvents, and tyr, groups 3 and 4 constitute the amino acids with lowest R_F's. For example, in the m-cresol system one

TABLE 2

R_F VALUES FOR AMINO ACIDS, CHROMATOGRAMMED ON PAPER IN VARIOUS SOLVENT SYSTEMS

Pyridine:H_2O 65:35 (ref. 5)	γ-Picoline:H_2O 60:40 (ref. 5)	1-Butanol: formic acid:H_2O = 12:1:1 (ref. 5)	m-Cresol cupron (ref. 6)	Phenol HCN (ref. 6)	Phenol:H_2O pH 5 (ref. 6)	Collidine: lutidine 1:1 (ref. 6)
Leu 68	Phe 67	Leu 81	Phe 82	Pro 91	Pro 86	Tyr 74
Phe 66	Try 64	Ilu 80	Try 76	Phe 89	Phe 84	Phe 67
Ilu 66	Tyr 63	Phe 75	Pro 73	Ilu 86	Leu 84	Try 66
Met 66	Ilu 62	Val 75	Leu 73	Leu 85	Ilu 84	Leu 65
Tyr 66	Leu 60	Met 72	Ilu 70	Try 83	Try 80	Ilu 62
Val 65	Met 59	Pro 64	Met 64	Met 80	Val 78	Met 61
Try 63	Val 55	Ala 61	Val 52	Val 77	Tyr 66	Val 53
Pro 56	Pro 43	Tyr 61	Tyr 35	His 69	His 64	Thr 43
Thr 56	Ala 40	Ser 51	His 34	Tyr 64	Gln 59	Pro 41
Ala 54	His 31	Thr 50	Ala 23	Arg 59	Ala 58	Ala 41
Ser 51	Gly 27	Gly 46	Thr 14	Ala 54	Arg 56	Ser 37
Glu 48	Asp 22	Glu 42	Gly 12	Thr 50	Thr 49	His 34
Asp 43	Arg 20	Asp 42	Ser 08	Lys 46	Lys 48	Gly 33
His 43	Lys 14	Arg 30	Arg 07	Gly 40	Asn 44	Gln 32
Gly 41	Glu 14	Lys 27	Cys 04	Ser 36	Gly 38	Asn 29
Arg 31		His 25	Lys 04	Cys 30	Ser 36	Glu 26
			Glu 01	Glu 25	Glu 31	Asp 24
			Asp 01	Asp 15	Asp 18	Arg 14
						Lys 14

finds the members of group 3 (cys, arg, ser, and gly) traveling as an unbroken group between R_F's of 0.04 and 0.12. (Except for arg—R_F = 0.59—group 3 travels together in a phenol–HCN system also—R_F's between 0.30 and 0.40.)

No conditions exist where the members of groups 4*a* or 4*b* travel as an uninterrupted group. However, many instances of subgroups of this class traveling together exist: (1) tyr and his are paired in m-cresol (0.34 vs. 0.35 R_F), and in phenol pH 5 (0.66 vs. 0.64); (2) lys and gln are paired in phenol (0.56 vs. 0.62),[5] while lys and glu are paired in γ-picoline-H_2O (0.14 vs. 0.14); (3) in collidine-lutidine mixtures asp, glu, asn, and gln move with R_F's of 0.24, 0.26, 0.29, and 0.32, respectively.

Next, if one looks for correlations, with codon assignment by grouping the amino acids according to the *third* position in the codon (rather than the first), the result is rather unspectacular, mainly because only *pairs* of amino acids result by this classification—phe-leu, ilu-met, tyr-gln, his-gln, asn-lys, and asp-glu. Table 3 shows the solvents in which these pairings correlate with R_F's.

Considering the third way of classing amino acids, i.e., changing only the *middle* nucleotide in the codon, there are probably no correlations with chromatography—with the possible exception of leu, pro, and arg traveling close together in some phenol systems.

It is interesting to note particularly the chromatographic behavior of serine. In some solvents, this amino acid travels with the group 2 amino acids, ala, thr, and pro, while in other solvents it is closely paired with gly (group 3). Serine is one of the few amino acids possessing unrelated codons, which classify it as *both* a group 2 *and* a group 3 amino acid.

I think no statistical analysis is required here to see that the codon assignments do indeed correlate with the chromatographic behavior of the amino acids. We are now faced with the question of whether any further correlations exist and with what all these correlations mean. I do not intend to go into these matters to any great extent now, for the amount of speculation involved might prove somewhat intoxicating. A few remarks will suffice for the present. There are some obvious correlations between codons for amino acids with related *structures* which involve the second position in the codon. Examples here are ser(UCU)-cys(UGU), thr(ACU)-ser (AGU), and ala(GCU)-gly(GGU) pairs. Also arg(CGU)-his(CAU) is of this class, but it is not so obvious until molecular models are constructed. Another very interesting correlation is worth noting at this time. The codon catalogue groups together the amino acids $\text{cys}^{\text{UGU}}_{\text{UGC}}$, $\text{arg}^{\text{CGU}}_{\text{CGC}}$, $\text{tyr}^{\text{UAU}}_{\text{UAC}}$, and $\text{his}^{\text{CAU}}_{\text{CAC}}$. The possession of π electrons and lone pairs links tyr, arg, and his. The S atom in cys likewise possesses lone pair electrons, and in terms of polarizability at least, its outer electrons resemble π electrons.

The main question is whether these amino acid-codon correlations mean a "codon-amino acid" logic to the code, i.e., amino acid-nucleotide interactions behind the

TABLE 3

R_F VALUES FOR AMINO ACIDS GROUPED BY CODON ASSIGNMENT OF POSITION III

Phe-leu	In most solvents, e.g., pyridine: H_2O	0.66 vs. 0.68
Ilu-met	In most solvents, e.g., o-cresol-cupron[6]	0.58 vs. 0.58
Tyr-gln	Phenol	0.63 vs. 0.62
His-gln	In pyridines, e.g., collidine: lutidine	0.34 vs. 0.32
Asn-lys	None good—phenol	0.42 vs. 0.56
Asp-glu	In many solvents, e.g., collidine: 2,4 lutidine[6]	0.24 vs. 0.26

codon assignments. Although the suggestion is certainly strong here, the alternative presented by Sonneborn has not yet been eliminated. To recapitulate the Sonneborn argument, selection pressure will bring about a code in which mutations produce a minimal amount of lethality. This means that (1) "nonsense" codons will be eliminated or minimized (some codons may have to serve a punctuation function), (2) the fraction of mutations which involve a codon change *not* leading to an amino acid change will be optimized, as will (3) the fraction of mutations involving codon changes leading to replacement of a given amino acid by a "functionally related" amino acid. I feel that while the end result of the Sonneborn scheme is plausible, there is no plausible mechanism by which the code could have evolved to this degree of perfection (without some underlying amino acid-nucleotide interactions). A scheme such as Sonneborn's would involve countless evolutionary trials and errors, and I feel that the possibilities for evolving into "blind alleys" (forms of the code having a far lower degree of order) so far outnumber the possibilities for evolving an optimal code (the one observed) that the latter could never have evolved in this way. However, it must be admitted that without a proper analysis of the Sonneborn model—such as a computer study—this counterargument remains feeble. Thus the question is not completely resolved at this time.

One thing is clear from the above, however. Extensions of amino acid chromatography in which one of the phases resembles more closely nucleic acid components may well turn out to be useful in elucidating possible amino acid-nucleotide interactions.

[1] Crick, F. H. C., in *Progress in Nucleic Acids* (New York: Academic Press, 1963), vol. 1, p. 163.

[2] Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal, these PROCEEDINGS, **53**, 1161 (1965).

[3] Woese, C., *Nature*, **194**, 1114 (1962); Woese, C., *ICSU Reviews*, **5**, 210 (1963).

[4] Sonneborn, T. M., in *Evolving Genes and Proteins*, Rutgers Symposium 1964, ed. H. Vogel (New York: Academic Press, 1965).

[5] Block, R., E. Durram, and G. Zweig, *Paper Chromatography and Paper Electrophoresis* (New York: Academic Press, 1955).

[6] Lederer, E., and M. Lederer, *Chromatography* (New York: Elsevier, 1957), 2nd ed.

[7] Roberts, R. B., these PROCEEDINGS, **48**, 897 (1962).

---

*PENICILLIN: ITS BASIC SITE OF ACTION AS AN INHIBITOR OF A PEPTIDE CROSS-LINKING REACTION IN CELL WALL MUCOPEPTIDE SYNTHESIS**

BY EDMUND M. WISE, JR.,† AND JAMES T. PARK

DEPARTMENT OF MICROBIOLOGY, TUFTS UNIVERSITY SCHOOL OF MEDICINE, BOSTON

Communicated by Rollin D. Hotchkiss, May 17, 1965

It has been known for some time that penicillin interferes with bacterial cell wall mucopeptide synthesis.[1-3] When a penicillin-sensitive cell grows in the presence of penicillin, the integrity of the cell wall is lost and the cell either ruptures or its membrane is damaged beyond repair.[4, 5]

Although this general picture of the over-all effect of penicillin on growing cells