UGA: A Third Nonsense Triplet in the Genetic Code

by
S. BRENNER
L. BARNETT
E. R. KATZ
F. H. C. CRICK

M.R.C. Laboratory of Molecular Biology,
Hills Road,
Cambridge

Two base triplets of the genetic code are known not to represent any amino-acid. It now appears that, in *Escherichia coli*, the UGA triplet of the bases uracil, guanine and adenine does not code for an amino-acid and is therefore also a "nonsense triplet".

Most of the sixty-four triplets of the genetic code[1] have been allocated to one or other of the twenty amino-acids. The two known nonsense triplets (UAA, *ochre* and UAG, *amber*) are believed to signal the termination of the polypeptide chain. The only other triplet so far unallocated is UGA, for which binding experiments give uncertain or negative results.

In this article we show that UGA is "unacceptable" in our system (*Escherichia coli* infected with bacteriophage *T4*) and present suggestive evidence that it is nonsense; that is, that it does not stand for any amino-acid. Theoretical arguments make it likely that there is no transfer RNA (*t*RNA) to recognize it. The reason for this apparent absence of function is not yet known. Neither is it known whether UGA is nonsense in other organisms.

Evidence that UGA may be nonsense in *E. coli* has also been presented by Garen *et al.*[2]. They investigated the reversion of *amber* and *ochre* mutants in the alkaline phosphatase gene of *E. coli*. *Amber* mutants (UAG) reverted, as expected, to seven different amino-acids including tryptophan which is coded by UGG. *Ochre* mutants (UAA) reverted to six of these amino-acids, but not to tryptophan. This negative result makes it unlikely that UGA stands for tryptophan (see also Sarabhai and Brenner[3]) and suggests that it might be a nonsense codon.

Mutant X655 contains UGA. Much of our genetic work has been concerned with the left-hand end of the B cistron of the *r*II region of bacteriophage *T4*. We have made extensive and detailed investigations of this region which are being reported elsewhere[4]. The mutant *X655* occurs in the middle of this region. In brief our proof that *X655* contains the triplet UGA consists in converting it to an *ochre* (UAA), using mutagens the behaviour of which is already known.

X655 was induced from wild type by 2-aminopurine, and identical mutants are also found after treatment of wild type phage with hydroxylamine. This shows that it differs from an acceptable triplet by a G–C to A–T base pair change in the DNA. It is not suppressed by any *amber* or *ochre* suppressor (Table 1) and is therefore neither UAG nor UAA. The reversion properties of *X655* are shown in Table 2. It is strongly induced to revert to r^+ by 2-aminopurine, as is expected, but there is no induction to r^+ by hydroxylamine. Thus the triplet in the DNA either contains no G–C pairs or, if it does contain one, it is connected to another unacceptable triplet by a G–C to A–T transition.

The triplet is in fact connected to UAA by a transition, because *X655* can be converted to an *ochre* and this change is induced by 2-aminopurine (Table 2). The nature of the transition is more precisely specified by the finding that the conversion to an *ochre* is induced by hydroxylamine and that the *ochre* triplet produced does not require any replication for expression. Using a previous argument[5] this result suggests that the change arises from a G→A change in the messenger RNA. Because *X655* is not an *amber*, this proves that it contains the triplet UGA. To confirm that an *amber* at the site of *X655* would be suppressed by *amber* suppressors the *X655 ochre* has been converted to an *amber* by mutation and its properties tested (Table 1).

Table 1. SUPPRESSION PROPERTIES OF *X655* AND ITS DERIVATIVES

			Amber suppressors			*Ochre* suppressors		
Mutant	Triplet	*su*⁻	su_I^+	su_{II}^+	su_{III}^+	su_3^+	su_5^+	su_7^+
X655	UGA	0	0	0	0	0	0	0
X655 ochre	UAA	0	0	0	0	+	+	+
X655 amber	UAG	0	+	+	+	+	+	+

Phage stocks were plated on the following strains*: *su*⁻, CA244; su_I^+, CA266; su_{II}^+, CA180; su_{III}^+, CA265; su_3^+, CA165; su_5^+, CA167; and su_7^+, CA248.

Table 2. REVERSION OF *X655*

	Reversion index (in units of 10^{-7})			
	Spontaneous	2-Aminopurine	Hydroxylamine direct	Hydroxylamine after growth
to r^+	4	312	5	6
to *ochre*	4	51	1,090	533

X655 was treated with 2-aminopurine and hydroxylamine as previously described[4,5]. Total phage was assayed on *E. coli B* and r^+ revertants on CA244 (*su*⁻). *Ochre* revertants were selected on CA248 (su_7^+) and distinguished from r^+ revertants by picking and stabbing about 300 plaques into CA248 and CA244.

Other occurrences of UGA. In three cases we have been able to produce the triplet UGA by selected phase shifts in our region. When (+ −) phase shifts are made over the first part of the B cistron, the two phase shift mutants frequently do not suppress each other. We have shown[4] that these barriers to mutual suppression are due to the generation of unacceptable triplets in the shifted frame. One of these barriers, b_9, has been identified as an *amber* and two others, b_3 and b_4, as *ochres*. Three barriers, b_2, b_5 and b_6, have now been identified as UGA by their base-analogue induced reversion to *ochres*. In each case the identification has been checked by converting the *ochre* to an *amber* at the same site.

Tryptophan is represented by the single codon UGG. It would therefore be expected to mutate by transitions to both UAG (*amber*) and UGA, and thus in such cases *amber* and UGA mutants should occur in close pairs. The *amber* mutant, *HB74*, which maps close to *X655*,

is an example of this. Genetic crosses between it, *X655*, and the *ochre* and *amber* derived from *X655*, show that *HB74* maps identically to the *amber* derived from *X655*, as expected (Table 3).

Table 3. RECOMBINATION BETWEEN VARIOUS MUTANTS

	X655	*X655 ochre*	*X655 amber*	*HB74*	Triplet
X655	0				UGA
X655 ochre	0	0			UAA
X655 amber	+	0	0		UAG
HB74	+	0	0	0	UAG

The phages were crossed in *E. coli B* and the complexes irradiated with ultra-violet light to stimulate recombination (see ref. 4). In the Table, 0 means that r^+ recombinants were not significantly above the reversion rate, which was between 2 and 9×10^{-7}; in those experiments where positive results were obtained (+), the frequency was between 2 and 6×10^{-6}.

So far we have found the expected pairs consisting of UGA and an *amber* in two other cases. In the A cistron, a mutant *X665** is found with the *amber* mutant *N97*, and in the B cistron, *N65* is paired with the *amber* mutant *X237*. Both *N97* and *X237* are likely to have arisen from UGG (tryptophan) which is confirmed by the finding that they respond only poorly to the *amber* suppressor su_{II}^+ which inserts glutamine[5]. Both *X665* and *N65* have been converted into *ochre* mutants, showing that they contain the triplet UGA. These *ochres* have also been converted to *ambers* at the same site. Mapping investigations, analogous to those in Table 3, are consistent with these allocations.

Unacceptability of UGA. There is very good evidence that the amino-acid sequence coded by the first part of the B cistron is not critical for the function of the gene[4]. It can be replaced by varying lengths of the A cistron using deletions that join the two genes. Moreover, an extensive (− +) frame shift can be made without noticeable effect on the function. Of the fifteen known base-analogue mutants in the region, thirteen are either *ochres* or *ambers*; one, *HD263*, is temperature sensitive and *X655* is UGA. The extreme bias towards *amber* and *ochre* chain-terminating mutants confirms the dispensability of the region[4]. These results make it unlikely that the unacceptability of UGA in *X655* and the three barriers results from the insertion of an amino-acid, and strongly suggest that it is nonsense.

In addition, the UGA mutant *X665* in the A cistron has been combined with the deletion *r1589* and has been found to remove the B activity of this phage. This is the test for nonsense originally used by Benzer and Champe[6].

In all these cases, however, it could be argued that UGA might code cysteine, especially as the two known triplets for cysteine are UGU and UGC. If the B protein already contained a cysteine essential for its function the effect of UGA elsewhere might be to produce an S–S bridge between the cysteine inserted by UGA and the (hypothetical) essential one, and thus inactivate the protein. Nevertheless we regard this as unlikely for two reasons, one genetic and one chemical.

The genetic evidence concerns the anomalous minutes produced by certain (+ +) combinations in the B cistron[4]. In some regions of the first part of the B cistron combinations of two (+) phase shift mutants are able to grow to some extent on the restrictive host, *E. coli K12*. The plaques produced are minute, however, showing that the wild type phenotype is very far from being completely restored. A detailed analysis of one set of these combinations showed that minutes are obtained only from pairs of (+) mutants which straddle barrier b_6. The presence of the barrier is obligatory because, if it is removed by mutation, the (+ +) doubles are unable to grow at all on *E. coli K12*. The minutes are clearly due to a phase error of one sort or another and the phase error is dependent on the barrier b_6 which we now know to be UGA. This result shows that UGA cannot be associated with any normal amino-acid reading and points strongly to the conclusion that it is nonsense.

The chemical reason for UGA not coding for cysteine comes from the work of Khorana *et al.*[7]. They have shown that poly $(UGA)_n$ when used as a messenger in a cell-free system derived from *E. coli* induces the production of poly methionine (corresponding to AUG) and also poly aspartic acid (corresponding to GAU). No other amino-acid appears to be incorporated. In particular, no poly cysteine was found. For various reasons this evidence is not completely decisive, but it at least makes it unlikely that UGA is cysteine.

Function of UGA. It might be thought that the sequence containing UGA was nonsense because it was the signal for the beginning or ending of a gene (or operon). In other words, that it produced its effect during the synthesis of the messenger RNA on the DNA template of the gene. This explanation is highly unlikely because the effects of UGA depend on it being read in phase. The phenotypic effect of *X655* can be removed when the mutant is placed in a (− +) shifted frame[4], and the barriers b_2, b_5 and b_6 are of course produced by phase shifts. That is, the base sequence UGA actually occurs at these places in the wild type messenger RNA but in such a way that it is out of phase when the message is read correctly. Because we have no reason to suspect that RNA polymerase synthesizes messenger RNA in groups of three bases at a time these results imply that the phenotypic effects of UGA must occur during protein synthesis.

It thus seems unlikely that UGA codes for any amino-acid, and in particular it does not appear to code for either cysteine (UGU and UGC) or tryptophan (UGG). The wobble theory of codon–anticodon interaction developed by one of us[8] makes the prediction that because of a wobble in the recognition mechanism at the third place of the codon no *t*RNA molecule can recognize XYA alone without at the same time recognizing either XYG or both XYU and XYC. Such theoretical arguments cannot be considered conclusive, but they certainly suggest that UGA is a triplet for which no *t*RNA exists. For this reason we think it unlikely that UGA produces the efficient termination of the polypeptide chain, but more direct evidence will be needed to establish this point.

Conclusion. We have thus established that in the phage-infected cell UGA is certainly "unacceptable" in the *r*II cistrons, although it remains to be seen whether this is true for other species. We have produced reasons why it is unlikely to code for any amino-acid. We are confident that there must be weighty reasons if even a single triplet is not used in the genetic code, because otherwise natural selection would have certainly allocated it to an amino-acid. At the moment we are inclined to believe that UGA may be necessary as a "space" to separate genes in a polycistronic message. It is possible to make a plausible theory for *E. coli* along these lines, but we prefer to leave the discussion of this until we have more experimental evidence to support it. This we are at present attempting to obtain.

We thank Drs. A. Garen, H. G. Khorana and A. Sarabhai for interesting discussions and for showing us their papers in advance of publication. One of the authors (E. R. K.) is a holder of a United States Churchill Foundation scholarship.

Received December 22, 1966.

* This is not a misprint for *X655*.

[1] For the structure of the genetic code and the evidence for nonsense triplets see the papers in the Cold Spring Harbor Symposium XXXI on "The Genetic Code", 1966 (in the press).
[2] Weigert, M. G., Lanka, E., and Garen, A., *J. Mol. Biol.*, **23**, 391 (1967).
[3] Sarabhai, A., and Brenner, S., in preparation.
[4] Barnett, L., Brenner, S., Crick, F. H. C., Shulman, R. G., and Watts-Tobin, R. J., *Phil. Trans. Roy. Soc.* (in the press).
[5] Brenner, S., Stretton, A. O. W., and Kaplan, S., *Nature*, **206**, 994 (1965).
[6] Benzer, S., and Champe, S. P., *Proc. U.S. Nat. Acad. Sci.*, **48**, 1114 (1962).
[7] Morgan, A. R., Wells, R. D., and Khorana, H. G., *Proc. U.S. Nat. Acad. Sci.* (in the press).
[8] Crick, F. H. C., *J. Mol. Biol.*, **19**, 548 (1966).
[9] Brenner, S., and Beckwith, J. R., *J. Mol. Biol.*, **13**, 629 (1965).