

"Two out of three": An alternative method for codon reading

(codon-anticodon recognition/translational fidelity/wobbling/organization of the genetic code)

ULF LAGERKVIST

Department of Medical Biochemistry, University of Gothenburg, S-400 33 Gothenburg, Sweden

Communicated by George Klein, February 6, 1978

ABSTRACT An alternative method for codon reading, whereby only the first two codon nucleotides are recognized by the anticodon, is discussed and the experimental evidence for this "two out of three" reading method is reviewed. Misreading of codons by the "two out of three" method could pose a significant threat to the fidelity of protein synthesis unless the genetic code is organized in such a way as to prevent this method from being used when it might compromise translational fidelity. Inspection of the genetic code shows that it is arranged in such a way that the "two out of three" reading method can be used without translational errors.

The genetic code is a universal, highly degenerate, three-letter code in which the first two positions of the codon are read by the anticodon strictly according to the rules of classic base pairing. The third position in the codon, however, introduces complications. Thus, there is a discrepancy between the large number of codons in a degenerate code and the limited number of anticodons available for the reading of these codons. It was to bridge this gap that Crick in 1966 put forward his ingenious wobble hypothesis (1). In this classic paper, Crick proposed that the nucleotide at the 5' end of the anticodon, the wobble position, can in fact make interactions with the third position of the codon that are not allowed by the rules of classic base pairing: U in the wobble position can recognize G in the third position of the codon; I can recognize both U and A; and G can recognize U. Nevertheless, there are a number of interactions that are not allowed even in the wobble position: U does not recognize U and C; C does not recognize U, C, or A; I does not recognize G; and G does not recognize A and G.

In his paper, Crick gave a number of structural reasons for these restrictions, and they are also supported by the results of ribosomal binding experiments. The results of such experiments have on the whole been as predicted by the wobble hypothesis with the possible exception that 5-oxyacetic acid uridine (U*) in the wobble position seems to recognize U in the third codon position (2). One can then ask whether the same restrictions apply to protein synthesis *in vivo*. An obvious answer would be that the restrictions of the wobble hypothesis must apply in all situations in which there otherwise would be a mistake in the synthesized protein. For instance, in the phenylalanine/leucine codon group (UUU, UUC, UUA, and UUG), an anticodon with G in the wobble position cannot recognize the codons UUA and UUG because that would lead to the mistaken introduction of phenylalanine instead of leucine.

We often think of the genetic code as being made up of groups of four codons that have their first two nucleotides in common, with the variation in the third position. When all four codons in such a group code for the same amino acid, this will be referred to as a codon family. Do the rules of the wobble hypothesis apply as strictly here as in the phenylalanine/leucine

case, in spite of the fact that it makes no difference to translational fidelity how the third position of the codon is read because the first two codon nucleotides are enough to specify the amino acid? To answer this question, the codon-anticodon recognition in the valine codon family has been investigated by using an *in vitro* system from *Escherichia coli* programmed with MS 2-RNA, in which 80-90% of the protein synthesized is made up of MS 2 coat protein. The primary structure of the coat protein cistron in MS 2-RNA has been determined and can be compared to the known amino acid sequence of the coat protein (3).

By measuring the incorporation of labeled valine from valyl-tRNAs with different anticodons into peptide positions corresponding to the four valine codons (GUU, GUC, GUA, and GUG), the ability of an anticodon to read a certain codon could be ascertained. The results of these experiments (4) can be summarized by saying that, under the conditions of *in vitro* protein synthesis, the valine anticodons U*AC, GAC, and IAC can each recognize all four valine codons. In other words, the third position in the valine codons did not seem to have any absolute discriminatory function. Assuming that the restrictions that the wobble hypothesis imposes on the reading of the third codon position are valid, our results would admit of only one conclusion: Under the conditions of *in vitro* protein synthesis, a codon can be read by recognition of only its first two nucleotides, the third position of the codon being disregarded. The existence of an alternative reading method has been postulated and in the following is referred to as reading "two out of three."

"Two out of three" reading cannot be infrequent *in vitro*, as judged by the outcome of experiments in which a valine anticodon, which can only read the codon by the "two out of three" method, competed against another anticodon that could read all three positions of the codon. In this competitive situation the anticodon that read all three codon nucleotides according to the rules of classic base pairing was only an order of magnitude more efficient than the anticodon, which read only two out of three.[†] One may then ask whether reading according to the "two out of three" method is limited to the valine codon family. In my opinion it is more likely that this type of reading can take place in many codon families and perhaps in all. In fact, there is some evidence that it can be used also in the glycine family. Ten years ago, Bergquist *et al.* (5) published data indicating that four different isoaccepting glycine tRNAs could recognize all glycine codons in the R 17 coat protein cistron. Viewed in the light of the information that we have today, these findings seem to indicate that the glycine codons can be read *in vitro* according to the "two out of three" method. The all important question is, of course, whether this method can also be used *in vivo*. Two lines of argument come to mind in this connection.

If the third codon position had an absolute discriminatory

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

[†] Unpublished data.

power *in vivo*, one would not expect amber (UAG) suppressors to be able to suppress ochre (UAA) mutations because this would involve the recognition of A in the third codon position by C in the wobble position. Nevertheless, there is evidence that such suppression takes place although the efficiency is only a few percent of that of the amber suppression (6, 7). This apparent contradiction can be resolved by assuming that the amber suppressor can read the ochre codon, by using the "two out of three" method, with a probability that is sufficient to sustain the low suppression observed.

The other line of argument comes from a study of glycine missense suppressors (8–10). According to the results of binding experiments, only one of the tRNA^{Gly} isoacceptors in *E. coli* recognizes the codon GGA. This tRNA^{Gly} is the product of a single gene on the *E. coli* chromosome that can be recovered from mutants in two different altered forms, both of which are missense suppressors that insert glycine in response to the arginine codon AGA. The loss of the wild-type function should leave the cell without a tRNA^{Gly} able to read the codon GGA, assuming that reading always requires the recognition of the third codon nucleotide and that the rules of the wobble hypothesis are strictly adhered to *in vivo*. Nevertheless, one of the mutants, a double mutant, grows fairly well in a minimal medium but fails to grow when transferred to a rich medium. The other mutant is extremely fastidious and difficult to grow but the main point is that it can be grown, albeit with the greatest difficulty. These results could be explained if we assume that other tRNA^{Gly} isoacceptors in the cell can read the codon GGA by using the "two out of three" method and in this way to some extent make up for the absence of an anticodon able to read the whole codon.

The data discussed so far indicate that it is possible for the translational machinery of the cell to read codons by the "two out of three" method, disregarding the third nucleotide. This is certainly so under the conditions of protein synthesis *in vitro* and possibly also *in vivo*. On the other hand, we have no way of predicting what the probability of reading according to the "two out of three" method actually is *in vivo*. In this context it is important to emphasize that reading by the "two out of three" method would normally be contained by competition with tRNAs having anticodons that could recognize all three codon nucleotides. In the examples cited above, which concern nonsense and missense suppressors, this would not be the case because there would be no competing tRNAs. Nevertheless, let us assume that, at least on some codons, reading by the "two out of three" method can occur *in vivo* with a frequency that is not negligible. If this is so, the cell would be faced with a certain probability of misreading which could mean a threat to translational fidelity if the "two out of three" method were to be used inappropriately—i.e., anywhere outside the codon families, where it could lead to mistakes in protein synthesis. This threat to fidelity would obviously have to be contained and, in what follows, possible methods that the cell might use to achieve this will be considered.

The discussion will take, as its starting point, the virtually self-evident prediction that the probability of reading a codon by the "two out of three" method must be a function of the strength of the interaction between the anticodon and the first two codon nucleotides. It also will be assumed that, in codon-anticodon recognition, an interaction of the G-C type, involving three hydrogen bonds, is stronger than an A-U interaction with only two bonds. This assumption could perhaps be disputed on the basis of data from anticodon-anticodon interactions between tRNAs in solution (11). It must be kept in mind, however, that these experiments have little resemblance to the actual

reading situation in which codon and anticodon interact with each other on the surface of the ribosome in a highly structured environment. In any case, in the absence of any direct evidence to the contrary, it seems more likely that the relative strength of the different types of nucleotide interactions involved in codon-anticodon recognition is similar to what is found in DNA and in helical regions of RNA.

In Fig. 1 the codons in the genetic code have been divided in the usual way into groups of four: the first two positions are the same for all four codons and the third position can be occupied by any of the nucleotides U, C, A, and G. Nucleotides in the first two positions, which give strong G-C type interactions with the anticodon, are shown as bold letters, and codon families are boxed in by heavy lines.

When we consider the general organization of the genetic code as revealed in Fig. 1, certain features are striking. Bear in mind that reading "two out of three" might constitute a threat to translational fidelity and that the probability of such misreading is a function of the strength of the interactions between the anticodon and the first two codon positions. Two strong G-C type interactions would give a maximal probability of misreading whereas two weak A-U type interactions would represent a minimal probability. One would then expect the maximal probability codons to be confined to families in which a misreading would have no effect on translational fidelity because the same amino acid would be introduced in any case. By the same argument, the minimal probability codons should all be found outside the families. From Fig. 1 it is apparent that this simple rule of thumb is strictly adhered to throughout the code.

But what about "mixed" codons which make one strong and one weak interaction in the first two positions? Let us first consider some structural characteristics of the code. The anticodons corresponding to codons in the left half of the codon square all have either two purines in the positions interacting with the two first codon nucleotides, or else have a purine and a pyrimidine in which case the purine is always in the middle of the anticodon. For those in the right half of the square, on the other hand, the same anticodon positions are occupied either by two pyrimidines or by a purine and a pyrimidine with the pyrimidine in the middle. From Fig. 1 it is immediately obvious that the distribution of the mixed codons with respect to the codon families is not random. In the left half of the codon square all mixed codons appear in families while in the right half they are all outside the families. It is tempting to speculate that this nonrandom distribution, which has been noted and commented on before by L. E. Orgel (personal communication), has a biological function. Let us assume, without trying to present any structural arguments for this assumption, that interactions between mixed codons and their anticodons are stronger in the left half of the codon square than in the right half. From this assumption it would follow that mixed codons in the left half would represent a greater probability of "two out of three" misreading than mixed codons in the right half. We would consequently expect to find the mixed codons in the left half of the codon square confined to the families, as indeed they are. One of our test cases, the valine codon family, is an example of this and we have evidence[†] that the probability of reading "two out of three" *in vitro* is much greater in this family than in the two lysine codons, which make only A-U type interactions in the first two positions and belong to the codons in the right half of the codon square.

Finally, it should perhaps be emphasized that this discussion has dealt with codon-anticodon interaction only in the strictest sense of this concept. In a wider perspective, it is obvious that

Strong mixed codon interactions		Weak mixed codon interactions	
U U U Phe	U C U Ser	U A U Tyr	U G U Cys
U U C "	U C C "	U A C "	U G C "
U U A Leu	U C A "	U A A Och	U G A Umb
U U G "	U C G "	U A G Amb	U G G Trp
C U U Leu	C C U Pro	C A U His	C G U Arg
C U C "	C C C "	C A C "	C G C "
C U A "	C C A "	C A A Gln	C G A "
C U G "	C C G "	C A G "	C G G "
A U U Ile	A C U Thr	A A U Asn	A G U Ser
A U C "	A C C "	A A C "	A G C "
A U A "	A C A "	A A A Lys	A G A Arg
A U G Met	A C G "	A A G "	A G G "
G U U Val	G C U Ala	G A U Asp	G G U Gly
G U C "	G C C "	G A C "	G G C "
G U A "	G C A "	G A A Glu	G G A "
G U G "	G C G "	G A G "	G G G "

FIG. 1. The genetic code.

other factors such as the conformation of the whole tRNA molecule (i.e., the structural context in which the anticodon presents itself to the codon) can be of great importance for the specificity of codon-anticodon recognition (12). A good example of this is the tryptophan nonsense suppressor with a mutation outside the anticodon (G25 to A25) which recognizes the terminator codon UGA without any corresponding change in the anticodon sequence (13, 14). This could be explained if one assumes that the mutation leads to a new tRNA conformation with an enhanced ability to read "two out of three." The suppressor would then be able to read the nonsense codon UGA by the "two out of three" method but would presumably be prevented from reading the cysteine codons UGU and UGC

because of competition with the cysteine tRNAs. This prediction has, in fact, been verified experimentally in a recent paper by Buckingham and Kurland (15).

I would like to suggest that misreading by the "two out of three" method could pose a significant threat to the fidelity of protein synthesis in the cell. To contain this threat, the codons of the genetic code have been laid out in such a way that codons that represent a high probability of reading "two out of three" are strictly confined to the codon families in which the "two out of three" method can be used with impunity. On the other hand, those places in the code where the "two out of three" method could lead to translational errors are exclusively occupied by low-probability codons. This organization of the code

and the competition with tRNAs having anticodons able to read all three positions of the codon would effectively prevent the "two out of three" method from being used when it might compromise translational fidelity.

I thank Drs. Charles G. Kurland and Jacques Ninio for stimulating and thought-provoking discussions. I am greatly indebted to Dr. Leslie E. Orgel for communicating his unpublished "rules" for codon-anticodon interactions to me.

1. Crick, F. H. C. (1966) *J. Mol. Biol.* **19**, 548-555.
2. Murao, K., Saneyoshi, M., Harada, F. & Nishimura, S. (1970) *Biochem. Biophys. Res. Commun.* **38**, 657-662.
3. Jou, W. M., Haegeman, G., Ysebaert, M. & Fiers, W. (1972) *Nature* **237**, 82-88.
4. Mitra, S. K., Lustig, F., Åkesson, B., Lagerkvist, U. & Strid, L. (1977) *J. Biol. Chem.* **252**, 471-478.
5. Bergquist, P. L., Burns, D. J. W. & Plinston, C. A. (1968) *Biochemistry* **7**, 1751-1760.
6. Strigini, P. & Brickman, E. (1973) *J. Mol. Biol.* **75**, 659-672.
7. Feinstein, S. I. & Altman, S. (1977) *J. Mol. Biol.* **112**, 453-470.
8. Hill, C. W., Foulds, J., Soll, L. & Berg, P. (1969) *J. Mol. Biol.* **39**, 563-581.
9. Hill, C. W., Squires, C. & Carbon, J. (1970) *J. Mol. Biol.* **52**, 557-569.
10. Carbon, J., Squires, C. & Hill, C. W. (1970) *J. Mol. Biol.* **52**, 571-584.
11. Grosjean, H., de Henau, S. & Crothers, D. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 610-614.
12. Kurland, C. G., Rigler, R., Ehrenberg, M. & Blomberg, C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4248-4251.
13. Hirsch, D. (1971) *J. Mol. Biol.* **58**, 439-458.
14. Hirsch, D. & Gold, L. (1971) *J. Mol. Biol.* **58**, 459-468.
15. Buckingham, R. H. & Kurland, C. G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5496-5498.