

5th April, 1965.

Dr. James D. Watson,
Harvard University,
The Biological Laboratories,
16 Divinity Avenue,
Cambridge, Mass. 02138,
U.S.A.

Dear Jim,

Very pleased to get your letter of 31st March. As you can see from the enclosed my version is fairly close to yours, except that I have more of Nirenberg's results (copy of his latest results enclosed). I would make the following comments:

1. As far as I know Nirenberg has not tested UGC for Cys, though he often talks as if he has. (Nor has he yet tested GUC, GCC, GCA or GCG.)
2. I have been more cautious than you in using the compositional data. For example, I have not accepted (GGC) and (GGA) for glycine, although I agree that they are reasonable.
3. My double ^{Ala} brackets for Ala and Gly mean "either position, or both" - based on mutagenesis.
4. Nirenberg finds that AGU works weakly for Ser, and AGC more weakly still. We also have shown that poly (AGU) incorporates Ser. Moreover we have been to some pains to show that it is not due to a C impurity.
5. Nirenberg tells me that UUA is difficult to assign, but that Ser is just possible for UCA. You will notice that several triplets containing GC work weakly for Ala, and several with AA weakly for Lys. It's unclear why CCA should give a weak effect for Lys.

6. You imply that Khorana has shown that poly AG incorporates Glu and Arg. I hope so, since I put in a lot of time telling him it ought to!
7. Yanofsky's cross of Arg x Val \rightarrow Gly + Ser may be explained as

A	U	
AGG	x	GUC

 with a few bits of special pleading.
8. The doubt about deducing "direction of translation" from Yanofsky's evidence is not really due to our partial ignorance of the code, since no plausible alternative version of the code leads to the opposite result, but rather concerns the outside markers. However he told me he now has closer outside markers, and is repeating the cross.
9. You probably know that George Streisinger's colleagues have evidence that a (+ -) pair of acridine mutants in the lysozyme of T4 alters a string of amino acids probably 6 of them in a row. This also enables one with luck to determine the direction of reading, as you will see if you think about it. Accepting their tentative data we deduced when I was there that it had to be as Ochoa has it i.e. from 5' to 3'. However, this cannot be accepted till both the wild-type sequence and the double mutant sequence are water-tight.
10. I certainly think that U = C is likely in the third position, but I am being cautious about A = G. A key triplet is AUA, which might be Ileu. Nirenberg tells me the binding results are "doubtful". Notice that he finds UGA doesn't work for Tryp. However, he has no positive result for Tryp to act as a control (he has not yet tested UGG), so we must be cautious in case the activating enzyme or the S-RNA was missing.
11. It's odd that all triplets of the form CU. and AG. work so poorly, and that if we rejected these results each amino acid would occur in only one square in the Table. I don't quite know what to make of this.

The copy of the code I have sent shows the codons for which we have reasonable evidence. This of course is not the same as the best guess. This is obtained by filling in each gap with the amino acid "nearest" to it. This assumes that there are no other nonsense triplets. The direct evidence about this is not at all clear. Notice that if U = C and A = G in the third position the code is practically solved already.

It is an interesting question whether one S-RNA recognizes both XYU and KYC. This could be done rather neatly if the anti-codon is GY'X', or IY'X', and allowing a bit of wobble. This would fit with the possible anti-codons (in yeast) of IGC for Ala (Holley) and IAC for Val (Ingram). In a similar way (but not the identical way) the anti-codon UY'X' might recognize XYA and XYG.

On general matters it seems to me that the cell-free system should give us a plausible version of the code. The best way to confirm this in vivo is by well-authenticated (i.e. repeated) mutagenic changes. In particular the following changes (if they exist) would be most helpful.

Arg - Ileu
Arg - Met
AspN - Ser
GluN - Leu
His - Leu
Ileu - Lys
Ileu - Ser

The other useful source of information will be the amino-acid sequences from (+ -) shifts. This will enable us to tell whether there is a U or a C (or an A or a G) in the third position (for that bit of the message). This is difficult to find in any other way.

Dr. James D. Watson

- 4 -

5th April, 1965.

l rules
Finally, the compositions of the anti-codons (when located) will help, but interpreting these data depends on the pairing ~~rates~~ between codon and anti-codon. I should be very surprised if the pairing in the first two positions is not the standard one.

Leslie Orgel had told me some weeks ago about your results on R17. However I am not clear whether the coat protein is produced when unloaded S-RNA from CR63 is added to the sus⁻ in vitro system, although that is what you imply. The key question, of course, is the origin of this new S-RNA.

Interested to hear about Wally's results, but sceptical about your theory, though there is no reason why it may not be right.

Odile has completely recovered from her wounds, and would be in fine spirits if she had not yesterday developed conjunctivitis in her right eye.

Looking forward to seeing you in June,

F. H. C. Crick

cc: Sydney Brenner
Leslie Orgel