A SPECULATION ON THE ORIGIN OF PROTEIN SYNTHESIS*

F. H. C. CRICK, S. BRENNER, A. KLUG, and G. PIECZENIK **

*Medical Research Council, Laboratory of Molecular Biology,
Hills Road, Cambridge, England*

Abstract. It is suggested that protein synthesis may have begun without even a primitive ribosome if the primitive tRNA could take up two configurations and could bind to the messenger RNA with five base-pairs instead of the present three. This idea would impose base sequence restriction on the early messages and on the early genetic code such that the first four amino acids coded were glycine, serine, aspartic acid and aspargine. A possible mechanism is suggested for the polymerization of the early message.

1. A Speculation on the Origin of Protein Synthesis

The origin of protein synthesis is a notoriously difficult problem. We do not mean by this the formation of *random* polypeptides but the origin of the synthesis of polypeptides directed, however crudely, by a nucleic acid template and of such a nature that it could evolve by steps into the present genetic code, the expression of which now requires the elaborate machinery of activating enzymes, transfer RNAs, ribosomes, factors, etc.

One solution is that the original mechanism was made mainly if not entirely of nucleic acid so that to express the earliest version of the genetic code (which was probably at that time both partial and rather inaccurate) little or no protein was required. It was suggested by Smithies (quoted in Crick, 1968) that in the beginning no activating enzymes were necessary because each primitive tRNA had a special cavity to hold its own amino acid. Woese (1967) made a similar suggestion. We shall not concern ourselves with this aspect of the problem here. It has also been suggested that the original ribosome was made entirely, or almost entirely, of nucleic acid. The hope has been that when the three-dimensional structure of the nucleic acid in the two portions of the present day ribosomes becomes known it may be possible to guess the structure of the primitive ribosome. For example the first ribosome may have consisted only of the ancestor of the present 5S RNA.

2. Protein Synthesis without Ribosomes

Here we consider an even more drastic simplification. We shall assume that originally no ribosome at all was necessary and that the ordering of amino acids in protein synthesis was accomplished using only messenger RNA and a few primitive tRNAs. This possibility has already been mentioned by Woese (1967 and 1972). The justification for this approach is that the synthesis of the basic clover-leaf structure of tRNA is not, on reasonable hypotheses, as improbable as might at first sight appear. This argument, first published by Orgel (1968) has

---

been made into an ingenious game by Eigen (1973). It is thus plausible to consider that in the primitive soup molecules existed not unlike the present tRNA molecules (though naturally without modified bases) many duplicate copies of which were produced from a nucleic acid template by some unspecified primitive copying mechanism.

3. General Requirements

There are a number of general requirements for a primitive system of protein synthesis. These are all aimed to reduce gross errors in the process while not necessarily removing minor errors. For example, the message must be read fairly consistently in the same phase since if the phase slips too often during the reading the resultant polypeptide will differ too much from the ideal one without any errors. On the other hand an occasional incorrect amino acid will not necessarily be unacceptable.

It seems likely that one such requirement is that, at any moment, the particular tRNA molecule to which the growing polypeptide chain is attached is bound to the messenger RNA by sufficiently strong bonds such that the two will not usually come apart until the polypeptide chain is transferred to the amino acid attached to the next tRNA. Otherwise polypeptide synthesis would be repeatedly interrupted and, worse, would usually resume again at the wrong place in the message.

The tRNA attached to an incoming amino acid, on the other hand, need not be bound to the messenger RNA so strongly and could perhaps come off and go on again before receiving the polypeptide chain since this would only slow the process rather than make a gross error in it. A tRNA with no amino acid attached should bind rather weakly, if at all, so that it will not interfere too much with the synthetic process.

It is possible to devise several rather involved schemes whereby each primitive tRNA was bound to the primitive messenger RNA by only the three bases of the anticodon. Since such an attachment by itself is unlikely to be stable one must invoke complicated interactions between tRNA molecules, adjacent on the message, in order to get a stable complex and in order that the message be read systematically in one direction. We shall not consider such schemes further here but will instead explore schemes in which the tRNA holding the polypeptide chain is held by 5 rather than by 3 base pairs.

4. Theoretical Assumptions

Our idea contains three main elements:

(1) That under the conditions then existing of temperature, salt, etc a tRNA molecule making *five* base pairs with a messenger RNA (rather than the present three) is stably attached for a sufficiently long time.

(2) That the anticodon loop of each primitive tRNA could take up two configurations. In the first of these (called by Woese (1970) the FH configuration because it was originally proposed by Fuller and Hodgson (1967)) the five bases at the 3' end of the seven base anticodon loop are stacked on top of each other. In
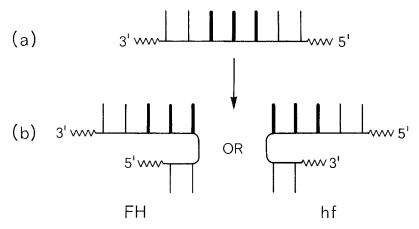
Fig. 1. The two configurations postulated for the anticodon loop, shown symbolically. (a) The seven bases of the anticodon loop drawn in a straight line. (b) The configuration proposed by Fuller and Hodgson (FH) is shown on the left. The other, the hf configuration suggested by Woese, is on the right. Each vertical line represents a base. The thick lines show the three bases of the present anticodon.
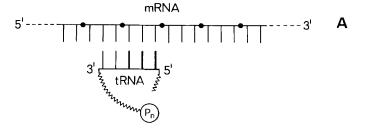
the second (labelled by Woese the hf configuration) the five bases at the 5' end form a stack (see Figure 1). The possibility of such a transition playing an important part in protein synthesis was first put forward by Woese in the ingenious paper quoted above. He also (Woese, 1972) suggested it might play a part in the primitive environment.
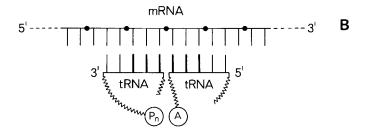
(3) We assume, following Woese, that when an amino acid is attached to a tRNA molecule the latter takes up the hf configuration; when a peptide is attached the configuration flips to FH. When neither is attached we make no special prediction – possibly both configurations can exist in equilibrium.
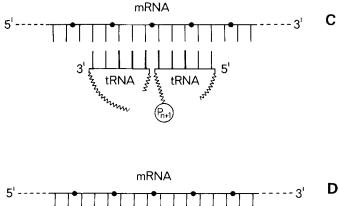
There is a fourth postulate which, if not absolutely necessary, makes the important conformation energetically more favourable .and thus several undesired arrangements less favourable. This assumes that there is a weak unspecific interaction between two tRNA molecules which are adjacent on the messenger RNA, the first being in the FH configuration and the second in the hf one.

5. The Suggested Mechanism

With these four assumptions the outlines of the mechanism are obvious. Consider first the state in the middle of the synthesis of a polypeptide chain when the tRNA (in the FH configuration) is held to the mRNA by five base pairs (the bases in the anticodon loop being unmodified) as shown in Figure 2A. The tRNA bearing the next amino acid coded for then enters the adjacent position, in the hf configuration, also making five base pairs, as in Figure 2B. Then, by proximity, probably aided by a general non-specific catalyst, the polypeptide chain is transferred to the new amino acid in the usual way, resulting in Figure 2C. This causes the tRNA which now has the polypeptide attached to flip to the HF configuration (Figure 2D) thus causing the previous tRNA to be held by only three base pairs, so that after an interval it falls off the mRNA. The process then repeats.
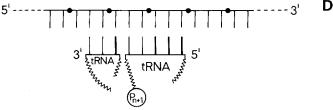
Fig. 2.   Each vertical line represents a base. The dots on the messenger RNA show the phase in which it should be read. The representation of the tRNA molecules has been greatly oversimplified. (A) The tRNA in the FH configuration with the nascent polypeptide, P_n, attached, sits on the mRNA making five base-pairs. (B) The tRNA carrying the next amino acid, A, goes onto the mRNA in the hf configuration, also making five base-pairs. (C) The polypeptide chain is transferred to the amino acid to give the polypeptide P_{n+1}. (D) The tRNA carrying the nascent peptide flips to the FH configuration. The tRNA which has given up its amino acid is now held by only three base-pairs so that it will shortly fall off, giving a situation similar to that of Figure 2A but moved along three bases. These figures are purely explanatory and show neither the correct scale nor the relative orientations of the components.

The primitive code, on this theory, was therefore a partially overlapping quin-tuplet code, the number five arising because a loop of seven bases (which we take as given) can have a stack of five bases on one side and two on the other, so that $5 = 7 - 2$. The movement along the mRNA of three bases at a time is produced because of the flip mechanism, since $3 = 5 - 2$.

It is almost essential, as has been emphasized before (Crick, 1968) for the primitive system to have moved along three bases at a time (rather than, say, two bases at a time) because of the principle of continuity. The fact that a sequence of five adjacent bases must be recognised places important restrictions on the base sequences of the early messages and of the primitive anticodons.

6. Possible Primitive Genetic Codes

We must now consider the implication of these ideas for the primitive genetic code. Here a fair number of possibilities exist. We shall only illustrate a few rather simple and indeed over-simplified possibilities.

We shall tentatively assume that the restrictions on the (unmodified) base sequences found in the present anticodon loops (Barell and Clark, 1974), are relics from the primitive tRNAs. These restrictions can be written

$$3' \quad NR\alpha\beta\gamma UY$$

(where the anticodon sequence is written backwards, with the 3' on the left) using the usual notation (and ignoring modified bases).

> N = any of the four bases, A, G, U, or C
>
> R = a purine, A or G
>
> Y = a pyrimidine, U or C

and where the α, β, γ stand for the three bases of the present anticodon, the third (or wobble) position (γ) being on the right.

To simplify discussion we now assume that some degree of "wobble" (that is, $U = G$ pairing) was possible in *all* positions and also that in the primitive tRNA the Y at the 5' end of the loop was a U (and not a C). Thus our primitive family of anticodon loops can be written

$$3' \quad NR\alpha\beta\gamma UU.$$

We now need to put restrictions on the messenger sequence so that five base pairs (normal or wobble) are always possible on both the FH and hf configurations of the tRNA. (The constraint arises because the bases adjacent to the anticodon must also pair with the message). Thus for the message we deduce the repeating family of sequences

$$\ldots\ldots, RRY, RRY, RRY, \ldots,$$

(where the commas are written to show the correct phase of reading) and for the anticodon the family

$$3' \quad UGYYRUU$$

the triplet part of the anticodon being in italics. Note that this symbolism does *not* imply that the message repeats exactly in groups of three but that the message must obey the purine-pyrimidine restrictions shown. Written out in full this becomes, for the mRNA

$$\ldots\ldots \begin{matrix} AAU \\ GGC \end{matrix} \quad \begin{matrix} AAU \\ GGC \end{matrix} \quad \begin{matrix} AAU \\ GGC \end{matrix} \ldots\ldots$$

and 3′ $UG^{UUA}_{CCG}UU$ for the anticodon loops where ^A_G represents A or G, etc.

The base pairs allowed are always either $A = U$, $G \equiv C$ or $G = U$ or their reversals. The pair $A - C$ is *not* allowed, nor are $A = G$ and $G - C$ (see Crick, 1966).

Notice two points:

(1) This restricted base sequence although written with commas for convenience of illustration, is comma-free (in the sense of Crick, Griffith and Orgel, 1957), that is, a tRNA with any of the possible loops specified above cannot go onto such a message in either of the two incorrect phases and make five base pairs whether the loop is in the FH or the hf configuration. The advantage, at this stage of the problem, in having a comma-free code is not just that the message cannot then be read in the two incorrect phases (which would only improve efficiency by a factor of three) but that a tRNA cannot go onto the message, out of phase, just ahead of the growing point and either block the whole process or shift the phase of reading.

(2) The codons allowed are those found in the present code in the bottom right-hand corner (as the codon table is usually written) and stand for

$$GG^U_C \quad GA^U_C \quad AG^U_C \quad AA^U_C$$
$$gly \quad\ \ asp \quad\ \ ser \quad\ \ asn$$

so that, for example, the anticodon loop for the glycine tRNA would be

3′ UG*CC*GUU

This is encouraging as most people would be willing to believe that at least three of these (gly, ser and asp) are among the more likely primitive amino acids.

The assumptions of wobble in all positions produces an asymmetrical lack of precision. Consider the two triplets coding for asn which are AA^U_C. These will be read unambiguously by the tRNA for asn having the anticodon

3′ UG*UU*GUU

and by no other tRNA of this limited set. Thus AA^U_C will code unambiguously for asn. The other three sets of codons will be read with varying degrees of ambiguity depending on how much wobble can occur in each position. Thus, because of wobble, the presumed anticodon loop for serine

3′ UG*UC*GUU

will read not only the codons AG^U_C but also, with less affinity, the codons GG^U_C, and thus occasionally insert serine by error into a glycine position.

These ideas should not be pressed too far. Our discussion is naive since we have made no allowance for $G \equiv C$ pairs being stronger than $A = U$ pairs, nor for stability being affected by stacking effects depending on base sequence. Further experiments are needed to allow correctly for these and other effects.

If we are prepared to relax the rule that there must always be five good base pairs in both the FH and the hf configurations then we can use for the anticodon loops the family

$$3' \quad UGNYR(U)U$$

which corresponds to the set of codons N_{GC}^{AU}, at the cost of occasional $U = C$ and $U = U$ pairs (which may be possible but rather weak (Crick, 1966)) in the position marked with a bracket. In the present code this adds the amino acids tyrosine, cysteine, histidine and arginine. A less likely alternative is the family

$$3' \quad (U)GYNRUU$$

which corresponds to the codon set $_{G}^{A}N_{C}^{U}$. The additional amino acids for these codons are at present isoleucine, threonine, valine and alanine. Both of these codon sets, separately, are comma-free. The second set is less attractive in that the possible weaker base pairing occurs not only in the hf configuration but also in the FH configuration. This latter is the configuration needed to hold the growing polypeptide chain to the mRNA and one might expect it to be the most stable of all. Note however that these codons might have included GC_{C}^{U} which now codes for alanine, another likely candidate for a primitive amino acid and that, since three $G \equiv C$ base pairs would give extra stability, the use of the codon GCC, combined with the four mentioned previously, is not unattractive. Whatever the details, the point is that new anticodons can be introduced by relaxation of the original rules.

7. A Difficulty

There is one possible difficulty with the type of scheme outlined above which should not be overlooked. The comma-free conditions largely prevents a tRNA going on in the wrong phase; that is, displaced by 1, 2, 4, 5, ... bases, but a tRNA can quite happily bind with 5 base-pairs displaced by 3 bases from the proper position. If it persisted there indefinitely, and if the nascent polypeptide chains could not be transferred to the amino acid of this tRNA then further synthesis would be blocked. This difficulty is not so great if there is a weak nonspecific affinity, as we have assumed, between two adjacent tRNAs, but not between two tRNAs spaced one or more bases apart on the mRNA. Indeed it would be better if a single tRNA in the hf configuration did not bind too strongly so that it could float away from the mRNA after a moderately short time. If this were so polypeptide synthesis would only be delayed rather than stopped completely should it have gone on in the wrong place. The additional binding of the entering tRNA, with its amino acid, when in the *correct* position next to the previous tRNA (having the nascent chain attached) would help stabilise this important complex.

In the latter stages of the evolution of the code a primitive ribosome might make it unnecessary for a tRNA to interact with more than three base pairs and all comma-free constraints would then be removed. At the same time modification of the anticodon loop might remove unwanted pairing outside the anticodon triplet itself, as is found in many tRNAs today. Once the comma-free restraints were removed

many other codons would be brought into play as these were demanded by mutation in the original rather simple messages.

Returning for the moment to the family of codons of the type $^{AAU}_{GGC}$ notice that the two possible out-of-phase readings of this class of message given the codon sets $^{AUA}_{GCG}$ and $^{UAA}_{CGG}$. The former is related to the present start codons $^{A}_{G}UG$ while the latter includes the present stop codons which are U^{AA}_{GG} if we ignore tryptophan (UGG) as being a later addition. Thus starting and stopping codons may originally have been evolved when the copying of the primitive message, with its restricted family of sequences, slipped out of phase.

8. Messenger Synthesis

Finally, we should consider how this original message, of the form ..., RRY, RRY, RRY, ... was synthesized. Apart from some repeated-slippage mechanism in the replication process a less obvious possibility is that the mRNA was initially formed *using the anticodon loops of the existing tRNA's molecules* as partial templates. This would be especially attractive if, under appropriate environmental conditions, there were a weak attraction between adjacent tRNA molecules and if tRNAs (without amino acids) could shift easily between the FH and the hf configurations.

Thus all that would be needed to get polypeptide synthesis started would be a single type of tRNA molecule to which a single amino acid was attached, though this would only produce a repeating homopolypeptide, such as polyglycine, from an equally simple message. By gene duplication and mutation (especially transitions) new, slightly different anticodon loops would be produced to pair with related codons and, hopefully, to attach to themselves new amino acids. Such simple pieces of chemical apparatus might well be enough to produce from a mutated message (or one synthesised by the mechanism suggested above) a few primitive proteins an occasional one of which might act to increase the accuracy and speed of the whole process. Given replication, natural selection could do the rest.

9. Concluding Remarks

Theories of the origin of life are usually fairly speculative and ours is no exception. The basic idea would be more credible if it could be shown that during present-day protein synthesis the tRNA does indeed occur in both the hf and the FH forms. At present the evidence on this point is weak and conflicting and so will not be reviewed here. If this flip mechanism turns out to be correct it may be possible to achieve template-directed synthesis in contemporary test-tubes without ribosomes by using (unmodified) tRNA molecules with carefully designed loops and having the appropriate amino acid attached to each one. This assumes that primitive tRNA molecules were very similar to present-day ones. The theory is thus to some extent open to experimental test.

References

Barell, B. G. and Clark, B. F. S.: 1974, *Handbook of Nucleic Acid Sequences*, Joynson-Bruvveri Ltd., Oxford.
Crick, F. H. C.: 1966, *J. Mol. Biol.*, **19**, 584.
Crick, F. H. C.: 1968, *J. Mol. Biol.*, **38**, 367.

Crick, F. H. C., Griffith, J. S., and Orgel, L. E.: 1957, *Proc. Nat. Acad. Sci. (U.S.A.)* **43**, 416.
Eigen, M.: 1973, in J. Mehra (ed.), *The Physicist's Conception of Nature*, D. Reidel Publishing Co., Boston, U.S.A.
Fuller, W. and Hodgson, A.: 1967, *Nature* **215**, 817.
Orgel, L. E.: 1968, *J. Mol. Biol.* **38**, 381.
Woese, C. R.; 1967, *The Genetic Code*, Harper and Row, New York, Evanston and London.
Woese, C. R.: 1970, *Nature* **226**, 817.
Woese, C. R.: 1972, in C. Ponnamperuma (ed.), *Exobiology*, North-Holland Publishing Co.