# On the Origin of Protein Biosynthesis

B. R. Thomas

Organic Chemistry 2, Lund University, Lund

(Z. Naturforsch. **30 c**, 323–326 [1975]; received December 10, 1974)

Evolution, Genetic Code, Codon-anticodon Site, Protein Biosynthesis

There is a very close steric relationship between the codon-anticodon site which accounts for the genetic code dictionary and a polynucleotide replicase site. Protein biosynthesis must therefore have arisen out of a primaeval polynucleotide replicase system.

The most important step in the evolution of life has been the bringing together of protein and polynucleotide systems to give a combination of the catalytic versatility and efficiency of enzymes with the stability of information storage in a polynucleotide-based genetic system. This step has hitherto remained obscure. It has indeed at times been regarded as paradoxical since enzymes have been assumed to be essential for the replication of polynucleotides and, in turn, polynucleotides have been regarded as essential to carry the information coding for the enzyme sequences.

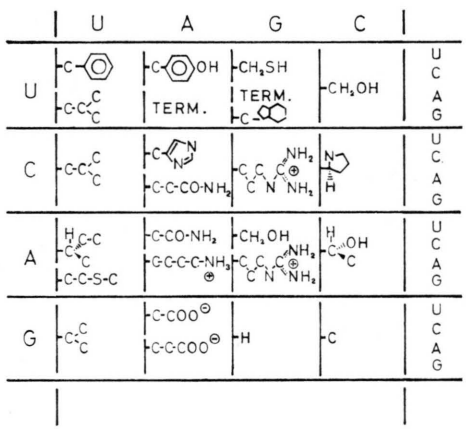


Fig. 1. The genetic code dictionary. The genetic code dictionary when written out with aminoacids represented by their side-chains shows clearly its stereochemical origin (see ref. 1). Thus for instance short aminoacids are all 4-codon except where Arg and Trp can compete. When the middle base pair of the codon-anticodon site is G:C or C:G the aminoacids selected are either short or very long, indicating the presence of the bulky amino group in the middle of the site. Longer aminoacids are nearly all polar and thus the longer medium-length aminoacids that are selected when the middle base pair is A:U give rise to a correlation of polar residues with second codon letter A. (The first codon letter is on the left-hand side, the third codon letter is on the right.)

Requests for reprints should be sent to Dr. B. R. Thomas, Organic Chemistry 2, Lund University, *S-220 07 Lund*, Sweden.

However it has been shown that a segment of polyribonucleotide double helix such as the codon-anticodon site of protein biosynthesis provides a surface (the face groove or "small" groove) that accounts for the specifications of the genetic code dictionary (see Fig. 1) [1]. It follows that a codon-anticodon site of this type or an analogue must have formed a primaeval "synthetase" site or at least a part of a synthetase system which selected amino acids for reaction with ATP. The present paper shows that the codon-anticodon site is closely related to a polynucleotide site and thus that protein biosynthesis must have arisen out of a polyribonucleotide replication system.

## Method and Materials

The steric relationships discussed in this paper have been checked with space-filling models (scale 10 cm = 1 nm) specially designed for these investigations (see ref. 1). These have included models of ATP, the AMP derivative of alanine, all the amino acids of the genetic code dictionary and a segment of polyribonucleotide double helix. Variation is possible in the period, diameter and base-pair tilt of a polynucleotide double helix. The double helix segment used here was constructed with base pairs placed at right angles to the helix axis and was adjusted to give best fit between an amino acid carboxylate group and the ribose hydroxyl pair that holds and orients the carboxylate group on the codon-anticodon site (see Fig. 2).

## Discussion

The correlations between amino acid structure and the codon-anticodon sites specified by the genetic code dictionary [1] are too many and the

Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung") beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition "no derivative works"). This is to allow reuse in the area of future scientific usage.

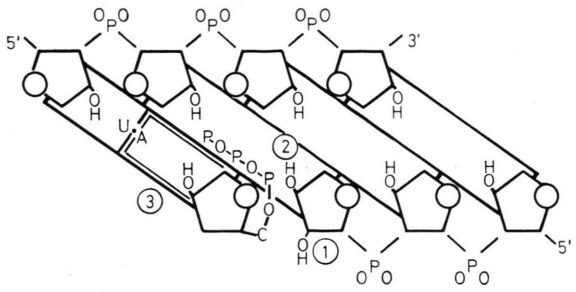


Fig. 2. The codon-anticodon "synthetase"/replicase site. The correlations of codon-anticodon site stereochemistry with aminoacid structure place the aminoacid carboxylate group at 2 with side chain extending to the right. Conversion of an aminoacid to its AMP derivative on a "synthetase" site requires ATP at 3 (double-edged base). On a replicase site the incoming base at 3, here ATP coded for by U, will react with the ribose-3′-hydroxyl at 1. The positions of ATP in both cases will be essentially identical apart from a slight shift in the position of the triphosphate group.

genetic code dictionary itself is too firmly based on experimental results (see ref. 2) for the correlations to be due to chance. The codon-anticodon site is of the correct size and provides a suitable range of variation. It has suitable chemical characteristics, *i. e.* pairs of ribose hydroxyls for hydrogen bonding to the carboxylate group and a phosphate group to serve as counteranion. Its variations correlate with short (4-codon), medium-length (2-codon) and long (1-codon, except Arg) amino acids. They correlate in profile with short, with flat and with the long and awkwardly shaped (Arg and Trp) amino acids. Five of the six pairs of two-codon amino acids of medium length show a correlation between the shorter amino acid of each pair and the presence of a pyrimidine in the third codon letter, presumably due to the steric hindrance of the pyrimidine carbonyl group at the near end of the third base pair. The discrimination between Phe and Tyr is apparently also a correlation of length with codon-anticodon site profile since the A:U base pair in the middle of the codon-anticodon site will slope upwards away from the carboxylate group allowing the longer Tyr to extend across a ribose grouping.

Although the synthetases that now carry out the amino acylation step are proteins (questions raised by the modern synthetases and their function are not of direct concern here and will be dealt with at a later opportunity), these codon-anticodon site correlations indicate that there must have been a primaeval amino acid selecting site of codon-anticodon type that has given rise to protein biosynthesis. (This need not be and probably is not the codon-anticodon site formed by engagement of tRNA with mRNA on the ribosomal peptidase site.) The primaeval synthetase site must of course have had a site for ATP in the immediate vicinity (this can be taken either as the simplest possible assumption or, possibly, as a definition). This ATP site should also be of polynucleotide structure. Such a site is provided by the free side of the first base pair of the codon-anticodon grouping.

Protein biosynthesis is a very complicated process. Even without the protein factors and synthetases that are now required, it involves two nucleotide triphosphates, twenty amino acids, forty or fifty tRNAs, a complicated ribosomal apparatus and an mRNA sequence. It follows that protein biosynthesis must have developed from some pre-existing process. That is, it must have arisen out of a polyribonucleotide replicase system. This is to be expected since, while both peptides and polynucleotides can reasonably be expected to have been components of the prebiological environment, only polynucleotides have properties that would facilitate copying. Further, polynucleotide replication requires only a coding strand and four nucleoside triphosphates and, given a conformation that accelerates copying, a polynucleotide particle is self-sufficient. A protein-based enzyme on the other hand must be a component of a relatively complicated system. Thus we can predict that the primaeval codon-anticodon site, or "synthetase" site on which amino acids were selected for reaction with ATP must be closely related to a polyribonucleotide replication site.

This prediction is easily tested. The test is in fact essentially trivial and can be quickly checked with the aid of a few simple space-filling models. On a replicase site there will be a double helix made up of coding and product strands extending to the replication point. The coding strand will extend beyond the replication point in a 5′-direction. The product strand will terminate with its 3′-end at the replication point. The next incoming base will stack on the 3′-end of the product strand so that it is paired with the corresponding base on the coding strand and can react with the 3′-hydroxyl of the terminal ribose of the product strand. On the "synthetase" site of protein biosynthesis the codon-anticodon pair will correspond to the last three base pairs of double helix and the ATP site will be equivalent to a replicase site with uridine as the first unpaired base along the replicase coding strand. The

prediction is then – that ATP when paired with uridine will be in good position to react with the carboxylate group of an amino acid on the codon-anticodon site (see Fig. 2).

Inspection of the models shows immediately that a slight displacement of the triphosphate grouping will be sufficient to cause the ATP to react, not with the ribose-3′-hydroxyl on the preceding base, but with the amino acid on the codon-anticodon site. This clear and simple relationship between replicase site and amino acid selecting site cannot be merely fortuitous. It therefore indicates a direct evolutionary relationship between protein biosynthesis and an earlier replication system for polyribonucleotides.

There is excellent support for a close relationship between protein biosynthesis and polyribonucleotide replication in the functions of tRNA in transcription control [3] and in the dual involvement of some auxiliary protein factors both in protein biosynthesis and in replication [4, 5]. The requirement for these protein factors in both types of system might be used as an argument for an evolutionary development proceeding from protein to polynucleotide systems but this is completely excluded by the complicated nature of the replicases and of the protein factors themselves, quite apart from other arguments. However, protein biosynthesis and replicase systems must have evolved along closely parallel lines for the protein factors to retain their double function, possibly because the two systems are closely interlocked.

Even more striking support for a very close relationship between protein biosynthesis and polynucleotide replication is the function of fMet-$tRNA_f^{met}$ as a primer for the replication of an RNA tumour virus with reverse transcriptase [6]. It has moreover been found that polymerisation begins here with the addition of dAMP to the 3′-end of the tRNA [7].

In a primitive pre-protein replicase the basic framework must have been a polynucleotide particle which had evolved from simple polynucleotide precursors by selection for improved stability and had developed some replication mechanism. In an initial mixture of polynucleotide particles produced by random polymerisation and depolymerisation the development of a conformation that gave some advantage in replication rate would of course lead to the production of large populations of similar particles and would favour the evolution of a replication mechanism. It is not possible at this stage to indicate the critical change that made protein biosynthesis possible. It would be necessary to alter the point of attack of the triphosphate grouping slightly and to "reverse" the direction of movement of the coding strand. These changes might well be associated.

The "reversal" of the coding strand can be accomplished in two ways. It might be direct, so that the coding strand becomes an mRNA moving in a 5′-direction. More probably, the product strand of the replicase would become an mRNA equivalent by the elaboration of an analogous codon-anticodon site separate from the replicase site, thus avoiding any change in the direction of movement of the polynucleotide strands through the system. Initially it would not be essential to provide for continuous translation or for the interlocking of translocation and peptide bond formation. Nor would it be necessary for a single particle to retain both functions provided the population contained replicases accepting non-self polynucleotide strands. It would however be necessary to collect aminoacyl groups on a peptidyl-tRNA.

The present paper is concerned solely with the replicase relationship. The further reactions required to transfer an amino acid residue from its AMP derivative to the 3′-end of a tRNA molecule and thence to the end of a peptide chain present a much more complicated topic which cannot be dealt with here (some possibilities have been discussed elsewhere [8]).

## Conclusion

Simple consideration of the chemical structures involved shows that protein biosynthesis must have arisen as a further development of polynucleotide replication. Although the questions of the order in which protein biosynthesis and polynucleotide replication arose and the way in which they were interconnected have given rise to inexplicable difficulty in the past, it seems that the evolution of life was a simple and straightforward chemical process. Since, as it now appears, conditions on the primaeval earth will have led to the formation of the precursors required for both polynucleotides and polypeptides, life must have appeared practically as soon as physicochemical conditions permitted polynucleotides sufficient stability for them to accumulate.

The relationship of protein biosynthesis to replication provides further confirmation of the basic correctness of the codon-anticodon site origin of the genetic code dictionary. It therefore finally eliminates the very unsatisfactory hypothesis of a "frozen accident" [9]. It also excludes the possibility that the genetic code has gone through a stage of evolution to reach its present form since all 64 potential codon-anticodon sites would have been available from the start. There might well of course have been minor changes in the set of amino acids competing for these sites.

It should be pointed out that the origin of protein biosynthesis is by no means the minor academic question it might have been if the synthetases were a random collection of amino acid-recognising enzymes. The mechanisms and reactions involved will reflect some of the biochemical potential of polynucleotide structures and will probably have formed the basis for the development of further fundamental biological systems. They are therefore of wide interest. The primitive polynucleotide particles among which protein biosynthesis arose will have had some resemblance to present day viruses, albeit without an organised protein coat and without devices for the penetration of a cell membrane. It would appear then that viruses are to be regarded not as degenerate and parasitic molecules but as representatives of the type of biological molecule present at this very early stage of evolution. Since a range of tRNAs are required simultaneously for protein biosynthesis, it follows that these must represent an important group in the primitive polynucleotide population. Ribosomal particles can also be expected to be related to some of the early particles which produced protein biosynthesis but they must have undergone considerable evolutionary development subsequently since they contain elaborate protein matrixes.

I am grateful to Professor Börje Wickberg for laboratory facilities.

[1] B. R. Thomas, Biochem. Biophys. Res. Commun. **40,** 1289 [1970].
[2] J. N. Davidson, The Biochemistry of the Nucleic Acids, 7th ed., Chapman and Hall, London 1972.
[3] U. Z. Littauer and H. Inouye, Ann. Rev. Biochem. **42,** 439 [1973].
[4] T. Blumenthal, T. A. Landers, and K. Weber, Proc. Nat. Acad. Sci. U. S. **69,** 1313 [1972].
[5] H. Weber, M. A. Billeter, S. Kahane, C. Weissman, J. Hindley, and A. Porter, Nature New Biology **237,** 166 [1972].
[6] K. T. Elder and A. E. Smith, Nature **247,** 435 [1973].
[7] A. J .Faras, J. M. Taylor, W. E. Levinson, H. M. Goodman, and J. M. Bishop, J. Mol. Biol. **79,** 163 [1973].
[8] B. R. Thomas, Biochimie **55,** 1325 [1973].
[9] F. H. C. Crick, J. Mol. Biol. **38,** 367 [1968].