CELL-FREE PROTEIN SYNTHESIS AND THE GENETIC CODE

by

J. Heinrich Matthaei and Marshall W. Nirenberg

National Institutes of Health, Bethesda, Md.

Introduction

Biochemistry is still largely in its analytical stage; people attempt to get exact knowledge of single reactions occurring in the intact cell. Therefore, they need simplified systems, containing possibly only compounds involved in the reaction under investigation. However, since protein synthesis involves a whole variety of chemical reactions, biochemists had to start working with crude systems, including extacts from Escherichia coli (a bacterium living in our intestines). These extracts can be obtained after grinding the fresh cells with aluminum oxide, followed by extraction with buffer solution and removing all remaining cells, protoplasts and debris by centri - fugation. One may call these extracts a "suspension of ribosomes* in a solution of enzymes and nucleic acids." All smaller molecules are thoroughly removed by dialysis to make it possible to have those

---

* Ultramicroscopical particles consisting of ribonucleic acids, very basic proteins and several enzyme-proteins.

compounds added in defined amounts, which are necessary for protein synthesis: 20 amino acids as the building blocks of proteins, ATP and GTP as energy sources for the reactions and salts for a certain ionic environment. Mercaptoethanol serves as a stabilizer which allows to prepare extracts and store these in frozen state without loss of activities required. The incorporation is measured by using C^{14}-amino acids and counting radioactivity in the protein precipitated with trichloroacetic acid.

Survey of processes involved in protein synthesis

All the major processes involved in protein synthesis seem to occur in this crude system (see fig. 1):

1. The activation of the amino acids, including an energy-transfer from ATP (Lipmann), a liberation of pyrophosphate (Keller and Hoagland) and esterification of the amino acid to its species of transfer-ribonucleic acid (RNA) (Holley, Lipmann/Zachau, Hoagland, and others).

2. The formation of "messenger"-RNA from 4 kinds of nucleoside-triphosphates complementary in its bases to the deoxyribonucleic acid-strand (DNA) of the gene (Hurwitz, Weiss, Stephens, and others). This messenger-RNA (postulated by Jacob and Monod), carries by means of the sequence of its 4 species of nucleotides the information from the gene to the factories of proteins; there, it determines the specific sequence of some 20 different amino acids within long polypeptide-chains of up to several hundred units. The demonstration of an inhibition of amino acid incorporation into protein by a DNA-destroying enzyme (DNAase) by Tissières, Novelli, and Matthaei and Nirenberg indicated that "messenger-RNA" might be produced in E. coli extracts. Further experiments with P^{32}-labeled nucleotides done in collaboration with Dr. R. Roberts, have shown that a messenger-like RNA is only made in the absence of DNAase. However, the direct evidence, a production of specific "messengers" upon the addition of the DNA from individual genes, is not established at the moment.

3. The complex-formation between messenger-RNA and some stage in the develolment of a ribosome remains to be studied. Isotope-labeled synthetic messenger-RNA, like polyuridylic acid, which we found to be highly active in stimulating amino acid incorporation, will be very useful for such investigations on the fate of the "messenger".

4. The amino acid-charged transfer-RNA is assumed to recognize and bind to specific places on "messenger"-RNA by the classical base-pairing mechanism which was discovered in DNA by Watson and Crick (A pairs with U, G with C). Thus, transfer-RNA carries as an adaptor specific amino acids to their proper places on messenger-RNA, so that the amino acids can be linked, in correct sequence, into protein.

5. Peptide-bonds are formed between the amino acids that are sequentially arranged along the messenger-RNA-strand. The steps involved are unknown and are being investigated in several laboratories. Here again, the use of polyuridylic acid coding

specifically for phenylalanine is very helpful, since it allows an enormous reduction of the precursors required for a model-synthesis of an extremely simple polypeptide chain.

The assay for "messenger"-RNA

In order to make all protein synthesis in our system dependant upon the addition of informational or "messenger"-RNA, we destroyed the DNA with DNAase and incubated the E. coli extracts until all of the endogenous "messenger"-RNA had been inactivated in some way. This treatment gave us the first "assay" system for "messenger"-RNA.

This "assay-system" seemed to copy informational RNA extracted from any organism; from E. coli as well as from yeast, tobacco mosaic virus (TMV), or Ascites tumors of mice. The synthetic RNA polyuridylic acid directed 100 times as much amino acid into protein than was ever observed in a cell-free system. This polymer contains only uridylic acid, one of the 4 kinds of nucleotides found generally in ribonucleic acids, and leads to the polymerization of only one

of the 20 protein-amino acids, phenylalanine. We showed that transfer-RNA carries this amino acid towards polyphenylalanine-synthesis.

Nucleotide-composition of RNA coding-units

This discovery opened the way towards the deciphering of the genetic code. The further production and use of synthetic polynucleotides let us find out which groups of nucleotides direct the other amino acids by means of their transfer-RNA-adaptors into their proper places in the specific proteins. When we had enzymatically synthesized RNA containing two, three of four different species of nucleotides in a random sequence, we could determine which amino acids would be "coded" by certain combinations of nucleic acid-bases: U, C, A or G. The results so far definitely established, are seen in Table 1. Ochoa and collaborators have reported similar findings.

TABLE 1

Amino Acid	Code word determined
Alanine	UCG...
Arginine	UCG...
Cysteine	UUG...
Glutamic acid	UAG...
Glycine	UGG...
Isoleucine	UUA...
Leucine	UUG... + UUC...
Lysine	UAA...(?)
Methionine	UAG...
Phenylalanine	UUU...
Proline	UCC...
Serine	UUC... + UCG...
Tryptophane	UGG...
Tyrosine	UUA...
Valine	UUG...

The number of nucleotides per coding unit, but not the sequence, was determined. For this purpose, we expressed the amount of each amino acid incorporated in percent of the phenylalanine, which was directed into protein by the same U-containing polynucleotide.

We calculated also the probability of any triplet of nucleotides in percent of UUU, coding presumably for phenylalanine. This must be done on the basis of the determined quantitative nucleotide-composition of the polynucleotides used. So we could correlate the observed incorporation of any amino acid to the triplet with the best-fitting statistically expected frequency. These determinations were done with many polynucleotides of varied ratios between the comparing nucleotides and essentially led to the same results.

On the basis of these calculations, however, we could not decide whether the coding units might contain uridylic acid residues in addition to those specified. If the number of letters per code word would be larger than three, a proportional increase in the number of words coding for the individual amino acids should be expected (= more degeneracy). The triplet-code is still not only the simplest, but also the most likely and experimentally uncontradicted concept. This conclusion comes from both biochemical and genetic (Crick) evidence.

The use of different methods in synthesizing RNA of well-defined nucleotide-sequences should allow us in the near future to get direct

evidence for the number and sequence of nucleotides in the code-words. Then, we might also find other coding units possibly existing in addition to the ones already determined for certain amino acids. The code could be more degenerate.

The high proportion which U takes in the coding units determined thus far, may disappear where other than random-polynucleotides will be used and show possibly U-less code words in addition. The present selection of partially known code words may be just the result of certain limitations inherent in the methods used.

Single-strandedness of coding units

Poly A base-pairs and forms double- and triple-stranded helices with poly-U. In this manner it totally inactivates poly-U added to our E. coli system. Poly-C does neither base-pair nor inactivate poly U. This may be a model for "repression" of the synthesis of certain proteins on the level of RNA. Such repressions occur in cells and if their mechanism is disturbed, uncontrolled synthesis of proteins might result.

Universality of the code?

These determinations of RNA coding units in a bacterial

system would be more significant, if different organisms used the same set of coding units, and if each unit would be correlated in every one to the same amino acid. The genetic code would then be "universal". This universality is favored by already existing observations:

1. Mutations of the TMV-protein, studied by Wittmann in Tübingen, and Tsugita and Fraenkel-Conrat in Berkeley, resulted after treatment of the TMV-RNA with nitrous acid. There occur only two mutagenic transitions of nucleic acid bases by oxidative deamination; U replaces C, or G replaces A in one single place somewhere along the RNA-chain. As a consequence, mutants occur, in which single amino acids replace certain other amino acids found in TMV-protein of the wild type. 12 out of 14 different types of amino acid-replacements can be explained by the base compositions determined with random-polynucleotides in the E. coli system. Chance should have led to less than 33% agreement.

2. Another approach, taken first by Lippmann and V. Ehrenstein, is to use amino acid-charged transfer-RNA from E. coli and let it deliver its amino acids in a cell-free system from another organism,

actually from rabbit-reticulocytes. The amino acids became incorporated into the proper places in rabbit-hemoglobin. Thus, at least part of the code seems to be universal. Further experiments done in many laboratories at the moment, shall finally answer this question. The general nature of these experiments is to put either messengers for the formation of a specific protein or amino acid-charged transfer-RNA into a cell-free system prepared from another organism. If the code is universal, transfer-RNA from one and messenger from another species have to fit for making correct amino acid sequences.

A fundamental concept, the transfer of inherited information being stored in certain nucleotide sequences for the ultimate translation into the broad variety of functioning proteins, has found its final experimental proof. Certain findings made during our work on the coding problem, have promoted the research of many laboratories on various other processes involved in protein synthesis.

SUMMARY

1. The base-composition of coding units for 15 amino acids was determined by means of random-polynucleotides of different base-composition.

2. The code is degenerate at least for the amino acids leucine and serine.
3. The informational part of RNA appeared to be single-stranded.
4. At least part of the code seems to be universal.
5. Important features of the code are still unknown: The total number and the sequence of nucleotides in the coding units, the amount of degeneracy, and the polarity and chemical nature of "starting points", from which the messages apparently (Crick) are read off.