RESEARCH

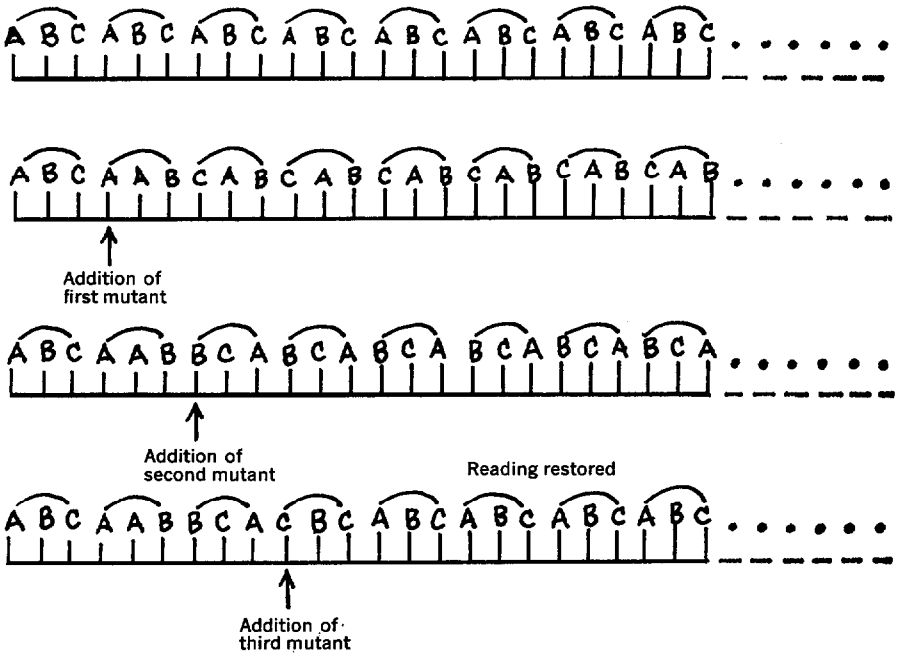
# Findings May Speed Solution to Genetic Code

## RNA stimulates amino acid incorporation into protein; sequence of bases in nucleic acid appears to be key to code

More and more information is gathering which soon may permit chemists to solve the mysteries of the genetic code. Two teams of scientists in the U.S. have shown experimentally that ribonucleic acid (RNA) stimulates amino acid incorporation into cell protein. They have also prepared a synthetic RNA that is specific for phenylalanine. A team of British scientists has evidence that the code is related to the sequence of bases along the nucleic acid of genetic material (C&EN, Jan. 8, page 43).

Two chemists at the National Institutes of Health, Bethesda, Md., have found new evidence showing that the hereditary instructions for the synthesis of proteins in living cells are carried by a "messenger" RNA. NIH's Dr. Marshall W. Nirenberg and Dr. J. Heinrich Matthaei have developed a stable, cell-free amino acid–incorporating system dependent upon messenger or soluble RNA (sRNA). They have also been able to substitute a simplified synthetic messenger for the natural product.

RNA has a long chainlike polymeric structure consisting of ribose units linked by phosphate bonds. Attached to each ribose unit is one of four key nucleotides: adenylic acid (A), guanylic acid (G), cytidylic acid (C), and uridylic acid (U) (C&EN, May 8, 1961, page 81). Chemists think that the sequence of these nucleotides on the RNA chain governs how RNA directs amino acid incorporation into protein. Chemists soon found that the amino acids form covalent links to sRNA and that the reaction is catalyzed by the same preparations that contain the activating enzymes.




**TRIPLETS.** Chart prepared by Dr. F. H. C. Crick and co-workers shows type of code involved in the nucleic acid of genetic material. They found evidence for a coding ratio of 3 by examining six triple mutants of the FC family. In all the six, the combination of three mutants restored the function of the gene-product. With all three mutants in combination the order of triplet reading is restored. A, B, and C each represent a different base of the nucleic acid. The sequence is such that it is read in sets of three starting on the left

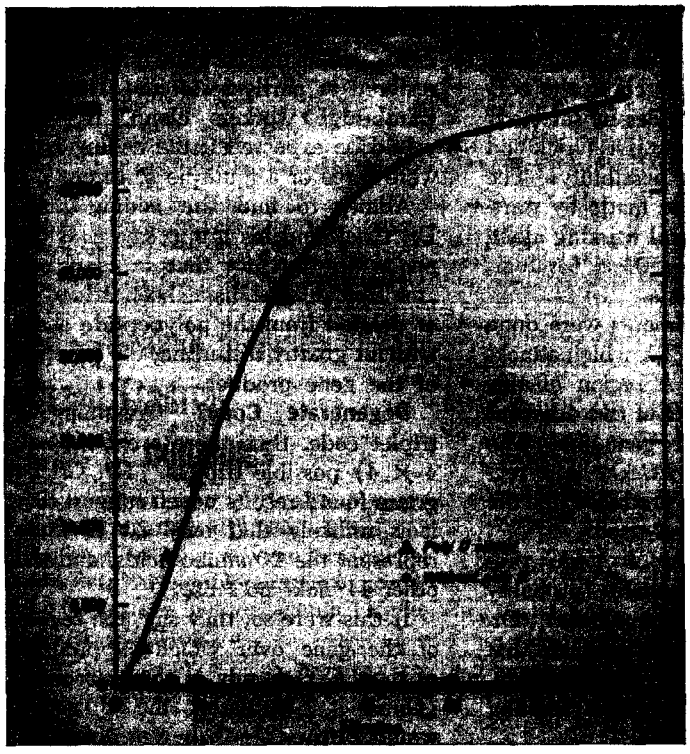
Controversy, however, exists concerning the role of sRNA in protein synthesis. Little information, too, is on hand as to the nature of the code on the sRNA chain (the number and sequence of nucleotides) that signals which amino acids will be incorporated into cell protein.

Dr. Nirenberg and Dr. Matthaei have succeeded in preparing from the colon bacillus (*Escherichia coli*) cell-free extracts which actively incorporate amino acids into protein. Because of its stability, the preparation is a stimulus to further research. In the past, fresh cell-free *E. coli* extracts had to be made for each experiment.

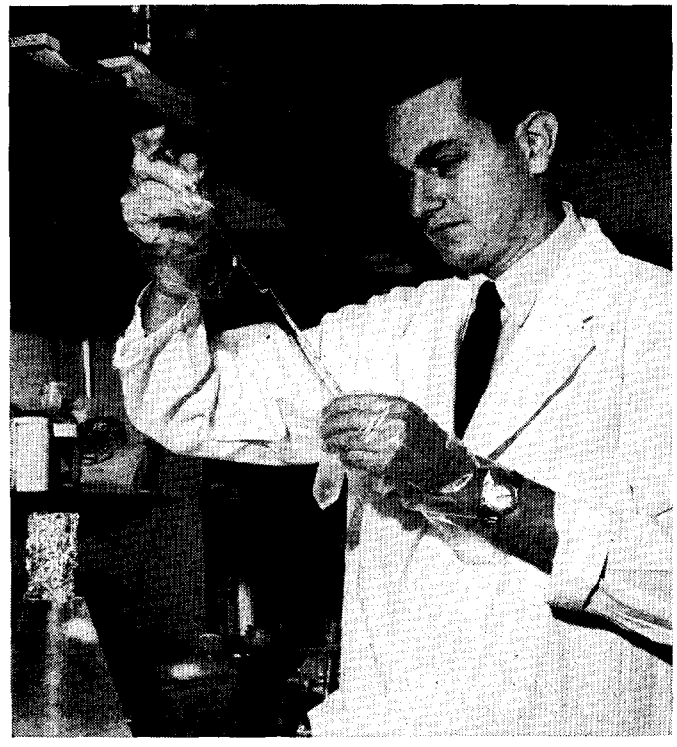
The NIH chemists' preparation can incorporate amino acids into protein, as evidenced by the following characteristics (typical of amino acid incorporation into protein):

- Both ribosomes and a certain fraction of the supernatant liquid after centrifugation (105,000 × g) are needed for amino acid incorporation.
- Adenosine triphosphate (ATP) and an ATP-generating system are necessary for protein synthesis.
- Incorporation of a particular amino acid is stimulated by a mixture of other amino acids.
- Certain substances known to inhibit amino acid incorporation into protein (puromycin, chloramphenicol, and RNAase) strongly inhibit the activity of the extract.

The NIH team's studies show that soluble RNA from *E. coli* stimulates amino acid incorporation into protein. In addition, they find that messenger RNA preparations from *E. coli* strongly stimulate incorporation of $C^{14}$-tagged valine into protein. When low messenger RNA concentrations are used, they obtain a linear relationship between messenger RNA concentration and $C^{14}$-valine incorporation into protein. The linear relationship between messenger RNA concentration and activity provides a valuable assay for messenger RNA.



**STIMULATION.** NIH graph shows that polyuridylic acid strongly stimulates phenylalanine incorporation into protein



**CELL-FREE.** Dr. Marshall Nirenberg removes sample of cell-free extract which incorporates amino acids into protein

When Dr. Nirenberg and Dr. Matthaei add a mixture of 20 amino acids to their preparation containing the ribosomal RNA, the amino acids stimulate the incorporation of valine into protein. Addition of small amounts of DNAase inhibited amino acid incorporation, suggesting that the genetic material was directing protein synthesis in this cell-free system.

In studying the effect of *E. coli* ribosomal RNA upon the incorporation of seven different amino acids (valine, threonine, methionine, arginine, phenylalanine, lysine, and leucine), the NIH team finds that the added messenger RNA increases the incorporation of every amino acid tested. RNA from other sources stimulates $C^{14}$-valine incorporation into protein, too. Tobacco mosaic virus RNA, for example, strongly stimulates amino acid incorporation.

**Phenylalanine Incorporation.** The NIH team prepared a synthetic RNA composed of only one of the four nucleotides to see if it might specifically incorporate one amino acid. They made polyuridylic acid (poly U) and added 10 micrograms of it per ml. of reaction mixture. Poly U, they found, strongly stimulates $C^{14}$-phenylalanine incorporation. They prepared other polynucleotides such as polyadenylic acid (poly A) and polycytidylic acid (poly C) but found that no other polynucleotide tested can replace poly U in activating phenylalanine. A solution of poly U and poly A (which forms triple-stranded helices) had no activity at all. Thus, single-strandedness is a necessary requirement for activity, the NIH chemists say.

Omitting a mixture of 19 other amino acids does not inhibit phenylalanine incorporation, showing that polyuridylic acid stimulates the incorporation of phenylalanine alone, Dr. Nirenberg and Dr. Matthaei say. Poly U, they find, has little effect in stimulating the incorporation of 17 other radioactive amino acids.

Because the addition of sRNA can't replace template RNA in their system, the NIH team says that template RNA (in the cell ribosomes) is a requirement for cell-free amino acid incorporation.

Dr. Nirenberg and Dr. Matthaei's work shows that poly U contains the information for making polyphenylalanine. One or more uridylic acid residues therefore appear to be the code for phenylalanine. Poly U, it seems, functions as a synthetic template or messenger RNA. And this stable, cell-free *E. coli* system may well synthesize any protein corresponding to meaningful information contained in added RNA, they add.

Another team of chemists is working with *E. coli* systems. Peter Lengyel, Dr. Joseph F. Speyer, and Dr. Severo Ochoa of the New York University school of medicine also find that various synthetic ribonucleic acids are active as messengers in an *E. coli* system and determine incorporation of different amino acids into an acid-insoluble product.

The NYU team studied the incorporation of different $C^{14}$-labeled amino acids into an acid-insoluble product by the *E. coli* supernatant plus ribosomes system, with and without the addition of poly U. Out of 19 amino acids tested individually, only the incorporation of phenylalanine is strongly stimulated by poly U. This agrees with the NIH team's findings. They also find that poly U exerts a small stimulation on the incorporation of leucine and isoleucine. Adding *E. coli* transfer RNA causes a further pronounced increase of phenylalanine incorporation. This indicates to them that the polymers affect the transfer of activated amino acid residues from transfer RNA to ribosomes and act as messenger or template RNA in this system. In one experiment they found that with poly U and phenylalanine, without addition of transfer RNA, rat liver ribosomes can be substituted for their *E. coli* counterparts.

**Other Polynucleotides.** The NYU group also studied the effects of other homopolymers on amino acid incorporation. Poly A, they find, does not stimulate incorporation of any of 19 amino acids tested. Addition of poly A to a system containing poly U completely inhibits poly U's effect. This, they say, is undoubtedly due to formation of the double-stranded, helical poly A + U complex. Polythiouridylic acid (which appears to be multistranded) has no effect, and polyfluorouridylic acid, which like poly U is single-stranded, had only a small effect on phenylalanine incorporation. Polycytidylic acid (poly C) has a small but consistent effect on the incorporation of proline but has no influence on that of any other amino acid.

Poly UC (U:C = 5:1), they find, promotes the incorporation of phenylalanine and serine, and poly UA (U:A = 5:1) stimulates the incorporation of phenylalanine and tyrosine.

**Year-End Goal.** The genetic code for proteins may be deciphered before the end of 1962. This is a prediction of four scientists working at the Cavendish Laboratory of the Medical Research Council Unit for Molecular Biology, Cambridge, England. They base it on the results of experimental work indicating the type of code that is involved in the sequence of bases along the nucleic acid of genetic material. The scientists are Dr. F. H. C. Crick, Leslie Barnett, Dr. S. Brenner, and Dr. R. J. Watts-Tobin.

Here's how they typify, in general terms, the genetic code:

- A group of three of the four bases (or less probably, a multiple of the three) codes one amino acid.
- The code is nonoverlapping.
- A fixed starting point is used to read the sequence of bases. This, the scientists say, determines how the long sequences of bases are to be correctly read off as triplets. There are no special "commas" (every fourth base) to show how to select the right triplets. Displacing the starting point by one base displaces the reading of triplets, and becomes incorrect.
- The code is probably "degenerate," or one amino acid can be coded by one of several triplets of bases.

**Nonoverlapping Code.** Pointing to the existing evidence that the genetic code is not overlapping, the Cambridge scientists contend that their experimental results rule out all overlapping simple codes.

But if there is no overlapping, how can the correct triplets along the continuous sequence of bases be selected? Of the possible answers, Dr. Crick and his group favor this possibility: The correct choice may be made by starting at a fixed point and working along the sequence of bases three (or four, or whatever) at a time.

The genetic experiments were done on the bacteriophage T4, which attacks strains of *E. coli.* A region of this bacteriophage consists of two adjacent genes, gene A and gene B. The mutant (designated *FC* O) in a segment of the B gene produced by the action of proflavin was used.

Assuming that the B gene produces a polypeptide chain (probably through an RNA intermediate), Dr. Crick and his co-workers imagined that the string of nucleotide bases be read, triplet by triplet, from a starting point left of the B gene.

They then supposed that the mutant *FC* O be produced by inserting an addition of a base in the sequence. This addition of a base at the *FC* O site will mean, the Cambridge workers say, that the reading of all triplets to the right of the *FC* O will be shifted along one base, and will be incorrect. So the amino acid sequence of the protein which the B gene is assumed to produce will be completely altered from that point. This is supported by experimental evidence where a gene undergoing such a mutation lacks any function.

Dr. Crick and his colleagues postulate that a suppressor of *FC* O (say *FC* 1) is formed by deleting a base. So when the *FC* 1 mutation alone is present, all triplets right of *FC* 1 will be read incorrectly, and so the function of the gene will again be absent.

But when the two mutations are present in the same piece of DNA, as in the double mutant *FC*(O + 1), then although the arrangement of triplets between *FC* O and *FC* 1 will be altered, the first reading will be restored to the rest of the gene. Dr. Crick and his group claim to have convincing evidence that the coding ratio is three or a multiple of three.

They constructed triple mutants having two forms, involving either three additions or three deletions. Taking a triple set of mutants, all of the addition type, they found that either mutating singly or in pairs resulted in an absence of function in the gene. But when the three mutants were combined in the same gene, the function of the gene protein was restored or partly restored. This, the Cambridge workers claim, is what would be expected if the coding ratio were three or a multiple of three.

Ability to find the coding ratio, Dr. Crick and his group say, thus depends on the fact that at least one amino acid must have been added to or deleted from the polypeptide chain without greatly disturbing the function of the gene product.

**Degenerate Code?** Assuming a triplet code, then there are 64 (4 × 4 × 4) possible triplets. Dr. Crick's group found results which suggest that it is unlikely that only 20 of these represent the 20 amino acids, and the other 44 make no sense.

If this were so, they say, the region of the gene over which suppressors of the *FC* O family of mutants occur (about a quarter of the B gene) would be smaller than was observed. But it depends on the size of the protein which Dr. Crick and his colleagues assumed the B gene to produce.

But, they say, the length of the B gene suggests that the protein may have about 200 amino acids; so the code is probably "degenerate," or generally more than one triplet code for each amino acid. It's well known, the British scientists say, that if this were so, it would also account for the major dilemma of the coding problem—namely, that while the base composition of the DNA can differ in different microorganisms, the amino acid composition of their proteins changes by only a moderate amount. However, exactly how many triplets code amino acids and how many have other functions isn't known.

"Dr. Nirenberg and Dr. Matthaei's work at NIH," say Dr. Crick and his co-workers, "implies that a sequence of uracils codes for phenylalanine, and our work suggests that it is probably a triplet of uracils."

Dr. Crick and his group recognize that it is possible by various devices to synthesize polyribonucleotides with defined or partly defined sequences. If these will produce specific polypeptides, then they acknowledge that the coding problem is wide open for experimental attack. "If the coding problem is indeed three, as our results suggest," they say, "and if the code is the same throughout nature, then the genetic code may well be solved within a year."