EUROPEAN SERVICE

First Broadcast 22.1.62.
English Service
at 1800 GMT

Pre-rec: 16.1.62.
7½ TBU 135549
Dur :

THE FRONTIERS OF KNOWLEDGE NO.256

'Cracking the Genetic Code'

by Dr. Francis H.C. Crick, F.R.S.,
of the Department of Physics,
Cavendish Laboratory, University
of Cambridge.

Why do people resemble their parents? In a broad way we already know the answer. We all start from the union of one sperm cell with one egg cell, each of which carries about half the genetic information. Now there must be a lot of this information since we're very complicated objects. Not only do we have many different organs, such as eyes and ears and a liver, but each of these is made from special cells, in a very intricate way, and each cell is also in itself a very complex ~~complicated object.~~ So clearly there must be a very great deal of genetic information stored in both the egg and the sperm.

And yet it's been known for a long time that a sperm is a very small object. Its head, which carries the genetic information, is no more than one hundredth of a millimetre across. The actual genetic material weighs only a few times a millionth of a millionth of a gram. However the information is recorded, it must be written on a very small scale if so much is to be got into such a very tiny space.

The obvious conclusion is that the genetic information is likely to be written at the atomic level, in terms of atoms and small molecules, and indeed all the experimental work of the last ten years suggests that this is true.

However, a general argument of this kind can only be suggestive. We shall need many more detailed facts if we're to know just exactly how the genetic message is stored, and how it's read.

For some time we've had a good idea of the sort of way in which the message is stored. In most organisms it's written on long thin molecules known collectively as the family deoxyribonucleic acids, or DNA. DNA is a polymer. That is to say, it has a regular

25 Lines : Cont'd..

repeating backbone, with side groups called "bases" projecting at regular intervals. However, the bases are not all the same. There are four kinds of them, and the genetic information is conveyed by the precise order of the different sorts of bases along the DNA. In other words the genetic message is written in a language of four letters. Incidentally, the total length of the message, for man, is not short. It is probably more than a thousand million letters long.

But what does the message mean? How does it act? Here again we know the answer in broad terms. Each section of the ~~messenger~~ message, each paragraph we might say, controls the production of a particular protein. Proteins are large molecules and their great importance in the living cell is that they are the "catalysts", which decide which chemical reactions will take place. For every chemical reaction the cell carried out, there's a special protein which accelerates that reaction and that reaction alone.

Now all proteins are built on the same ground plan. Each is a polymer, a long molecule, again with a regular backbone (although quite different from the backbone of DNA), a backbone to which side groups are attached at regular intervals. However, in protein there are 20 kinds of side-groups, so that proteins, the machine tools of the cell, are written in a 20-letter language.

What we did not know, until recently, was how the cell managed the translation from the 4-letter language of the genetic material to the 20-letter language of the gene products, the proteins. Within the last six or eight months there's been a spectacular breakthrough in this problem and it now looks as if we shall know the answer within a year or so.

Although it's obviously important to know the details of the reading machanism - in other words to discover the various biochemical steps - and although we know quite a bit about some of these steps, especially those leading up to the actual act of protein synthesis, the part of the problem I want to spotlight is what one might call the "dictionary" used in the process. Each of the 20 units —

they're called amino acids, by the way - which go to make up the protein must be represented by a group of letters in the four-letter genetic language. Let's call these letters, A,B,C, and D, to make it easier. Then one amino acid, one of the twenty, might be coded by ABB, and another by CDA and so forth. So the first question we can ask is how many letters stand for one amino acid. It could hardly be only two, because there are only 16 possible different pairs of letters to be made from A,B,C, and D, whereas we need 20 combinations. So three is the minimum number. Actually, as we shall see, it looks as if three is correct, and that the genetic code has to be read off in triplets.

You may wonder why one can't get at the code in a different manner. One would need to know the amino acid sequence of a particular protein and also the base-sequence of the piece of DNA which carried the message for that protein, and then simply compare the two. But unfortunately this is not technically possible. In favourable cases, and with a lot of hard work, one may be able to find the sequence of amino acids in a certain protein, but it's almost impossible at the present time to find the base sequence of a piece of DNA, or even to obtain pure the particular piece one is after.

One indirect method is to make a chemical change in the DNA, and see how much of the protein is changed. This has been done recently, particularly by Dr. Wittmann at Tubingen, by using the plant virus, Tobacco Mosaic Virus. This has the related ribonucleic acid, RNA, as its genetic material. Using nitrous acid it's possible to alter just one base of the RNA, at random, and then to find how the protein produced by the virus, which makes up its shell, is altered. The typical change is to one amino acid only.

This result is important because it shows that one letter of the RNA only affects a single amino acid. It's quite easy to

30 lines (Cont'd.)

imagine a case where, say, the first three letters would code the first amino acid, and the second amino acid would be coded by letters 2, 3 and 4 in the sequence, not by no's 4, 5 and 6. Such overlapping codes now look highly unlikely.

But how do we know that a set of three letters codes an amino acid, and not a set of four, or five or any other number? Some very recent work by my Cambridge colleagues, Mrs. Leslie Barnett, Dr. Sydney Brenner, Dr. Richard Watts-Tobin, and myself have made it very probable that the correct answer is three. Our work was not biochemical but genetical, and was done for convenience on a virus, or bacteriophage as it is called, which attacks bacteria; the T4 virus. We had been studying the effects produced by a certain chemical which we suspected added or removed a base from the DNA of the virus. We found we had several different viruses which seemed to be altered in the same sort of way, but in different places in the DNA, all rather close together. For simplicity let's suppose that in each case a single base had been added to the DNA, but at different points.

Now by genetic methods we can put together viruses which have any two of these defects in their DNA, or even all three. If the virus had any of the single defects, or any of the double defects, it wouldn't grow on a particular strain of bacteria. The gene in which the defects occurred would not function. But if we put all three defects into one gene the function came back again.

We explained this as follows. We assumed that the code was read in groups of three letters, starting from a fixed point and working along three at a time. There is nothing in the string of letters to show where one triplet ends and the next begins. Now if a letter is added anywhere, the whole reading goes wrong from there on, since after the added letter, the letters fall into the wrong sets of three. Thus most of the message becomes

31 lines (Cont'd.)

nonsense. The effect is just the same if two letters are added. However, if three letters are added, although the message is wrong in that neighbourhood, the whole of the rest of the message now falls correctly into sets of three and makes sense again, so that in favourable cases the gene will work, after a fashion.

Our results also suggest that most of the 64 possible triplets stand for an amino acid, so that usually an amino acid can be coded by more than one triplet, but exactly how this is done we can't say.

However, the real break-through in the problem is due to two young biochemists at the National Institutes of Health (at Bethesda in America) called Nirenberg and Matthaei. This August, at the Biochemical Congress in Moscow, they reported a spectacular result. They were working on a cell-free system, made from broken bacterial cells, which will synthesise protein. This system has been developed in a number of laboratories. To this system they added a special RNA. RNA is a close relation of DNA and is believed to carry the genetic message from the DNA, in the nucleus, to the cytoplasm where the protein is actually made. The DNA is the master-copy and the RNA is the working copy - "DNA makes RNA and RNA makes protein" as the phrase goes.

The RNA they added, which was made by a special enzyme, was peculiar in that all the bases were the same. That is, the messenger read BBBBB... and so on, let us say. To everyone's surprise this simple repetitive message was translated by the fragments of the cell, and a protein was produced which had all its amino acids the same, so that it read ϒϕϒϒ... let us say. In other words the triplet BBB stands for ϒ. In strict chemical terms, three uracil bases code for one phenylalanine molecule.

This result has now been repeated in labs all over the world. Moreover, other artificial RNA's can be produced and tested on the system. Already very interesting results have been reported,

31 lines (Cont'd.)

especially by Ochoa and his colleagues in New York, and the general feeling is that with luck the whole code may be obtainable in this way.

We still don't know whether the code is universal. The same 20 amino acids are used in proteins throughout nature, from virus to man, but it is not yet certain that the same triplets code them in all organisms, although preliminary evidence suggests this is probable. If so, we shall have the key to the molecular organisation of all living things on Earth.

But on Mars, I wonder? Will there be life, or the remains of life, on Mars? And will it be DNA and RNA and protein all over again? The same languages perhaps, with the same code connecting them? Who knows?

E N D

12 lines

TOTAL 160 lines