CHAPTER III

GENERAL NATURE OF THE GENETIC CODE*

F. H. C. Crick, *University of Cambridge, Cambridge, England*

I have taken as my title "The *General* Nature of the Genetic Code" because we do not yet know the genetic code in detail, although we certainly hope to do so before very long.

We believe that genes are made of nucleic acid. It need not concern us at this stage whether it is DNA or RNA, double-stranded or single-stranded. We also believe that one of the main functions of genes is to determine the amino acid sequence of proteins. There may be other functions but this function is certainly an important one.

In addition, proteins have to take up their correct secondary and tertiary structure. I shall assume here without further discussion that the folding of a protein is mainly dictated by its primary structure.

The genetic code, then, covers the problem of how the sequence of bases in a particular piece of nucleic acid determines the sequence of amino acids in a particular protein. There appear to be four common bases in nucleic acid: guanine, adenine, thymine (or uracil) and cytosine. It is true there are a few rarer ones, most of which can be described as having additional methyl or hydroxymethyl groups. We cannot be sure that they do not have some important function in the genetic code, but this is not very likely.

It is an interesting speculation as to whether we could have nucleic acid with six bases; that is, three base-pairs instead of two base-pairs. Dr. Alex Rich, a little time ago, did make a suggestion along these lines (Rich, 1963), but we can say fairly categorically that at the moment we have no indication of a third base-pair in any biological material.

The polypeptide chains of proteins are made from a standard set of twenty amino acids. Again you could argue about the number twenty. It does not include hydroxyproline, phosphoserine, and certain other amino acids which are certainly found in true proteins. Nevertheless, for reasons which I have explained elsewhere (Crick, 1963), I think the standard set of twenty is probably the correct one.

I shall assume, therefore, that we have four standard bases in nucleic acids and twenty standard amino acids in proteins. Moreover the bases and

---

*An address presented before "The Robert A. Welch Foundation Conferences on Chemical Research. VIII. Selected Topics in Modern Biochemistry," which was held in Houston, Texas, November 16-18, 1964.

the amino acids are basically the same throughout Nature. It does not matter what organism one looks at — animal, plant, or micro-organism — one always finds in them the same standard set of amino acids and the same four bases.

It might be thought that solving the genetic code would be straightforward, but this is not so, since although it is possible, in favourable circumstances, to determine the amino acid sequence of a protein, it is at the moment technically almost impossible to discover the base sequence of the corresponding piece of nucleic acid. Consequently less direct methods than a detailed comparison of two corresponding sequences have to be used.

The first assumption I shall make is that the genetic code is fairly simple. In other words, that the relationship between the sequence of four things in the nucleic acid and the twenty things in the protein is not an elaborate one but is relatively straightforward. The main justification for this is that the code has to be translated by biochemical machinery, and although we know that this is fairly complicated we do not think it is exceedingly ingenious and intricate. I shall assume that for each amino acid there will be a small group of bases on the nucleic acid which codes it, and I shall call this group a "codon".

All codons cannot consist of just two bases, because we have only four alternatives in each position, so we can only construct a maximum of sixteen different pairs of dinucleotides, whereas we need twenty for the amino acids and certainly one or two more in addition as punctuation marks. Groups of three bases would provide sixty-four possible codons, and this is now believed to be the correct answer.

It has been widely assumed that the genetic message and its translation were co-linear. That is, that the order of the codons on the nucleic acid was the same as the order of the corresponding amino acids in the protein. Until recently there was very little evidence for this, although it was confidently predicted that it would soon be found.

Strictly speaking we still do not have such evidence but it has been shown recently by two groups of workers that the genetic map of a gene is co-linear with the amino acid sequence of the protein controlled by that gene.

A genetic map can be constructed by the standard methods of genetics. The essence of the matter is that two different versions of the same organism, which differ in having mutations in somewhat different places, are mated together in some way so that genetic recombination takes place. It is found that when these two differences are far apart on the genetic map recombination is common: on the other hand, when these differences are close together,

recombination is rather rare. By this method, and extensions of this method, it is possible to put genes in order on the chromosome and in more recent years, using micro-organisms, to put different sites within a single gene into a linear order. The classic work on mapping within a gene is that of Benzer on the rII locus of bacteriophage T4 (Benzer, 1959 and 1961). In these genes he found several hundred different sites and by an ingenious method (called deletion mapping) he was able to show that they could be put in a linear order on the genetic map.

Dr. Yanofsky and his colleagues (Yanofsky *et al.* 1964) have worked on the A protein of the enzyme tryptophan synthetase of *Escherichia coli.* A series of mutants was picked up, mapped genetically and thus put into a definite order. The corresponding protein was isolated, purified and part of its amino acid sequence determined. For each mutant the amino acid alteration was pinpointed in the sequence. As was expected the two orders, that of the mutants on the genetic map and that of the corresponding alterations in the polypeptide chain, were the same.

Shortly after this work was published some colleagues of mine at Cambridge, Dr. Sarabhai, Dr. Brenner and Dr. Stretton together with Dr. Bolle from Geneva (Sarabhai *et al.* 1964) also produced some evidence which led to the same result.

They used a special sort of mutant called "amber" mutants. Such mutants are found in all sorts of genes, not only in the genes of the virus, but also in the genes of the host. They are characterized by the fact that they can be suppressed. They were able to show that what an amber mutant does is to terminate the polypeptide chain at a particular point. In other words, instead of producing a complete polypeptide chain the cell makes the protein (from the amino end), but only up to a certain point, and then releases it.

Using proteoytic enzymes they chopped the polypeptide chain into a number of peptides and they looked to see, in each case, whether certain peptides were made or not. By this means they were able to tell, for each mutant, whether it made a long polypeptide chain or a short one, and so they could put all the mutants in order, using only the methods of protein chemistry. At the same time they constructed a genetic map, using only genetic methods, and found, as expected, that the two orders were the same.

We thus have two examples to make us feel reasonably happy that the gene and the protein it controls are colinear, and I think that in the course of time more evidence, from different proteins, in different organisms, will accumulate.

It was thought at one time that the genetic code might be of the overlapping type. That is, that a particular base might belong to more than

one codon. This now seems very unlikely, since almost all mutants which have been studied show a change of only a single amino acid. If the code were an overlapping one we should certainly expect to find mutants which were changed in two or three (adjacent) places in the polypeptide chain. The evidence comes from the work of Dr. Yanofsky already mentioned and from studies on abnormal human haemoglobins and the protein of Tobacco Mosaic Virus (Wittmann and Wittmann-Liebold, 1963; Tsugita, 1962.)

It is not yet certain that the bases making up a codon are adjacent on the polypeptide chain, although we have no reason to think they are not (Dr. Stent, in discussion, pointed out that the recent biochemical work of Leder and Nirenberg (1964) makes it very likely that codons consist of three *adjacent* bases).

If the code is of the non-overlapping type there must be something which decides how the message is split up correctly into groups of three bases. Otherwise the message might be read "out of phase" and thus be totally incorrect. There are, of course, two ways of reading a triplet code out of phase, one corresponding to beginning one base too early and the other to beginning one too late. From evidence (Crick *et al.* 1961) obtained by studying two adjacent genes joined together by a genetic deletion we believe this is done in the most obvious way. That is, by beginning at some fixed point in the sequence and reading along from there in steps of three bases at a time.

This genetic work also suggested very strongly that codons consist of three bases, rather than four or five or some other number. The evidence was obtained using acridine mutants of the rII locus of phage T4. These mutants behave as though they were due to the addition or subtraction of a base (or bases) in the genetic message, rather than to one base being changed into another, though this interpretation will be more certain when it becomes possible to study the amino acid changes produced by such mutations.

It was found that when three mutants of like type (each the addition of a single base, say) were combined by genetic methods into one gene, the gene still functioned, whereas the same mutants, either singly or in pairs, made the gene completely inactive. Such results are most easily understood if every codon consists of three bases. The recent biochemical work of Leder and Nirenberg (1964) appears to confirm this.

Our recent more detailed knowledge of the genetic code has come from experiments using cell-free systems which synthesize proteins in the test-tube. The advantages of such systems are not only that the separate steps of protein synthesis can be more easily studied in this way, but also that one can add molecules to them which cannot penetrate easily, if at all, into intact cells. The system most used is obtained from *E. coli*, though other sources, including mammalian ones are also employed. In outline it consists of ribosomes, S-RNA molecules, activating enzymes and various "soluble" factors, which are probably protein, in addition to amino acids, ATP (as an energy source) and GTP. The recent review by Watson (1964) should be consulted for details.

In the cell the genetic information in DNA is copied onto single-stranded RNA by the enzyme RNA polymerase, and the messenger RNA (m-RNA) thus produced is then "read" by the ribosomes. In the cell-free system various artificial m-RNA molecules can be added and their effect on amino acid incorporation studied. The break-through came when Nirenberg and Matthaei (1961) showed that polyuridylic acid, used as m-RNA, promoted the incorporation of phenylalanine into polypeptide chains. Thus one codon for phenylalanine is presumably UUU. Subsequent work mainly by Nirenberg and his colleagues and by Ochoa and his group have shown that polyadenylic acid codes lysine, and that polycytidylic acid probably codes proline. In addition by using mixed polynucleotides, with random or semi-random sequences, it has been possible to allocate a fair number of the 64 possible triplets to the various amino acids, though triplets high in guanine have proved difficult to study, since polynucleotides tend to have too much secondary structure if their guanine content is high. Of course by these methods alone it has not been possible to obtain the *order* of the three bases in a particular codon, but only the composition of the codon. Two reviews (Nirenberg *et al.* 1963; Speyer *et al.* 1963) provide a good summary of the results obtained by these methods.

Among other things the evidence strongly suggested that there is, in general, more than one codon for each amino acid; that is, that the code is "degenerate". It is suspected that there is one type of S-RNA for each codon, but this has not yet been proved rigorously.

It is not yet known whether the genetic code is identical in all living things, but preliminary studies (see the article by Weinstein, 1963) show that it is certainly rather similar in a number of rather widely separate organisms.

In order to obtain the order of bases within any codon it is necessary to use polynucleotides having their bases in a known order. Khorana's group is attempting to do this by synthesizing polynucleotides with repeating sequences. Doty's group is also studying polynucleotides with known or partly known sequences. Meanwhile Nirenberg has hit upon an ingenious method for obtaining codons using trinucleotides. Only preliminary results have so far been published (Leder and Nirenberg, 1964) but it is known that Nirenberg and his colleagues have already determined many codons in this way and the results will probably be in print before this article appears.

The method depends upon the binding of specific S-RNA molecules to ribosomes in the presence of a particular trinucleotide. Thus GUU will help to bind a species of valine RNA, whereas the trinucleotides UGU and UUG will not. (The convention is that the 5′ position is written on the left.) It is not yet clear how clean and reliable this method will become, but it is certainly one of great promise.

Even when the majority of codons have been obtained it will be necessary to confirm them by quite different methods to make sure that no artifacts have crept into the cell-free system. In particular the code should explain the various amino acid changes produced by single-step mutations, of which many different types have already been observed. Eventually this should enable us to see how various chemical mutagens act, as direct chemical methods may sometimes be misleading.

There is also additional evidence to be satisfied. Dr. Yanofsky, for example, has been able to obtain genetic recombination between bases within the same codon. In particular he has crossed a mutant having arginine at a particular place with one that has glutamic acid there, and has obtained a genetic recombinant having glycine at that point, which was the amino acid there in the wild type (Yanofsky, 1963). Any allocation of codons should be able to explain such a result.

We shall also need to know the so-called "nonsense" codons, though my own belief is that all 64 codons will have some function or another. My colleagues at Cambridge are studying the codon associated with the "amber" mutants, which, you recall can in certain cells terminate the polypeptide chain. They know that this mutant has to be read "in phase", and that it is related, by single-step mutations, to glutamine, tryptophane and tyrosine. When it is suppressed the amino acid inserted is serine. This and other evidence has led Brenner and his colleagues to suspect that the amber triplet is UAG, and that the related nonsense triplet (now called an "ochre" mutant) is UAA. Whether these are the usual triplets for chain termination, and whether there are any more of them remains to be seen.

When the details of the code are fully established we shall want to examine its general structure. By this I mean the way in which codons are allocated to the different amino acids, and in particular the relationship (if any) between the various codons for any one amino acid (see for example, Eck, 1963, and the review by Lanni, 1964). Is there something in the nature of, say, leucine which makes it inevitable that it will have certain codons, or is the present code the result of a series of historical accidents? In the former case it should be possible to deduce the whole of the code from theoretical principles. In the latter case this would be impossible.

This will certainly lead us to the question of how the code originated. It is not too difficult to see how nucleic acid synthesis could arise, but it is almost inconceivable that the elaborate machinery for protein synthesis arose all in one step. It must have evolved from something simpler. We do not know whether in the primitive system there were all twenty amino acids, or even all four bases. Perhaps only one pair of bases was present at the start.

The crucial point in all discussions about the origin of life is to see at which point natural selection could have begun to operate. For that you must have a replicating system of some kind, although what it was we can, at this moment, only guess.

However, our more immediate aim should be to determine unambiguously the complete code for at least one organism, say *E. coli*. We are likely to have a moderately accurate, moderately complete version of the code before long, but it is important that each triplet should be established beyond reasonable doubt. This will open the way to many other problems, of which the biochemical nature of control mechanisms is probably the most important.

REFERENCES

Benzer, S. (1959) Proc. Nat. Acad. Sci. U.S. **45**, 1607.
Benzer, S. (1961) Proc. Nat. Acad. Sci. U.S. **47**, 403.
Crick, F. H. C. (1963) in Progress in Nucleic Acid Research, Ed. J. N. Davidson and W. E. Cohn, **Vol. 1**, Academic Press, Inc., New York.
Crick, F. H. C., Barnett, L., Brenner, S. and Watts-Tobin, R. J. (1961) Nature, **192**, 1227.
Eck, R. V. (1963) Science, **140**, 477.
Lanni, F. (1963) in Advances in Genetics, **Vol. 12**, Academic Press, New York.
Leder, P. and Nirenberg, M. W. (1964) Proc. Nat. Acad. Sci. U.S. **52**, 420.
Nirenberg, M. W., Jones, O. W., Leder, P., Clark, B. F. C., Sly, W. S. and Pestka, S. (1963) Cold Spring Harbor Symp. Q. B. **Vol. 28**, p. 549.
Nirenberg, M. W. and Matthaei, J. H. (1961) Proc. Nat. Acad. Sci. U.S. **47**, 1588.
Rich, A. (1963) in Horizons in Biochemistry, Ed. A. Kasha and B. Pullman, Academic Press, New York.
Sarabhai, A. S., Stretton, A. O. W., Brenner, S. and Bolle, A. (1964) Nature, **201**, 13.
Speyer, J. F., Lengyel, P., Basilio, C., Wahba, A. J., Gardner, R. S. and Ochoa, S. (1963) Cold Spring Harbor Symp. Q. B. **Vol. 28**, p. 559.
Tsugita, A. (1962) J. Mol. Biol. **5**, 284.
Watson, J. D. (1964) in Volume du cinquantenaire de la société de chimie biologique, Maurice Declume, Lons-le-Saunier.
Weinstein, I. B. (1963) Cold Spring Harbor Symp. Q. B. **Vol. 28**, p. 579.
Wittmann, H. G. and Wittmann-Liebold, B. (1963) Cold Spring Harbor Symp. Q. B. **Vol. 28**, p. 589.
Yanofsky, C. (1963) Cold Spring Harbor Symp. Q. B. **Vol. 28**, p. 581.
Yanofsky, C., Carlton, B. C., Guest, J. R., Helinski, D. R. and Henning, V. (1964) Proc. Nat. Acad. Sci. U.S. **51**, 266.

INTRODUCTION AND DISCUSSION

Dr. Stanley: Ladies and gentlemen, we are ready to begin our afternoon session with another real treat in store. Francis Crick, so far as I know, is not particularly trigger-happy, but that word came to my mind when I was thinking about a terminology that would describe his mind. I must say, I have never met anybody who was so quick mentally as I've known Francis Crick to be over the years.

I imagine this is what paid off a few years ago when Jim Watson came to Cambridge for a little stay, and it was the interplay of these two minds that resulted in a conception which I regard as being one of the truly great ones of modern times, and that's the Watson-Crick structure, the double helical structure, for DNA. It certainly has had a tremendous impact upon the minds of mankind, and scientists, of course, particularly. There is an extraordinary amount of work which has stemmed from this concept.

Dr. Crick came into this work as a physicist. Dr. Watson, who, unfortunately, is not here, came in as a biologist, a microbiologist. He received his degree with Dr. Luria. So I think the two were ideally equipped to deal with the problem of the structure of DNA at that time.

Dr. Crick has ceased to become a straight x-ray man. When I was in his laboratory in 1961 for a few weeks, he was very busy pouring bacterial plates hour after hour, sometimes until the wee hours of the morning. He was studying acridine mutants in an attempt to establish the mode of reading of the DNA. Nobody watching him in those days would have thought that he had started out as a physicist, because he was an out and out biologist at that time.

Well, I need not tell you anything more about Dr. Crick. You have his biography here in the booklet. I shall now call on Dr. Crick to give his discussion of the "General Nature of the Genetic Code":

[The following discussion took place after Dr. Crick presented his address.]

Dr. Stanley: Thank you very much, Dr. Crick, for this quite fascinating exposé on the "General Nature of the Genetic Code." I should like to make a comment or two about the work that has gone on in the Virus Laboratory with tobacco mosaic virus. The active portion is, of course, an RNA, and it can be regarded as a messenger RNA. Tobacco mosaic virus, a nucleoprotein, has been available for years in pure form and from this a quite pure active RNA can be obtained. This viral RNA carries the message or information for making the protein overcoat component of the virus.

One of our objectives, which was started some years ago, was to obtain

the amino acid sequence of the protein of tobacco mosaic virus. This work was stimulated by Sanger's work on insulin. And as a result, about 1960, I think it was, we obtained the complete amino acid sequence of the 158 amino acids of the protein component of tobacco mosaic virus.

Then Schuster in Tubingen, Germany reported that you could convert cytosine to uracil in RNA through deamination. A great many nitrous acid mutants of tobacco mosaic virus RNA were made in our laboratory, as well as in the Tubingen Laboratory. This gave an opportunity, in a preliminary way, to make some of the verifications that Dr. Crick has mentioned relating triplets to amino acids.

We have, between the two laboratories, I think around 50 or 60 mutants, where the exact amino acid can be pinpointed in the 158 amino acids, that has been changed, presumably by virtue of a C to U conversion.

The difficulty with tobacco mosaic virus is that you cannot conduct genetic experiments; you cannot do the necessary crossing experiments. This was the reason I am particularly pleased with the recent work of Yanofsky of Stanford, for he combined chemistry and genetics in the very beautiful way that Dr. Crick has indicated. I had hoped that Dr. Yanofsky would be here, but as has already been indicated, he stays in the laboratory at work and doesn't tend to come to meetings.

Biochemists are not nearly as backward as our friend, the physicist-turned-molecular-biologist, would indicate. I happen to know a graduate student who, two or three years ago, had separated and purified 15 or 20 of these triplets obtained from the hydrolysis of yeast nucleic acid.

Unfortunately, things don't move quite as rapidly as sometimes the minds of some people move, but I am sure that all 64 triplets will be available eventually, and means will be found to test their activity in amino acid incorporating systems. We already, for example, know much about the terminal nucleotides in the tobacco mosaic virus nucleic acid and that may provide useful information.

Now, in view of the vast amount of work that Dr. Ochoa has done on the genetic code, I wonder if he would care to make a few comments, or ask a question.

Dr. Severo Ochoa (Speaker), *New York University School of Medicine:* It was, as usual, a fascinating presentation by Francis Crick, which we all have enjoyed very much. Now, as Dr. Stanley mentioned, a number of trinucleotides are available. They have been isolated from ribonuclease digests of RNA by Wendell M. Stanley, Jr., who is a post-doctoral student in our laboratory.

We have not looked into the binding of aminoacyl-sRNA's by ribosomes in the presence of nucleotide triplets very thoroughly, but as far as we went, we had no difficulty in confirming Nirenberg's results with ApApA and UpUpU. One does get specific binding of lysyl-sRNA and phenylalanyl-sRNA, respectively, to the ribosomes. With ApApU we had negative results with each of three aminoacyl-sRNA's, one of which would have been expected to be specifically bound by ribosomes in the presence of this triplet. But, I repeat, we have not tried very extensively.

The extent to which binding occurs is quite small. We have made a few trials (using ApApA and UpUpU) with ribosomes from bacteria other than *E. coli* but the binding was not better.

The cell-free *E. coli* system, referred to by Crick as purified, is really very crude. Among other things it contains large amounts of nucleases that cleave the messenger RNA, whether natural or artificial. The main culprit is an exonuclease which apparently cleaves polyribonucleotides from the 3′-end releasing 5′-nucleoside monophosphates.

We put triplets of known sequence at the 3′-end (the right-hand side end of the polynucleotide in the conventional way of writing these compounds). The terminal base in these triplets had radioactive label. For example, we had a poly ApApA . . . pApAC* in which C was labeled with C^{14} or H^3. When this polynucleotide was incubated with the system of *E. coli* supernatant and ribosomes, the radioactivity became acid-soluble exceedingly rapidly. This explained why we would not obtain any specific amino acid incorporation (other than that of lysine) with such a polynucleotide. We have, therefore, put some effort into getting a system which might be reasonably free of exonuclease and of the so-called latent ribonuclease. Albert Wahba and Wendell Stanley have to a great extent succeeded in doing so by using a purification procedure of the *E. coli* ribosomes which removes contaminating ribonuclease and exonuclease. The purification involves treatment of the ribosomes with ammonium chloride solutions overnight and chromatography on DEAE-cellulose columns. These so-called "clean" ribosomes are essentially nuclease-free.

We next tried to purify the *E. coli* supernatant, which contains 75 per cent of the exonuclease of the whole extract. However, to remove the nuclease leaving behind the 20-odd amino acid activating enzymes as well as the transfer enzymes, is not easy. We only made a few unsuccessful attempts and rapidly dropped this approach. We had an idea that maybe organisms other than *E. coli* might be free of exonuclease. The first one we tried, *Lactobacillus arabinosus*, proved to meet this specification. The *L. arabinosus* supernatant proved to couple fairly well with the purified *E. coli* ribosomes to give amino acid incorporation into polypeptides, at least with synthetic

polynucleotides such as poly A, poly U, poly AU, and so forth. With this system poly ApApA . . . . pApApC*, which stays intact during the incubation, leads to the incorporation of lysine and asparagine. The ratio of asparagine to lysine incorporated with a polymer of about 20 nucleotide residues was in the neighborhood of 5 to 6 per cent. These results tentatively assign the sequence AAC to one of the asparagine triplets. Experiments are in progress to determine whether the asparagine thus incorporated into short lysine polypeptides is C-terminal or N-terminal. This should definitely answer the question as to the direction of translation of the message.

The combined purified *E. coli* ribosomes-*L.arabinosus* system may be a good one for the study of the synthesis of natural proteins with natural messengers by cell-free systems. Since the combined system has little or no nuclease activity the messengers should remain intact. No matter what one does, one usually first runs into complications. Our combined "clean" *E. coli* ribosomes-*L.arabinosus* system did not work with a natural messenger, namely MS2 phage RNA, but neither did the usual system of (non-purified) *E. coli* ribosomes + *E. coli* supernatant. However, as used by other people, an extract obtained by centrifugation at 30,000 *g* (the so-called *E. coli* S-30 extract) was fairly active with MS2 RNA. All this goes to show that a number of difficulties will have to be overcome before one obtains systems purer than are now available for the study of protein synthesis *in vitro*.

Dr. Stanley: Thank you, Dr. Ochoa. Do you want to comment now, Dr. Crick?

Dr. Crick: I was very interested to hear about your poly A with a C at the end, because we tried this and got very funny results. It looked as if the message slipped.

As to your comment on UUA, let me just say there are reasons to suspect, on genetic and mutagenic grounds, that not all the mixed U and A triplets stand for amino acids, in spite of the result with the poly UA. We have a strong suspicion that chain termination may occur with a triplet containing only U's and A's. We would be very interested to know which triplet you failed with, because that might be the one we are particularly interested in.

Dr. Ochoa: I believe the one we tried was AAU not UUA. Your indications that one of the triplets having U and A does not stand for an amino acid are interesting because, for a long time, we thought that in the case of poly AU we had a polymer which contains hundred per cent sense, that is, all of its triplets code for one amino acid. I was, therefore, interested to hear what you had to say in this respect.

Dr. Crick: Bretscher, when he was in our laboratory, obtained some (unpublished) evidence about this. He looked into the question of what makes the polypeptide chain terminate. It is well known that if you use poly U as a messenger, the polyphenylalanine produced is not released but is still bound to the soluble RNA. To a large extent this is also true if you use poly A as a messenger. Although short oligolysines are produced, they are still bound to soluble RNA. He, therefore, used poly AU as a messenger and obtained a release of up to 50 per cent, although it was not very reproducible.

Dr. Stanley: Dr. Rich, I think you had a comment.

Dr. Alexander Rich (Discussion Leader), *Massachusetts Institute of Technology:* At the beginning of his lecture Dr. Crick mentioned the fact that DNA is made of four bases which form two base pairs; toward the end of his lecture he referred to the fact that in the early history of life, it is possible that only one base pair was found.

In fact, only one base pair is needed, because this is able to carry enough information, to make a very complex molecule. It is like a molecular Morse Code using, instead of dots and dashes, the two bases of the pair.

We may ask if it is possible to have a DNA-like genetic code involving more than two base pairs. To answer that question one has to specify several relations.

First of all, we believe that one of the reasons that these particular base pairs are used in DNA is the fact that both pairs can fit into the double helix in either orientation, either A and T, or vice versa, with the same geometry and the same angles between the glycosidic bonds going to the chains.

Now, if we restrict the question to a molecule of the same geometry, there is a possibility of using one other base pair. Only one additional base pair would have the form which would allow it to fit into the DNA helix and still have specificity between its hydrogen bonds. This involves a pair of bases which are closely related to the guanine-cytosine pair. These are the bases isoguanine and isocytosine. Whereas guanine has a carbonyl oxygen at position 6, and an amino group at position 2, isoguanine has these reversed. Likewise, cytosine has an amino group at position 4 and a carbonyl group at position 2, while these are reversed in isocytosine. These would form a hydrogen bonded pair which could fit into the DNA helix if the tautomeric form of isoguanine has a proton on position N_1.

So, in principle, there could be another pair that might fit into the DNA molecule. Now, we don't know whether this can, in fact, be done.

We are trying to tackle this problem experimentally at the present time by making essentially methylated or ethylated derivatives of isoguanine and isocytosine. We are attempting to co-crystallize these together to see if they will in fact form this hydrogen bonded pair. If we can do that, then we will try to introduce them into the DNA molecule.

Dr. Stanley: Thank you, Dr. Rich. Are there other comments? Dr. Scheraga, we are glad to see a non-nucleic-acid chemist join us.

Dr. H. A. Scheraga (Discussion Leader), *Cornell University:* I would like to ask a question about the role of binding in Nirenberg's experiments, that is, whether binding is really the only thing that is involved?

If binding were all that is involved, this would imply that a triplet on the S-RNA sees a triplet on the messenger, and hooks on without the amino acid knowing anything about it. Yet you cited an experiment in which the S-RNA took the amino acid to the site on the messenger, but it took the wrong amino acid to the site after the amino acid was chemically modified. This would imply that the amino acid doesn't know anything about the binding properties.

On the other hand, mustn't the S-RNA come off the messenger; therefore, doesn't this imply that there may be a conformational change involved?

Dr. Crick: Of course the S-RNA must come off, but the only signal that may be necessary for it to do so is the breaking of the link between the amino acid and the S-RNA. Suppose we had to start from first principles. It is obviously necessary to reject the S-RNA when you have finished with it, but it is not at all clear how it is done. There is no reason why the *triplet* has to see the amino acid to do that, though perhaps the ribosome has to be aware of it in some way.

However it need not be the side chain of the amino acids which is recognized, but the part which is common to all amino acids. Consequently, changing one side chain into another may make no difference, because it has got to be a mechanism which is going to work for all amino acids.

Dr. Stanley: There is also an unfortunate lack of specificity in the Nirenberg experiments, too. Isn't that right? I mean, the background is pretty high, isn't it?

Dr. Crick: I think we should be cautious about this new technique. Even if it does work fairly well it will still need confirmation. I quite agree that we need rather better discrimination. Even if you know some of the answers in advance, and can check the method because you know the answers, you have got to find a method of checking it without knowing the answers.

Dr. Stanley: Which is a little difficult!

Dr. Ochoa: I think the question raised by Dr. Scheraga, on the release of the polypeptide chain when it is finished, is a very interesting one. It would seem that there must be an enzyme that does this job, that is, cleave the last aminoacyl-sRNA bond when the chain is completed. Now, as Dr. Crick mentioned, there may be at the end of the messenger, or in intercistronic positions in polycistronic messengers, special release triplets. It may be that when the polypeptidyl sRNA chain hits that spot it comes off and now becomes accessible for the activity of the sRNA release enzyme.

In regard to the universality of the genetic code, I believe Dr. Crick forgot one of the very dramatic experiments that bear upon this point; the experiments of Pamela Abel in which she was able to infect *B. subtilis* with vaccinia virus, so vaccinia virus multiplied inside an "unnatural" host. This is the sort of thing one would expect from universality of the genetic code. If a messenger nucleic acid, viral nucleic acid let us say, can get into a cell, any cell, it will give rise to the production of virus in a cell which is not normally its host. The selectivity of a host is probably at the level of the entry of the nucleic acid into the cell, and that may be why specific viruses with specific hosts.

Dr. Stanley: I think one of the amazing things of the past few years has been the discovery, which Dr. Ochoa just mentioned, of Pamela Abel and Thomas Trautner on vaccinia virus growing in *B. subtilis.*

There are increasing evidences, as people work with viruses, that virus nucleic acids will now grow or replicate or at least give a one-growth cycle, in a whole variety of cells, whereas just a few years ago we hadn't the vaguest idea that these viruses had such a potential. And now, with the isolation of several biologically active viral nucleic acids and the preparation of protoplasts, you have great opportunities.

It is my own personal feeling that this whole area is just beginning to be explored. It is really in its infancy. I think we are in for some real surprises about nucleic acids of one kind, growing in a whole variety of cells. And this will be very interesting, as are these experiments which have already been referred to.

Dr. Crick: I am sorry I didn't mention Abel and Trautner, but I was pressed for time at the end of my talk. I do agree with both of you on the importance of it.

I would say—going back to chain termination for the moment—that I would be surprised if there were not an enzyme involved, although we have discussed how it could be done without an enzyme. I think the interesting question is whether there is a special soluble RNA for chain termination, what it will be like, and how you would look for it.

Dr. Gunther S. Stent (Discussion Leader), *University of California:* I would just like to point out that, if Nirenberg's experiments on the ribosomal binding of transfer RNA by trinucleotides are correct, the interlacing codes, of which you said that they are still possible, would be eliminated because under such codes nucleotides no smaller than pentanucleotides would be required for the binding reaction.

I would also like to point out that whereas the code appears to be universal, or quasi-universal, the elements that are responsible for regulation of enzyme synthesis are not. Experiments have been carried out in which both structural and regulatory genes of some enzymes in one bacterial genus were introduced into the cells of another genus, to produce a kind of partial hybrid diploid or heterogenote. In these intergeneric heterogenotes, the regulatory properties of the enzymes carried by the foreign genetic element are completely upset, showing that some specific, non-universal cytoplasmic elements participate in these regulatory processes.

Dr. Crick: I think the question is whether any universal base sequence is involved. We all suspect there may be proteins involved in regulation and these might possibly have unusual features. There may be a special base sequence which occurs at the beginning of all genes, saying where to start reading. But in addition, there may be a special mechanism which turns on and off that particular gene.

Dr. Stanley: Are there other comments or questions from the audience in general that you would like to put to Dr. Crick?

Dr. F. R. Duke, Texas A & M University: Now, the only way that I can see that you can have these little lines drawn every three times in these polycodons is to start at one end. And in that case it would seem that the synthesis of these proteins would be rather slow, because you would have to put in one amino acid, then the next and the next and the next, and I wonder if such rate studies have been done, and whether or not they are reasonable on the basis of this one-at-a-time mechanism?

Dr. Crick: The evidence the proteins are synthesized sequentially from one end, one amino acid after the other is really pretty good, and shows that it starts at the amino end.

So the next question you can ask is: Do you accumulate the amino acids a little ahead of the growing point? You can show quite easily this will give you an advantage in rates. If you accumulate 10 positions ahead you can go up to a factor of about four and a half faster. There isn't any evidence that this happens, though I wouldn't say that the present evidence really rules it out.

Given, therefore, that the protein is synthesized from the amino end one step at a time, you can ask whether the reaction rates are reasonable. I think that you can't make a dilemma to show that the rates are totally unreasonable, since one doesn't know enough about the factors involved. In fact, you don't make proteins terribly fast. What you do make terribly fast is nucleic acid.

Dr. Duke: Did you speculate that some of the triplets or codons not necessary for fixing the amino acid sequence might be intermediate starting points in protein synthesis?

Dr. Crick: Yes. This is one of the speculations. That is what I meant by saying there might be a triplet for starting a polypeptide chain. That was exactly what was implied. Whether there is such a thing, I wouldn't like to say. I suppose I hear the thing discussed, roughly, every other day in the laboratory over a period of six months, and I know we haven't reached a conclusion yet. So I think everyone would agree the evidence is more suggestive than conclusive, according to which way you take it, one way or the other.

Dr. Henry Eyring, University of Utah: I am afraid mine is a long involved question with no answer. Apparently there is no direct connection in the sense of hand-in-glove relation between the DNA carrying the genetic code and the amino acids. Such a relation seems to be ruled out by the things that you have told us.

The other point that seems to be true is that the genetic code is a universal language, or at least this is probably true. This has interesting implications. Thus, the development of the code probably came after the development of the individual enzymes that activate the amino acids prior to their incorporation into the proteins.

There is one other point I would like to raise. We never find anything but the *L*-amino acids incorporated in the proteins of any known living species. Now in order for the activating enzyme to distinguish between *L* and *D* amino acids it must fit tightly against the *D* hydrogen of the *L* amino acid. Such universality seems to bespeak an early common origin of this selective property.

Now, the only way that you can have the exclusive use in proteins of *L*-amino acids persisting during the entire evolutionary process is through a dynamic stability which operates in spite of mutations. One must suppose that the spasmodic mutations that occur never can produce individuals that can successfully compete. Competition would only be possible if there were simultaneous wholesale mutations in the organism and its environment.

I finally come to my point. Apparently we haven't yet reached the stage where we can make very much out of the relative rates on the basis of how

substrate and enzymes fit in selecting the amino acids. The genetic code is apparently a completely learned language presumably developed in response to natural selection. If so, it becomes all the more remarkable that there has been no confusion of tongues during the whole course of evolution, and that we should end up with a universal or nearly universal genetic code for living things. I have said several things, I would like to be corrected at any point where you think I'm obviously wrong.

Dr. Crick: To take the question of handedness. I think you can ask first, why should there not be mixed *L* and *D* amino acids? The answer I would expect is that the mechanism is such that at some stage stereochemistry comes in, and since it's a repetitive process you want them all to be the same.

The next question you can ask is: Why are they all *L* amino acids? Why couldn't they all be *D*? You can look at protein and ask: Would it make much difference to the way proteins fold up, if you folded up alpha helices right or left handed? It turns out that there is not *too* much difference between making a lefthanded one and a righthanded one.

However, when we look at nucleic acid the position is quite different. You cannot easily make a lefthanded DNA structure using the present handedness of sugar. This is because the sugar has got a number of asymmetric carbon atoms, which are a very integral part of the backbone, so there is a very strong handedness imposed in nucleic acid.

One has a feeling that if the choice between *L* and *D* originated by chance, it operated on the nucleic acid rather than the protein, although I could not defend this point if I were pressed hard.

The evolution of the code I don't understand. It would be very easy to understand if there were a structural reason for it all. If it isn't that, if it were a series of historical accidents, then we and the whole of nature have probably all descended from one small population of organisms, all of which had the same code. What isn't easy to see is the stage just before that. Why didn't the code diversify? You can easily set up mathematical equations to show that in a siutation like this one version of the code will get the upper hand, but this simplified picture is not always borne out if you look around you. It's not true to say that the reptiles have all disappeared because the mammals are superior. Of course, if at each step there is a clear selective advantage to one choice or the other there would be no problem. But I find it just a little difficult to believe that, that if you started with, say, half a dozen codons, (which might be determined structurally), and expanded to sixty, taking about fifty or so distinct steps, at each of which you added on a new codon, that when there were two alternatives, one was selected and the other was completely eliminated by competition.

So there are the problems, and I personally have left them a little on one side until we know the structure of genetic code, since that may give us a hint one way or the other.

Dr. Eyring: I have the impression that the development of secondary structure greatly stabilized protein molecules. The irreversible destruction of enzymes frequently involves a big activation energy. For example the irreversible inactivation of the enzyme luciferase, important in bacterial luminescence, occurs via a reaction having an overall activating energy. This fantastically large inactivation energy is made up of a large heat associated with the preliminary reversible denaturation plus an additional modest heat of activation of the reversibly denatured molecule. The reversible denaturation presumably involves the disappearance of some tertiary and probably some secondary protein structure. Analogously, I wonder if the ability of nucleic acids to last depends on whether or not they can protect themselves by cooperative hydrogen bonding.

Dr. Crick: Well, I think it is a protection in the present context because you tend to get enzymes developing that chop them off, but I don't know any other reason.

Dr. Eyring: But is unfolding a protection for enzymes?

Dr. Crick: Now, I wouldn't believe that chemical loss would be affected very much by whether, let us say, DNA was single stranded or double stranded. But you don't get a chemical breakage. All you are saying is that there is a cooperative effect in destroying the tertiary structure.

Dr. Eyring: Excuse me. In luminescent bacteria there is an activation energy of 91 kilocalories for destroying luciferase. Most of it is the activation energy for destroying structure, that is, of reversible denaturation. You can inhibit the enzyme very easily with sulfanilamide, for example, and do no harm to it, because this inhibition is not due to reversible denaturation. But if you inhibit luminescence with something like alcohols or ethers that make it reversibly denatured you get an accompanying rapid destruction.

Thus the structure that is lost in reversible denaturation is a very great protection against subsequent reactions which irreversibly destroy the enzymes activity.

Dr. Crick: Well, I think I agree. The tertiary structure, of course, stabilizes the tertiary structure. But I don't see what else you are saying.

Dr. Eyring: Oh, excuse me. Then I haven't made myself clear. The irreversible destruction is something that is in addition to and only follows after reversible denaturation.

Dr. Crick: I agree. What's the evidence that you break the backbone?

Dr. Eyring: Whether the irreversible reactions involve something like breaking S-S linkages or breaking the backbone I cannot say, but you can put your light out by warming the enzyme and so reversibly denaturing it, and during that period of reversible denaturation there is a greatly enhanced tendency for it to break irreversibly.

So the irreversible structure has a known activation energy, a very large one, and the overall process is something in addition to the reversible unfolding, because the enzyme will unfold and go back again very nicely if it is not kept too long in the reversibly inactivated state.

Dr. Crick: Let me just put you on the spot. Are you trying to say that you can break the primary structure of a protein by unfolding it?

Dr. Eyring: No, indeed.

Dr. Crick: Then I don't see what you are saying.

Dr. Eyring: I am saying that, preliminary to what may be primary breaking of protein structure but in any case is irreversible inactivation, there is an unfolding which makes it vulnerable to this subsequent reaction which destroys the enzyme.

Dr. Crick: But is there evidence that you easily break the primary structure—of course, if you digest with trypsin, it's well known.

Dr. Eyring: There is very good evidence that you can reversibly put the enzyme out of business by something that can only be interpreted, at least in my opinion, as a preliminary unfolding of the protein. And you can bring about this state where it breaks very fast by warming it or with chemical denaturants.

Dr. Crick: What breaks?

Dr. Eyring: Some chemical bond (or bonds) break essential to the functioning of the enzyme. There is an irreversible change which follows and is contingent upon a preliminary reversible change.

Dr. Stanley: There was a hand way back there. Would you like to ask a question?

Mr. Ed. Peeples, The University of Texas: There seems to be a little bit of a question about the physical relationships here, and relationships, say to the ribosomes as they come off the messenger RNA, their movement, if any movement, and also the relationship of the soluble RNA. In other words, these things have to get to the proper place at the proper time.

You mentioned grouping them together in tens and so forth, but what is the mechanism of movement here? Does there need to be postulated a movement mechanism?

Dr. Crick: The first question concerns the movement of the ribosome along the messenger, or vice versa, since there is no important distinction. We don't know how that happens. There are speculations that you need GTP (which may be split in the process) and also two special enzymes to produce the actual movement, but in fact we don't know how it happens. It is obviously a problem we have to answer.

The second question is how the soluble RNA gets in the right place, and this is a quite different problem. When I was a physicist I used to worry about how the substrate got onto the enzyme, and the answer is it just bangs around and bangs around and bangs around until it hits it in the right way. And that is what we believe to be the answer in this case. Whether the rates, as I was asked earlier, are reasonable, we cannot say for sure. We know that they are not totally unreasonable.

Dr. Stanley: Well, Dr. Crick, there is evidence, is there not, on the rate of synthesis of hemoglobin?

Dr. Crick: Yes, but you have to estimate how fast the soluble RNA will diffuse to a ribosome and so on. Even the figures that we have on the rate for hemoglobin synthesis are not agreed to within a factor of two. If everything were very favorable there is no reason why synthesis shouldn't go at the rate it does, but one can't be more precise than that.

Dr. Stanley: Are there other questions or comments?

Dr. T. R. Henderson, University of Arkansas School of Medicine: Dr. Crick, I would like to ask you if you really think that base sequence alone can account for many of the phenomena, such as operator regions in the DNA and for spaces in between the gene messages? Or do you think it would also be necessary to consider it a possibility that some of the regions in DNA may vary from a simple double helix, that is, the type of hydrogen bonding which I believe poly dG and poly dC can undergo, according to Dr. Inman? Do you think something of this type as well would have to be considered in these mechanisms?

Dr. Crick: Let me make two points. First of all, in the last analysis. I believe it's the base sequence that counts, in this sense: I believe the only way you can copy a base sequence is by specific pairing. Anything else. secondary structure or methyl groups, and so on, will not affect the specificity of that copying process.

On the other hand. we can ask: Can the expression of a gene be medi-

ated by something which is not strictly the base sequence, but is a result of it? For example, by putting methyl groups on by a certain enzyme, which puts them on only where they are certain sequences and not others, or because of the secondary structure, in the way you have mentioned. I think that is quite possible.

In particular there is a speculation that the base pairing might be different if you get runs of A's and T's, and that this might be the signal for the beginning of a gene. It is known that when you mix poly A with poly U, you make a structure having one poly A with *two* poly U's, one of which pairs by the standard base pairing and one of which pairs by a different one.

The question you have to ask is: What do you get in a structure having one poly A and *one* poly U? This is not known. It could be that in DNA if you have a stretch of A's on one chain and U's on the other that the structure clicks over into the other base pairing. This would expose the normal base pairing and would be a very good way of beginning making messenger.

Dr. Stanley: Are there other questions, or comments?

Dr. Paul D. Boyer (Discussion Leader), *University of California at Los Angeles:* In connection with the last question, I wondered if the implication was that there might be linkages other than the three-five phosphodiester that would be involved, such as the linkage in the circular phage, ϕX-174. If so, might such linkages have functional genetic significance?

Dr. Crick: One always suspects there may be occasional links of that kind which may do special things, but I think that one would expect that these would be special. The argument in theoretical terms is this: The replication process is going to be a repetitive one, since this is very economical. It means you can copy by the same mechanism very, very many different sorts of genes. If you are going to do something special you have probably got to make an enzyme to do that part of the job. Of course, there is no reason why you shouldn't, but it is only likely to be done in special cases. It's not likely to be general because it is *ad hoc.*

Dr. Boyer: I would think the implication is that the special linkage takes a special function, like that of beginning reading or ending reading.

Dr. Crick: Yes, but there are lots of other functions. One of the things you need, for example, is to keep genes quiet. The reason, I think, most things are double-stranded is that you can't easily read messenger off them.

Often I think these mechanisms are not so much to do something as to prevent things happening. This falls within what you were saying, of course, but just emphasizes a different aspect of it.

Dr. Stanley: Most of the information in our genes presumably is being suppressed most of the time, and we use very little of it at any one time. Are there any more comments or questions to be put to Dr. Crick?

Dr. Alfred Marshak, Tulane University Medical School: I wonder whether Dr. Crick would like to comment on the possible role that DNA might play in such structures as mitochondria and plastids.

Dr. Crick: Can I pass the topic on to Dr. Rich?

Dr. Rich: The question is what is the DNA doing in subcellular organelles, such as chloroplasts or mitochondria?

The facts are briefly the following: It is possible to isolate from chloroplasts a DNA species with a different base ratio than that found in the nucleus of the cell. We were able to show this in our laboratory about two or three years ago.

And more recently workers have discovered that there is also another species of DNA in mitochondria.

Now, for many years people have believed that the chloroplast is a self-replicating organelle, and that it had what has been called "extra chromosomal inheritance." It is an attractive hypothesis to believe that some of the information needed to replicate this subcellular organelle is found in the organelle itself, in a DNA which is not integrated into the great mass of DNA in the chromosome.

These are higher organisms, of course, and in a sense this might be like an episome or a fragment of DNA which is not integrated into the genome of a bacteria.

The evidence about the origin of this DNA is not very clear. I think what has been shown clearly is that these different classes of DNA exist. At a later stage, one would hope that people would take this DNA and try to show that messenger RNA made from it produces some of the structural protein of the sub-cellular organelle. At that stage one would say, that indeed, this is the function of the extra chromosomal DNA.

Dr. Stanley: Are there other questions?

Dr. Eyring: Do you picture the appropriate folding or pairing of chromosome-like material as a control which stops growth, and that damaging this folding up mechanism may lead to uncontrolled growth?

Dr. Stanley: The possibility of nucleic acids operating in the cytoplasm independently of the genome is, of course, a very interesting one, having, perhaps, implications in the cancer field. Because we now have, particularly in work with the Rous sarcoma virus as well as with the polyoma virus, good evidence that the viral nucleic acid replicates endlessly in cells from which you can get no active viral nucleic acid.

The big problem today is, what is the status of the nucleic acid in these so-called non-producer cells? Is it integrated with the nucleic acid of the host, or is it not? And is it there in only one copy, or is it there as several copies? Some of the things you heard this morning, the techniques that Dr. Ochoa has devised, I think will be useful in elucidating this particular problem, which I think will be of extreme importance in connection with the cancer viruses I mentioned, and quite possibly with human cancer.

If there are no further questions, then, I should like to close this session by thanking our excellent speaker this morning, Professor Ochoa, and our very stimulating speaker this afternoon, Dr. Crick, for their wonderful, really brilliant presentations. I don't know when I have heard two such good talks as I have heard today. I should like to thank the discussion leaders for entering in, in such a very useful manner, and providing leadership in the discussion.

Again thank you, the audience, for being with us during the day. We hope to see you all tomorrow morning at ten o'clock.