

The Biochemistry of Genetics

FRANCIS H. C. CRICK

THERE IS A SENSE in which genetics is the most important part of biology. The most characteristic feature of biological systems is their organized complexity. Such a degree of complexity, we believe, can only arise by the action of Natural Selection. It is therefore important to understand the biochemical mechanisms which allow Natural Selection to take place, and this necessarily involves us in the biochemistry of genetics.

Classical Genetics started as a "black box" subject. It was found that the pattern of inheritance of certain characteristics showed a striking and unexpected regularity. This led to the concept of Mendelian factors, or "genes." It was proved at an early stage that such genes were located on the visible chromosomes. By genetic methods it could be shown that certain genes were "linked" and that each linkage group was located on a particular chromosome. By making linkage maps the genes were shown to occur in a linear order along the chromosome and were inferred to have a material structure. Classical cytogenetics was limited by the resolving power of the light microscope. It tended to be mainly concerned with such questions as the structure of chromosomes, the mechanisms of mitosis and meiosis and in particular of genetic recombination. Most geneticists believed that the gene was made of protein.

It is worth noting that these questions of the large-scale organization of the genetic material are among those that we cannot yet answer. What we *can* answer is a rather different set of questions, such as: What do we mean by a gene? What is it made of chemically? How is it copied?

FRANCIS H. C. CRICK, Medical Research Council Laboratory of Molecular Biology, Cambridge, England.

How does it act? These questions were of course asked at an early stage, but without much hope that they would be answered. The answers have come about mainly due to the rise of two quite different disciplines; on the one hand, biochemistry and on the other, the genetics of microorganisms.

Biochemistry has let us approach the problem at the atomic level, and since it has turned out that the general plan of the macromolecules involved is rather simple, it has been remarkably successful. On the genetical side a very high genetic "resolving power" is often needed. To obtain this, very large populations must be scored, and this is most easily and most rapidly done with microorganisms, especially bacteria and viruses. Their nutritional requirements are simple and for many genes it has proved quite easy to devise convenient selective techniques. In addition, the genetic material of these organisms is itself simpler than that of higher organisms in that there is much less of it per cell, and it appears less highly organized. Finally, most of what we have found out about genes in microorganisms is probably true for all genes in all organisms.

The genes of higher organisms are, we suspect, basically the same as those of microorganisms, but with additional complications. For example, all organisms have nucleic acid, but only the higher ones have histones as well.

In this review, therefore, I shall concentrate almost entirely on microorganisms, with only an occasional reference to organisms with proper chromosomes and the classical mitotic apparatus. On the other hand, I shall also leave on one side many of the rather strange methods of genetic recombination found in microorganisms (transduction, for example) except in so far as they

provide useful information on other aspects of the subject. I shall also omit two important topics, mainly because our knowledge of them is at the moment confused. These are the mechanism of genetic recombination and the mechanism of the control of gene action. Of the former it may be said that the process of "breakage and rejoin" now seems to occur in most cases, but the details of it are obscure. About control mechanisms we know that small molecules can turn genes on and off (probably indirectly) and that genes can be regulated together in groups. We also know that some aspects of control affect genes on one chromosome, but not those alleles on the sister chromosome in the same cell. That is, they act *cis* rather than *trans*. Otherwise our knowledge is, on the genetic side, rather confused, and on the biochemical side practically non-existent. Finally, for reasons of space, I shall not consider the phenomenon known as inter-allelic complementation.

THE QUESTIONS

I shall thus address myself to answer three broad questions.

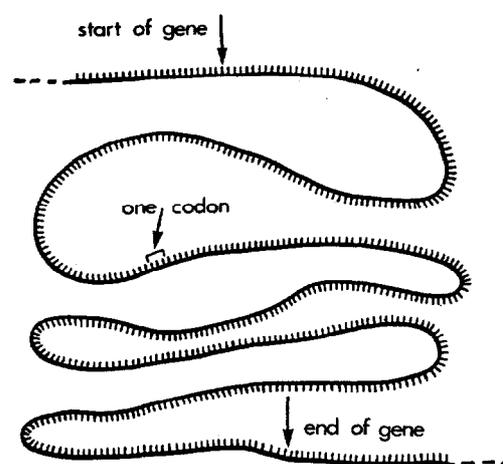
1. What is the genetical material?
2. How is it replicated?
3. How does it function?

I shall assume that the reader is familiar with the elementary chemistry and physical chemistry of protein and nucleic acid.

THE ANSWERS IN OUTLINE

There is now much experimental evidence that the main genetic material is nucleic acid. In most organisms it is double-helical DNA. This is duplicated in a relatively simple way. Each chain of the double helix directs the synthesis of a new, specific, companion chain. The correct sequence of bases on the newly formed chain is obtained by selecting the correct base (at each step) for the new chain, using the standard base-pairing.

The basic genetic material is thus a very long piece of nucleic acid. A gene, defined rather



The size of a small gene

FIG. 1. The figure represents part of a piece of (single-stranded) nucleic acid. Each short line represents one base. Notice that there is nothing special in the structure (except the base sequence) to mark the beginning or end of the gene.

loosely as the "unit of function," is then a particular stretch of this nucleic acid, having a characteristic base sequence, usually about a thousand or so bases long (see Fig. 1). There is no break in the structure where one gene ends and the next begins. This is presumably signalled by a special base-sequence at that point. Genetic recombination can take place on a very small scale, probably between adjacent bases, so that the gene can be subdivided genetically into many sites arranged in a linear order.

A stable mutation is then any change in the base sequence which can be copied at the next replication. It may be that one or more bases are changed, or that a base or sequence of bases is added or deleted.

The most important *function* of genes is to direct the production of protein molecules. This is done by producing a copy of the gene in the form of "messenger RNA" (m-RNA). This m-RNA is then "read" by the protein-making machinery of the cell in such a way that a polypeptide chain is synthesized having a defined

amino acid sequence. It is believed that this polypeptide chain then spontaneously folds on itself to produce the final conformation of the protein.

Note that the duplication of DNA and the synthesis of RNA—a process known as *transcription*—are comparatively simple processes, needing only a supply of activated monomers and a specific enzyme in order to occur. On the other hand the processes of protein synthesis, in which the base-sequence of m-RNA directs the formation of the amino acid sequence of the polypeptide chain—a process known as *translation*—is a complicated process, involving bulky biochemical machinery (the ribosomes) and many different associated components (transfer RNA, activating enzymes, etc.).

GENES ARE MADE OF NUCLEIC ACID

The first evidence that this might be true was provided in 1944, in the epoch-making paper of Avery, MacLeod and McCarty (1), which showed that DNA was the essential component of the Transforming Factor in *Pneumococcus*. In transformation one gives an organism a new version of a gene by adding to it DNA taken from another strain. This has now been done in several species of bacteria, and for numerous different genes, many of which are known to control the production of particular proteins. There is now no longer any serious doubt that the active material is DNA itself.

Most viruses consist mainly (if not entirely) of a length of nucleic acid and a protein coat. In several cases it has been demonstrated that the nucleic acid alone is infective and enables the infected cell to reproduce complete virus containing both nucleic acid and the specific protein of the coat.

This work has shown that while nucleic acid is the essential genetic component, it does not have to be double-helical DNA. It can be DNA or RNA, double-stranded or single-stranded, depending upon which virus we are considering (see (2) for some recent references). The only generalization appears to be that, so far, no large viruses are known which are single-stranded.

One would surmise that a single-stranded nucleic acid would be more delicate than a double-stranded one, and that DNA would be degraded less easily than RNA. Thus the use of double-stranded DNA for the main genetic material of larger organisms is only to be expected. Notice that there is certainly no need for both strands from the point of view of information. Since the sequence of bases on one strand implies that of its companion (because of the specific base-pairing), the second strand adds nothing, except an insurance against damage.

In no case has it been possible to show for a virus that its protein component alone is infective. Such a component may kill its host, but there is no evidence that it can reproduce. It is more difficult to be quite certain that a *small* amount of protein is not essential in order for the nucleic acid to be infective. Indeed, on the basis of our present ideas this may well occur in some cases. However, one would be surprised if it were essential for all viruses. Moreover, one would not expect the detailed genetic information needed to make such a protein to be carried by the protein itself, but rather by a nucleic acid gene of the virus (or, less likely, of the host). In short, we are reasonably certain that the most important genetic material is nucleic acid. Other minor components are difficult to exclude completely.

The above remarks should strictly be taken to apply to the genes proper. Two reservations must be made. There may be other systems in a cell, especially in higher organisms, which are extrachromosomal. It is thought that such systems occur, for example, in mitochondria and in chloroplasts. These organelles probably contain DNA of their own. It remains to be shown whether such DNA, inside the appropriate cell, determines them completely. On the other hand it is conceivable that some structural elements of a cell can only form ("crystallize") if such a structure (or perhaps the nucleus of such a structure) already exists. This possibility has been discussed recently by Sonneborn (3).

Nucleic acid, then, is the main genetic material. However, it does not follow that *all* nucleic

acid has a highly specific genetic function. We do not even know whether this is true for all the DNA on the chromosomes, and indeed there is a strong suspicion that some of it may be rather unspecific. Certain crabs, for example, have an appreciable fraction of their DNA with an unusual composition, consisting mainly of adenine and thymine in a predominantly alternating sequence (see (4) for references). The function of this special DNA is not known.

Then again there is probably nucleic acid which will never be copied. Egg cells frequently have large quantities of DNA in their cytoplasm. It is surmised that this is a storage material, which is broken down to mononucleotides, which are then used for DNA replication. Moreover, the balance of evidence suggests that in most normal cells RNA molecules are not copied at all. For bacterial cells there is good evidence that all types of RNA (messenger RNA, ribosomal RNA and soluble RNA) are copies of the DNA of the chromosome. It is plausible that all this RNA is a direct copy, and not a copy of a previous copy. However, one must be cautious at this point. It is certain that in a cell infected with an RNA virus there is an RNA-RNA copying mechanism. There is nothing in principle against such a mechanism in normal cells, and whereas it may perhaps be absent in most organisms it could well be present in some.

THE SIZE AND SHAPE OF DNA

It used to be thought that DNA "molecules" were poly-disperse with an average molecular weight of about, say, 10 million. We now know that this is an artefact. Long polymer molecules are easily broken by hydrodynamic shear. Their fragility increases as the cube of their length, and they tend to break near the middle. The DNA of the T even phages, for example, is all in one long piece (molecular weight $\approx 130 \times 10^6$), having a total length of about 70μ . Even relatively gentle handling tends to break this in half, and normal handling will break it even further (5-7). It is now believed that the DNA of bacterial cells may also be one long piece (8, 9). For *E.*

coli the estimated length is 1 mm, which is about a thousand times as long as the size of the cell.

An unexpected feature is that the DNA of several viruses and bacteria has been shown to be circular (for references, see (10)). This is true for the polyoma virus and rabbit papilloma virus (though filaments also occur) and for the single-stranded DNA of the virus Φ X-174, as well as its double-stranded replicative form found in the infected cell. The lysogenic virus λ appears to have "sticky ends," and can form a circle, or several can join together end to end into a large circle or a longer straight line. By autoradiography, the long DNA chromosome of *E. coli* has been shown to be circular (8).

So far no RNA virus has been shown to have circular RNA. It is unlikely that the RNA of rod-shaped viruses like tobacco mosaic virus is circular, but there is no reason why some spherical viruses should not have circular RNA.

In some cases it is known that the genetic map of an organism is circular, but the genetic map can be circular even in cases in which the DNA of the virus is a straight line, as in the T even phages (11). The subject is of great interest but it is too complex to discuss further here.

THE REPLICATION OF NUCLEIC ACID

By replication we mean a highly specific biochemical process by which the base sequence of the nucleic acid used as the template dictates the base sequence of the product. We know of three such processes which occur in Nature: 1) the production of DNA, using a DNA template; 2) the production of RNA, using an RNA template; 3) the production of RNA, using a DNA template. The first two occur in the replication of genetic material. The third is the main mechanism of gene function. A fourth possible process, the replication of DNA on an RNA primer, has not yet been discovered *in vivo*, although a special case has been demonstrated *in vitro* (12).

THE REPLICATION OF DNA

It is easiest to consider first the replication of double-helical DNA. As is well known, this struc-

ture consists of two polynucleotide chains wound round one another, as symbolized by the blue and red ribbons on the badge of this Congress. The two backbones are anti-parallel. Although much of the energy to stabilize the structure comes from the stacking of the bases, the *specific* forces holding the structure together are the hydrogen bonds between the specific pairs of bases: adenine with thymine, guanine with cytosine, as shown in Fig. 2.

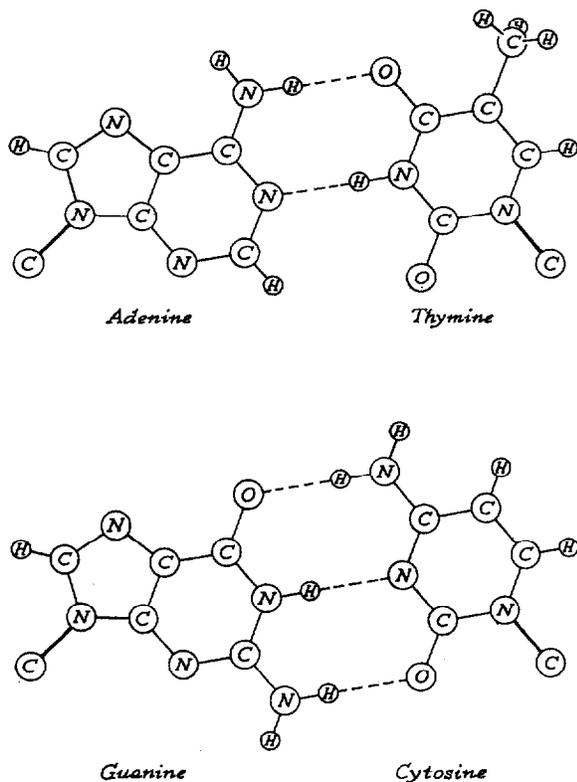


FIG. 2. The two standard base-pairs believed to occur in DNA. From *Molecular biology of bacterial viruses*, Gunther Stent, Freeman, 1964, Figs. 3-9.

The pairing of the bases is the key to the proposed replication mechanism, since it allows a single chain to act as a template and direct the selection of the bases for a new companion chain. The mechanism has two important features: 1) The two original chains come apart, so that the two daughter double-helices are "hybrid,"

each consisting of one old and one new chain. Such a mechanism is described as "semi-conservative" (see Fig. 3). 2) The base-sequence of each new chain is "complementary" to that of one of the old ones; this is achieved during replication by the specific base-pairing mechanism. In my view the second of these is the really essential feature.

There is now good experimental evidence that both these features are correct. That the DNA of *E. coli* behaves as a duplex during replication was shown by the now classic experiment of Meselson and Stahl (13). There has been some controversy as to whether the two parts of the duplex are indeed the single chains of the double helix. A contrary view has been put forward, with experimental evidence, by Cavalieri (14), but it has not found much support. The balance of evidence suggests that in fact the two chains separate and that replication is semi-conserv-

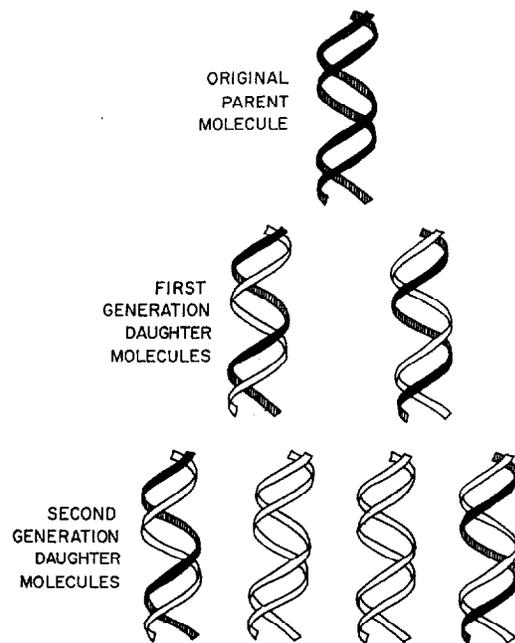


FIG. 3. Illustrating the semi-conservative replication of DNA. In conservative replication the two original chains never come apart. In semi-conservative replication they separate and a hybrid is formed consisting of one old and one new chain (cf. Meselson and Stahl (13), Fig. 6).

ative. There is also evidence that the DNA of the chromosomes of plants such as *Vicia* replicates as a duplex (see the review by Taylor (15)) but this is outside my present subject.

as the DNA template used, and the same nearest-neighbour base frequencies. Moreover, the results obtained with purine and pyrimidine analogues are compatible with the base-pairing mechanism

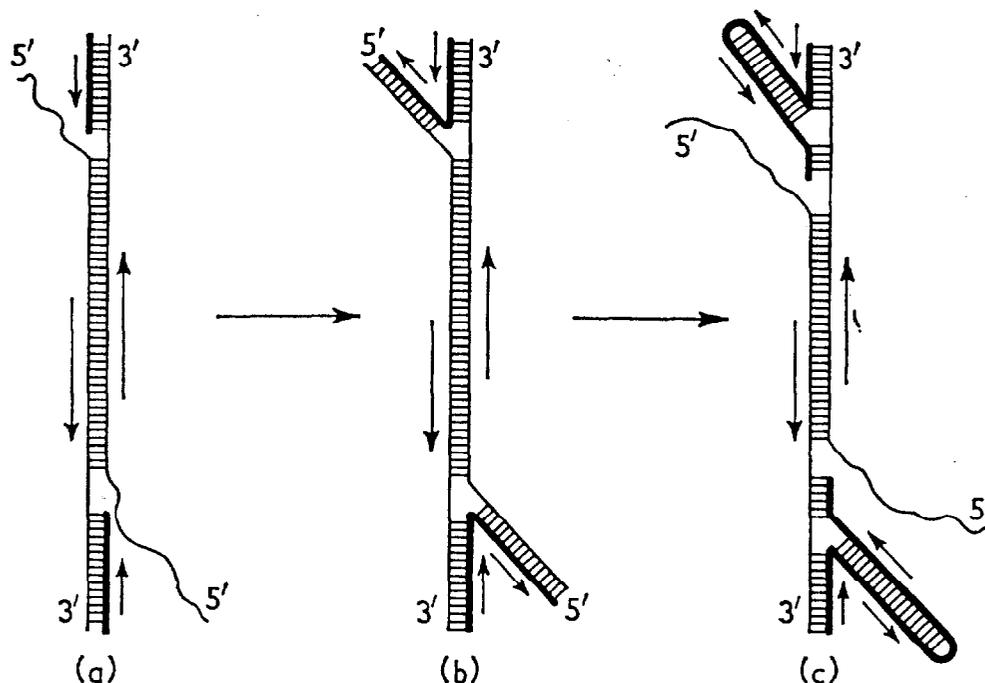


FIG. 4. A hypothetical model to account for the non-denaturability of synthesized DNA. The primer is indicated by thin lines, the produced DNA by heavy lines. Hydrogen bonds between strands are shown as horizontal lines. According to this scheme, replication proceeds by pairing with the base at the 3'-hydroxyl end of the template strand; the direction of the arrows shows the direction of synthesis and polarity of the strand (cf. Schildkraut *et al.* (17), Fig. 9).

The replication of DNA can be carried out in the test-tube, as first achieved by Kornberg and his colleagues (see references in (16)). The experimental results are in fairly good agreement with the mechanism proposed by Watson and myself. A special enzyme, DNA polymerase, is needed which requires as substrates all four types of deoxynucleotide triphosphate. During the synthesis pyrophosphate is split off. The enzyme from *E. coli* will copy any DNA added to it. The product is double-stranded DNA of high molecular weight. It has the same base composition

we suggested. The nearest neighbour frequencies show that two strands of the double helix are anti-parallel rather than parallel.

In general the physical state required for the template DNA varies somewhat with the species from which the enzyme is prepared. In all cases single-stranded DNA will act as a template, but in some cases double-stranded DNA will not act.

Extensive synthesis can be obtained with the polymerase from *E. coli*, but the product is unlike native DNA in at least two respects. It appears branched in the electron microscope, and

it renatures fully after heating and quick cooling. The exact reasons for this behaviour are not yet known, but a plausible model has recently been presented by Schildkraut, Richardson, and Kornberg (17), and is described in Fig. 4. They call this the multiple hair-pin structure. These eccentricities may explain the failure to replicate biologically active DNA in a convincing manner.

The enzyme shows two unexpected additional reactions. It will, in certain cases, synthesize a special DNA without any template being present, but only after a long lag. Thus if given only the adenine and thymine precursors, it will produce a polymer of high molecular weight having a strictly alternating base-sequence and a base-paired double-helical structure. Given only the guanine and cytosine precursors, it will synthesize chains of two types: either all guanine or all cytosine.

The other peculiarity is that if the enzyme is provided with Mn^{++} instead of Mg^{++} it will accept ribonucleoside triphosphates as well as the corresponding deoxyribonucleoside compounds (17). It is thus possible to synthesize a polymer having a mixture of ribose and deoxyribose in its backbone. So far no polymer of this type has been found in Nature.

The enzyme for *E. coli* has, or has closely associated with it, several nucleases (see Richardson *et al.* (19) for a summary and references; also the review by Lehman (20)). It is an interesting speculation that these may really exist to increase either the accuracy or the efficiency of the synthesis, but this is too complicated a subject to discuss here. DNA polymerase from *Bacillus subtilis* appears to lack one of these activities (21).

It should be noticed that single-stranded DNA, such as that of the virus $\Phi X-174$, can be replicated both *in vivo* and *in vitro*. This underlines a feature which was not so clear to us at first, namely that the polymerase enzyme must play a very important part in the specificity of the replication. On our scheme it must hold the backbones both of the growing new chain and of the old chain near the growing point in such a position that only the standard base-pairs can

form. The geometrical rigidity of the double helix is not enough by itself to do this with high efficiency.

If the two chains of the double helix are separated during replication, they must be unwound, since the X-ray evidence shows beyond any reasonable doubt that the two backbones are twisted together round a common axis. Exactly how this happens is still not known. There is no difficulty about the two chains coming apart when the structure is "melted" by raising the temperature. The important "annealing" technique whereby separated chains can be made to form a double helix again with their complementary chains by slowly cooling them together, shows that, given the chance, two chains can wind up as well as unwind. What is puzzling is how this is achieved for molecules of DNA of very great length; how the energy is supplied and how it is done in a neat manner.

To approach this problem we need to know how the two chains behave during replication. An early speculation was that the chains did not come fully apart before they were copied, but that the synthesis and the unwinding went together, so that only a small part of the chains near the growing point was ever single-stranded at any one time. This is the so-called Y mechanism, illustrated diagrammatically in Fig. 5. Because the two chains of a double helix run in opposite directions this implies that the two new chains are being synthesized in chemically opposite directions. Presumably this would require that one of the two growing chains would have a triphosphate on the 5'-position of its terminal deoxyribose.

Until recently there was no evidence to support such an idea, and direct chemical evidence is still lacking. However, by autoradiography Cairns has produced pictures (8) of the replicating circular chromosome of *E. coli* which certainly suggest that the Y mechanism, or something rather like it, is operating, and implies that the chromosome is synthesized by starting at a fixed point and going systematically along the entire chromosome. There is other evidence which indicates that this happens, and indeed it

markable that not only can a circle of DNA of 1 mm length be replicated but that the two daughter molecules can separate *into separate regions* without getting hopelessly tangled together. Some insight into how this is done in bacterial cells might ultimately help us to understand the structure of the much more complicated chromosomes of higher organisms.

Finally it is worth remarking that, whereas in the cell DNA synthesis proceeds very rapidly, at the rate of a couple of thousand base-pairs per second, *in vitro* the DNA polymerase works much more slowly. This may well be because it cannot carry out the proper Y mechanism (Fig. 5) for some reason. For example, the enzyme may be a dimer in the cell, but has become a monomer under the conditions used in the test-tube. Alternatively, there may be some difficulty in starting or maintaining a chain with a triphosphate on its growing end. Dark doubts have been expressed that the DNA polymerase we have is "only a repair enzyme." If this were so, a mutant lacking it could perhaps survive in a fashion. It would certainly be more comforting if the *in vitro* synthesis could be achieved more rapidly and also without the hairpin artefacts shown in Fig. 4; but this is easier said than done. We verge here on the fascinating but difficult topic of the *control* of DNA synthesis in the cell, but this is outside the scope of this lecture.

THE REPLICATION OF SINGLE-STRANDED VIRAL RNA

In no case has it yet been shown that the replication of an RNA virus goes through a DNA intermediate, and in several cases the evidence suggests that it does not.

In two cases it has been demonstrated that there is a special form of RNA in the virus-infected cell which has the properties expected of double-stranded RNA, and which is probably the replicative form of the virus. It has also been shown in several cases that upon virus infection at least one new enzyme appears in the infected cell which can synthesize RNA using RNA as a primer. This activity has been found in HeLa

cells infected with polio virus and in L-cells infected by Mengovirus (see the review by Baltimore and Franklin (22)). It has been extensively studied in *E. coli* infected with the phage MS 2 (see Weissman *et al.* (23) and Ochoa *et al.* (24) for references). In this case the new enzyme is found associated with the double-stranded replicative form. This appears to be good double-helical RNA, and it is inferred that the base sequence of the new strand is complementary to the parental one in the usual way.

The action of the enzyme attached to the double-stranded RNA has been studied *in vitro*. It requires the four ribonucleoside triphosphates as substrates, and splits off pyrophosphate. It appears to synthesize mainly the parental type strand, some of the product staying in the double-stranded form, and some appearing as single-stranded RNA.

If the double-stranded RNA is indeed an obligatory replicative form, as the evidence suggests, there must be a maturation process by which parental-type single strands are produced and then incorporated in the mature phage particle. Exactly how this is done is not at the moment clear, but it is interesting that the *in vitro* system appears to synthesize only chains of the parental type and not the complementary chain.

It will be interesting to see if this general scheme for the replication of single-stranded RNA viruses is generally applicable. Two reports at the Congress claimed that a double-stranded replicative form of tobacco mosaic virus has been discovered recently.

THE SYNTHESIS OF MESSENGER RNA

It has been shown for many types of cell that there is a species of single-stranded RNA, heterogeneous in size, which is rapidly labelled and has a base-composition complementary to that of the DNA of the cell. In some cases this RNA has been hybridized with single strands of the cell's DNA, thus showing that it is a faithful copy of the base-sequence of at least part of the DNA. Such an RNA is also found attached to ribosomes, stringing them together into poly-

somes. In bacterial cells this RNA turns over rapidly. In mammalian cells it turns over much less rapidly and in reticulocytes probably not at all. This RNA is surmised to be the messenger which carries the genetic information from the DNA of the gene to the site of protein synthesis, the ribosomes. It has been called messenger RNA, or m-RNA for short (for reviews see Jacob and Monod (25) and Lipmann (26)).

The concept of messenger RNA implies that the RNA of the ribosomes itself is *not*, as previously believed, the genetic message, and that the ribosome is a non-specific part of the mechanism of protein synthesis. It is not too much to say that this idea is the most important addition to our thinking in the last ten years. It has clarified a host of puzzling observations and stimulated experimental work on protein synthesis, the genetic code, the biochemistry of embryology and many other subjects.

The enzyme RNA polymerase which we believe synthesizes m-RNA was isolated by several groups of workers. It requires nucleic acid as a template and (with reservations) the presence of all four ribonucleoside triphosphates as precursors. The product is RNA of a reasonably high molecular weight, having a base composition and nearest neighbour analysis complementary to the nucleic acid used as template, and is able to hybridize with the latter (see the review by Hurwitz and August (27) and recent papers (28-31)). During the reaction pyrophosphate is split off. The yield of RNA can be many times greater than the amount of DNA template used. The results with base analogues are crudely what one would expect, but the detailed pattern of incorporation is complicated and suggests that other effects are operating in addition to the pairing of bases.

The requirements for the template are rather complicated. Both single-stranded and double-stranded DNA can act, as can small oligodeoxynucleotides. In the latter case the RNA product can be considerably longer than the template. When double-stranded DNA is used as a template, the two strands are not separated at the end of the reaction, even after extensive RNA

synthesis. This suggests that if the two strands separate at all during the reaction they do so only transiently. In other words the synthesis is "conservative." In addition, the evidence shows that *in vitro* both strands of the DNA are copied, since the RNA produced can be annealed to form a double helix with itself. However, it has been established in several cases that *in vivo* only one strand of the DNA is copied, as would be expected on theoretical grounds. A very neat experiment has been reported recently by Hayashi, Hayashi and Spiegelman (32). They show that if intact, circular, double-stranded DNA, of the replicative form of Φ X-174, is used as a template for the RNA polymerase, only one strand is copied, whereas if the circle is disrupted, RNA copies of both strands are made. The RNA produced by the intact circular form is the complement of the complement of the parental single-stranded DNA, as would be expected from earlier studies of the RNA produced inside the virus-infected cell.

When the single-stranded DNA of the virus Φ X-174 is used as a template, the first product of the reaction is a hybrid DNA-RNA complex; free RNA is produced later. It is suspected that the RNA chains produced may be rather shorter than the length of the template DNA. Similar results are obtained with thermally denatured DNA from λ dg and from calf thymus. If the DNA-RNA complex is itself used as a template the reaction is at least partly semi-conservative. That is, some of the RNA of the hybrid is displaced by the new RNA being synthesized (29). The precise reasons for this behaviour are not yet clear.

The system shows two further types of behaviour which are unexpected. Given a double-stranded DNA template and a single nucleoside triphosphate the enzyme can produce a homopolymer. The optimum conditions for this reaction differ somewhat from those for the usual synthesis. Chamberlain and Berg (33) suggested that the formation of homopolymers is due to the template action of short clusters of identical complementary bases on the template nucleic acid used. Thus the formation of polyadenylic acid

would be directed by short stretches of thymidylate. The available evidence, such as it is, is compatible with this idea.

The other surprising feature is that RNA itself can act as a template for the enzyme. The reaction is similar to the DNA-primed reaction, though there are minor differences. All four ribonucleoside triphosphates are required, pyrophosphate is released and the product is complementary to the RNA template used. RNA homopolymers, complementary RNAs and natural RNA will all act as templates, but in decreasing order of efficiency. The natural double-helical RNA of reovirus also acts as a very effective template (34). However, there is as yet no evidence that any of these reactions take place *in vivo*.

The formation of homopolymers also occurs with single-stranded or double-helical RNA when the RNA polymerase from *E. coli* is used.

THE SYNTHESIS OF RIBOSOMAL RNA AND s-RNA

It has been shown by specially sensitive hybridization techniques that in bacterial cells at any rate, these two classes of RNA are copied from parts of the DNA, presumably by the RNA polymerase, in the same sort of way that m-RNA is synthesized (see Spiegelman and Hayashi (35) for a review and references). Rough estimates have been made of the amount of DNA involved in the case of *E. coli*—about 0.1 to 0.2% of the genome for ribosomal RNAs and about 0.023% for the s-RNA molecules. There appear to be different genes for the 16S and 23S ribosomal RNA.

Although both of these classes of RNA have a similar base composition throughout Nature, despite large variations in the over-all base ratios of DNA from one species to another, their detailed base sequence is peculiar to each species, since RNA from one source will not readily hybridize with the DNA from a quite unrelated organism.

It has not yet been established for either type of RNA that the DNA does not first make some

master copies of complementary RNA, which are then themselves copied by a further synthesis to give large amounts of the complement of the complement, but so far there is no evidence to suggest this.

PROTEIN SYNTHESIS

Protein synthesis is too large a subject to cover properly in this review. I shall give a brief description of what we now know about it, and then discuss certain aspects of it more fully. For references see the recent review by Watson (36) from which the schematic diagram of Fig. 6 is taken.

Proteins are synthesized on large structures, of the order of 200A diameter, known as ribosomes. Ribosomes are found in all cells, in all organisms, in which protein synthesis is occurring. All ribosomes are in two parts, a larger one and a smaller one. Each ribosome is made of about equal parts of protein and RNA (the exact proportion varies from organism to organism). The smaller particle contains one piece of RNA, sedimentation 16S; the larger particle has one piece of about 23S. The function of the RNA of the ribosome is not understood. The exact nature of the proteins of the ribosome is at the moment a matter of controversy. It is agreed that they are fairly basic, rather small, and that there is probably more than one kind in any cell. I shall not discuss them further here.

In order to be incorporated into a protein, an amino acid must be activated by a special activating enzyme, using ATP. In bacterial cells there appears to be one and only one activating enzyme for each amino acid. This enzyme then transfers the amino acid to the terminal ribose of a special RNA, known as soluble or transfer RNA. In general there are several different kinds of soluble RNA for each amino acid. The soluble RNA then carries the amino acid to the site of protein synthesis.

The amino acids are joined up in the correct order on the ribosome, under the direction of a piece of messenger RNA which combines with the smaller part of the ribosome. There appear to be two special sites on the larger part of the

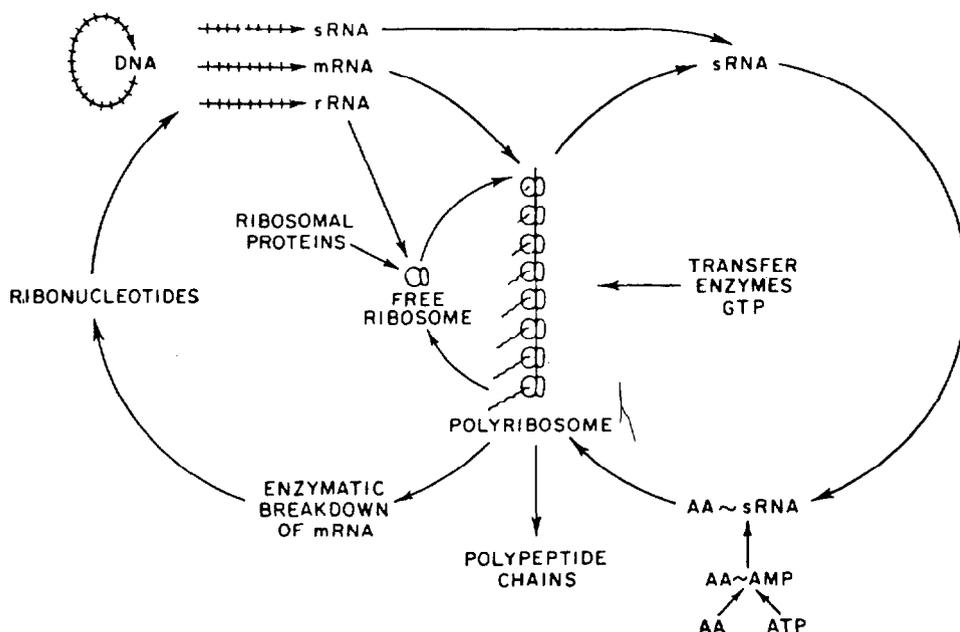


FIG. 6. A diagrammatic outline of protein synthesis (cf. Watson (36), Fig. 7).

ribosome. In one site a molecule of s-RNA sits, to which is joined the part of the polypeptide chain already made. The carboxyl end of the polypeptide chain is held by an ester link to the terminal ribose of the s-RNA. The other site holds the s-RNA which is carrying the amino acid to be added next to the chain. The correct s-RNA is presumed to be selected by its combining, probably by base-pairing, with the next triplet of bases on the m-RNA. It is then imagined that the growing chain is removed from the s-RNA molecule in the first site and joined on to that in the second site. The polypeptide chain is thus lengthened by one amino acid. The apparatus then resets in some way and the process is repeated. In addition to the components mentioned, two soluble proteins and GTP are somehow involved in this whole process.

The growth of the polypeptide chain thus starts from the amino end, and proceeds by successive terminal addition to the carboxyl end of the growing chain. This has been established in several cases by direct experiment.

As the synthesis of the polypeptide chain pro-

ceeds the ribosome moves along the m-RNA. After it has cleared the beginning of the m-RNA another ribosome can go on there and follow the second one, "reading" the m-RNA and making a polypeptide chain. Eventually a whole series of ribosomes will be travelling along the m-RNA, one behind the other, each making one protein molecule. Such a structure is called a polyribosome. The polyribosomes involved in the synthesis of haemoglobin typically contain about five or six ribosomes, but much larger polyribosomes are known to exist in other cells.

It is fairly probable that, to act as a messenger, the nucleic acid must be single-stranded RNA. If it has too much secondary structure it will not function, nor is there any evidence that single-stranded DNA is acceptable. Single-stranded viral RNA may act as a messenger, forming enormously large polysomes.

THE GENETIC CODE

The sequence of bases on the m-RNA is thus in some sense a code for the amino acid sequence

of the protein whose synthesis it directs. It is assumed that this code is likely to be a relatively simple one. Although the code is not yet known in detail it seems likely that it is a *non-overlapping, non-interlacing, degenerate, triplet code* (for references see (37)). This means that the codon—the group of nucleotides which codes one amino acid—is a group of three adjacent nucleotides and that the next three nucleotides on the m-RNA code for the next amino acid. “Degenerate” implies that usually more than one triplet codes for each amino acid, there being 64 possible triplets and only 20 different amino acids found in proteins.

By using artificial m-RNA, made enzymatically with a known composition but a random sequence, to direct polypeptide synthesis in a cell-free system, it has been possible to arrive at the probable composition of most of the codons (38-39), though it is not yet known how trustworthy these assignments are.

Until very recently the order of the bases within a codon was not known. However, at this Congress Dr. Marshall Nirenberg reported that he and Dr. Leder have devised a method whereby trinucleotides could be used to decipher the code.

They were able to show, for example, that the trinucleotide GUU promoted the binding of the s-RNA for valine to ribosomes, whereas UGU and UUG did not. Thus GUU is presumably a codon for valine (40). This development, together with the work of Khorana (41), who is synthesizing polymers with known repeating sequences, and of Doty and his colleagues (42), who are testing short polynucleotides with a known sequence, makes it likely that most of the genetic code will soon be established.

It is reasonably certain (43) that the code is very similar in all organisms tested, not only because no significant difference has been found between cell-free systems from different organisms in their response to artificial m-RNA, but also because of the evidence from mixed systems. However, it is still possible that the code differs in some respects from species to species.

THE GENE AND ITS PROTEIN ARE COLINEAR

It has been surmised for many years that the gene and its protein were colinear, although other more complicated schemes were not entirely ruled out. However, it has taken far longer than expected to prove it. This has now been done by two groups of workers.

The ideal experiment is still not technically possible. In this one would determine the actual sequence of bases along the nucleic acid of the gene, and compare it with the sequence of amino acids in the polypeptide chain for which it codes. Instead, sites within the gene, recognizable because each causes the alteration of a particular amino acid, are ordered by genetic methods and the order so found is compared with the order found in the corresponding alterations to the polypeptide chain. The resolution needed for the genetic work is so high that at the moment it is

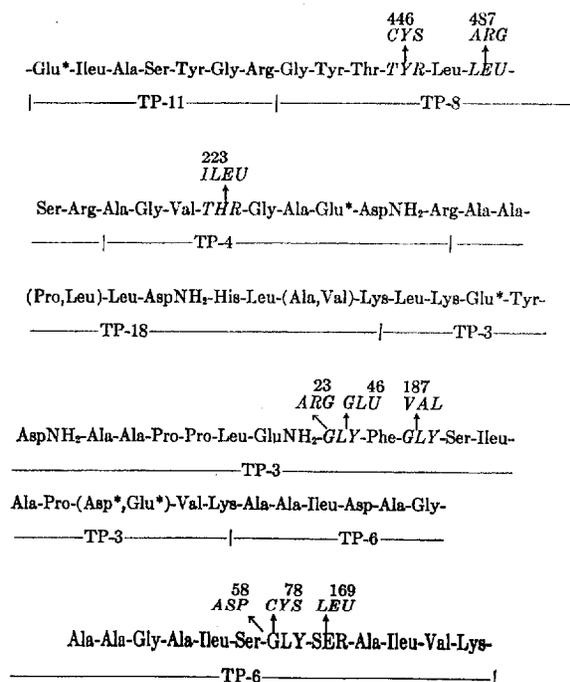


FIG. 7. Part of the amino acid sequence of the A protein of tryptophan synthetase. Those amino acids in capital letters show the different mutations which have been found and the points at which they occurred (cf. Yanofsky *et al.* (44), Fig. 3).

only technically possible to do this using micro-organisms.

Essentially this method has been employed by Dr. Yanofsky and his co-workers (44), using the A protein of tryptophan synthetase from *Escherichia coli*. They have studied 9 distinct sites, affecting 7 different amino acids, and covering a stretch of about 60 amino acids. Their results are shown in Fig. 7. As expected the two orders are the same.

The same result has been obtained in our laboratory by Dr. Sydney Brenner and his co-workers (45). They studied the protein of the head of phage T₄. They used a special class of mutants (known as "amber" mutants) which act by terminating the polypeptide chain. This action can be suppressed by growing the phage in a special host strain, a fact in itself of considerable interest. They showed, in effect, that mutants towards one end of the gene produced only a short length of the protein, whereas those towards the other end produced longer lengths. In all they studied about 10 mutants, and were able to order eight of them. Again the two orders, the genetic order and the polypeptide order, were the same. In this case the distance covered was estimated to be several hundred amino acids, so their results were on a coarser scale than Dr. Yanofsky's.

Generalising happily from these two examples we may surmise that any gene and its protein will be colinear. All this is very satisfactory and fits in very well with the quite independent evidence that proteins are assembled sequentially from the amino end.

Although the gene and its protein are colinear on a coarse scale, it is still possible that the bases which code for one amino acid are not adjacent, but spaced out on the polynucleotide chain. For example successive amino acids might be coded by bases numbered:

(1, 3, 5,
4, 6, 8,
7, 9, 11,
etc.

Such a code is called an "interlacing" code and is illustrated in Fig. 8. This possibility cannot

be ruled out, but there is no evidence which suggests that it may be true.

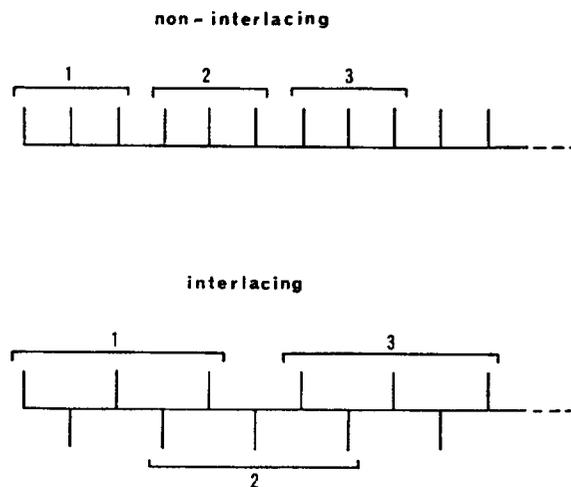


FIG. 8. To illustrate the meaning of "interlacing." Only the simplest type is shown. Each short vertical line represents a base. They have been shown alternatively up and down in the lower figure merely to make the diagram simpler. Each number shows the codon for an amino acid.

GENETIC RECOMBINATION CAN OCCUR WITHIN A CODON

This too has been demonstrated by Dr. Yanofsky, using the system just referred to. At a certain place in the polypeptide chain the wild type strain puts in glycine. In two distinct mutants one has glutamic acid and the other has arginine. If these two mutants are crossed together the rather rare recombinants have glycine at this point. Glu crossed with Arg gives Gly. Even more striking if a further mutant containing valine is crossed with the arginine mutant the recombinants are of two types; some have glycine, like the wild type, whereas some have serine at this particular place. However, if the valine mutant is crossed with the glutamic acid mutant no recombinants are formed at all, suggesting that in this case the two mutations are at the same place in the nucleotide sequence of

the gene (see Yanofsky (46) for a fuller description).

This result would have been received with astonishment a dozen years ago. Today we realize that it is exactly the sort of thing to be expected. Information of this type is invaluable for helping to decipher the genetic code. Incidentally, in the course of this work Yanofsky has picked up no less than seven different amino acids at this one position in the polypeptide chain. In all cases the adjacent amino acids are unaltered.

GENERAL OBSERVATIONS ON THE REPLICATION OF NUCLEIC ACID

The unique feature of nucleic acid replication is that one enzyme can make an exact (complementary) copy of a vast number of different molecules; that is, of nucleic acids having a great variety of base-sequences. The importance of this as a foundation for a genetic mechanism can hardly be overestimated, since it allows Natural Selection to operate very powerfully, yet in a simple manner. Not only does it permit the organism to have a selection from an almost infinite number of different kinds of genes, but for each gene a very large number of alterations to it (mutations) can be faithfully copied. We know of no other biochemical synthesis which has this character, and we strongly suspect we shall not find one which is capable of such variety.

The essential feature of this replication is the pairing of the bases. In all the cases that we know, the base sequence of the product is complementary to that of the template provided guanine pairs with cytosine and adenine with thymine (or uracil). Notice two points: 1) The pairing is complementary. So far no case of like-with-like pairing has been discovered in a replication mechanism. 2) There is no evidence of any third base-pair.

This second point needs some comment. For a third base-pair to be useful it should be a quite distinct pairing. That is, the six bases should be such that, given the usual geometrical restraints,

it should be possible for each of the six bases to pair only with its unique partner. Rich has looked for such a pair, assuming the usual hydrogen-bond type of arrangement. He has suggested (47) that isoguanine and isocytosine might be suitable, but there is some doubt about the stability of the various tautomeric forms of these bases. In any event, there is as yet no hint of their occurrence in natural nucleic acid. On the grounds that if a third base-pair were possible Nature would have certainly had learnt to use it, I propose as a hypothesis:

No new base-pair will be discovered which could not be confused with the two already known.

Other bases than the standard four do occur in natural nucleic acid. Hydroxymethyl cytosine (with some glucose attached) is found in the T even phages instead of cytosine. One phage has uracil in its DNA instead of thymine. Pseudouracil is found in s-RNA, and numerous methyl derivatives of the standard bases occur, usually in small amounts. The subject is too complex to review here. As far as I know, all the data so far are consistent with the following generalization:

No base will be found in natural nucleic acid which cannot form one of the standard base-pairs, unless it was formed by modification of the nucleic acid after replication and unless this nucleic acid is not going to be replicated further.

This does not answer the question how these modified bases are formed and whether they occur in random or in special positions in the base-sequence, but that is a topic in itself.

We have seen that strictly only one chain of the nucleic acid is needed to carry the base-sequence information, and that single-chain viruses exist, of both the RNA and the DNA type. We have also seen that in several cases there is suggestive evidence that such single-stranded viruses pass through a double-stranded replicative form.

The implication of the ideas just discussed is that *all* nucleic acid will be replicated using the

standard base-pairs. This leads us to the further plausible generalization:

All biological replication of nucleic acid takes place by forming (at least transiently) a double-helix using the standard base-pairs.

This generalization, if accepted, leads to the previous one, assuming that no nucleic acid is first modified and then de-modified.

The standard base-pairs, then, look as if they may be a uniform feature of nucleic acid replication. Of course, it is well known that other base-pairs can occur in double-helical structures. Thus polyadenylic acid (poly A) forms a double helix with itself, and polyinosinic acid (poly I) and polyguanylic acid (poly G) form multiple helices. In all these structures, however, the backbones are parallel, rather than anti-parallel as they are in DNA. The only "new" base-pair so far found in an anti-parallel double-helical structure is the pair (poly A plus poly I), and this has the two backbones further apart than they are in DNA.

Two further points should be made. First, it does not follow that other base-pairs cannot occur in important places in certain situations, for example, in transfer RNA, or during the presumed combination of transfer RNA with messenger RNA. Second, the exact nature of the base-pairs is still not completely clear. There is little doubt about the pair (G ÷ C). Three hydrogen bonds are formed between the bases, as mentioned tentatively by Watson and myself (48), but first clearly suggested by Pauling and Corey (49). This structure has now been found in two single crystals (50, 51).

The uncertainty concerns the pair (A ÷ T). The standard form suggested by Watson and myself is that shown in Fig. 2. However, the first single crystal of a base-pair to be solved, that of 9-methyl adenine and 1-methyl thymine, by Hoogsteen (52), contained the pairing illustrated in Fig. 9. This pairing, or a variant of it, has since been found in three other crystals (53, 54). Neither of these two new base-pairs can easily be fitted into a double helix which also has G-C pairs without making the backbone

irregular. However, it is by no means ruled out that stretches of A-T pairs could click over into the new pairing. It has been speculated that this might be a special signal; to start an operon, for example. It is thus of some interest to discover

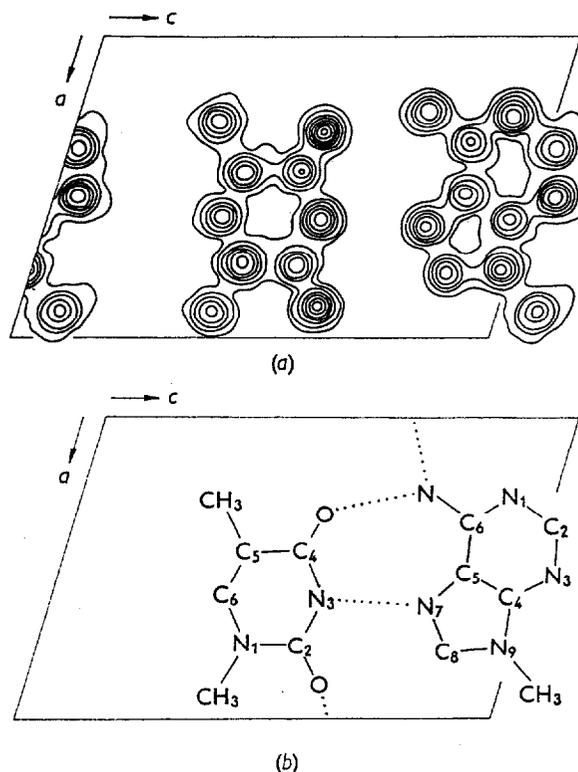


FIG. 9. The crystal structure of the base-pair 9-methyl adenine and 1-methyl thymine, from Hoogsteen (52). Compare the pairing shown in the upper half of Fig. 2 (cf. Hoogsteen (52)).

the exact structures for the double helix (poly A and poly U) and the triple helix (poly A and 2 poly U). A recent paper (55), dealing with the infrared evidence, discusses this problem more fully.

Given that DNA is replicated semi-conservatively and that the product is base-paired double-helical DNA, it follows that the new chain of each hybrid DNA molecule is complementary to the old one. As pointed out in the review by Cavalieri and Rosenberg (56), this

does not decide the *mechanism* by which this is attained. When one comes right down to it, this is as uncertain as the details of the action of almost any enzyme. However, the obvious idea is that the selection of the correct new base is, at each step, dictated by specific hydrogen-bonding between the next nucleotide to be added to the growing chain and its destined companion on the template chain. Cavalieri and Rosenberg (56) suggest that the pairing forces are mainly due to London-van der Waal's forces between bases temporarily stacked one above the other. The theoretical advantages they claim for their method of replication appear to me unconvincing, and as far as I can see, there is no direct experimental evidence which supports this idea against the pure hydrogen-bonding scheme.

It should be noticed that in all cases the precursors are the nucleoside-5'-triphosphates. Neither the 3'-compounds nor the diphosphates will act. The probable reason for the latter point was originally suggested by Lipmann, who pointed out that if the diphosphates were used, the equilibrium would not be well over in the direction of polynucleotide synthesis. By using the triphosphate esters, and then splitting the pyrophosphate released with a pyrophosphatase, the synthesis is made effectively irreversible. It is worth noting that the only enzymatic "synthesis" of nucleic acid which does use the diphosphate esters is that due to polynucleotide phosphorylase. This does not accept directions from a template and in the cell is probably a degradative enzyme. Only with the peculiar substrate concentrations provided in the test-tube will it carry out synthesis.

There is one feature of the behaviour of informational nucleic acid in the test-tube which is somewhat disturbing. This is the rather common occurrence of aberrant behaviour—that is, of behaviour which we suspect does not happen inside the intact cell. It is not merely that, given Mn^{++} , the DNA polymerase will accept ribose precursors as well as deoxyribose ones. It is the whole pattern of synthesis, of the kind illustrated in Fig. 4. The RNA polymerase, while it only copies one of the chains of DNA in the cell

usually copies both of them in the test-tube. In protein synthesis also one finds a lack of precision. The fact that under most circumstances polyuridylic acid will promote, in addition to the incorporation of phenylalanine, the incorporation of variable amounts of leucine is suspected to be another example of aberrant behaviour. Thus, one needs to show considerable caution in assuming that all the details of *in vitro* behaviour will also apply *in vivo*.

Meanwhile it is important to find out the reasons for such behaviour. In some cases the ionic environment in the cell may have been inadequately reproduced in the test-tube. In others one suspects that the trouble lies in the control mechanisms. The cell must not only carry out the reaction, but must control its rate, its timing and its point of initiation. Such mechanisms may be lost or damaged in the extraction process.

It seems very probable that one of the reasons for the double-strandedness of DNA is that most of the double-stranded DNA is relatively inert as a starting point for the RNA polymerase, except perhaps at the open ends of the double helix. Thus, in the cell the ends of DNA molecules may be kept to a minimum by having the DNA in very long pieces. Presumably some special base sequence at various points on the DNA shows the polymerase when the copying should start, and at the same time decides which of the two DNA chains is copied. It is significant that single-stranded viral DNA does not act effectively as a template for the RNA polymerase (57). Single-stranded viral RNA, on the other hand, appears in several cases to work directly as a messenger (for example, see (58)). The RNA synthetase found in *E. coli* after infection with phage MS 2 responds very poorly to *added* single-stranded RNA. It is found attached to the double-stranded replicative form (23). This suggests that there may be a special sequence at one end of MS 2 RNA which is necessary as a starting point for the enzyme.

ON PROTEIN SYNTHESIS

As far as we know, nucleic acid has only two major functions. It can serve as a template for

the replication of more nucleic acid, or it can direct protein synthesis by acting as m-RNA. No authentic case of nucleic acid acting as an enzyme has yet been discovered.

Proteins, on the other hand, are as a class immensely versatile. Every biochemist knows that all known enzymes are proteins. What protein apparently cannot do easily is to form a regular double helix suitable for a simple replication mechanism. Nevertheless, it is not too difficult to imagine a rather complicated system which could replicate protein. I am not completely clear why this has not occurred (for there is certainly no evidence for it). It may be that there are theoretical reasons why one type of molecule should act as the genetic material, of which basically only one copy is required, and another type should act as enzymes, many copies of which are needed. Then again, the action of proteins demands that they form a relatively stable three-dimensional structure. To be copied, this would have to be unfolded. Whatever the reason, Nature has found it better to make nucleic acid the genetic material, although this means that very elaborate biochemical machinery is needed in order that it can direct protein synthesis. It is not easy to see how this could have arisen. It must have evolved by successive stages, but this fascinating problem is outside my present topic.

What is not clear at the moment is how much the details of the genetic code are dictated by stereochemistry or how much they are an arbitrary feature of the translation machinery. On the latter view, the fact that UUU codes for phenylalanine is due to the "accident" that the s-RNA molecule which recognizes this triplet on the m-RNA happens to have a part of its structure which fits the activating enzyme for phenylalanine. From the former viewpoint (43), there is something about the shape or chemistry of phenylalanine which necessarily makes it be coded by the triplet UUU. No definite stereochemical model has so far been presented to justify this idea (43), and in what follows I shall not consider it further.

The experimental evidence suggests that the

genetic code is universal, or nearly so, whereas the relationship between an s-RNA molecule and the activating enzyme is often species specific. The s-RNA molecules carrying leucine, which will accept leucine from the activating enzyme of the same species, may or may not accept leucine for the leucine-activating enzyme of a different species (see (59) for references). Nevertheless, the triplets for leucine appear, to a first approximation, to be the same in all species.

The paradox is easily resolved on two assumptions: 1) That the recognition site for the activating enzyme on the s-RNA is not identical to the coding site which recognizes the m-RNA. 2) That events requiring very many *simultaneous* mutations are very unlikely to occur, even over long periods of time.

The latter assumption would explain why the code is invariant. If the meaning of a triplet is changed, most of the genes in the organism will be affected, and the change is almost certain to be lethal. Thus, the code will never change once organisms have reached a certain minimum size. On the other hand, to change the relationship between an s-RNA and its activating enzyme only a few simultaneous changes might be needed, some on the gene which controls the s-RNA and some compensating change on the gene for the activating enzyme. Over very long periods of time we might well expect a drift of this sort to occur.

This argument leaves open the possibility that at any early stage in the evolution of the code there may have been some special relationship between a few of the amino acids and certain base sequences on the nucleic acid. It merely states that once an arbitrary (or partly arbitrary) code has arisen, it may be difficult to alter, although parts of the coding machinery may change nevertheless.

CONCLUDING REMARKS

The general scheme we have been considering for gene structure, gene replication and gene action appears to be basically the same throughout Nature. It combines simplicity with com-

plexity in a striking and characteristic manner. The device of carrying the genetic information on a polymer, using a language of four side-groups, is one of great power and elegance. The information is carried on a one-dimensional structure. The existence of the two base-pairs allows the genetic replication to be accomplished in a simple manner.

Complexity appears with the proteins. These are necessarily complex because proteins, as a class, need to carry out many different functions, both enzymatic and structural, and Natural Selection has refined them so that they do so with great precision. Their elaborate structure is

achieved by the way each protein folds up to form its unique three-dimensional conformation. Nevertheless, the underlying chemical plan is simple: that of twenty standard units, the amino acids, joined up in a uniform manner into polypeptide chains. However, they must be joined up in the correct sequence, and rather elaborate biochemical machinery is needed to enable the nucleic acid to direct this. The whole arrangement provides an immensely powerful instrument by which Natural Selection can operate. Allowing for the necessary limitations of macromolecular structure, it may well be the best that could be devised.

REFERENCES

For a lecture covering so much ground it has not been easy to provide adequate references. I have quoted a few rather recent papers, but mainly I have sent the reader to three special sources, of which the first has proved the most useful. References to these volumes have been abbreviated as follows.

CSH *The Cold Spring Harbor symposium on quantitative biology, Vol. 28, "Synthesis and Structure of Macromolecules"*, Long Island Biological Assoc., Cold Spring Harbor, Long Island, New York, 1963.

PNAR I *Progress in nucleic acid research*, Eds. J. N. DAVIDSON AND W. E. COHN, Academic Press Inc., New York. Vol. I, 1963; Vol. II, 1964.

Abst. *Abstracts of the VIth international congress of Biochemistry*, New York, 1964, IUB Vol. 32.

1. AVERY, O. T., MACLEOD, C. M., AND McCARTY, M., *J. Exp. Med.*, **79**, 137 (1944).
2. MARKHAM, R., Abst., p. 1.
3. SONNEBORN, T. M., *Proc. Natl. Acad. Sci. U.S.*, **51**, 915 (1964).
4. SMITH, M., *J. Mol. Biol.*, **9**, 17 (1964).
5. RUBENSTEIN, I., THOMAS, JR., C. A., AND HERSHEY, A. D., *Proc. Natl. Acad. Sci. U.S.*, **47**, 1113 (1961).
6. DAVISON, P. F., FREIFELDER, D., HEDE, R., AND LEVINTHAL, C., *Proc. Natl. Acad. Sci. U.S.*, **47**, 1123 (1961).

7. RIS, H., AND CHANDLER, B. L., CSH, p. 1.
8. CAIRNS, J., CSH, p. 43.
9. CAIRNS, J., Abst., p. 199.
10. KLEINSCHMIDT, A. K., Abst., p. 193.
11. EPSTEIN, R. H., BOLLE, A., STEINBERG, C. M., KELLENBERGER, E., BOY DE LA TOUR, E., CHEVALLEY, R., EDGAR, R. S., SUSMAN, M., DENHART, G. H., AND LIELAUSIS, A., CSH, p. 375.
12. LEE-HUANG, S., AND CAVALIERI, L. F., *Proc. Natl. Acad. Sci. U.S.*, **50**, 1116 (1963).
13. MESELSON, M., AND STAHL, F. W., *Proc. Natl. Acad. Sci. U.S.*, **44**, 671 (1958).
14. CAVALIERI, L. F., AND ROSENBERG, B. H., *Ann. Rev. Biochem.*, **31**, 247 (1962).
15. TAYLOR, J. H., in J. H. TAYLOR (Editor), *Molecular genetics, Part I*, Academic Press, Inc., New York, 1963, p. 65.
16. KORNBERG, A., in *Enzymatic synthesis of DNA*. Ciba Lectures in Microbial Biochemistry, John Wiley and Sons, New York, 1961, p. 14.
17. SCHILDKRAUT, C. L., RICHARDSON, C. C., AND KORNBERG, A., *J. Mol. Biol.*, **9**, 24 (1964).
18. BERG, P., FANCHER, H., AND CHAMBERLAIN, M., in H. VOGEL, V. BRYSON AND J. LAMPEN (Editors), *Symposium on informational macromolecules*, Academic Press, New York, 1963, p. 467.
19. RICHARDSON, C. C., LEHMAN, I. R., AND KORNBERG, A., *J. Biol. Chem.*, **239**, 251 (1964).
20. LEHMAN, I. R., in PNAR II, p. 84.

21. OKAZAKI, T., AND KORNBERG, A., *J. Biol. Chem.*, **239**, 259 (1964).
22. BALTIMORE, D., AND FRANKLIN, R. M., CSH, p. 105.
23. WEISSMANN, C., BORST, P., BURDON, R. H., BILLETTER, M. A., AND OCHOA, S., *Proc. Natl. Acad. Sci. U.S.*, **51**, 890 (1964).
24. OCHOA, S., WEISSMANN, C., BORST, P., BURDON, R. H., AND BILLETTER, M. A., *Federation Proc.*, in press.
25. JACOB, F., AND MONOD, J., *Cold Spring Harbor symposium on quantitative biology*, Vol. 26, Long Island Biological Assoc., Cold Spring Harbor, Long Island, New York, 1961, p. 193.
26. LIPMANN, F., in PNAR I, p. 135.
27. HURWITZ, J., AND AUGUST, J. T., in PNAR I, p. 59.
28. HURWITZ, J., EVANS, A., BABINET, C., AND SKALKA, A., CSH, p. 59.
29. CHAMBERLAIN, M., BERG, P., CSH, p. 67.
30. FOX, F. C., AND WEISS, S. B., *J. Biol. Chem.*, **239**, 175 (1964).
31. STEVENS, A., AND HENRY, J., *J. Biol. Chem.*, **239**, 196 (1964).
32. HAYASHI, M., HAYASHI, M. N., AND SPIEGELMAN, S., *Proc. Natl. Acad. Sci. U.S.*, **51**, 351 (1964).
33. CHAMBERLAIN, M., AND BERG, P., *Proc. Natl. Acad. Sci. U.S.*, **48**, 81 (1962).
34. GOMATOS, P. J., KRUG, R. M., AND TAMM, I., *J. Mol. Biol.*, **9**, 193 (1964).
35. SPIEGELMAN, S., AND HAYASHI, M., CSH, p. 161.
36. WATSON, J. D. in *Volume du cinquantenaire de la société de chimie biologique*, Maurice Declume, Lons-le-Saunier, 1964, in press.
37. CRICK, F. H. C., in PNAR I, p. 164.
38. NIRENBERG, M. W., JONES, O. W., LEDER, P., CLARK, B. F. C., SLY, W. S., AND PESTKA, S., CSH, p. 549.
39. SPEYER, J. F., LENGYEL, P., BASILIO, C., WAHBA, A. J., GARDNER, R. S., AND OCHOA, S., CSH, p. 559.
40. LEDER, P., AND NIRENBERG, M., *Proc. Natl. Acad. Sci. U.S.*, **52**, 420 (1964).
41. KHORANA, H. G., Abst., p. 3.
42. THACH, R. E., SUNDARARAJAN, T. A., AND DOTY, P., Abst., p. 90.
43. WEINSTEIN, I. B., CSH, p. 579.
44. YANOFSKY, C., CARLTON, B. C., GUEST, J. R., HELINSKI, D. R., AND HENNING, V., *Proc. Natl. Acad. Sci. U.S.*, **51**, 266 (1964).
45. SARABHAI, A. S., STRETTON, A. O. W., BRENNER, S., AND BOLLE, A., *Nature*, **201**, 13 (1964).
46. YANOFSKY, C., CSH, p. 581.
47. RICH, A., in A. KASHA AND B. PULLMAN (Editors), *Horizons in biochemistry*, Academic Press, New York, 1962, p. 103.
48. CRICK, F. H. C., AND WATSON, J. D., *Proc. Roy. Soc. (London) A*, **223**, 80 (1954).
49. PAULING, L., AND COREY, R. B., *Arch. Biochem. Biophys.*, **65**, 164 (1956).
50. O'BRIEN, E. J., *J. Mol. Biol.*, **7**, 107 (1963).
51. SOBELL, H. M., TOMITA, K., AND RICH, A., *Proc. Natl. Acad. Sci. U.S.*, **49**, 885 (1963).
52. HOOGSTEN, K., *Acta Cryst.*, **12**, 822 (1959).
53. HASCHEMEYER, A. E. V., AND SOBELL, H. M., *Proc. Natl. Acad. Sci. U.S.*, **50**, 872 (1963).
54. MATHEWS, F. S., AND RICH, A., *J. Mol. Biol.*, **3**, 89 (1964).
55. MILES, H. T., *Proc. Natl. Acad. Sci. U.S.*, **51**, 1104 (1964).
56. CAVALIERI, L. F., AND ROSENBERG, B. H., PNAR II, p. 2.
57. WOOD, W. B., AND BERG, P., CSH, p. 237.
58. NATHANS, D., NOTANI, G., SCHWARTZ, J. H., AND ZINDER, N. D., *Proc. Natl. Acad. Sci. U.S.*, **48**, 424 (1962).
59. BENNETT, T. P., GOLDSTEIN, J., AND LIPMANN, F., CSH, p. 233.