ON PROTEIN SYNTHESIS

By F. H. C. Crick
Medical Research Council Unit for the Study of Molecular Biology,
Cavendish Laboratory, Cambridge

I. INTRODUCTION

Protein synthesis is a large subject in a state of rapid development. To cover it completely in this article would be impossible. I have therefore deliberately limited myself here to presenting a broad general view of the problem, emphasizing in particular well-established facts which require explanation, and only selecting from recent work those experiments whose implications seem likely to be of lasting significance. Much very recent work, often of great interest, has been omitted because its implications are not clear. I have also tried to relate the problem to the other central problems of molecular biology—those of gene action and nucleic acid synthesis. In short, I have written for the biologist rather than the biochemist, the general reader rather than the specialist. More technical reviews have appeared recently by Borsook (1956), Spiegelman (1957), and Simkin & Work (1957*b* and this Symposium).

The importance of proteins

It is an essential feature of my argument that in biology proteins are uniquely important. They are not to be classed with polysaccharides, for example, which by comparison play a very minor role. Their nearest rivals are the nucleic acids. Watson said to me, a few years ago, 'The most significant thing about the nucleic acids is that we don't know what they do.' By contrast the most significant thing about proteins is that they can do almost anything. In animals proteins are used for structural purposes, but this is not their main role, and indeed in plants this job is usually done by polysaccharides. *The main function of proteins is to act as enzymes.* Almost all chemical reactions in living systems are catalysed by enzymes, and all known enzymes are proteins. It is at first sight paradoxical that it is probably easier for an organism to produce a new protein than to produce a new small molecule, since to produce a new small molecule one or more new proteins will be required in any case to catalyse the reactions.

I shall also argue that the main function of the genetic material is to control (not necessarily directly) the synthesis of proteins. There is a little direct evidence to support this, but to my mind the psychological drive behind this hypothesis is at the moment independent of such evidence.

Once the central and unique role of proteins is admitted there seems little point in genes doing anything else. Although proteins can *act* in so many different ways, the way in which they are *synthesized* is probably uniform and rather simple, and this fits in with the modern view that gene action, being based upon the nucleic acids, is also likely to be uniform and rather simple.

Biologists should not deceive themselves with the thought that some new class of biological molecules, of comparable importance to the proteins, remains to be discovered. This seems highly unlikely. In the protein molecule Nature has devised a unique instrument in which an underlying simplicity is used to express great subtlety and versatility; it is impossible to see molecular biology in proper perspective until this peculiar combination of virtues has been clearly grasped.

II. THE PROBLEM

Elementary facts about proteins

(1) *Composition.* Simple (unconjugated) proteins break down on hydrolysis to amino acids. There is good evidence that in a native protein the amino acids are condensed into long polypeptide chains. A typical protein, of molecular weight about 25,000, will contain some 230 residues joined end-to-end to form a single polypeptide chain.

Two points are important. First, the actual chemical step required to form the covalent bonds of the protein is always the same, irrespective of the amino acid concerned, namely the formation of the peptide link with the elimination of water. Apart from minor exceptions (such as S—S links and, sometimes, the attachment of a prosthetic group) *all* the covalent links within a protein are formed in this way. Covalently, therefore, a protein is to a large extent a linear molecule (in the topological sense) and there is little evidence that the backbone is ever branched. From this point of view the cross-linking by S—S bridges is looked upon as a secondary process.

The second important point—and I am surprised that it is not remarked more often—is that only about twenty different *kinds* of amino acids occur in proteins, and that these same twenty occur, broadly speaking, in *all* proteins, of whatever origin—animal, plant or micro-organism. Of course not every protein contains every amino acid—the amino acid tryptophan, which is one of the rarer ones, does not occur in insulin, for example—but the majority of proteins contain at least one of each of the twenty amino acids. In addition all these twenty amino acids (apart from glycine) have the L configuration when they occur in genuine proteins.

There are a few proteins which contain amino acids not found else-

where—the hydroxyproline of collagen is a good example—but in all such cases it is possible to argue that their presence is due to a modification of the protein after it has been synthesized or to some other abnormality. In Table 1, I have listed the standard twenty amino acids believed to be of universal occurrence and also, in the last column, some of the exceptional ones. The assignment given in Table 1 might not be agreed by everyone,

Table 1

The magic twenty amino acids found universally in proteins		Other amino acids found in proteins
Glycine	Asparagine	Hydroxyproline*
Alanine	Glutamine	Hydroxylysine
Valine	Aspartic acid	Phosphoserine
Leucine	Glutamic acid	Diaminopimelic acid
Isoleucine	Arginine	Thyroxine and related molecules
Proline*	Lysine	
Phenylalanine	Histidine	Cystine†
Tyrosine	Tryptophan	
Serine	Cysteine†	
Threonine	Methionine	

* These are, of course, imino acids. This distinction is not made in the text.
† This classification implies that all the cystine found in proteins is formed by the joining together of two cysteine molecules.

as the evidence is incomplete, but more agreement could be found for this version than for any other. Curiously enough this point is slurred over by almost all biochemical textbooks, the authors of which give the impression that they are trying to include as many amino acids in their lists as they can, without bothering to distinguish between the magic twenty and the others. (But see a recent detailed review by Synge, 1957.)

(2) *Homogeneity.* Not only is the composition of a given protein fixed, but we have every reason to believe that the exact order of the amino acid residues along the polypeptide chains is also rigidly determined: that each molecule of haemoglobin in your blood, for example, has exactly the same sequence of amino acids as every other one. This is clearly an overstatement; the mechanism must make mistakes sometimes, and, as we shall see, there are also interesting exceptions which are under genetic control. Moreover, it is quite easy, in extracting a protein, to modify some of the molecules slightly without affecting the others, so that the 'pure' protein may appear heterogeneous. The exact amount of 'microheterogeneity' of proteins is controversial (see the review by Steinberg & Mihalyi, 1957), but this should not blind one to the astonishing degree of homogeneity of most proteins.*

(3) *Structure.* In a native globular protein the polypeptide chain is not

* The γ-globulins and other antibody molecules are exceptions to these generalizations. They are probably heterogeneous in folding and possibly to some extent in composition.

fully extended but is thrown into folds and superfolds, maintained by weak physical bonds, and in some cases by covalent —S—S— links and possibly some others. This folding is also thought to be at least broadly the same for each copy of a particular protein, since many proteins can be crystallized, though the evidence for *perfect* homogeneity of folding is perhaps rather weak.* As is well known, if this folding is destroyed by heat or other methods the protein is said to be 'denatured'. The biological properties of most proteins, especially the catalytic action of enzymes, must depend on the exact spatial arrangement of certain side-groups on the surface of the protein, and altering this arrangement by unfolding the polypeptide chains will destroy the biological specificity of the proteins.

(4) *Amino acid requirements.* If one of the twenty amino acids is supplied to a cell it can be incorporated into proteins; amino acids are certainly protein precursors. The only exceptions are amino acids like hydroxyproline, which are not among the magic twenty. The utilization of peptides is controversial but the balance of evidence is against the occurrence of peptide intermediates. (See the discussion by Simkin & Work, this Symposium.)

If, for some reason, one of the twenty amino acids is not available to the organism, protein synthesis stops. Moreover, the continued synthesis of those parts of the protein molecules which do not contain that amino acid appears not to take place. This can be demonstrated particularly clearly in bacteria, but it is also true of higher animals. If a meal is provided that lacks an essential amino acid it is no use trying to make up for this deficiency by providing it a few hours later.

Very little is known about the accuracy with which the amino acids are selected. One would certainly expect, for example, that the mechanism would occasionally put a valine into an isoleucine site, but exactly how often this occurs is not known. The impression one gets from the rather meagre facts at present available is that mistakes occur rather infrequently.

In recent years it has been possible to introduce amino acid analogues into proteins by supplying the analogue under circumstances in which the amino acid itself is not easily available (see the review by Kamin & Handler, 1957). For example in *Escherichia coli* fluorophenylalanine has been incorporated in place of phenylalanine and tyrosine (Munier & Cohen, 1956) and it has even proved possible to replace completely the sulphur-containing amino acid methionine by its selenium analogue (Cohen & Cowie, 1957). Of the enzymes produced by the cell in these various ways some were active and some were inactive, as might have been expected.

(5) *Contrast with polysaccharides.* It is useful at this point to contrast proteins with polysaccharides to underline the differences between them.

* See previous footnote.

(I do not include nucleic acids among the polysaccharides.) Polysaccharides, too, are polymers, but each one is constructed from one, or at the most only about half-a-dozen kinds of monomer. Nevertheless many different monomers are found throughout Nature, some occurring here, some there. There is no standard set of monomers which is always used, as there is for proteins. Then polysaccharides are polydisperse—at least so far no monodisperse one has been found—and the order of their monomers is unlikely to be rigidly controlled, except in some very simple manner. Finally in those cases which have been carefully studied, such as starch, glycogen and hyaluronic acid, it has been found that the polymerization is carried out in a straightforward way by enzymes.

(6) *The genetics and taxonomy of proteins.* It is instructive to compare your own haemoglobin with that of a horse. Both molecules are indistinguishable in size. Both have similar amino acid compositions; similar but not identical. They differ a little electrophoretically, form different crystals, and have slightly different ends to their polypeptide chains. All these facts are compatible with their polypeptide chains having similar amino acid sequences, but with just a few changes here and there.

This 'family likeness' between the 'same' protein molecules *from different species* is the rule rather than the exception. It has been found in almost every case in which it has been looked for. One of the best-studied examples is that of insulin, by Sanger and his co-workers (Brown, Sanger & Kitai, 1955; Harris, Sanger & Naughton, 1956), who have worked out the complete amino acid sequences for five different species, only two of which (pig and whale) are the same. Interestingly enough the differences are all located in one small segment of one of the two chains.

Biologists should realize that before long we shall have a subject which might be called 'protein taxonomy'—the study of the amino acid sequences of the proteins of an organism and the comparison of them between species. It can be argued that these sequences are the most delicate expression possible of the phenotype of an organism and that vast amounts of evolutionary information may be hidden away within them.

There is, however, nothing in the evidence presented so far to prove that these differences between species are under the control of Mendelian genes. It could be argued that they were transmitted cytoplasmically through the egg. On the other hand, there is much evidence that genes do affect enzymes, especially from work on micro-organisms such as *Neurospora* (see Wagner & Mitchell, 1955). The famous 'one gene—one enzyme' hypothesis (Beadle, 1945) expresses this fact, although its truth is controversial (personally I believe it to be largely correct). However, in none of these cases has the protein (the enzyme, that is) ever been obtained pure.

There are a few cases where a Mendelian gene has been shown unambiguously to alter a protein, the most famous being that of human sickle-cell-anaemia haemoglobin, which differs electrophoretically from normal adult haemoglobin, as was discovered by Pauling and his co-workers (1949). Until recently it could have been argued that this was perhaps not due to a change in amino acid sequence, but only to a change in the folding. That the gene does in fact alter the amino acid sequence has now been conclusively shown by my colleague, Dr Vernon Ingram. The difference is due to a valine residue occurring in the place of a glutamic acid one, and Ingram has suggestive evidence that this is the *only* change (Ingram, 1956, 1957). It may surprise the reader that the alteration of one amino acid out of a total of about 300 can produce a molecule which (when homozygous) is usually lethal before adult life but, for my part, Ingram's result is just what I expected.

The nature of protein synthesis

The basic dilemma of protein synthesis has been realized by many people, but it has been particularly aptly expressed by Dr A. L. Dounce (1956):

> My interest in templates, and the conviction of their necessity, originated from a question asked me on my Ph.D. oral examination by Professor J. B. Sumner. He enquired how I thought proteins might be synthesized. I gave what seemed the obvious answer, namely, that enzymes must be responsible. Professor Sumner then asked me the chemical nature of enzymes, and when I answered that enzymes were proteins or contained proteins as essential components, he asked whether these enzyme proteins were synthesized by other enzymes and so on *ad infinitum*.
>
> The dilemma remained in my mind, causing me to look for possible solutions that would be acceptable, at least from the standpoint of logic. The dilemma, of course, involves the specificity of the protein molecule, which doubtless depends to a considerable degree on the sequence of amino acids in the peptide chains of the protein. The problem is to find a reasonably simple mechanism that could account for specific sequences without demanding the presence of an ever-increasing number of new specific enzymes for the synthesis of each new protein molecule.

It is thus clear that the synthesis of proteins must be radically different from the synthesis of polysaccharides, lipids, co-enzymes and other small molecules; that it must be relatively simple, and to a considerable extent uniform throughout Nature; that it must be highly specific, making few mistakes; and that in all probability it must be controlled at not too many removes by the genetic material of the organism.

The essence of the problem

A systematic discussion of our present knowledge of protein synthesis could usefully be set out under three headings, each dealing with a flux:

the flow of energy, the flow of matter, and the flow of information. I shall not discuss the first of these here. I shall have something to say about the second, but I shall particularly emphasize the third—the flow of information.

By information I mean the specification of the amino acid sequence of the protein. It is conventional at the moment to consider separately the synthesis of the polypeptide chain and its folding. It is of course possible that there is a special mechanism for folding up the chain, but the more likely hypothesis is that the *folding is simply a function of the order of the amino acids*, provided it takes place as the newly formed chain comes off the template. I think myself that this latter idea may well be correct, though I would not be surprised if exceptions existed, especially the γ-globulins and the adaptive enzymes.

Our basic handicap at the moment is that we have no easy and precise technique with which to study how proteins are folded, whereas we can at least make some experimental approach to amino acid sequences. For this reason, if for no other, I shall ignore folding in what follows and concentrate on the determination of sequences. It is as well to realize, however, that the idea that the two processes can be considered separately is in itself an assumption.

The actual chemical step by which any two amino acids (or activated amino acids) are joined together is probably always the same, and may well not differ significantly from any other biological condensation. The unique feature of protein synthesis is that only a single standard set of twenty amino acids can be incorporated, and that for any particular protein *the amino acids must be joined up in the right order*. It is this problem, the problem of 'sequentialization', which is the crux of the matter, though it is obviously important to discover the exact chemical steps which lead up to and permit the crucial act of sequentialization.

As in even a small bacterial cell there are probably a thousand different kinds of protein, each containing some hundreds of amino acids in its own rigidly determined sequence, the amount of hereditary information required for sequentialization is quite considerable.

III. RECENT EXPERIMENTAL WORK

The role of the nucleic acids

It is widely believed (though not by every one) that the nucleic acids are in some way responsible for the control of protein synthesis, either directly or indirectly. The actual evidence for this is rather meagre. In the case of deoxyribonucleic acid (DNA) it rests partly on the T-even bacteriophages, since it has been shown, mainly by Hershey and his colleagues, that

whereas the DNA of the infecting phage penetrates into the bacterial cell almost all the protein remains outside (see the review by Hershey, 1956); and also on Transforming Factor, which appears to be pure DNA, and which in at least one case, that of the enzyme mannitol phosphate dehydrogenase, controls the synthesis of a protein (Marmur & Hotchkiss, 1955). There is also the indirect evidence that DNA is the most constant part of the genetic material, and that genes control proteins. Finally there is the very recent evidence, mainly due to the work of Benzer on the rII locus of bacteriophage, that the functional gene—the 'cistron' of Benzer's terminology—consists of many sites arranged strictly *in a linear order* (Benzer, 1957) as one might expect if a gene controls the order of the amino acids in some particular protein.

As is well known, the correlation between ribonucleic acid (RNA) and protein synthesis was originally pointed out by Brachet and by Caspersson. Is there any more direct evidence for this connexion? In particular is there anything to support the idea that the sequentialization of the amino acids is controlled by the RNA?

The most telling evidence is the recent work on tobacco mosaic virus. A number of strains of the virus are known, and it is not difficult to show (since the protein sub-unit of the virus is small) that they differ in amino acid composition. Some strains, for example, have histidine in their protein, whereas others have none. Two very significant experiments have been carried out. In one, as first shown by Gierer & Schramm (1956), the RNA of the virus alone, although completely free of protein, appears to be infective, though the infectivity is low. In the other, first done by Fraenkel-Conrat, it has proved possible to separate the RNA from the protein of the virus and then recombine them to produce virus again. In this case the infectivity is comparatively high, though some of it is usually lost. If a recombined virus is made using the RNA of one strain and the protein of another, and then used to infect the plant, the new virus produced in the plant resembles very closely *the strain from which the RNA was taken*. If this strain had a protein which contained no histidine then the offspring will have no histidine either, although the plant had never been in contact with this particular protein before but only with the RNA from that strain. In other words *the viral RNA appears to carry at least part of the information which determines the composition of the viral protein*. Moreover the viral protein which was used to infect the cell was not copied to any appreciable extent (Fraenkel-Conrat, 1956).

It has so far not proved possible to carry out this experiment—a model of its kind—in any other system, although very recently it has been claimed that for two animal viruses the RNA alone appears to be infective.

Turnover experiments have shown that while the labelling of DNA is homogeneous that of RNA is not. The RNA of the cell is partly in the nucleus, partly in particles in the cytoplasm and partly as the 'soluble' RNA of the cell sap; many workers have shown that all these three fractions turn over differently. It is very important to realize in any discussion of the role of RNA in the cell that it is very inhomogeneous metabolically, and probably of more than one type.

The site of protein synthesis

There is no known case in Nature in which protein synthesis proper (as opposed to protein modification) occurs outside cells, though, as we shall see later, a certain amount of protein can probably be synthesized using broken cells and cell fragments. The first question to ask, therefore, is whether protein synthesis can take place in the nucleus, in the cytoplasm, or in both.

It is almost certain that protein synthesis can take place in the cytoplasm without the presence of the nucleus, and it is probable that it can take place to some extent in the nucleus by itself (see the review by Brachet & Chantrenne, 1956). Mirsky and his colleagues (see the review by Mirsky, Osawa & Allfrey, 1956) have produced evidence that some protein synthesis can occur in isolated nuclei, but the subject is technically difficult and in this review I shall quite arbitrarily restrict myself to protein synthesis in the cytoplasm.

In recent years our knowledge of the structure of the cytoplasm has enormously increased, due mainly to the technique of cutting thin sections for the electron microscope. The cytoplasm of many cells contains an 'endoplasmic reticulum' of double membranes, consisting mainly of protein and lipid (see the review of Palade, 1956). On one side of each membrane appear small electron-dense particles (Palade, 1955). Biochemical studies (Palade & Siekevitz, 1956; among others) have shown that these particles, which are about 100–200 Å. in diameter, consist almost entirely of protein and RNA, in about equal quantities. Moreover the major part of the RNA of the cell is found in these particles.

When such a cell is broken open and the contents fractionated by centrifugation, the particles, together with fragments of the endoplasmic reticulum, are found in the 'microsome' fraction, and for this reason I shall refer to them as microsomal particles.

These microsomal particles are found in almost all cells. They are particularly common in cells which are actively synthesizing protein, whereas the endoplasmic reticulum is most conspicuously present in (mammalian) cells which are secreting very actively. Thus both the cells

of the pancreas and those of an ascites tumour contain large quantities of microsomal particles, but the tumour has little endoplasmic reticulum, whereas the pancreas has a lot. Moreover, there is no endoplasmic reticulum in bacteria.

On the other hand particles of this general description have been found in plant cells (Ts'o, Bonner & Vinograd, 1956), in yeast, and in various bacteria (Schachman, Pardee & Stanier, 1953); in fact in all cells which have been examined for them.

These particles have been isolated from various cells and examined in the ultra-centrifuge (Petermann, Mizen & Hamilton, 1952; Schachman *et al.* 1953; among others). The remarkable fact has emerged that they do not have a continuous distribution of sedimentation constants, but usually fall into several well-defined groups. Moreover some of the particles are probably simple aggregates of the others (Petermann & Hamilton, 1957). This uniformity suggests immediately that the particles, which have 'molecular weights' of a few million, have a definite structure. They are, in fact, reminiscent of the small spherical RNA-containing viruses, and Watson and I have suggested that they may have a similar type of substructure (Crick & Watson, 1956).

Biologists should contrast the older concept of *microsomes* with the more recent and significant one of *microsomal particles*. Microsomes came in all sizes, and were irregular in composition; microsomal particles occur in a few sizes only, have a more fixed composition and a much higher proportion of RNA. It was hard to identify microsomes in all cells, whereas RNA-rich particles appear to occur in almost every kind of cell. In short, microsomes were rather a mess, whereas microsomal particles appeal immediately to one's imagination. It will be surprising if they do not prove to be of fundamental importance.

It should be noted, however, that Simpson and his colleagues (Simpson & McLean, 1955; Simpson, McLean, Cohn & Brandt, 1957) have reported that protein synthesis can take place in mitochondria. It is known that mitochondria contain RNA, and it would be of great interest to know whether this RNA is in some kind of particle. Mitochondria are, of course, very widely distributed but they do not occur in lower forms such as bacteria. Similar remarks about RNA apply to the reported incorporation in chloroplasts (Stephenson, Thimann & Zamecnik, 1956).

Microsomal particles and protein synthesis

It has been shown by the use of radioactive amino acids that during protein synthesis the amino acids appear to flow through the microsomal particles. The most striking experiments are those of Zamecnik and his

co-workers on the livers of growing rats (see the review by Zamecnik *et al.* 1956).

Two variations of the experiment were made. In the first the rat was given a rather large intravenous dose of a radioactive amino acid. After a predetermined time the animal was sacrificed, the liver extracted, its cells homogenized and the contents fractionated. It was found that the microsomal particle fraction was very rapidly labelled to a constant level.

In the second a very small shot of the radioactive amino acid was given, so that the liver received only a pulse of labelled amino acid, since this small amount was quickly used up. In this case the radioactivity of the microsomal particles rose very quickly *and then fell away*. Making plausible assumptions Zamecnik and his colleagues have shown that this behaviour is what one would expect if most of the protein of the microsomal particles was metabolically inert, but 1 or 2% was turning over very rapidly, say within a minute or so.

Very similar results have been obtained by Rabinovitz & Olson (1956, 1957) using intact mammalian cells, in this case rabbit reticulocytes. They have also been able to show that the label passed into a well-defined globular protein, namely haemoglobin. Experiments along the same general lines have also been reported for liver by Simkin & Work (1957*a*).

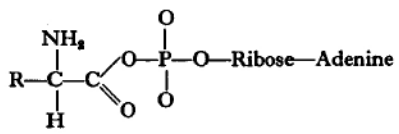
We thus have direct experimental evidence that the microsomal particles are associated with protein synthesis, though the precise role they play is not clear.

Activating enzymes

It now seems very likely that the first step in protein synthesis is the activation of each amino acid by means of its special 'activating enzyme'. The activation requires ATP, and the evidence suggests that the reaction is

$$\text{amino acid} + \text{ATP} = \text{AMP} - \text{amino acid} + \text{pyrophosphate}.$$

The activated amino acid, which is probably a mixed anhydride of the form



in which the carboxyl group of the amino acid is phosphorylated, appears to be tightly bound to its enzyme and is not found free in solution.

These enzymes were first discovered in the cell-sap fraction of rat liver cells by Hoagland (Hoagland, 1955; Hoagland, Keller & Zamecnik, 1956) and in yeast by Berg (1956). They have been shown by DeMoss & Novelli (1956) to be widely distributed in bacteria, and it is surmised that they occur

in all cells engaged in protein synthesis. Recently Cole, Coote & Work (1957) have reported their presence in a variety of tissues from a number of animals.

So far good evidence has been found for this reaction for about half the standard twenty amino acids, but it is believed that further research will reveal the full set. Meanwhile Davie, Koningsberger & Lipmann (1956) have purified the tryptophan-activating enzyme. It is specific for tryptophan (and certain tryptophan analogues) and will only handle the L-isomer. Isolation of the tyrosine enzyme has also been briefly reported (Koningsberger, van de Ven & Overbeck, 1957; Schweet, 1957).

The properties of these enzymes are obviously of the greatest interest, and much work along these lines may be expected in the near future. For example, it has been shown that the tryptophan-activating enzyme contains what is probably a derivative of guanine (perhaps GMP) very tightly bound. It is possible to remove it, however, and to show that its presence is not necessary for the primary activation step. Since the enzyme is probably involved in the next step in protein synthesis it is naturally suspected that the guanine derivative is also required for this reaction, whatever it may be.

In vitro *incorporation*

In order to study the relationship between the activating enzymes and the microsomal particles it has proved necessary to break open the cells and work with certain partly purified fractions. Unfortunately it is rare to obtain substantial net protein synthesis from such systems, and there is a very real danger that the incorporation of the radioactivity does not represent true synthesis but is some kind of partial synthesis or exchange reaction. This distinction has been clearly brought out by Gale (1953). The work to be described, therefore, has to be accepted with reservations. (See the remarks of Simkin & Work, this Symposium.) It has been shown, however, in the work described below, that the amino acid is incorporated into true peptide linkage.

Again the significant results were first obtained by Zamecnik and his co-workers (reviewed in Zamecnik *et al.* 1956). The requirements so far known appear to fall into two parts:

(1) The activation of the amino acids for which, in addition to the labelled amino acid, one requires the 'pH 5' fraction, containing the activating enzymes, ATP and (usually) an ATP-generating system. There appears to be no requirement for any of the pyrimidine or guanine nucleotides.

(2) The transfer to the microsomal particles. For this one requires the previous system plus GTP or GDP (Keller & Zamecnik, 1956) and of

course the microsomal particles; the endoplasmic reticulum does not appear to be necessary (Littlefield & Keller, 1957).

Hultin & Beskow (1956) have reported an experiment which shows clearly that the amino acids become bound in some way. They first incubate the mixture described in (1) above. They then add a great excess of *un*labelled amino acid before adding the microsomal particles. Nevertheless some of the labelled amino acid is incorporated into protein, showing that it was in some place where it could not readily be diluted.

Very recently an intermediate reaction has been suggested by the work of Hoagland, Zamecnik & Stephenson (1957), who have discovered that in the first step the 'soluble' RNA contained in the 'pH 5' fraction became labelled with the radioactive amino acid. The bond between the amino acid and the RNA appears to be a covalent one. This labelled RNA can be extracted, purified, and then added to the microsomal fraction. In the presence of GTP the labelled amino acid is transferred from the soluble RNA to microsomal protein. This very exciting lead is being actively pursued.

Many other experiments have been carried out on cell-free systems, in particular by Gale & Folkes (1955) and by Spiegelman (see his review, 1957), but I shall not describe them here as their interpretation is difficult. It should be mentioned that Gale (reviewed in Gale, 1956) has isolated from hydrolysates of commercial-yeast RNA a series of fractions which greatly increase amino acid incorporation. One of them, the so-called 'glycine incorporation factor' has been purified considerably, and an attempt is being made to discover its structure.

RNA turnover and protein synthesis

From many points of view it seems highly likely that the *presence* of RNA is essential for cytoplasmic protein synthesis, or at least for specific protein synthesis. It is by no means clear, however, that the *turnover* of RNA is required.

In discussing this a strong distinction must be made between cells which are growing, and therefore producing new microsomal particles, and cells which are synthesizing without growth, and in which few new microsomal particles are being produced.

This is a difficult aspect of the subject as the evidence is to some extent conflicting. It appears reasonably certain that not *all* the RNA in the cytoplasm is turning over very rapidly—this has been shown, for example, by the Hokins (1954) working on amylase synthesis in slices of pigeon pancreas, though in the light of the recent work of Straub (this Symposium) the choice of amylase was unfortunate. On the other hand Pardee (1954)

has demonstrated that mutants of *Escherichia coli* which require uracil or adenine cannot synthesize β-galactosidase unless the missing base is provided.

Can RNA be synthesized without protein being synthesized? This can be brought about by the use of chloramphenicol. In bacterial systems chloramphenicol stops protein synthesis dead, but allows 'RNA' synthesis to continue. A very interesting phenomenon has been uncovered in *E. coli* by Pardee & Prestidge (1956), and by Gros & Gros (1956). If a mutant is used which requires, say, leucine, then when the external supply of leucine is exhausted both protein and RNA synthesis cease. If now chloramphenicol is added there is no effect, but if in addition the cells are given a small amount of leucine then rapid RNA synthesis takes place. If the chloramphenicol is removed, so that protein synthesis restarts, then this leucine is built into proteins and then, once again, the synthesis of both protein and RNA is prevented. In other words it appears as if 'free' leucine (i.e. not bound into proteins) is required for RNA synthesis. This effect is not peculiar to leucine and has already been found for several amino acids and in several different organisms (Yčas & Brawerman, 1957).

As a number of people have pointed out, the most likely interpretation of these results is that protein and RNA require *common intermediates* for their synthesis, consisting in part of amino acids and in part of RNA components such as nucleotides. This is a most valuable idea; it explains a number of otherwise puzzling facts and there is some hope of getting close to it experimentally.

For completeness it should be stated that Anfinsen and his co-workers have some evidence that proteins are not produced from (activated) amino acids in a single step (see the review by Steinberg, Vaughan & Anfinsen, 1956), since they find unequal labelling between the same amino acid at different points on the polypeptide chain, but this interpretation of their results is not accepted by all workers in the field. This is discussed more fully by Simkin & Work (this Symposium).

Summary of experimental work

Both DNA and RNA have been shown to carry some of the specificity for protein synthesis. The RNA of almost all types of cell is found mainly in rather uniform, spherical, virus-like particles in the cytoplasm, known as microsomal particles. Most of their protein and RNA is metabolically rather inert. Amino acids, on their way into protein, have been shown to pass rapidly through these particles.

An enzyme has been isolated which, when supplied with tryptophan and ATP, appears to form an activated tryptophan. There is evidence that

there exist similar enzymes for most of the other amino acids. These enzymes are widely distributed in Nature.

Work on cell fractions is difficult to interpret but suggests that the first step in protein synthesis involves these enzymes, and that the subsequent transfer of the activated amino acids to the microsomal particles requires GTP. The soluble RNA also appears to be involved in this process.

Whereas the presence of RNA is probably required for true protein synthesis its rapid turnover does not appear to be necessary, at least not for all the RNA. There is suggestive evidence that common intermediates, containing both amino acids and nucleotides, occur in protein synthesis.

IV. IDEAS ABOUT PROTEIN SYNTHESIS

It is an extremely difficult matter to present current ideas about protein synthesis in a stimulating form. Many of the general ideas on the subject have become rather stale, and an extended discussion of the more detailed theories is not suitable in a paper for non-specialists. I shall therefore restrict myself to an outline sketch of my own ideas on cytoplasmic protein synthesis, some of which have not been published before. Finally I shall deal briefly with the problem of 'coding'.

General principles

My own thinking (and that of many of my colleagues) is based on two general principles, which I shall call the Sequence Hypothesis and the Central Dogma. The direct evidence for both of them is negligible, but I have found them to be of great help in getting to grips with these very complex problems. I present them here in the hope that others can make similar use of them. Their speculative nature is emphasized by their names. It is an instructive exercise to attempt to build a useful theory without using them. One generally ends in the wilderness.

The Sequence Hypothesis

This has already been referred to a number of times. In its simplest form it assumes that the specificity of a piece of nucleic acid is expressed solely by the sequence of its bases, and that this sequence is a (simple) code for the amino acid sequence of a particular protein.

This hypothesis appears to be rather widely held. Its virtue is that it unites several remarkable pairs of generalizations: the central biochemical importance of proteins and the dominating biological role of genes, and in particular of their nucleic acid; the linearity of protein molecules (considered covalently) and the genetic linearity within the functional gene, as shown by the work of Benzer (1957) and Pontecorvo (this Symposium); the simplicity

of the composition of protein molecules and the simplicity of the nucleic acids. Work is actively proceeding in several laboratories, including our own, in an attempt to provide more direct evidence for this hypothesis.

The Central Dogma

This states that once 'information' has passed into protein *it cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein.

This is by no means universally held—Sir Macfarlane Burnet, for example, does not subscribe to it—but many workers now think along these lines. As far as I know it has not been *explicitly* stated before.

Some ideas on cytoplasmic protein synthesis

From our assumptions it follows that there must be an RNA template in the cytoplasm. The obvious place to locate this is in the microsomal particles, because their uniformity of size suggests that they have a regular structure. It also follows that the synthesis of at least some of the microsomal RNA must be under the control of the DNA of the nucleus. This is because the amino acid sequence of the human haemoglobins, for example, is controlled at least in part by a Mendelian gene, and because spermatozoa contain no RNA. Therefore, granted our hypotheses, the information must be carried by DNA.

What can we guess about the structure of the microsomal particle? On our assumptions the protein component of the particles can have no significant role in determining the amino acid sequence of the proteins which the particles are producing. We therefore assume that their main function is a structural one, though the possibility of some enzyme activity is not excluded. The simplest model then becomes one in which each particle is made of the same protein, or proteins, as every other one in the cell, and has the same basic *arrangement* of the RNA, but that different particles have, in general, different base-sequences in their RNA, and therefore produce different proteins. This is exactly the type of structure found in tobacco mosaic virus, where the interaction between RNA and protein does not depend upon the sequence of bases of the RNA (Hart & Smith, 1956). In addition Watson and I have suggested (Crick & Watson, 1956), by analogy with the spherical viruses, that the protein of microsomal particles is probably made of many identical sub-units arranged with cubic symmetry.

On this oversimplified picture, therefore, the microsomal particles in a cell are all the same (except for the base-sequence of their RNA) and are metabolically rather inert. The RNA forms the template and the protein supports and protects the RNA.

This idea is in sharp contrast to what one would naturally assume at first glance, namely that the protein of the microsomal particles consists entirely of protein being synthesized. The surmise that most of the protein is structural was derived from considerations about the structure of virus particles and about coding; it was independent of the direct experimental evidence of Zamecnik and his colleagues that only a small fraction of the protein turns over rapidly, so that this agreement between theory and experiment is significant, as far as it goes.

It is obviously of the first importance to know how the RNA of the particles is arranged. It is a natural deduction from the Sequence Hypothesis that the RNA backbone will follow as far as possible a spatially regular path, in this case a helix, essentially because the fundamental operation of making the peptide link is always the same, and we therefore expect any template to be spatially regular.

Although we do not yet know the structure of isolated RNA (which may be an artifact) we do know that a pair of RNA-like molecules can under some circumstances form a double-helical structure, somewhat similar to DNA, because Rich & Davies (1956) have shown that when the two polyribotides, polyadenylic acid and polyuridylic acid (which have the same backbone as RNA) are mixed together they wind round one another to form a double helix, presumably with their bases paired. It would not be surprising, therefore, if the RNA backbone took up a helical configuration similar to that found for DNA.

This suggestion is in contrast to the idea that the RNA and protein interact in a complicated, irregular way to form a 'nucleoprotein'. As far as I know there is at the moment no direct experimental evidence to decide between these two points of view.

However, even if it turns out that the RNA is (mainly) helical and that the structural protein is made of sub-units arranged with cubic symmetry it is not at all obvious how the two could fit together. In abstract terms the problem is how to arrange a long fibrous object inside a regular polyhedron. It is for this reason that the structure of the spherical viruses is of great interest in this context, since we suspect that the same situation occurs there; moreover they are at the moment more amenable to experimental attack. A possible arrangement, for example, is one in which the axes of the RNA helices run radially and clustered in groups of five, though it is always possible that the arrangement of the RNA is irregular.

It would at least be of some help if the approximate location of the RNA in the microsomal particles could be discovered. Is it on the outside or the inside of the particles, for example, or even both? Is the microsomal particle a rather open structure, like a sponge, and if it is what size of molecule can diffuse in and out of it? Some of these points are now ripe for a direct experimental attack.

The adaptor hypothesis

Granted that the RNA of the microsomal particles, regularly arranged, is the template, how does it direct the amino acids into the correct order? One's first naïve idea is that the RNA will take up a configuration capable of forming twenty different 'cavities', one for the side-chain of each of the twenty amino acids. If this were so one might expect to be able to play the problem backwards—that is, to find the configuration of RNA by trying to form such cavities. All attempts to do this have failed, and on physical-chemical grounds the idea does not seem in the least plausible (Crick, 1957*a*). Apart from the phosphate-sugar backbone, which we have assumed to be regular and perhaps linked to the structural protein of the particles, RNA presents mainly a sequence of sites where hydrogen bonding could occur. One would expect, therefore, that whatever went on to the template in a *specific* way did so by forming hydrogen bonds. It is therefore a natural hypothesis that the amino acid is carried to the template by an 'adaptor' molecule, and that the adaptor is the part which actually fits on to the RNA. In its simplest form one would require twenty adaptors, one for each amino acid.

What sort of molecules such adaptors might be is anybody's guess. They might, for example, be proteins, as suggested by Dounce (1952) and by the Hokins (1954) though personally I think that proteins, being rather large molecules, would take up too much space. They might be quite unsuspected molecules, such as amino sugars. But there is one possibility which seems inherently more likely than any other—that they might contain nucleotides. This would enable them to join on to the RNA template by the same 'pairing' of bases as is found in DNA, or in polynucleotides.

If the adaptors were small molecules one would imagine that a separate enzyme would be required to join each adaptor to its own amino acid and that the specificity required to distinguish between, say, leucine, isoleucine and valine would be provided by these enzyme molecules instead of by cavities in the RNA. Enzymes, being made of protein, can probably make such distinctions more easily than can nucleic acid.

An outline picture of the early stages of protein synthesis might be as

follows: the template would consist of perhaps a single chain of RNA. (As far as we know a single isolated RNA backbone has no regular configuration (Crick, 1957*b*) and one has to assume that the backbone is supported in a helix of the usual type by the structural protein of the microsomal particles.) Alternatively the template might consist of a pair of chains. Each adaptor molecule containing, say, a di- or trinucleotide would each be joined to its own amino acid by a special enzyme. These molecules would then diffuse to the microsomal particles and attach to the proper place on the bases of the RNA by base-pairing, so that they would then be in a position for polymerization to take place.

It will be seen that we have arrived at the idea of common intermediates without using the direct experimental evidence in their favour; but there is one important qualification, namely that the nucleotide part of the intermediates must be specific for each amino acid, at least to some extent. It is not sufficient, from this point of view, merely to join adenylic acid to each of the twenty amino acids. Thus one is led to suppose that after the activating step, discovered by Hoagland and described earlier, some other more specific step is needed before the amino acid can reach the template.

The soluble RNA

If trinucleotides, say, do in fact play the role suggested here their synthesis presents a puzzle, since one would not wish to invoke too many enzymes to do the job. It seems to me plausible, therefore, that the twenty different adaptors may be synthesized by the *breakdown* of RNA, probably the 'soluble' RNA. Whether this is in fact the same action which the 'activating enzymes' carry out (presumably using GTP in the process) remains to be seen.

From this point of view the RNA with amino acids attached reported recently by Hoagland, Zamecnik & Stephenson (1957), would be a halfway step in this process of breaking the RNA down to trinucleotides and joining on the amino acids. Of course alternative interpretations are possible. For example, one might surmise that numerous amino acids become attached to this RNA and then proceed to polymerize, perhaps inside the microsomal particles. I do not like these ideas, because the supernatant RNA appears to be too short to code for a complete polypeptide chain, and yet too long to join on to template RNA (in the microsomal particles) by base-pairing, since it would take too great a time for a piece of RNA twenty-five nucleotides long, say, to diffuse to the correct place in the correct particles. If it were only a trinucleotide on the other hand, there would be many different 'correct' places for it to go to (wherever a valine was required, say), and there would be no undue delay.

Leaving theories on one side, it is obviously of the greatest interest to know what molecules actually pass from the 'pH 5 enzymes' to the microsomal particles. Are they small molecules, free in solution, or are they bound to protein? Can they be isolated? This seems at the moment to be one of the most fruitful points at which to attack the problem.

Subsequent steps

What happens after the common intermediates have entered the microsomal particles is quite obscure. Two views are possible, which might be called the Parallel Path and the Alternative Path theories. In the first an intermediate is used to produce both protein and RNA at about the same time. In the second it is used to produce either protein, or RNA, but not both. If we knew the exact nature of the intermediates we could probably decide which of the two was more likely. At the moment there seems little reason to prefer one theory to the other.

The details of the polymerization step are also quite unknown. One tentative theory, of the Parallel Path type, suggests that the intermediates first polymerize to give an RNA molecule with amino acids attached. This process removes it from the template and it diffuses outside the microsomal particle. There the RNA folds to a new configuration, and the amino acids become polymerized to form a polypeptide chain, which folds up as it is made to produce the finished protein. The RNA, now free of amino acids, is then broken down to produce fresh intermediates. A great variety of theories along these lines can be constructed. I shall not discuss these further here, nor shall I describe the various speculations about the actual details of the chemical steps involved.

Two types of RNA

It is an essential feature of these ideas that there should be *at least two types of RNA in the cytoplasm.* The first, which we may call 'template RNA' is located inside the microsomal particles. It is probably synthesized in the nucleus (Goldstein & Plaut, 1955) under the direction of DNA, and carries the information for sequentialization. It is metabolically inert during protein synthesis, though naturally it may show turnover whenever microsomal particles are being synthesized (as in growing cells), or breaking down (as in certain starved cells).

The other postulated type of RNA, which we may call 'metabolic RNA', is probably synthesized (from common intermediates) in the microsomal particles, where its sequence is determined by base-pairing with the template RNA. Once outside the microsomal particles it becomes 'soluble RNA' and is constantly being broken down to form the common intermediates

with the amino acids. It is also possible that some of the soluble RNA may be synthesized in a random manner in the cytoplasm; perhaps, in bacteria, by the enzyme system of Grunberg-Manago & Ochoa (1955).

One might expect that there would also be metabolic RNA in the nucleus. The existence of these different kinds of RNA may well explain the rather conflicting data on RNA turnover.

The coding problem

So much for biochemical ideas. Can anything about protein synthesis be discovered by more abstract arguments? If, as we have assumed, the sequence of bases along the nucleic acid determines the sequence of amino acids of the protein being synthesized, it is not unreasonable to suppose that this inter-relationship is a simple one, and to invent abstract descriptions of it. This problem of how, in outline, the sequence of four bases 'codes' the sequence of the twenty amino acids is known as the coding problem. It is regarded as being independent of the biochemical steps involved, and deals only with the transfer of information.

This aspect of protein synthesis appeals mainly to those with a background in the more sophisticated sciences. Most biochemists, in spite of being rather fascinated by the problem, dislike arguments of this kind. It seems to them unfair to construct theories without adequate experimental facts. Cosmologists, on the other hand, appear to lack such inhibitions.

The first scheme of this kind was put forward by Gamow (1954). It was supposedly based on some features of the structure of DNA, but these are irrelevant. The essential features of Gamow's scheme were as follows:

(*a*) Three bases coded one amino acid.

(*b*) Adjacent triplets of bases overlapped. See Fig. 1.

(*c*) More than one triplet of bases stood for a particular amino acid (degeneracy).

In other words it was an overlapping degenerate triplet code. Such a code imposes severe restrictions on the amino acid sequences it can produce. It is quite easy to disprove Gamow's code from a study of known sequences—even the sequences of the insulin molecule are sufficient. However, there are a very large number of codes of this general type. It might be thought almost impossible to disprove them all without enumerating them, but this has recently been done by Brenner (1957), using a neat argument. He has shown that the reliable amino acid sequences already known are enough to make *all* codes of this type impossible.

Attempts have been made to discover whether there are any obvious restrictions on the allowed amino acid sequences, although the sequence data available are very meagre (see the review by Gamow, Rich & Yčas,

1955). So far none has been found, and the present feeling is that it may well be that none exists, and that any sequence whatsoever can be produced. This is very far from being established, however, and for all we know there may be quite severe restrictions on the neighbours of the rarer amino acids, such as tryptophan.

If there is indeed a relatively simple code, then one of the most important biological constants is what Watson and I have called 'the coding ratio' (Crick & Watson, 1956). If B consecutive bases are required to code A consecutive amino acids, the coding ratio is the number B/A, when B and A are large. Thus in Gamow's code its value is unity, since a string of 1000 bases, for example, could code 998 amino acids. (Notice that when the coding ratio is greater than unity stereochemical problems arise, since a polypeptide chain has a distance of only about $3\frac{1}{2}$ Å. between its residues, which is about the minimum distance between successive bases in nucleic acid. However, it has been pointed out by Brenner (personal communication), that this difficulty may not be serious if the polypeptide chain leaves the template as it is being synthesized.)

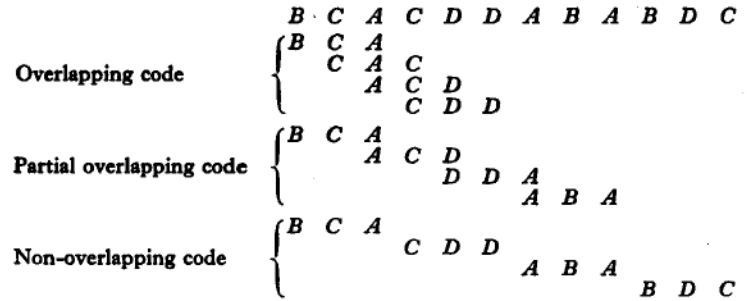


Fig. 1. The letters A, B, C, and D stand for the four bases of the four common nucleotides. The top row of letters represents an imaginary sequence of them. In the codes illustrated here each set of three letters represents an amino acid. The diagram shows how the first four amino acids of a sequence are coded in the three classes of codes.

If the code were of the non-overlapping type (see Fig. 1) one would still require a triplet of bases to code for each amino acid, since pairs of bases would only allow $4\times4=16$ permutations, though a possible but not very likely way round this has been suggested by Dounce, Morrison & Monty (1955). The use of triplets raises two difficulties. First, why are there not $4\times4\times4=64$ different amino acids? Second, how does one know which of the triplets to read (assuming that one doesn't start at an end)? For example, if the sequence of bases is ..., *ABA*, *CDB*, *BCA*, *ACC*, ..., where A, B, C and D represent the four bases, and where *ABA* is supposed to code one amino acid, *CDB* another one, and so on, how could one read it correctly if the commas were removed?

Very recently Griffith, Orgel and I have suggested an answer to both these difficulties which is of some interest because it *predicts* that there should be only twenty kinds of amino acid in protein (Crick, Griffith & Orgel, 1957). Gamow & Yčas (1955) had previously put forward a code with this property, known as the 'combination code' but the physical assumptions underlying their code lack plausibility. We assumed that some of the triplets (like *ABA* in the example above) correspond to an amino acid—make 'sense' as we would say—and some (such as *BAC* and *ACD*, etc., above) do not so correspond, or as we would say, make 'nonsense'.

We asked ourselves how many amino acids we could code if we allowed all possible sequences of amino acids, and yet never accidentally got 'sense' when reading the wrong triplets, that is those which included the imaginary commas. We proved that the upper limit is twenty, and moreover we could write down several codes which did in fact code twenty things. One such code of twenty triplets, written compactly is

$$A\ B\ \begin{matrix}A\\B\end{matrix} \qquad \begin{matrix}A\\B\end{matrix}\ C\ \begin{matrix}A\\B\\C\end{matrix} \qquad \begin{matrix}A\\B\\C\end{matrix}\ D\ \begin{matrix}A\\B\\C\\D\end{matrix}$$

where $A\ B\ \begin{matrix}A\\B\end{matrix}$ means that two of the allowed triplets are *ABA* and *ABB*, etc. The example given a little further back has been constructed using this code. You will see that *ABA*, *CDB*, *BCA* and *ACC* are among the allowed triplets, whereas the false overlapping ones in that example, such as *BAC*, *ACD* and *DBB*, etc., are not. The reader can easily satisfy himself that no sequence of these allowed triplets will ever give one of the allowed triplets in a false position. There are many possible mechanisms of protein synthesis for which this would be an advantage. One of them is described in our paper (Crick *et al.* 1957).

Thus we have deduced the magic number, twenty, in an entirely natural way from the magic number four. Nevertheless, I must confess that I find it impossible to form any considered judgment of this idea. It may be complete nonsense, or it may be the heart of the matter. Only time will show.

V. CONCLUSIONS

I hope I have been able to persuade you that protein synthesis is a central problem for the whole of biology, and that it is in all probability closely related to gene action. What are one's overall impressions of the present state of the subject? Two things strike me particularly. First, the existence of general ideas covering wide aspects of the problem. It is remarkable that one can formulate principles such as the Sequence Hypothesis and the Central Dogma, which explain many striking facts and yet for which proof is completely lacking. This gap between theory and experiment is a great stimulus to the imagination. Second, the extremely active state of the subject experimentally both on the genetical side and the biochemical side. At the moment new and significant results are being reported every few months, and there seems to be no sign of work coming to a standstill because experimental techniques are inadequate. For both these reasons I shall be surprised if the main features of protein synthesis are not discovered within the next ten years.

It is a pleasure to thank Dr Sydney Brenner, not only for many interesting discussions, but also for much help in redrafting this paper.

REFERENCES

Beadle, G. M. (1945). *Chem. Rev.* **37**, 15.

Benzer, S. (1957). In *The Chemical Basis of Heredity*. Ed. McElroy, W. D. & Glass, B. Baltimore: Johns Hopkins Press.

Berg, P. (1956). *J. Biol. Chem.* **222**, 1025.

Borsook, H. (1956). *Proceedings of the Third International Congress of Biochemistry, Brussels, 1955* (C. Liébecq, editor). New York: Academic Press.

Brachet, J. & Chantrenne, H. (1956). *Cold Spr. Harb. Symp. Quant. Biol.* **21**, 329.

Brenner, S. (1957). *Proc. Nat. Acad. Sci., Wash.* **43**, 687.

Brown, N. H., Sanger, F. & Kitai, R. (1955). *Biochem. J.* **60**, 556.

Cohen, G. N. & Cowie, D. B. (1957). *C.R. Acad. Sci., Paris*, **244**, 680.

Cole, R. D., Coote, J. & Work, T. S. (1957). *Nature, Lond.* **179**, 199.

Crick, F. H. C. (1957*a*). In *The Structure of Nucleic Acids and Their Role in Protein Synthesis*, p. 25. Cambridge University Press.

Crick, F. H. C. (1957*b*). In *Cellular Biology, Nucleic Acids and Viruses*, 1957. New York Academy of Sciences.

Crick, F. H. C., Griffith, J. S. & Orgel, L. E. (1957). *Proc. Nat. Acad. Sci., Wash.* **43**, 416.

Crick, F. H. C. & Watson, J. D. (1956). Ciba Foundation Symposium on *The Nature of Viruses*. London: Churchill.

Davie, E. W., Koningsberger, V. V. & Lipmann, F. (1956). *Arch. Biochem. Biophys.* **65**, 21.

DeMoss, J. A. & Novelli, G. D. (1956). *Biochim. Biophys. Acta*, **22**, 49.

Dounce, A. L. (1952). *Enzymologia*, **15**, 251.

Dounce, A. L. (1956). *J. Cell. Comp. Physiol.* **47**, suppl. 1, 103.

Dounce, A. L., Morrison, M. & Monty, K. J. (1955). *Nature, Lond.* **176**, 597.

Fraenkel-Conrat, H. (1956). *J. Amer. Chem. Soc.* **78**, 882.

Gale, E. F. (1953). *Advanc. Protein Chem.* **8**, 283.

Gale, E. F. (1956). *Proceedings of the Third International Congress of Biochemistry, Brussels, 1955* (C. Liébecq, editor). New York: Academic Press.

Gale, E. F. & Folkes, J. (1955). *Biochem. J.* **59**, 661, 675 and 730.

Gamow, G. (1954). *Nature, Lond.* **173**, 318.

Gamow, G., Rich, A. & Yčas, M. (1955). *Advanc. Biol. Med. Phys.* **4**. New York: Academic Press.

GAMOW, G. & YČAS, M. (1955). *Proc. Nat. Acad. Sci., Wash.* **41**, 1101.
GIERER, A. & SCHRAMM, G. (1956). *Z. Naturf.* **11***b*, 138; also *Nature, Lond.* **177**, 702.
GOLDSTEIN, L. & PLAUT, W. (1955). *Proc. Nat. Acad. Sci., Wash.* **41**, 874.
GROS, F. & GROS, F. (1956). *Biochim. Biophys. Acta*, **22**, 200.
GRUNBERG-MANAGO, M. & OCHOA, S. (1955). *J. Amer. Chem. Soc.* **77**, 3165.
HARRIS, J. I., SANGER, F. & NAUGHTON, M. A. (1956). *Arch. Biochem. Biophys.* **65**, 427.
HART, R. G. & SMITH, J. D. (1956). *Nature, Lond.* **178**, 739.
HERSHEY, A. D. (1956). *Advances in Virus Research.* IV. New York: Academic Press.
HOAGLAND, M. B. (1955). *Biochim. Biophys. Acta*, **16**, 288.
HOAGLAND, M. B., KELLER, E. B. & ZAMECNIK, P. C. (1956). *J. Biol. Chem.* **218**, 345.
HOAGLAND, M. B., ZAMECNIK, P. C. & STEPHENSON, M. L. (1957). *Biochim. Biophys. Acta*, **24**, 215.
HOKIN, L. E. & HOKIN, M. R. (1954). *Biochim. Biophys. Acta*, **13**, 401.
HULTIN, T. & BESKOW, G. (1956). *Exp. Cell Res.* **11**, 664.
INGRAM, V. M. (1956). *Nature, Lond.* **178**, 792.
INGRAM, V. M. (1957). *Nature, Lond.* **180**, 326.
KAMIN, H. & HANDLER, P. (1957). *Annu. Rev. Biochem.* **26**, 419.
KELLER, E. B. & ZAMECNIK, P. C. (1956). *J. Biol. Chem.* **221**, 45.
KONINGSBERGER, V. V., VAN DE VEN, A. M. & OVERBECK, J. TH. G. (1957). *Proc. K. Akad. Wet. Amst.* B, **60**, 141.
LITTLEFIELD, J. W. & KELLER, E. B. (1957). *J. Biol. Chem.* **224**, 13.
MARMUR, J. & HOTCHKISS, R. D. (1955). *J. Biol. Chem.* **214**, 383.
MIRSKY, A. E., OSAWA, S. & ALLFREY, V. G. (1956). *Cold Spr. Harb. Symp. Quant. Biol.* **21**, 49.
MUNIER, R. & COHEN, G. N. (1956). *Biochim. Biophys. Acta*, **21**, 592.
PALADE, G. E. (1955). *J. Biochem. Biophys. Cytol.* **1**, 1.
PALADE, G. E. (1956). *J. Biochem. Biophys. Cytol.* **2**, 85.
PALADE, G. E. & SIEKEVITZ, P. (1956). *J. Biochem. Biophys. Cytol.* **2**, 171.
PARDEE, A. B. (1954). *Proc. Nat. Acad. Sci., Wash.* **40**, 263.
PARDEE, A. B. & PRESTIDGE, L. S. (1956). *J. Bact.* **71**, 677.
PAULING, L., ITANO, H. A., SINGER, S. J. & WELLS, I. C. (1949). *Science*, **110**, 543.
PETERMANN, M. L. & HAMILTON, M. G. (1957). *J. Biol. Chem.* **224**, 725; also *Fed. Proc.* **16**, 232.
PETERMANN, M. L., MIZEN, N. A. & HAMILTON, M. G. (1952). *Cancer Res.* **12**, 373.
RABINOVITZ, M. & OLSON, M. E. (1956). *Exp. Cell Res.* **10**, 747.
RABINOVITZ, M. & OLSON, M. E. (1957). *Fed. Proc.* **16**, 235.
RICH, A. &. DAVIES, D. R. (1956). *J. Amer. Chem. Soc.* **78**, 3548.
SCHACHMAN, H. K., PARDEE, A. B. & STANIER, R. Y. (1953). *Arch. Biochem. Biophys.* **43**, 381.
SCHWEET, R. (1957). *Fed. Proc.* **16**, 244.
SIMKIN, J. L. & WORK, T. S. (1957*a*). *Biochem. J.* **65**, 307.
SIMKIN, J. L. & WORK, T. S. (1957*b*). *Nature, Lond.* **179**, 1214.
SIMPSON, M. V. & MCLEAN, J. R. (1955). *Biochim. Biophys. Acta*, **18**, 573.
SIMPSON, M. V., MCLEAN, J. R., COHN, G. I. & BRANDT, I. K. (1957). *Fed. Proc.* **16**, 249.
SPIEGELMAN, S. (1957). In *The Chemical Basis of Heredity.* Ed. McElroy, W. D. & Glass, B. Baltimore: Johns Hopkins Press.
STEINBERG, D. & MIHALYI, E. (1957). In *Annu. Rev. Biochem.* **26**, 373.
STEINBERG, D., VAUGHAN, M. & ANFINSEN, C. B. (1956). *Science*, **124**, 389.

STEPHENSON, M. L., THIMANN, K. V. & ZAMECNIK, P. C. (1956). *Arch. Biochem. Biophys.* **65**, 194.
SYNGE, R. (1957). In *The Origin of Life on the Earth.* U.S.S.R. Acad. of Sciences.
TS'O, P. O. B., BONNER, J. & VINOGRAD, J. (1956). *J. Biochem. Biophys. Cytol.* **2**, 451.
WAGNER, R. P. & MITCHELL, H. K. (1955). *Genetics and Metabolism.* New York: John Wiley and Sons.
YČAS, M. & BRAWERMAN, G. (1957). *Arch. Biochem. Biophys.* **68**, 118.
ZAMECNIK, P. C., KELLER, E. B., LITTLEFIELD, J. W., HOAGLAND, M. B. & LOFTFIELD, R. B. (1956). *J. Cell. Comp. Physiol.* **47**, suppl. 1, 81.