THE CHEMISTRY OF NUCLEIC ACIDS

T. R. SESHADRI

Department of Chemistry, University of Delhi

NUCLEIC acids are probably the most vitally important of natural polymers from the point of view of the living system. In several ways, their chemistry has been far more complex than that of other well-known polymers like polysaccharoses and proteins. That, in this difficult field, a stage of definite achievement has been reached at the present time is largely due to the great work of Prof. Sir Alexander Todd and his collaborators during the past fifteen years and this has been recognised by the recent award of the Nobel Prize to him.

Occurrence

The first isolation of a nucleic acid was made from pus cells and egg yolk in 1871 by Miescher. He found that they were usually associated with proteins and had to be liberated by means of alkali or enzymes. He noted also their marked acid property and insolubility in usual organic solvents. Many other sources of nucleic acids have subsequently been found, such as yeast, thymus gland and fish sperm. They are widely distributed in plants and animals and their existence in bacteria and viruses has considerable interest.

At one time it was thought that there was definite distinction between plant nucleic acids and animal nucleic acids and that the former was based on ribose and the latter on desoxyribose. The names Ribo Nucleic Acids (RNA) and Desoxyribo Nucleic Acids (DNA) were given to them. Later discoveries have shown that this is not so, and both types are found in plants, as well as in animals. But each is present in a different part of the cells; for example, the cytoplasm contains RNA, and the nucleus DNA; and each group has probably a different function to perform.

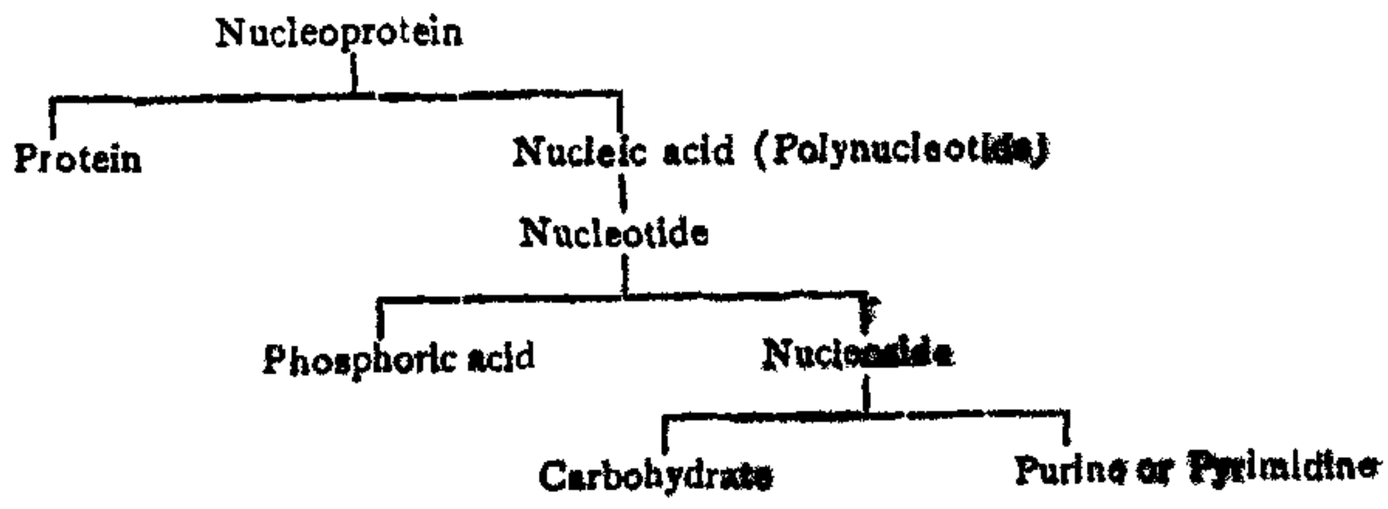
Isolation

As already mentioned, the nucleic acids occur in combination with proteins (nucleoproteins) intimately associated with the cells. The main stages in the isolation involve the destruction of the tissues, the separation of nucleic acids from proteins and finally precipitation and purification of the nucleic acids. For the first two stages sodium hydroxide in the cold or in the hot is employed and precipitation effected with hydrochloric acid. This works well for the preparation of yeast nucleic acid. The use of alkali has a great disadvantage, since the nucleic acid molecule is liable to break down during the treatment and hence alternative methods such as extraction with hot aqueous sodium chloride, heat denaturation and disruption by ultrasonic waves have been recommended. Their purification is extremely difficult and homogeneous preparations are rarely obtained.

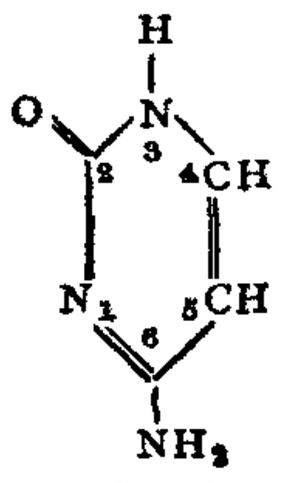
Analytical Studies

We owe a great deal to the pioneering and extensive work of Levene. As already mentioned, nucleic acids are large molecules made up of a number of smaller units called nucleotides. The conversion of nucleic acids into nucleotides is conveniently effected by mild hydrolysis, *e.g.*, heating with dilute ammonia at 115° C. for 1 hr. These nucleotides occur free to some extent and they have been shown to be made up of three components, a purine or pyrimidine base, a sugar, and phosphoric acid. When treated with sodium hydroxide phosphoric acid alone is cleaved leaving the other two parts together, in what is called nucleoside (glycoside). Further acid hydrolysis liberates the free purine or pyrimidine base and D-ribose in the case of RNA and D-desoxyribose in the case of DNA.

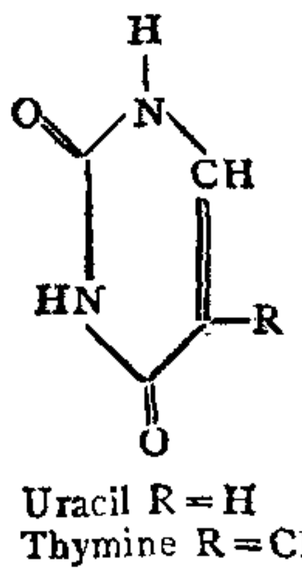
The basic part of different nucleotides varies. It may either belong to the purine group such as adenine and guanine or to the pyrimidine group such as cytosine, uracil and thymine. The constitution of the purines and pyrimidines



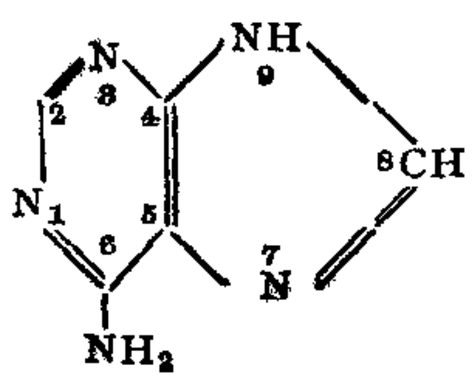
had been established earlier mainly by the work of Emil Fischer's school.



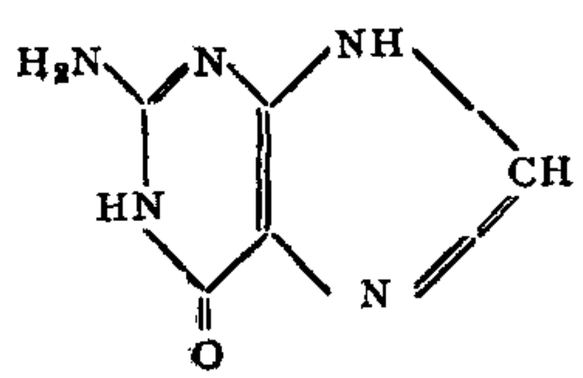
Cytosine



Uracil R = H
Thymine R = CH_3



Adenine



Guanine

D-ribose and D-desoxyribose were first identified by Levene and Jacobs (1908-11) as hydrolytic products of nucleic acids. This was the first discovery of these sugars in Nature. Later they have been found to be present in other sources too. It is interesting to note that D-ribose was known much earlier as a synthetic product, synthesised by Emil Fischer in 1901, in the course of his study of carbohydrates. He also synthesised desoxyribose in 1913. These are probably two of the most important carbohydrates occurring in Nature. In 1935, Levene and Tipson showed that these sugars are linked with the bases in the form of glycosides and the rate of hydrolysis of the nucleosides is like those of other glycosides. They also showed that in the combination they exist in the furanose ring form. This information was obtained by complete methylation and hydrolysis. The resulting trimethyl ribose yielded dimethyl tartaric acid on oxidation. Later on, Todd and his co-workers used periodic acid titration and showed that in ribonucleosides there are only two hydroxyl groups in the neighbouring positions thus confirming the furanose structure. This is a handy method of diagnosis of the ring structure present in natural as well as synthetic ribosides and has also been ingeniously used for determining the β-configuration of the glycosidic linkage.

Though the nature of the ultimate units composing the nucleic acids was known fairly early, there has been considerable difficulty in finding out the precise manner in which these are linked together to constitute the nucleosides, the nucleotides and eventually the nucleic acids. First there was the question regarding the position of linking of the purines and the pyrimidines with the sugars. Gulland and co-workers studied this problem by comparing the ultra-violet absorption spectra of the natural ribonucleosides with those of synthetic alkyl substituted bases and concluded that in purines it was the 9 position and in pyrimidines the 3 position. In view of the close resemblance between the ribose and desoxyribose derivatives the same positions have been considered to be occupied by the desoxyribose units also.

Synthetic Work

As already mentioned, the discovery of nucleic acids was made more than eighty years ago. However, it is only during the last decade that definite progress in the study of their structures has been possible. This long delay was due to the lack of precise knowledge of the structure and properties of the simple nucleosides and nucleotides, formed as products of hydrolysis. A marked difference exists between nucleic acids and other well-known high polymers occurring in Nature. In carbohydrates and proteins, the monomers are bifunctional simple entities like mono-saccharoses and amino acids. In the case of nucleic acids, the nucleotide monomers consist of three different parts and the inter-linkages were difficult to establish. Further, other difficulties existed, *e.g.*, the nucleotides which are polar substances are insoluble in common organic solvents and are very difficult to handle by the classical methods of organic chemistry. The intensive development of nucleotide chemistry had therefore to await the introduction of new experimental techniques. The important synthetic studies can be considered under three heads.

Structure and Synthesis of Nucleosides

The total synthesis of the four ribonucleosides was a great achievement and established their structure. For the synthesis of purine nucleosides a possible route was that of Emil Fischer in which the silver salt of the purine is made to react with an aceto-halogeno-sugar. This method required the unknown (at that time) acetohalogenoribofuranose and was not unambiguous as regards the location of the sugar residue in the product obtained. A new type of synthesis was therefore developed by Todd

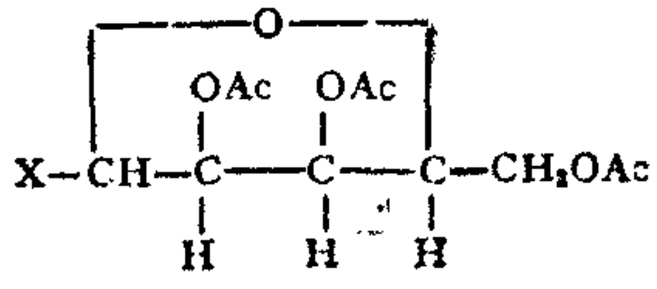
and his co-workers. It involved the preparation of a glycoside of a pyrimidine, with subsequent completion of the purine nucleus by building up the second ring. Though this method needed long and difficult exploration, it was successfully developed leading to the final synthesis of adenosine. The essential stages are outlined below :

Adenosine
I

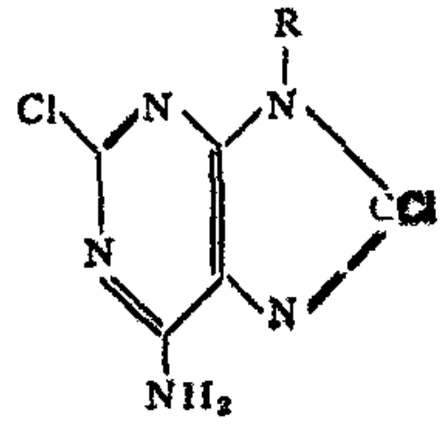
Adenosine (I) was also synthesised by an extension of Fischer's method. Acetochlororibofuranose (II) prepared from 5-trityl ribose was treated with silver 2 : 8-dichloroadenine (III, R = Ag) and the product deacetylated, giving 2 : 8-dichloro-9-β-D-ribofuranosidyl adenine (III, R = ribose unit). Removal of the chlorine atoms by hydrogenation yielded adenosine (I).

Partial reduction of dichloro compound (III, R = ribose unit) to 2-chloroadenosine followed by deamination with nitrous acid and replacement of the chloro group by amino, gave guanosine (IV, R = ribose).

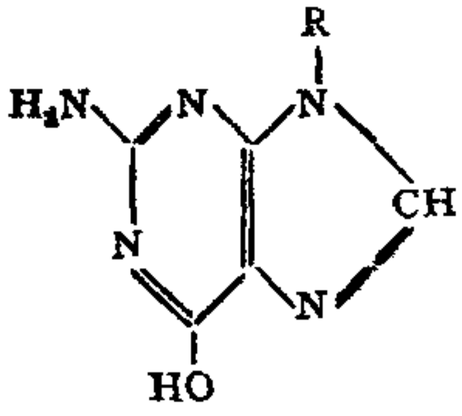
For the synthesis of cytidine (VI *b*), acetobromoribofuranose (II, X = Br) was treated with 2 : 6-diethoxypyrimidine (V *a*). This involves a reaction characteristic of heterocyclic compounds in which quaternary ammo-



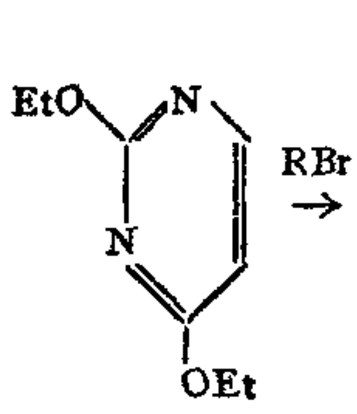
II X=Cl or Br



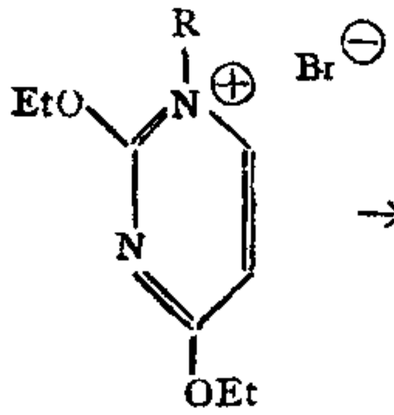
III



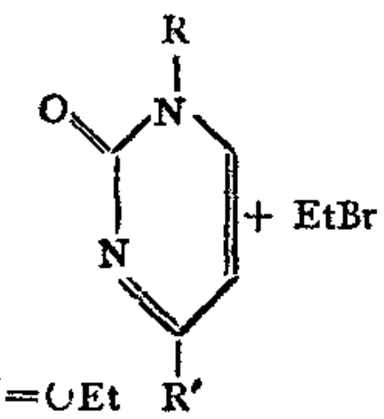
IV



V *a*



V *b*



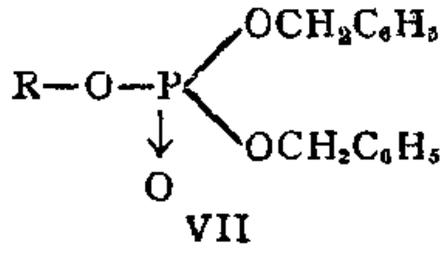
a, R′=OEt
b, R′=NH_2
c, R′=OH VI

nium salts (V *b*) undergo conversion into dihydroderivatives. The stages are indicated in the formulæ given above. Treatment of the product with ammonia gave cytidine (VI *b*). Uridine (VI *c*) could be obtained by deamination of cytidine with nitrous acid.

They are in each case β-D-ribofuranosides, the sugar residue being attached at N-9 in the purine and at N-3 in the pyrimidine units respectively. The β-form is the stablest and tends to be formed easily; this is attributed to the favourable influence of the basic nitrogen atom to which the sugar group is attached. Similar structures hold good for the desoxyribonucleosides also, though their synthesis has not so far been effected, largely due to the unavailability of desoxyribose.

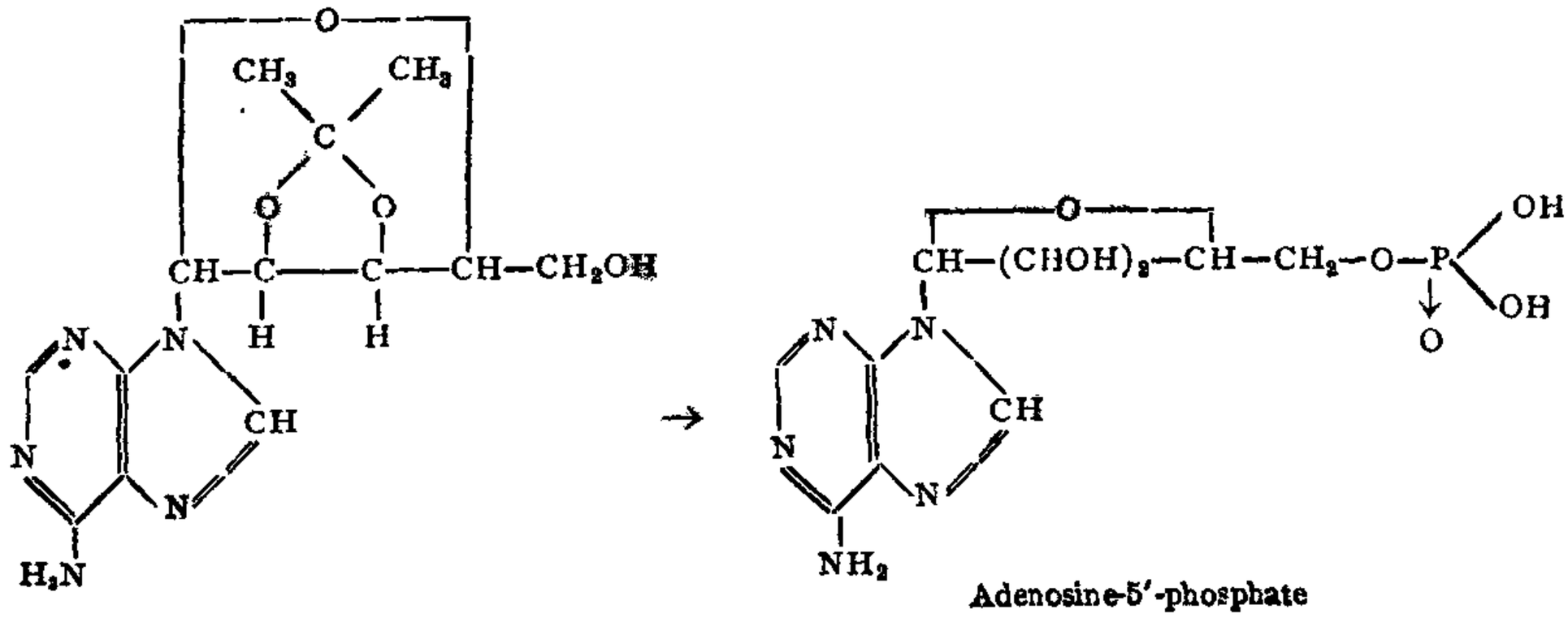
Phosphorylation of Nucleosides (Nucleotide Synthesis)

Thirty years ago Levene established by titration studies that the nucleic acids were made up of nucleoside units linked together by phosphate residues. He also demonstrated that ether and pyrophosphate linkages were absent. However, no further progress could be made because of the lack of detailed knowledge of nucleosides and the chemistry of the esters of phosphoric acid. Before 1949, it was generally held that ribonucleic acid (RNA) yielded on alkaline hydrolysis only four simple nucleotides (adenylic acid, guanylic acid, uridylic acid and cytidylic acid) believed to be 3′-phosphates. Using ion-exchange chromatography, Cohn showed during this year (1949) that the hydrolysates really contained 8 simple nucleotides made up of four pairs of isomers. By that time Todd and his co-workers had been devising methods for the unambiguous synthesis of mononucleotides of adenosine, guanosine, uridine and cytidine. Earlier methods of phosphorylation were unsatisfactory. An efficient method should proceed in good yields under mild conditions and no hydrolytic process should be involved which might damage sensitive glycosides. They found by extensive study that dibenzyl chlorophosphonate serves the purpose best. This substance reacts readily with alcohols at room temperature in the presence of tertiary bases yielding esters of type (VII). From these products the benzyl groups can be removed smoothly by catalytic hydrogenation.



VII

The 5′-phosphates were obtained by the phosphorylation of the 2′ : 3′-isopropylidene derivatives of the nucleosides followed by the removal of the protecting groups.



Adenosine-5′-phosphate

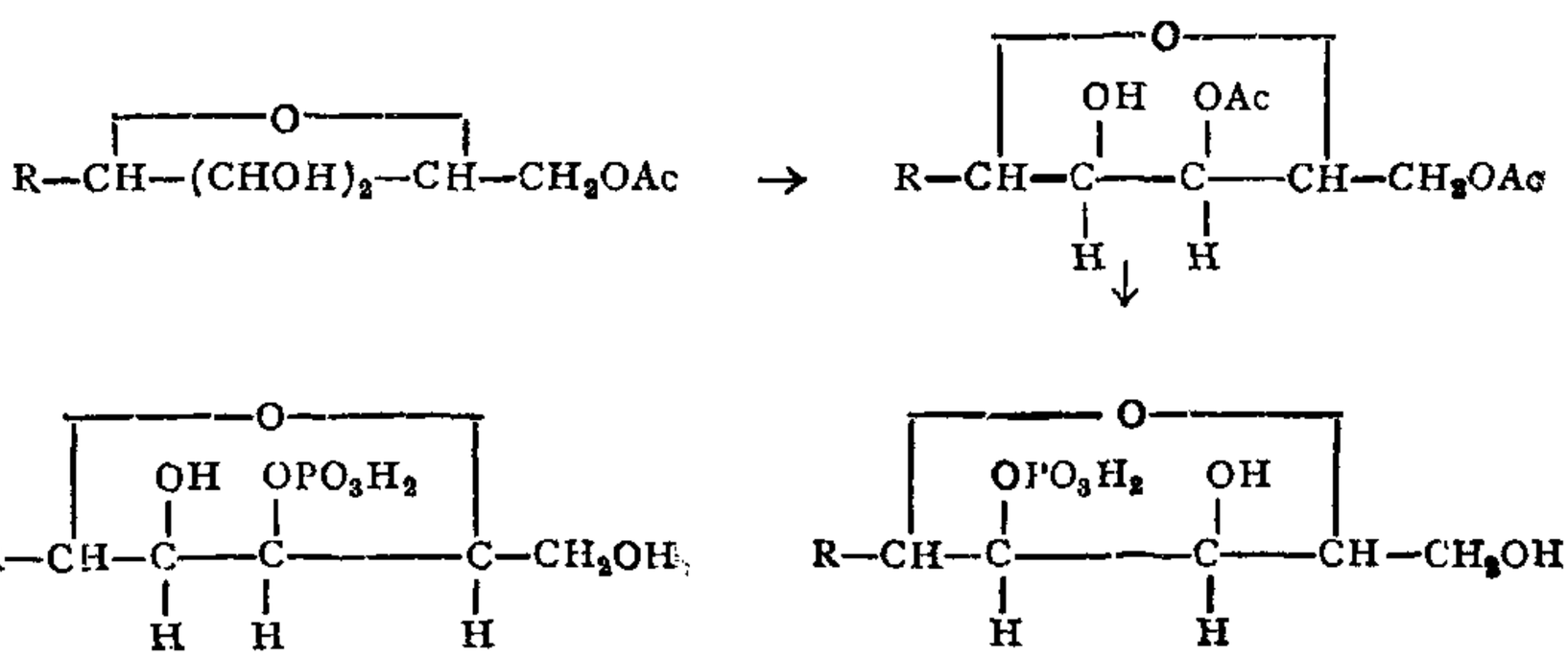
Subsequently Brown and Todd synthesised adenosine-2', and adenosine-3'-phosphates by phosphorylating 5'-trityladenosine and separating the mixture of products by ion-exchange chromatography. They were identical with the adenylic acids -*a* and -*b* of Cohn, but which was the 2' and which the 3' it was impossible to say with certainty. This was first indicated by the hydrolysis of the isomers with an acid ion-exchange resin. The former yielded a product which was identified as the 2'-phosphate and the latter the 3'-phosphate of ribose. Confirmation was provided by the partial acetylation of 5'-acetyl adenosine to 3' : 5'-diacetyl derivative, the phosphorylation of which followed by deacetylation yielded adenylic acid-*a* alone; it should therefore be 2'-phosphate of adenosine and adenylic acid-*b* should therefore be 3'-phosphate. This point has also been established by X-ray crystallographic methods.

phate considered to be an intermediate. The process of hydrolysis is indicated in the following scheme and the peculiar reaction is attributed to the presence of a *cis*-hydroxyl group in the concerned system.

The structural analogy between these nucleotide mono-esters and poly-nucleotides was recognised by Brown and Todd and they used it to explain the ready alkaline hydrolysis of RNA and the stability of DNA (which do not have the 2'-hydroxyl) under same conditions. They also advanced the general formulation of both types of nucleic acids as 3'-5'-linked polynucleotides.

Formation of Polynucleotides

The general structures at present accepted for DNA and RNA are indicated in formulæ (VIII) and (IX). In these representations the sugar part intimately connected with the inter-

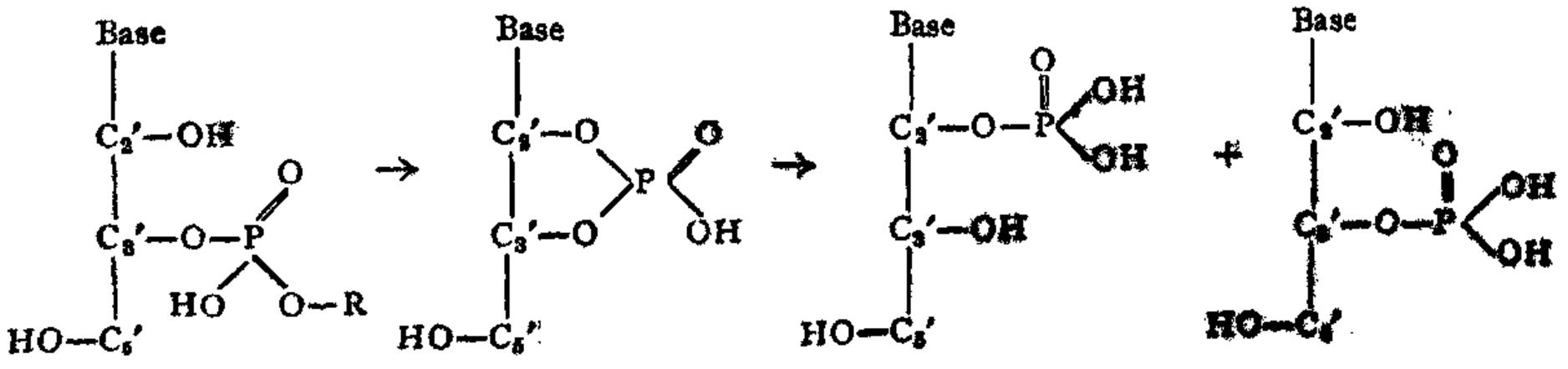


Adenosine-3' phosphate
(Adenylic acid-*b*)

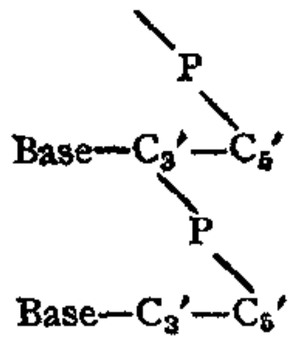
Adenosine-2'-phosphate
(Adenylic acid-*a*)

Explanation of the results of Cohn who obtained pairs of phosphates of each nucleoside was provided by the study of the migration of the phosphate groups. A significant observation was the lability of nucleotide mono-esters towards alkali. These undergo hydrolysis readily and yield mixtures of the unesterified 2' and 3'-phosphates along with a cyclic 2'-3'-phos-

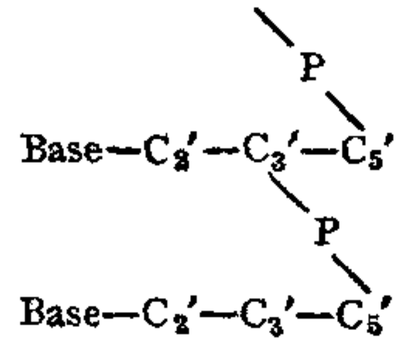
nucleotide linkage and bearing free hydroxyl groups alone is specified. In the case of DNA evidences are derived from the stability towards alkali and the course of enzymic and acid hydrolysis. In regard to RNA definite proofs have been more difficult to obtain but they have been provided by studying the action of specific enzymes on these nucleic acids.



R = Alkyl or nucleotide chain

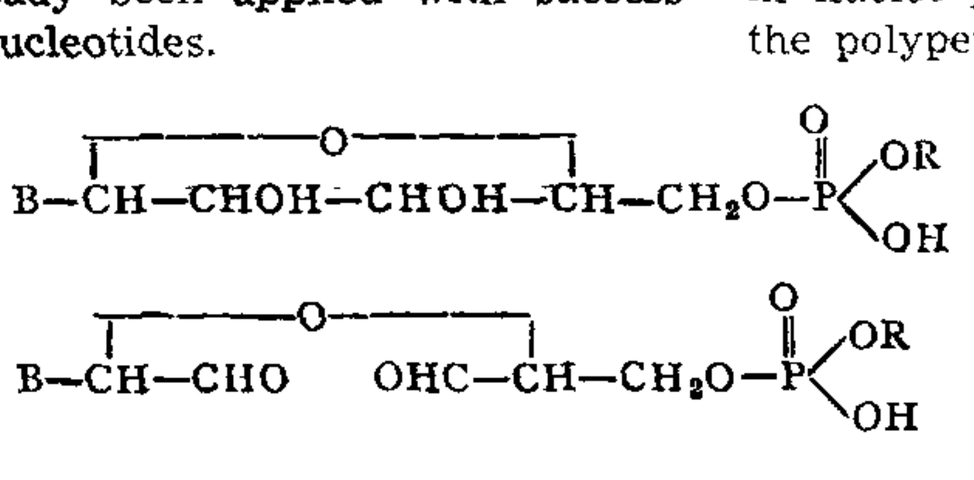


VIII DNA



IX RNA

The linear structures represented above for both types of nucleic acids are in accord with their properties and X-ray evidence. Individual nucleic acids may be expected to differ substantially in the sequence of the nucleotide residues and hence methods of sequence determination are necessary in order to have a complete picture. In the analogous case of polypeptides stepwise degradation from an end has been successfully employed for this purpose. But the problem is far more difficult in nucleic acids because of the greater instability of the phosphate linkages. An interesting method has been suggested by Todd and co-workers for the study of ribonucleic acid. This involves the use of periodate fission which will affect only one end group and subsequent elimination; the method has already been applied with success to small oligo-nucleotides.



R = Nucleotide chain

An alternative procedure for getting this information will be stepwise synthesis. This has also been shown to be possible. Michelson and Todd have synthesised a dithymidine-dinucleotide containing the 3′-5′-inter-nucleotidic linkage and have shown that the behaviour of the synthetic material towards enzymes is exactly the same as that of dinucleotidic fragments obtained by degrading desoxyribonucleic acids.

Another important recent discovery is that of Ochoa (1955) that an enzyme system can rapidly synthesise polynucleotides from nucleoside 5′-pyro-phosphates with the elimination of orthophosphate. This is akin to the technical production of fibre-forming poly esters and lends support to the view that the monomers involved in the formation of natural nucleic acids are nucleoside 5′-phosphates.

As the next step the conformation of nucleic acid macromolecules could be considered; this has received a great deal of attention. Salts of DNA can be obtained in crystalline fibre form and are quite suited for X-ray studies. They appear to have the form of a double helix. According to Watson and Crick (1953) they exist as a kind of double molecules consisting of two polynucleotide chains in the form of right-handed helices coiled round the same axis and held together by hydrogen bonds. The phosphate groups are on the outside and the purine and pyrimidine bases lie inside the helix at right angles to the long axis. An important feature is that the purine and pyrimidine bases are not at random but are hydrogen-bonded in specific pairs. A third coaxial chain is present in nucleo-proteins and seems to be formed by the polypeptide chain of the protein component.

This picture seems to offer an important clue to the way in which hereditary patterns are passed on during cell division and also how mutation can take place. DNA being characteristic components of chromosomes seem to play a vital part in heredity transmission. They have a key function in controlling the synthetic processes in the cell and this is obviously related to protein synthesis. This function seems to be carried out through the help of the RNA which occur largely in the cytoplasm. Thus the three, DNA, RNA and protein, seem to be closely related. Todd has suggested a mechanism of specific protein synthesis based in the Watson-Crick model and the known behaviour of mixed anhydrides. According to this the nucleic acid macromolecule should serve as a template to guide and facilitate the synthesis of specific proteins.